Mirrored Frames: LLM-Based Narrative Framing Analysis of Thai-Cambodian Border Conflict Coverage

Anonymous ACL submission

Abstract

Framing in news media shapes how events, actors, and causal responsibility are interpreted, yet computational framing analysis remains limited for lower-resource languages and fast-evolving international conflicts. In this work, we analyze narrative framing in news coverage of the 2025 Thailand–Cambodia border conflict across English, Thai, and Khmer news articles. To support benchmarking for this task, we manually annotate 51 conflict-related articles using a stakeholder-oriented Hero–Villain–Victim (HVV) framework and source attribution annotations. We then benchmark strong proprietary large language models (LLMs) as zero-shot annotators for both tasks against our manually annotated gold standard. Our evaluation reveals strong performance on Villain recognition, but substantially weaker performance on sympathetic narrative roles such as Hero and Victim. Applying the evaluated pipeline across both active-conflict and post-conflict periods further reveals largely mirrored partisan framing patterns in Thai and Cambodian national media, with outlets on both sides predominantly portraying their own country favorably while assigning antagonistic roles to the opposing side.

1 Introduction

Framing in media refers to the selective emphasis of information to promote particular interpretations while suppressing competing, incompatible perspectives (Entman, 1993). In news discourse, framing determines how events, actors, and causal relations are interpreted. A report on the same geopolitical event, for example, may frame one side as a hero, victim, or villain, depending on which details are foregrounded and which are omitted. Computational analysis of framing enables large-scale analyses of how narratives differ across target language (Sinelnik and Hovy, 2024) or across political leanings (Otmakhova and Frermann, 2025). Such analyses often require the model to handle multiple languages at scale.

The 2025 Thailand–Cambodia border conflict illustrates the importance of multilinguality especially for lower-resource languages. This conflict was a major escalation of a long-running territorial dispute along the Thai–Cambodian border. Tensions intensified in May 2025 after a fatal clash between Thai and Cambodian soldiers in a disputed border area, and the situation escalated into sustained armed confrontations in July 2025 involving artillery, rocket fire, and air strikes. Thai and Cambodian news outlets constructed divergent narratives around aggression, ceasefire violations, and diplomatic responsibility. Computational framing analysis of this international conflict was limited by the fact the majority of the news was covered by Thai and Cambodian presses in Thai and Khmer respectively. The question remains whether most of the population, who live miles away from the border, ever have access to unbiased, nonpartisan news on this matter. In this work, we want to analyze the narrative framing of the news coverage of the 2025 Thailand-Cambodia border conflict from Thai, Cambodian, and third-country news outlet.

Large language models (LLMs) have emerged as a strong and increasingly popular approach for automatic framing analysis due to their zero-shot capability for discourse-level understanding (Otmakhova and Frermann, 2023, 2025). Computational analysis of news is particularly valuable when conducted in a timely manner, making fully supervised approaches less practical in fast-evolving conflict settings where annotated data are rarely available in advance. In such contexts, zero-shot LLM-based analysis offers a promising alternative. However, benchmark datasets for framing analysis remain limited, particularly for lower-resource languages that are underrepresented in state-of-the-art and proprietary LLMs. To address this gap, we manually annotate 51 conflict-related

news articles in English, Thai, and Khmer using a stakeholder-based Hero–Villain–Victim (HVV) framework grounded in the Narrative Policy Framework (Jones and McBeth, 2010; Shanahan et al., 2018). In addition to narrative role annotations, we annotate source attribution to capture the entities or sources to which claims and reports are attributed within each article.

We evaluate strong proprietary LLMs as a zero-shot annotator for the two tasks against our manually annotated gold standard. Our evaluation identifies systematic patterns in LLM annotation behavior. The model performs strongly on antagonistic framing, with Villain detection achieving an F1 score of 0.90. On the contrary, performance is substantially lower for sympathetic roles. Hero and Victim detection achieves F1 scores of 0.53 and 0.61, respectively, with precision falling far below recall. Source attribution achieves a higher overall F1 of 0.71. Through qualitative analysis, we characterize two recurrent failure modes in narrative-role labeling. *Overreliance on local signals* occurs when localized sympathetic or suffering cues trigger article-level role assignment for peripheral entities. *Narrative salience misattribution* occurs when the model fails to identify the primary narrative protagonist of a given article.

Substantively, applying the evaluated pipeline across both batches reveals largely symmetric, "mirrored" partisan framing patterns. Thai and Cambodian national outlets both predominantly employ self-favoring narrative framing; that is, Thai outlets consistently cluster on Thailand-Hero and Cambodia-Villain framing, while Cambodian counterparts mirror this with Cambodia-Hero/Victim and Thailand-Villain framing. This mirrored adversarial framing persists across both active-conflict and post-conflict periods, and remains largely the same after active conflict.

Our contributions are as follows:

- We introduce a multilingual benchmark for analyzing narrative framing in the 2025 Thailand–Cambodia border conflict, covering English, Thai, and Khmer news articles annotated with stakeholder-oriented HVV roles and source attribution labels.[1]

- We evaluate proprietary LLMs as zero-shot framing annotators and show that, while strong on antagonistic framing, the models show consistent weaknesses in identifying sympathetic narrative roles and discourse-level narrative salience.

- We apply the model to examine the framing of the conflict across news outlets and find that Thai and Cambodian outlets construct largely mirrored self-favoring narratives that persist even beyond periods of active military escalation.

2 Related Work

Framing theory conceptualizes how communicators define problems, attribute causes, make moral evaluations, and propose remedies. Frames operate through selective foregrounding, highlighting certain attributes of an issue while omitting, obscuring, or downplaying others (Entman, 1993; Boydstun et al., 2014), particularly in conflict settings in which competing interpretations emerge. Computational work has expanded this tradition by distinguishing multiple levels of framing (e.g. emphasis, equivalency, connotation, and narrative framing) and analyzing their linguistic realizations (Otmakhova et al., 2024a).

Narrative framing approaches model political communication as stories structured around actors assigned moral roles. The Narrative Policy Framework (NPF) formalizes three core roles: Heroes, Villains, and Victims (Jones and McBeth, 2010). Later extensions also introduce Beneficiaries, who gain from a proposed policy solution (Weible et al., 2016), and Allies, who support the hero's goals (McBeth et al., 2005). Recent studies have adapted narrative framing to textual corpora through decomposing narratives into characters, conflicts, and cultural story structures (Otmakhova and Frermann, 2025).

Early research in computational journalism established foundational resources such as the Media Frames Corpus (Card et al., 2015), which contains over 20,000 news articles annotated across 15 framing dimensions (e.g. quality-of-life, political, and morality frames) derived from the taxonomy proposed by Boydstun et al. (2014). Subsequent studies applied supervised (Nguyen et al., 2015; Ji and Smith, 2017), semi-supervised (Khanehzar et al., 2019), and unsupervised (Card et al., 2016) machine-learning- and neural-network-based approaches to frame detection and classification tasks. In multilingual settings, embedding-based meth-

[1]The dataset will be made available on GitHub upon acceptance.

ods (Field et al., 2018) and transfer learning techniques (Akyürek et al., 2020) have also been employed. Comprehensive surveys of computational approaches to media framing are provided by Ali and Hassan (2022) and Otmakhova et al. (2024b).

More recently, large language models (LLMs) have come forth as a promising approach to automated framing analysis due to their strong performance in zero-shot and few-shot classification settings (Brown et al., 2020). Del Barrio et al. (2024) showed that prompt-engineered LLMs can achieve human–LLM annotation agreement rates of up to 43%. Islam and Goldwasser (2025) demonstrated that integrating LLMs into annotation workflows can both improve annotation accuracy and reduce cognitive load for human annotators. Notably, Alizadeh et al. (2023) and Gilardi et al. (2023) used zero-shot learning to classify Card et al.'s (Card et al., 2015) policy frames, with both studies demonstrating that GPT-family models and open-source LLMs consistently outperform crowd-annotators, respectively.

3 Data and Annotation

3.1 Dataset

We compile 51 conflict-relevant articles published during the active border clashes (June to September 2025). The articles are retrieved through keyword searches of each outlet's internal search engine. Extracted metadata includes URL, title, publication date, outlet, language, location, and full article text. Non-English articles are machine-translated into English prior to annotation. The downloaded data is used for research purposes only. The dataset is balanced by source country (26 Thai, 25 Cambodian) and spans ten news outlets, as outlined in Table 1. The released dataset will not include the news text but only the URLs that we download the news from.

Table 1: News outlets included in the dataset.

Country	Outlet	Articles
Thailand	ThaiPBS	6
Thailand	The Standard	5
Thailand	Bangkok Post	5
Thailand	BBC Thai	5
Thailand	Thairath	5
Cambodia	DAP News	5
Cambodia	ThmeyThmey	5
Cambodia	Cambodia Daily	5
Cambodia	Kampuchea Thmey	5
Cambodia	RFI Cambodia	5
	Total	51

3.2 Annotation Scheme

3.2.1 Narrative-Role Annotation

Annotators are instructed to identify the main narrative frames in each article along two dimensions, Nation and Narrative Role. The Nation dimension includes `TH_THAILAND`, `KH_CAMBODIA`, and `OTHER`, the latter of which is used when other actors are central to the narrative. The Narrative Role dimension comprises `HERO` (defending the nation, resolving conflict, peacekeeping, or being praised), `VILLAIN` (causing harm, escalating conflict, or being condemned), and `VICTIM` (the central actor portrayed as suffering harm).

Annotation is based on narrative salience; that is, a nation is labeled only if it is a central driver of action, blame, praise, or suffering in the narrative. Multiple nations may be annotated within the same article, and a single nation may receive multiple narrative role labels when supported by the narrative framing. Annotations are recorded as structured nation–label pairs, e.g. `OTHER-HERO`.

3.2.2 Source Attribution Annotation

Annotators are instructed to identify all source categories present in an article. A source is defined as any entity that provides information, opinions, statements, or reported attribution within the text. This includes entities that are directly quoted, paraphrased, or referenced as the origin of information or reporting. Entities that are only the subject of the story, but do not communicate information, are excluded.

Source labels comprise `MEDIA` (news outlets, media organizations, or journalists providing information or commentary), `GOVERNMENT` (government bodies, agencies, or officials making statements or taking action), and `MILITARY` (armed forces or military personnel, current or former, acting or speaking in a military capacity) stakeholders from `TH_THAILAND`, `KH_CAMBODIA`, and `OTHER` nations. Unverified sources and mentions of organizations such as ASEAN are excluded.

Annotators must first attempt to identify valid source categories, and are able to assign `NO_SOURCE` only if no valid source appears anywhere in the article. The `NO_SOURCE` label is mutually exclusive with all other categories and cannot co-occur with them. Outputs must always contain at least one label.

3.2.3 Inter-Annotator Agreement

We recruit two trained annotators proficient in English and familiar with Thai political discourse. To assess annotation reliability, the entire dataset is double-annotated by the two annotators. Inter-annotator agreement scores, measured using Cohen's κ, are 0.68 for narrative-role labeling and 0.73 for source attribution. Disagreements are resolved through adjudication to produce the final gold-standard labels.

4 Experimental Setup

We use GPT-5.2, Gemini 3.5 Flash, and Gemini 3.1 Pro as a zero-shot annotator for both narrative-role labeling and source attribution. All inference is performed at decoding temperature 0 to maximize output stability and reduce stochastic variance across runs. Each article is processed independently in two separate API calls, one for each annotation task. The articles that are not written in Egnlish were translated into English with GPT-5.2. The model is given only the full article text without in-context examples or fine-tuning. Both prompts share a common structure, i.e. a role specification (e.g. "narrative framing analyst"), a definition of the label space, a set of decision rules and constraints (see Section 3.2), and a strict JSON output format. Full prompt texts are provided in Appendices A and B.

We evaluate the LLM's predictions on the dataset against the gold standard. Both tasks are treated as multi-label classification at the article level. For narrative-role assignment, each article can have any subset of the nine possible labels. For source attribution, each article can have any subset of ten possible labels. Predictions are deduplicated before evaluation so that each label can be present at most once per article.

5 Results

Across models, overall recall sits at 0.96 for Gemini Flash, 0.92 for Gemini Pro, and 0.91 for GPT, while precision remains substantially lower at 0.45, 0.52, and 0.51, respectively, yielding overall F_1 scores of 0.61, 0.66, and 0.65 (Table 2). These results suggest that the models are consistently recall-biased, tending to over-trigger rather than omit gold labels.

Aggregated by role, the strongest and most consistent performance is observed for `VILLAIN` (Table 3). GPT achieves an F_1 score of 0.90 on `VILLAIN`, compared to 0.79 for Gemini Pro and 0.69 for Gemini Flash. By contrast, `HERO` and `VICTIM` remain substantially weaker across all models. For `HERO`, F_1 ranges from 0.53 to 0.59, while `VICTIM` remains comparatively stable at 0.61–0.62. For `THAILAND-VILLAIN`, F_1 sits at 0.77 (Gemini Flash), 0.88 (Gemini Pro), and 0.95 (GPT). Likewise, `CAMBODIA-VILLAIN` achieves 0.62, 0.73, and 0.86 across the three models. `THAILAND-HERO`, `THAILAND-VICTIM`, `CAMBODIA-HERO`, and `CAMBODIA-VICTIM` generally exhibit precision between 0.37 and 0.50, despite recall frequently exceeding 0.90. Collectively, these results suggest that the models are substantially better at identifying who is being blamed or condemned in an article (i.e. `VILLAIN`) than at distinguishing actors who are framed as sympathetic (i.e. `HERO` and `VICTIM`).

Performance on `OTHER` categories are treated as preliminary due to sparseness. `OTHER-HERO` contains nine gold instances and attains moderate F_1 scores between 0.51 and 0.59, albeit with relatively low precision. `OTHER-VICTIM` appears only twice in the gold data and yields unstable scores across models, while `OTHER-VILLAIN` contains a single gold instance and is not successfully identified by any model. These categories should therefore be interpreted with caution.

The performance of source-attribution task is somewhat stronger than narrative-role labeling, with overall F_1 scores of 0.74 (Gemini Flash), 0.77 (Gemini Pro), and 0.71 (GPT) (Table 4). Across models, recall sits at or near 1.00 for most government and military categories with non-trivial support, while precision varies more substantially due to over-prediction. For instance, `THIRD_COUNTRY_GOVERNMENT` reaches perfect recall across all three models but only moderate precision, yielding F_1 scores of 0.69 (Gemini Flash), 0.78 (Gemini Pro), and 0.60 (GPT). Similarly, `THAI_MILITARY` achieves recall between 0.93 and 1.00 but lower precision ranging from 0.58 to 0.82. `CAMBODIA_GOVERNMENT`, the most frequent category (20 gold instances), is detected reliably across models, with recall between 0.90 and 1.00 and F_1 between 0.81 and 0.84. Overall, the results suggest that source attribution is easier than narrative-role labeling but remains recall-biased. The models find nearly every real source instance, but also flags sources that are not actually attributed in the text.

Label	Gemini Flash			Gemini Pro			GPT			Qwen3-8B			Mistral-7B			Support
	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1	
THAILAND-HERO	0.39	0.92	0.55	0.42	0.83	0.56	0.42	0.92	0.58	0.476	0.833	**0.606**	0.270	0.833	0.408	12
THAILAND-VICTIM	0.39	1.00	0.56	0.42	1.00	**0.59**	0.39	0.90	0.55	0.310	0.900	0.462	0.263	1.000	0.417	10
THAILAND-VILLAIN	0.62	1.00	0.77	0.78	1.00	0.88	0.90	1.00	**0.95**	0.800	0.889	0.842	1.000	0.667	0.800	18
CAMBODIA-HERO	0.39	0.90	0.55	0.50	0.90	**0.64**	0.37	0.70	0.48	0.500	0.400	0.444	0.429	0.300	0.353	10
CAMBODIA-VICTIM	0.47	1.00	0.64	0.48	0.93	0.64	0.50	0.93	0.65	0.647	0.733	**0.688**	0.727	0.533	0.615	15
CAMBODIA-VILLAIN	0.44	1.00	0.62	0.57	1.00	0.73	0.75	1.00	**0.86**	0.407	0.917	0.564	0.353	1.000	0.522	12
OTHER-HERO	0.36	0.89	0.52	0.44	0.89	**0.59**	0.35	1.00	0.51	0.067	0.111	0.083	0.185	0.556	0.278	9
OTHER-VICTIM	1.00	1.00	**1.00**	1.00	0.50	0.67	1.00	0.50	0.67	0.000	0.000	0.000	0.500	0.500	0.500	2
OTHER-VILLAIN	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.000	0.000	0.000	0.000	0.000	0.000	1
Micro-average	0.45	0.96	0.61	0.52	0.92	**0.66**	0.51	0.91	0.65	0.413	0.697	0.519	0.343	0.685	0.457	89

Table 2: Per-label narrative-role labeling performance across models.

Label	G-Flash	G-Pro	GPT	Gold
HERO	0.54	**0.59**	0.53	31
VICTIM	**0.62**	**0.62**	0.61	27
VILLAIN	0.69	0.79	**0.90**	31
Overall	0.61	**0.66**	0.65	89

Table 3: Narrative-role labeling performance aggregated by HVV role across nations (F_1 scores).

6 Error Analysis

We identify a few failure modes observed across the two tasks by inspecting false-positive predictions. All failure modes share a common origin, in that the models are recall-biased and produce labels from local cues rather than discourse-level synthesis. Zero-shot LLM annotation for framing tasks may benefit from prompts that explicitly force article-level judgments before label assignment, though we leave systematic testing of this hypothesis to future work.

6.1 Overreliance on Local Signals

The model frequently assigns Hero or Victim labels based on individual sentences although the broader article frames the entity as Villain or non-central character to the narrative. Background mentions of compliance, peace intentions, casualty counts, or aid efforts trigger Hero label assignment for entities that are not the article's narrative main characters.

To exemplify this pattern of error, consider a Thai article on ceasefire violations whose narrative consistently frames Cambodia as the aggressor and Thailand as the victim of those violations. The gold annotation labels KH-VILLAIN and TH-VICTIM. The model produces six labels: KH-VILLAIN, TH-HERO, TH-VICTIM, KH-HERO, KH-VICTIM, and OTHER-HERO. We found a few sentences that might lead the overproduction of labels. The phrase "Cambodia insists it remains fully compliant with the agreement" likely triggers KH-HERO, despite the article rejecting this claim. "The Thai army stated that Thailand strictly complied with the agreement" likely triggers TH-HERO, even though Thailand's compliance is mentioned only as part of the wider accusation against Cambodia. A single mention of ASEAN Chair Anwar Ibrahim brokering ceasefire negotiations triggers OTHER-HERO, even though Anwar is mentioned only once and is peripheral to the narrative.

The narrative-role labeling prompt instructs the model to identify "central drivers of action, blame, praise, or suffering." The phrasing of Hero ("defending the nation, peace-keeping, or being praised") and Victim ("the central group suffering harm") admits a wide range of weak surface cues. Compared to Villain ("causing harm, escalating conflict, or being condemned"), which has sharper antagonistic lexical anchors, Hero and Victim definitions can fire on local sympathetic phrases without article-level grounding. We hypothesize that this is a primary source of the recall-precision asymmetry observed in the evaluation.

To probe the sentence-level mechanism, we removed flagged trigger sentences from four articles and re-ran classification using GPT-5.2. Across the four articles, 4 of 13 false postives (FPs) (31%) were eliminated by sentence removal. The remaining 9 persisted, with the model's stated reasoning citing distributed sympathetic content elsewhere in the article. The eliminated cases corresponded to FPs anchored to a single sympathetic mention; the persistent cases corresponded to FPs supported by multiple distributed cues. This suggests that the failure mode operates on a spectrum from localized to distributed triggers, with the latter closer to narrative salience misattribution (Section 6.2). Full ablation under matched API conditions and multi-run replication remains future work.

Category	Gemini Flash P	Gemini Flash R	Gemini Flash F1	Gemini Pro P	Gemini Pro R	Gemini Pro F1	GPT P	GPT R	GPT F1	Gold
THAI_MEDIA	0.67	0.77	0.71	0.83	0.77	**0.80**	0.73	0.85	0.79	13
CAMBODIA_MEDIA	0.62	1.00	0.77	1.00	1.00	**1.00**	1.00	0.80	0.89	5
THIRD_COUNTRY_MEDIA	0.50	1.00	0.67	0.50	1.00	0.67	0.60	1.00	**0.75**	6
THAI_GOVERNMENT	0.62	1.00	**0.77**	0.62	0.87	0.72	0.64	0.93	0.76	15
CAMBODIA_GOVERNMENT	0.70	0.95	0.81	0.78	0.90	**0.84**	0.71	1.00	0.83	20
THIRD_COUNTRY_GOVERNMENT	0.53	1.00	0.69	0.64	1.00	**0.78**	0.43	1.00	0.60	9
THAI_MILITARY	0.71	1.00	0.83	0.82	0.93	**0.88**	0.58	1.00	0.73	15
CAMBODIA_MILITARY	0.17	1.00	**0.29**	0.00	0.00	0.00	0.00	0.00	0.00	1
THIRD_COUNTRY_MILITARY	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
NO_SOURCE	1.00	0.43	**0.60**	0.75	0.43	0.55	0.00	0.00	0.00	7
Overall	0.62	0.91	0.74	0.70	0.86	**0.77**	0.60	0.87	0.71	91

Table 4: Per-label source attribution performance across models.

6.2 Narrative Salience Misattribution

In some articles, the model fails to identify which entity is the actual protagonist of the narrative and instead assigns roles to entities that appear only as background context. This is distinct from overreliance on local signals in that the failure is not driven by a single triggering sentence, but by the model's inability to weight narrative centrality across the article.

To exemplify this pattern of error, we look to an article which describes a U.S.-based lobbyist accused of bias in the Thai-Cambodia conflict and subsequently barred from entering Thailand. The narrative protagonist is the lobbyist, with Thailand framed as the entity that bars and accuses him. The gold annotation labels OTHER-VICTIM, TH-VILLAIN, and KH-VICTIM. The model produces KH-VICTIM, KH-HERO, TH-VILLAIN, TH-HERO, and OTHER-HERO, but misses OTHER-VICTIM entirely. Background details, such as Thailand having received Cambodian refugees during the Pol Pot era and the lobbyist deciding to "abandon all politics with a focus on humanity", may have triggered spurious Hero labels for Thailand and the lobbyist, while the article's actual protagonist (the lobbyist as victim of Thai retaliation) is missed.

This failure mode suggests that the model treats role assignment as a per-mention judgment rather than a discourse-level synthesis. Without explicit instructions to identify the article's protagonist or weight central versus peripheral entities, the model defaults to surface-level role attribution.

6.3 Subject-Source Conflation

The model frequently assigns source status to entities that are merely subjects of the article rather than originators of information. The prompt explicitly excludes "entities that are only the subject of the story but do not speak or provide info", but this distinction is not reliably honored.

News outlet	Active conflict	Post conflict	Total
Cambodia	536	766	1302
Thailand	852	918	1770
Grand Total	1388	1684	3072

Table 5: Distribution of articles across conflict periods and news outlets included in the analysis

The pattern is most visible in categories with a high ratio of predicted to gold positives. THIRD_COUNTRY_GOVERNMENT has 9 gold instances and 21 predictions, a ratio of more than 2:1. THAI_MILITARY has 15 gold and 26 predictions. In each case, the additional predictions correspond to entities that the article describes (e.g. a foreign president whose policies are discussed) but who do not themselves provide quoted or paraphrased information.

Source attribution requires a distinction between entities that *produce* information and entities that are *discussed in* said information. This distinction is discourse-level and depends on the structural role of each mention in the article. The model's tendency to conflate the two suggests that surface mentions of any salient entity are weighted as potential sources, even when the prompt provides explicit exclusion criteria.

7 Analysis of Narrative Framing across News Outlets

We want to investigate if narrative framing by Thai and Cambodian news outlets changes after the active conflict. After the period which constitutes the main timeframe of the dataset used in this study (i.e.

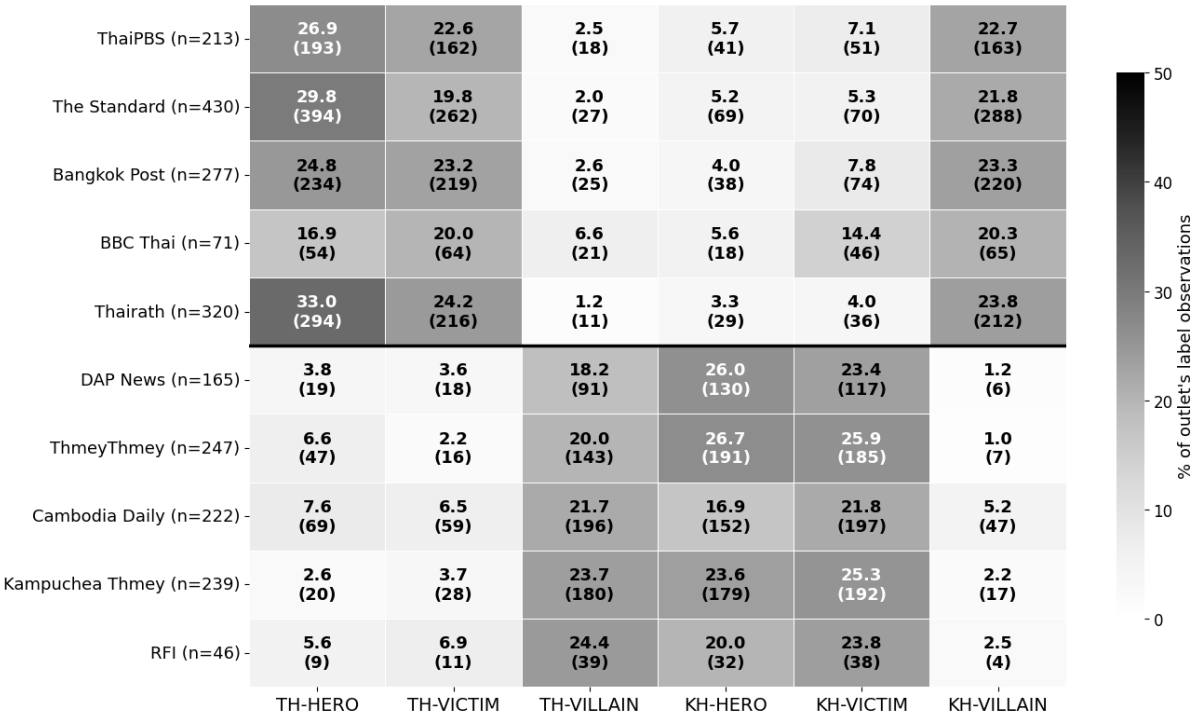

Figure 1: GPT-predicted narrative-role label distribution per outlet, normalized as a percentage of each outlet's total label observations. The horizontal line separates Thai and Cambodian outlets. TH = Thailand target, KH = Cambodia target.

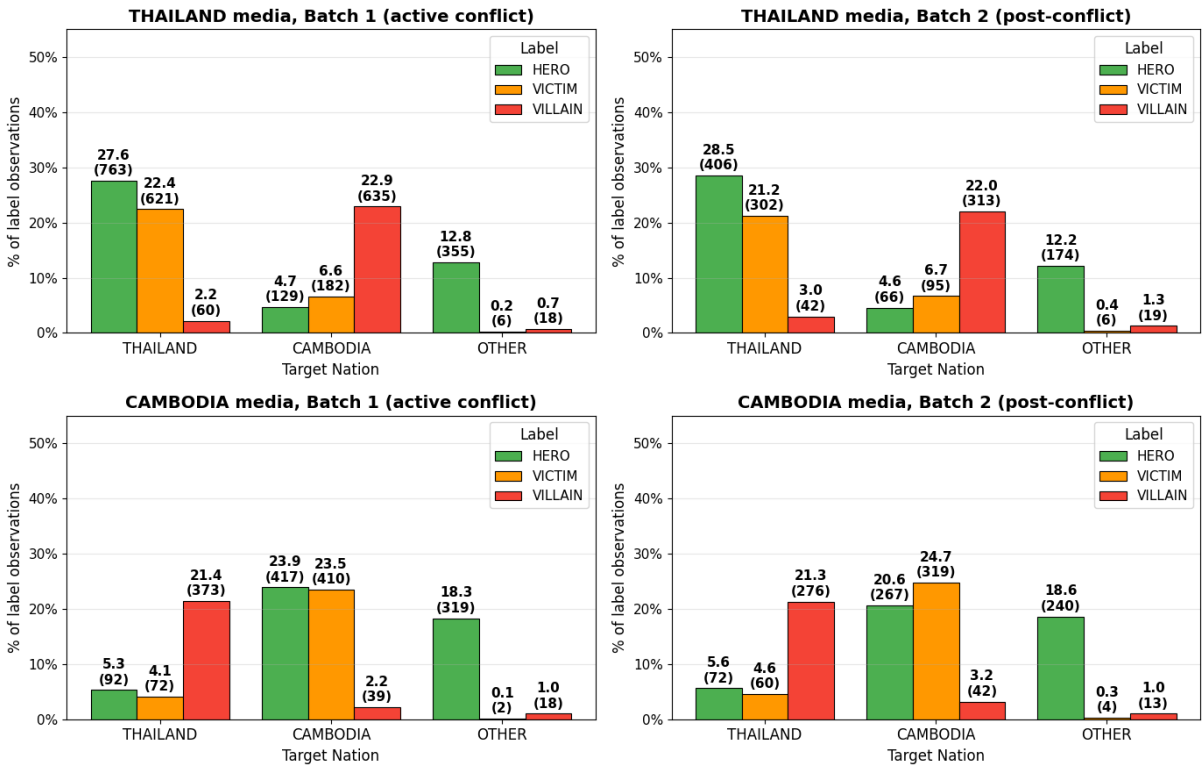

Figure 2: Predicted narrative-role label distribution by outlet country and batch, based on GPT results. Top row: Thai outlets, active-conflict (June to September 2025) and post-conflict (December 2025 to April 2026). Bottom row: Cambodian outlets, same split. Bar values show percentage of label observations and raw count in parentheses.

June to September 2025), a ceasefire agreement, i.e. the Kuala Lumpur Peace Accord, was signed on October 26, 2025 during the 47th ASEAN Summit. A new wave of media coverage after December 2025 focuses on ceasefire fragility and humanitarian consequences. We collect and filter news articles about the Thai-Cambodia conflict during the active conflict period (June to September 2025) and the post-active conflict period (December 2025 to April 2026) from the same set of Thai and Cambodia news outlets. This dataset consists of 1,388 articles collected during the active-conflict period and 1,684 during the post-conflict period. Thai news outlets contribute 1,770 articles and Cambodian news outlets contribute 1,302 articles. Additionally, we collect 32 articles from BBC Thai as a third-country news outlet because other international news outlet only publishes 1 - 3 articles on this conflict.

These articles are processed only by the LLM pipeline presented in the previous section and used for exploratory comparative framing analysis. We use GPT-5.2 only, excluding the Gemini-family models for ease of interpretation. We apply the same prompts and decoding settings to the post-conflict set without further modification. Running the GPT-5.2 API on the entire dataset costs around $8. We use the resulting predictions on both datasets to characterize the narrative framing by Thai and Cambodian news outlets.

7.1 Per-Outlet Framing Patterns

The narrative framing by Thai and Cambodian news outlets shows a 'mirrored' pattern (Figure 1). Thai outlets cluster on `TH-HERO` and `KH-VILLAIN` framing: Thai PBS, The Standard, Thairath, and Bangkok Post each devote around 25 - 33% of their predicted labels to `TH-HERO` and 20 - 24% to `KH-VILLAIN`, with Cambodian-target Hero and Victim labels largely absent. Although UK-based, BBC Thai show the same patterns because the agency is staffed by local Thais. This suggests that Thai news outlets predominantly frame Thailand as a legitimate or heroic actor while portraying Cambodia as the aggressor, reflecting a strongly nationalized narrative perspective with limited representation of Cambodian-centered frames. Cambodian outlets mirror this pattern. The Cambodian news outlets use `KH-HERO` around 17 - 27% and `TH-VILLAIN` around 18 - 24%.

7.2 Active Conflict vs. Post-Conflict Comparison

Both Thai and Cambodian outlets do not soften the "mirrored" partisan pattern during the post-conflict period (Figure 2). A chi-square test of independence showed no significant difference in the overall framing-label distribution between active-conflict period and post-conflict for Thai media ($\chi^2(8) = 11.53, p = 0.173$). Similarly, the distribution of narrative-role labels are the same for both period for Cambodian news outlets ($\chi^2(8) = 8.42, p = 0.393$). Initially, we hypothesize that mirrored framing pattern would decrease as active conflict ends. However, the directional pattern is consistent across both sides and multiple label categories, suggesting that adversarial framing persists or even strengthens after active fighting has ended.

8 Conclusion

In this paper, we introduced a multilingual benchmark for narrative framing analysis of the 2025 Thailand–Cambodia border conflict, comprising English, Thai, and Khmer news articles annotated for stakeholder-oriented narrative-role labels and source attribution. We evaluated proprietary LLMs as zero-shot framing annotators, finding that they do well on *Villain* framing but fall short on identifying *Hero* and *Villain* and discourse-level narrative salience. Applying one of the models to cross-outlet analysis, we showed that Thai and Cambodian media construct largely mirrored, self-favoring narratives which remain even after periods of active military escalation.

9 Limitations

LLM outputs are inherently non-deterministic and may vary across prompts and model families. However, despite these differences, the evaluated models display similar performance profiles and recurring error patterns. The findings should therefore be interpreted mainly in terms of recurring trends across models and articles, rather than exact outputs for any particular article.

References

Afra Feyza Akyürek, Lei Guo, Randa Elanwar, Prakash Ishwar, Margrit Betke, and Derry Tanti Wijaya. 2020. Multi-label and multilingual news framing analysis. In *Proceedings of ACL 2020*.

Mohammad Ali and Naeemul Hassan. 2022. A survey of computational framing analysis approaches. In *Proceedings of EMNLP 2022*, pages 9335–9348, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Meysam Alizadeh, Maël Kubli, Zeynab Samei, Shirin Dehghani, Juan Diego Bermeo, Maria Korobeynikova, and Fabrizio Gilardi. 2023. Open-source large language models outperform crowd workers and approach chatgpt in text-annotation tasks. *arXiv preprint arXiv:2307.02179*.

David Alonso del Barrio, Max Tiel, and Daniel Gatica-Perez. 2024. Human interest or conflict? leveraging llms for automated framing analysis in tv shows. In *Proceedings of the ACM International Conference on Interactive Media Experiences (IMX '24)*.

Amber E. Boydstun, Dallas Card, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2014. Tracking the development of media frames within and across policy issues. APSA 2014 Annual Meeting Paper.

Tom B. Brown, Benjamin Mann, Nick Ryder, and 1 others. 2020. Language models are few-shot learners. In *Proceedings of NeurIPS 2020*, pages 1877–1901. Curran Associates.

Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of ACL-IJCNLP 2015 (Volume 2: Short Papers)*, pages 438–444, Beijing, China. Association for Computational Linguistics.

Dallas Card, Justin Gross, Amber Boydstun, and Noah A. Smith. 2016. Analyzing framing through the casts of characters in the news. In *Proceedings of EMNLP 2016*, pages 1410–1420, Austin, Texas. Association for Computational Linguistics.

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. *arXiv preprint arXiv:1808.09386*.

Fabrizio Gilardi, Meysam Alizadeh, and Maël Kubli. 2023. Chatgpt outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30).

Tunazzina Islam and Dan Goldwasser. 2025. Uncovering latent arguments in social media messaging by employing LLMs-in-the-loop strategy. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 7412–7444, Albuquerque, New Mexico. Association for Computational Linguistics.

Yangfeng Ji and Noah A. Smith. 2017. Neural discourse structure for text categorization. *arXiv preprint arXiv:1702.01829*.

Michael D. Jones and Mark K. McBeth. 2010. A narrative policy framework: Clear enough to be wrong? *Policy Studies Journal*, 38(2):329–353.

Shima Khanehzar, Andrew Turpin, and Gosia Mikolajczak. 2019. Modeling political framing across policy issues and contexts. In *Proceedings of the 17th Annual Workshop of the Australasian Language Technology Association*, pages 61–66.

Mark K. McBeth, Elizabeth A. Shanahan, and Michael D. Jones. 2005. The science of storytelling: Measuring policy beliefs in greater yellowstone. *Society & Natural Resources*, 18(5):413–429.

Viet-An Nguyen, Jordan Boyd-Graber, Philip Resnik, and Kristina Miler. 2015. Tea party in the house: A hierarchical ideal point topic model and its application to republican legislators in the 112th congress. In *Proceedings of ACL-IJCNLP 2015 (Volume 1: Long Papers)*, pages 1438–1448.

Yulia Otmakhova and Lea Frermann. 2023. Conflicts, villains, resolutions: Towards models of narrative framing in the news. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics.

Yulia Otmakhova and Lea Frermann. 2025. Narrative media framing in political discourse. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 9167–9196, Vienna, Austria. Association for Computational Linguistics.

Yulia Otmakhova, Shima Khanehzar, and Lea Frermann. 2024a. 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.

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

Elizabeth A. Shanahan, Michael D. Jones, and Mark K. McBeth. 2018. How to conduct a narrative policy framework study. *The Social Science Journal*, 55(3):332–345.

Antonina Sinelnik and Dirk Hovy. 2024. Narratives at conflict: Computational analysis of news framing in multilingual disinformation campaigns. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 131–143, Bangkok, Thailand. Association for Computational Linguistics.

Christopher M. Weible, Kristin L. Olofsson, Daniel P. Costie, Juniper M. Katz, and Tanya Heikkila. 2016. Enhancing precision and clarity in the study of policy

narratives: An analysis of climate and air issues in delhi, india. *Review of Policy Research*, 33(4):420–441.

Appendix

A Narrative-Role Annotation Prompt

The full prompt provided to each model for narrative-role labeling is reproduced below.

```
You are an expert narrative framing analyst.

Your task is to identify the MAIN NARRATIVE
    FRAMES in this news article using ONLY the
    following dimensions:

NATION:
- THAILAND
- CAMBODIA
- OTHER (third countries or international actors,
     if central)

LABEL:
- HERO: Fixing the problem, defending the nation,
     peace-keeping, or being praised.
- VILLAIN: Causing harm, escalating conflict, or
     being condemned.
- VICTIM: The central group suffering harm (
    narrative focus).

CRITICAL RULES:

1. NARRATIVE SALIENCE ONLY
- Include a nation ONLY if it is a central
    driver of action, blame, praise, or
    suffering in the narrative.
- If multiple nations are central, include all
    of them.

2. MULTIPLE LABELS ALLOWED
- The same nation MAY appear more than once with
     different labels.

3. STRICT OUTPUT RULE
- Output JSON ONLY.
- No explanations.
- No extra text.

Output format:
[
    {"nation": "THAILAND", "label": "HERO"},
    {"nation": "THAILAND", "label": "VICTIM"},
    {"nation": "CAMBODIA", "label": "VILLAIN"},
    {"nation": "OTHER", "label": "HERO"}
]

Input Text:
"{article_text}"
```

B Source Attribution Annotation Prompt

The full prompt provided to each model for source attribution is reproduced below.

```
You are an expert news analyst specializing in
    Source Attribution. Your task is to identify
     all SOURCE CATEGORIES present in this news
    article.

1. DEFINITION: WHAT IS A SOURCE?
- A SOURCE is any entity that provides
    information, opinions, or is quoted in the
    text.
- INCLUDE:
  - Entities directly quoted (Minister X said
    ...).
  - Entities paraphrased (According to the Army
    ...).
  - Origins of reports (Reported by Khaosod...,
    Viral on Facebook...).
- EXCLUDE:
  - Entities that are only the subject of the
    story but do not speak or provide info.

2. STAKEHOLDER CATEGORIES (Strictly use these
    definitions):
- THAI_MEDIA: Refers to Thai news outlets, media
     organizations, or journalists providing
    information or commentary.
- CAMBODIA_MEDIA: Refers to Cambodian news
    outlets, media organizations, or journalists
     providing information or commentary.
- THIRD_COUNTRY_MEDIA: Refers to foreign or
    international news outlets or journalists
    outside Thailand and Cambodia.
- THAI_GOVERNMENT: Refers to Thai government
    bodies, agencies, or officials making
    statements or taking action.
- CAMBODIA_GOVERNMENT: Refers to Cambodian
    government bodies, agencies, or officials
    making statements or taking action.
- THIRD_COUNTRY_GOVERNMENT: Refers to government
     representatives or agencies from countries
    other than Thailand or Cambodia.
- THAI_MILITARY: Refers to Thailand's armed
    forces or military personnel, current or
    former, acting or speaking in a military
    capacity.
- CAMBODIA_MILITARY: Refers to Cambodia's armed
    forces or military personnel, current or
    former, acting or speaking in a military
    capacity.
- THIRD_COUNTRY_MILITARY: Refers to countries
    other than Thailand or Cambodia's armed
    forces or military personnel, current or
    former, acting or speaking in a military
    capacity.
- NO_SOURCE: ONLY use this stakeholder when no
    source is REALLY referred.

3. EXCLUSIVE FALLBACK RULE (CRITICAL):
- You must first attempt to find valid sources
    from the categories above (Government, Media
    , Military, etc.).
- IF AND ONLY IF you find ZERO valid sources in
    the entire text, then (and only then) output
     a single source category:
    {"source_category": "NO_SOURCE"}
- NEVER output "NO_SOURCE" if you have already
    identified another valid source category.
- A list containing both a valid source and "
    NO_SOURCE" is strictly FORBIDDEN.

4. STRICT FILTERING & DEDUPLICATION RULE:
- IGNORE ANY ENTITY that does not fall into one
    of the STAKEHOLDER CATEGORIES listed above.
- If a person or organization is mentioned but
    they are not a government, media, or
    military entity from Thailand, Cambodia, or
    a Third Country, DO NOT include them in the
    output.
- Output a unique list of categories found. If a
     category appears multiple times in the text
```

```
            , only list it ONCE in your output.

        5. ANTI-EMPTY RULE (CRITICAL):
        - Returning an empty JSON object or empty list
            is STRICTLY UNACCEPTABLE.
        - Your response MUST ALWAYS contain at least one
             categorization.
        - If you cannot find any valid categories, you
            must strictly follow Rule 3 and output "
            NO_SOURCE".

        Output Format (JSON List):
        [
            {"source_category": "THAI_GOVERNMENT"},
            {"source_category": "CAMBODIA_MEDIA"}
        ]
        OR (if absolutely no sources exist):
        [
            {"source_category": "NO_SOURCE"}
        ]

        Input Text:
        "{article_text}"
```