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

Siwakorn Bubphasawan and Attapol T. Rutherford*
Department of Linguistics
Faculty of Arts,
Chulalongkorn University

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 benchmark proprietary and open-weight large language models (LLMs), compare original-language and translated input, evaluate salience-focused prompt variants, and test supervised cross-dataset transfer. Original-language inference and explicit article-level salience instructions improve the best micro-averaged F_1 from 0.66 to 0.73. Applying this improved pipeline across active-conflict and post-conflict periods suggests largely mirrored partisan framing patterns in Thai and Cambodian national media.

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 multilingual analysis, 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 is complicated by the fact that most coverage was published by Thai and Cambodian outlets in Thai and Khmer, respectively, limiting access to cross-national perspectives. We therefore analyze how Thai, Cambodian, and third-country outlets frame the conflict.

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, we annotate source attribution to capture the entities or

*Corresponding author.

Findings of the Association for Computational Linguistics: IJCNLP-AACL 2026, pages 871–884
November 6-10, 2026 ©2026 Association for Computational Linguistics

sources to which claims and reports are attributed within each article. We evaluate proprietary and open-weight LLMs as zero-shot annotators against our manually annotated gold standard. Our evaluation identifies systematic patterns in LLM annotation behavior. Through qualitative analysis, we identify errors consistent with two hypotheses: *overreliance on local signals*, in which localized cues may contribute to excess article-level roles, and *narrative salience misattribution*, in which predictions do not reflect the primary narrative protagonist.

Applying the evaluated pipeline across both batches suggests 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 directional pattern appears in both the active-conflict and post-conflict periods.

Our contributions are as follows:

- We create a dataset 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 benchmark proprietary and open-weight zero-shot LLMs, original-language versus translated input, salience-focused prompting strategies, and supervised cross-dataset transfer.
- We apply the model to examine framing across news outlets and find evidence of largely mirrored self-favoring narratives in both active-conflict and post-conflict coverage.

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., 2024).

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 methods (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. (2024).

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

[1]The dataset is publicly available on GitHub: https://github.com/nlp-chula/mirrored-frame.

studies demonstrating that GPT-family models and open-source LLMs consistently outperform crowd-annotators, respectively. Closest to our role inventory, Mahmoud et al. (2025) introduce a multilingual news corpus for entity framing and role portrayal. Their hierarchical taxonomy groups fine-grained archetypes under three broad roles, i.e. Protagonist, Antagonist, and Innocent, and provides a relevant source domain for evaluating cross-dataset transfer to our Hero, Villain, and Victim labels.

3 Data and Annotation

3.1 Dataset

We compile 51 conflict-relevant articles published during the active border clashes (June to September 2025). We select prominent Thai and Cambodian outlets that regularly reported on the conflict and retrieve articles through keyword searches of each outlet's internal search engine. After conflict-relevance filtering, the retrieval pool contains more than 1,000 articles. The annotated benchmark is sampled from this pool and is approximately balanced across outlets, with five articles per outlet except for six from ThaiPBS. Extracted metadata includes URL, title, publication date, outlet, language, location, and full article text. English-language reporting was included where available, but sparse coverage by international outlets, typically only one to three relevant articles per outlet, prevented us from constructing a separate balanced international-news stratum. The dataset is balanced by source country (26 Thai, 25 Cambodian) and spans ten outlets, as outlined in Table 1. This outlet-defined collection does not estimate a fixed percentage of the entire Thai or Cambodian media market. The final gold standard contains 89 narrative-role assignments across the 51 articles. The downloaded data is used for research purposes only. The released dataset will contain article URLs, metadata, and annotations, but not the copyrighted article text.

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

Table 1: News outlets included in the dataset.

3.2 Annotation Scheme

We distinguish an individual narrative *role* from the broader concept of *framing*. A role is an article-level label describing how a particular national actor is portrayed, such as Hero, Villain, or Victim. Narrative framing refers to the wider interpretive structure created by the combination, distribution, and salience of these roles within an article or corpus. We operationalize narrative framing through the stakeholder-oriented role assignments described below.

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. During adjudication, we operationalize salience by requiring a candidate actor to be discussed in at least three sentences, thereby distinguishing sustained framing from isolated or background mentions. 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, me-

dia 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. We focus on these institutional categories to examine how authority is invoked in conflict reporting. Civilians, residents, witnesses, non-governmental organizations, unverified sources, and mentions of organizations such as ASEAN fall outside the present label inventory. Consequently, this task does not measure which civilian or non-institutional voices an outlet selects or omits.

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 Annotation Process

We recruit two trained Thai annotators proficient in English and familiar with Thai political discourse. They annotate the 26 Thai-source articles in their original publication language and the 25 Khmer-source articles from English machine translations. To assess annotation reliability, the entire dataset is double-annotated. Inter-annotator agreement scores, measured using Cohen's κ, are 0.68 for narrative-role labeling and 0.73 for source attribution. For each disagreement, the annotators and authors discuss the competing labels with reference to the guidelines until consensus is reached. Most borderline cases concern whether an actor is sufficiently salient to receive an article-level role, motivating the three-sentence operational criterion described above.

Because translation may alter politically or culturally loaded framing cues, we conduct an additional validation of all 25 Khmer articles. A native Khmer speaker familiar with Cambodian news and political discourse independently annotates the original Khmer texts while blinded to both the English translations and the existing annotations. Agreement between this annotation and the previously adjudicated labels is $\kappa = 0.63$ for narrative-role labeling and $\kappa = 0.69$ for source attribution. The remaining disagreements are adjudicated with the existing annotations to produce the final gold standard used in our evaluation.

4 Experimental Setup

We use GPT-5.2, Gemini 3.5 Flash, and Gemini 3.1 Pro as zero-shot annotators for both narrative-role labeling and source attribution. For narrative-role labeling, we additionally report open-weight baselines using Qwen3-8B and Mistral-7B. Proprietary-model 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 models receive only the full article text, without in-context examples or fine-tuning. The original prompts translate the annotation guidelines into a role specification (e.g. "narrative framing analyst"), a definition of the label space, decision rules and constraints (see Section 3.2), and a strict JSON output format.

We conduct two additional narrative-role experiments with Gemini Pro. First, we compare English machine translations with the articles in their original publication languages. Second, guided by the error analysis, we evaluate two prompt variants designed to reduce assignments based on isolated local cues. Prompt 1 explicitly requires dominant article-level framing and instructs the model to prefer fewer, well-supported labels. Prompt 2 adds detailed role definitions and a Hero–Victim tie-breaking rule. Because these prompt revisions were motivated by errors observed on the same 51-article benchmark, their results measure prompt refinement on the available data rather than performance on an independently held-out test set. Full prompt texts are provided in Appendices B, C, and E.

As a supervised transfer baseline, we fine-tune XLM-R on the entity-framing dataset of Mahmoud et al. (2025), mapping Protagonist, Antagonist, and Innocent to Hero, Villain, and Victim, respectively. We evaluate the resulting model on both English translations and original-language articles. Under *union* decoding, we retain all predicted roles, while under *argmax* decoding, each nation retains only its highest-probability role.

We evaluate the LLM predictions against the final 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 the proprietary models, overall recall is 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 5). These results suggest that the models are consistently recall-biased, tending to over-trigger rather than omit gold labels. The open-weight Qwen3-8B and Mistral-7B baselines obtain lower micro-averaged F_1 scores of 0.52 and 0.46, respectively, although Qwen3-8B is competitive on several individual Thailand- and Cambodia-target labels.

Aggregated by role, the strongest and most consistent performance is observed for VILLAIN (Table 2). 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).

Results for the OTHER categories are treated as preliminary because of their sparse support. OTHER-HERO contains nine gold instances and attains moderate F_1 scores between 0.51 and 0.59 for the proprietary models, 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 are included in the reported micro-average but should not be interpreted as reliable category-level estimates.

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

Input	Prompt	Micro-F_1
English translation	Original	0.66
Original language	Original	0.68
Original language	Prompt 1	**0.73**
Original language	Prompt 2	0.69

Table 3: Gemini Pro narrative-role performance by input language and prompt variant.

5.1 Input Language and Prompt Refinement

Running Gemini Pro directly on the original Thai and Khmer texts improves micro-averaged F_1 from 0.66 to 0.68 relative to English machine translations (Table 3). Prompt 1 further raises performance to 0.73, whereas the more elaborate Prompt 2 reaches 0.69. The improvement from Prompt 1 supports the error analysis: explicit instructions to prioritize dominant framing and omit weakly supported labels reduce the tendency to over-trigger from local cues. Prompt 2's additional definitions do not provide a further benefit, suggesting that more detailed instructions are not necessarily more effective.

5.2 Cross-Dataset Transfer

Table 7 reports the XLM-R transfer baseline. Argmax decoding improves overall F_1 from 0.42 to 0.45 on English translations and from 0.46 to 0.50 on original-language input. Nevertheless, the strongest XLM-R configuration remains 0.23 F_1 points below the best zero-shot LLM. This gap indicates limited transfer between annotation settings that differ in domain, language, annotation unit, and role taxonomy, and should not be interpreted as a general comparison between supervised and zero-shot learning. Full role-level results are reported in Appendix D.

Performance on source attribution is somewhat stronger than on narrative-role labeling, with overall F_1 scores of 0.74 (Gemini Flash), 0.77 (Gemini Pro), and 0.71 (GPT) (Table 6). Across models, recall is at or near 1.00 for most government and military categories with non-trivial

support, while precision varies more substantially because of 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 flag sources that are not actually attributed in the text.

6 Error Analysis

Qualitative inspection of false-positive predictions suggests that the models may over-weight localized cues and under-weight article-level narrative salience. Because the original prompts request labels without rationales or supporting spans, this inspection cannot establish which passage caused an individual prediction. We therefore treat the examples below as hypothesis-generating. The prompt-refinement experiment in Section 5 provides aggregate behavioral evidence consistent with this hypothesis: explicit article-level salience instructions improve micro-averaged F_1 from 0.68 to 0.73 on original-language input. This improvement does not directly reveal the model's internal decision process.

6.1 Overreliance on Local Signals

Under the original prompt, the model frequently produces additional Hero or Victim labels when an article contains localized favorable or suffering-related descriptions of actors that are peripheral to its dominant narrative. Because the model output contains no rationale or evidence span, we cannot attribute an individual false positive to a particular sentence.

To exemplify this pattern, 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. The article contains localized favorable statements about Cambodian compliance, Thai compliance, and ASEAN Chair Anwar Ibrahim's role in ceasefire negotiations. These passages illustrate the type of cue that is consistent with our hypothesis, but the category-only output does not show that any particular passage caused a specific prediction.

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 sympathetic descriptions. Compared with Villain ("causing harm, escalating conflict, or being condemned"), which has sharper antagonistic lexical anchors, these definitions may make article-level centrality harder to enforce. The improvement from Prompt 1 is consistent with this interpretation, but does not establish a sentence-level causal mechanism.

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, we consider an article describing 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. The article also contains localized favorable descriptions of Thailand and the lobbyist that coexist with its dominant framing. These descriptions are consistent with the local-cue hypothesis, but cannot be causally linked to the additional labels from the category-only output.

This example is consistent with the hypothesis that the model insufficiently distinguishes central from peripheral actors. However, because the prompt requests labels without rationales, it does not establish that the model performs independent per-mention judgments. The improvement

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

Table 4: Distribution of articles across conflict periods and news outlets included in the analysis.

obtained with Prompt 1 indicates that explicitly directing the model toward dominant article-level framing reduces these errors in aggregate while leaving the underlying mechanism unobserved.

6.3 Category-Level Source Overprediction

The models frequently predict source categories that are absent from the gold annotations. The prompt explicitly excludes "entities that are only the subject of the story but do not speak or provide info", but the category-level output does not identify which article mention, if any, motivates an additional prediction.

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, while `THAI_MILITARY` has 15 gold instances and 26 predictions. Manual inspection of selected false positives identifies articles in which entities from the predicted category are discussed but do not provide quoted or paraphrased information. This pattern is consistent with subject–source conflation, but does not confirm that those entities caused the prediction.

Source attribution requires a distinction between entities that *produce* information and entities that are *discussed in* that information. This distinction is discourse-level and depends on the structural role of each mention in the article. A prompt that requests entity spans or supporting evidence in addition to category labels would be required to test the subject–source conflation hypothesis directly.

7 Analysis of Narrative Framing across News Outlets

We investigate whether narrative framing by Thai and Cambodian news outlets changes after the active conflict. After the period that constitutes the main timeframe of the annotated dataset (June to September 2025), 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 articles about the Thai–Cambodia conflict during the active-conflict period (June to September 2025) and the post-conflict period (December 2025 to April 2026) from the same set of Thai and Cambodian outlets. This corpus consists of 1,388 articles from the active-conflict period and 1,684 from the post-conflict period. Thai outlets contribute 1,770 articles and Cambodian outlets contribute 1,302 articles. We additionally collect 32 articles from BBC Thai as an internationally affiliated outlet. Other international outlets generally publish only one to three articles about the conflict.

These articles are processed only by the LLM pipeline presented in the previous section and used for exploratory comparative framing analysis. We use Gemini Pro with Prompt 1 and process Thai and Khmer articles in their original languages. This configuration obtains the strongest benchmark performance (micro-averaged $F_1 = 0.73$). We apply the same prompt and decoding settings to both periods without further modification. Because the benchmark performance remains imperfect, the resulting proportions contain non-trivial measurement error. We therefore interpret broad, repeated contrasts across outlet groups rather than individual predictions or small percentage differences.

7.1 Per-Outlet Framing Patterns

The predicted label distributions exhibit a "mirrored" pattern (Figure 1). Thai outlets cluster on `TH-HERO` and `KH-VILLAIN` framing: ThaiPBS, The Standard, Thairath, and Bangkok Post each devote approximately 25–33% of their predicted labels to `TH-HERO` and 20–24% to `KH-VILLAIN`, with Cambodian-target Hero and Victim labels largely absent. BBC Thai's predicted distribution resembles those of the Thai outlets. Cambodian outlets show the opposite directional pattern, with approximately 17–27% of predicted labels assigned to `KH-HERO` and 18–24% to `TH-VILLAIN`. These model-derived distributions suggest strongly nationalized and self-favoring narrative perspectives on both sides.

7.2 Active Conflict vs. Post-Conflict Comparison

The model-derived distributions do not show a significant softening of the "mirrored" pattern during the post-conflict period (Figure 2). A chi-square test of independence finds no significant differ-

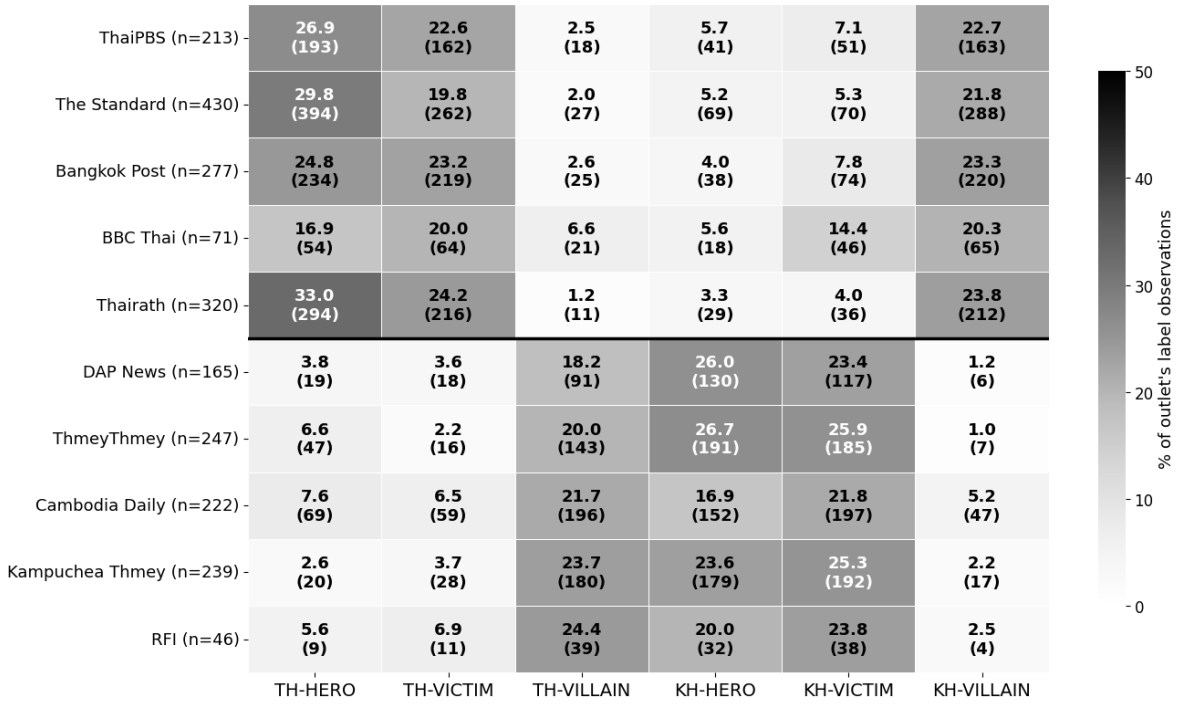


Figure 1: Gemini-Pro-predicted narrative-role label distribution per outlet using Prompt 1 and original-language input, normalized as a percentage of each outlet's total label observations. The horizontal line separates Thai and Cambodian outlets. TH = Thailand target, KH = Cambodia target.

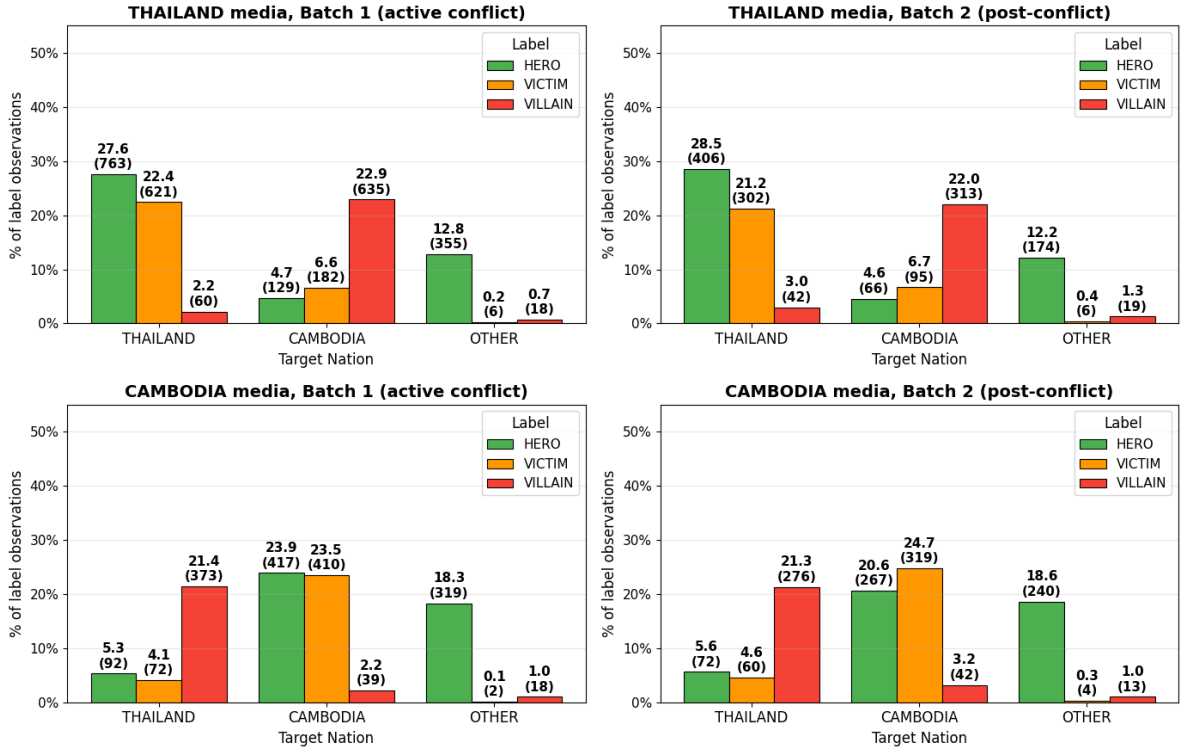


Figure 2: Gemini-Pro-predicted narrative-role label distribution by outlet country and period using Prompt 1 and original-language input. Top row: Thai outlets, active-conflict (June to September 2025) and post-conflict (December 2025 to April 2026). Bottom row: Cambodian outlets, same split. Bar values show percentage of label observations and raw count in parentheses.

ence in the overall predicted label distribution between the active-conflict and post-conflict periods for Thai media ($\chi^2(8) = 11.53, p = 0.173$). Similarly, we find no significant period difference for Cambodian outlets ($\chi^2(8) = 8.42, p = 0.393$). We initially hypothesized that mirrored framing would decrease after active conflict ended. Although the tests do not establish that the distributions are identical, the same directional pattern appears across both periods and outlet groups.

8 Conclusion

In this paper, we created a multilingual dataset for analyzing narrative framing in news coverage of the 2025 Thailand–Cambodia border conflict. The dataset covers English, Thai, and Khmer articles annotated with stakeholder-oriented Hero–Villain–Victim roles and source attribution labels. We benchmarked proprietary and open-weight zero-shot LLMs, compared original-language and translated input, evaluated salience-focused prompting strategies, and tested supervised cross-dataset transfer. Original-language inference and explicit article-level salience instructions improved Gemini Pro's micro-averaged F_1 from 0.66 to 0.73, while the transferred XLM-R baseline reached 0.50. Finally, we applied the best-performing configuration to examine framing across news outlets and found evidence of largely mirrored, self-favoring narratives in both active-conflict and post-conflict coverage.

9 Limitations

The manually annotated benchmark contains only 51 articles and 89 narrative-role assignments. It supports an initial evaluation of article-level framing, but not precise comparisons across every language, outlet, or rare label. In particular, `OTHER-VICTIM` and `OTHER-VILLAIN` have only two and one gold instances, respectively. Results for these categories should not be treated as stable estimates.

Khmer articles were initially machine-translated into English, whereas the Thai-source articles could be annotated in their original publication language. Native-Khmer reannotation produced moderate agreement and was incorporated through adjudication, but this validation cannot establish that every culturally or politically loaded cue was preserved. Translation with the same model family used in part of the evaluation may also introduce correlated model-specific biases.

We evaluate two salience-focused variants of the narrative-role prompt, but the error analysis that motivated these variants and their evaluation use the same 51-article benchmark. The improvement to 0.73 may therefore overestimate generalization to an independently held-out sample. In addition, the three-sentence salience criterion was introduced during human adjudication, whereas the model prompts express salience qualitatively rather than as a numerical threshold. The results characterize the reported prompts and annotation procedure rather than an invariant upper bound on LLM framing analysis. LLM outputs may also vary across model versions and inference environments despite deterministic decoding settings.

The analysis of 3,072 additional articles relies on automatic labels whose benchmark performance is imperfect. Its cross-outlet patterns should be treated as exploratory corpus-level trends, not as verified labels for every article or as evidence about the factual correctness or intentions of individual outlets. Finally, our source schema excludes civilian and non-institutional voices and therefore cannot measure whose civilian testimony is selected or omitted. Collection through outlet search engines may also omit relevant articles and limits claims of corpus exhaustiveness.

10 Ethical Considerations

This study analyzes public reporting about an ongoing, politically sensitive conflict. Hero, Villain, and Victim labels refer to narrative roles within articles, not judgments about factual responsibility or national populations. Automatic framing labels may reproduce translation errors, model biases, and nationalist assumptions. We therefore report aggregate patterns and release URLs, metadata, and annotations rather than copyrighted article text. The dataset and models should not be used to assess the loyalty or credibility of individual journalists or communities.

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, Mohammadmasiha Zahedivafa, Juan Diego Bermeo, Maria Korobeynikova, and Fabrizio Gilardi. 2023. Open-source LLMs for text annotation: A practical guide for model setting and fine-tuning. *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. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3570–3580, Brussels, Belgium. Association for Computational Linguistics.

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. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 996–1005, Vancouver, Canada. Association for Computational Linguistics.

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.

Tarek Mahmoud, Zhuohan Xie, Dimitar Iliyanov Dimitrov, Nikolaos Nikolaidis, Purificação Silvano, Roman Yangarber, Shivam Sharma, Elisa Sartori, Nicolas Stefanovitch, Giovanni Da San Martino, Jakub Piskorski, and Preslav Nakov. 2025. Entity framing and role portrayal in the news. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 302–326.

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. 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.

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 Detailed Per-Label Results

B 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}"
```

C Narrative-Role Prompt Variants

Prompt 1 adds the following rules to the original narrative-role prompt:

```
RULES:
1. SALIENCE: label a nation only if it is a
    central driver of the narrative.
2. FRAME, NOT MENTION: a label needs the article
    's dominant framing, not one cue.
3. PREFER FEWER, WELL-SUPPORTED LABELS. If
    unsure a role is central, omit it.
4. The same nation may take multiple labels only
     if each is independently central.
```

Prompt 2 includes the rules from Prompt 1 and replaces the role definitions with the following expanded definitions and tie-breaking instructions:

```
LABEL:
- HERO: portrayed favorably across MULTIPLE
    beats -- brokering/keeping peace,
  showing restraint or de-escalating, protecting
     civilians, or mounting a
  legitimate defense. A single self-serving
    quote is NOT enough; the favorable
  framing must be sustained. Sustained favorable
     framing counts as HERO even
  when the article is one-sided.
- VILLAIN: causing harm, escalating, violating
    agreements, or condemned -- as a
  recurring theme.
- VICTIM: the central party whose SUFFERING is
    foregrounded as a main narrative
  purpose (displacement, appeals for help, harm
    centered as the story's point).

HERO vs VICTIM TIEBREAKER:
A nation can be BOTH attacked AND actively
    responding, so both may apply.
- If the DOMINANT frame is its suffering, label
    VICTIM.
- If the DOMINANT frame is its active, praised
    response (defending, de-escalating,
  protecting others), label HERO.
- Assign BOTH only if each is independently
    sustained across the article.
Do not default to VICTIM whenever harm is
    mentioned; check whether the article's
weight is on the harm or on the response.
```

D XLM-R Cross-Dataset Transfer Results

E 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.
```

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 5: Per-label narrative-role labeling performance using the original prompt and English-translated input.

Category	Gemini Flash			Gemini Pro			GPT			Gold
	P	R	F1	P	R	F1	P	R	F1	
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 6: Per-label source attribution performance across models.

	Hero	Victim	Villain	Overall
EN–Union	0.381	0.400	0.458	0.416
EN–Argmax	0.457	0.357	0.506	0.451
Original–Union	0.412	0.487	0.487	0.461
Original–Argmax	0.507	0.357	0.588	0.500
Best LLM	0.643	0.737	0.800	**0.728**

Table 7: Role-level F_1 for XLM-R cross-dataset transfer and the best zero-shot LLM. EN denotes English-translated input.

```
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}"
```