

# Conflicts, Villains, Resolutions: Towards models of Narrative Media Framing

Lea Frermann<sup>1</sup> Jiatong Li<sup>2</sup> Shima Khanehzar<sup>1,3</sup> Gosia Mikolajczak<sup>4</sup>

<sup>1</sup>School of Computing and Information Systems, The University of Melbourne

<sup>2</sup>Department of Computing, The Hong Kong Polytechnic University <sup>3</sup>CSIRO Data61

<sup>4</sup>The Global Institute for Women’s Leadership, The Australian National University

lfrermann@unimelb.edu.au

jiatong.li@connect.polyu.hk

shima.khanehzar@data61.csiro.au

Gosia.Mikolajczak@anu.edu.au

## Abstract

Despite increasing interest in the automatic detection of media frames in NLP, the problem is typically simplified as single-label classification and adopts a topic-like view on frames, evading modelling the broader document-level narrative. In this work, we revisit a widely used conceptualization of framing from the communication sciences which explicitly captures elements of narratives, including *conflict* and its *resolution*, and integrate it with the narrative framing of key entities in the story as *heroes*, *victims* or *villains*. We adapt an effective annotation paradigm that breaks a complex annotation task into a series of simpler binary questions, and present an annotated data set of English news articles, and a case study on the framing of climate change in articles from news outlets across the political spectrum. Finally, we explore automatic multi-label prediction of our frames with supervised and semi-supervised approaches, and present a novel retrieval-based method which is both effective and transparent in its predictions. We conclude with a discussion of opportunities and challenges for future work on document-level models of narrative framing.<sup>1</sup>

## 1 Introduction

Media discourse around contested issues is often biased by experiences or interests of the news outlets and different stakeholders they give voice to. News framing by the media has been formalized and examined on many levels in communication and social sciences, ranging from selection of information (Levitt, 1981) over discourse-centric (Pan and Kosicki, 1993) and entity focused approaches (Lawlor and Tolley, 2017). While a growing body of work in NLP attempts to automatically detect framing in the news or social media, most work adopts well-defined yet oversimplifying approaches like topic modeling or simple

<sup>1</sup>Code and data are available at <https://github.com/phenixace/narrative-framing>.

---

Resolution:	solution/alleviation of the issue
Conflict:	disagreements between individuals, groups, institutions, countries, etc.
Human Interest:	emotionalization and dramatization of an issue through the lens of affected individuals
Moral:	moral or religious references
Economic:	economic consequences for individuals, groups, institutions, countries, etc.

---

Hero:	entity contributing to/responsible for issue resolution
Villain:	entity contributing to/responsible for issue cause
Victim:	entity suffering the consequences of an issue

---

Table 1: Five frames (top) and three narrative roles (bottom) considered in this paper.

classifiers (see Ali and Hassan (2022) for a recent review); formalize the task as single-label classification ignoring co-existence and interactions of different frames, and focus on localized emphasis frames rather than the full story (Card et al., 2015; Field et al., 2018; Khanehzar et al., 2021).

This paper addresses the above shortcomings by considering framing through the lens of narratives. We adopt a small set of high-level framing devices established in the communication literature (Neuman et al., 1992; Semetko and Valkenburg, 2000), and integrate them with narrative roles assigned to key actors (entities) in the discourse. Table 1 defines the frames and associated entity roles. We argue that more nuanced and transparent automated models of framing are essential to meaningfully support social studies into a systematic understanding of the viewpoints presented by different stakeholders in contested issues such as climate change. Our contributions to this end are (1) introducing an established frame inventory and annotation procedure from the communication sciences into NLP; (2) a labeled data set; (3) a case study on the framing of climate change, showcasing the potential of our annotations for large-scale media analysis; and (4) experiments on automatic frame prediction, including an effective and transparent retrieval-based classifier to predict multiple

frames per article.

Table 1 (top) summarizes the frames used in our work. Our framework departs from existing NLP approaches in two ways: first, we adopt a multi-label classification paradigm, allowing for a more nuanced analysis and avoiding the oversimplification of framing to a single label per item. Secondly, our framework emphasizes the narrative structure of the *full* article: frames such as *conflict*, *resolution* or *human interest* are central building blocks of narratives, and are dominant in news coverage of contested issues (Semetko and Valkenburg, 2000). Building on components of the Narrative Policy Framework (Shanahan et al., 2018), we identify the key entities responsible for the issue (*villains*), those who are affected (*victims*) and those who can resolve the issue (*hero*); see Table 1 (bottom).

We apply our framework to the issue of climate change, a pressing global challenge with wide-reaching impacts (Pew Research Center, 2022), which remains politically contested in terms of the understanding of its urgency, causes, and possible solutions (Sparkman et al., 2022). Importantly, studies show that climate ‘skeptics’ (and those lacking scientific backing) are cited almost twice as often in the mainstream news media as those calling for climate action (Wetts, 2020), rendering the examination of media framing and its effects on public support for climate change mitigation a pressing goal. While a substantive body of work on climate change framing emerged in the social and communication sciences (Nisbet, 2009; Wolters et al., 2022), the issue has attracted surprisingly little attention in the NLP community to date. Exceptions include work on stance detection in news (Luo et al., 2020) or social media (Vaid et al., 2022) or models of scepticism detection (Bhatia et al., 2021), whereas we focus on the narrative framing of the issue across political leanings. To recap, in this paper we present:

- The concept of “Narrative Media Framing” formalized through a set of frames about conflicts, their effects and resolutions which are integrated with narrative roles assigned to key actors (Section 3).
- The narrative frames corpus of 428 English news articles on climate change labeled with frame devices (Table 1). Following Semetko and Valkenburg (2000), annotators answered binary indicator questions, and the final frame

labels were derived from the answer set (Section 3).

- A detailed analysis of our annotated data set, highlighting the interaction of frames and narrative roles, and differences across media outlets with different political bias (Section 4).
- Experiments on automatic frame prediction, including semi-supervised and supervised methods, including a new simple and transparent, yet effective method which combines retrieval with classification (Section 5).

## 2 Background

Media framing refers to the deliberate presentation of information in order to elicit a desired response or shift in reader’s attitude. We introduce into NLP five high-level frames (Table 1 top), identified by Semetko and Valkenburg (2000) as covering the dominant framing in reporting on contested issues with the aim to attract reader’s attention (Mendelsohn et al., 2021). These categories have been applied via manual content analysis to a variety of issues and events, ranging from the media coverage of the Egyptian revolution (Fornaciari, 2012), over the MH370 crash (Bier et al., 2018) to climate change (Dotson et al., 2012; Feldman et al., 2017). To identify each frame, Semetko and Valkenburg (2000) proposed a set of binary indicator questions which improved annotation quality and portability of the framework across studies. We construct the first publicly available data set annotated with this framework, cover a larger and more diverse set of news articles than prior work, and link the frames with narrative roles assigned to key entities appearing in the story.

Media framing may manifest through the narrative roles – Hero, Villain or Victim – assigned to key entities in a document, and this phenomenon has been widely studied in the communication sciences in general (Shanahan et al., 2018) and in the context of climate change in particular (Lück et al., 2018). We draw on this work as well as work which identified key stakeholder categories in the climate change discourse (Haigh and Griffiths, 2009; Ahchong and Dodds, 2012; Chen et al., 2023) to analyze the framing of entities in news articles along the political spectrum.

NLP studies on framing have predominantly focused *emphasis framing*, the strategic inclusion or omission of aspects of an issue, such as legality

or public opinion (Card et al., 2015) or, to a lesser extent, on *equivalence framing* as different expressions of identical concepts (“alien” vs “immigrant”, Lee et al. (2022); Ziems et al. (2022a)). Both perspectives focus on local, lexical signals. Emphasis framing is typically formalized as a single-label prediction of the most dominant frame in a news article (Card et al., 2015), headline (Liu et al., 2019; Akyürek et al., 2020), or a social media post (Johnson et al., 2017; Hartmann et al., 2019). The media frames corpus (Card et al., 2015) is one of the most comprehensive frame-labeled data sets comprising several thousand news articles across five contested issues. While the data includes span-level labels which *could* be used for multi-label classification, work using the MFC predominantly attempts document-level prediction of a single “primary” article frame, disregarding span labels (Ji and Smith, 2017; Khanehzar et al., 2021).<sup>2</sup> More broadly, our work complements emphasis frames by considering more abstract frames around conflict, resolution, and personal, moral and economic impacts. Both formalizations of framing have a strong foundation in the communication literature, and studying their interaction at scale with NLP methodology is an interesting avenue for future work.

Mendelsohn et al. (2021) consider a variety of framing strategies in the context of tweets, but with less of a focus on story structure due to the short document lengths. Our framework complements their work in three ways: i) we approach framing as multi-label classification, relaxing the assumption of a single frame per article; ii) we present a set of frames that are abstract with evidence distributed across a document, requiring higher-level document model; and iii) we link frames with narratives via entity roles in a unified annotation framework consisting of a series of binary indicator questions, allowing us to study the interplay of framing and narratives.

### 3 The Narrative Frames Corpus

We identified 17.9K English-language news articles on climate change published in 2017–2019 in the UK and in the US by matching a set of climate change-specific keywords in articles from the NELA corpora (Horne et al., 2018; Nørregaard et al., 2019; Gruppi et al., 2020). See Appendix A for more details. For each article, NELA provides

<sup>2</sup>Although see Field et al. (2018) for an exception.

metadata about its publication date, media outlet, and its associated political leaning as identified by the Media Bias Fact Check (MBFC) website.<sup>3</sup> We manually annotated a subset of 428 articles of this data set, balanced across the three years and the four most dominant MBFC categories: left, center-left, right and questionable source.<sup>4</sup>

We recruited four on-site annotators, all English native speakers with a background in the social/political sciences. The annotators went through an extensive training phase including several rounds of feedback. Details of annotator remuneration can be found in the Ethics statement.

**Frame annotations** We adapted Semetko and Valkenburg (2000)’s frame indicator questions. We added a pre-screening question to confirm that an article is predominantly (>70%) about climate change, removed one question about visual information (as we focus on text only), and changed wording specific to the ‘government’ to ‘any entity’ to align with our broad definition of stakeholder entities, discussed below. The full questionnaire is shown in Appendix B. Annotators were presented with the full article text together with the questionnaire, but no explicit meta information such as outlet name or date of publication.

The raw annotations provided answers to a list of binary indicator questions. We verified that the mapping (i.e., the factor structure) between the five frames in Table 1 and their associated indicator questions in Semetko and Valkenburg (2000) replicates in our annotated data set. To do so we ran a confirmatory factor analysis (CFA; Brown and Moore (2012)).<sup>5</sup> We removed all items with a factor loading < 0.3 (as not fitting well into any of the five factors), retaining a total of 13 indicators, with 2-3 indicators loading on a given frame. These questions are listed in Table 2. The final model fitted the data well (CFI=.945; RMSEA=.052[.039, .065], p=.370; SRMR=.059), confirming the five-factor structure. An article was then labeled with a frame if  $\geq 2$  indicator questions for that frame were answered ‘yes’ by  $\geq 2$  annotators.<sup>6</sup> This resulted

<sup>3</sup><https://mediabiasfactcheck.com/>

<sup>4</sup>The ‘center-right’ category was very rare in the set of sampled articles, and hence merged with ‘right’.

<sup>5</sup>The input data for the CFA is based on majority voting, i.e., at least two out of three annotators agreeing on a given response. Prior to the main analysis, we converted raw dichotomous (0-1) scores into a polychoric correlation matrix which served as an input into CFA.

<sup>6</sup>Except for Moral, where only 1 question had been answered ‘yes’ by a majority of annotators due to the rarity of the label.

RE	(1) Does the story suggest a solution(s) to the issue/problem?; (2) Does the story suggest that some entity could alleviate the problem?
CO	(1) Does the story reflect disagreement between political parties/individuals/groups/countries?; (2) Does one party/individual/group/country reproach another?; (3) Does the story refer to two sides or more than two sides of the problem or issue?
HI	(1) Does the story provide a human example or a "human face" on the problem/issue?; (2) Does the story employ adjectives or personal vignettes that generate feelings of outrage, empathy-caring, sympathy, or compassion?; (3) Does the story go into the private or personal lives of the entities involved?
MO	(1) Does the story contain any moral message?; (2) Does the story make reference to morality, God, and other religious tenets?
EC	(1) Is there a mention of financial losses or gains now or in the future?; (2) Is there a mention of the costs/degree of the expense involved?; (3) Is there a reference to the economic consequences of pursuing (or not) a course of action?

Table 2: Binary indicators for the five frames: Resolution (RE), Conflict (CO), Human Interest (HI), Moral (MO), Economic (EC).

in a multi-label data set with articles covering zero (12%), one (39%), two (32%) three (15%) or four (3%) frames. See Appendix F for additional data set statistics.

**Entity annotations** Entities were annotated as part of the binary indicator questionnaire introduced above. Three indicator questions assessed whether an article contained an entity that could *alleviate the problem* (Hero); *was responsible for the problem* (Villain) or *was negatively affected by the issue* (Victim). If an annotator answered ‘yes’ to any of these questions, they were asked to identify the most appropriate entity in the text. An entity was meant to be selected only if the article was explicit about that entity’s role (e.g., a politician was depicted as "the only person who could save the planet") and strictly based on the entity’s presentation in the article, rather than the annotator’s opinion about that entity.<sup>7</sup> We included all entities extracted by our annotators as part of our published data set.

**Annotator agreement** Krippendorff’s  $\alpha$  across four annotators and 13 frame indicator questions is 0.52, indicating fair agreement as expected for a complex task like frame annotation. Average pairwise agreement without chance-correction is 0.78 (min=0.75, max=0.81).

A total of 2,185 entities were extracted across all narrative roles. Average pairwise agreement on *existence* of a role in an article was 0.59 (Krippendorff’s  $\alpha = 0.40$ ). To assess agreement on the *identity* of entities for roles which were attested by at

<sup>7</sup>E.g., if an article presented Trump as a person who could mitigate climate change, the annotator was supposed to tag him as a Hero, even if they didn’t personally agree with that interpretation.

least two annotators, we computed the exact string match of associated entities, after basic text normalization. Entities match exactly 41% of the time. We also computed more lenient metrics based on token overlap (average Rouge-L=0.45) and embedding similarity (average BertScore=0.91) between pairs of extracted entities.

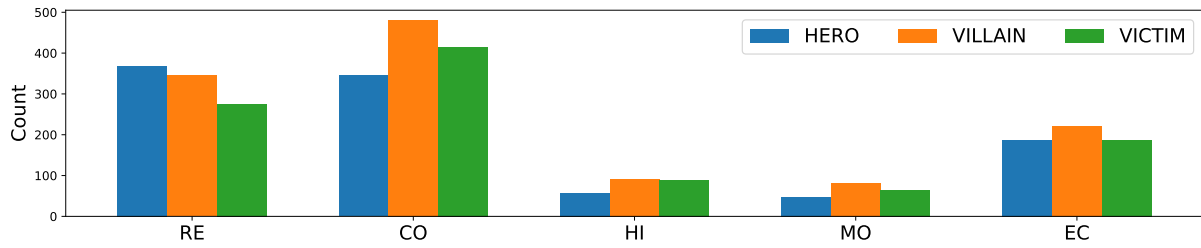
The agreement for both role detection and entity-role assignment was low overall, suggesting that the task is challenging. In this paper, we use the narrative role labels in the exploratory analysis in Section 4 and discuss future work on computational modeling of narrative roles in Section 6.

### 3.1 Stakeholder categories

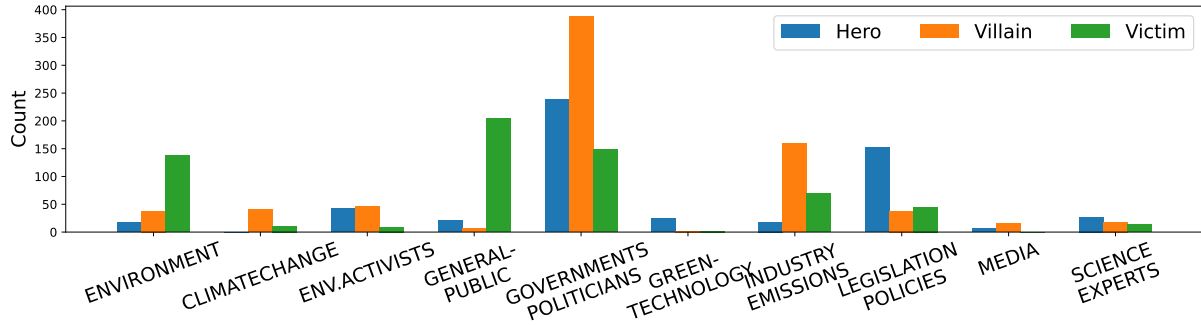
We grouped the  $>2K$  extracted entities into a smaller set of stakeholder categories to ease analysis. We identified 10 such categories from the previous literature (Ahchong and Dodds, 2012; Blair and McCormack, 2016; Chen et al., 2023; Haigh and Griffiths, 2009), adopting a broad definition of stakeholders which includes groups or entities that ‘affect or are affected by’ the issue of climate change (Freeman, 1984). The set of stakeholder categories is shown in Figure 1b, and Appendix E provides additional details. One annotator assigned each unique extracted entity to its most appropriate stakeholder category (or a generic category ‘Other’ if no other category fit).

In sum, the narrative frames corpus consists of 428 English news articles labeled with (1) multi-label frame categories; (2) narrative roles for specific entities; and (3) their associated stakeholder category; as well as meta-data about the article’s date of origin, outlet, and associated political leaning.





(a) Co-occurrence (count) of roles with frames.



(b) Co-occurrence (count) of roles with stakeholder groups.

Figure 1: Association of roles with different frames (top) and stakeholder groups (bottom).

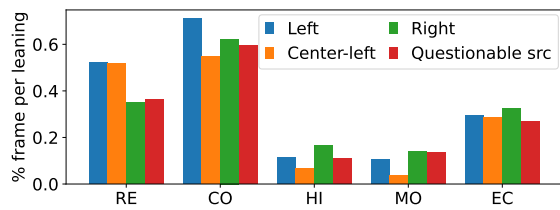


Figure 2: Distribution of frames across political leanings of news outlets (as attested by MBFC).

#### 4 Narrative Framing of Climate Change

We conduct an exploratory analysis on the framing of climate change in media outlets with different political leaning, as well as the interplay of frames, narrative roles and stakeholder categories.

**Framing and political leaning** Figure 2 shows the proportion of articles mentioning each frame by the media outlets’ political leaning.<sup>8</sup> Conflict (CO) and Resolution (RE) are most prevalent across all leanings. The Moral frame (MO) is least prevalent throughout. This pattern is partially consistent with previous research. Dirix and Gelders (2010) found Resolution, but not Conflict, to dominate in climate change reporting in the Netherlands and France in early 2000s, which might suggest that the discourse on climate change has be-

<sup>8</sup>Noting that the numbers do not sum to one due to the multi-label nature of our annotations.

come more polarized over time, in particular in our data set of US and UK news coverage where the media landscape is strongly partisan. For example, in a more recent study involving four major US newspapers, Kim and Wanta (2018) show that Conflict is the most common frame in the context of US immigration.

Resolution (RE) is more prevalent in the left-leaning outlets (left, left\_center), while the opposite is true for Human Interest (HI): right-leaning (and questionable) outlets are more likely to refer to personal stories and use language evoking empathy. These findings are partially consistent with prior work, e.g, Feldman et al. (2017) show that both Economic and Conflict are more likely to be used in conservative outlets, while we find Conflict prevalent across the board. However, Feldman et al. (2017) only included three major US news papers, in contrast to 41 in our analysis.

**Frames, roles and stakeholders** Figure 1 illustrates the association of narrative roles with different frames (1a) and stakeholders (1b). Unsurprisingly, the Hero, an entity presented with the ability to fix or alleviate the issue under discussion, is the most prevalent role in the Resolution frame. The Villain dominates most other frames, except for the Human Interest frame where Victim is equally dominant. This aligns with a well-known “negativity bias” in news reporting, i.e., a domi-

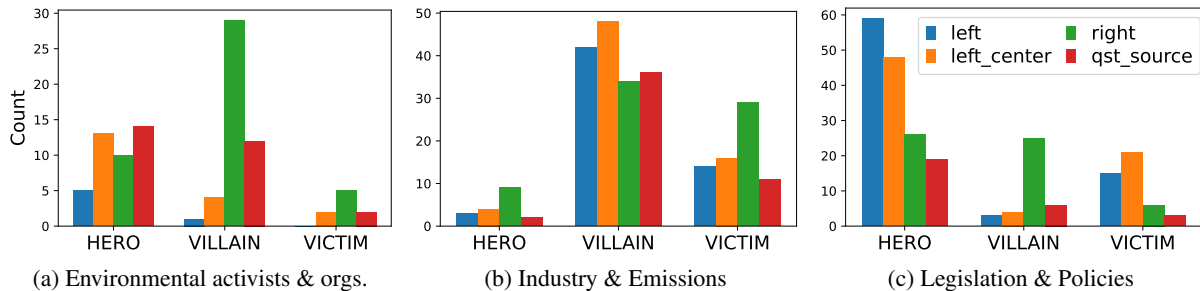


Figure 3: Narrative roles assigned to stakeholder categories in news outlets with different political leaning.

nance of negative content with a focus on problems, conflicts and their causes and victims (Soroka et al., 2019).

We explore the distribution of roles across stakeholder categories in Figure 1b. Overall, Governments & Politicians are the most dominant stakeholder category, typically depicted as the Villain (of all stakeholder categories they are also most likely to be depicted as the Hero pointing to ambivalent attitudes toward this category). The Environment and the General Public dominate the Victim role, somewhat unexpectedly followed by Governments & Politicians and Industry & Emissions. We explain this phenomenon next by disentangling the labels by political leaning.

Figure 3 reveals how the framing of a particular stakeholder category can vary with political leaning of the source. Right-leaning media are more likely to depict Environmental Activists & Organisations, and Legislation as the Villain and Industry and Emissions as either the Hero or the Victim in the context of climate change news. Conversely, left-leaning media are more likely to frame Legislation as a Hero, cover Environmental Activists less frequently overall, and predominantly frame the Industry as a Villain.

## 5 Narrative Frame Prediction

Predicting narrative frames automatically and with high quality would open new possibilities for scaling media framing analyses to larger data sets, longer time spans or more languages. Given the political sensitivity of automated media analysis, models should not only be reliable but also transparent in their predictions. To this end, we present Retrieval-Based Frame prediction (RBF), which incorporates an embedding-based retrieval module into supervised classifiers. We compare RBF against a range of neural classifiers on multi-label frame prediction. RBF not only outperforms off-

the-shelf fine-tuned transformers on this task, but also increases interpretability by predicting frames for a given article together with the most relevant article sentences for the frame as evidence. Section 6 discusses additional modelling tasks supported by our data set to be addressed in future work.

### 5.1 RBF: Retrieval-Based Frame Prediction

We propose a simple method, retrieval-based frame prediction (RBF), which combines pre-trained language model embeddings with a retrieval objective. Similar approaches have been previously proposed in the context of word-sense disambiguation and semantic frame predictions (Jiang and Riloff, 2021; Blevins and Zettlemoyer, 2020). We embed (i) short frame descriptions  $f_1 \dots f_C$ <sup>9</sup> and (ii) sentences from an input news article  $s_1 \dots s_N$  in a joint space, and retrieve sentences most proximate to the frame embedding:

$$\begin{aligned}
 h_i^s &= emb(s_i) \\
 h_j^f &= emb(f_j) \\
 rel(s_i, f_j) &= \cos(h_i^s, h_j^f),
 \end{aligned} \tag{1}$$

where  $h_i^s$  and  $h_j^f$  are the embeddings of sentence  $s_i$  and frame  $f_j$ , respectively, the relevance  $rel$  of  $s_i$  to  $f_j$  corresponds to their cosine similarity. We use SentenceBert (Reimers and Gurevych, 2019) as our embedding method  $emb$ . Given an article, we obtain  $J$  frame-specific relevance-rankings of all sentences in the input article.

We then train a linear classifier to predict the presence or absence of a frame in an article based on the most relevant sentences by our measure above. We include five input channels: channels (1)–(3) are the three sentences most relevant for a

<sup>9</sup>RE: Solution or alleviation of the problem, CO: Human interest, emotion or dramatization of events, HI: Conflict or disagreement between two or more sides, MO: Morality or religion, EC: Economic consequences

Model	Macro-Pr	Macro-Re	F1
Random	0.33 ( $\pm 0.025$ )	0.49 ( $\pm 0.023$ )	0.39
Majority	0.14 ( $\pm 0.030$ )	0.24 ( $\pm 0.080$ )	0.18
KNN	0.40 ( $\pm 0.049$ )	0.61 ( $\pm 0.071$ )	0.49
BERT	0.44 ( $\pm 0.109$ )	0.63 ( $\pm 0.059$ )	0.52
Longformer	0.48 ( $\pm 0.054$ )	0.63 ( $\pm 0.052$ )	0.53
Snippext	<b>0.56</b> ( $\pm 0.074$ )	0.65 ( $\pm 0.074$ )	0.60
RBF	0.51 ( $\pm 0.044$ )	<b>0.76</b> ( $\pm 0.131$ )	<b>0.61</b>
RBF -a	0.50 ( $\pm 0.047$ )	0.70 ( $\pm 0.077$ )	0.58
RBF -a,t	0.48 ( $\pm 0.041$ )	0.63 ( $\pm 0.085$ )	0.55

Table 3: Frame prediction results of baselines (top) representative supervised and semi-supervised methods, and RBF (middle), as well as an ablation of RBF channels (bottom). We report macro-averaged precision and recall across the five labels with standard deviation (in brackets), and their harmonic mean (F1).

frame according to RBF relevance; channel (4) includes all sentences exceeding relevance threshold  $\theta > 0.15$ , except for sentences (1)–(3), concatenated with a [SEP] token;<sup>10</sup> and (5) contains the news article truncated at 256 tokens. Each channel is encoded with the Longformer (Beltagy et al., 2020), final hidden state embeddings are concatenated and passed into the classifier. Longformer parameters are fine-tuned during the training process. We evaluate the importance of different channels by ablating the impact of the full article channel (5) (RBF -a) and additionally the threshold sentence channel (4) (RBF -a,t).

RBF combines two desiderata: First, it identifies multiple sentences relevant to a target frame, capturing key evidence that may be distributed across the article rather than locally. Second, RBF’s frame-based sentence retrieval backbone can be interpreted as an ‘explicit attention mechanism’, customized to the frame label to be predicted, with sentences serving as evidence. We evaluate the retrieved sentences in terms of their interpretability in Section 5.3.1.

## 5.2 Experimental Setup

Given the small size of the Narrative Frames Corpus, we adopt the simplest formalization of multi-label classification: for each model class (row in Table 3) we train five individual binary classifiers, one per frame label.<sup>11</sup>

<sup>10</sup>The threshold  $\theta$  was tuned on the dev set in preliminary experiments.

<sup>11</sup>This outperformed multi-label classifiers which shared a subset of parameters in preliminary experiments, presumably because the small number of article prevented models from learning useful interactions.

**Data set** We split the Narrative Frames Corpus set into five random folds of 60/20/20 train/dev/test data. As the proportion of articles mentioning each frame varies widely (see Appendix F), within each fold, we balanced the number of articles that do/do not contain each frame. E.g., for a frame that was featured in the majority of articles, we randomly up-sampled articles *not* featuring the frame to a ratio of 1:1.

**Comparison models** We compare our method against 1. a **Random** baseline; 2. a **Majority** baseline which predicts the majority class per frame (1 for frames which occur in the majority of labelled articles, 0 for others); 3. a non-neural method which embeds articles based on TF-IDF representations, and trains one **KNN** classifier per frame; 4. **BERT-medium** (Devlin et al., 2019) fine-tuned for binary frame prediction; 5. **Longformer-base-4096** (Beltagy et al., 2020) fine-tuned for binary frame prediction; and 6. an adaptation of the **Snippext** model (Miao et al., 2020; Berthelot et al., 2019), a method for semi-supervised fine-tuning of pre-trained language models which was originally proposed in computer vision, but recently adapted to semi-supervised opinion mining (Miao et al., 2020). Snippext fine-tunes BERT using an interpolation of a small amount of gold-labeled data, and a much larger set of unlabeled data with predicted, soft labels,<sup>12</sup> drawing on the MixMatch strategy recently proposed in computer vision (Berthelot et al., 2019). We augment our small labeled training data set with the  $\approx 17.5K$  unlabelled climate-related articles (cf., Section 3). The input to all transformer models is truncated to 256 tokens.<sup>13</sup> Detailed training settings and model parameters are provided in Appendix G.

**Metrics** We evaluate models on correctly predicting the *presence* of frames in articles. We report macro-averaged precision and recall over the five frames, assigning equal importance to each frame label; as well as their harmonic mean (F1 score).

<sup>12</sup>The original MixMatch additionally applies data augmentation to enhance consistency, which we disable due to the difficulty of DA in language, and for model simplicity.

<sup>13</sup>In preliminary experiments we tested truncation at {125, 256, 512 and 1024}. We found that 256 worked best, presumably reflecting the pyramid structure of news with important information presented upfront/in the lead paragraphs. Very few articles were longer than 1024 tokens.

RE	(1) However the study finds that no single solution will avert the dangers, so <b>a combined approach is needed</b> . (2) The key element is that these three <b>solutions must be implemented together</b> ." (3) We also looked at increasing <b>the efficiency of water use</b> , and we looked at <b>better monitoring and recycling of fertiliser</b> - lots of it is lost and it runs off into rivers and causes dead zones in the oceans."
CO	(1) <b>Their answers – and reactions to them</b> – foreshadowed the <b>fight</b> ahead with conservatives and industry regardless of who becomes the next president. (2) Democrats vying for president <b>revealed a fundamental split over</b> how <b>aggressively</b> the US should tackle climate change [...] in a seven-hour town hall meeting on Wednesday. (3) [...] held after the Democratic National Committee <b>refused to sanction an official climate debate between candidates</b> and amid unprecedented <b>pressure from young activists and the Democratic voting base</b> to tackle the climate crisis.
HI	(1) <b>To Janet</b> , this is a moral issue. (2) These were matters that we have historically agreed on, if for no other reason than <b>the sake of our children and grandchildren’s future</b> . (3) And <b>in Jordan’s case</b> , that would be Social Security, Medicare, education, health care and the like: Programs that benefit folks in his district.
MO	(1) It is, after all, the <b>measure of one’s moral fitness to value some things</b> (say, forgiveness) over others (vengeance). (2) <b>When organized religion fades</b> and its would-be adherents are left to search for meaning, does the <b>god of the environment</b> end their <b>search for a moral authority</b> ? (3) That statement not only describes <b>Judas’s moral disorder</b> but also reminds the audience that any concern, <b>holy as it may be</b> — poverty reduction, environmental protection, or any other <b>leathly mission</b> — that does not give a preferential deference to <b>God, His creation, and acts of beauty such as that of Mary Magdalene</b> are sure signs of misaligned priorities.
EC	(1) Both can impact the relative <b>financial attractiveness of future energy options</b> . (2) "Even Milton Friedman understood the existence of market externalities, the fact that damage to our environment is <b>not accounted for in the free market</b> without <b>placing some sort of price signal</b> ." (3) They end up <b>helping certain wealthy people to the disadvantage of the less fortunate</b> ."

Table 4: Three top relevant sentences (right) extracted by RBF for articles which were correctly predicted as containing the frame (left), in order of decreasing relevance. Highlights of relevant phrases manually added in bold.

### 5.3 Main results

Table 3 shows the frame prediction results. All models significantly outperform the random and majority baselines. All neural methods perform better than the non-neural KNN. BERT performs worse than the Longformer, presumably due to the Longformer’s higher capacity with  $1.5\times$  the parameters of BERT. RBF is best overall, suggesting that combining Longformer embeddings with a relevance based sentence retrieval backbone helps the models to focus on frame-relevant context.

We ablate the impact of the different channels in RBF in Table 3 (bottom). The model performance drops with the removal of each input channels, suggesting that the input channels are complementary and each contributes to the performance. Snippet and RBF perform comparably, with inverse emphasis on precision and recall, however, only RBF offers explicit evidence for prediction (which we explore in the next section). A semi-supervised extension of RBF is a promising avenue for future work.

Given the multi-label nature of our data set, a natural question is how often models predict all and only the annotated frames for an article (exact match). RBF does so 18% of the time. Appendix H provides more detailed results and analyses of per-frame and per-label performance.

#### 5.3.1 Qualitative analysis

For each frame, we inspect sentences retrieved as highly relevant by RBF. Table 4 displays these sentences. We boldfaced the most relevant phrases for ease of exposition. The selected sentences align closely with the definition of each frame: for Human Interest, they refer to the struggle of affected individuals and evoke empathy; for Moral they refer to god, religion and moral values; and for Resolution they mention explicit solutions. One intriguing direction for future work will be to study the differences in *manifestation* of different frames across outlets from different sides along the political spectrum.

## 6 Discussion

What are the recurring narratives that frame the public discourse about contested issues like climate change? Existing NLP approaches to frame prediction fall short of answering this question due to a focus on localized signals. Drawing on theories from the social and communicative sciences, we introduced a set of narrative framing devices to NLP, and integrated them with narrative roles assigned to the central entities in the news articles.

We applied our framework to the issue of climate change, and annotated  $>400$  English-language news articles from major outlets with different polit-



ical leanings with multi-label frames and narrative roles of entities and their stakeholder categories. Our exploratory analysis demonstrated how our framework can be utilized to study multiple levels of framing, including differences across outlets; co-occurrence of frames and narrative roles; and assignment of narrative roles to stakeholder categories.

With the ultimate goal of scaling such analyses to larger, unlabeled data sets, we introduced RBF, an effective and interpretable retrieval-based frame classifier. The ‘explicit attention’ module of RBF not only improved performance over its backbone, the vanilla Longformer, but also naturally provides evidence for its predictions as a list of relevance-ranked article sentences.

Our work addresses a disconnect between the complexity of framing acknowledged in the communication science literature, and models of framing in NLP. As recently surveyed by [Ali and Hassan \(2022\)](#), NLP approaches to framing predominantly focus on topic models or frequency-based methods, leaning heavily on local lexical signals as indicators for the presence or absence of a single frame per unit of analysis. However, the framing of a news article typically emerges from indicators spread throughout the text; frames can co-exist and interact with each other within a single news story. This paper takes one step towards such an integrated notion of framing in NLP in considering narrative frames and roles at the article level and adopting multi-label task formalization.

Our work and results suggest many avenues for future research. The Narrative Frames Corpus supports research on joint models of framing and entity roles: the presence of an entity with a specific role (e.g., the Hero) should render the presence of certain frames (e.g., Resolution) more likely. Conversely, frames like Conflict impact the probability of the existence of the number and kind of different roles (e.g., the Hero and the Villain). A joint model of frames and narrative roles could incorporate role labels with soft confidence weights as latent signal into a frame classification model.

Annotator (dis)agreement and aggregation of answers to indicator questions into frame tags provide fertile grounds for future work. We echo a line of recent work on acknowledging label variation as a signal of genuine complexity rather than noise. This holds true particularly for complex tasks like frame annotation which inevitably retain a level

of subjective variation ([Pavlick and Kwiatkowski, 2019](#); [Plank, 2022](#)). In this paper we aggregated indicator labels into a hard frame label by voting, however we release the raw annotations as part of our data set. Future work could explore soft aggregation methods, delineate genuine variation from noise, and adopt disagreement-aware models and evaluation metrics.

The comparatively small size of the Narrative Frames Corpus and the competitive semi-supervised Snippet suggest further exploration of semi-supervised approaches. Integrating RBF with a Snippet-inspired semi-supervised framework, most simply by soft labeling articles as a function of their retrieved sentences and RBF relevance scores, would allow to leverage large unlabeled data sets while retaining RBF’s interpretability. Alternatively, one could adapt models from different domains, for instance by drawing on the literature of modeling narrative roles in folk tales ([Valls-Vargas et al., 2014](#); [Jahan et al., 2021](#)).

## 7 Limitations

We acknowledge a range of limitations of our work.

As discussed in Sections 3 and 6, overall annotator agreement ranged from fair (frame annotations) to low (entity role annotations). We do not view this as a limitation per se, again pointing to the recent literature on the value of human label variance pointing at a potential loss of valuable information if we overly focus on arriving at a single gold label per instance, with high confidence ([Plank, 2022](#); [Pavlick and Kwiatkowski, 2019](#)). Future modeling work involving entity labels should, however, carefully inspect the role label variation, and potentially remove or aggregate selected annotations, before incorporating the labels as signal into predictive models. We explicitly refrained from training a model in this paper to avoid the risk of training a predictor on an unfavorable noise-to-signal ratio.

Our data set focuses on English-language news reports, sampled from 2017 to 2019 in mainstream media outlets in the US and UK, and as such focuses on cultures and communities which are already well-resourced and well studied. With climate change being a global challenge, broadening data sets, annotations and models to more languages is an important direction for future work. We explicitly caution against projecting annotations across languages without careful validation as we expect the manifestation of framing, views

on entities (or sheer set of dominant entities) to vary widely across countries and communities.

Even within our English study, we acknowledge that the size of annotated data is small for NLP scales, and an extension in the future is desirable. A related current limitation is the focus on just a single issue (climate change) and validation of our narrative framing framework for other issues is an important direction for the future. Finally, the annotation process was slow and costly, relying on trained, highly educated annotators with constant monitoring, rendering larger scale annotations challenging, on the one hand. On the other hand, we will release upon acceptance our annotation procedure including the full codebook with instructions, which have been optimized over several rounds of annotations and we hope can support more efficient annotation in the future.

### Ethics statement

This study was approved by the University of Melbourne ethics board (Human Ethics Committee LNR 3B), Reference Number 2023-22109-37029-4, and data acquisition and analysis has been taken out to the according ethical standards. We hired four local annotators who were paid an hourly rate of \$53 AU in line with the casual research assistant hourly rates set up in the University of Melbourne collective agreement.

We will release the Narrative Framing Corpus comprising of 428 news articles annotated with frame labels, entities, their narrative roles and stakeholder categories. We also publish the raw (non-aggregated) annotations. Our data set builds on news articles from the NELA corpora 2017-2019, which were released to the public domain (license CC0 1.0).<sup>14</sup> We release our code and Narrative Frames Corpus under a MIT license.

### Acknowledgements

We gratefully acknowledge the work of our annotators Candy Chu, Aarushi Kaul, Dana Pjanic, and Lloyd Rouse and the helpful feedback of the anonymous reviewers.

### References

Katrina Ahchong and Rachel Dodds. 2012. Anthropogenic climate change coverage in two canadian

<sup>14</sup>2017: <https://doi.org/10.7910/DVN/ZCXSKG>; 2018: <https://doi.org/10.7910/DVN/ULHLCB>; 2019: <https://doi.org/10.7910/DVN/O7FWPO>

newspapers, the toronto star and the globe and mail, from 1988 to 2007. *Environmental Science & Policy*, 15(1):48–59.

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.

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

Iz Beltagy, Matthew E Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*.

David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. 2019. Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, 32.

Shraey Bhatia, Jey Han Lau, and Timothy Baldwin. 2021. [Automatic classification of neutralization techniques in the narrative of climate change scepticism](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2167–2175, Online. Association for Computational Linguistics.

Lindsey M Bier, Sejin Park, and Michael J Palenchar. 2018. Framing the flight mh370 mystery: A content analysis of malaysian, chinese, and us media. *International Communication Gazette*, 80(2):158–184.

Benjamin D Blair and Larkin McCormack. 2016. Applying the narrative policy framework to the issues surrounding hydraulic fracturing within the news media: A research note. *Research & Politics*, 3(1):2053168016628334.

Terra Blevins and Luke Zettlemoyer. 2020. [Moving down the long tail of word sense disambiguation with gloss informed bi-encoders](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1006–1017, Online. Association for Computational Linguistics.

Timothy A Brown and Michael T Moore. 2012. Confirmatory factor analysis. *Handbook of structural equation modeling*, 361:379.

Dallas Card, Amber E. Boydston, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. [The media frames corpus: Annotations of frames across issues](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language*

- Processing (Volume 2: Short Papers)*, pages 438–444, Beijing, China. Association for Computational Linguistics.
- Kaiping Chen, Amanda L Molder, Zening Duan, Shelley Boulianne, Christopher Eckart, Prince Mallari, and Diyi Yang. 2023. How climate movement actors and news media frame climate change and strike: evidence from analyzing twitter and news media discourse from 2018 to 2021. *The International Journal of Press/Politics*, 28(2):384–413.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- Astrid Dirikx and Dave Gelders. 2010. To frame is to explain: A deductive frame-analysis of dutch and french climate change coverage during the annual conferences of the parties. *Public understanding of science*, 19(6):732–742.
- Devin M Dotson, Susan K Jacobson, Lynda Lee Kaid, and J Stuart Carlton. 2012. Media coverage of climate change in chile: A content analysis of conservative and liberal newspapers. *Environmental Communication: A Journal of Nature and Culture*, 6(1):64–81.
- Lauren Feldman, P Sol Hart, and Tijana Milosevic. 2017. Polarizing news? Representations of threat and efficacy in leading US newspapers’ coverage of climate change. *Public Understanding of Science*, 26(4):481–497.
- 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 *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3570–3580, Brussels, Belgium. Association for Computational Linguistics.
- Federica Fornaciari. 2012. Framing the Egyptian Revolution: A content analysis of Al Jazeera English and the BBC. *Journal of Arab & Muslim Media Research*, 4(2-3):223–235.
- R Edward Freeman. 1984. Stakeholder management: framework and philosophy. *Pitman, Mansfield, MA*.
- Maurício Gruppi, Benjamin D. Horne, and Sibel Adali. 2020. **Nela-gt-2019: A large multi-labelled news dataset for the study of misinformation in news articles**.
- Nardia Haigh and Andrew Griffiths. 2009. The natural environment as a primary stakeholder: the case of climate change. *Business Strategy and the Environment*, 18(6):347–359.
- Mareike Hartmann, Tallulah Jansen, Isabelle Augenstein, and Anders Søgaard. 2019. **Issue framing in online discussion fora**. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1401–1407, Minneapolis, Minnesota. Association for Computational Linguistics.
- Benjamin D Horne, Sara Khedr, and Sibel Adali. 2018. Sampling the news producers: A large news and feature data set for the study of the complex media landscape. In *Twelfth International AAAI Conference on Web and Social Media*.
- Labiba Jahan, Rahul Mittal, and Mark Finlayson. 2021. Inducing stereotypical character roles from plot structure. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 492–497.
- 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.
- Tianyu Jiang and Ellen Riloff. 2021. **Exploiting definitions for frame identification**. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2429–2434, Online. Association for Computational Linguistics.
- Kristen Johnson, Di Jin, and Dan Goldwasser. 2017. Modeling of political discourse framing on Twitter. In *Eleventh International AAAI Conference on Web and Social Media*.
- Shima Khanehzar, 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.
- Jeesun Kim and Wayne Wanta. 2018. News framing of the us immigration debate during election years: Focus on generic frames. *The Communication Review*, 21(2):89–115.
- Andrea Lawlor and Erin Tolley. 2017. Deciding who’s legitimate: News media framing of immigrants and refugees. *International Journal of Communication*, 11:25.
- Nayeon Lee, Yejin Bang, Tiezheng Yu, Andrea Madotto, and Pascale Fung. 2022. **NeuS: Neutral multi-news summarization for mitigating framing bias**. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3131–3148, Seattle, United States. Association for Computational Linguistics.



- Cyril Levitt. 1981. "The Whole World Is Watching: Mass Media in the Making and Unmaking of the New Left." *JSTOR*.
- 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.
- Julia Lück, Hartmut Wessler, Antal Wozniak, and Diógenes Lycarião. 2018. Counterbalancing global media frames with nationally colored narratives: A comparative study of news narratives and news framing in the climate change coverage of five countries. *Journalism*, 19(12):1635–1656.
- 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.
- Julia Mendelsohn, Ceren Budak, and David Jurgens. 2021. [Modeling framing in immigration discourse on social media](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2219–2263, Online. Association for Computational Linguistics.
- Zhengjie Miao, Yuliang Li, Xiaolan Wang, and Wang-Chiew Tan. 2020. Snippet: Semi-supervised opinion mining with augmented data. In *Proceedings of The Web Conference 2020*, pages 617–628.
- W Russell Neuman, Russell W Neuman, Marion R Just, and Ann N Crigler. 1992. *Common knowledge: News and the construction of political meaning*. University of Chicago Press.
- Matthew C Nisbet. 2009. Communicating climate change: Why frames matter for public engagement. *Environment: Science and policy for sustainable development*, 51(2):12–23.
- Jeppe Nørregaard, Benjamin D Horne, and Sibel Adali. 2019. Nela-gt-2018: A large multi-labelled news dataset for the study of misinformation in news articles. In *Proceedings of the international AAAI conference on web and social media*, volume 13, pages 630–638.
- Zhongdang Pan and Gerald M Kosicki. 1993. Framing analysis: An approach to news discourse. *Political communication*, 10(1):55–75.
- Ellie Pavlick and Tom Kwiatkowski. 2019. [Inherent Disagreements in Human Textual Inferences](#). *Transactions of the Association for Computational Linguistics*, 7:677–694.
- Pew Research Center. 2022. [Climate change remains top global threat across 19-country survey](#). Technical report, Washington, D.C.
- Barbara Plank. 2022. The 'problem' of human label variation: On ground truth in data, modeling and evaluation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*.
- Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. 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 3982–3992.
- Holli A Semetko and Patti M Valkenburg. 2000. Framing european politics: A content analysis of press and television news. *Journal of communication*, 50(2):93–109.
- Elizabeth A Shanahan, Michael D Jones, Mark K McBeth, and Claudio M Radaelli. 2018. The narrative policy framework. In *Theories of the policy process*, pages 173–213. Routledge.
- Stuart Soroka, Patrick Fournier, and Lilach Nir. 2019. Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences*, 116(38):18888–18892.
- Gregg Sparkman, Nathan Geiger, and Elke U Weber. 2022. Americans experience a false social reality by underestimating popular climate policy support by nearly half. *Nature communications*, 13(1):1–9.
- Roopal Vaid, Kartikey Pant, and Manish Shrivastava. 2022. [Towards fine-grained classification of climate change related social media text](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 434–443, Dublin, Ireland. Association for Computational Linguistics.
- Josep Valls-Vargas, Jichen Zhu, and Santiago Ontanón. 2014. Toward automatic role identification in unannotated folk tales. In *Tenth Artificial Intelligence and Interactive Digital Entertainment Conference*.
- Rachel Wetts. 2020. In climate news, statements from large businesses and opponents of climate action receive heightened visibility. *Proceedings of the National Academy of Sciences*, 117(32):19054–19060.
- Erika Allen Wolters, Michael D. Jones, and Kathryn Duvall. 2022. A narrative policy framework solution to understanding climate change framing research. In Michael D. Jones, Mark K. McBeth, and Elizabeth A. Shanahan, editors, *Narratives and the Policy Process: Applications of the Narrative Policy Framework*, chapter 9, pages 218–239. MT: Montana State University Library.
- Jing Yi Xie, Renato Ferreira Pinto Junior, Graeme Hirst, and Yang Xu. 2019. [Text-based inference of moral sentiment change](#). 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 4654–4663, Hong Kong, China. Association for Computational Linguistics.

Caleb Ziems, Minzhi Li, Anthony Zhang, and Diyi Yang. 2022a. [Inducing positive perspectives with text re-framing](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3682–3700, Dublin, Ireland. Association for Computational Linguistics.

Caleb Ziems, Jane Yu, Yi-Chia Wang, Alon Halevy, and Diyi Yang. 2022b. [The moral integrity corpus: A benchmark for ethical dialogue systems](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3755–3773, Dublin, Ireland. Association for Computational Linguistics.

## A Article selection

We identify climate-related news articles in the NELA corpora (2017–2019) by searching for keywords identified in the Wikipedia climate change glossary<sup>15</sup>. A few generic terms were removed ('weather') as too broad for our query. We consider an article as relevant either if  $\geq 1$  keywords are found in the article title or  $\geq 3$  mentions of climate keywords are found in the article body.

## B Codebook

Our full codebook of 22 fine-grained questions presented to annotators. The questions used in the final factor analysis model and subsequent analyses are boldfaced. We show the indicator distribution, i.e., the fraction of annotated articles in which two or more annotators answered 'yes' to a given question, in brackets.

ID	Question	Annotation rules
RE1 (0.51)	<b>2. Does the story suggest a solution(s) to the issue/problem?</b>	<b>Mark 'yes' if (a) solution(s), or a strategy to mitigate the problem, is explicitly mentioned.</b>
RE2	3. Is this problem/issue resolved in the story?	Mark 'yes' if the story explicitly mentions that the problem has been resolved.
RE3	4. Is there any hope in the story for future resolution of the problem/issue?	Mark 'no' if the story is about a failed attempt to tackle the issue under discussion.
RE4	5. Does the story suggest that the issue/problem requires urgent action?	Mark 'yes' if there is an explicit call for action, or ongoing efforts to alleviate the problem are (1) explicitly described and (2) raise a sense of urgency
RE5 (0.55)	<b>6. Does the story suggest that some entity could alleviate the problem? If your answer is "yes", please select the most appropriate entity.</b>	<b>Mark 'yes' if at least one entity in the story is described as actively alleviating, or planning to alleviate, the problem. If multiple options: select the one that's most central/prevalent in the article (in terms of #mentions / mentions in the central parts like title and opening.)</b>
RE6	7. Does the story suggest that some entity is responsible for the issue/problem? If your answer is "yes", please select the most responsible entity.	Mark 'yes' if at least one entity in the story is described as actively causing, or having caused, the problem. If multiple options: select the one that's most central/prevalent in the article (in terms of the number of mentions / mentions in the central parts like title and lead paragraphs).

<sup>15</sup>[https://en.wikipedia.org/wiki/Glossary\\_of\\_climate\\_change](https://en.wikipedia.org/wiki/Glossary_of_climate_change)

<b>HI1 8. Does the story provide a (0.27) human example or a "human face" on the problem/issue?</b>	Select 'yes' if the story uses "dramatization" i.e., explicitly refers to how the issue impacts the personal life living entities (including animals).
<b>HI2 9. Does the story employ ad- (0.21) jectives or personal vignettes that generate feelings of outrage, empathy-caring, sympathy, or compassion?</b>	Mark 'yes' if the story uses emotional language (original Q is already very concrete).
<b>HI3 10. Does the story emphasize how one or more entities are NEGATIVELY affected by issue/problem? If your answer is "yes", please select the most negatively affected entity.</b>	Select 'yes' if the story explicitly refers to how one or more entity/ies suffer from the problem/issue.
<b>HI4 11. Does the story emphasize how one or more entities are POSITIVELY affected by the issue/problem? If your answer is "yes", please select the most positively affected entity.</b>	Select 'yes' if the story explicitly refers to how one or more entity/ies benefit from the problem/issue.
<b>HI5 12. Does the story go into the (0.07) private or personal lives of the entities involved?</b>	Mark 'yes' if the story explicitly refers to the personal life of at least one entity.
<b>CO1 13. Does the story reflect (0.73) disagreement between political parties/ individuals /groups/ countries?</b>	Select 'yes' even if the story describes a disagreement or a conflict in a passive/observational manner.
<b>CO2 14. Does one party/ individual/ (0.40) group/ country reproach another?</b>	Select 'yes' if the story explicitly refers to the active conflict between two or more entities – past or present.
<b>CO3 15. Does the story refer to two (0.54) sides or more than two sides of the problem or issue?</b>	Select 'yes' if the story explicitly mentions at least two viewpoints on the current issue (even if they're not presented in a balanced, objective manner).
<b>CO4 16. Does the story refer to winners and losers? If your answer is "yes", please select the most appropriate winner/loser entity.</b>	Select 'yes' if the story explicitly refers to one or more 'winners' and/or 'losers' which emerged from an active conflict/argument/war. Note, in some stories an entity can be both a winner and a loser.
<b>MO1 17. Does the story contain any (0.07) moral message?</b>	Select 'yes' if the story explicitly applies standards or judgments of right or wrong to entities, actions or events.
<b>MO2 18. Does the story make ref- (0.05) erence to morality, God, and other religious tenets?</b>	Select 'yes' if the story explicitly refers to religious tenets or moral obligations framed through the lens of obligations to a spiritual community. Select 'yes' also if the mention is indirect e.g., through a quote or a metaphor.
<b>MO3 19. Does the story offer specific social prescriptions about how to behave?</b>	Select 'yes' if the story explicitly mentions expectations around norms of conduct, limitations or prohibitions on actions or events.

<b>EC1</b> (0.28) <b>20. Is there a mention of financial losses or gains now or in the future?</b>	Select ‘yes’ if the story explicitly refers to the financial impacts of the issue.
<b>EC2</b> (0.37) <b>21. Is there a mention of the costs/degree of the expense involved?</b>	Select ‘yes’ if the story explicitly refers to the amount of loss or gain (e.g., “\$100,000”, “enormous cost”).
<b>EC3</b> (0.25) <b>22. Is there a reference to the economic consequences of pursuing or not pursuing a course of action?</b>	Select ‘yes’ if the story explicitly mentions the impacts of action or inaction on the economy.

## C Annotation Interface

Figure 4 shows our annotation interface split into an answer form (left) and an article display with (optional) highlighting of prevalent entities (right, in color).

**Framing Annotation**

Welcome victoria !

You are annotating article B\_347: [Plain text](#) and [With highlighted entities !](#)

0: Does this article predominantly (>70%) discuss a problem/issue related to climate change?

yes  
 no

1: Does the article title reflect the core problem/issue underlying the article?

yes  
 no

If you selected "yes", paste original title here:   
If you selected "no", type alternative title here:

2: Does the story suggest a solution(s) to the issue/problem?

yes  
 no

3: Is the problem/issue resolved in the story?

yes  
 no

4: Is there any hope in the story for future resolution of the problem/issue?

yes  
 no  
 n/a: problem is resolved

5: Does the story suggest that the issue/problem requires urgent action?

yes  
 no

6: Does the story suggest that some entity could alleviate the problem?

yes  
 no

House **lawmakers 2** sent a **letter 1** Monday to the environmental **group 0** Earthjustice demanding documents regarding **its 0** ties to foreign officials and environmental activists , The Daily Caller News **Foundation 6** has learned .

The **letter 1** from top Republicans on the House Committee on Natural Resources is the fourth sent to environmental groups over their ties to foreign governments . **It 1** 's the second **letter 1** related to environmental opposition to the U.S. military presence in Okinawa , Japan .

**Lawmakers 2** say **Earthjustice 0** 's political activities may require **them 2** to register as a foreign agent , according to a copy of the **letter 1** obtained by **TheDCNF 6** . **Earthjustice 0** is an environmental law firm often represents environmental activists in litigation . RELATED : The Trump Administration Will Side With Environmentalists And Defend An Obama - Era Policy Before The US Supreme Court

\* The Committee is concerned that your **organization 0** 's political activities within the United States in opposition to the **relocation 4** of Marine Corps Air Station Futenma and the continued military presence in Okinawa may require compliance with " the Foreign Agents Registration Act , Republicans wrote to **Earthjustice 0** president Abigail Dillen .

The **committee 1** 's letter to **Earthjustice 0** stems from a **letter 1** sent in June to the **Center 7** for Biological Diversity over **its 7** opposition to the **relocation 4** of the Futenma base . **Environmentalists 8** say the proposed relocation site would harm the dugong , an endangered marine mammal .

Figure 4: Our annotation interface. Left: Excerpt of the annotation form which covers the 22 binary indicator questions, and free-text fields to record role-specific entities. Right: the news article with prevalent entities highlighted (based on automatic entity recognition and co-reference resolution.)

## D Media Outlets

Table 5 lists all media outlets in the labeled data set, together with number of articles and MBFC political leaning.

## E Entity groups

Table 6 lists our set of entity groups, together with some representative examples and the total number (tokens) of instances assigned to each group.

## F Label statistics

Figure 5 shows label distributions in our data set. In terms of prevalence of our five frames individually, Conflict is present in 63%; Resolution in 45%; Economic in 30%; Human Interest in 11% and



**left\_bias:** thehuffingtonpost (47), vox (22), cnn (18), politicususa (16), motherjones (10), shareblue (5), dailykos (4), talkingpointsmemo (3), slate (2), palmerreport (1)

**left\_center\_bias:** bbc (22), theguardian (20), npr (14), usatoday (10), cbsnews (9), yahoonews (9), thenewyorktimes (8), pbs (4), fusion (4), cnbc (3), thehill (3)

**right\_bias:** thedailycaller (39), drudgereport (27), foxnews (13), theblaze (10), nationalreview (9), redstate (6), newsbusters (4), thepoliticalinsider (2), therightscoop (1)

**questionable\_source:** Breitbart (38), rt (18), dailymail (12), bipartisanreport (2), thedurant (1)

Table 5: Media outlets in our data, grouped by their political leaning (based on the Media Bias Fact Check portal; rows) and with associated article count (in brackets).

Entity Group	Count	Example entities
GOVERNMENTS_POLITICIANS_POLIT.ORGs	828	democrats, trump, the EPA
INDUSTRY_EMISSIONS	271	fossil fuels, carbon pollution, plastic, capitalism
LEGISLATION_POLICIES_RESPONSES	251	US climate response, Green New Deal, paris agreement
GENERAL PUBLIC	241	americans, indigenous people, public health
ANIMALS_NATURE_ENVIRONMENT	213	air quality, the ocean, the earth
ENV.ORGs_ACTIVISTS	105	Greta Thunberg, youth activists, extinction rebellion
SCIENCE_EXPERTS_SCI.REPORTS	59	the stupid bloody academics, professors, University of Auckland study
CLIMATE CHANGE	52	climate change, global warming
OTHER	49	meat consumption, resilience, god
AMBIGUOUS	37	deniers, response
GREEN TECHNOLOGY_INNOVATION	31	wind energy, geoengineering, renewables
MEDIA_JOURNALISTS	26	journalists, media outlets, CNN

Table 6: Stakeholder groups with token count and examples of assigned entities.

Moral in 10% of all annotated articles.

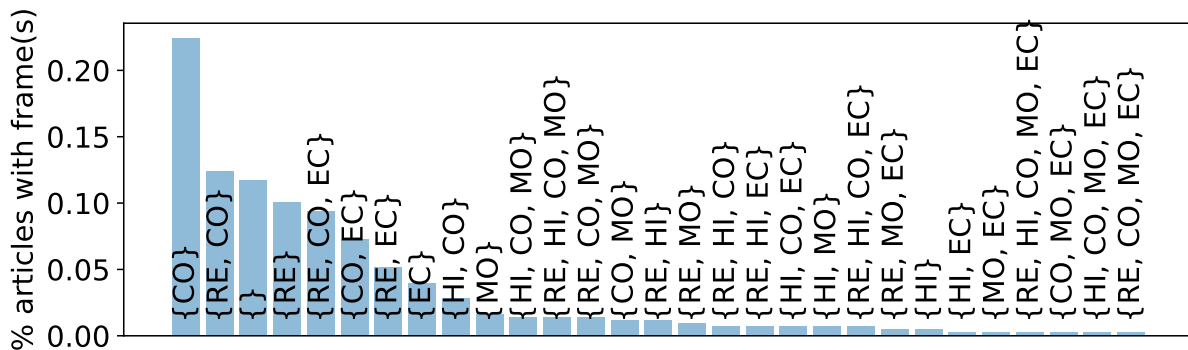


Figure 5: Distribution of the most frequent frame labels.

## G Hyperparameters, model sizes and compute costs

Table 7 lists the hyperparameters for all neural models, and their total number of parameters. For KNN,

Item	RBF	BERT	Longformer	Snippext
random_seed	1042	1042	1042	–
learning_rate	2e-6	2e-6	2e-6	2e-6
max_len	256	256	256	256
batch_size	8	16	16	8
epochs	20	20	20	20
total # parameters	149 M	109 M	149 M	149M

Table 7: Hyper parameters and model size for all neural models.

K was tuned on the dev set for each fold. L2 distance was used as a distance function and for all other parameters default values were used from this implementation: <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>.

All neural models were trained on a single Nvidia A100 GPU. The total compute budget covering all experiments were 26 hours.

## H Additional Frame Prediction Results

**Frame-wise analysis** Table 8 shows frame-wise prediction performance averaged over five runs, for the three best performing models in Section 5.3. We observe that Human Interest (HI) is the most challenging frame for all models, while access to the unlabelled data appears to be effective to boost performance for Snippext on this challenging class. The most prevalent Conflict (CO) frame is predicted most reliably.

	Longformer			RBF			Snippext		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
RE	0.52	0.80	0.63	0.52	0.83	0.64	0.54	0.71	0.61
HI	0.22	0.28	0.24	0.20	0.59	0.30	0.48	0.47	0.48
CO	0.70	0.92	0.80	0.70	0.89	0.78	0.76	0.83	0.80
MO	0.50	0.46	0.48	0.51	0.69	0.59	0.51	0.55	0.53
EC	0.46	0.70	0.55	0.63	0.80	0.70	0.51	0.68	0.58

Table 8: Frame-level prediction results of the three best performing models in Section 5.3, averaged over five folds.

**Label-wise performance** We analyze for RBF, which multi-labels are predicted best (worst). Overall, RBF achieves 18% exact match accuracy. Table 9 lists the best (left) and worst (right) performing multi-labels, with an occurrence (N) of at least three in the gold labeled data set. We can see, in line with Table 8 that the model struggles with multi-labels involving the Human Interest or Moral frame. We also find that performance does not correlate with the number of frames: both the top and bottom performing sets include single-labels and high multi-labels.

### H.1 Error analysis

**Overprediction of Human Interest** We first analyze the false positive prediction of Human Interest (HI). We list sentences extracted by RBF most confidently extracted as evidence for an incorrect positive HI label, some examples include context for clarity, in gray:

Label	Best		Label	Worst	
	N	% correct		N	% correct
RE;CO;EC	40	0.4	HI;CO	12	0.0
RE;CO	53	0.34	MO	7	0.0
RE;HI;CO;EC	3	0.33	RE;CO;MO	6	0.0
RE	43	0.21	RE;HI;CO;MO	6	0.0

Table 9: The five best-predicted multi-labels by RBF (left) and worst-predicted (right) with >1 occurrences in the labeled data. For each label we show its prevalence in the data set (N) and % predicted exact match.

- (i) *“In its research, SPARK Neuro measured physiological data such as brain activity and palm sweat to quantify people’s emotional reactions to stimuli.”* (Breitbart, 2019-09-02)
- (ii) *“‘Something must have changed in the debate that so many young people are speaking up and so many young people are being targeted,’ Thunberg told Yahoo News in response to the mockery her movement has received from leaders like Trump and Russian President Vladimir Putin. ‘They can sense that we are making an impact’ ”* (Yahoo News, 2019-10-17)
- (iii) *“[...] it might be tempting to criticize or dismiss activists supporting it. But Amy Myers Jaffe hopes older, more experienced policymakers won’t do that. ‘We need not to discourage them,’ she says. ‘They have an energy and will to innovation that is not only infectious, but inspiring.’ ”* (NPR, 2019-02-08)<sup>16</sup>

Sentence (i) refers to emotions, which are often linked to a human interest perspective, but not in this quote which refers to a scientific study. In case (ii) ‘they’ refers to members of the Russian government, pointing to an error in contextualization. Resolving ‘they’ as a cataphora to ‘so many young people’ may be plausible in a superficial, yet incorrect read, as evidence for a Human Interest frame. Sentence (iii) is about shaping a strategic response to activists, and does not imply any human impact or emotion.

**Overprediction of Resolution** is often due to quotes, or criticisms, of previously proposed policies without framing them as actual solutions in the context of the article, for instance

- (iv) *“In the 2016 presidential election, Sanders staked out the most ambitious climate platform of any candidate, vowing to slash carbon dioxide pollution 40 percent by 2030, end fossil fuel subsidies and ban fracking. ”* (Huffington Post, 2018-04-12)

where Sanders’ ambitions are discussed in the context election campaigns of different candidates.

**Underprediction of Moral** . Moral issues can take very different forms, for instance an article from The Daily Caller (2018-12-12) refers to the importance of *‘tolerance of [a] diversity of viewpoints’*, and automatic prediction of a link to moral questions is near impossible without a fundamental understanding of the concept of morality. Millenia of theoretical arguments on the conceptualization of moral, and a much shorter yet active stream of research in NLP/ML confirms the intrinsic challenge with this frame (Xie et al., 2019; Ziems et al., 2022b). Our model to a large extent relies on explicit mentions of the keyword ‘moral’ or religious terms.

## ACL 2023 Responsible NLP Checklist

---

### A For every submission:

- A1. Did you describe the limitations of your work?  
*Section 7*
- A2. Did you discuss any potential risks of your work?  
*Section 7*
- A3. Do the abstract and introduction summarize the paper’s main claims?  
*Section 1 and abstract*
- A4. Have you used AI writing assistants when working on this paper?  
*Left blank.*

### B Did you use or create scientific artifacts?

*Section 3*

- B1. Did you cite the creators of artifacts you used?  
*We introduce (and use) a new artifact / data set*
- B2. Did you discuss the license or terms for use and / or distribution of any artifacts?  
*the ethics statement*
- B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?  
*limitations section (7)*
- B4. Did you discuss the steps taken to check whether the data that was collected / used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect / anonymize it?  
*Not applicable. we use only publicly available news articles*
- B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?  
*Sections 1 and 3*
- B6. Did you report relevant statistics like the number of examples, details of train / test / dev splits, etc. for the data that you used / created? Even for commonly-used benchmark datasets, include the number of examples in train / validation / test splits, as these provide necessary context for a reader to understand experimental results. For example, small differences in accuracy on large test sets may be significant, while on small test sets they may not be.  
*Left blank.*

### C Did you run computational experiments?

*Section 5*

- C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?  
*Appendix G*

---

*The Responsible NLP Checklist used at ACL 2023 is adopted from NAACL 2022, with the addition of a question on AI writing assistance.*



- C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?  
*Appendix G*
- C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?  
*Section 5, we report averages + variance over five runs and a qualitative analysis. Further results and error analyses in Appendix H.*
- C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?  
*Appendix G*
- D**  **Did you use human annotators (e.g., crowdworkers) or research with human participants?**  
*Section 3*
- D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?  
*Codebook in appendix B, annotation interface screenshot in Appendix C. Full instructions as part of the data set (supp material)*
- D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?  
*Section 3 and ethics statement*
- D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?  
*Consent form is provided as part of the data set (supp material)*
- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?  
*Ethics statement. The study was approved by the university*
- D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?  
*Section 3*