# Multi-Modal Framing Analysis of News

**Arnav Arora*** **Srishti Yadav***
**Maria Antoniak** **Serge Belongie** **Isabelle Augenstein**
University of Copenhagen
{aar, srya, maan, s.belongie, augenstein}@di.ku.dk

## Abstract

Automated frame analysis of political communication is a popular task in computational social science that is used to study how authors select aspects of a topic to *frame* its reception. So far, such studies have been narrow, in that they use a fixed set of pre-defined frames and focus only on the text, ignoring the visual contexts in which those texts appear. Especially for framing in the news, this leaves out valuable information about editorial choices, which include not just the written article but also accompanying photographs. To overcome such limitations, we present a method for conducting multi-modal, multi-label framing analysis at scale using large (vision-) language models. Grounding our work in framing theory, we extract latent meaning embedded in images used to convey a certain point and contrast that to the text by comparing the respective frames used. We also identify highly partisan framing of topics with issue-specific frame analysis found in prior qualitative work. We demonstrate a method for doing scalable integrative framing analysis of both text and image in news, providing a more complete picture for understanding media bias.

Title: Fear, Politics, and the Fight for Immigration Reform


Article: The immigration debate in the United States has always been politically charged, but in the years since Donald Trump took office, it has become a defining fault line in American society [...] Reports of **overcrowded detention centers**, family separations, and asylum seekers being forced to wait in **dangerous conditions in Mexico** drew widespread condemnation. Human rights groups labeled the policies cruel, while immigration hardliners defended them as necessary steps to secure the border. [...] There was also an **economic reality** to consider. Millions of undocumented immigrants in the U.S. contribute to key industries like agriculture, construction, and food service. Crackdowns on immigrant labor led to **workforce shortages** in some sectors, raising concerns among **business owners and economists** [...] This rhetoric, however, came at a cost. Studies have shown that immigrants, both documented and undocumented, **commit crimes at lower rates** than native-born citizens. Yet [...]

Figure 1: News can be intentionally *framed* to affect reader perception. Editorial choices decide what is communicated through words and images. Our approach systematically detects this framing.

## 1 Introduction

Frames are conceptual tools that both communicators and audiences use to interpret and categorize issues and events (Gitlin, 1980; Eko, 1999; Pan and Kosicki, 1993; Reese et al., 2001). By highlighting specific elements of a topic and minimizing others, communicators *frame* messages in ways they believe will resonate with audiences (Goffman, 1974) and can shape the way the topic is perceived by readers or viewers (Schudson, 2003). In the field of journalism, framing is a core narrative device by which news consumption is framed within an interpretive perspective (Card et al., 2015). Most prior work on computational frame analysis has focused on linguistic structure and content analyses via text elements (Ali and Hassan, 2022). However, framing is not solely textual; visual elements also play a crucial role in conveying implicit and explicit messages (Messaris and Abraham, 2001).

News media often employ images alongside text to reinforce or contrast the intended frame, leveraging the affective and cognitive impact of visuals (Cope et al., 2005; Keib et al., 2018; Grabe and Bucy, 2009; Geise et al., 2025). An example can be seen in Figure 1. The image depicts protesters

[*] denotes equal contribution



*Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 31531–31553

with signs with police presence around, using the public opinion and security framing. The article text, on the other hand, talks about the quality of life of migrants, economic implications, crime, and policy framing. Such differences in portrayal can affect readers' perception of the crisis and their induced emotions. Further, while communication in text is more explicit through linguistic framing, images encode frames in more subtle and implicit ways, requiring sophisticated interpretation models to capture their meaning (Aiello and Parry, 2019). As such, when conducting media analysis, the communicated framing across both the image and the text should be considered.

While visual framing analysis has been explored in communication science and journalism studies (Wessler et al., 2016; Powell et al., 2017), existing studies doing computational framing analysis have largely ignored this crucial aspect (Ali and Hassan, 2022). Further, they have primarily focused on a fixed set of labels for framing analysis, performing prediction in a multi-class setting. This substantially limits the information one can derive from predictions as an article can convey several frames (Figure 1) and fine-grained analysis of framing within a topic necessitates frames specific to a particular issue. Large vision and language models are particularly well suited for conducting this task at scale, considering the semantic understanding embedded into them through large scale pre-training. This led us to our overall research questions: **(RQ1)** Can LLMs and VLMs reliably detect framing in news articles? **(RQ2)** Are there differences in framing conveyed through text vs the images? and **(RQ3)** How do these framings vary across topics and publishers?

Our **contributions** include the following.

- We present the first computational study of multi-modal framing in the news, outlining a methodology using open-weight LLMs and VLMs.
- We provide a large-scale dataset[1] of 500k U.S.-based news articles for framing analysis, automatically labelled with generic frames, validated through human annotations. We further provide issue-specific frames for article texts, allowing for topic specific fine-grained analysis.
- We conduct a thorough analysis of generic and issue-specific frames , finding meaningful differences in the framing used in images and texts, across political leanings and topics.
- Our approach is scalable, generalisable, robust, reliable, and grounded in framing theory. It allows for multi-modal framing analysis for both social science and NLP scholars.

[1]Released under MIT license: `https://huggingface.co/datasets/copenlu/mm-framing`

## 2 News Framing

Framing of news articles has been studied widely in communication studies. There are several definitions of framing and scholars often disagree on the method for extraction and its role in public communication, leading to it being termed a "fractured paradigm" (Entman, 1993). The definition from Entman for framing at large, however, is the most widely accepted one. He defined framing as *"making some aspects of reality more salient in a text in order to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described"*. Narrowing the framework to news, De Vreese (2005) lays out a typology of news framing. He states that there are *generic* news frames and *issue-specific* news frames. *Generic* news frames *"transcend thematic limitations and can be identified in relation to different topics, some even over time and in different cultural contexts"*. They are particularly useful for uncovering broad patterns within or across countries. Issue-specific frames allow for richer, more fine-grained analysis of various aspects highlighted within a particular issue.

**Framing Through Visuals** Photographs are an important vehicle of framing as a reader may process textual and visual messages differently. Readers may focus on photographs without also reading an accompanying story (Miller, 1975) or might select which news story to read depending on the image thumbnail. Images are potent framing tools because they evoke immediate emotional responses, provide contextual cues, and sometimes contradict textual narratives (Geise and Xu, 2024). For visual data, however, automation is more challenging because computational image analysis often struggles with connotative and symbolic elements that are readily discernible to human annotators (Rodriguez and Dimitrova, 2011). While some recent advancements in machine learning and computer vision offer promising avenues for automated visual framing analysis, the complexity of symbolic and ideological meanings typically requires human interpretation (Coleman, 2010).

**Automated Framing Analysis** In computational studies, framing has been studied with the help of machine learning methods for content analysis at scale. Ali and Hassan (2022); Vallejo et al. (2024); Otmakhova et al. (2024) survey these efforts towards on computational framing analysis, focusing on different aspects like methods, varying conceptualisations and inter-disciplinary connections of framings. The most widely used resource is one by Card et al. (2015), who present the Media Frames Corpus (MFC), a dataset of US based news headlines annotated for generic frames. Most studies use it in a supervised, multi-class prediction setting with MFC frames for analysis of social networks (Mendelsohn et al., 2021) or discussion forums (Hartmann et al., 2019). There are also English-only and multi-lingual SemEval tasks on frame detection (Piskorski et al., 2023a,b; Sajwani et al., 2024). Unsupervised approaches to framing analysis use framing lexica (Field et al., 2018), clustering (Burscher et al., 2016; Ajjour et al., 2019), and topic models (Nguyen, 2015). All of these approaches have focused on texts alone.

**Integrative Framing Analysis** Integrative framing analysis is when both images and text are observed separately, but the results are integrated to form a more complete picture of framing analysis (Dan, 2017). Although there is broad agreement that visual and verbal elements should be studied side by side (Coleman, 2010; Rose, 2022), relatively few studies have effectively combined the two. Understanding multimodal framing is important for several reasons: it allows for a more granular examination of media bias, as textual analysis alone may miss the ideological and emotional undertones of visual elements (Wessler et al., 2016; Geise et al., 2025). It improves fact-checking and detection of misleading content by identifying inconsistencies between textual claims and their supporting images. It also finds applications in political communication, journalism studies, and public policy through tracing changes in framing over time and across outlets (Baumgartner et al., 2008).

## 3 Dataset

We crawl news articles along with corresponding images from 28 US-based news agencies, extracting data for a 12 month period between May 2023 and April 2024 using the newsplease library (Hamborg et al., 2017). Our selected news sources reflect the entire political spectrum, based on data from AllSides Media Bias,[2] an organization that assigns a rating of political leaning to each media outlet. Our list of sources along with corresponding political leanings are listed in Appendix A. We query the publicly available Common Crawl archives[3] for the corresponding publishers and extracted each article's text, headline, publication date, image_urls and other metadata in JSON format. Post scraping, we filter extremely short and long articles, images of logos and other noise. Further details about the filtering process are provided in Appendix C. Our final dataset includes about 500K articles and corresponding images [4]. We show the distribution of the data across the time period of data collection in Figure 3 and across topics in Figure 2.

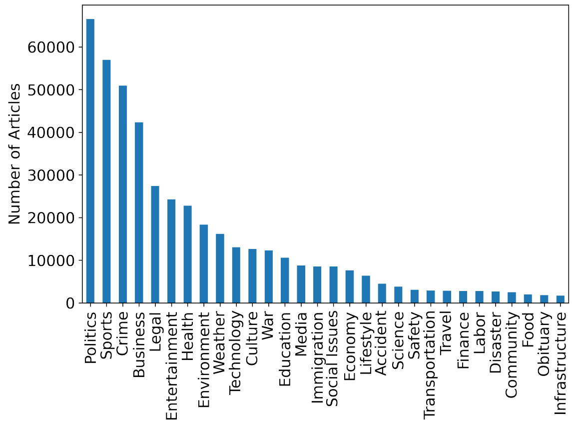


Figure 2: Distribution of data across the top 30 topics

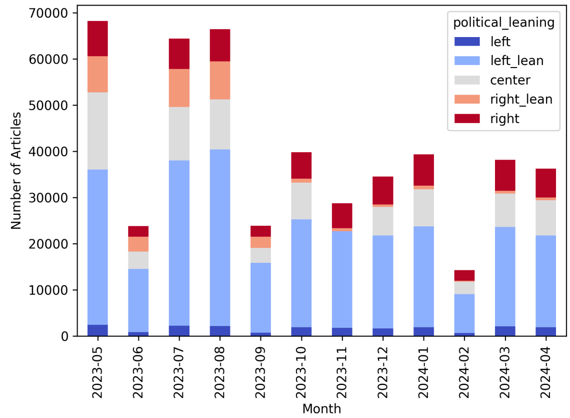


Figure 3: Distribution of data across the time-period of collection, broken down by political leaning.

[2] https://www.allsides.com/media-bias/ratings
[3] https://commoncrawl.org/news-crawl
[4] Full dataset available upon request for non-commercial research purposes

## 4 Method

### 4.1 Model Annotation

We use Large Language Models (LLM) and Vision-Language Models (VLM) to label several aspects of the text and image from news articles, including generic, issue-specific framing and the topic of the article. In Appendix B, we list the different aspects extracted per modality from the news articles. We also generate explanations for the decisions along each prediction for reducing hallucination and facilitating downstream qualitative analysis. For extracting generic frame across both modalities, we use the framing dimensions outlined by Boydstun et al. (2014), as provided in the Media Frames Corpus (Card et al., 2015), which includes 15 generic frames appropriate for analysis of US news. These frames are *Economic, Capacity & Resources, Morality, Fairness & Equality, Legality, Constitutionality & Jurisprudence, Policy Prescription & Evaluation, Crime & Punishment, Security & Defense, Health & Safety, Quality of Life, Cultural Identity, Public Opinion, Political, External Regulation & Reputation* and *Other*. The descriptions of each of the frames are provided in Appendix D. Differing from most previous studies on automated framing analysis, which assign a single frame to an article, we conduct the frame extraction in a multi-label setting, i.e., an article or an image can have more than one frame. This setting is much closer to the setup scholars use when conducting qualitative studies on framing (Dan, 2017).

### 4.2 Models and Prompts

We use Mistral-7B for text annotations and Pixtral-12B (multi-modal) for image (See Appendix E for more details). We carefully craft prompts for the extraction of each aspect from the article. For extraction of frames expressed in the articles and images, we experiment with including the Entman and Gamson definitions of framing. We also experiment with short, medium, and long descriptions of each frame and different output formats. We benchmark these different prompts on a validation set and qualitatively examine model predictions and explanations, iteratively improving the prompt by based on categories of errors. For instance, the Pixtral model has a tendency to predict the *economic* frame every time an entity in professional attire appears in the image, associating professional attire with being wealthy – we instruct the model to avoid such reliance. We report performance on a held-out test set in Section 4.3. For issue-specific frames, we prompt the model by providing a definition of framing and some examples of issue-specific frames across topics. The task was then to analyse the article and generate an issue-specific frame (described in 2-3 words) w.r.t. the topic of the article. More details are provided in Section 6. The full prompts for the extraction of each aspect are provided in Appendix (Listing 1 and Listing 2).

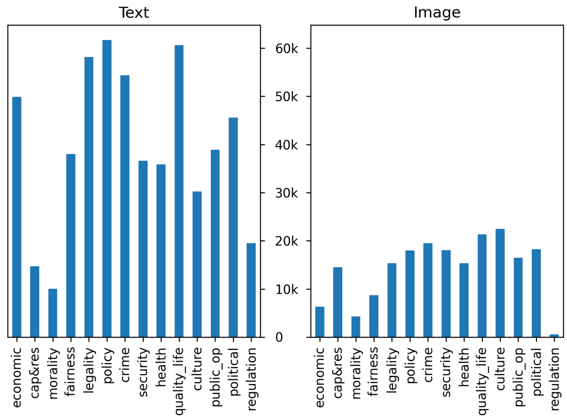


Figure 4: Frequency of predicted generic frames across all articles for texts and images.

### 4.3 Model Evaluation

**Text Framing** To assess the quality of output frames from our approach, and compare against prior work, we evaluate our text frame extraction model on the existing large scale benchmark Media Frames Corpus (Card et al., 2015). This corpus provides frame labels by each annotator for over 32,000 articles from US news on topics like immigration, smoking, and same-sex marriage. We take the union of all frame labels assigned by the annotators and assign the top 3 most frequent frames assigned by the annotators as the labels for an article. We run our text framing analysis model on this dataset, providing the article of the text while allowing the model to generate multiple frames per article. We calculate the intersection between the sets of Mistral model annotated frames and human-annotated frames for each article in the dataset. 95.7% of the articles had a non-zero intersection, with at least one overlapping frame label, demonstrating that the model outputs accurate frames in most cases. Our model received a micro averaged F1 score of 0.5, with an averaged precision of 0.42 and averaged recall of 0.62. For comparison, we also computed a random baseline (avg. over 10 runs with different random seeds) for the multi-label classification task on the full MFC dataset. When allowing a random number of predictions per sample, the random baseline achieved an F1-macro of 0.080 and an F1-micro of 0.092. We also compute this for mean number of predictions for the gold set (3), and get an F1-macro of 0.159 and an F1-micro of 0.182. For prediction across 15 labels in a multi-label setting and a task as subjective as framing, we believe the model performs quite well, outperforming baseline by a substantial margin. On qualitative inspection of a sample of erroneous predictions, several of them can be attributed to subjective interpretation, as also exemplified by the disagreement amongst the annotators in the gold labels (Krippendorff's $\alpha = [0.25-0.6]$). We provide an error analysis of the model with per-label metrics, frequent misclassifications and their examples in Appendix G.

**Image Framing** For images, there is no existing benchmark with frame labels. Two of the authors (graduate students between the ages of 28-35) of this study manually annotated 600 images for generic frames across a stratified (time-period and topics) sample of the dataset. We use annotation guidelines released by Mendelsohn et al. (2021), adapting them to the image annotation setting with examples of images for each framing category. We set up a multi-label classification platform (Figure 12) where annotators select one or more frames from the list given an image. To calculate agreement, we compute both Krippendorff's alpha ($\alpha = 0.393$) and mean Jaccard Index (0.614) as measures of inter-annotator agreement. While Krippendorff's alpha provides label-level reliability, Jaccard Index is particularly well-suited to our multi-label setting, as it evaluates instance-level agreement and gives partial credit when annotators agree on some but not all labels. Framing analysis is a highly subjective task, as is well established by prior work in communication studies as well as NLP (Card et al., 2015). For images, the subjectivity increases substantially, given the limited amount of context available and requiring cultural and conceptual familiarity more so than needed for text (Geise and Baden, 2015). Our agreement scores are thus in line with prior work on framing.

To minimise the effect of subjectivity of the annotations on our findings, we take the union of the frame annotations by two annotators as our gold set. The model is thus tasked with generating *all* labels annotated by the different annotators. We calculate the intersection of model predictions and the gold set for each image. The proportion of non-zero intersection instances between the model and human annotations is 84.2%, i.e., at least one correctly predicted frame most of the time, demonstrating the model aligning with the human framing interpretations in a majority of the cases, with the most frequent error being a 'None' (cases where the model output format was invalid or the predicted frames fell outside the closed set) prediction. We conduct a thorough analysis of frequent misclassifications by the model in Appendix G.

**Topics** To evaluate the topics output by the model, we extract the 20 most frequent topics for our analysis, but then discard one (*Media*) as most of its included documents were publication boilerplate, leaving us with 19 topics to analyse. Two of the authors hand-annotate a set of 190 articles (10 per topic), marking whether the model's topic prediction was acceptable. The overlap between the annotators was 83.5%, substantially higher than random chance (50%). When calculated against the labels assigned by each annotator, the overall accuracy for topics predicted by the model was found to be 86% and 87%. We list these topics, with examples and accuracies, in Table 12.

## 5 Multi-Modal Framing Analysis

### 5.1 Generic Framing Analysis

We first analyse the generic frames predictions across the dataset[5] in Figure 4. Overall, we can see that there are many more predicted frames for text compared to the images (a mean of 3.6 predicted frames per text and 1.3 predicted frames per image). This is intuitive (and can be seen in the Figure 1) since the text of the article can more easily express several distinct frames. Looking at their relative distributions, we can see that in the text of the article, *quality of life*, *legality*, and *policy* appear more frequently, while *morality* and *capacity & resources* are more rare. For images, *external regulation & reputation* has very low frequency, while *culture* and *quality of life* are the most frequent. Contrasting the two modalities, *economic*

[5]For framing analysis, we perform additional filtering, excluding articles with 'None' frame predictions. We further removed articles less than 100 words and articles on the topics 'sports' and 'media', as we observed while conducting qualitative analysis that articles on sports only used the *Cultural Identity* frame in the images and several articles with the media topic only had paywall text. This resulted in a reduced set of 154k articles


Figure 5: Examples of generic frame prediction in images vs texts about immigration across political leanings.

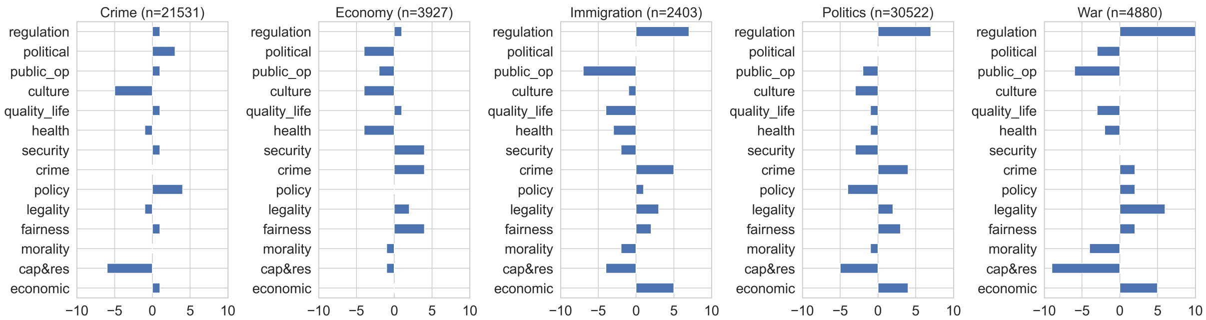


Figure 6: Comparison of generic frame prediction frequencies in images vs texts. The x-axis represents the subtracted rank of predictions for that frame between the two modalities: **positive** scores indicate that the frame was more often predicted in **texts**, **negative** scores indicate that the frame was more often predicted in **images**.

and *policy* frames much more frequently appear in text, while *capacity & resources* and *culture* are more prevalent in images.

We also compare frames by modality across topics in Figure 6. We observe substantial differences in how topics are framed across images and texts. When covering war, news outlets focus more on the *external regulation & reputation* and *crime* framing in the text, but the images that are used convey *public opinion* and *capacity & resources* framing, depicting people giving speeches and of military equipment. For articles on the topic of economy, there is more focus on *fairness* and *security & defense* spending in texts, while the images depict the *health and safety*, *culture*, and *political* frames.

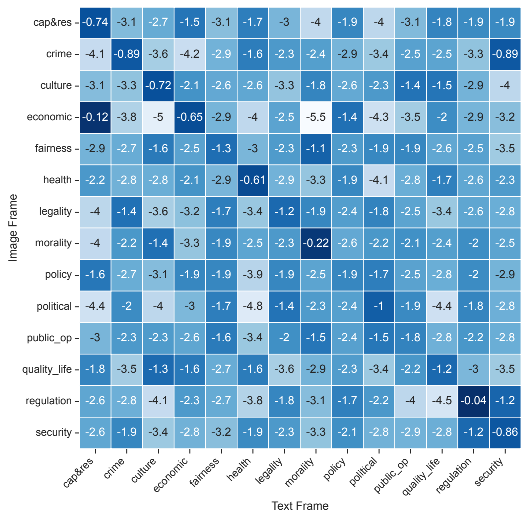


Figure 7: Pointwise mutual information between text and image frames across the dataset. Some frames are used consistently in texts and images for the same article (dark cells in the diagonal), other frames differ widely.

**Frame Co-Occurrence** To understand what is highlighted in the text of the article compared to the image, we plot the pointwise mutual information (PMI) of image frames and text frames across the entire dataset in Figure 7. We can see the presence of a diagonal, demonstrating that there is often alignment between frames across modalities. However, there are many deviations, some of which are intuitive. For example, depictions of *quality of life* in the images when writing about cultural topics and vice versa is quite common. *Quality of life* framing in the images is also used when the text is using an *economic* frame. *Political* framing in the text is associated with *policy* and *public opinion* framing, and *legal* framing in text is associated with *political* framing in the images. To get a finer-grained understanding of co-occurrence of frames across the modalities, we analyse the percentage

| crime | | quality of life | |
|---|---|---|---|
| $z$-score | bigram | $z$-score | bigram |
| 14.7 | year old | 5.5 | disney world |
| 13.6 | police said | 5.0 | mother day |
| 12.3 | police department | 4.9 | ice cream |
| 7.8 | county sheriff | 4.8 | morning brew |
| 7.5 | police officers | 4.5 | prime day |
| 7.3 | sheriff office | 4.5 | walt disney |
| 7.3 | old man | 4.0 | memorial day |
| 7.2 | police chief | 4.0 | privacy policy |
| 6.6 | police say | 3.9 | black friday |
| 6.5 | law enforcement | 3.6 | day deals |
| −4.4 | justice department | −3.1 | getty images |
| −4.4 | biden administration | −3.1 | former president |
| −4.6 | united states | −3.1 | tropical storm |
| −5.0 | hunter biden | −3.2 | police said |
| −5.3 | joe biden | −3.7 | taylor swift |
| −5.4 | president donald | −4.0 | health care |
| −5.9 | supreme court | −4.3 | interest rates |
| −6.0 | white house | −4.4 | interest rate |
| −7.5 | donald trump | −4.4 | social security |
| −7.8 | former president | −4.7 | student loan |

Table 1: The bigrams most associated with the **images** (higher $z$-scores) and **texts** (lower $z$-scores) for the *crime* and *quality of life* frames.

| Issue Frame | Top Topics |
|---|---|
| Humanitarian Crisis (1723) | War (437), Politics (284), Immigration (272) |
| Political Crisis (1470) | Politics (1071), Legal (294), Immigration (49) |
| Public Health Crisis (1176) | Health (797), Environment (155), Crime (84) |
| Political Persecution (1134) | Politics (606), Legal (519), Crime (2) |
| Political Corruption (1064) | Politics (843), Legal (207), Crime (12) |
| Public Safety Concern (1053) | Crime (732), Legal (42), Safety (37) |
| Political Manipulation (931) | Politics (667), Legal (179), Immigration (44) |
| Natural Disaster (907) | Environment (328), Weather (253), Disaster (105) |
| National Security Threat (824) | Politics (375), War (180), Immigration (63) |
| Political Power Struggle (736) | Politics (649), Legal (60), War (10) |
| Cultural Celebration (627) | Culture (285), Entertainment (178), no_topic (58) |
| Political Scandal (622) | Politics (446), Legal (166), Crime (3) |
| Security Threat (604) | War (244), Politics (139), Immigration (80) |
| Natural Disaster Threat (589) | Environment (297), Weather (112), Nat. Disasters (79) |
| Economic Struggle (587) | Economy (364), Business (124), Politics (38) |
| Economic Burden (587) | Economy (152), Business (128), Immigration (69) |
| Financial Opportunity (578) | Finance (268), Business (229), Economy (48) |
| Tragedy (566) | Crime (217), Accident (108), Transportation (46) |
| Legal Battle (558) | Legal (460), Politics (69), Crime (11) |
| Criminal Threat (557) | Crime (512), Immigration (26), Legal (9) |

Table 2: Most frequent issue-frames across the dataset along with the top 3 article topics they are encountered in. Counts are provided in parenthesis.

of co-occurrences per topic, as shown in Figure 8. For articles about crime, when using the *criminal* framing in images, the texts also tend to highlight the *security*, *quality of life*, and the *legal* frames. For war, the images consistently highlight the *security & defense* framing, even when the article text is highlighting the *policy* or *legality* frames.

**Lexical Comparison** Our results above indicate that frames are used in images and texts in different ways, but *how* those uses differ is unclear. To explore this question, we perform a lexical analysis of the words used in articles whose image or text used the same frame. We use the "Fightin' Words" algorithm from Monroe et al. (2008), a comparison metric which takes into account disproportionate numbers of samples as well as rare words, to find the bigrams from the article texts most associated with a single frame for images vs texts. Table 1 shows the bigrams sorted by their $z$-scores (prior=0.01, frequency$\geq$ 5) for two frames. Qualitatively, we observe that words associated more with image frames tend to be concrete ("ice cream") and associated with a single meaning more easily recognized to the predicted frame ("police department"). Prior work suggests that more tightly clustered and recognizable images are associated with more concrete topics (Hessel et al., 2018). See the Appendix for a full list of the frames.

**Political Leaning** We also analysed the correlation between political leaning and text/image frames across topics. For each topic, we compute how often each frame appears for each political leaning. Figure 9 shows the proportional frequency of frames per leaning, allowing us to compare which frames dominate among different leanings for a specific topic. We observe that for articles on Crime and War, the images mostly use the *crime* and *security* framing. In text, we see additional frames like *legality*, *policy*, *regulation*. Topics like immigration and the economy have more variation. For Immigration, right-leaning outlets have more relative focus on *crime* and *security* framing in text as compared to *fairness* for left-leaning outlets. For articles on the Economy, the portrayal in the images is on *quality of life* for the left whereas on *politics* for the right.

## 6 Issue Frame Analysis

As outlined in Section 4.1, we extract open-ended issue specific frames by prompting the model by giving the model example frames and instructing it to generate appropriate issue frame based on article's topic. The generated outputs are limited to a few words, but are open-ended as the model is not instructed to choose from a given set of frames but rather generate them inductively. Issue frames are tailored towards fine-grained analysis of specific issues, so repetition of same labels is relatively sparse across topics, producing over 56k unique issue frames across the entire dataset. We show the top 20 predicted issue frames by frequency, along with the top 3 broad-level topics across which they appear in Table 2. To demonstrate the quality of

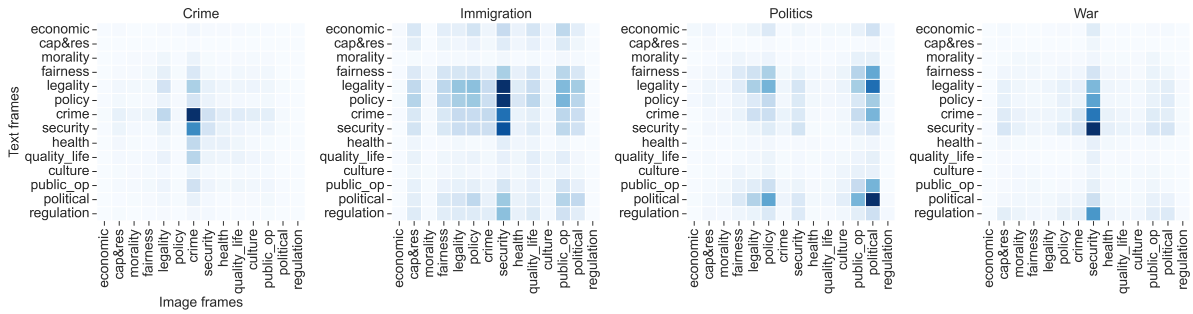


Figure 8: Frequency of frame co-occurrence between text and image frames across four selected topics.

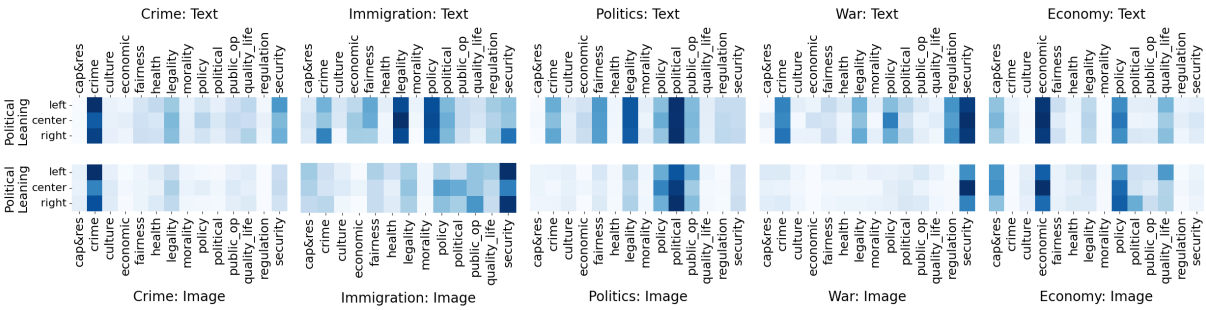


Figure 9: Comparison of text and image frame distributions across political leanings for five topics.

our issue frames, we looked at the 5 most frequent issue frames per topic as show in Table 11. We tested their coherence with intruder-test (Chang et al., 2009), a test typically used for evaluation of open-ended output like topics, with 3 annotators. In the test, for each topic, a random frame from above was replaced with an intruder. Annotators then identify the intruder, with a random baseline of 20%. The annotators identified the intruder frames 89.5%, 94.7% and 100% of the time, respectively, indicating the high coherence of our issue frames.

## 7 Case Study: Immigration

So far, we demonstrated the differences in generic framing across across modalities, topics, and political leaning across an entire corpus of articles. But our approach also allows for a more fine-grained analysis, and we demonstrate this in a case study focused on one topic: immigration. In Figure 6, we show that articles about immigration use the *crime and punishment*, *external regulation & reputation*, and *economic* framing much more frequently than the images. These articles often focus on deals with other countries, their contribution to the economy, the cost of their deportation, and/or the crimes that they commit. On the other hand, images tend to use the *capacity & resources* framing, showing migrants in camps, or the *public opinion* framing, showing people giving speeches. There are also differences across the political spectrum, as can be seen in Figure 9, with the right focusing much more on the *security & defense* framing compared to sources from the left or center across both modalities. To highlight this further, we show the finer-grained issue specific frames per political leaning in Figure 10. In our analysis, we further look at frame frequencies normalised by each political leaning and see clear differences and fine-grained signals about which specific framing publishers leaning to different sides of the political spectrum use. These observations are in line with previous studies (Mendelsohn et al., 2021; Hayes, 2008) that study how political leaning of the news outlet affects the portrayal of immigrants in news The left and center tend to highlight the *humanitarian crisis* framing much more, the gap is smaller for right-leaning publishers, while on the right, the focus is more on immigrants being an *economic burden* or a *national security threat*.

## 8 Discussion

Being exposed to selective information can significantly bias our world view. For news, it can lead to problems at an individual level such as miscommunication; at a societal level, it can lead to misinformation, political polarisation, or avoidance

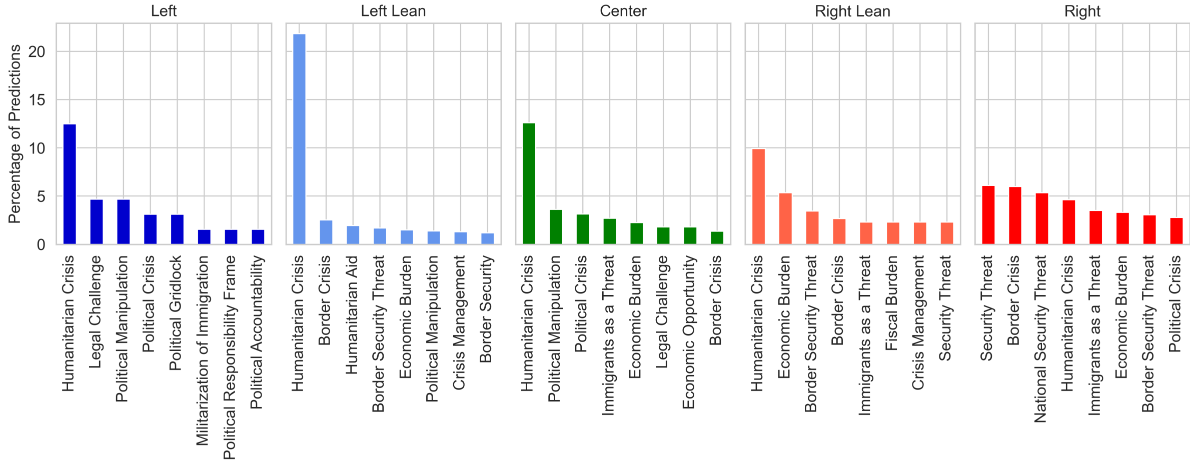


Figure 10: Normalised count of top 10 issue frames used by news publishers in the article text across the political spectrum in their reporting on immigration.

of news (Lecheler, 2018; Iyengar, 2017). This problem is aggravated considering news consumers tend to receive their news from social media, further creating "curation bubbles" (Green et al., 2025), where the dynamics of the networks on social media further contribute to curation, and in effect, exposure to selective partisan information. Thus, it is important to conduct analyses of news framing at scale to understand its effects (Lecheler, 2018). Restricting computational framing analysis to text or single labels overlooks important information, as articles often contain multiple frames, and images might convey ideological or emotional undertones that text alone may miss (Wessler et al., 2016; Geise et al., 2025). While analysing framing in images is subjective, we find it possible to reach meaningful agreement among trained annotators and detect frames with reasonable accuracy at scale.

Our analysis of multimodal framing in the news reveals that texts yield more predicted frames than images, text and image frames have distinct lexical patterns, and that frames can be used for detailed topic analysis, both at a corpus level and at a more fine-grained level by focusing on a single issue. We show how news publishers across the political spectrum use different issue-specific frames to shape narratives, potentially reinforcing echo chambers and influencing misinformation and societal polarisation. The differences we uncover between the use of text and image frames emphasize that it is crucial to take images into account when analysing framing in the news.

The cause of these different framings in images compared to the text is multi-faceted as they stem from multiple factors, including reliance on copyright-free images, algorithmic bias in image search, editorial intent, individual and institutional biases, and corporate ownership structures. Prior work classifies this as frame-building which shapes how framing appears in news (De Vreese, 2005). Our method allows for an analysis of the outcome of this process, allowing for a more thorough investigation and analysis at scale.

## 9 Conclusion

Framing analysis is an important task for content analysis at large, with implications for several fields in the social sciences, particularly for media bias analysis. In this work, we propose a methodology for leveraging open-weight large (vision-) language models to conduct integrative framing analysis of news articles at scale. While models are sensitive to prompting and over-rely on surface level visual features, we show that they can be prompted to have reasonable accuracy on human-annotated framing datasets. Our approach allows for efficient, customisable, and replicable analyses of images and text in large-scale corpora. We demonstrate its usefulness by conducting multi-faceted framing analyses that capture generic and issue-specific frames, textual and visual framing, and multiple framings per article. We find stark differences in visual & textual framing of same topics, as well as differences in framing across publications across the political spectrum. We release a large scale dataset of 500k news articles with model annotations for article text as well as images, which we hope is useful for future work on framing analysis.

## 10 Acknowledgments

This research was co-funded by a DFF Sapere Aude research leader grant under grant agreement No 0171-00034B and a DFF Research Project 1 under grant agreement No 9130-00092B, and supported by the Pioneer Centre for AI, DNRF grant number P1.

## 11 Limitations

Despite the advancements in automating multimodal framing analysis, our approach has certain limitations. Firstly, our method assumes that both textual and visual elements contribute meaningfully to framing, yet in some cases, framing might be driven predominantly by one modality. This is akin to a news reader focusing more on one modality while consuming news. The challenge of evaluating the dominance of one modality remains an open problem when text and image signals conflict or reinforce different interpretations.

Similarly, our approach also assumes equivalence across several frame predictions within an article i.e. all predicted frames have the same salience, however, in reality, one frame might be more dominant over another within the text or the image. We chose this design due to method constraints, accurately predicting multiple frames in an article or an image is already an very hard problem, detecting their relative salience is much harder. We ran some initial experiments with naively prompting for predictions in order of saliency but found the outputs to be quite noisy while doing qualitative analysis. Further, the outputs were hard to assess systematically since no such dataset with dominance scores exists for frame prediction.

Another limitation is that the interpretability of automated framing predictions remains limited. While our approach can identify patterns aligned with theoretical framing literature, it does not fully replicate the depth of qualitative human analysis. Ensuring transparency in model decisions and improving explainability remains an important avenue for future research. Finally, our study focuses on a specific set of framing methodologies and datasets, meaning generalizability to other media contexts or platforms requires further investigation. The dynamic nature of framing, influenced by evolving cultural and political landscapes, suggests that models must be continuously updated to remain effective.

## 12 Ethical Considerations

The models used in this study have been trained on large-scale datasets that may encode existing societal biases, which can impact the fairness and accuracy of framing predictions. If left uninvestigated, these biases may reinforce dominant narratives while marginalizing alternative perspectives, particularly in politically sensitive or socially divisive topics. Though having a scalable and automated approach can help aid news organizations and journalism scholars, it must be kept in mind that the generated responses are based on probabilistic patterns, hence we should allow error margins for some misclassifications. News is dynamic in nature but models can become static in nature, hence any method for news framing necessitates continuous updates of these large models.

## References

Giorgia Aiello and Katy Parry. 2019. *Visual communication: Understanding images in media culture*. SAGE Publications Ltd.

Yamen Ajjour, Milad Alshomary, Henning Wachsmuth, and Benno Stein. 2019. Modeling Frames in Argumentation. 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 2922–2932, Hong Kong, China. 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*, pages 9335–9348, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Frank R Baumgartner, Suzanna L De Boef, and Amber E Boydstun. 2008. *The decline of the death penalty and the discovery of innocence*. Cambridge University Press.

Amber E. Boydstun, Dallas Card, Justin Gross, Paul Resnick, and Noah A. Smith. 2014. Tracking the Development of Media Frames within and across Policy Issues. *American Political Science Association*.

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

Dallas Card, Amber Boydstun, Justin H Gross, Philip Resnik, and Noah A Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444.

Jonathan Chang, Sean Gerrish, Chong Wang, Jordan Boyd-Graber, and David Blei. 2009. Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems*, 22.

Renita Coleman. 2010. Framing the pictures in our heads: Exploring the framing and agenda-setting effects of visual images. In *Doing news framing analysis*, pages 249–278. Routledge.

Jay Cope, Andeelynn Fifrick, Douglas Holl, Marlon Martin, David Nunnally, Donald Preston, Paul Roszkowski, Amy Schiess, and Allison Tedesco. 2005. Image impact in print media: A study of how pictures influence news consumers. *Impact of News Images*, pages 1–40.

Viorela Dan. 2017. *Integrative Framing Analysis: Framing Health through Words and Visuals*. Number 4 in Routledge Research in Communication Studies. Routledge, New York.

Claes H De Vreese. 2005. News framing: Theory and typology. *Information design journal+ document design*, 13(1):51–62.

Lyombe Eko. 1999. Framing and priming effects. *Clarifying communication theories: A hands-on approach*, pages 276–288.

Robert M Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4):51–58.

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.

Stephanie Geise and Christian Baden. 2015. Putting the image back into the frame: Modeling the linkage between visual communication and frame-processing theory. *Communication Theory*, 25(1):46–69.

Stephanie Geise, Diana Panke, and Axel Heck. 2025. From news images to action: the mobilizing effect of emotional protest images in news coverage. *Frontiers in Political Science*, 6.

Stephanie Geise and Yi Xu. 2024. Effects of visual framing in multimodal media environments: A systematic review of studies between 1979 and 2023. *Journalism & Mass Communication Quarterly*.

Todd Gitlin. 1980. *The whole world is watching*. University of California Press.

Erving Goffman. 1974. *Frame analysis: An essay on the organization of experience.* Harvard University Press.

Maria Elizabeth Grabe and Erik Page Bucy. 2009. *Image bite politics: News and the visual framing of elections*. Oxford University Press.

Jon Green, Stefab Mccabe, Sarah Shugars, Hanyu Chwe, Luke Horgan, Shuyang Cao, and David Lazer. 2025. Curation bubbles. *American Political Science Review*, page 1–19.

Felix Hamborg, Norman Meuschke, Corinna Breitinger, and Bela Gipp. 2017. news-please: A generic news crawler and extractor. In *Proceedings of the 15th International Symposium of Information Science*, pages 218–223.

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.

Danny Hayes. 2008. Media frames and the immigration debate. In *Annual Meeting of the Midwest Political Science Association, Chicago*.

Jack Hessel, David Mimno, and Lillian Lee. 2018. Quantifying the visual concreteness of words and topics in multimodal datasets. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2194–2205, New Orleans, Louisiana. Association for Computational Linguistics.

Shanto Iyengar. 2017. A typology of media effects. *The Oxford handbook of political communication*, pages 59–68.

Kate Keib, Camila Espina, Yen-I Lee, Bartosz W Wojdynski, Dongwon Choi, and Hyejin Bang. 2018. Picture this: The influence of emotionally valenced images, on attention, selection, and sharing of social media news. *Media Psychology*, 21(2):202–221.

Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*.

Sophie Lecheler. 2018. *News Framing Effects*. Routledge.

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.

Paul Messaris and Linus Abraham. 2001. The role of images in framing news stories. In *Framing public life*, pages 231–242. Routledge.

Susan H Miller. 1975. The content of news photos: Women's and men's roles. *Journalism Quarterly*, 52(1):70–75.

Burt L Monroe, Michael P Colaresi, and Kevin M Quinn. 2008. Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16(4):372–403.

Viet An Nguyen. 2015. *Guided Probabilistic Topic Models for Agenda-setting and Framing*. Ph.D. thesis, University of Maryland, College Park.

Yulia Otmakhova, Shima Khanehzar, and Lea Frermann. 2024. Media framing: A typology and survey of computational approaches across disciplines. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15407–15428, Bangkok, Thailand. Association for Computational Linguistics.

Zhongdang Pan and Gerald M Kosicki. 1993. Framing analysis: An approach to news discourse. *Political communication*, 10(1):55–75.

Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. 2023a. SemEval-2023 Task 3: Detecting the Category, the Framing, and the Persuasion Techniques in Online News in a Multi-lingual Setup. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2343–2361, Toronto, Canada. Association for Computational Linguistics.

Jakub Piskorski, Nicolas Stefanovitch, Nikolaos Nikolaidis, Giovanni Da San Martino, and Preslav Nakov. 2023b. Multilingual Multifaceted Understanding of Online News in Terms of Genre, Framing, and Persuasion Techniques. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3001–3022, Toronto, Canada. Association for Computational Linguistics.

Thomas Edward Powell et al. 2017. *Multimodal news framing effects*. Ph.D. thesis, Amsterdam School of Communication Research (ASCoR), University of Amsterdam.

Stephen D Reese, Jr Gandy, and August E Grant. 2001. Prologue—framing public life: A bridging model for media research. In *Framing public life*, pages 23–48. Routledge.

Lulu Rodriguez and Daniela V Dimitrova. 2011. The levels of visual framing. *Journal of visual literacy*, 30(1):48–65.

Gillian Rose. 2022. *Visual methodologies: An introduction to researching with visual materials*. Sage publications.

Ahmed Sajwani, Alaa El Setohy, Ali Mekky, Diana Turmakhan, Lara Hassan, Mohamed El Zeftawy, Omar El Herraoui, Osama Afzal, Qisheng Liao, and Tarek Mahmoud. 2024. FRAPPE: FRAming, Persuasion, and Propaganda Explorer. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, pages 207–213, St. Julians, Malta. Association for Computational Linguistics.

M. Schudson. 2003. *The Sociology of News*. Contemporary societies. Norton.

Gisela Vallejo, Timothy Baldwin, and Lea Frermann. 2024. Connecting the dots in news analysis: Bridging the cross-disciplinary disparities in media bias and framing. In *Proceedings of the Sixth Workshop on Natural Language Processing and Computational Social Science (NLP+CSS 2024)*, pages 16–31, Mexico City, Mexico. Association for Computational Linguistics.

Hartmut Wessler, Antal Wozniak, Lutz Hofer, and Julia Lück. 2016. Global multimodal news frames on climate change: A comparison of five democracies around the world. *The International Journal of Press/Politics*, 21(4):423–445.

## A Source selection

We provide the list of all sources along with their corresponding political leanings in Table 3.

## B Extracted Aspects

In Table 4, we provide a list of extracted aspects per modality. For the main subject of the articles, the models were instructed to extract entities central to the text or the image, if there is one. We extract captions from the images and release it as part of the dataset but do not use it for our subsequent framing analysis.

## C Data Filtering

The scraped data carried a lot of noise. There were single line or extremely long articles, images that only depicted logos of news websites, which are not useful for analysis. To remove these, we filtered the articles whose lengths below the 5th percentile and above the 95th percentile, removing the outliers. We also removed articles that were not in English, an information available via news-please library. For images, we similarly removed the files for which the size was above the 95th percentile. There were several images in the dataset that were

| Leaning | News Domain |
|---|---|
| Left | alternet.org, editor.cnn.com, democracynow.org, dailybeast.com, huffpost.com, theintercept.com, jacobin.com, motherjones.com, newyorker.com, slate.com, msnbc.com, vox.com |
| Left Leaning | abcnews.com, apnews.com, theatlantic.com, bloomberg.com, cbsnews.com, insider.com, nbcnews.com, thenytimes.com, npr.com, politico.com, propublica.org, time.com, washingtonpost.com, yahoonews.com, usatoday.com, theguardian.com |
| Center | axios.com, forbes.com, newsweek.com, reuters.com, realclearpolitics.com, thehill.com |
| Right Leaning | thedispatch.com, theepochtimes.com, foxbusiness.com, ijr.com, nypost.com, thepostmillennial.com, washingtonexaminer.com, washingtontimes.com |
| Right | theamericanconservative.com, theamericanspectator.com, breitbart.com, dailycaller.com, dailywire.com, foxnews.com, newsmax.com, oann.com, thefederalist.com |

Table 3: List of news sources used for our dataset split by their political leaning

only logos of news organisations due to the scraper picking the wrong image from the webpage. These only had the logo and no other content, making them irrelevant for our framing analysis. There was no simple way to remove these images, we approximated their identification based on the image size. We distributed the images per their size into bins and plotted 50 randomly sampled images from them. After qualitatively assessing images from each bin, we concluded that 90% of the visualised images below the 5000 bytes threshold mostly constituted of logos only, while it was much lower for higher threshold. Therefore, we remove all figures with a size of less than 5000 bytes.

## D Frame Descriptions

In Table 7, we provide the names and corresponding descriptions for the generic frames used in our work. There are minor differences in the frame names provided here and the ones used in our dataset. This was due to us relying on the version released by Card et al. (2015) with the Media Frames Corpus dataset, which builds on the same dimensions.

| Modality | Aspect | Description |
|---|---|---|
| Text | Topic | Broad topic of the article |
| | Main subject | Subject of the text |
| | Generic Frame | One or more of the 15 generic media frames |
| | Issue Frame | Inductive, open-ended framing of a topic |
| Image | Caption | Caption of the image |
| | Main Subject | Subject in the image |
| | Generic Frame | One or more of the 15 generic media frames |

Table 4: List of aspects extracted per modality.

## E Experimental Details

**Mistral**: We used Mistral-7B-Instruct-v0.3 available via Hugging Face[6] (available under Apache-2.0 ). We set the parameters as: temperature=0.2, max_tokens=4000, dtype='half' and max_model_len=8096.

**Pixtral**: We used Pixtral-12B-2409 available via Hugging Face[7] (available under Apache-2.0 ). We set the parameters as: temperature=0.2, max_tokens=1024, dtype='half' and max_model_len=7000. Before processing the images via Pixtral, we also resized the images to 512 x 512 for computational efficiency.

We use the vLLM library (Kwon et al., 2023), available under Apache-2.0, for high-throughput inference, allowing us to conduct analysis on the entire dataset in a few days, demonstrating the scalability of our approach. For vision annotations, we ran Pixtral on 8 Nvidia A100s which enabled us to finish the computation in 5 days. For text annotations, we ran Mistral on a mix of Nvidia TitanRTX and A100s and finished the computation in similar time as vision annotations.

## F Article Subject Portrayal

We additionally explore how individual entities in the articles are portrayed. For this analysis, we take the subset where the article and the image are portraying the same main entity. The entities and sentiment are extracted by prompting the LLM and VLM to identify the main subject in the article or the image, as shown in the prompt in List-

[6] https://huggingface.co/mistralai/Mistral-7B-v0.3
[7] https://huggingface.co/mistralai/Pixtral-12B-2409

| Label | Precision | Recall | F1-score |
|---|---|---|---|
| cap&res | 0.39 | 0.34 | 0.36 |
| crime | 0.50 | 0.87 | 0.63 |
| culture | 0.38 | 0.37 | 0.37 |
| economic | 0.43 | 0.69 | 0.53 |
| fairness | 0.17 | 0.74 | 0.28 |
| health | 0.48 | 0.48 | 0.48 |
| legality | 0.53 | 0.87 | 0.66 |
| morality | 0.30 | 0.63 | 0.41 |
| policy | 0.40 | 0.73 | 0.51 |
| political | 0.68 | 0.53 | 0.60 |
| public_op | 0.32 | 0.55 | 0.40 |
| quality_life | 0.28 | 0.36 | 0.31 |
| regulation | 0.26 | 0.48 | 0.34 |
| security | 0.30 | 0.45 | 0.36 |
| micro avg | 0.42 | 0.62 | 0.50 |
| macro avg | 0.39 | 0.58 | 0.45 |
| weighted avg | 0.45 | 0.62 | 0.51 |
| samples avg | 0.46 | 0.63 | 0.51 |

Table 5: Metrics per frame label (multi-label) for our text frame classifier on the MFC dataset

ing 1. Figure 14 shows peculiar differences when contrasting the image portrayal from the portrayal in the articles. A general pattern that can be observed is that entities were portrayed more positively in the images, compared to the text, e.g., Rishi Sunak, Prince Harry, Elon Musk, Joe Biden, Benjamin Netanyahu. For some, there is a negative portrayal overall that is much more explicit in the text, e.g., Donald Trump, Hunter Biden, Vladimir Putin, Rudy Giuliani.

## G Frame Classifier Error Analysis

**Text** We provide the per label performance of the framing classifier for text on the MFC dataset in Table 5. The performance varies substantially across the different frames. We can see that the model has the least performance on detecting *fairness*, *quality of life* framing. We additionally show examples of misclassifications on the dataset in Table 10. Here, while the intersection of the prediction and label sets is null, the explanation of the model for identifying those framings is reasonable and faithful to the article text. To understand patterns in these misclassifications, we analyse labels that the model often gets confused with. Since we're operating in a multi-label setting, a traditional confusion matrix cannot be applied. A multi-label confusion matrix also treats the labels as binary and not give insight into confusion patterns. To provide insight into these specific confusions, we calculate a mismatch frequency matrix where for each missed ground truth label in an instance (rows in the table below), we mark the erroneous predictions as 1 (columns). We then sum these for the entire dataset, giving us an overview of which labels are often mis-predicted for each missed gold label. For the text frames, the mismatch frequency against MFC is shown in Table 8. The most frequent mismatch is with the *political* frame, with the model predicting *legality, policy* or *crime* framing instead. *Culture* is another frame often mislabeled, albeit with much lower frequencies.

**Image** Similarly, for the image error analysis, we calculate the mismatch frequency matrix as shown in Table 9. As can be seen, most mispredictions involve the *None* label, with the model over or under-predicting it. Other labels often confused are *public opinion* and *policy prescription*, the latter being particularly hard to detect in images even during human annotations.

We examine the 15% of examples where the model's predicted frames had no overlap with the human-labeled (gold-standard) frames. Among these zero-overlap cases, the most common human-assigned label was 'none', which accounted for 40.9% of the instances. This indicates that in many such cases, the model over-predicted by assigning frames even when annotators found no meaningful framing. The most frequently missed substantive frame was 'public opinion' (11.8%), followed by 'cultural identity' (10.8%). Other frames commonly missed included 'quality of life', 'political', 'capacity and resources', and 'health and safety', each appearing in approximately 7.5% of the zero-overlap cases. Less frequently missed frames were 'external regulation and reputation' (3.2%), and 'economic', 'morality', and 'crime and punishment', each at 1.1%.

We also note that some frames were relatively infrequent in the human-labelled data overall. For instance, 'fairness and equality' appeared in only 4.1% of annotated examples, followed by 'morality' (2.31%), 'external regulation and reputation' (2.31%), 'economic' (1.92%), 'legality, constitutionality and jurisprudence' (1.92%), and 'policy prescription and evaluation' (1.54%). As a result, their low miss rates may reflect either their lower frequency in the dataset or higher accuracy in the

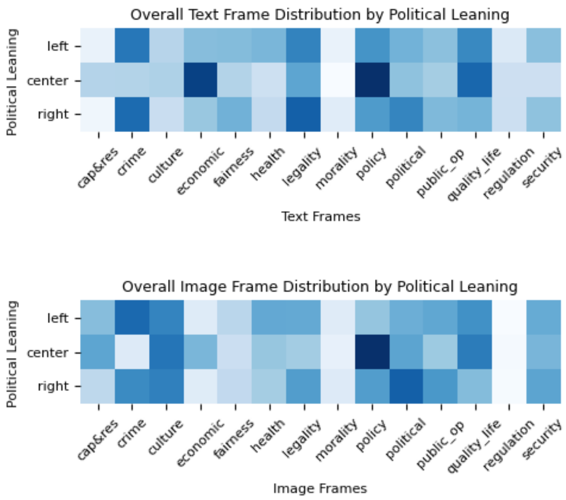


Figure 11: Overall comparison of text and image frame distributions across political leanings

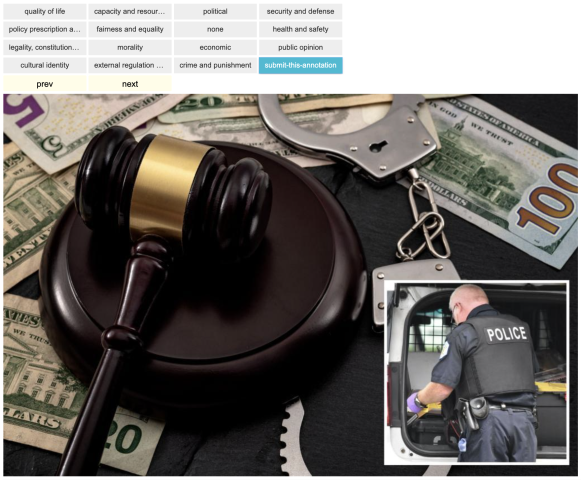


Figure 12: Image annotation UI used for annotating images frames by annotators

model's predictions.

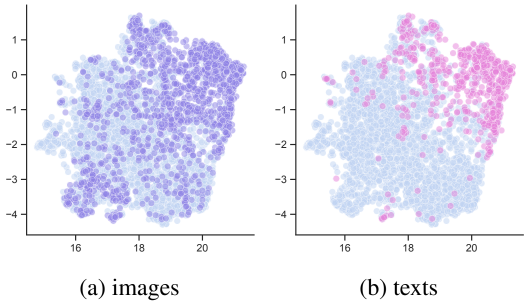


Figure 13: Image and text frames cover different parts of the topic space. UMAP reduction (n_neighbors=200, cosine distance) of a 5k sample of the generated topic descriptions (TF-IDF weighted token vectors: max_features=5000, min_df=5, max_df=0.95) of the articles. Highlighted points represent the *political* frame.

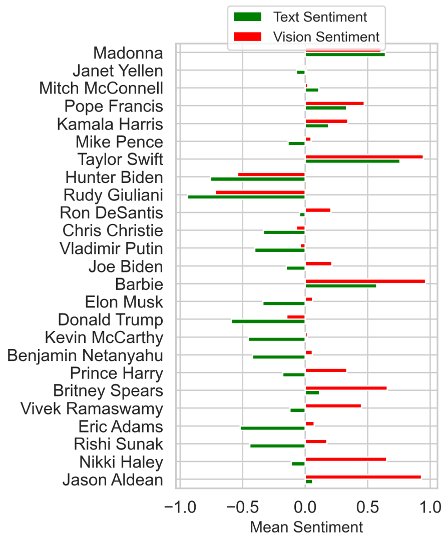


Figure 14: Difference in sentiment of the main subject's portrayal across text and images

| | legality | | morality | | security | | regulation | |
|---|---|---|---|---|---|---|---|---|
| | $z$-score | bigram | $z$-score | bigram | $z$-score | bigram | $z$-score | bigram |
| more associated with **images** | 7.6 | supreme court | 4.7 | pope francis | 5.7 | air force | 6.6 | secretary general |
| | 6.3 | bankman fried | 4.1 | san francisco | 3.2 | follow twitter | 5.3 | united arab |
| | 5.6 | hunter biden | 3.7 | supreme court | 3.1 | border patrol | 5.3 | arab emirates |
| | 5.1 | years prison | 3.3 | new jersey | 3.1 | fire department | 4.9 | biden said |
| | 4.8 | attorney office | 3.1 | two years | 2.8 | united airlines | 4.9 | prime minister |
| | 4.4 | clarence thomas | 3.1 | law enforcement | 2.6 | safety board | 4.8 | join nato |
| | 4.2 | justice department | 3.0 | family members | 2.5 | climate change | 4.3 | saudi arabia |
| | 4.1 | district attorney | 2.9 | catholic church | 2.5 | federal government | 4.2 | official said |
| | 4.1 | rights act | 2.8 | high school | 2.4 | russian forces | 4.1 | general jens |
| | 3.9 | high school | 1.9 | took place | 2.4 | natural gas | 4.1 | jens stoltenberg |
| more associated with **texts** | −2.8 | north korea | −2.1 | president donald | −2.8 | debt limit | −0.8 | donald trump |
| | −2.8 | trump said | −2.1 | president biden | −2.9 | attempted murder | −0.8 | officials said |
| | −2.9 | loan forgiveness | −2.1 | vice president | −3.3 | anyone information | −0.9 | president biden |
| | −2.9 | interest rates | −2.1 | social media | −3.3 | york city | −1.1 | last week |
| | −3.0 | gov ron | −2.1 | joe biden | −3.4 | old man | −1.1 | news app |
| | −3.2 | nbc news | −2.1 | request comment | −3.5 | san francisco | −1.1 | click get |
| | −3.3 | ron desantis | −2.5 | white house | −3.7 | hong kong | −1.2 | get fox |
| | −3.4 | prime minister | −3.1 | donald trump | −5.1 | police department | −1.2 | united states |
| | −3.6 | new hampshire | −3.2 | fox news | −5.3 | year old | −1.6 | new york |
| | −3.6 | national security | −3.2 | former president | −7.5 | police said | −2.1 | fox news |

| | culture | | fairness | | health | | public opinion | |
|---|---|---|---|---|---|---|---|---|
| | $z$-score | bigram | $z$-score | bigram | $z$-score | bigram | $z$-score | bigram |
| more associated with **images** | 8.1 | taylor swift | 5.8 | los angeles | 5.2 | long term | 6.5 | supreme court |
| | 7.3 | eras tour | 5.2 | student debt | 5.1 | term care | 5.9 | sag aftra |
| | 6.1 | getty images | 5.1 | monthly payments | 4.9 | care insurance | 5.0 | artificial intelligence |
| | 5.2 | las vegas | 4.8 | san francisco | 4.3 | year year | 4.5 | biden administration |
| | 4.9 | swift eras | 4.8 | fain said | 4.2 | weight loss | 4.3 | anti israel |
| | 4.6 | box office | 4.4 | student loan | 3.9 | long covid | 4.0 | cbs news |
| | 4.2 | los angeles | 4.3 | auto workers | 3.5 | five year | 4.0 | san francisco |
| | 4.1 | fourth july | 4.3 | general motors | 3.4 | prime day | 3.8 | auto workers |
| | 3.7 | instagram post | 4.1 | writers strike | 3.0 | health insurance | 3.8 | mar lago |
| | 3.6 | last year | 4.1 | official said | 2.8 | disease control | 3.8 | big three |
| more associated with **texts** | −3.0 | ads content | −2.7 | south carolina | −2.4 | former president | −2.9 | state police |
| | −3.0 | parties information | −2.7 | nikki haley | −2.5 | county sheriff | −3.2 | police department |
| | −3.0 | charities online | −2.9 | special counsel | −2.5 | tropical storm | −3.2 | eras tour |
| | −3.0 | contain info | −3.3 | ron desantis | −2.5 | medical examiner | −3.3 | social media |
| | −3.0 | information see | −3.4 | attorney general | −2.6 | cause death | −3.5 | window share |
| | −3.0 | online ads | −3.6 | hunter biden | −2.7 | cbs essentials | −3.7 | new window |
| | −3.0 | newsletters may | −3.7 | president donald | −2.8 | sheriff office | −3.7 | opens new |
| | −3.8 | young people | −3.9 | new hampshire | −3.1 | mother day | −4.3 | new hampshire |
| | −3.8 | new hampshire | −5.4 | donald trump | −3.4 | office said | −4.5 | bud light |
| | −4.2 | supreme court | −5.4 | former president | −5.0 | year old | −5.3 | taylor swift |

| | policy | | capacity & resources | | political | | economic | |
|---|---|---|---|---|---|---|---|---|
| | $z$-score | bigram | $z$-score | bigram | $z$-score | bigram | $z$-score | bigram |
| more associated with **images** | 6.9 | white house | 5.1 | climate change | 10.2 | former president | 13.1 | credit card |
| | 6.6 | hunter biden | 5.0 | year old | 7.4 | donald trump | 12.2 | personal loan |
| | 6.3 | ron desantis | 4.9 | officials said | 6.8 | student loan | 12.2 | credit score |
| | 6.0 | house republicans | 4.2 | news app | 6.5 | white house | 12.2 | interest rates |
| | 5.8 | mortgage rates | 4.1 | click get | 6.3 | president donald | 11.5 | interest rate |
| | 5.6 | president biden | 4.1 | oil gas | 5.8 | continue reading | 10.6 | card debt |
| | 5.3 | house speaker | 3.9 | fox news | 5.4 | debt ceiling | 9.0 | social security |
| | 5.1 | kevin mccarthy | 3.7 | said statement | 5.3 | supreme court | 8.1 | lower interest |
| | 5.0 | prime minister | 3.6 | united states | 5.0 | joe biden | 8.0 | high yield |
| | 5.0 | donald trump | 3.5 | supreme court | 5.0 | primary ballot | 7.7 | yield savings |
| more associated with **texts** | −3.1 | recent years | −3.3 | card debt | −3.3 | taylor swift | −2.3 | news digital |
| | −3.1 | social security | −3.3 | wall street | −3.5 | police officer | −2.4 | click get |
| | −3.2 | told cbs | −3.4 | savings accounts | −3.6 | city council | −2.4 | climate change |
| | −3.2 | per month | −3.6 | cash flow | −3.8 | san francisco | −2.5 | news app |
| | −3.2 | police said | −3.8 | federal reserve | −4.2 | los angeles | −2.5 | former president |
| | −3.3 | los angeles | −4.5 | long term | −4.3 | police department | −2.6 | social media |
| | −3.4 | officials said | −4.5 | credit score | −4.3 | window share | −2.6 | white house |
| | −3.5 | border patrol | −4.7 | interest rate | −4.4 | social media | −2.6 | loan forgiveness |
| | −4.3 | mental health | −5.0 | credit card | −4.6 | new window | −2.8 | privacy policy |
| | −4.9 | year old | −6.0 | interest rates | −4.6 | opens new | −3.7 | fox news |

Table 6: The bigrams most associated the **images** (higher $z$-scores) and **texts** (lower $z$-scores) for all the frames except crime and quality of life (shown in Table 1).

| Frame Name | Description |
|---|---|
| Economic | The costs, benefits, or monetary/financial implications of the issue (to an individual, family, community or to the economy as a whole) |
| Capacity and resources | The lack of or availability of physical, geographical, spatial, human, and financial resources, or the capacity of existing systems and resources to implement or carry out policy goals. |
| Morality | Any perspective—or policy objective or action (including proposed action)—that is compelled by religious doctrine or interpretation, duty, honor, righteousness or any other sense of ethics or social responsibility. |
| Fairness and equality | Equality or inequality with which laws, punishment, rewards, and resources are applied or distributed among individuals or groups. Also the balance between the rights or interests of one individual or group compared to another individual or group. |
| Constitutionality and jurisprudence | The constraints imposed on or freedoms granted to individuals, government, and corporations via the Constitution, Bill of Rights and other amendments, or judicial interpretation. This deals specifically with the authority of government to regulate, and the authority of individuals/corporations to act independently of government. |
| Policy prescription and evaluation | Particular policies proposed for addressing an identified problem, and figuring out if certain policies will work, or if existing policies are effective. |
| Law and order, crime and justice | Specific policies in practice and their enforcement, incentives, and implications. Includes stories about enforcement and interpretation of laws by individuals and law enforcement, breaking laws, loopholes, fines, sentencing and punishment. Increases or reductions in crime. |
| Security and defence | Security, threats to security, and protection of one's person, family, in-group, nation, etc. Generally an action or a call to action that can be taken to protect the welfare of a person, group, nation sometimes from a not yet manifested threat. |
| Health and safety | Healthcare access and effectiveness, illness, disease, sanitation, obesity, mental health effects, prevention of or perpetuation of gun violence, infrastructure and building safety. |
| Quality of life | The effects of a policy on individuals' wealth, mobility, access to resources, happiness, social structures, ease of day-to-day routines, quality of community life etc. |
| Cultural identity | The social norms, trends, values and customs constituting culture(s), as they relate to a specific policy issue |
| Public opinion | References to general social attitudes, polling and demographic information, as well as implied or actual consequences of diverging from or getting ahead of public opinion or polls. |
| Political | Any political considerations surrounding an issue. Issue actions or efforts or stances that are political, such as partisan filibusters, lobbyist involvement, bipartisan efforts, deal-making and vote trading, appealing to one's base, mentions of political manoeuvring. Explicit statements that a policy issue is good or bad for a particular political party. |
| External regulation and reputation | The United States' external relations with another nation; the external relations of one state with another; or relations between groups. This includes trade agreements and outcomes, comparisons of policy outcomes or desired policy outcomes. |
| Other | Any frames that do not fit into the above categories. |

Table 7: Frame name and descriptions for each frame used in the dataset

```
SYS_PROMPT = f"""You are an intelligent and logical journalism scholar conducting analysis of images associated with
      news artciles."""

prompt_entity = """Your task is to identify the main subject or entity in the image that is central to the image.
      Entities are people or organizations that play a central role in the image. If there are multiple entities,
      choose the one that is central to the image. if there are no central or clearly identifiable entities, choose
      "None".
Additionally, analyse the image and output the sentiment with which the subject is portrayed in the image. The
      sentiment can be "positive", "negative" or "neutral". In case of no entity, output "None. Output only the json
      and no other text.
Format your output as a json entry as follows:
{"entity-name": "<name of the entity>", "sentiment": "<sentiment towards the entity>", "sentiment-reason": "<reasoning
      for the portrayed sentiment>"}
<image>
For the image above, provide the name of the entity, sentiment towards the entity, and reasoning for the chosen
      sentiment."""
FRAMES = f"""
Economic - costs, benefits, or other finance related. The image can includes things including but not limited to
      money, funding, taxes, bank, meetings with a logo of a financial institution. If you are using logo of a
      financial instituion to classify it as economic, make sure it is clearly visible. If it is not clearly visible,
      it should be classified as 'None'. A professional attire in itself doesn not mean economic frame.
Capacity and resources - availability of physical, human, or financial resources, and capacity of current systems. In
      the image, we can see things including but not limited to a geographical area, farmland, agriculture land,
      labour, people working in an institution, or images that convey scarcity or surplus in some way.
Morality - religious or ethical implications. In the image, we can see things including but not limited to god, death,
      priests, church, protests related to moral issues.
Fairness and equality - balance or distribution of rights, responsibilities, and resources. In the image, we can see
      things including but not limited to the fight for civil or political rights, LGBTQ,  or calls to stopping
      discrimination.
Legality, constitutionality and jurisprudence - legal rights, freedoms, and authority of individuals, corporations,
      and government. In the image, we can see things including but not limited to , prisons, laws, judges in robes,
      courtrooms, legal documents, and prison facilities. This does not include sports contexts, such as referees or
      players enforcing or breaking game rules.
Policy prescription and evaluation - discussion of specific policies aimed at addressing problems. In the image, we
      can see things including but not limited to discussions on rule, rule making bodies, people in formal settings
      such as boardrooms or legislative halls - actively debating, and reviewing policy drafts or proposals. You might
      see official charts, graphs, or official documents. People in formal attire with no other information should not
      be classified as policy prescription and evaluation.
Crime and punishment - effectiveness and implications of laws and their enforcement. In the image, we can see things
      including but not limited to criminal activities, violence, police officers making arrests, crime scenes with
      investigators, courtrooms during criminal trials, prisons with detainees. This frame specifically excludes
      contexts involving sports, such as referees, players, or rule enforcement within games, which are not related to
      societal law violations or legal punishment.
Security and defense - threats to the individual, community, or nation. In the image, we can see things including but
      not limited to military uniforms, defense personnel, border patrol, war, soldiers, military equipment like tanks
      or fighter jets, border walls, or surveillance systems monitoring wide areas.
Health and safety - health care, sanitation, public safety. Images with objects like coffee, drinks, food items or
      activities like sports which a clear and literal message that it affects health and safety positively or
      negatively should be classified as health and safety, otherwise it should be classified as 'None'. E.g. a person
      drinking coffee does not mean health and safety, but a person drinking a medicine or having cigarette does. A bus
      does not mean health and safety, but a bus with a warning sign does.  In the image, we can see things including
      but not limited to doctors, nurses, injury, disease, or events with environmental impact that may impact health
      and safety.
Quality of life - threats and opportunities for the individual's wealth, happiness, and well-being. In the image, we
      can see things that improves happiness or demonstrates quality of life in some form. It also includes things that
      demonstrate deterioration of quality of life by showing hardships of people, homelessness etc. This may also
      include happy children, food items that demonstrate good quality of life or people enjoying a nice meal.
Cultural identity - traditions, customs, or values of a social group in relation to a policy issue. In the image, we
      can see things including but not limited to concerts, cultural dance, sports, art, celebrities, artists and
      prominent people related to these topics. Examples, celebrities, traditional dress, sports with clear countriy
      specific detail e.g. jerseys/flags, culural events, cultural art etc. Otherwise, it should be classified as
      'None'.
Public opinion - attitudes and opinions of the general public, including polling and demographics. Includes generic
      protests, people (non-celebrities) engaging with large crowds, riots, and strikes and including but not limited
      to sharing petitions and encouraging people to take political action. It will also include news broadcasts, talk
      shows, and interviews with people that are related to public opinion at large.
Political - considerations related to politics and politicians, including lobbying, elections, and attempts to sway
      voters. In the image, we can see things related to politicians, elections, voting, political campaigns. Just
      formal clothing does not mean political frame. If the images does not have a political person which is
      recognizable, it should not be classified as political. A formal attire with no political information should be
      classified as 'None'.
External regulation \& reputation - international reputation or foreign policy. In the image, we can see things
      including but not limited to international organizations, global discussions/meetings, foreign policy, flags from
      multiple countries, or delegates at a cross-country forum discussing reputation and regulation. If you use a logo
      of a global organization to classify it as external regulation and reputation, make sure it is clearly visible in
      the image. If it is not clearly visible, it should be classified as 'None'.
None - no frame could be identified because of lack of information in the image. This should be selected when no other
      frame is applicable. Example, a handshake with no other information, a logo of a company with no other
      information, a landscape with no other information, a person in a photo album with no other information, a person
      speaking with no other information about the content of the speech or person's identity, a formal event with  no
      other information, a person in formal attire with no other information, a news logo with no news, a sports event
      with no additional information, simple objects like vehicle/car/pen/paper/sign-boards/objects etc with no other
      information etc."""

FRAMING_PROMPT = "A set of generic news frames with an id, name and description are: \n"
FRAMES_TASK_PROMPT = """
Given the list of frames, and the image.
<image>
Your task is to carefully analyse the image and choose the appropriate frames from the above list.
Output your answer in a json format with the format:
{"frames-list": "[<All frame names that apply from list provided above>], "reason": "<reasoning for the frames
      chosen>"}
Output only the json and no other text.
"""
```

Listing 1: The prompt template for frame prediction using the pixtral model

```
SYS_PROMPT = f"You are an intelligent and logical journalism scholar conducting analysis of news articles. Your task
      is to read the article and answer the following question about the article. Only output the json and no other
      text.\n"

TOPIC_PROMPT = "Output the topic of the article, along with a justification for the answer. The topic should be a
      single word or phrase. Format your output as a json entry with the field 'topic_justification' and 'topic'."

ENTITY_PROMPT = """Your task is to identify the main subject or entity in the article that is central to the article.
      Entities are people or organizations that play a central role. If there are multiple entities being discussed,
      choose the one that is central to the article. If there are no central or clearly identifiable entities, choose
      "None".
Additionally, analyse the image and output the sentiment with which the subject is portrayed in the image. The
      sentiment can be "positive", "negative" or "neutral". In case of no entity, output "None. Output only the json
      and no other text.
Format your output as a json entry as follows:

{"entity-name": "<name of the entity>", "sentiment": "<sentiment towards the entity>", "sentiment-reason": "<reasoning
      for the portrayed sentiment>"}

For the given article, provide the name of the entity, sentiment towards the entity, and reasoning for the chosen
      sentiment."""

FRAMES = """
A list of frame names and their descriptions used in news is:
Economic - costs, benefits, or other financial implications,
Capacity and resources - availability of physical, human, or financial resources, and capacity of current systems,
Morality - religious or ethical implications,
Fairness and equality - balance or distribution of rights, responsibilities, and resources,
Legality, constitutionality and jurispudence - rights, freedoms, and authority of individuals, corporations, and
      government,
Policy prescription and evaluation - discussion of specific policies aimed at addressing problems,
Crime and punishment - effectiveness and implications of laws and their enforcement,
Security and defense - threats to welfare of the individual, community, or nation,
Health and safety - health care, sanitation, public safety,
Quality of life - threats and opportunities for the individual's wealth, happiness, and well-being,
Cultural identity - traditions, customs, or values of a social group in relation to a policy issue,
Public Opinion - attitudes and opinions of the general public, including polling and demographics,
Political - considerations related to politics and politicians, including lobbying, elections, and attempts to sway
      voters,
External regulation and reputation - international reputation or foreign policy of the U.S,
None - none of the above or any frame not covered by the above categories."""

GENERIC_FRAMING_PROMPT = f"""Framing is a way of classifying and categorizing information that allows audiences to
      make sense of and give meaning to the world around them (Goffman, 1974).
Entman (1993) has defined framing as "making some aspects of reality more salient in a text in order to promote a
      particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the
      item described".
Frames serve as metacommunicative structures that use reasoning devices such as metaphors, lexical choices, images,
      symbols, and actors to evoke a latent message for media users (Gamson, 1995).
A set of generic news frames with an id, name and description are: {FRAMES}.
Your task is to code articles for one of the above frames and provide justification for it. Format your output as a
      json entry with the fields 'frame_justification', 'frame_id', 'frame_name'.
'frame_name' should be one of the above listed frames. Only output one frame per article."""

GENERIC_FRAMING_MULTIPLE_PROMPT = """
Given the list of news frames, and the news article.
Your task is to carefully analyse the article and choose the appropriate frames used in the article from the above
      list.
Output your answer in a json format with the format:
{"frames-list": "[<All frame names that apply from list provided above>], "reason": "<reasoning for the frames
      chosen>"}.
Only choose the frames from the provided list of frames. If none of the frames apply, output "None" as the answer.
"""

ISSUE_FRAMING_PROMPT = """
Entman (1993) has defined framing as "making some aspects of reality more salient in a text in order to promote a
      particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the
      item described".
Frames serve as metacommunicative structures that use reasoning devices such as metaphors, lexical choices, images,
      symbols, and actors to evoke a latent message for media users (Gamson, 1995).
There are several ways to cover a specific issue in the news. For instance, the issue of climate change can be framed
      as a scientific, a political, a moral, or a health issue etc. with issue-specific frames such as "Global Doom",
      "Local Tragedies", "Sustainable future".
Similarly, articles related to immigration can frame immigrants as a hero, a victim, or a threat with frames such as
      "Economic Burden", "Cultural Invasion", "Humanitarian Crisis".
Based on the topic of the article, come up with an issue-specific frame that is relevant to the topic of the article.
      Provide a justification for the frame.
Format your output as a json entry with the fields 'issue_frame_justification' and 'issue_frame'."""

POST_PROMPT = " Output only the json and no other text. Make sure to add escape characters where necessary to make it
      a valid json output."
```

Listing 2: The prompt template for text prediction using the mistral model

| | none | economic | cap&res | morality | fairness | legality | policy | crime | security | health | quality_life | culture | public_op | political | regulation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| none | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| economic | 5 | 0 | 6 | 198 | 724 | 897 | 970 | 798 | 167 | 178 | 152 | 152 | 408 | 261 | 180 |
| cap&res | 20 | 1205 | 0 | 95 | 438 | 953 | 1497 | 380 | 349 | 913 | 854 | 213 | 443 | 140 | 309 |
| morality | 2 | 207 | 55 | 0 | 340 | 318 | 392 | 282 | 73 | 103 | 164 | 163 | 235 | 150 | 69 |
| fairness | 0 | 80 | 20 | 120 | 0 | 177 | 246 | 189 | 37 | 32 | 55 | 88 | 139 | 148 | 29 |
| legality | 2 | 296 | 153 | 421 | 664 | 0 | 479 | 329 | 182 | 100 | 165 | 285 | 420 | 265 | 100 |
| policy | 12 | 282 | 126 | 567 | 1262 | 1042 | 0 | 649 | 210 | 175 | 308 | 426 | 424 | 478 | 132 |
| crime | 1 | 204 | 72 | 158 | 498 | 364 | 387 | 0 | 98 | 64 | 80 | 164 | 268 | 349 | 77 |
| security | 1 | 341 | 108 | 117 | 524 | 866 | 809 | 806 | 0 | 60 | 183 | 148 | 397 | 276 | 182 |
| health | 8 | 541 | 194 | 435 | 871 | 1065 | 1218 | 1074 | 507 | 0 | 448 | 300 | 669 | 341 | 194 |
| quality_life | 8 | 639 | 175 | 563 | 1462 | 713 | 1181 | 445 | 197 | 237 | 0 | 563 | 892 | 505 | 181 |
| culture | 15 | 797 | 241 | 576 | 1036 | 1189 | 1766 | 1179 | 390 | 438 | 634 | 0 | 897 | 563 | 319 |
| public_op | 4 | 278 | 77 | 383 | 954 | 951 | 796 | 846 | 219 | 127 | 289 | 364 | 0 | 406 | 142 |
| political | 21 | 1529 | 406 | 535 | 1944 | 3172 | 3199 | 2308 | 1143 | 824 | 1136 | 569 | 1243 | 0 | 421 |
| regulation | 6 | 401 | 62 | 52 | 220 | 399 | 529 | 205 | 184 | 168 | 221 | 108 | 179 | 79 | 0 |

Table 8: Mismatch frequency matrix for the text annotations on the MFC dataset

| | cap&res | crime | culture | economic | fairness | health | legality | morality | none | policy | political | public_op | quality_life | regulation | security |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cap&res | 0 | 3 | 0 | 0 | 0 | 3 | 1 | 1 | 14 | 2 | 0 | 4 | 1 | 0 | 6 |
| crime | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 3 | 0 | 0 | 1 | 0 | 0 | 0 |
| culture | 2 | 0 | 0 | 0 | 4 | 3 | 2 | 3 | 16 | 2 | 0 | 3 | 2 | 0 | 4 |
| economic | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 1 | 0 | 0 |
| fairness | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| health | 2 | 4 | 1 | 0 | 0 | 0 | 3 | 3 | 11 | 2 | 0 | 4 | 1 | 0 | 7 |
| legality | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| morality | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| none | 4 | 3 | 7 | 0 | 3 | 4 | 9 | 2 | 0 | 21 | 4 | 6 | 4 | 0 | 8 |
| policy | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| political | 0 | 0 | 0 | 0 | 1 | 0 | 3 | 2 | 9 | 9 | 0 | 0 | 0 | 1 | 3 |
| public_op | 1 | 0 | 1 | 0 | 0 | 0 | 6 | 0 | 7 | 11 | 1 | 0 | 0 | 1 | 0 |
| quality_life | 3 | 0 | 1 | 0 | 2 | 3 | 2 | 3 | 21 | 1 | 0 | 2 | 0 | 0 | 0 |
| regulation | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 0 | 0 | 0 | 0 | 3 |
| security | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 0 | 0 | 1 | 0 | 0 | 0 |

Table 9: Mismatch Matrix for Vision annotations across the human annotated dataset

| **Topic** | **Prediction** | **Label** | **Model reasoning** |
|---|---|---|---|
| gun_control | ‘crime’, ‘policy’, ‘security’ | ‘political’, ‘health’ | The article discusses a gun ban for people on the no-fly list, which is related to crime and punishment and policy prescription and evaluation. Additionally, the context of the discussion is in response to a security threat, making it relevant to the Security and defense frame. |
| climate_change | ‘economic’, ‘health’, ‘quality_life’, ‘policy’, ‘regulation’ | ‘cap&res’, ‘political’ | The article discusses the warming of the oceans over the last 40 years, which has implications for economic costs (e.g., potential damage to marine life and industries), health and safety (e.g., potential impact on human health and marine ecosystems), quality of life (e.g., potential effects on the overall well-being of individuals and communities), policy prescription and evaluation (e.g., the study provides evidence for or against computer models used to predict Earth‘s warming), and external regulation and reputation (e.g., international implications of climate change and the role of the U.S. in addressing it). |
| gun_control | ‘economic’, ‘crime’, ‘policy’, ‘public_op’, ‘culture’ | ‘regulation’, ‘political’, ‘quality_life’ | The article discusses the financial implications of gun violence (Economic), the legal and enforcement aspects of gun control (Crime and punishment), the proposed policy changes to address the issue (Policy prescription and evaluation), the public opinion and sentiments towards gun control (Public sentiment), and the cultural differences between Japan and America in terms of gun control (Cultural identity). |
| climate_change | ‘economic’, ‘security’, ‘quality_life’ | ‘cap&res’, ‘health’, ‘culture’ | The article discusses the potential impact of global warming on Antarctica‘s ice, which could lead to a rise in sea levels around the world, affecting coastal cities and potentially causing economic, security, and quality of life issues. Additionally, the article mentions the climate change aspect related to the melting of Antarcticas ice. |
| Tobacco | ‘health’, ‘legality’, ‘fairness’ | ‘policy’, ‘economic’, ‘quality_life’ | The article discusses the health implications of smoking, the legality of banning smoking in bars and restaurants, and the potential for unequal treatment of smokers compared to non-smokers. |
| climate_change | ‘economic’, ‘legality’, ‘policy’ | ‘political’, ‘regulation’, ‘cap&res’ | The article discusses the financial implications of adhering to the Kyoto treaty on global warming for the U.S. economy (Economic). The President’s statement that he couldn’t in good faith sign Kyoto because it does not apply to other nations implies a discussion of the treaty’s legality and constitutionality (Legality, constitutionality and jurisprudence). The article also mentions the White House disputing the science behind the conclusion that greenhouse gas emissions contribute to global warming, which suggests a policy prescription and evaluation (Policy prescription and evaluation). |
| Immigration | ‘economic’, ‘legality’, ‘fairness’, ‘policy’, ‘public_op’ | ‘quality_life’, ‘morality’, ‘political’ | The article discusses the financial implications of cutting federal assistance to legal immigrants (Economic), the constitutionality of denying assistance to legal residents (Legality, constitutionality and jurisprudence), the balance or distribution of rights and resources for legal immigrants compared to other residents (Fairness and equality), the discussion of specific policies aimed at addressing the issue (Policy prescription and evaluation), and the public sentiment towards allowing elderly immigrants to starve or children to live on the street (Public sentiment). |
| climate_change | ‘economic’, ‘policy’, ‘legality’, ‘public_op’, ‘political’, ‘regulation’ | ‘culture’, ‘cap&res’, ‘health’ | The article discusses the decision of a TV station manager to not cover global warming stories, citing financial implications (Economic), the evolution of the issue into politics (Policy prescription and evaluation, Political), the perceived lack of conclusive scientific evidence (Legality, constitutionality and jurisprudence), public sentiment towards the issue (Public sentiment), and international implications of the U.S.’s stance on global warming (External regulation and reputation). |

Table 10: Example misclassifications on the MFC dataset along with the model reasoning for the predictions

| **Topic** | **Issues** |
|---|---|
| Politics | Political Crisis, Political Corruption, Political Manipulation, Political Power Struggle, Political Persecution |
| Crime | Public Safety Concern, Criminal Threat, Urban Violence, Law and Order, Criminal Activity |
| Business | Financial Opportunity, Economic Opportunity, Economic Recovery, Investment Opportunity, Economic Burden |
| Legal | Political Persecution, Legal Battle, Political Crisis, Political Accountability, Political Corruption |
| Entertainment | Cultural Phenomenon, Cultural Icon, Cultural Celebration, Cultural Legacy, Musical Legacy |
| Health | Public Health Crisis, Health Crisis, Health Risk, Medical Breakthrough, Public Health Concern |
| Environment | Natural Disaster, Natural Disaster Threat, Public Health Crisis, Environmental Health Crisis, Environmental Crisis |
| War | Humanitarian Crisis, Military Conflict, Security Threat, National Security Threat, Military Aggression |
| Culture | Cultural Celebration, Community Celebration, Cultural Preservation, Cultural Icon, Cultural Enrichment |
| Economy | Economic Struggle, Economic Recovery, Economic Stability, Economic Burden, Economic Instability |
| Technology | Privacy and Consent, Cybersecurity Threat, National Security Threat, Technological Advancement, Privacy Concerns |
| Education | Financial Burden, Professional Development, Financial Relief for Student Loan Borrowers, Economic Burden, Educational Opportunity |
| Social Issues | Humanitarian Crisis, Racial Discrimination, Community Support, Housing Crisis, Cultural Conflict |
| Lifestyle | Culinary Delight, Culinary Innovation, Fashion Innovation, Cultural Celebration, Fashion Trend |
| Immigration | Humanitarian Crisis, Border Crisis, Security Threat, Economic Burden, National Security Threat |
| Finance | Financial Opportunity, Financial Burden, Financial Empowerment, Financial Struggle, Financial Insecurity in Retirement |
| Weather | Natural Disaster, Natural Disaster Threat, Natural Disaster Preparedness, Weather Alert, Natural Disaster Risk |
| Accident | Road Safety Concern, Tragedy, Tragic Accident, Safety Concern, Unforeseen Accident |
| Travel | Luxury Travel Experience, Nature's Disruption, Luxury Escapism, Economic Recovery, Luxury Tourism |

Table 11: 5 most frequent predicted issue frames per topic

| Topic Name | Accuracies | Example |
|---|---|---|
| **Accident** | 90%, 100% | *YONKERS, N.Y. – A 70-year-old woman was struck and killed by a car while walking on the sidewalk in Yonkers. It happened Sunday night on North Broadway. Investigators say the car then went over a retaining wall...* |
| **Crime** | 90%, 90% | *MINNEAPOLIS – The person suspected of causing a crash that killed five young women is in custody at the Hennepin County Jail. WCCO is not naming the man until he's charged with a crime, which prosecutors say could happen as soon as Tuesday...* |
| **Culture** | 70%, 70% | *CHICAGO (CBS) – You can start your summer with a pop of color at the new Andy Warhol exhibition on the campus of the College of DuPage. With more than 200 original photographs...* |
| **Economy** | 100%, 100% | *Andrew Ross Sorkin grilled White House economic advisor Heather Boushey on Wednesday over whether the Biden administration planned for increased inflation when the president passed several spending packages. The panel was discussing....* |
| **Education** | 100%, 90% | *BALTIMORE - Westminster National Golf Course hosted 100 third graders from Westminster Elementary School on Thursday for a hands-on cross-curricular STEM-related field trip. The students learned all about golf and the science...* |
| **Entertainment** | 100%, 100% | *A new animated fantasy comedy movie that follows the adventures of a preteen Latina who wants to do her own thing while surrounded by her multigenerational Mexican American family premieres Friday on Netflix...* |
| **Finance** | 100%, 100% | *WASHINGTON, May 30 (Reuters) - The former head of Wells Fargo & Co's (WFC.N) retail bank agreed to pay a $3 million penalty to settle the U.S. Securities and Exchange Commission's fraud charges for misleading investors...* |
| **Health** | 100%, 90% | *LOS ANGELES (AP), Madonna has postponed her career-spanning Celebration tour due to what her manager called a "serious bacterial infection." Manager Guy Oseary wrote on Instagram Wednesday that the singer had spent several days...* |
| **Immigration** | 100%, 100% | *A U.S. Citizenship and Immigration Services (USCIS) district office in New York City. USCIS expects to accept and approve a low number of H-1B registrations from the H-1B lottery, first selection round...* |
| **Legal** | 80%, 80% | *Former President Donald Trump stubbornly rejected his legal team's efforts last year to settle the classified documents case and prevent him from being indicted by a federal grand jury, according to a bombshell report. Christopher Kise, one of Trump's attorneys in the fall of 2022...* |
| **Lifestyle** | 80%, 80% | *The courgettes are roasting sweetly in the oven, half of them for lunch today dressed with sultanas, pine kernels and honey, the rest to serve as a salad tomorrow. This is something I also do with aubergines, red onions and sweet potatoes. There are so many vegetables to roast right now...* |
| **Politics** | 90%, 100% | *Rep. Matt Gaetz (R-Fla.) criticized House Republicans' recent effort to impeach President Joe Biden saying it was not done in a "legitimate" or "serious" way, a video obtained by NBC News shows, raising questions of whether he will support Rep. Jim Jordan (R-Ohio), who is a leading candidate to become House speaker...* |
| **Safety** | 90%, 50% | *The Sacramento region has some of the highest numbers of fatal traffic collisions in the state. Sacramento police say that last year, more than 50 people died on city streets. Now, as part of National Passenger Safety Week, there's an effort to reduce fatal collisions...* |
| **Social Issues** | 70%, 70% | *REDWOOD CITY, The San Mateo County Board of Supervisors is continuing to explore ways to provide more housing for farmworkers in the county, nearly four months after a mass shooting in Half Moon Bay exposed an urgent need for more living options for agricultural workers with low income...* |
| **Technology** | 70%, 100% | *America's spending on artifical intelligence in public safety is projected to increase from* $9.3billionin2022to71$ *billion by 2030, according to a new analysis by the Insight Partners research firm. The projected seven-year boom is expected to be fueled by global and domestic terrorism...* |
| **War** | 100%, 100% | *The Ukrainian air-assault force,Äôs 25th Brigade just got a lot heavier. Photos that appeared online this week confirm the brigade has re-equipped with German-made Marder infantry fighting vehicles. The 25th is the second Ukrainian air-assault brigade, after the 82nd, to get 31-ton Marders from German stocks...* |
| **Weather** | 100%, 100% | *Tropical Storm Nigel is expected to become a hurricane as soon as Monday, the National Hurricane Center said Sunday, and could be the latest tropical storm in the Atlantic this season to rapidly intensify to major hurricane status...* |

Table 12: The 19 topics used in our analysis. Accuracies represent the average acceptability judgments of two annotators (two of the paper authors) over a set of 10 predicted examples for each topic.