All Things Considered: Detecting Partisan Events from News Media with Cross-Article Comparison

Anonymous ACL submission

Abstract

Public opinion is shaped by the information news media provide, and that information may be shaped by ideological leaning of media outlets. While much attention has been devoted to media bias via overt ideological language, a more profound way the media shape opinion is via the strategic inclusion or omission of partisan events that may support one side or the other. We develop a latent variable model to predict news articles’ ideology and identify partisan events by same-story article comparison. Our study validates the existence of partisan event selection and shows that cross-document comparison helps detect partisan events and article ideology. Our results reveal a high-level form of media bias, which is present even among mainstream media with strong norms of objectivity and nonpartisanship.

1 Introduction

News media play a critical role in society not merely by supplying information, but also by selecting and shaping the content they report (DellaVigna and Kaplan, 2007; DellaVigna and Gentzkow, 2009; Perse and Lambe, 2016). To understand how media bias affects media consumers (Gentzkow and Shapiro, 2006), we must understand not just how media ideology affects the presentation of news stories on a surface level such as the usage of partisan phrases, but also the less obvious process of content selection (Fan et al., 2019; Enke, 2020).

There are three common strategies to sway readers (Broockman and Kalla, 2022). Agenda setting (McCombs and Shaw, 1972) refers to when the public perception of a topic’s significance is shaped by the amount of news coverage (Grimmer, 2010; Quinn et al., 2010; Field et al., 2018). Framing concerns highlighting some aspects of the same reality to make them more salient to the public (Entman, 1993; Baumer et al., 2015; Card et al., 2015; Liu et al., 2019a). Partisan coverage filtering is when media selectively report content that is flattering to their copartisans. Content selection has recently become a focus in political science. However, existing work either requires manual inspection (Broockman and Kalla, 2022), or relies on simple tools for coarse analyses, such as overall slant and topic (Baum and Groeling, 2008; Grossman et al., 2022). As such, these studies are limited to a short time period and unable to provide a detailed understanding of content selection bias.

In this paper, we investigate into partisan coverage filtering: how media ideology affects their selection of which events to include for news reporting. In line with Broockman and Kalla (2022), we define partisan events as selectively reported events that are flattering to copartisans or unfavorable to opponents. Among many relevant events, which subset is reported fundamentally affects how readers interpret the story and can reveal a media outlet’s stance and ideology (Mullainathan and Shleifer, 2005; McCombs and Reynolds, 2008; Entman, 2007). One example of event-selection bias is shown in Fig. 1, where a Washington Post article includes a survivor’s request to impose gun control (pro-gun control), whereas a New York Post article claims Biden’s statement as false (pro-gun rights).

This paper has two major goals: (1) examine
the relation between event selection and media ideology, and (2) formulate a task for partisan event detection in news and develop computational methods to automate the process. For the first goal, we verify the existence of partisan event selection by measuring how event selection affects the performance of media ideology prediction. Specifically, we denote articles by the set of reported events and entities (Fig. 1). We conduct two studies. We first compare article-level ideology prediction performance by using events within a single article vs. contextualizing them with events in other articles on the same story but reported by media of different ideologies. We show that the latter setup yields higher F1 scores, suggesting that cross-article comparison can identify partisan events and thus produce better ideology prediction. Second, we annotate an evaluation dataset of 50 articles that around two political issues. Using this dataset, we show that removing partisan events from the articles hurts ideology prediction significantly more than removing similar amounts of randomly selected events. For the second goal, we use latent variables to represent if an event is partisan or not, and propose to jointly infer partisan events and an article’s ideology. We adopt Chen et al. (2018) and explore two methods for further improvement: (1) steering the model toward events that are selected only by one side, which are more likely to be partisan, and (2) providing prior knowledge about event ideology. We conduct experiments on two news article datasets (Liu et al., 2022; Fan et al., 2019) and our evaluation-only data with newly annotated partisan events. Results indicate that latent variable models outperform all competitive baselines on both partisan event detection and ideology prediction, where cross-article comparison is shown to be critical for both tasks. To the best of our knowledge, we are the first to develop computational methods for studying media bias at the event selection level. Our results provide insights into a high-level form of media bias that may even exist in apparently nonpartisan news, enabling new understanding of how media content is produced and shaped.

2 Event Selection Effect Study

We verify the existence of partisan event selection by examining its influence on ideology prediction. We design a model that predicts ideology using events (§2.1), based on the assumption that comparing events included by different media may reveal their ideological leanings. On a manually annotated dataset with partisan events in news stories (§2.2), we show that cross-article content comparison can reveal potential partisan events and removing partisan events hurts ideology prediction (§2.3).

2.1 Ideology Prediction with Events

We extend the narrative embedding model (Wilner et al., 2021) to include story level context by adding article segment, event frequency, and event position signals. This enables gauging the effect of partisan events’ presence/absence on ideology prediction.

Given \( N \) input articles that report on the same news story, we use an event extraction model to extract events in \( i \)th article as \( \{x_1^{(i)}, \ldots, x_{L_i}^{(i)}\} \), which are encoded as \( e \) by DistilRoBERTa (Sanh et al., 2019). Events in one article or all articles on the same story are encoded by a Transformer (Vaswani et al., 2017) to obtain contextualized events \( c \):

\[
[c_{1:L_1}; \ldots; c_{N:L_N}] = \text{Encoder}(e_1^{(1)}; \ldots; e_{L_1}^{(N)} + E)
\]

where \( E \) contains three types of embeddings: Article embeddings associate the index of an article with its events. Frequency embeddings signal if an event appears in only one article, more than one but not all articles, or all articles on the same story. Position embeddings denote the relative position of an event in an article. Finally, the model predicts article’s ideology with mean pooling over events in an article. Full details are in Appendix B.

2.2 Partisan Event Dataset Annotation

Since there is no dataset with partisan event annotations for news articles, we manually label a Partisan Event (PEV) dataset with 50 articles (1867 sentences) covering two controversial issues in the U.S. in 2022: a mass shooting in Texas, and the overturn of Roe v. Wade. Note that the event dataset contains articles from a separate and later time than the training data, and is used for evaluation purposes only. We collect articles from AllSides,\(^2\) where groups of three articles that report the same news story are carefully selected by editors. For each story, we discard the center ideology article to focus on partisan media coverage. The remaining two articles, together with extracted events, are provided to two college students who have gone through similar annotation tasks. They

are underlined. 

Table 1: Macro F1 scores for article ideology prediction (average of 5 runs). Best results are in bold and second best are underlined. art., fre., and pos. refer to article, frequency, and position embeddings in §2.1.

<table>
<thead>
<tr>
<th>Model</th>
<th>AllSides</th>
<th>Basel</th>
<th>PEvent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-article +pos.</td>
<td>64.10 ± 5.51</td>
<td>55.08 ± 6.01</td>
<td>44.37 ± 2.60</td>
</tr>
<tr>
<td>Single-article +art.</td>
<td>64.37 ± 0.75</td>
<td>54.78 ± 2.38</td>
<td>45.77 ± 3.46</td>
</tr>
<tr>
<td>Multi-article +art.</td>
<td>79.52 ± 1.52</td>
<td>64.91 ± 1.78</td>
<td>76.64 ± 3.16</td>
</tr>
<tr>
<td>Multi-article +art. + fre.</td>
<td>88.64 ± 0.84</td>
<td>67.30 ± 2.45</td>
<td>85.19 ± 2.28</td>
</tr>
<tr>
<td>Multi-article +art. + fre. + pos.</td>
<td>88.49 ± 0.74</td>
<td>68.50 ± 2.07</td>
<td>83.59 ± 1.67</td>
</tr>
</tbody>
</table>

3 We intentionally annotate articles’ ideology rather than using media-level ideology to ensure accurate ideology labels.

compared to removing random events. Moreover, the more partisan events are removed, the larger the difference is, which confirms that models exploit partisan events to discern ideology (in Fig. 2).

3 Latent Variable Models for Partisan Event Detection

3.1 Task Overview

Our data is in the form of \((a, y)\), where \(y\) is the ideology for article \(a\). We extract events \(x = (x_1, \ldots, x_L)\) from each article. We define a binary variable \(m_i \in \{0, 1\}\) for each event \(x_i\), and \(m_i = 1\) indicates \(x_i\) is a partisan event. The ideology prediction task aims at predicting \(y\) using \(x\). The partisan event detection task focuses on predicting partisan indicators \(m = (m_1, \ldots, m_L)\).

3.2 Latent Variable Models

We draw on rationale extraction literature, where rationale is defined as part of inputs that justifies model’s prediction (Lei et al., 2016). We adopt the formulation in Chen et al. (2018). In details, assume a positive number \(k\) is given, the goal is to extract \(k\) events with the highest mutual information with \(y\) and treat them as partisan events. In other words, our partisan indicator \(m\) satisfies \(|m| = k \times |L|\). Since optimizing mutual information is intractable, Chen et al. (2018) provides a variational lower bound as the objective instead:

\[
\max_{E_\theta,q_\phi} \sum_{(x,y) \in D} \mathbb{E}_{m \sim E_\theta(x)} [\log q_\phi(y | m \odot x)]
\]

where extractor \(E_\theta\) models the distribution of \(m\) given \(x\), \(q_\phi\) is a predictor of \(y\) given partisan events, \(D\) is training set, and \(\odot\) is element-wise product. Intuitively, the model selects ideology indicative events as partisan events.

We parameterize both \(E_\theta\) and \(q_\phi\) exactly the same as in §2.1. For the extractor, we first get the embedding \(e\) for all events and then pass it to the transformer encoder. A linear layer converts output contextualized representations to logits, from which we sample \(k\%\) of them following the subset sampling method (Xie and Ermon, 2019)—a differentiable sampling method that allows us to train the whole system end-to-end. At inference, we select the top \(k\%\) of events with the largest logits by the extractor. For the predictor, we input \(m \odot e\) to the encoder so that it only sees the sampled subset of events. There are single- and multi-article variants, depending on if \(E_\theta\) and \(q_\phi\) access all events in a story or in an article only (details in Appendix D).
3.3 Improving Partisan Event Detection

Restricting from Picking Common Events. Since background events and main events should not be viewed as partisan events (Fig. 1), we prohibit models from selecting these events. Precisely, we use the same lexical matching method as in §2.1 to find common events in a story. At training, an auxiliary objective is added to minimize probabilities of the extractor selecting events that appear in both articles as partisan events, driving models to prefer events reported by only one side. We only apply this constraint to multi-article models since it requires story-level context to locate common events.

Pretraining to Add Event Ideology Priori. Prior knowledge, especially the media’s stance on controversial topics, plays an important role in partisan content detection. Given that the AllSides training set is small, it is hard for the model to acquire such knowledge on a broad range of topics. We pretrain a model on BIGNEWSALIGN (Liu et al., 2022) to gain prior knowledge about events. The model takes each individual event as input and predicts its ideology without context information. Intuitively, it counts the reporting frequency of each event.

4 Experiments

All models are trained solely on AllSides. For evaluation metrics, we measure article-level macro-F1 for ideology prediction, and binary-F1 (partisan events as positives) for partisan event detection.

Baselines. (1) Randomly predict partisan events with a 0.3 probability, and article ideology at 0.5. (2) Event-prior is the pretrained event model with ideology priori in §3.3, which predicts the chance of each event being left and right. We take 30%\(^4\) of events with the most skewed distribution as partisan events. We infer article’s ideology using the majority vote among partisan events. Intuitively, this baseline utilizes the prior knowledge of event ideology to detect partisanship. (3) Non-latent is the best performing multi-article model in §2.3, with no latent variables. Built upon this model, we have two variants for partisan event detection.

For Attention-based method (Wang et al., 2016), we consider the top 30% of events with the largest attention weights (sum over all heads and positions) as partisan events. For perturbation-based method (Li et al., 2016), we remove one event at a time, and treat the 30% of events that lead to the largest output change as partisan events.

Results. Table 2 presents the results. We find that multi-article models outperform single-article models on both tasks, showing that story-level context is essential to cross-document event comparison.

On partisan event detection (last column of Table 2), latent variable models outperform all baselines, showing the effectiveness of training with article ideology labels. Moreover, restricting models from selecting common events improves the performance, which validates the intuition that common events are less likely to be partisan. Prior knowledge of event ideology further boosts on both tasks, especially for single-article models, illustrating the benefits of prior knowledge when the context is limited. Combining the two improvements, the multi-article model achieves the best performance across the board. It is also important to point out that this model only uses 30% of events to predict ideology, but it still outperforms models that see full articles, which suggests that a good modeling of events in the article could be more helpful than raw text representations when predicting ideology.

Error analysis of the extracted partisan events in Appendix E reveals key challenges in detecting implicit nuanced sentiments and discerning event relations (e.g., main vs. background events).

5 Conclusion

Partisan event selection is an important form of media bias that even exist in apparently nonpartisan news. We first verify the existence of partisan event selection and then jointly detect partisan events and article’s ideology using latent variable models. Experiments show that our models identify partisan events that reasonably align with human judgment.

\(^4\)We provide pertaining details in Appendix D.

\(^5\)k = 30% since 27% of events in PEevent are partisan.

<table>
<thead>
<tr>
<th>Ideology Prediction</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>AllSides</td>
<td>Basel</td>
</tr>
<tr>
<td>Random</td>
<td>49.83±1.16</td>
</tr>
<tr>
<td>Event-prior</td>
<td>63.39±0.90</td>
</tr>
<tr>
<td>Non-latent-attr</td>
<td>88.49±0.74</td>
</tr>
<tr>
<td>Non-latent-pert</td>
<td>88.49±0.74</td>
</tr>
<tr>
<td>Single-article</td>
<td>66.75±2.35</td>
</tr>
<tr>
<td>+pri.</td>
<td>81.50±1.52</td>
</tr>
<tr>
<td>Multi-article</td>
<td>86.43±0.50</td>
</tr>
<tr>
<td>+res.</td>
<td>85.68±0.32</td>
</tr>
<tr>
<td>+pri.</td>
<td>91.93±0.72</td>
</tr>
<tr>
<td>+res. + pri.</td>
<td>91.56±0.25</td>
</tr>
</tbody>
</table>

Table 2: F1 scores (avg. of 5 runs) for ideology prediction and partisan event detection. res.: restrict models from common events; pri.: event-level prior knowledge. Non-latent models have the same ideology prediction scores since they are the same model.
6 Limitations

We investigate the impact of event selection on models’ ideology prediction performance, to verify the existence of event selection in news media. The results, however, do not state a causal relation between media ideology and reported events.

We analyze the model output and discuss in details two major limitations of our latent variable models in §E. Apart from those two errors, we also observe that events detected by the model as partisan may not align with the model’s prediction of the article’s ideology. In other words, the model could identify right-leaning events as partisan events while predicting the article as left-leaning (Table 3). Although the methods we adopt in this paper identify events that are indicative of ideology (Chen et al., 2018; Yu et al., 2019), they do not provide further justifications for how these events interact to reflect the ideology. For instance, the extractor could detect a right event and several left events that attack it. To further understand the event selection effect, future work may consider incorporating event-level ideology to model the interplay among events.

Although our models that include cross-article context can be extended to any number of articles without modification, they may be restricted by the GPU memory limit in practice. Particularly, the Transformer encoder that contextualizes all events in a story requires computational resources to scale quadratically with the number of events, which is infeasible for stories that contain many articles. Future work may consider designing special attention patterns based on the discourse role of each event in the article (van Dijk, 1988; Choubey et al., 2020).

Finally, due to the cost of manual labeling, we only evaluate our partisan event detection models on a dataset that covers two specific political issues. It remains to be seen whether methods introduced in this paper can be generalized to a broader range of issues. We call for the community’s attention to design and evaluate partisan event detection models on more diverse topics.

7 Ethical Considerations

7.1 Dataset Collection and Usage

Partisan Event Dataset Collection. We conform with the terms of use of the source websites and the intellectual property and privacy rights of the original authors of the texts when collecting articles.

Title: Biden calls for assault weapons ban, making gun manufacturers liable for shootings

President Biden on Thursday made an emotional appeal for ambitious new gun laws, including a ban on military-style rifles . . . On the other side of the aisle, Republicans bristled at Democrats’ equating support for the Second Amendment with tolerating mass murder. “You think we don’t have hearts,” said Rep. Louis Gohmert, Texas Republican.

<table>
<thead>
<tr>
<th>Ideology label: right</th>
<th>Prediction: left</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article snippet</td>
<td></td>
</tr>
</tbody>
</table>

We do not collect any sensitive information that can reveal original author’s identity. We also consult Section 107 of the U.S. Copyright Act and ensure that our collection action fall under the fair use category.

Datasets Usage. Except the partisan event dataset collected in this work, we get access to the Basil dataset by direct download. For AllSides, we contact with the authors and obtain the data by agreeing that we will not further distribute it.

7.2 Usage in Application

Intended Use. The model developed in this work has the potential to assist the public to better understand and detect media bias in news articles. The experiments in §4 show that our model is able to identify partisan events on two controversial issues that moderately align with human judgement. The detected events can be presented to show different perspectives from both ends of the political spectrum, thus providing readers with a more complete view of political issues.

Failure Modes. Our model fails when it mistakenly predicts a non-partisan event as a partisan event, misses out the partisan events, or predicts the wrong ideology for an article. They may cause misperception and misunderstanding of an event. For vulnerable populations (e.g., people who maybe not have the specific knowledge to make the right judgements), the harm could be amplified if they blindly trust the machine outputs.

Biases. The training dataset is roughly balanced in the number of left and right articles, so the model is not trained to encode bias. However, the dataset is relatively small and does not cover all possible

\[https://www.copyright.gov/title17/92chap1.html\#107\]
political topics. Particularly, most of the news articles in the training set are related to U.S. politics, thus the model is not directly applicable to other areas in the world. **Misuse Potential.** Users may mistakenly take the model outputs as ground truth. We recommend any usage of our model displaying an “use with caution” message to encourage users to cross-check the information from different sources and not blindly trust a single source.

**References**


Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. “O’Reilly Media, Inc.”.

David Broockman and Joshua Kalla. 2022. The manifold effects of partisan media on viewers’ beliefs and attitudes: A field experiment with Fox News viewers.


Appendix A  Implementation Details

For all experiments in this paper, our implementation is based on Pytorch (Paszke et al., 2019) and HuggingFace transformers (Wolf et al., 2020) library, and we preprocess all articles using Stanza (Qi et al., 2020). All experiments are conducted on 4 NVIDIA RTX A6000 GPUs.

Appendix B  Event-based Ideology Prediction Models

B.1 Event Extraction

We follow the scheme in TimeML which defines events as “situations that happen or occur” (Pustejovsky et al., 2003). We train an event extraction model on the MATRES data (Ning et al., 2018), as its event annotation is not limited to predefined event types, and thus is applicable to the open-domain scenario. We use RoBERTa-large (Liu et al., 2019b) that predicts a binary label for each word, deciding whether the word is an event predicate or not. To provide surrounding context, we split articles into groups of 4 sentences and process 4 sentences together. We follow previous work on using TimeBank and AQUAINT sections in MATRES as training set and Platinum section as test set (Ning et al., 2019). Table 4 shows the hyperparameters for model architecture and training process. On the same train and test split, our model achieves an F1 score of 89.53, which is on par with the state-of-the-art performance of 90.5 F1 score (Zhang et al., 2019b) that predicts a binary label for each word, matching semantic role labeling tools (Shi and Lin, 2019; Gardner et al., 2018),\(^7\) and we only keep predicates that match our event extraction results.

Multiple events can exist in one sentence with overlapping predicates and arguments. We hence remove the shorter event if there is an overlap, as we find that shorter events tend to be less informative. For example, it is easier to determine the partisanship of the event “the leak of a draft opinion” than a shorter one on “the leak of a draft opinion.” Therefore, we remove an event if its predicate is covered by another event’s arguments.

\(^7\)github.com/CogComp/SRL-English for nouns.

We then input all events in one article or all articles on the same story to another transformer encoder (Vaswani et al., 2017) to get contextualized predicates. If a sentence has multiple events, we mask out other events’ tokens when encoding one event, so that the information in one event does not leak to others. We then get \(e = W[e_{\text{pred}}, e_{\text{arg0}}, e_{\text{arg1}}]\) for event by taking the average of last-layer token embeddings. If a sentence has multiple events, we mask out other events’ tokens when encoding one event, so that the information in one event does not leak to others. We then get \(e = W[e_{\text{pred}}, e_{\text{arg0}}, e_{\text{arg1}}]\) for event.

B.2 Ideology Prediction

Given \(N\) input articles that report on the same news story, we extract events in \(i\)th article as \(\{x_{i}^{(1)}, \ldots, x_{i}^{(L_i)}\}\), where \(L_i\) is the number of events. We first use a DistilRoBERTa model to get the embedding \(e\) for an event (Sanh et al., 2019): We input the sentence that contains the event to DistilRoBERTa and get the embeddings \(e_{\text{pred}}, e_{\text{arg0}}, e_{\text{arg1}}\) for predicate, ARG0, and ARG1 by taking the average of last-layer token embeddings. If a sentence has multiple events, we mask out other events’ tokens when encoding one event, so that the information in one event does not leak to others. We then get \(e = W[e_{\text{pred}}, e_{\text{arg0}}, e_{\text{arg1}}]\), where \(,\) means concatenation and \(W\) is learnable.\(^8\)

We then input all events in one article or all articles on the same story to another transformer encoder (Vaswani et al., 2017) to get contextualized \(e\) for each event:

\[
\{e_{1}^{(1)}; \ldots; e_{1}^{(N)}\} = \text{Encoder}\left(\{e_{1}^{(1)}; \ldots; e_{1}^{(N)}\} + E\right)
\]

where the three embeddings are:

- **Article embedding** indicates the index of the article that contains the event, with one embedding per article index. The datasets we experiment with in this paper have at most 3 articles in each story. During training, we randomly shuffle the articles in each story.

- **Frequency embedding** informs the model whether the event appears in only one article, at least two but not all articles, or all articles in the story. We have one embedding per category. We find common events through lexical\

\(^8\)We use a zero vector if ARG0 or ARG1 does not exist.
Remove partisan events - Remove random events

Figure 2: Performance difference (average of 10 runs) between removal of partisan events and of random events. A negative value indicates more severe performance regression when dropping partisan events, compared with dropping the same amount of random events.

Table 5: Hyperparameters used for the event-based ideology prediction model.

<table>
<thead>
<tr>
<th>Hyperparameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of epochs</td>
<td>5</td>
</tr>
<tr>
<td>maximum learning rate</td>
<td>5e-5</td>
</tr>
<tr>
<td>learning rate scheduler</td>
<td>linear decay with warmup</td>
</tr>
<tr>
<td>warmup percentage</td>
<td>6%</td>
</tr>
<tr>
<td>optimizer</td>
<td>AdamW</td>
</tr>
<tr>
<td>weight decay</td>
<td>1e-4</td>
</tr>
<tr>
<td>transformer hidden dimension</td>
<td>768</td>
</tr>
<tr>
<td>transformer # heads</td>
<td>12</td>
</tr>
<tr>
<td># transformer layer</td>
<td>4</td>
</tr>
<tr>
<td># FFNN layer</td>
<td>2</td>
</tr>
<tr>
<td>hidden layer dimension in FFNN</td>
<td>768</td>
</tr>
<tr>
<td>dropout in FFNN</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Table 6: Statistics for AllSides training set, Basil, and PartisanEvent. AllSides test set contains 1,416 articles.

<table>
<thead>
<tr>
<th></th>
<th>AllSides</th>
<th>Basil</th>
<th>PEvent (ours)</th>
</tr>
</thead>
<tbody>
<tr>
<td># stories</td>
<td>2,221</td>
<td>67</td>
<td>25</td>
</tr>
<tr>
<td># articles</td>
<td>5,361</td>
<td>134</td>
<td>50</td>
</tr>
<tr>
<td># events detected per article</td>
<td>66.82</td>
<td>48.71</td>
<td>60.70</td>
</tr>
</tbody>
</table>

Appendix C Partisan Event Annotation

Data Collection. We manually collect 25 stories, each with three articles from AllSides that relate to the mass shooting in Texas and the overturn of Roe v. Wade. We extract events from each article and only keep the left and right article in each

Appendix C

We search threshold values from 0.2 to 0.5 by manually inspecting identified common events in 6 articles. A value of 0.4 can identify common events accurately while still allowing variations such as variants of mentions (e.g., president vs. president Biden).

Each story on AllSides contains three articles from left, center, and right respectively. We only include the left and right articles in our dataset.

positions. We multiply the relative position of the event (a real number in [0, 1]) with a learnable embedding.

We further train a [SEP] token that separates the events from different articles. Finally, average representation of all events in an article is used to predict the article’s ideology. Table 5 includes the hyperparameters of the model.

The entire model contains 106M parameters. On average, the training takes 25 minutes on a single NVIDIA RTX A6000 GPU.

Position embedding represents the relative position of the event in the article. We multiply the relative position of the event (a real number in [0, 1]) with a learnable embedding.

We further train a [SEP] token that separates the events from different articles. Finally, average representation of all events in an article is used to predict the article’s ideology. Table 5 includes the hyperparameters of the model.

The entire model contains 106M parameters. On average, the training takes 25 minutes on a single NVIDIA RTX A6000 GPU.

Position embedding represents the relative position of the event in the article. We multiply the relative position of the event (a real number in [0, 1]) with a learnable embedding.

We further train a [SEP] token that separates the events from different articles. Finally, average representation of all events in an article is used to predict the article’s ideology. Table 5 includes the hyperparameters of the model.

The entire model contains 106M parameters. On average, the training takes 25 minutes on a single NVIDIA RTX A6000 GPU.

Position embedding represents the relative position of the event in the article. We multiply the relative position of the event (a real number in [0, 1]) with a learnable embedding.

We further train a [SEP] token that separates the events from different articles. Finally, average representation of all events in an article is used to predict the article’s ideology. Table 5 includes the hyperparameters of the model.

The entire model contains 106M parameters. On average, the training takes 25 minutes on a single NVIDIA RTX A6000 GPU.
• Estimate the sentiment toward each entity in the event. Sentiments can be reflected in words, quotations, and the relations between entities.

• Use entities and sentiments to decide whether the event is sided with the article’s ideology. If it does, label it as a partisan event. Ex. Label an event as partisan in the left article, only if its “left” entity has a positive sentiment, or its “right” entity has negative sentiment. Also, it may be possible for events to be purely factual, which means there is no strong sentiment toward entities in events. For these kinds of events, try your best to estimate whether these events indirectly present any sentiment toward entities in the article.

Two annotators label all 50 articles, and a third annotator compares their annotations and resolve conflicts. We calculate inter-annotator agreement on all 50 articles and numbers can be found in §2.2.

Partisan Events Distribution. We further investigate the distribution of position of partisan events in the article. Fig. 3 shows the percentage of partisan events that belong to each quartile of an article. As can be observed, right articles have more partisan events that appear in later parts of an article, whereas partisan events in left articles are evenly distributed in the article.

Appendix D  Latent Variable Models

Implementation Details. For both extractor and predictors, we use the same model architecture as in §B.2 with hyperparameters listed in Table 5. The two-player model contains 213M parameters. On average, the training takes 50 minutes on a single NVIDIA RTX A6000 GPU.

Pretrained Model for Event Representation. We use the BIGNEWSALIGN dataset (Liu et al., 2022) to pretrain a model with prior event ideology knowledge. We remove stories in the dataset that contain duplicate articles and downsample articles in each story so that the number of left and right articles are balanced. Table 7 shows the statistics of the pretraining dataset. We then train a DistrilRoBERTa model that takes each event as input and predicts the event’s ideology, where we use the article ideology as the event’s ideology. We train this model on BIGNEWSALIGN for 2 epochs and use it to initialize our latent variable models.

Appendix E  Additional Error Analysis

Table 8 and Table 9 present the predictions by the two-player model on the multi-article setup with one-sided restriction and prior knowledge for events. Two major types of errors are observed. First, the model struggles when an article attacks a statement from the opposite side with an implicit sentiment. For instance, “threw,” “continue,” and “had” in Table 8 are events or statements from the right, but the author reports them with an implicit negative sentiment (e.g., “not a thing!”), making the event flatter to the left. Future models need to have an enhanced understanding of implicit sentiment along with the involving entities (Deng and Wiebe, 2015; Zhang et al., 2022). Second, the model still frequently selects main events as partisan content, as shown by the “delivered” event in Table 9, maybe because models need to include it as necessary context. The constraint introduced in §3.3 fails in this case because the other article describes this event differently (i.e., “Biden made an emotional appeal”), thus suggesting future research direction that leverages cross-document event coreference.

<table>
<thead>
<tr>
<th></th>
<th># articles</th>
<th># events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>128, 481</td>
<td>6, 280, 732</td>
</tr>
<tr>
<td>Right</td>
<td>123, 380</td>
<td>4, 986, 165</td>
</tr>
</tbody>
</table>

Table 7: Statistics for the BIGNEWSALIGN pretraining dataset.
Table 8: Article snippets of model predictions (multi-article two-player model with both improvements) and annotations. Colored spans denote events, with the predicate bolded. Blue: model predictions; red: human annotations; purple: annotations and predictions.

Title: At the NRA Convention, People Blame Mass Shootings on Everything But Guns
The nation has been plunged into despair and mourning . . . in Houston, the National Rifle Association still threw a party . . . Two messages emerged from the assembled throngs and the drooping politicians in attendance, just 300 miles from Uvalde: 1) People must continue to enjoy the right to acquire any damn firearm they choose, without meddling from the state; and 2) the massacre had absolutely nothing—not a thing!—to do with the untrammeled commerce in guns . . .

Ideology label: left Prediction: left

Table 9: Article snippets of human annotations and model predictions (multi-article two-player model with both improvements). Highlighted spans denote events, with the predicate bolded. Blue: model predictions; purple: human annotations and predictions.

Title: “Enough”: Biden Exhorts Congress To Pass Gun Control Laws
President Joe Biden delivered the second evening address of his presidency on Thursday night, almost begging Congress to pass gun control legislation . . . However, Biden cited former Supreme Court Justice Antonin Scalia—a conservative icon—who had declared that the Second Amendment was “not unlimited.”

Ideology label: left Prediction: left