Media Attitude Detection via Framing Analysis with Events and their Relations

Jin Zhao and **Jingxuan Tu** and **Han Du** and **Nianwen Xue**
Department of Computer Science
Brandeis University
Waltham, Massachusetts, USA
{jinzhao,jxtu,handu,xuen}@brandeis.edu

Abstract

Framing is used to present some selective aspects of an issue and make them more salient, which aims to promote certain values, interpretations, or solutions (Entman, 1993). This study investigates the nuances of media framing on public perception and understanding by examining how events are presented within news articles. Unlike previous research that primarily focused on word choice as a framing device, this work explores the comprehensive narrative construction through events and their relations. Our method integrates event extraction, Cross-Document Event Coreference (CDEC), and causal relationship among events to extract framing devices employed by the media to assess their role in framing the narrative. We evaluate our approach with a media attitude detection task and show that the use of event mentions, event cluster descriptors, and their causal relations effectively captures the subtle nuances of framing, thereby providing deeper insights into the attitudes conveyed by news articles. The experimental results show the framing device models surpass the baseline models and offer a more detailed and explainable analysis of media framing effects. We make the source code and dataset publicly available.[1]

1 Introduction

Media framing significantly shapes public perception and opinion by highlighting or downplaying certain aspects of events (Entman, 1993). In news articles reporting Putin's 2024 election win worldwide, supportive articles emphasize the legitimacy of the election and present it as the reflection of the majority of Russian voters' will. These articles argue that Putin has brought economic growth and a strong international presence to Russia. In contrast, skeptical articles highlight incidents of electoral fraud and manipulation and mention deaths of Putin's political opponents and Ukrainian War. In this paper, we conduct framing analysis on highly contentious international news articles by strategically extracting and encoding event information to detect the media attitude reflected in the news articles towards the main event.

Framing analysis introduces rigorous and systematic methods for studying media content. This approach enriches research in communication, journalism, and social sciences by providing tools to dissect and interpret narratives within news articles (Goffman, 1974). Identifying and understanding different perspectives through framing analysis can encourage more balanced and fair reporting.

Previous work only focused on the extraction for frames that are more aligned with topics and the choice of words out of many framing devices. They answered "what is presented" (the issue) instead of "how it is presented" (framing) (Ali and Hassan, 2022). Narratives in media framing can be constructed by linking multiple events together (Goffman, 1974). This process involves creating a broader storyline or context that connects individual events to convey a more comprehensive theme. By analyzing how events are used to frame a topic, we can provide a deeper understanding of the attitude of a news article. This approach goes beyond equating frames with topics and text span based framing detection that are common in computational framing research (Ali and Hassan, 2022), offering deeper insights into narrative construction and media attitudes.

We employ a pipeline approach for detecting media attitudes through framing analysis using events and their relationships. Figure 1 illustrates the full pipeline for the media attitude detection task. The process begins with the collection of news articles on various topics. For each article, we extract all events mentioned from its main content. Then, we cluster coreferent events from all articles of the same topic and generate a shared event descrip-

[1] https://github.com/jinzhao3611/attitude-detection-with-framing

Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, pages 17197–17210
November 12-16, 2024 ©2024 Association for Computational Linguistics

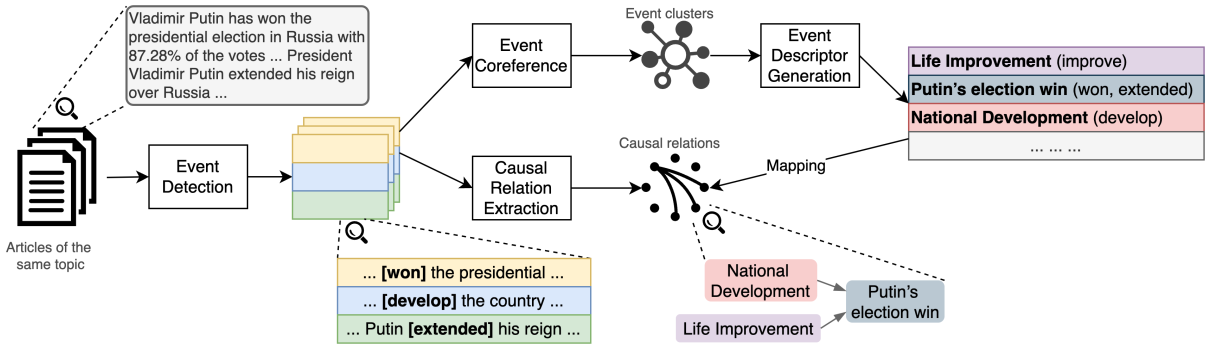

Figure 1: Media attitude detection pipeline.

tor for each cluster. This descriptor represents all event mentions within that cluster across the documents, ensuring consistency and clarity in event representation. Given the event mentions, we extract causal relations among the events for each article, mapping event mentions to their respective cluster descriptors.

We then evaluate the effectiveness of encoding coreferent event descriptors, encodings of event mentions, and causal relationships as framing devices. This is achieved by detecting the attitudes of the news articles by fine-tuning small language models and prompting Large Language Models (LLMs). Our results demonstrate that these framing devices are not only effective in capturing framing nuances but also provide explainable insights into how media frames the narratives. The experimental results indicate that all three methods of encoding framing devices outperform the baseline models. Specifically, in the case of GPT-4o, device encodings improve over the baseline encoding by ranging from 13 to 28 points. The best-performing models are the fine-tuned models incorporating framing devices. We summarize the major contributions of this paper are as follows:

Framing device concepulization We address three open questions in computational framing posed by Ali and Hassan (2022): First, we extend beyond the traditional focus on word choice as a framing device. Our computational methods capture events and their causal relations, facilitating a more nuanced exploration of frames. Second, we employ the cross-document event coreference technique to capture semantic relations within a single document and how these relations are informed by those in other documents. This approach provides a broader and more integrated understanding of frames across multiple documents. Third, we broaden the scope of framing devices to include the selection or omission of events, as well as the causal relations among events, in addition to the conventional focus on word choice. This allows for a more comprehensive analysis of how framing is achieved through different mechanisms.

New dataset for media attitude detection We provide a robust dataset annotated with media attitudes. It contains over 1,600 articles from three topics. This dataset serves as a resource for further research in computational framing analysis and related fields.

Better performing models with framing devices Our models significantly outperform the baselines, showing an improvement over the baseline encoding. They achieve higher overall accuracy, demonstrating that the three framing devices are effective for compressing information in news texts. Additionally, our input length is significantly shorter than that of the baseline models, resulting in reduced computing time and more efficiency. Moreover, our proposed encodings, namely the event descriptors, event mentions, and causal relationships enhance the explainability of the models' results.

The rest of the paper is organized as follows. In §2, we discuss related work on computational framing in terms of conceptualization and computational methods. §3 describes the media attitude detection task and our adoption of three framing devices in encoding. §4 describes the data collection and data annotation process. §5 describes event detection, CDEC, event descriptor generation and event causal relationship extraction we adopted in the pipeline. §6 describes the setup of the four media attitude detection models, presents the results and analysis. Finally we conclude in §7.

2 Related Work

The Policy Frames Codebook (Boydstun and Gross, 2013) is a codebook that was used by Card et al. (2015) to develop the Media Frames Corpus. They established 15 framing dimension categories that tend to resemble topics more than actual frames. The Gun Violence Frame Corpus (Liu et al., 2019a) annotated news headlines but did not fully elaborate on the framing aspects defined according to Entman (1993). Instead, it used keyword clues to categorize the framing dimensions.

Computational methods have been increasingly adopted in framing analysis. DiMaggio et al. (2013) employed LDA topic modeling to explore frames, but this approach is more about theme, lacking the ability to capture framing nuances such as causal interpretations. (Nguyen et al., 2015) introduced "Supervised Hierarchical Latent Dirichlet Allocation (SHLDA)," which generates a hierarchy of topics with first-level nodes as agendas and second-level nodes as frames. This subtopic framing approach does not align with framing conceptualizations (Entman, 1993). Additionally, Burscher et al. (2016) applied k-means clustering on a dataset, using word frequencies as frame features in their cluster analyses. However, it fails to capture how events are interconnected through cause-and-effect relationships, which are essential for understanding the underlying narrative and motivations presented in the text.

Furthermore, Mendelsohn et al. (2021) employed fine-tuned RoBERTa models to create classifiers for identifying frames in news articles. Although their approach proved effective, the conceptualization of framing and the categorization of their dataset could be improved by incorporating more detailed and nuanced framing aspects. Typically, computational methods tend to only partially capture the complexity of framing, often relying solely on single framing devices such as word usage. This approach overlooks the multiple dimensions and layers of framing analysis.

In a related study, Liu et al. (2023) extracted events and assessed their impact on the stance of news articles by identifying events indicative of ideology. However, the authors noted they did not further explore how these events relate to each other to convey ideology. For instance, their extraction method could detect a right-leaning event alongside several left-leaning events that oppose it, without explaining the interaction between these events.

In narrative framing research, the perception to narratives is influenced by event temporal structure and shared event scripts (Baldassano et al., 2018). Furthermore, narratives act as tools for social construction, shaping public opinion and policy by framing events in particular ways (Johnson-Cartee, 2004).

3 Media Attitude Detection

3.1 Task Definition

Given a set of news articles with the same target event (topic), we formalize the task as one of classifying the attitude of each article towards the target event into one of three categories: supportive, skeptical, or neutral. Let $\mathcal{D} = \{d_1, d_2, \ldots, d_n\}$ be a set of n news articles. We encode each article d_i in the following encoding methods $\mathcal{E}$:

Event Mention Coreference Clusters (X^c) Let $E_{i1}, E_{i2}, \ldots, E_{ik}$ be the events extracted from d_i. Let $C(E_{i1}), C(E_{i2}), \ldots, C(E_{ik})$ be the event coreference clusters for the extracted events, each of which is associated with a event discriptor. The concatenated encoding of article d_i is represented as:

$$X_i^c = [C(E_{i1}) \| C(E_{i2}) \| \cdots \| C(E_{ik})]$$

Event Mentions (X^m) Let $E_{i1}, E_{i2}, \ldots, E_{ik}$ be the events extracted from d_i. The concatenated encoding of article d_i is represented as:

$$X_i^m = [E_{i1} \| E_{i2} \| \cdots \| E_{ik}]$$

where $\|$ denotes the concatenation operator.

Event Causal Relationships (X^r) Let $\{(E_{ca_{ij}}, E_{ef_{ij}})\}_{j=1}^{k}$ be the set of k cause-effect event pairs extracted from d_i, where $E_{ca_{ij}}$ denotes the cause event and $E_{ef_{ij}}$ denotes the effect event in the j-th pair. The concatenated encoding of article d_i is represented as:

$$X_i = [(C(E_{ca_{i1}}) \to C(E_{ef_{i1}})) \| \cdots \\ \| (C(E_{ca_{ik}}) \to C(E_{ef_{ik}}))]$$

Feature Matrix (X) For a chosen encoding method $\mathcal{E}_k$, construct the feature matrix $X^k \in R^{n \times m_k}$ where each row X_i^k corresponds to the encoding of article d_i

Attitude (y) : Each article d_i has an associated attitude towards the main event $y_i \in \{\text{supportive}, \text{skeptical}, \text{neutral}\}$. The classification task is to predict the attitude y_i of each article d_i using its encoded representation X_i^k.

3.2 Conceptualization: Represent Framing Devices with Encodings

The frame manifests itself in media content through various framing devices. Framing devices are the various tools and techniques used to shape how information is presented and perceived (Van Gorp, 2006). These devices help construct a particular interpretation or understanding of events, issues, or individuals. We set up our experiments to investigate how effective the following three framing devices are in shaping the attitudes of the news articles.

Device 1: Selection and Omission of Events Selecting which events to report and which to ignore can significantly influence public perception (Entman, 1993). Highlighting certain events makes them seem more important, while omitting others downplays their significance and shapes the overall story (e.g. presence vs. absence of Putin's opponents deaths). We extract event mentions from the article and generate summaries for each event coreference cluster, thereby capturing the selection and omission of events.

Device 2: Linguistic Information The specific language used in reporting can significantly shape audience perception (Entman, 1993). For instance, labeling a protest as a "riot" versus a "demonstration" leads to different interpretations. Metaphors and analogies frame complex issues in relatable terms, such as describing an economic downturn as a "storm," thereby guiding audience conceptualization. Euphemisms soften the perception of skeptical events (e.g., "collateral damage" instead of "civilian casualties"), while dysphemisms make events seem worse (e.g., "regime" versus "government"). We incorporate the original event trigger words and associated arguments of event triggers in the article to include this linguistic information.

Device 3: Cause and Effect Establishing causal links between events shapes the interpretation of their causes and consequences, allowing media to influence audience understanding (Gamson and Modigliani, 1989). By suggesting that one event caused another, the media can create a narrative that influences how audiences understand the situation (e.g., "economic development" → "putin election win" vs. "surpression of opponents" → "putin election win"). We use extracted causal event pairs, represented by their cluster descriptors, to capture causal relations between events in the article.

4 Dataset

We describe the data collection and annotation process for creating the media attitude detection dataset.

4.1 Data collection

While extensive research has focused on framing, it has predominantly centered on U.S. politics, with limited application to non-English languages, other social contexts, or information warfare (Park et al., 2022). To address this gap, we collect news articles on three highly contentious international events: *Putin's election win in March 2024*, *Israel's Al-Shifa Hospital raid in November 2023*, and *the Hong Kong July 1 protest in 2019*. There are 495, 643, 471 articles respectively. We source these articles from seven news aggregators and 98 news sources. These sources include a variety of Western media outlets such as CNN and BBC, state-backed Russian media like Sputnik, Chinese media such as Xinhua News, and Arabic and Israeli sources like Al Jazeera and The Times of Israel.

The news articles were either fully scraped or, for longer articles, segmented under section titles. The articles were in English, Chinese, and Russian, with non-English articles translated into English using the Google Translate API[2]. To prevent redundancy, we remove documents with 60% sentence overlap with other documents. We also exclude the following types of news articles from our data: editorials, blog posts, critique analyses, self-media articles, opinion columns, articles with an episodic frame focusing on a narrow individual example, and press conference transcripts. In addition, articles that focus on moral judgments and lack event mentions are excluded. Any meta-information present in the articles are also removed.

4.2 Annotation

All news articles in our dataset were annotated by two independent annotators, who are graduate students from computational linguistics department of a US-based university. Both annotators were selected based on their familiarity with media attitude and framing analysis. They were provided with extensive training prior to the annotation task.

Training Process The annotators received an overview of the annotation task, including an explanation of the media attitude task and the event top-

[2] https://github.com/vitalets/google-translate-api

ics they would be working with. Annotators were provided with annotation guidelines, which outlined the annotation categories (supportive, skeptical, neutral), how to interpret different media attitudes through framing clues. Ambiguous cases were discussed in detail to help the annotators resolve edge cases. Following the guidelines, the annotators annotated 50 trial examples using a spreadsheet. The trial examples covered various topics and edge cases to expose annotators to different complexities within the data. After completing the trial annotations, they received feedback to improve their understanding. Once the annotators achieved at least 90% accuracy on assessment examples, they were deemed ready for the main annotation task.

Annotation Process Each article in the dataset was annotated by the same two annotators across all topics. Annotators were tasked with classifying each article into one of three predefined categories: supportive, skeptical, or neutral. These categories corresponded to the article's stance on the main event described in the topic. For example, an article discussing Putin's election victory could be classified as supportive if it omitted allegations of electoral malfeasance, or skeptical if it highlighted such issues. We provide annotation guidelines of an example topic to the annotators. This example illustrates how the annotators approached the task and what factors they considered when classifying articles. After completing their independent annotations, the annotators discuss to resolve discrepancies through an adjudication process.

Inter-Annotator Agreement We calculated the Cohen's kappa coefficient, which yielded a score of 0.92. Although this is a high agreement score, except our well-prepared annotations, several factors contributed to this result: The nature of the selected international topics, which involved clear and pronounced attitudes (e.g., support for or opposition to specific events), likely facilitated higher agreement between annotators. Contrary to initial expectations, many articles displayed strong stances, making classification relatively straightforward. For instance, articles skeptical of the Al-Shifa raid consistently omitted key military justifications mentioned in supportive articles. Similarly, articles supportive of the Hong Kong protests rarely attributed the unrest to foreign influence, as skeptical articles often did.

Data Overview As shown in Table 1, the dataset consists of over 470 articles per topic, with the Al-Shifa topic containing the most articles. However, these articles tend to be shorter in length compared to other topics (232 vs. 297 words per article). In addition to the number of articles per topic, we report the number of extracted events, event clusters, and causal relations in the dataset. More details on the extraction procedure are provided in the following section.

Counts	Putin	Al-Shifa	Hong Kong
Articles	495	643	471
Avg. tokens	314	232	297
Clusters	321	310	450
Avg. events	7	8	8
Avg. relations	7	9	9

Table 1: Statistics of the dataset for the media attitude task. The number of events and causal relations are reported after the filter.

5 Data Preparation

To evaluate the media attitude detection via framing analysis, we prepare the dataset by automatically creating encodings of different framing devices from extracted events and relations.

5.1 Event Detection

We adopt the event detection model from Yao et al. (2021) to extract event trigger words and their modality information. Event modality indicates the certainty of the speaker or author regarding whether the event occurred. We filter out events with trigger words from the list of aspectual verbs (e.g., *begin to*), copular verbs (e.g., *become*), mental action verbs (e.g., *understand*), modal-like verbs (e.g., *need to*), and reporting verbs (e.g., *said*), as these provide limited information on media attitude. To extract the contextual information of each event, we set each event trigger as the predicate, and adopt the SRL parser[3] to extract its arguments. Each event is formatted as `[ARG-0] [trigger] [ARG-1] [ARG-2] [ARG-3]` for the Device 2 encoding.

5.2 Cross Document Event Coreference

We adopt the pairwise cross-encoder recently proposed by Yu et al. (2022) for the CDEC task. It first generates the score for each pair of event mentions, and then utilizes agglomerative clustering to form

[3] https://huggingface.co/cu-kairos/propbank_srl_seq2seq_t5_large

coreference event clusters (Yu et al., 2022; Held et al., 2021; Caciularu et al., 2021; Barhom et al., 2019). This approach effectively balances performance and cost. For example, while Caciularu et al. (2021) implemented a larger context window for cross-encoding that might improve CDEC performance, it proved costly to train.

Since the most significant events carry substantial framing information to determine an article's attitude, we prioritized precision over recall. In a collection of documents on a single topic, pairing up thousands of events to compute tens of millions of pairs would be impractical. Therefore, we only cluster event pairs from the same group where the event trigger is the same and the count of the events is greater than five, assuming that the frequent pairs with the same trigger are more likely to corefer. We also include cross-trigger pairs from less frequent groups whose triggers are connected in the AIDA ontology (Tracey et al., 2022) or are concrete synonyms in VerbNet (Schuler, 2006) and the concrete word list (Brysbaert et al., 2014).

5.3 Event Descriptor Generation

Event coreference cluster descriptors are short summaries that describe and distinguish different clusters of event mentions that refer to the same underlying event across a corpus of documents. These descriptors help in identifying and grouping event mentions that are semantically or contextually equivalent, even if they are expressed differently in the text. Consider the example in (1). The cluster contains the following two coreference event mentions from two news article respectively.

(1) **Sentence 1**
Russian Central Election Commission announces election results,...
Sentence 2
... election authorities declaring Putin the winner.
Cluster descriptor
Official announcement of Putin's election win

Inspired by the Dense Paraphrasing method proposed by Tu et al. (2023) and Rim et al. (2023), we apply GPT-4o (OpenAI, 2023) to write one descriptor for each event coreference cluster in each topic. We formulate the prompt by instructing the model to generate a concise and informative summary based on the context from up to five sentences that contain the event mentions from the same cluster. The generated event descriptors are used for Device 1 encoding.

5.4 Event Causal Relation Extraction

Recent work shows that prompt-based models can achieve competitive performance in determining pairwise causal relationships with high accuracy (Shen et al., 2022; Liu et al., 2023; Kıcıman et al., 2023). Similar to the descriptor generation, we adopt GPT-4o to extract event causal relations from the the articles. We formulate the prompt by instructing the model to identify the causal relations from the article with the event mentions marked up. In each identified causal relation, we replace the event mention with its cluster descriptor, and use it for Device 3 encoding.

6 Experiments

We evaluate our proposed framing device encodings under different model settings. We experiment on the media attitude detection task with model inputs encoded from different framing devices, and compare them to the baseline setting with raw texts as the input. For each topic, we randomly select 70% articles from the dataset for training and the remaining 30% for testing. We use accuracy as the primary evaluation metric.

6.1 Model Setup

Fine-tuning models We evaluate our methods on media attitude detection on two task configurations, text classification and text generation. We adopt the pretrained RoBERTa$_{\text{BASE}}$ (Liu et al., 2019b) for the classification configuration. For each article, the model input is encoded by concatenating all the relevant components from each framing device. We adopt T5$_{\text{BASE}}$ (Raffel et al., 2020) as the generation model. As is the common practice, we recast the media attitude detection as a Question Answering (QA) task. We prepare the model input by appending the task-specific prefix to the question and context text: "`question: {What is the article's stance?} context: {device_str}`". The `device_str` is the concatenation of all the relevant components from each framing device. We fine-tune each model on the training set and evaluate on the testing set.

Prompting models We also compare our methods using LLMs under a zero-shot prompt learning setting. We evaluate two types of models, (1) FlanT5 (Chung et al., 2022), an open-source encoder-decoder model; (2) GPT (Brown et al., 2020; OpenAI, 2023), a proprietary autoregressive decoder model. We experiment with the XL version

Topic	Method	Fine-tuning		Prompting	
		RoBERTa$_{BASE}$	T5$_{BASE}$	FlanT5$_{XL}$	GPT-4o
Putin Election Win	Baseline	75.00	77.70	56.77	59.46
	Device 1	**82.07**	**83.45**	**70.69**	**81.38**
	Device 2	77.24	76.56	62.66	72.41
	Device 3	80.69	79.79	65.07	75.86
Al-Shifa Hospital Raid	Baseline	81.87	73.06	40.41	52.33
	Device 1	80.89	**75.56**	**73.89**	**80.00**
	Device 2	81.63	74.61	70.73	76.36
	Device 3	**82.25**	71.54	68.82	78.44
Hong Kong Protest	Baseline	**97.18**	**96.49**	53.50	52.52
	Device 1	93.02	91.80	**65.45**	**78.17**
	Device 2	93.71	87.43	60.45	73.54
	Device 3	94.41	89.97	63.32	77.48

Table 2: Evaluation results on the attitude detection task for each topic. We compare the baseline with inputs encoded from different devices. Accuracy from each model setting is reported.

Topic	Baseline	Device 1	Device 2	Device 3
Putin Election Win	314	40 (87%)	46 (85%)	110 (65%)
Al-Shifa Hospital Raid	232	54 (78%)	63 (68%)	132 (43%)
Hong Kong Protest	297	38 (87%)	70 (76%)	120 (59%)

Table 3: Average input token counts and for each framing device. The compression rate over the baseline input length is also reported.

of FlanT5 (FlanT5$_{XL}$), and use the OpenAI API with version `gpt-4o`. We formulate the prompt by instructing the model to predict the attitude based on the framing device information from each article. Refer to Appendix A.2 for all the prompts we use in our experiments.

6.2 Results

Evaluation on fine-tuned models We show the model results in Table 2. The model input for the baseline is the raw text of each article. Across the three topics, RoBERTa$_{BASE}$ with baseline input performs the best on *Hong Kong Protest*, followed by *Al-Shifa Hospital Raid* and *Putin Election Win*, suggesting the variety of the topics and different levels of difficulty for the fine-tuned models to learn patterns of the data from each topic. T5$_{BASE}$, in spite of being a larger model, does not show superior results than RoBERTa$_{BASE}$ overall, suggesting that the generative model may not be fit for the media attitude detection task with preset labels under the fine-tuning setting. Inputs encoded from framing devices achieve mixed yet comparable results with the baseline. RoBERTa$_{BASE}$ with Device 1 outperforms the baseline by 7 points on the *Putin Election Win*, while underperforming the baseline by 4 points. Overall, the results exhibit that training with framing devices is able to achieve competitive results on certain news topics, and improves the training efficiency by having much shorter model inputs. We discuss this in greater detail in Section 6.3.

Evaluation on LLMs Table 2 also shows the experimental results from LLMs. Compared with the fine-tuned models, LLMs struggle with the task when only having the baseline inputs. Both FlanT5$_{XL}$ and GPT-4o perform much worse than the supervised models by at least 20 points. This result is expected because the task is domain-specific and no training or in-context examples are provided to the LLMs. However, we observe that the framing devices are able to significantly help LLMs better understand the task and article contexts by pinpointing the essential information for detecting media attitude. For example, GPT-4o with Device 1 improves the baseline by 22 points on the *Putin Election Win*. FlanT5$_{XL}$ also shows improvements over the baseline, but generally performs worse than GPT-4o. This is expected because GPT-4o is a larger and more powerful model. Overall, we show the utility of framing devices in media attitude detection under zero-shot learning scenarios. LLMs with framing device information can substantial outperform the baseline and achieve comparable results with supervised models.

6.3 Analysis

Data compression from device encodings We present the length of the input encoded from each framing device in Table 3, and illustrate the amount of input text reduction for each framing device. On average, the baseline input of each article has 280 tokens. By using device encodings, the average length of the input has been reduced by at least 43%. While the input is shortened, the performance of the fine-tuned models with framing devices is not compromised. For example, the average score change from all devices with RoBERTa$_{BASE}$ is +5%, +0.4%, -3.5% for the topics Putin Election Win, Al-Shifa Hospital Raid and Hong Kong Protest.

Qualitative Analysis We present predictions from LLMs with different types of inputs to analyze the limitations of the framing devices for the attitude detection task. Table 4 illustrates three common errors in the test data. CDEC errors affect all three devices. In the first example, CDEC incorrectly clustered "Israel destroy facilities" and "Hamas destroy facilities". The cluster descriptor choose a general term "soldiers" to represent both.

Model Input	Device	Label	Error type
Soldiers destroy hospital facility. ...	1	supportive / skeptical	CDEC error
Israel Defense Forces seizes weapons. ...	2	skeptical / supportive	SRL error
Implementation of extradition bill → restoration of order. ...	3	skeptical / supportive	Causal error

Table 4: Examples of common error types in the test set. We compare the predicted labels from GPT-4o with gold labels.

If Hamas were destroying facilities, this would generate a supportive attitude for Israel; however, the original article indicates that it is Israel destroying facilities. In the second example from Device 2, the location argument is crucial but missing. Finding weapons in Al-Shifa Hospital can justify Israel's raid, making this event a strong indicator of support for the raid. Without the location, it is a neutral event. The original article of the third example argues that the implementation of the extradition bill caused the protest and the violence, thus suggesting that restoring order is necessary. This causal relation is inferred incorrectly, and lead to incorrect classification of the models. These examples highlight how specific errors can affect the interpretation and classification of news articles, emphasizing the need for precision in event clustering, semantic role labeling, and causal relation extraction to accurately capture the intended attitude.

Framing Devices Comparison We demonstrate how the framing devices can help the model understand the article's attitude in Table 5. Compared to the baseline, Device 1 input compresses the main event from the baseline input. Descriptors are summaries of multiple event mentions in an event coreference cluster, which can encompass a range of attitudes. Therefore, many descriptors are in a neutral tone. Device 1 summarizes the murder as "Navalny's death". While some information is lost, GPT-4o still manages to capture the purpose the news article mentions Navalny's death is to raise doubt in Putin's election win. In comparison, Device 2 uses the original event trigger and arguments, retaining the skeptical connotation of the original words, as "murder" implies that Putin might be involved. The specified event mention eliminates much of the noise in the article and helps GPT-4o better understand the article's attitude. Device 3 input, compared to the previous two, shows the cause and effect of events. In this example, Device 3 explicitly specifies the consequence of Navalny's death, which is Putin's election win. This causal relation directly suggests that the election is rigged, hence indicating a skeptical attitude towards Putin's win. Our error analysis shows that all three devices help GPT-4o find the most useful information from noisy news articles, thereby improving its ability to understand and classify the article's attitude.

Learning Patterns or Knowledge While the framing devices are able to significantly improve the performance of theLLMs on the task, they are not as effective on the fine-tuned models. With the assumption that supervised models are more about learning matching text patterns instead of the encoded knowledge from the devices in our task, we analyze the text similarity between the train and test splits with different input. Specifically, we get the count distribution of the top 200 frequent tokens[4] from the train and test splits, and compute the Jensen–Shannon divergence (JSD) to quantify the distance between the distributions (Dagan et al., 1997). Since all three devices are encoded based on the events, we also compute JSD between the distribution of the top 200 frequent event triggers. As shown in Table 6, the JSD between the token distributions is much smaller than that between the event distributions, indicating that it is easier for the supervised models to learn the matching text patterns from the baseline input than the device input. LLMs with prompts, however, exhibit the capability to infer and understand the knowledge from the framing devices in our task, because they have no direct access to the in-domain examples for training.

6.4 Model Details

We use OpenAI API to run GPT models. The pricing for the models used in the paper is on the OpenAI website.[5]. We fine-tune RoBERTa$_{\text{BASE}}$ and T5$_{\text{BASE}}$ with different device configurations on a single 40GB Telsa V100 GPU. We use the hyperparameter setting with epoch 6, batch size 16, input length 512 and learning rate 5e-2. All the other hyperparameters remain default. The training time

[4]After removing the stop words and punctuation.
[5]https://openai.com/api/pricing/

Model Input	Device	GPT-4o Label
Navalny is murdered on February 16 in the Arctic prison where he was serving a 19-year sentence, Russia's prison service said...	Baseline	neutral
Navalny's death, ...	1	skeptical
Navalny is murdered in the Arctic prison ...	2	skeptical
Navalny's death → Putin's election win,...	3	skeptical

Table 5: Examples of GPT-4o with model input of baseline and different framing devices. We compare the predicted labels with gold labels.

Topic	Baseline (Tokens)	Devices (Events)
Putin Election Win	8.4	13.8
Al-Shifa Hospital Raid	7.8	14.8
Hong Kong Protest	8.7	15.2

Table 6: JSD (multiplied by 100) between the token/mention frequency distributions from the train and test split from each topic.

is less than 10 minutes for each model run. We use FlanT5$_{\text{XL}}$ for inference only, and run it on a single 40GB Telsa V100 GPU.

7 Conclusion

Framing analysis helps uncover the underlying attitude that news outlets may have. By analyzing how events are framed, we can better understand the intentions and viewpoints of the media sources, providing insight into how information is presented to the public. Our methods extend beyond words by conceptualizing framing devices and integrating event extraction, CDEC, and causal relationship extraction. Our media attitude detection through framing analysis demonstrates that all three framing devices effectively captures the subtle nuances of framing, providing deeper insights into the attitudes conveyed by news articles. Our results indicate that training with framing devices achieves competitive results on some topics and improves training efficiency by utilizing much shorter and more explainable model inputs. LLMs incorporating framing device information significantly outperform the baseline and achieve results comparable to those of supervised models. We hope that our analysis and improvement on the media attitude detection task can facilitate further developments by future researchers.

8 Limitations

Our pipeline is not well-suited for processing news articles that do not follow a narrative structure. Some news articles have too many singleton events, which do not align with the existing framing structures typical in Western reporting. These articles, often lacking clear supportive or skeptical connotations. For instance, coverage on the Xinjiang issue often includes numerous rebuttal articles that focus on a single point of contention —accusations of Western fabrication — alongside many official statements. Similarly, articles on Mariupol from the Russian perspective entirely deny the events, leading to a lack of narrative structure and framing complexity.

We acknowledge the challenges associated with translating non-English articles, particularly when cultural nuances significantly impact interpretation. For instance, the term Pǔjīng Dàdì, literally translated as "Putin the Emperor," appears in some Chinese news articles. In Chinese cultural context, referring to someone as an emperor often emphasizes their dignity, elegance, and respectability, thus conveying a tone of support. However, this translation may inadvertently suggest a connection to dictatorship to English-speaking readers, a connotation absent in the original Chinese text. In future research, we aim to improve our handling of multilingual document collections. This will involve exploring more advanced multilingual language models that better capture such cultural and contextual nuances.

Our method also encounters bottlenecks with event coreference and causal relationship extraction, which is crucial for accurate attitude detection. Additionally, our current method is unable to detect sarcasm. Sarcasm can completely invert the perceived meaning of a statement, leading to misclassification of the article's attitude. This inability to identify sarcasm limits the accuracy of our analysis in articles where sarcastic remarks are prevalent.

Overall, while our pipeline is effective for traditional narrative news articles, it struggles with non-narrative formats, translation nuances, event models complexities, and the detection of sarcasm. These limitations highlight areas for future improvement to enhance the robustness and accuracy of our framing analysis.

Acknowledgement

This work is supported by grants from the CNS Division of National Science Foundation (Awards no: NSF_2213804) entitled "Building a Broad Infrastructure for Uniform Meaning Representations". Any opinions, findings, conclusions or recommendations expressed in this material do not necessarily reflect the views of NSF. We also wish to extend our appreciation to Cloudbank, which provided an indispensable computational resource for our experiments.

References

Mohammad Ali and Naeemul Hassan. 2022. A survey of computational framing analysis approaches. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9335–9348, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Christopher Baldassano, Uri Hasson, and Kenneth A. Norman. 2018. 1. representation of real-world event schemas during narrative perception. *The Journal of Neuroscience*.

Shany Barhom, Vered Shwartz, Alon Eirew, Michael Bugert, Nils Reimers, and Ido Dagan. 2019. Revisiting joint modeling of cross-document entity and event coreference resolution. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4179–4189, Florence, Italy. Association for Computational Linguistics.

Amber E. Boydstun and Justin H. Gross. 2013. Identifying media frames and frame dynamics within and across policy issues.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior Research Methods*, 46(3):904–911.

Bjorn Burscher, Rens Vliegenthart, and Claes H. de Vreese. 2016. Frames beyond words: Applying cluster and sentiment analysis to news coverage of the nuclear power issue. *Social Science Computer Review*, 34(5):530–545.

Avi Caciularu, Arman Cohan, Iz Beltagy, Matthew E. Peters, Arie Cattan, and Ido Dagan. 2021. Cdlm: Cross-document language modeling. *Preprint*, arXiv:2101.00406.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *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)*, pages 438–444, Beijing, China. Association for Computational Linguistics.

Hyung Won Chung, Le Hou, S. Longpre, Barret Zoph, Yi Tay, William Fedus, Eric Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Dasha Valter, Sharan Narang, Gaurav Mishra, Adams Wei Yu, Vincent Zhao, Yanping Huang, Andrew M. Dai, Hongkun Yu, Slav Petrov, Ed Huai hsin Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. 2022. Scaling instruction-finetuned language models. *ArXiv*, abs/2210.11416.

Ido Dagan, Lillian Lee, and Fernando Pereira. 1997. Similarity-based methods for word sense disambiguation. In *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, pages 56–63, Madrid, Spain. Association for Computational Linguistics.

Paul DiMaggio, Manish Nag, and David Blei. 2013. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of u.s. government arts funding. *Poetics*, 41(6):570–606. Topic Models and the Cultural Sciences.

Robert M. Entman. 1993. Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*, 43(4):51–58.

W.A. Gamson and A. Modigliani. 1989. Media discourse and public opinion on nuclear power: A constructionist approach. *American Journal of Sociology*, 95:1–37.

Erving Goffman. 1974. Frame analysis: An essay on the organization of experience. *Philosophy and Phenomenological Research*, 39(4):601–602.

William Held, Dan Iter, and Dan Jurafsky. 2021. Focus on what matters: Applying discourse coherence theory to cross document coreference. In *Proceedings of the 2021 Conference on Empirical Methods in*

Natural Language Processing. Association for Computational Linguistics.

Karen S. Johnson-Cartee. 2004. *2. News Narratives and News Framing: Constructing Political Reality*.

Emre Kıcıman, Robert Ness, Amit Sharma, and Chenhao Tan. 2023. Causal reasoning and large language models: Opening a new frontier for causality. *Preprint*, arXiv:2305.00050.

Siyi Liu, Lei Guo, Kate Mays, Margrit Betke, and Derry Tanti Wijaya. 2019a. Detecting frames in news headlines and its application to analyzing news framing trends surrounding U.S. gun violence. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 504–514, Hong Kong, China. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019b. Roberta: A robustly optimized bert pretraining approach. *ArXiv*, abs/1907.11692.

Yujian Liu, Xinliang Frederick Zhang, Kaijian Zou, Ruihong Huang, Nick Beauchamp, and Lu Wang. 2023. All things considered: Detecting partisan events from news media with cross-article comparison. *Preprint*, arXiv:2310.18827.

Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. Modeling framing in immigration discourse on social media. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2219–2263, Online. Association for Computational Linguistics.

Viet-An Nguyen, Jordan Boyd-Graber, Philip Resnik, and Kristina Miler. 2015. Tea party in the house: A hierarchical ideal point topic model and its application to republican legislators in the 112th congress. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1438–1448, Beijing, China. Association for Computational Linguistics.

OpenAI. 2023. Gpt-4 technical report. *ArXiv*, abs/2303.08774.

Chan Young Park, Julia Mendelsohn, Anjalie Field, and Yulia Tsvetkov. 2022. Challenges and opportunities in information manipulation detection: An examination of wartime Russian media. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 5209–5235, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Kyeongmin Rim, Jingxuan Tu, Bingyang Ye, Marc Verhagen, Eben Holderness, and James Pustejovsky. 2023. The coreference under transformation labeling dataset: Entity tracking in procedural texts using event models. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 12448–12460, Toronto, Canada. Association for Computational Linguistics.

Karin Kipper Schuler. 2006. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph.D. thesis, University of Pennsylvania.

Shirong Shen, Heng Zhou, Tongtong Wu, and Guilin Qi. 2022. Event causality identification via derivative prompt joint learning. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 2288–2299, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Jennifer Tracey, Ann Bies, Jeremy Getman, Kira Griffitt, and Stephanie Strassel. 2022. A study in contradiction: Data and annotation for AIDA focusing on informational conflict in Russia-Ukraine relations. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1831–1838, Marseille, France. European Language Resources Association.

Jingxuan Tu, Kyeongmin Rim, Eben Holderness, Bingyang Ye, and James Pustejovsky. 2023. Dense paraphrasing for textual enrichment. In *Proceedings of the 15th International Conference on Computational Semantics*, pages 39–49, Nancy, France. Association for Computational Linguistics.

Baldwin Van Gorp. 2006. The Constructionist Approach to Framing: Bringing Culture Back In. *Journal of Communication*, 57(1):60–78.

Jiarui Yao, Haoling Qiu, Jin Zhao, Bonan Min, and Nianwen Xue. 2021. Factuality assessment as modal dependency parsing. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1540–1550, Online. Association for Computational Linguistics.

Xiaodong Yu, Wenpeng Yin, and Dan Roth. 2022. Pairwise representation learning for event coreference. In *Proceedings of the 11th Joint Conference on Lexical and Computational Semantics*, pages 69–78, Seattle, Washington. Association for Computational Linguistics.

A Appendix

A.1 Annotation Guidelines for Classifying Articles on Putin's 2024 Election Win

Annotation Guideline Objective The purpose of this annotation task is to classify news articles about Vladimir Putin's 2024 election win into one of the three categories: supportive/skeptical/neutral.

Supportive An article is classified as supportive if it frames Putin's election victory in a positive or favorable light, endorsing his leadership or justifying the legitimacy of his win. These articles may portray Putin as a strong or capable leader and may downplay or omit any criticisms or allegations of electoral fraud. For example:

> ...Putin's decisive victory in the 2024 election is a clear reflection of the Russian people's confidence in his leadership and his steady hand in navigating the country through global uncertainties....

Skeptical An article is classified as skeptical if it questions the fairness, legitimacy, or democratic nature of Putin's election victory. These articles often highlight criticisms, allegations of voter suppression, or doubts about the transparency of the electoral process.

> ...Despite Putin's claimed landslide victory in the 2024 election, opposition parties and international observers have raised serious concerns about widespread voter suppression and allegations of election rigging that call into question the integrity of the results...

Neutral An article is classified as neutral if it provides a balanced or objective report on the election win, without showing clear bias toward or against Putin. These articles typically present facts without significant editorializing or framing that strongly supports or criticizes the election outcome. Lack of emotive or evaluative language. Mention of different perspectives on the election without prioritizing one over the other.

> ...In the 2024 Russian election, Vladimir Putin secured another term with 70% of the vote, according to official results. While government officials hailed the outcome as a victory for stability, opposition leaders have raised concerns about fairness, citing restricted media access during the campaign...

A.2 Sample prompts and responses of GPT-4o

```
What is the article's stance towards Putin's Election Win?
Please answer as either supportive, skeptical, or neutral in one word.

Some events might reflect a defensive and negative stance,
but they are supportive of Putin's perspective as it portrays him as a target of external forces,
which could garner sympathy and support, thus the article is considered supportive.
While skeptical articles highlight repression and lack of freedom of expression.

Input:President Vladimir Putin extended his reign over Russia in a landslide election ...

Output: Skeptical
```

Figure 2: GPT-4o prompt for baseline media attitude detection.

```
Each line below is the summary of an event mentioned in an article about "Putin's Election Win".
Based on these event summaries, determine the article's attitude towards the topic "Putin's Election Win".
Some events might reflect a defensive and negative stance,
but they are supportive of Putin's perspective as it portrays him as a target of external forces,
which could garner sympathy and support, thus the article is considered supportive.
While skeptical articles highlight repression and lack of freedom of expression.
Indicate whether the attitude is supportive, skeptical, or neutral in one word.

Input:
Nomination of candidates for Russian elections.
Putin delivering a speech.
Promotion of national development by Vladimir Putin.
State of the Nation Address outlining national development goals.
Russia's progression and strengthening.

Output: supportive
```

Figure 3: GPT-4o prompt media attitude detection with framing device1.

```
Each line below is the summary of an event mentioned in an article about "Putin's Election Win".
Based on these event summaries, determine the article's attitude towards the topic "Putin's Election Win".
Some events might reflect a defensive and negative stance,
but they are supportive of Putin's perspective as it portrays him as a target of external forces,
which could garner sympathy and support, thus the article is considered supportive.
While skeptical articles highlight repression and lack of freedom of expression.
Indicate whether the attitude is supportive, skeptical, or neutral in one word.

Input: Putin deliver a speech to the media.
his vote in the Russian election high.
Putin consolidate Russia ’s core leadership position in the post - Soviet space.
people in the Luhansk region participate in voting in the Russian presidential election.

Output: supportive
```

Figure 4: GPT-4o prompt media attitude detection with framing device2.

```
Each line below is the summary of a causal relation extracted in an article about "Putin's Election Win".
Based on these causal relations, determine the article's attitude towards the topic "Putin's Election Win".
Some events might reflect a defensive and negative stance,
but they are supportive of Putin's perspective as it portrays him as a target of external forces,
which could garner sympathy and support, thus the article is considered supportive.
While skeptical articles highlight repression and lack of freedom of expression.
Indicate whether the attitude is supportive, skeptical, or neutral in one word.

input:
Explosion at or near polling stations. It leads to Prohibition of certain activities or resources.
Russian national election voting. It leads to Incidents involving disruptions during a voting period.

output: supportive
```

Figure 5: GPT-4o prompt media attitude detection with framing device3.

```
Below is showing a list of mentions of the same event in different sentence contexts.
Each event mention is tagged by appending [EVENT] to it.
Please provide a single phrase that summarizes the event based on contexts.
The phrase should be concise, focusing only on the event and its participants.

Input:
... Putin almost certain to win[EVENT] a fifth term in office.
... Putin extended[EVENT] his reign over Russia ...
... trust in him as he celebrates his election win[EVENT]...

Output: Putin election win
```

Figure 6: GPT-4o prompt for generating descriptors for event coreference clusters.

```
Please analyze the following event mentions, where each event trigger is tagged with
[cluster-\\d]. Extract and list all causal relations among the events, using the format
[cluster-\\d] -> [cluster-\\d] to indicate the direction of the causal relationships.

Input:
... President Vladimir Putin extended[cluser-2] his reign over Russia in a landslide election...
Output: [cluster-3] -> [cluster-2]
        [cluster-4] -> [cluster-2]
```

Figure 7: GPT-4o prompt for extracting causal relationships among events.