

# Talking Point based Ideological Discourse Analysis in News Events

Nishanth Nakshatri<sup>✳</sup> Nikhil Mehta<sup>✳</sup> Siyi Liu<sup>◇</sup> Sihao Chen<sup>◇</sup>  
Daniel J. Hopkins<sup>◇</sup> Dan Roth<sup>◇</sup> Dan Goldwasser<sup>✳</sup>

<sup>✳</sup>Purdue University <sup>◇</sup>University of Pennsylvania  
{nnakshat, mehta52, dgoldwas}@purdue.edu  
{siyiliu, sihaoc, danhop, danroth}@upenn.edu

## Abstract

Analyzing ideological discourse even in the age of LLMs remains a challenge, as these models often struggle to capture the key elements that shape real-world narratives. Specifically, LLMs fail to focus on characteristic elements driving dominant discourses and lack the ability to integrate contextual information required for understanding abstract ideological views. To address these limitations, we propose a framework motivated by the theory of ideological discourse analysis to analyze news articles related to real-world events. Our framework represents the news articles using a relational structure—*talking points*, which captures the interaction between entities, their roles, and media frames along with a topic of discussion. It then constructs a vocabulary of repeating themes—*prominent talking points*, that are used to generate ideology-specific viewpoints (or partisan perspectives). We evaluate our framework’s ability to generate these perspectives through automated tasks—ideology and partisan classification tasks, supplemented by human validation. Additionally, we demonstrate straightforward applicability of our framework in creating event snapshots, a visual way of interpreting event discourse. We release resulting dataset and model to the community to support further research<sup>1</sup>.

## 1 Introduction

One of the signs of the growing social and political polarization is the formation of insulated information bubbles (Gentzkow and Shapiro, 2011; Quattrociocchi et al., 2016; Dubois and Blank, 2018; Garimella et al., 2018), in which news media discourse is shaped around ideological lines, often intended to shape the readers’ views. Understanding this phenomenon better, to the extent we can examine the degree to which members of the two

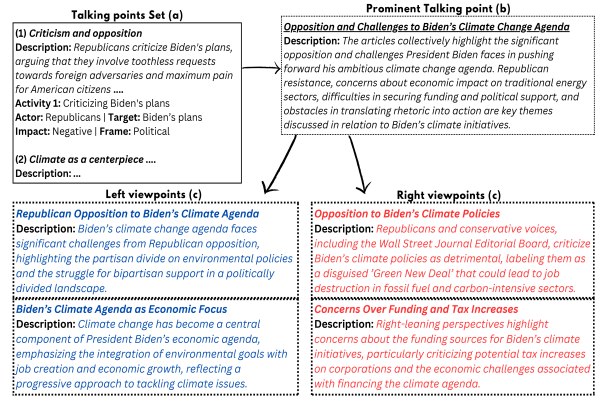


Figure 1: Shows how each political orientation discusses an event about *Climate Change*. (a) Shows a collection of *talking points* extracted from all the news articles about the event. (b) Shows an example of a *Prominent Talking Point* (PTP), a repeating theme discussed by both political ideologies. (c) Shows *left-* and *right-* ideological interpretations of the PTP.

communities hold opposite accounts of reality, requires computational methods that can compare the narratives of both sides and identify the points in which their accounts converge and diverge (example shown in Figure 1).

Past work typically focused on discrete aspects, such as stance and bias detection (Liu et al., 2022; Luo et al., 2020; Kiesel et al., 2019; Li and Goldwasser, 2019, 2021), political news framing (Mendelsohn et al., 2021; Roy and Goldwasser, 2020; Field et al., 2018; Card et al., 2015b), sentiment toward relevant entities (Park et al., 2021; Rashkin et al., 2016), which while relevant, fall short of providing the comprehensive view needed to analyze political discourse. The rise of Large Language Models (LLMs) has a transformative potential for enabling complex discourse analysis connecting these discrete aspects and explaining their relationship. However, realizing it is not straightforward, as demonstrated by several recent works analyzing political texts, either as a straightforward stance prediction task (Ziems et al., 2024), or mapping political positions to specific stances on policy issues (Santurkar et al., 2023).

<sup>1</sup>Data available at: <https://github.com/nnakshat/TP-IDA>

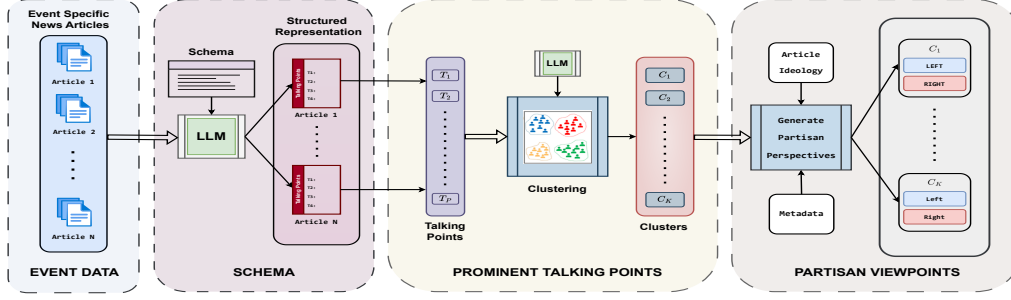


Figure 2: **Framework overview:** Event-specific news articles are organized into a relational structure, referred to as *talking points*. These are clustered to identify *Prominent Talking Points* (PTPs), a vocabulary of repeating themes that are relevant to the event. Each PTP is infused with ideological information to identify *left-leaning* and *right-leaning* viewpoints, referred to as *partisan perspectives*.

To address these challenges, we suggest a structured method to characterize ideological discourse by analyzing repeating themes that may be shared across political sides. We achieve this by defining a relational structure—*talking point*, which builds upon foundational research in political discourse analysis (Van Dijk, 1996; Van Dijk et al., 1997).

A *talking point* captures a key aspect of a news article through a short summary and a set of properties. These properties include (1) the lens through which it is discussed, identified using media-frames (Boydston et al., 2014), (2) relevant entities mentioned, (3) roles assigned to these entities, and (4) attitudes towards them (Khanehzar et al., 2021; Roy et al., 2021). By organizing news articles into this structure, we develop a vocabulary of repeating themes and use these properties to compare variation in discourse across political sides. This approach enables us to detect potential *agenda-setting* attempts (McCombs and Shaw, 1972; Scheufele and Tewksbury, 2007) by identifying themes overwhelmingly discussed by one side. In addition, it helps uncover *areas of consensus and polarization*, based on examining how both sides engage with the same repeating theme—either similarly or in a contrasting manner. Figure 4 exemplifies this analysis for Covid-19.

To implement this approach, we develop a pipeline to characterize how each political orientation discusses a *news event* (illustrated in Figure 2). First, an LLM extracts main topics of discussion from each event-related article and organizes them according to our structured schema — *talking points*. Next, we identify repeating themes relevant to the event, which we refer to as *Prominent Talking Points* (PTPs). These PTPs are high-level abstractions of frequently repeating *talking points* and are commonly discussed across the political spectrum.

To obtain PTPs, we cluster the extracted *talking points* and use an LLM to characterize each cluster. We further refine these clusters by eliminating redundancies and inconsistencies. Once the PTPs are established, we use an LLM to reason about how different political leanings (*left* and *right*) interpret each PTP. The resulting *left-leaning* and *right-leaning* viewpoints are collectively referred to as *partisan perspectives*, capturing differences in framing and attitudes towards entities, as shown in Figure 1. Finally, to obtain a comprehensive view of how each political leaning discusses the entire event, we compile the viewpoints associated with each PTP. Specifically, we obtain a *left-leaning perspective* on the event by aggregating all *left-leaning* viewpoints for each PTP, and the *right-leaning perspective* on the event is constructed similarly.

This process generates resources validated by both human and automated techniques, beneficial for researchers. Specifically, we provide a dataset comprising a structured representation of 6,141 articles on 24 events related to 4 contested political issues, sourced from 126 outlets coded for bias. Further, we also identify PTPs for each event, and provide their respective *partisan perspectives*.

We suggest two automated methods to evaluate the generated *partisan perspectives*. These evaluations occur at two levels: (1) event-level granularity, and (2) individual PTPs. At the event level, we perform an *ideology classification task* on previously unseen news articles. This task validates whether our method can generate broader partisan discourse that effectively captures the ideological nuances of different political groups. At the PTP level, we perform a *partisan classification task* to measure the degree to which the generated perspectives exhibit political leaning towards a specific ideology.

To summarize, our main contributions are: **(1)**

We propose a new way to characterize ideological discourse in an event via a vocabulary of repeating *talking points*. (2) We suggest an LLM-based pipeline to generate *partisan perspectives*. (3) We perform automated and human validation of generated *partisan perspectives*, resulting in a novel dataset that captures partisan views over multiple events, which can be used to drive future research.

## 2 Related Work

Prior work on studying partisan perspectives in NLP has primarily focused on *frames*. While a contested concept, framing is commonly conceived of as a communicative structure in which the speaker highlights specific aspects of an issue to promote a political viewpoint (Goffman, 1974; Entman, 1993; Kinder, 1998; Chong and Druckman, 2007). (Card et al., 2015b) proposes the Media Frame Corpus that has 15 generic media frames defined by (Boyd-stun et al., 2014), such as economics or public opinion. In a polarized media environment, frames serve as instrumental mechanisms to promote political agendas through the selective coverage of events (informational bias) and the manipulation of their presentation (lexical bias) (Gentzkow and Shapiro, 2006; Jamieson et al., 2007; Fan et al., 2019). Prior work has also explored approaches to automatically detect and mitigate framing biases. (Liu et al., 2019; Akyürek et al., 2020) identify frames through news headlines, (Ji and Smith, 2017; Khanehzar et al., 2021) detect frames at a document level, and (Lee et al., 2022; Liu et al., 2023) mitigate framing biases using multi-domain summarization and graphs. However, the formalization of frames oversimplifies the intricacies of partisan perspectives and falls short in capturing the nuance of how political agendas are deliberately conveyed in news articles. In this work, we look closer at news articles, and represent them with a predefined structure of *talking points* – capturing a key aspect of discussion along with interaction between with respective entities, and cluster them to identify repeating themes, to collectively shape the *partisan perspectives*. Identifying *talking points* can be thought of similar to using LLMs to generate explicit representations that helps in assessing arguments (Hoyle et al., 2023). Recent work has also explored finer analysis in news articles/political biases. (Lawlor and Tolley, 2017) presents an entity-focused study of media news framing. (Spinde et al., 2021) detects media biases at the word and sentence level, and (Frermann et al.,

2023) identifies and uses multi-label frames. Our work complements these by introducing a framework that allows us to establish repeating themes based on talking points to unveil the *partisan perspectives* within an event.

## 3 Our Framework

In this section, we describe our systematic approach towards generating *partisan perspectives* for a real-world news event. An overview of our approach is illustrated in Figure 2.

### 3.1 Schema

**Background.** Political groups often simplify or obscure their ideological differences. Therefore, to analyze them, theory of ideological discourse analysis, as proposed by (Van Dijk, 1996; Van Dijk et al., 1997), introduces semantic discourse structures that can capture the abstract ideological positions of political groups across various social issues and contexts. In particular, (Van Dijk, 1996) suggested that ideological discourse can be predicted by examining local meanings and implications through goal and activity-descriptions, such as questions like “What are our tasks?” and “What are our roles?” etc. Additionally, (Van Dijk et al., 1997) proposed to characterize political discourse by defining semantically relevant structures known as *topics*, which are contextualized by their respective topical actors to better understand persuasive communication.

**Structured Representation.** Inspired by these foundational works, rather than directly analyzing news articles, we organize them into a set of *talking points*, a relational structure to characterize ideological differences.

Similar to the discourse structures in (Van Dijk, 1996; Van Dijk et al., 1997), a *talking point* consists of a key discussion element from a news article, which helps to maintain the semantic relevance to the article. To capture nuanced ideological positions and local meanings, we contextualize the discussion by extracting *metadata* that can help analyze ideological differences. This *metadata* is constructed by first identifying the set of entities associated with the discussion. We then capture the relationships between these entities and their influence on each other by identifying the set of activities linked to the discussion. An activity is characterized by the structure *who did what to whom*—it includes a brief description, an *actor* who

initiates the activity, a *target* affected by the actor, and the sentiment towards the target entity, indicating whether the impact is positive or negative. Additionally, we follow the nomenclature in (Card et al., 2015a), to identify the media frame associated with each activity. Thus, we formally define a *talking point* as a relational structure consisting of a key discussion element from a news article, contextualized with the *metadata* (exemplified in Figure 1(a)).

**Extraction.** We represent each news article related to a news event using its identified *talking points*. Given a set of  $n$  news articles  $\{d_z\}_{z=1}^n$  relevant to an event  $\mathcal{E}$ , our proposed schema distills each article  $d_z$  into a set of up to four key *talking points*, denoted as  $\{t_i\}_{i=1}^m$ , where  $m \leq 4$ .

Building on (Stammach et al., 2022), which found GPT-3 to be at least 63% accurate in extracting entities and its roles, we prompt an LLM to generate structured representations per our schema. We let the LLM decide the number of *talking points* in a news article. Rather than generating a fixed number of *talking points*, we request the LLM to generate at most 4 points. This adaptive approach allows the LLM to generate fewer talking points (or no points at all) when there is no useful information in the article. The prompt template is provided in Figure 6 in the Appendix.

### 3.2 Characterize Partisan Perspectives

After structuring the event-related news articles according to our schema, we outline the methodology used to obtain *left-leaning* and *right-leaning* perspectives on the event  $\mathcal{E}$ .

#### 3.2.1 Prominent Talking Points Identification

Although *talking points* are equipped to analyze nuanced political messaging within a news article, they do not directly capture broader ideological discourse at the granularity of a news event  $\mathcal{E}$ . To capture this, we identify and focus on *Prominent Talking Points* (PTPs) that shape broader narratives. PTPs are repeating themes that are potentially discussed on both sides of the political spectrum.

To obtain PTPs, we first compile a comprehensive talking points set  $\mathcal{T} = \{t_s\}_{s=1}^p$  by aggregating all the *talking points* in the articles about  $\mathcal{E}$ . Then, we cluster  $\mathcal{T}$  to identify groups of frequently repeating *talking points*. Each group is labeled using an LLM. The label denotes a higher-level abstraction of the *talking points* within that group. We refer to

each label as a PTP, which is a topical representation of the frequently occurring *talking points* that constitute the narrative of an event. Algorithm 1 outlines the procedure used to obtain PTPs.

**Clustering.** We embed each point in  $\mathcal{T}$  using a dense retriever model  $f$  (Ni et al., 2021), and use HDBSCAN (Campello et al., 2013) to group them into candidate *clusters*. For each cluster  $c$ , we prompt the LLM to assign a label comprising two components: an *aspect* and a brief *description*. The aspect represents a broad concept evident in the top-5 talking points nearest to the cluster centroid, and the description provides a short summary of these points. Note that we exclude metadata during clustering to ensure the clusters accurately capture meaningful topic-based abstractions of *talking points*. Including metadata introduces noise and affects the clarity of these representations.

The initial clustering uses conventional distance metrics and is imperfect. To improve it, we further refine the cluster labels by merging redundant clusters and removing inconsistent ones. Redundancy is addressed by greedily comparing and merging cluster pairs that share the same *aspect*. The resulting refined labels represent the set of PTPs for  $\mathcal{E}$ . Next, each talking point  $t_s \in \mathcal{T}$  is assigned to a PTP based on the cosine similarity between its embedding and the PTP’s embedding. This process results in clusters  $\{\mathcal{C}_j\}_{j=1}^k$  and their associated labels  $\{\mathcal{L}_j\}_{j=1}^k$ , referred to as PTPs. More details about the clustering process are in Appendix D.3.

---

#### Algorithm 1 Identify prominent talking points

---

```

1: Input: Talking points  $\mathcal{T} = \{t_s\}_{s=1}^p$ 
2: Initialize: embeddings  $\mathcal{Z} = \{z_s = f(t_s)\}_{s=1}^p$ ,  $n \leftarrow$ 
   no. of news articles,  $\mathcal{C} \leftarrow \{\}$ ;
3: while  $|\mathcal{Z}| > 0.1 * n$  do
4:   clusters  $\leftarrow \text{Clustering}(\mathcal{Z})$ ;
5:   labelSet  $\leftarrow []$ ;
6:   for  $c$  in clusters do
7:     Compute centroid  $\mu_c$  by averaging;
8:      $Z' \leftarrow \text{getTopKPoints}(c, \mu_c)$ ;
9:     cLabel  $\leftarrow \text{getClusterLabel}(Z')$ ;
10:    Append cLabel to labelSet;
11:   end for
12:   updatedLabels  $\leftarrow \text{updateLabelSet}(\text{labelSet})$ ;
13:    $\mathcal{S} \leftarrow \text{TalkingPtMembership}(\mathcal{T}, \text{updatedLabels})$ 
14:    $\mathcal{T}' \leftarrow \text{getClusteredDocs}(\mathcal{S})$ 
15:    $\mathcal{T} \leftarrow \mathcal{T} \setminus \mathcal{T}'$ ;
16:    $\mathcal{Z} \leftarrow \mathcal{Z} \setminus \{\text{embeddings of } \mathcal{T}'\}$ ;
17:   Append clusters in  $\mathcal{S}$  to  $\mathcal{C}$ 
18: end while
19: Output:  $k$  clusters  $\mathcal{C} = \{\mathcal{C}_j\}_{j=1}^k$  with cluster labels
    $\{\mathcal{L}_j\}_{j=1}^k$ 

```

---



### 3.2.2 Generate Partisan Perspective

The clustering process identifies a set of PTPs relevant to the event  $\mathcal{E}$ . To analyze how different political ideologies—such as *left* and *right*—discuss  $\mathcal{E}$ , we examine how each ideology engage with these PTPs. Specifically, for each PTP, we define interpretations from the *left* ideology as *left-leaning* viewpoints and interpretations from the *right* ideology as *right-leaning* viewpoints. Collectively, these are referred to as *partisan perspectives*. To construct a comprehensive *left-leaning perspective* on the event  $\mathcal{E}$ , we aggregate the left-leaning viewpoints associated with each PTP. The *right-leaning perspective* on  $\mathcal{E}$  is formed in a similar manner.

**Ideological Interpretation of PTP.** PTPs obtained from the clustering process do not inherently capture the ideological specificity needed to define *partisan perspectives*. To address this, we first assign an ideology label —  $\{left, right\}$ , to each *talking point* within the PTP cluster, denoted as  $\mathcal{C}_j$ . The ideology label assigned to a *talking point* corresponds to the ideology label of the news article it originates from. Using this labeling, we systematically extract *left-leaning* and *right-leaning* viewpoints for each PTP based on the metadata associated with the *talking points*.

Particularly, to construct a *partisan perspective*, we illustrate the generation of *left-leaning* viewpoints (*right-leaning* ones follow an analogous process). For each PTP, our goal is to generate viewpoints that clearly distinguish the *left-view* from the *right*. To achieve this, we instruct the LLM to create *left-leaning* viewpoints in a contrastive manner. Specifically, we prompt the LLM with characteristic *talking points*: the top- $K$  left-biased points alongside the top- $M$  right-biased ones.<sup>2</sup> This allows the model to highlight contrasts and identify defining features of the *left-viewpoint*. The characteristic *talking points* are selected based on their cosine similarity to the PTP. Note that these *talking points* include metadata such as key actors, their targets, sentiments toward these targets, and the media frame. Incorporating these details ensures that generated *left-leaning* viewpoints capture nuanced relationships between the involved entities.

However, simply using the top- $K$  left-biased *talking points* is not enough, as these may not capture the broader ideological bias in the entire article. To mitigate this, we also include article summaries

Target Entity	Most common Actor	Sentiment on Target	Media frame
United States economy and job market	U.S. (represented by President Biden)	Positive	Economic
American workers	U.S. (represented by President Biden)	Positive	Policy
Biden	Republican Party	Negative	Political
Opponents	Biden	Negative	Political
Biden's climate change agenda	Republicans	Negative	Political
Joe Biden and his administration	CNN	Positive	Policy
Biden's climate change agenda	United Mine Workers of America	Positive	Economic
Joe Biden and his administration	Wall Street Journal Editorial Board	Negative	Policy
Biden's climate change agenda	Republican leaders in West Virginia	Negative	Political
Biden's climate vision & infrastructure bill	Opponents (specifically Republicans)	Negative	Political

Figure 3: Shows Aggregated metadata corresponding to the PTP shown in Figure 1. Each target is paired with its most common actor, sentiment, and media frame within the actor-target context. Metadata is color-coded: **Blue** for *left* ideology and **Red** for *right* ideology.

associated with top- $K$  points in the prompt. To ensure that these summaries reflect ideological biases and relevant information from the *talking points*, we instruct the LLM to generate summaries conditioned on: (1) ideology label of the article (2) *title* associated with the PTP. A detailed prompt template for generating these viewpoints is shown in Figure 7 in the Appendix.

**Metadata Aggregation for each PTP.** Beyond *partisan perspectives*, for each ideology, metadata from the top 50% of *talking points* is aggregated to identify frequent and distinctive entity pairs with their sentiments. Specifically, we extract the top-3 targets with positive and negative sentiments, their most common associated actors, and the dominant media frame for each actor-target pair. This aggregated view supplements *partisan perspectives* by offering insight into broader dynamics of clusters.

An example of left- and right-viewpoints for a PTP is shown in Figure 1, with its aggregated metadata provided in Figure 3. Interestingly, few right-biased sources like *Washington Times* highlight support for *Biden's Climate Agenda* from *United Mine Workers*. This support is based on the agenda's focus on securing jobs for displaced miners.

## 4 Dataset

To illustrate the effectiveness of our proposed framework, we use the KeyEvents dataset (Nakshatri et al., 2023). This dataset is constructed by segmenting the archive of news articles from NELA-2021 (Horne et al., 2022) into a set of temporally motivated news events. To construct these events, Nakshatri et al. (2023) dynamically analyzed the

<sup>2</sup>Based on empirical validation, we set  $M = 3$ ,  $K = 5$ .

Issue	No. of Articles	No. of Events
Climate Change	579	8
Capitol Insurrection	1,609	4
Immigration	1,137	4
Coronavirus	2,816	8
<b>Total Count</b>	6,141	24

Table 1: The dataset we use for testing our proposed framework. It is sampled from Nakshatri et al. (2023).

temporal trend of news articles published for a given issue, and identified the temporal landmarks that could signify the presence of an important news event. Then, the news articles published in and around the temporal landmarks were clustered to identify all the documents relevant to the news event. In this manner, Nakshatri et al. (2023) proposed a dataset comprising of 40k news articles with 611 key news events from 11 issues.

As our goal is to analyze political discourse and characterize partisan perspectives at an event-level granularity, this dataset can be directly applicable to test the efficacy of our framework. Thus, we manually sample a set of six issues and a set of events which have the highest number of news articles per event from this dataset. Table 1 shows the detailed statistics of our final dataset.

## 5 Evaluation

Our goal is to evaluate how effectively our framework generates *partisan perspectives*. To measure its performance, we use two types of automated evaluations (described below) along with human assessments. Additional evaluation of intermediate framework steps is detailed in the Appendix A.3.

**Ideology Classification.** In this evaluation, we assess how effectively our framework characterizes ideological discourse at the event level. Specifically, we examine whether the generated perspectives capture event-level political discourse and encode nuanced ideological cues. To achieve this, we consider an *ideology classification task* over previously *unseen* news articles related to the event  $\mathcal{E}$ . Here, we use *only* the generated *left-leaning* and *right-leaning* perspectives of  $\mathcal{E}$  to predict the ideological stance of the article. A strong performance on this task would indicate that the generated perspectives are event-centric rather than being overly specific to individual news articles. Furthermore, it would demonstrate that these perspectives effectively capture the ideological characteristics of different political sides. Our objective is not to outperform existing classification baselines, but rather to determine whether the generated perspectives

provide a faithful representation of the broader ideological discourse surrounding an event.

**Partisan Classification.** While ideology classification evaluates perspectives at the event level, this task focuses on assessing their quality at the PTP cluster level (as defined in Section 3.2.2). Within each PTP cluster, *left-leaning* viewpoints should align with how *left-* political ideology interprets the issue relative to the PTP, and the same applies to *right-leaning* viewpoints. As a consequence, *left-leaning* viewpoints should be primarily supported by *left-leaning* news articles and not appear in *right-leaning* articles, and vice-versa. This evaluation measures how well the generated perspectives follow this expected pattern.

## 6 Experiments & Results

In this section, we discuss the experimental details and findings linked to both automated and human evaluations. We use ChatGPT<sup>3</sup> as the LLM to obtain *partisan perspectives*, following the approach detailed in Section 3. This process yields structured representations for articles from every event in our dataset, and we release these, along with the original dataset, to the community.

### 6.1 Ideology Classification

**Task Setup.** Given an *unseen* news article related to the event but not part of the initial dataset, the task is to predict its ideology label (*left*, *right*). The central idea involves using *only* the generated *partisan perspectives* for classification. To do this, first, we consider the comprehensive *left-leaning* and *right-leaning perspective* on  $\mathcal{E}$ . Then, we identify the three most similar *left-* and *right-leaning* viewpoints to the input article using cosine similarity. We denote this as *TopK Event Partisan View*. For classification, we provide the input article along with the *TopK Event Partisan View* to an LLM and prompt it to assign the article to the viewpoint it aligns with most closely. In this process, we replace the explicit labels *{left-, right-}* viewpoints with neutral placeholders—*{summary1, summary2}*, to avoid any inherent bias of LLM. This ensures that the classification relies only on the LLM’s ability to align the article with the closest summary. For this task, we curated 481 unseen event-related articles (see Appendix C.1 for details).

**Comparison Methods.** We benchmark task performance with two LLMs—ChatGPT and

<sup>3</sup>gpt-3.5-turbo-0125 (OpenAI, 2022)

LLAMA3<sup>4</sup> (Dubey et al., 2024), against these methods: **1. Direct Prompting**, where the LLM is directly prompted for the article’s ideology label. **2. TopK Event Partisan View**, and **3. TopK Event Partisan View + Metadata**, which adds metadata to the topK view. For each ideology, metadata is aggregated from the top 50% of *talking points* closest to PTP (obtained from Section 3.2.2).

**Results.** Table 3 presents ideology classification task results. *TopK Event Partisan View* outperforms the direct prompting baseline, improving F1-scores by +4.5 for ChatGPT and +1 for LLama. This suggests that the partisan perspectives effectively capture event-level partisan signals that are essential for predicting the ideology of an unseen article. In addition, we observe that the inclusion of metadata can help improve classification task performance. This improvement suggests that metadata produced by our method aids in ideological discourse understanding by capturing the relationship between entities involved.

### 6.1.1 Additional validation of perspectives

**Narrative-LLAMA.** We further validate the effectiveness of generated perspectives by investigating if we could directly fine-tune an LLM to identify event-level discourse signals from individual news articles. Given a news article and its closest ideological viewpoints, the task is to train the model to generate these viewpoints without relying on intermediate steps in our framework. To achieve this, we used the *partisan perspectives* generated by our framework as *training data* to fine-tune a LLAMA3-8B-instruct model using PEFT<sup>5</sup> and LORA (Hu et al., 2021). We compiled 1100 examples pairing articles with their *left-* or *right-*viewpoints from PTP cluster, and trained the model using DPO (Rafailov et al., 2024) (model training details can be found in Appendix C.2). To evaluate the fine-tuned model, we tested it on ideology classification task using *unseen* articles. We assess classification via: **Prompting with Generated Perspective**, obtaining perspectives with Narrative-LLAMA and prompting for article ideology.

**Results.** From Table 2, it is evident that the fine-tuned Narrative-LLAMA outperforms the LLAMA3-8b-instruct model by increasing the F1 score by +3.6 points in the ideology classification task. This suggests that (1) *partisan perspectives*

LLM	Method	Prec.	Recall	F1
LLAMA	Prompting w	87.8	84.9	84.9
	Generated Perspective			
Narrative-LLAMA	Prompting w	89.5	88.4	<b>88.5</b>
	Generated Perspective			

Table 2: Narrative-LLAMA (fine-tuned model) improves over Llama3 (+3.6 F1-score) in ideology classification.

LLM	Method	Prec.	Recall	F1
ChatGPT (0-shot)	Direct Prompting	81.38	74.66	73.52
	TopK Event Partisan View	77.12	76.64	76.61
	+Metadata	81.20	79.78	<b>79.69</b>
ChatGPT (2-shot)	Direct Prompting	80.47	78.23	78.09
	TopK Event Partisan View	83.02	82.61	<b>82.65</b>
	+Metadata	83.79	82.50	82.52
LLAMA (2-shot)	Direct Prompting	79.54	81.0	79.77
	TopK Event Partisan View	80.45	81.71	80.78
	+Metadata	81.62	82.62	<b>81.93</b>

Table 3: Averaged results for ideology classification across 4 issues. *TopK Event Partisan View* effectively captures event-level ideological discourse.

from our framework that was used as *training data* encodes ideology-specific nuances. (2) generated perspectives capture broader event-level signals than focusing on specific details of the article (as illustrated in Figure 5 in the Appendix). Furthermore, using the Llama3-70B-instruct model, we assess the quality of generated perspectives by verifying if they actually appear in the original news articles (Appendix A.2). Notably, hallucination rates were found to be below 6%, and we intend to release the model parameters to the community.

## 6.2 Partisan Classification

**Task Setup.** Within each PTP cluster, this task focuses on evaluating whether the *left-leaning* viewpoints are supported by left-biased news articles and not by right-biased ones, and vice-versa. To achieve this, we first identify the news articles corresponding to the *talking points* within each PTP cluster. Then, for a PTP, given a news article along with the *left-* and *right-leaning* viewpoints for that PTP, we prompt an LLM to assign the article to the viewpoint it most closely aligns with. If the model successfully categorizes left-biased news articles to *left-leaning* viewpoints, then it suggests that these viewpoints are more aligned with how *left-* political ideology discusses the PTP (similarity for the *right*). As discussed in Section 6.1, we replace the terms *{left, right}* with *{summary1, summary2}* in this task. This substitution helps mitigate potential bias in the LLM. The prompt template used for this process is provided in Figure 8 in the Appendix.

<sup>4</sup>meta-llama/Meta-Llama-3-8B-Instruct

<sup>5</sup><https://github.com/huggingface/peft>



**Comparison Methods.** We benchmark the task performance using ChatGPT against the following methods: **1. Partisan Perspectives**, represents the perspectives generated by our framework at each PTP. **2. Partisan Perspectives + Metadata**, adds metadata aggregated from the top 50% of *talking points* closest to PTP. **3. Topically Relevant Points (TRPs)**—We construct a baseline based on the talking points used in the clustering process prior to generating *left-* and *right-leaning* interpretations of a PTP (Section 3.2.1). First, we label each talking point based on the ideology label of the news article it originates from. Then, for each political ideology, we consider the top-3 talking points closest to the PTP label to obtain *left-* and *right-* viewpoints for that PTP (collectively denoted as **TRPs**).

**Results.** Table 4 shows the results of partisan classification task, with issue-specific details available in Appendix A.1. We observe that *Partisan Perspectives* outperforms the baseline by +11 F1-score points, indicating that *left-* and *right-leaning* viewpoints tend to be strongly aligned with news articles from their respective ideologies within each PTP cluster. Adding metadata to *Partisan Perspectives* further enhances performance on the task. However, **TRPs** themselves do not inherently encode ideology-specific information. As a result, the LLM struggles to correctly assign news articles to their respective viewpoints.

Approach	Prec.	Recall	F1
TRPs (baseline)	73.44	73.33	73.37
Partisan Perspectives	85.03	84.61	84.76
Partisan Perspectives + Metadata	<b>85.93</b>	<b>86.14</b>	<b>85.98</b>

Table 4: Averaged results for *partisan classification task* across all issues. At PTP cluster level, *partisan perspectives* align with how each political ideology interprets the PTP.

### 6.3 Human Evaluation

We conduct a human evaluation on a small sample of data collected from 3 randomly selected events (see Table 11 in the appendix for event description details). Our evaluation focuses on measuring the quality of the generated *partisan perspectives* using: *summary coherence* and *mapping quality*.

(1) **Summary Coherence:** For each PTP, we evaluate whether the *left-leaning* viewpoints accurately capture key ideas from the top- $K$  left-biased *talking points* in that cluster (denoted as *L-Coherence*). *R-Coherence* is defined in a similar manner. The goal is to ensure that the generated viewpoints are not entirely random, but consistent

with the set of *talking points* used to construct it. (2) **Mapping Quality:** For each PTP, this criterion validates whether the generated *left-leaning* viewpoints are expressed in left-biased *news articles* within that PTP cluster—denoted as *L-MQ*. *R-MQ* is defined in a similar manner for the *right-leaning* viewpoints. For mapping quality, annotators must compare the generated viewpoints with excerpts from news articles to determine if the viewpoints are reflected in the article text. In total, our evaluation covered *left-* and *right-leaning* viewpoints from 84 PTP clusters across the 3 selected events.

We note that evaluating *mapping quality* requires reading lengthy article excerpts and thus, scaling it up is challenging. To address this, we also evaluated *mapping quality* using LLM and human validation (denoted as *MQ\_LLM*). Specifically, we prompted GPT-4o with 4 predefined questions designed to extract evidence from news articles supporting the presence of the generated viewpoints. These questions focus on the topic, key entities and their associated sentiments. Instead of manually reviewing long excerpts, human annotators compare generated viewpoints and extracted evidence to validate if the evidence aligns with the viewpoints. Using this approach, we annotated 92 article-viewpoint pairs for the *Climate Change*-related news event. Detailed discussion on the evaluation setup for all the metrics is provided in Appendix A.4.

**Discussion.** Table 5 shows the results for both *mapping quality* and *summary coherence*. Our findings indicate that generated viewpoints, for both political ideologies, have high coherence scores, suggesting that they align well with *talking points* from which they were created. Further, high mapping quality scores suggest that these viewpoints are accurately reflected in the news articles. However, a manual review suggests that these viewpoints can be sometimes incorrect. These typically occurs when the LLM produces poor summaries of news articles, leading to inaccuracies in generated viewpoints (as shown in Figure 10). In addition, the LLM occasionally overlooks information cited in the news articles, leading to incorrect summaries.

Table 6 shows *MQ\_LLM* results. We report the *evidence alignment score* — proportion of instances where GPT-4o extracted evidence aligns with the generated viewpoints, according to human validation for each question type. We observe a high score between GPT-4o extracted evidence



and viewpoints, indicating that the generated viewpoints are expressed in the news articles. However, we observe a relatively low score for the question type — entities are viewed negatively. In this case, manual inspection shows that GPT-4o sometimes misses relevant evidence from the articles, exemplified in Table 12 in the Appendix.

Issue	L-Coherence(%)	R-Coherence(%)	L-MQ(%)	R-MQ(%)
Climate Change	85.71	100	75.00	76.92
Coronavirus	100	90.90	90.90	70.00
Immigration	93.33	100	84.62	94.44

Table 5: Results from 84 annotations for *summary coherence* and *mapping quality*.

## 7 Broader Impact

**Analyzing Event-Level Discourse.** Our framework enables us to qualitatively visualize news events, offering insights about the *areas of consensus* and *polarization* by analyzing highly contested repeating themes - PTPs. For each PTP, we use GPT-4o to systematically analyze the agreement and disagreement between its *left-* and *right-* interpretations through a series of questions shown in the Appendix A.5. This analysis enables us to construct an *event-discourse snapshot*, a visualization that captures the political discourse surrounding a specific event. Figure 4 illustrates a *snapshot* for an event concerning Coronavirus - *Biden’s COVID-19 Relief Bill*. Our visualization maps each PTP as a circle along the X-axis where its position reflects its bias, with balanced view placing it in the center. Circle size indicates its frequency. Most frequent PTPs with high positive or negative x-axis values is evidence for "different realities", i.e., focusing on very different aspects of the topic. This visualization helps in analyzing discourse by categorizing the PTPs into different types — *agreement*, *disagreement*, *one-sided*, etc. Table 13 in Appendix provides an example for such categorization. For instance, talking point ID 2 (*Enhanced Vaccination Efforts*), have been discussed the most, and both political parties indicate disagreement (as it is below X-axis), signifying a strong *partisan battle*. While the *left* focus their discussions on *equitable vaccine distribution and healthcare reform*, the *right* focus

Question Type	Evidence Alignment Score(%)
Topic	97.82
Entities viewed negatively	80.43
Entities viewed positively	92.39
Angle of discussion	98.91

Table 6: Shows the GPT-4o extracted evidence and viewpoints alignment score for 92 article-viewpoint pairs (for *Climate Change*-related event).

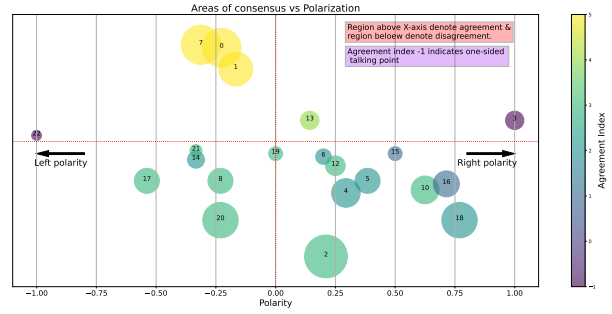


Figure 4: Discourse snapshot for a *Covid*-related event. Numbers represent PTP IDs. Points near center denote bipartisan discussion.

mostly on *criticizing Biden’s vaccine distribution decisions*. This structured approach enhances the ideological discourse understanding and facilitates broader studies on media polarization and framing.

## 8 Conclusion

In this paper, we introduced an LLM-based framework for analyzing ideological discourse in news events. We demonstrated the effectiveness of our approach through both automated tasks—ideology and partisan classification, and human validation. Additionally, we showcased the basic applications of our framework in event-discourse analysis, with plans for further investigation into its broader applicability in future work.

## Acknowledgements

We thank the anonymous reviewers of this paper for all of their vital feedback. The project was partially funded by NSF award IIS-2135573, and in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via 2022-22072200003. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

## Limitations and Ethics

We identify the following limitations from our study.

**Closed-source LLM.** As a first, the perspectives generated from our framework are based on an LLM model, ChatGPT, which is closed source, and the details of its construction are unclear (Spirling,

2023). We chose this model because it delivered higher quality output than other options available at the time.

**Talking point extraction.** Our framework allows the LLM to decide the key talking points from the news article, although it is possible that it could overlook a prominent talking point. While this is a potential limitation, we believe that if a talking point is really prominent, then it will repeat in many articles, to shape the narrative. Thus, there is a high chance that the LLM would identify that talking point in other articles, even if the model failed to recognize the prominent point in the given article.

**Labeling the talking point.** Further, we assume that all the talking points from a left-leaning news source are actually *left-biased*, and vice-versa. However, in reality, it need not be the case (Kim et al., 2022). Our approach performs fairly well, even with this assumption primarily because we are only interested in identifying salient talking points from each ideology, and less frequent talking points are rejected.

**Cautious use of our framework.** We acknowledge that our framework produces partisan viewpoints through a language model, which can occasionally generate inaccurate or misleading representations, often referred to as LLM-hallucinations. These inaccuracies may not be easily identified through automated evaluation methods. Consequently, we strongly advise users to seek expert validation before using our framework.

While our system offers numerous practical applications, we urge users to approach its usage with caution. Although it can effectively identify ideologies, there is a potential for misuse, such as targeting individuals based on their expressed beliefs or affiliations. For these reasons, as well as many others, it is crucial for users to thoroughly evaluate both the advantages and potential risks associated with deploying our framework.

**Ethics.** To the best of our knowledge, we did not violate any widely held ethical precepts when producing this paper. All results are from a machine learning model and should be interpreted as such. We present all implementation and dataset details for reproducibility of our study (some parts are in the appendix). The datasets and LLMs used in this paper are publicly available and allowed for scientific research.

## References

- Afra Feyza Akyürek, Lei Guo, Randa Elanwar, Prakash Ishwar, Margrit Betke, and Derry Tanti Wijaya. 2020. [Multi-label and multilingual news framing analysis](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8614–8624, Online. Association for Computational Linguistics.
- Amber E. Boydston, Dallas Card, Justin H. Gross, Paul Resnick, and Noah A. Smith. 2014. [Tracking the development of media frames within and across policy issues](#).
- Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. 2013. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 160–172. Springer.
- Dallas Card, Amber Boydston, Justin H Gross, Philip Resnik, and Noah A Smith. 2015a. 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.
- Dallas Card, Amber E. Boydston, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015b. [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.
- Dennis Chong and James N Druckman. 2007. Framing theory. *Annu. Rev. Polit. Sci.*, 10:103–126.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. [The llama 3 herd of models](#). *arXiv preprint arXiv:2407.21783*.
- Elizabeth Dubois and Grant Blank. 2018. The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5):729–745.
- Robert M. Entman. 1993. [Framing: Toward Clarification of a Fractured Paradigm](#). *Journal of Communication*, 43(4):51–58.
- Lisa Fan, Marshall White, Eva Sharma, Ruisi Su, Prafulla Kumar Choubey, Ruihong Huang, and Lu Wang. 2019. [In plain sight: Media bias through the lens of factual reporting](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6343–6349, Hong Kong, China. Association for Computational Linguistics.

- Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov. 2018. Framing and agenda-setting in russian news: a computational analysis of intricate political strategies. In *2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Lea Frermann, Jiatong Li, Shima Khanehazar, and Gosia Mikolajczak. 2023. [Conflicts, villains, resolutions: Towards models of narrative media framing](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics.
- Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the 2018 World Wide Web Conference*, pages 913–922. International World Wide Web Conferences Steering Committee.
- Matthew Gentzkow and Jesse M Shapiro. 2006. Media bias and reputation. *Journal of political Economy*, 114(2):280–316.
- Matthew Gentzkow and Jesse M Shapiro. 2011. Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839.
- Erving Goffman. 1974. *Frame analysis: An essay on the organization of experience*. Harvard University Press.
- Benjamin Horne, Mauricio Gruppi, and Sibel Adali. 2022. [NELA-GT-2021](#).
- Alexander Hoyle, Rupak Sarkar, Pranav Goel, and Philip Resnik. 2023. Natural language decompositions of implicit content enable better text representations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 13188–13214.
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.
- Kathleen Hall Jamieson, Bruce W Hardy, and Daniel Romer. 2007. The effectiveness of the press in serving the needs of american democracy.
- Yangfeng Ji and Noah A. Smith. 2017. [Neural discourse structure for text categorization](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 996–1005, Vancouver, Canada. Association for Computational Linguistics.
- Shima Khanehazar, Trevor Cohn, Gosia Mikolajczak, Andrew Turpin, and Lea Frermann. 2021. [Framing unpacked: A semi-supervised interpretable multi-view model of media frames](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2154–2166, Online. Association for Computational Linguistics.
- Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein, and Martin Potthast. 2019. [SemEval-2019 task 4: Hyperpartisan news detection](#). In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 829–839, Minneapolis, Minnesota, USA. Association for Computational Linguistics.
- Eunji Kim, Yphtach Lelkes, and Joshua McCrain. 2022. Measuring dynamic media bias. *Proceedings of the National Academy of Sciences*, 119(32):e2202197119.
- Donald R Kinder. 1998. Communication and opinion. *Annual review of political science*, 1(1):167–197.
- Andrea Lawlor and Erin Tolley. 2017. [Deciding who’s legitimate: News media framing of immigrants and refugees](#). *International Journal of Communication*, 11(0).
- Nayeon Lee, Yejin Bang, Tiezheng Yu, Andrea Madotto, and Pascale Fung. 2022. [Neus: Neutral multi-news summarization for mitigating framing bias](#).
- Chang Li and Dan Goldwasser. 2019. [Encoding social information with graph convolutional networks for Political perspective detection in news media](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2594–2604, Florence, Italy. Association for Computational Linguistics.
- Chang Li and Dan Goldwasser. 2021. Using social and linguistic information to adapt pretrained representations for political perspective identification. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4569–4579.
- Siyi Liu, Lei Guo, Kate Mays, Margrit Betke, and Derry Tanti Wijaya. 2019. [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.
- Siyi Liu, Hongming Zhang, Hongwei Wang, Kaiqiang Song, Dan Roth, and Dong Yu. 2023. [Open-domain event graph induction for mitigating framing bias](#).
- Y Liu, X Zhang, D Wegsman, N Beauchamp, and L Wang. 2022. Politics: Pretraining with same-story article comparison for ideology prediction and stance detection. *Findings of the Association for Computational Linguistics: NAACL 2022*.
- Yiwei Luo, Dallas Card, and Dan Jurafsky. 2020. Detecting stance in media on global warming. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3296–3315.

- Maxwell E McCombs and Donald L Shaw. 1972. The agenda-setting function of mass media. *Public opinion quarterly*, 36(2):176–187.
- 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.
- Davoud Moulavi, Pablo A Jaskowiak, Ricardo JGB Campello, Arthur Zimek, and Jörg Sander. 2014. Density-based clustering validation. In *Proceedings of the 2014 SIAM international conference on data mining*, pages 839–847. SIAM.
- Nishanth Nakshatri, Siyi Liu, Sihao Chen, Dan Roth, Dan Goldwasser, and Daniel Hopkins. 2023. Using llm for improving key event discovery: Temporal-guided news stream clustering with event summaries. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 4162–4173.
- Jianmo Ni, Chen Qu, Jing Lu, Zhuyun Dai, Gustavo Hernández Ábrego, Ji Ma, Vincent Y Zhao, Yi Luan, Keith B Hall, Ming-Wei Chang, et al. 2021. Large dual encoders are generalizable retrievers. *arXiv preprint arXiv:2112.07899*.
- OpenAI. 2022. [GPT-3.5 \(ChatGPT\)](#). Computer software.
- Kunwoo Park, Zhufeng Pan, and Jungseock Joo. 2021. Who blames or endorses whom? entity-to-entity directed sentiment extraction in news text. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4091–4102.
- Walter Quattrociocchi, Antonio Scala, and Cass R Sunstein. 2016. Echo chambers on facebook. *Available at SSRN 2795110*.
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2024. [Direct preference optimization: Your language model is secretly a reward model](#).
- Hannah Rashkin, Sameer Singh, and Yejin Choi. 2016. Connotation frames: A data-driven investigation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 311–321.
- Shamik Roy and Dan Goldwasser. 2020. Weakly supervised learning of nuanced frames for analyzing polarization in news media. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7698–7716.
- Shamik Roy, María Leonor Pacheco, and Dan Goldwasser. 2021. Identifying morality frames in political tweets using relational learning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9939–9958.
- Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cino Lee, Percy Liang, and Tatsunori Hashimoto. 2023. Whose opinions do language models reflect? In *International Conference on Machine Learning*, pages 29971–30004. PMLR.
- Dietram A Scheufele and David Tewksbury. 2007. Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of communication*, 57(1):9–20.
- Timo Spinde, Lada Rudnitskaia, Jelena Mitrović, Felix Hamborg, Michael Granitzer, Bela Gipp, and Karsten Donnay. 2021. [Automated identification of bias inducing words in news articles using linguistic and context-oriented features](#). *Information Processing & Management*, 58(3):102505.
- Arthur Spirling. 2023. Why open-source generative ai models are an ethical way forward for science. *Nature*, 616:413.
- Dominik Stammbach, Maria Antoniak, and Elliott Ash. 2022. [Heroes, villains, and victims, and GPT-3: Automated extraction of character roles without training data](#). In *Proceedings of the 4th Workshop of Narrative Understanding (WNU2022)*, pages 47–56, Seattle, United States. Association for Computational Linguistics.
- Teun A Van Dijk. 1996. Ideological discourse analysis. *MOARA–Revista Eletrônica do Programa de Pós-Graduação em Letras ISSN: 0104-0944*, (06):13–45.
- Teun A Van Dijk et al. 1997. What is political discourse analysis. *Belgian journal of linguistics*, 11(1):11–52.
- Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2024. Can large language models transform computational social science? *Computational Linguistics*, pages 1–55.



## A Extended Results

### A.1 Partisan Classification

Table 7 shows the partisan classification results for each issue and the total aggregated performance.

### A.2 Ideology Classification Task

**Narrative-LLAMA Quality Test.** We assess whether the perspectives generated by Narrative-LLAMA are actually expressed in the input news articles. To achieve this, we use an economical automated evaluation approach with a *LLAMA3-70B-instruct* model. Specifically, for all 481 unseen input articles, we first generate high-level ideological viewpoints using Narrative-LLAMA. We then prompt LLAMA3-70B-instruct to check if these viewpoints are represented in the corresponding news articles. We use two metrics for success: **hard** and **soft**. For each article, the **hard** metric scores 1 if *all* generated viewpoints are reflected in the article (indicating no hallucination). The **soft** metric calculates the overall proportion of viewpoints conveyed in the input articles. Table 8 presents these findings per issue, revealing that on an average, in 94% of cases, *all* generated perspectives are included in the input news articles, as evaluated by the 70B-instruct model.

### A.3 Evaluate clustering – PTP

We evaluate our framework’s ability to effectively cluster the talking points to form PTPs using two metrics – *coverage*, and *topic diversity*.

**Coverage.** We define coverage as the proportion of talking points included in the clustering process. Ideally, achieving 100% coverage would mean that every input talking point is fully accounted for when generating PTPs. Higher coverage is desirable as it ensures a more comprehensive representation of the input. Table 9 presents the average coverage for each issue. Our results indicate that the identified PTPs consistently cover at least 80% of the input talking points, demonstrating that they effectively represent the original set  $\mathcal{T}$ .

**Topic Diversity.** Ideally, we want every PTP to be unique, as this would indicate good cluster separation. To this end, we formulate *topic classification task*: Given a talking point and a set of  $K'$  PTP labels, assign the talking point to the most topically relevant PTP  $k^*$ , where  $k^* \in K'$ . Note that the talking point is associated with only one of the  $K'$  labels, and the rest of the labels are randomly

sampled negative examples. In this case,  $k^*$  helps assess how well the talking point assignments map to their respective clusters, whereas the remaining negative labels,  $K' \setminus \{k^*\}$ , help measure the degree of separation between the clusters.

**Setup.** We first split the data in each PTP cluster into 4 quartiles, where the 1st quartile refers to the top 25% closest talking points to a PTP label in the embedding space, the 2nd quartile the top 50%, etc. We randomly sample half the talking points from each quartile for this experiment. We include 3 negative labels along with a correct label for each talking point ( $|K'| = 4$ ). We prompt ChatGPT to assign the talking point to its most topically relevant label.

**Discussion.** Table 10 shows the performance of the *topic classification task*. We see that all quartiles have a decent performance, and the documents closer to the PTP show strongest topical relevance to the PTP. A strong performance from  $Q4$  suggests PTPs capture diverse topics, and each PTP captures a unique aspect when compared to others.

Further, on comparing both coverage and topic diversity, we observe that PTPs are diverse and span at least 80% of the input talking point set. This suggests that our approach forms reasonable set of PTPs to create *partisan perspectives*.

### A.4 Human Evaluation

We perform human evaluation on a set of 3 events. We randomly choose an event from three different issues - {Coronavirus, Climate Change, Immigration}. We annotate a total of 84 PTPs, and describe the annotation procedure for each metric – *summary coherence*, and *mapping quality*. Note that our annotators were graduate STEM-students who were not the authors of the paper and were under the age of 30.

**Summary Coherence.** For this metric, we compare the *left*- and *right*- interpretations of each PTP label against the elements that were used to construct it. Specifically, we compare these interpretations against the top- $K$  talking points and news article summaries that were used to construct them.

We explain the procedure for the *left* political ideology, and the same process is repeated for the *right* ideology as well. First, we explain the task to the annotators by providing an example. Annotators are provided with a *left-leaning* viewpoint along with three to five left-biased talking points and their respective news article summaries. We

Issue	Approach	Avg. Precision	Avg. Recall	Avg. F1-score
Climate Change	Topically Relevant Points	84.11	84.23	84.17
	Partisan View	91.73	89.46	90.29
	Partisan View + Metadata	92.43	90.86	<b>91.49</b>
Capitol Insurrection	Topically Relevant Points	69.50	71.62	69.18
	Partisan View	79.33	80.93	<b>79.93</b>
	Partisan View + Metadata	81.04	78.08	79.12
Immigration	Topically Relevant Points	69.14	74.64	69.92
	Partisan View	85.38	86.36	85.85
	Partisan View + Metadata	88.27	86.17	<b>87.15</b>
Coronavirus	Topically Relevant Points	73.11	72.60	72.77
	Partisan View	83.34	81.76	82.21
	Partisan View + Metadata	83.78	84.20	<b>83.92</b>
<b>Overall Performance</b>	Topically Relevant Points	73.44	73.33	73.37
	Partisan View	85.03	84.61	84.76
	Partisan View + Metadata	85.93	86.14	<b>85.98</b>

Table 7: Averaged results for *partisan classification task* shows the efficacy of partisan perspectives in capturing ideology-specific information at the cluster-level granularity.

<p><b>LLAMA:</b></p> <p>1. Joe Biden's Executive Orders: The article reports that President-elect Joe Biden plans to sign a raft of Executive Orders on his first day in office, covering issues such as COVID-19, climate change, racial equity, and economic recovery. These actions aim to address the four crises identified by Biden's transition team.</p> <p>2. Biden's Priorities: The article highlights Biden's focus on addressing the COVID-19 crisis, economic recovery, climate change, and racial equity. His administration plans to take decisive action to deliver relief to millions of Americans struggling with these crises.</p>	
<p><b>ChatGPT:</b></p> <p>1. Biden's Executive Orders: President-elect Joe Biden plans to sign several Executive Orders during his first days in office, addressing key issues such as COVID-19, climate change, student loan relief, and reversing Trump-era policies.</p> <p>2. Focus on Four Major Crises: Biden's administration will prioritize tackling four major crises—COVID-19, economic recovery, climate change, and racial equity—through immediate executive actions.</p>	
<p><b>Narrative-LLAMA:</b></p> <p>1. Biden's Executive Order Overreach: Biden's Use of Executive Orders to Implement His Agenda without Congressional Input\nThe article criticizes President-elect Biden's plan to implement his agenda through Executive Orders, bypassing Congress and potentially reversing Trump's policies, highlighting concerns about Biden's willingness to circumvent legislative process and the potential impact on his ability to implement his agenda effectively.</p>	

Figure 5: Shows an example of the generated perspectives for a *right-leaning* news article related to *Biden's executive orders*. The generated perspectives from Narrative-LLAMA capture broader ideological discourse as opposed to its counterparts.

Issue Name	% Success Rate	
	Soft	Hard
Climate Change	97.70	93.53
Capitol Insurrection	95.97	92.15
Immigration	98.28	95.73
Coronavirus	98.39	95.06

Table 8: LLAMA-3-70B-instruct evaluates issue-specific mapping quality, finding that on average, 94% of *all* the generated viewpoints are found in the input news articles.

ask the annotators to validate if the *left-leaning* viewpoint can be derived either from the news ar-

Issue	Avg. Coverage per event	Avg. # clusters
Climate Change	83.17	10
Capitol Insurrection	86.70	24
Immigration	90.55	21
Coronavirus	78.18	16

Table 9: Averaged results for coverage.

ticle summaries or the talking points. If it can be derived, then the response is 1, otherwise it is 0. In the cases where the annotators are not sure, the response is -1.

**Mapping Quality.** For this metric, we compare *left-* and *right-* interpretations of each PTP label

Issue	Q1	Q2	Q3	Q4
Climate Change	91.19	87.47	83.66	80.00
Capitol Insurrection	91.78	89.34	84.56	80.27
Immigration	91.96	88.69	85.01	80.34
Coronavirus	94.07	89.11	84.10	79.94
<b>Avg. Accuracy</b>	<b>92.74</b>	<b>88.90</b>	<b>84.37</b>	<b>80.12</b>

Table 10: Averaged results for each quartile for the *topic classification task* indicates that our prominent talking points capture diverse information.

against the news articles to determine if they are expressed in the news articles.

We explain the procedure for the *left* political ideology, and the same process is repeated for the *right* political ideology as well. In this case, we provide the annotators with a *left-leaning* viewpoint and a corresponding news article that is the most relevant to the viewpoint (measured based on cosine similarity distance in the embedding space). We segment the news article into sentences of 7, and we only provide the most relevant 7 sentences from the news article to the annotators. First, we let the annotators know that there are at most three bullet points in the provided *left-leaning* viewpoint. Then, we ask them to compare the viewpoint with the news article excerpt to validate if at least one of the points in the summary is expressed in the article. If it is, then the response is 1, otherwise it is 0. In cases where annotators are not sure, the response is  $-1$ .

**Mapping Quality – MQ\_LLM.** For this metric, human annotators validate the extracted evidence from an LLM. For the climate change event shown Table 11, we examine the PTP clusters. From each PTP cluster and for each ideological-viewpoint, we randomly selected up to 5 news articles, resulting in 92 article-viewpoint pairs (comprising 48 *left* pairs and 44 *right* pairs). For each pair, GPT-4o was tasked with quoting relevant sentences from the articles to answer specific questions. Note that in the following questions, *summary* denotes the *viewpoints*.

1. Is the summary discussing the same topic as the news article?
2. In the summary and the news article, are there any entities in common that are viewed negatively from the same perspective?
3. In the summary and the news article, are there any entities in common that are viewed positively from the same perspective?
4. Does the news article cover the views presented in the summary from the same angle?

Then, for each article-viewpoint pair, we present the viewpoints and the extracted evidence from the article to a human. Human is expected to validate whether the retrieved evidence aligns with the summary. Note that we proposed this experiment as an alternative evaluation scheme for mapping quality, and we limit the scale to one event considering the API cost associated with GPT-4o model.

### A.5 Visualizing Partisan Discourse in Events

To create *event-discourse snapshots*, we would need to obtain agreement and disagreement between the *left*- and *right*- interpretations of each PTP. To obtain this, we define a scale that aids in characterizing agreement between the *left*- and *right*- viewpoints of a PTP. We begin by prompting GPT-4o to assign a binary label  $-0/1$ , for each of the following questions. Note that in the following questions, *summary* denotes the *viewpoints*.

1. Do both summaries have at least one common aspect of discussion?
2. Are the summaries discussing about similar entities?
3. Are the entities in common viewed in the same manner? For example, is the entity viewed positively or negatively in both the summaries?
4. Do both the summaries talk about the event from the same perspective?
5. If the summaries are viewing the event from different angles, do the summaries have atleast some agreement with each other?

We obtained 2 PTPs with a cumulative score of 1; 5 PTPs with a score of 2; 8 PTPs with a score of 3; 4 PTPs with a score of 4; and 3 PTPs with a cumulative score of 5. We note that higher scores indicate that *left*- and *right*- interpretations of PTPs are closer to being in agreement with each other. Lower scores imply that these are mostly disagreeing with each other. For a score of 3, we manually inspected the outputs from the model and deduced that the two summaries shared a common aspect, discussed similar entities and had some agreements with each other. However, the entities were not viewed in the same manner due to which we assigned the ideological interpretations PTPs to disagree with each other.

## B Prompt Templates

Figures 6, 7, 8 shows all the prompt templates used in our work.

Issue	News Event
Climate Change	<b>Event Title:</b> Biden Announces Ambitious Greenhouse Gas Emissions Cut <b>Event Description:</b> This is about President Joe Biden’s announcement of an ambitious cut in greenhouse gas emissions as he looks to put the US back at the center of the global effort to address the climate crisis and curb carbon emissions.
Coronavirus	<b>Event Title:</b> Biden’s COVID-19 Vaccination Mandate <b>Event Description:</b> This is about President Joe Biden’s announcement of new COVID-19 vaccination requirements for federal government employees, healthcare workers, and companies with 100 or more employees, and his criticism of politicians who are undermining trust in COVID vaccines.
Immigration	<b>Event Title:</b> Biden’s Refugee Cap Decision <b>Event Description:</b> This is about the criticism faced by President Biden for his decision to not raise the cap on refugees allowed to enter the US this year, which he had promised to do during his campaign.

Table 11: Events considered for human evaluation.

Right-leaning viewpoints	Evidence From Article
Uncertainty in global cooperation and skepticism towards US leadership. Concerns persist over the uncertainty of international support, especially from major carbon emitters like China, India, and Russia, towards America’s climate initiatives. Differing views on the urgency of climate action and skepticism towards US leadership may hinder effective global collaboration on climate change.	Evidence: Both the viewpoints and the news article mention skepticism towards US leadership and the challenges in global cooperation. The summary states, "Concerns persist over the uncertainty of international support, especially from major carbon emitters like China, India, and Russia, towards America’s climate initiatives." The news article similarly notes, "Russian President Vladimir Putin and Chinese President Xi Jinping are two notable leaders who have both confirmed their attendance at the summit, underscoring the wide range of leaders attending," indicating the importance of their participation and potential skepticism.

Table 12: GPT-4o fails to correctly identify the relevant evidence from the news article. The negative sentiment expressed towards the countries in the viewpoints are not extracted as evidence from the news articles.

## C Experiments Related

### C.1 Dataset extraction

Here, we describe the process used for extracting the set of *unseen* news articles. We note that (Nakshatri et al., 2023) used NELA-2021 dataset for segmenting the news articles into a set of temporally motivated news events. In this process, (Nakshatri et al., 2023) used a temporal window of 3 in order to obtain coherent news events.

In order to obtain *unseen* news articles, yet relevant to the events under consideration, we extend this temporal window to 7 days, and retrieve all the news articles for that time period from NELA-2021 dataset. We filter out all the news articles that part of our clustering process. Then, we consider the all the unseen articles that are closest to the event centroid in the embedding space (threshold  $\geq 0.86$ ). Note that we obtain event centroid by averaging the embedding of all news articles relevant to the event. In this way, we extracted 481 relevant news articles

for the events under consideration, of which 234 news articles are from right-leaning news sources, and the rest are from the left-leaning news sources.

### C.2 Narrative-LLAMA Training Details

Our goal is to develop a model capable of capturing higher-level ideological views within news articles, rather than focusing on minor details that are not apparent in the broader event discourse. To eliminate these minor details and enable the model to focus on the event-level ideological discourse, we train an open-sourced LLAMA3-8B model with the partisan viewpoints generated by our proposed framework. Specifically, we fine-tune the LLAMA-8B-instruct model using news article as input, and its corresponding partisan viewpoint as output. In this process, we observe that the partisan perspectives provided in training are from our framework and generated by ChatGPT, a larger LLM with superior reasoning ability compared to LLAMA3-8B. Therefore, in a way, we are distill-



Categorization	PTP ID	Left-leaning Viewpoints (only titles)	Right-leaning Viewpoints (only titles)
Agreement	1	Rejection of splitting COVID relief bill into separate components	Resistance to breaking down relief package into separate bills
Disagreement	10	Emphasis on Transparency and Improved Vaccine Distribution	Questioning Biden's Vaccine Distribution Transparency
Agenda Setting	16	Biden's Travel Restrictions and Bans for Public Health	Criticism of Biden's Executive Order on Pandemic Language (for banning term - 'China Virus')
Partisan Battle	2	Biden administration's emphasis on equitable vaccine distribution and healthcare reform	Criticism of Biden administration's vaccine distribution decisions
Right Only	3	-	Economic Impact of \$15 Minimum Wage

Table 13: Shows an example of PTP categorization for the PTP IDs shown in Figure 4. We show the *left-leaning* and *right-leaning* viewpoints for each PTP.

Discussion - maintains semantic relevance to the article	You are tasked with discerning the key talking points from the following 'NEWS ARTICLE'. Your objective is to condense the contents of the 'NEWS ARTICLE' into a succinct list of up to four primary talking points, each accompanied by a brief description. 'NEWS ARTICLE': <article>
Entities	For each 'Talking Point', identify up to three pivotal entities associated with it, and assess whether the author of the 'NEWS ARTICLE' exhibits a bias either against or in favor of the mentioned entities. In instances where no discernible bias is evident, categorize the entity as neutral. The goal is to compile a list of entities along with their entity types, categorizing them as either against, in favor, or neutral, and accompany each categorization with a brief explanation.
Activity	For each 'Talking Point' and its associated 'Entities', identify the primary activities linked to it. For every identified 'Activity', pinpoint the entity assuming the role of 'Actor', driving the said 'Activity', and the entity acting as the 'Target', which is influenced by the 'Actor'. Assess whether the impact on the 'Target' is positive, negative, or neutral, providing a rationale for the impact. Focus only on pivotal 'Activities' closely related to the 'Talking Point'.
Media Frame	For each 'Talking Point' and its corresponding 'Activity', predict its media frame, and categorize it into one of 15 labels: Economic, Capacity & Resources, Morality, Fairness & Equality, Legality, Policy, Crime, Security, Health, Quality of Life, Cultural, Public Opinion, Political, External Regulation, or Other. With respect to the predicted 'Frame', provide a short explanation on how it is related to the main 'Activity'.

Figure 6: Illustrates the prompt template used to derive our relational structure, referred to as a *talking point*. We begin by prompting an LLM to extract the key discussion elements — repurposing the general notion of a "talking point" to identify the main topics within a news article. We then enrich this information by contextualizing it with relevant entities, their associated roles (via activity), and the surrounding media frame. In our work, this contextualized relational structure is what we define as a *talking point*.

You are provided with an aspect of discussion related to a news event, along with biased talking points from left and right political ideologies discussing the same aspect. Each talking point is associated with its respective news article summary, and metadata that includes actions, actors, targets, impacts, and framing. On comparing and analyzing the talking points from both ideologies, the objective is to refine and condense left-biased talking points into at most three unique points, such that the new points clearly capture the political bias towards left ideology. Redundant points and those not aligning well with left political ideology should be excluded.
#### Input #### Aspect of discussion: (aspect) Left-biased talking points: (left-biased points) ## includes metadata for each point Left-biased news article summaries: (left-biased summaries) Right-biased talking points: (right-biased points)

Figure 7: Shows the prompt template used to generate *left-leaning* viewpoints for a PTP. *Right-leaning* viewpoints can be obtained in a similar manner.

ing knowledge from a larger model (ChatGPT) to a smaller one (LLAMA3-8B), thereby instructing the smaller model to abstract away minor details and concentrate on event-level narratives.

**Dataset.** We select a portion of the dataset from our framework to fine-tune the model. Specifically, for each event, we consider the clusters from Section 3.2.2 and extract news articles linked to the talking points in each cluster, that are the most

Given a segment of a 'news article' from a (ideology)-biased media source and two summaries derived from the same news event, your task is to perform binary classification by assigning the news article to one of the two summaries. Each summary has a set of talking points about the event. Each summary is also accompanied by metadata that includes frequently occurring actors, targets, sentiment on the target entities, and media frame associated with the context of the talking points. Your goal is to use the associated metadata to better determine if the provided news article segment has a viewpoint that is more similar to 'summary1' or 'summary2'. The response should strictly be 'summary1' when the 'news article' segment has a consistent viewpoint with 'summary1'; otherwise, it should be 'summary2', indicating the 'news article' has consistent viewpoint with 'summary2'. Refrain from providing any additional information.
### Input to analyze ### 'News article': (article) 'Summary1': (summl) 'Summary1 Metadata': (summlMetadata1) 'Summary2': (sum2) 'Summary2 Metadata': (sum2Metadata2)

Figure 8: Shows the prompt template used for partisan classification task.

nearest to their respective ideology-specific views of that cluster. We take the top 25% of these articles to construct a dataset of 1100 examples, with the news articles as inputs and their corresponding partisan viewpoints as outputs. Each input-output pair includes a news article featuring two points of view: one matching its ideological label and the other from an opposing ideology, which is crucial for training the LLM using preference optimization.

**Training.** We follow a two-stage training process. First, we start by instruction-tuning the model using SFT, where the news article is provided as the input and the viewpoints matching the ideological label of the article serves as the output. This enables the model to adjust its output probability distribution to focus on event-level discourse. For this, we train the model using PEFT and LORA adapter parameters for 1 epoch (with hyperparameters set to  $r = 64, \alpha = 16, learning-rate = 1e^{-4}$ ). Next, we train the model using Direct Preference Optimization (DPO) for 1 epoch, to align the model's generated outputs to prefer the viewpoints that match the ideology label of the article, while contrasting them with the viewpoints from the opposing ideology (hyperparameters:  $r = 64, \alpha = 16, learning-rate = 5e^{-5}$ ). We fine-tuned the model (16-bit precision) using 8 Tesla V100 GPUs with 32GB VRAM each, and the training process took about 3 hours, for a batch size of 2 per device.

**Prompt to characterize a cluster:**  
 Given a set of news article excerpts taken from the same news event, the task is to analyze the articles with the intent to identify a high-level concept that captures the key aspect of discussion related to that event. The concept should be indicative of one of the main discussion angles related to the event, and not very specific to entities mentioned in the articles. The concept should be accompanied by a summary, which should not be a mere concatenation of articles.

**Prompt to Merge two clusters:**  
 Given two aspects from the same news event, you need to analyze them with the intent to understand if they are identical. You should compare the key emphasis of the aspects and their implications to decide if they are the 'same', or 'different'. Refrain from providing any additional explanations other than the label.

**Prompt to remove inconsistent clusters:**  
 You are provided with a few news article excerpts and a key aspect of discussion, all of which are from the same news event. The task is to analyze if all of the provided news article excerpts are discussing the given key aspect. Respond with 'yes' if the central theme of discussion in each excerpt align in meaning with the key aspect, and 'no' if there is any variance, refraining from offering any additional explanation.

Figure 9: Shows the prompt template used to characterize a cluster, remove redundant, and incoherent clusters.

## D Clustering the talking points

As described in 3.2.1, we cluster the talking points to identify the PTPs. In this process, we merge redundant clusters and remove incoherent clusters. The details of this process is outlined in this section. Figure 9 shows the prompt templates for the same.

### D.1 Merge Redundant Clusters

In order to merge redundant clusters, we devise a simple greedy algorithm. We consider pairwise combinations of cluster labels, and prompt the LLM to verify if the clusters can be merged based on the prompt template shown in Figure 9.

We start by constructing a set of pairwise cluster labels  $\mathcal{S} = \{(s_1, s_2), \dots, (s_{n-1}, s_n)\}$ . Precisely, for every cluster, we consider all possible pairwise combinations for the top-7 closest labels to that cluster in the embedding space. For each element in  $\mathcal{S}$ , we prompt LLM to infer if the pair of labels are discussing about the same aspect. If the aspects, say  $(s_1, s_2)$ , are equivalent, then we merge these aspects, and update the set  $\mathcal{S}$  by removing every element in the set that contains  $s_1$  or  $s_2$ . In the second iteration, we construct a new set,  $\mathcal{S}'$ , that holds every combination of updated cluster labels, and repeat the previous step. We run the algorithm for two iterations or halt if there are no merges after the first iteration.

### D.2 Remove Incoherent Clusters

We note that HDBSCAN algorithm provides us with an initial set of candidate clusters. For each candidate, we use the aspect associated with the cluster label to validate if the top-3 members that are closest to the cluster label in the embedding space are discussing the same high-level concept. We prompt the LLM using the prompt shown in Figure 9 to remove incoherent clusters.

### D.3 Talking Point Membership

After obtaining the cluster labels, which characterize the space of possible talking points. We consider each talking point from the set of all the talking points and assign the closest cluster label based on cosine similarity score. If this score is beyond a threshold value of 0.85, we assign the talking point to that cluster label. Otherwise, it is discarded but retained in the unclustered pool of talking points.

### D.4 Clustering-related Hyperparameters

Note that we are interested in identifying the dense regions in the embedding space associated with talking points, as these are the potential candidate topic indicators. Due to this, we choose HDBSCAN method as our clustering algorithm, which does not require any prior number of clusters. However, we are still required to tune a few hyperparameters in order to obtain a decent performance. We use a data-driven approach to estimate the best number of topics by maximizing the DBCV score (Moulavi et al. (2014)). We retain the default settings for *cluster\_selection\_method*, and *metric\_parameters*, while we change the *min\_cluster\_size* and *min\_samples* to get more sensible topics. This number is selected based on a grid search whose values are sensitive to the number of input talking points. Suppose  $|X|$  denote the number of talking points, then the grid parameters for HDBSCAN used in our method include  $5, 7, 9, 0.01 * |X|, 0.02 * |X|, \dots 0.04 * |X|$ .

For our algorithm’s talking point membership module, we choose a similarity threshold of 0.76 based on manually inspecting the prominent talking points, outputs for the cluster redundancy and removal of cluster incoherence operations for 3 events related to the issue - *Climate Change*.

## E Examples

Figure 10 shows an example where the generated viewpoints are incorrect — they do not align with the *right* ideology views.

## F Temporal Analysis - Case Study

Here, we provide a simple case study to show how the data obtained from our framework could be utilized to study the *left* and *right* perspectives for an entire issue. To do this, we consider 7 events related to the issue *Coronavirus* at various points in time,

Right-leaning viewpoints	<p>81: Title: Criticism of Biden's vaccine mandate as dictatorial Description: Dr. Fauci's characterization of President Biden's vaccine mandate as 'moderate' is challenged by right-leaning perspectives, labeling the mandate as dictatorial and overly lenient, emphasizing concerns about individual liberty being compromised.</p> <p>82: Title: Opposition to perceived leniency in vaccine mandate Description: Right-leaning voices critique Dr. Fauci and President Biden for what they view as a lenient approach to the vaccine mandate, supporting a need for stricter measures to ensure compliance and public health protection.</p> <p>83: Title: Advocating for a stricter vaccination-only policy Description: Right-leaning perspectives highlight a preference for a stricter vaccination-only policy over the compromise of allowing testing as an alternative in President Biden's vaccine mandate, emphasizing the importance of prioritizing vaccination for public health benefits.</p>
Right-biased Talking Points	<p>1. Title: President Biden's vaccine mandate is considered somewhat 'moderate' by Dr. Anthony Fauci. Description: Dr. Fauci believes that President Biden's vaccine mandate strikes a balance between the need for vaccination and the concerns of those who do not want to get vaccinated but also do not want to lose their jobs.</p> <p>2. Title: Dr. Fauci's stance on the vaccine mandate. Description: Dr. Fauci is quoted as saying that the President's approach is somewhat moderate and a compromise, but the article disagrees, stating that Fauci's true stance on vaccine mandates is even less moderate.</p> <p>3. Title: President Biden's vaccine mandates. Description: Dr. Fauci describes President Biden's vaccine mandates as a moderate approach, allowing for testing as an alternative to vaccination for those who are unwilling to get vaccinated.</p>
Associated news article summaries	<p>1. Dr. Fauci describes Biden's vaccine mandate as 'moderate' on CNN, emphasizing the need for options for those hesitant to get vaccinated. Republican governors criticize the mandate as an infringement on individual liberty and plan to challenge it in court. The mandate's impact on those previously infected with COVID sparks debate, with Fauci acknowledging the complexity of the issue.</p> <p>2. A right-leaning article criticizes Dr. Fauci for supporting what they view as President Biden's overly lenient vaccine mandate approach, highlighting Fauci's perceived lack of stringency and labeling Biden's actions as dictatorial, while emphasizing the need to expose the true intentions of political figures like Fauci and Biden.</p> <p>3. Dr. Fauci characterizes President Biden's vaccine mandate as moderate, emphasizing the option for testing as a compromise for those hesitant to get vaccinated, reflecting a right-leaning perspective on the level of stringency in vaccine mandates.</p>

Figure 10: Shows a negative example. The generated *right-leaning* viewpoints are incorrect. This is primarily attributed to inconsistent news article summaries (2 and 3), and LLM's failure to identify cited information within the news article.

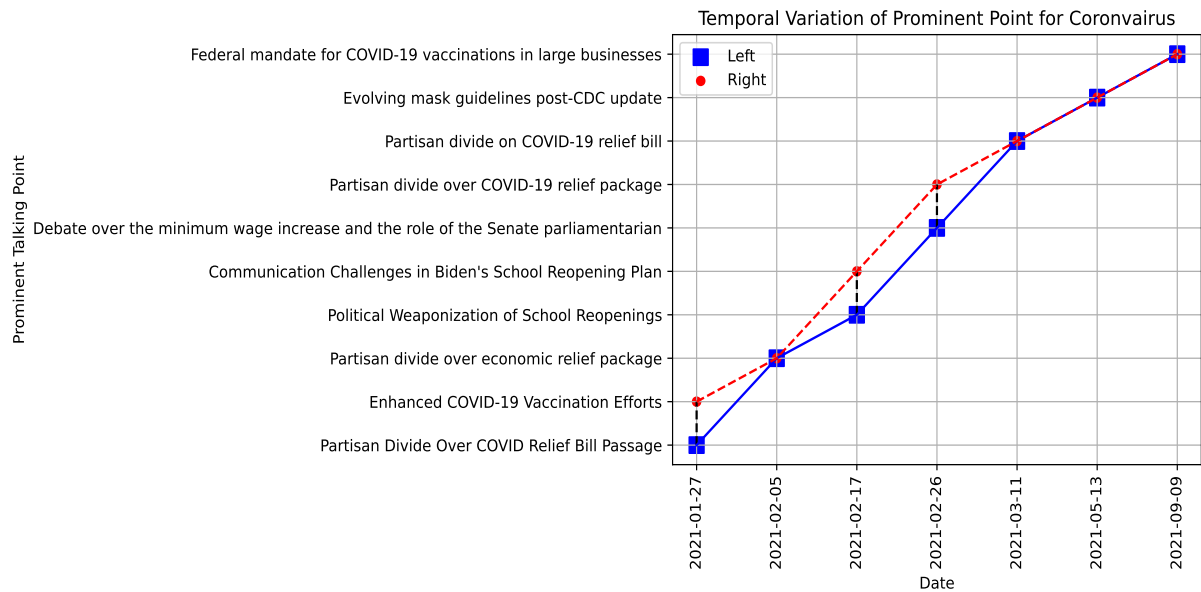
and at every point, analyze the most frequently repeating prominent talking point from each political ideology.

Figure 11(a) shows a dynamic evolution of prominent talking point of each political party for the issue - *Coronavirus*. We observe that frequently discussed prominent point of each political party is different from one another in 3 out of the 7 events under consideration. However, both political parties predominantly discuss the same prominent point in the remaining cases. Note that Figure 11(a) shows only the *aspect* associated with each prominent point for data visualization clarity.

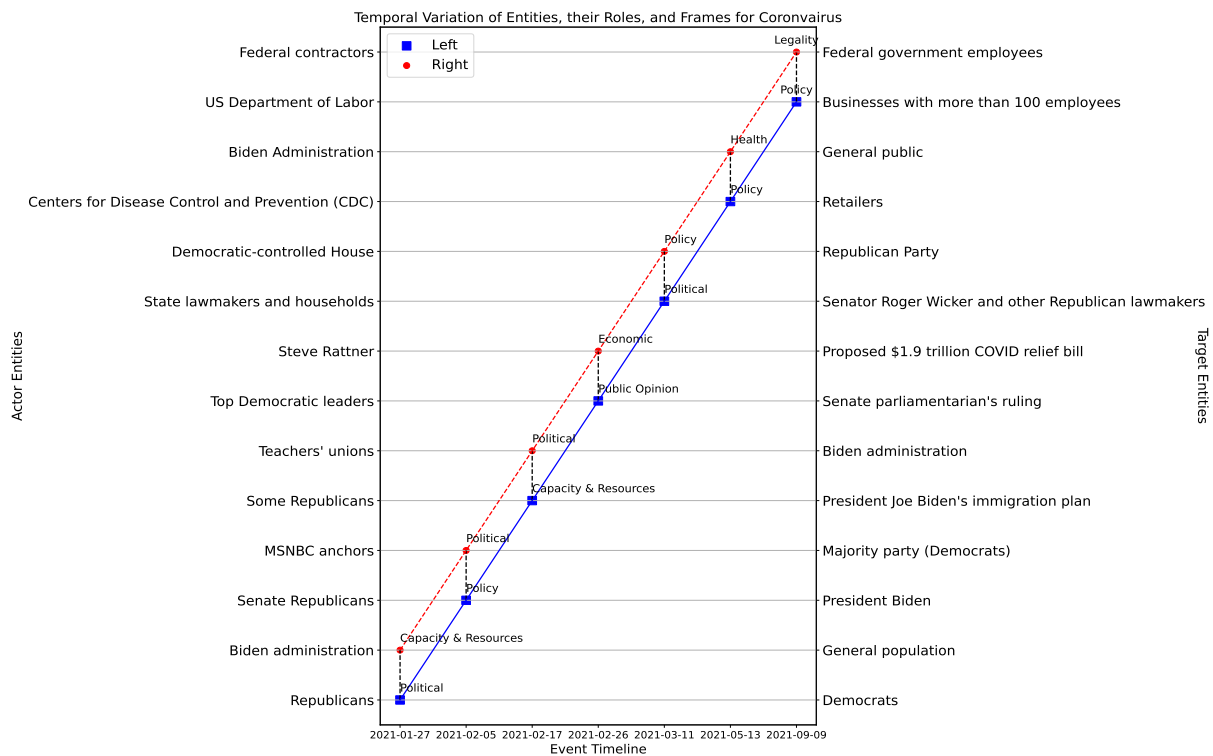
In the cases where both political parties discuss the same prominent point, we can further investigate the manner in which they talk about the prominent point by observing its corresponding partisan summary (*left-* and *right-leaning* viewpoints). For instance, let us consider the prominent point with the aspect - *Evolving mask guidelines post-CDC update*, that is commonly discussed by both political parties. While both the parties criticize the ambiguity in CDC's mask guidance, the *left-leaning* articles emphasizes more on pointing out the discrepancies with state and local mandates, and how it is impacting businesses. However, *right-leaning* sources focus on delayed response by CDC in updating mask mandates for vaccinated individuals and raises concern about the leadership.

We can further analyze this prominent point discussed by both parties through its associated metadata. The entity viewed as a *target* by an ideology, its corresponding *actor*, and the associated *media frame* can help analyze the differences in the viewpoints across political parties. For the same prominent point with the aspect *Evolving mask guidelines post-CDC update*, we observed that *left-leaning*

news sources viewed the entity *Centers for Disease Control and Prevention (CDC)* to have negatively impacted the target entity *Retailers*. Further investigation revealed that it was due to the criticism associated with changing mask guidelines, where *CDC* removed mask mandates for the vaccinated individuals, and left-leaning sources criticized *CDC* for creating ambiguity amongst the *retailers* regarding the mask guidelines. We note that *left-leaning* news sources commonly used *Policy* as the media frame of discussion in the context of this actor-target pair. In this way, the metadata associated with the prominent point of interest can further help distinguish left and right perspectives. To obtain an overall global view of variation in metadata for the entire issue, Figure 11 (b) shows a dynamic analysis over the actor/target entities for each prominent point across the two political parties over time.



(a) Compares the temporal variation of most frequent prominent point for each political ideology, and across 7 events related to the issue - *Coronavirus*. X-axis indicates event-timeline, and Y-axis shows the most prominent talking point label. Frequently discussed prominent point across the two ideologies intersect in 4 out of 7 cases.



(b) Illustrates the temporal variation in the metadata formatted as *actor-negatively impacting-target*. We show frequently repeating target entity with a negative sentiment for each ideology, and across 7 events for the issue - *Coronavirus*. For each target entity, its corresponding actor, and the associated media frame is also shown.

Figure 11: Temporal analysis of prominent points along with its respective metadata for the issue - *Coronavirus*.