Fane at SemEval-2025 Task 10:
Zero-Shot Entity Framing with Large Language Models

Enfa Fane[*]   **Mihai Surdeanu**[*]   **Eduardo Blanco**[*]   **Steven R. Corman**[†]

[*]University of Arizona   [†]Arizona State University

[*]{enfageorge, msurdeanu, eduardoblanco}@arizona.edu

[†]{steve.corman}@asu.edu

Abstract

Understanding how news narratives frame entities is crucial for studying media's impact on societal perceptions of events. In this paper, we evaluate the zero-shot capabilities of large language models (LLMs) in classifying framing roles. Through systematic experimentation, we assess the effects of input context, prompting strategies, and task decomposition. Our findings show that a hierarchical approach of first identifying broad roles and then fine-grained roles, outperforms single-step classification. We also demonstrate that optimal input contexts and prompts vary across task levels, highlighting the need for subtask-specific strategies. We achieve a Main Role Accuracy of 89.4% and an Exact Match Ratio of 34.5%, demonstrating the effectiveness of our approach. Our findings emphasize the importance of tailored prompt design and input context optimization for improving LLM performance in entity framing.

1 Introduction

News is produced and consumed at an unprecedented scale, yet the expectation of neutrality and objectivity in journalism (Schudson, 2001) often contrasts with the reality that media selectively determines what is newsworthy (Eliaz and Spiegler, 2024). Through inclusion, omission, or emphasis of specific details, journalists shape public interpretation of events (Iyengar, 1990), promoting particular recommendations (Entman, 1993), which in turn influences public opinion and policy decisions (Ziems and Yang, 2021).

When such selective framing is used to shape perceptions of an entity, it is known as entity framing (van den Berg et al., 2020). SemEval-2025 Task 10 introduces a benchmark for analyzing entity framing (Piskorski et al., 2025; Stefanovitch et al., 2025), examining how linguistic choices from word selection to narrative structure affect perception.

Example in Figure 1 demonstrates how strategic language choices portray the entity *Bill Gates* as

---

Gates *claimed* that because he continues to spend billions of dollars on climate change activism, his *carbon footprint isn't an issue* [. . .] Elsewhere during the *carefully constructed interview*, Gates said he was surprised that he was *targeted by 'conspiracy theorists' for pushing vaccines* during the pandemic [. . .] While the BBC interview *was set up* to look like Gates was being challenged or grilled, he wasn't asked about his *close friendship with the elite pedophile Jeffrey Epstein*

Figure 1: An example document excerpt from the dataset. The author strategically uses loaded language, contrastive framing, and selective emphasis to shape the reader's perception of Bill Gates as corrupt and deceptive. By highlighting contradictions in his public advocacy, casting doubt on the authenticity of his media appearances, and referencing controversial associations, the text reinforces a negative framing.

an Antagonist. Despite mentioning his substantial investments in climate change initiatives, the narrative emphasizes his high carbon footprint, highlighting hypocrisy. Phrases such as "carefully constructed interview" and references to his connections with Epstein further reinforce an image of deception and evading accountability. This selective emphasis shapes the reader's perception of Gates as a Deceiver and Corrupt figure, demonstrating how framing subtly influences interpretation.

Recent advances in large language models (LLMs) have demonstrated superior performance in text classification, frequently outperforming traditional machine learning approaches on complex tasks (Kostina et al., 2025). In this paper, we explore if this extends to entity framing. Relying solely on prompting, we ask whether LLMs accurately identify the framing roles assigned to entities in news articles. Given the sensitivity of LLM predictions to prompt variability (Zhuo et al., 2024), we assess how prompt engineering strategies such as role-based prompting, incorporating task related information such as label definitions, and rationale generation shape classification performance.

To this end, we develop a modular zero-shot approach that relies exclusively on prompting. Our method decomposes the task into two stages, first predicting broad narrative roles (Protagonist, Antagonist, Innocent), and then refining into fine-grained roles. We systematically vary the input context comparing full articles, entity-centered excerpts, and framing-preserving summaries and design prompt templates incorporating expert personas and task related information.

The main contributions of this paper are:

(1) *Systematic evaluation of LLMs for zero-shot entity framing:* We assess how well LLMs infer framing roles from implicit narrative cues without fine-tuning.

(2) *A multi-step prompting strategy:* We show that decomposing the task into two stages, predicting broad roles first, then fine-grained roles, yields better results than a single-step approach.

(3) *Strategic prompt design:* We demonstrate that effective prompt design, incorporating role-based guidance and task definitions, improves classification. A carefully engineered prompt enables a smaller model to perform comparably to a larger one.

We report a Main Role Accuracy of 89.4% and an Exact Match Ratio of 34.5% on the SemEval 2025 Task 10 Subtask 1 for English, placing sixth on among the participating teams. We make our codebase and associated prompts publicly available.[1]

2 Background: Task & Dataset

The SemEval 2025 Task 10: *Multilingual Characterization and Extraction of Narratives from Online News* introduces three subtasks for analyzing narrative elements in news articles across five languages: Bulgarian, English, Hindi, European Portuguese, and Russian. This paper focuses on *Subtask 1: Entity Framing* for *English*.

The goal of *Entity Framing* is to determine how a named entity is framed within a news article. Framing is represented at two levels: *Main Role* and *Fine-Grained Roles*. The main role broadly categorizes the entity as a Protagonist, Antagonist, or Innocent, while fine-grained roles provide more

detailed labels.[2] An entity may be framed with multiple fine-grained labels, making this a multi-class, multi-label classification problem.

For further details on the dataset, including domain coverage, corpus statistics, and annotation methodology, see (Piskorski et al., 2025; Stefanovitch et al., 2025). We now describe our approach to leveraging large language models (LLMs) for entity framing classification.

3 Related Work

Language shapes interpretation by selectively emphasizing aspects of reality (Entman, 1993). Prior research identifies two key linguistic mechanisms underlying framing effects: *naming conventions*, where respectful or formal titles correlate strongly with positive sentiment (van den Berg et al., 2019, 2020); and *narrative framing*, reflecting partisan differences in media portrayals of events (Ziems and Yang, 2021). However, these studies primarily treat framing as an implicit phenomenon correlating with other tasks, rather than explicitly classifying or predicting specific framing roles or labels.

Beyond text, in the multimodal domain, Sharma et al. (2022) introduce HVVMemes, a meme dataset covering COVID-19 and US politics, annotated with coarse-grained framing roles: *hero*, *villain*, *victim*, or *none*. In contrast, our work focuses on news text and classifies entities at a more fine-grained, hierarchical roles.

Large language models (LLMs) have recently achieved strong performance in text classification tasks, often surpassing traditional machine learning methods on complex datasets (Kostina et al., 2025). However, significant challenges remain, particularly concerning sensitivity to prompt phrasing and input structure (Zhuo et al., 2024). Structured prompting strategies, such as role conditioning (Kong et al., 2024) and instruction reframing (Mishra et al., 2022), have been proposed. Nevertheless, the effectiveness of these strategies remains highly task-dependent and often requires careful adaptation to specific problem settings (Atreja et al., 2024).

Guided by these insights, we develop and systematically experiment with multiple prompting strategies for hierarchical entity framing, aiming to improve both coarse and fine-grained entity framing classification.

---

[1] https://github.com/beingenfa/semeval2025task10

[2] The fine-grained labels are listed in Appendix A

```
{expert_phrasing}
You will be provided with {input_format_phrase}.
Your task is to {task_definition}.
{task_instructions}
{output_format_with_example}
{input_context}
```

Figure 2: Structure of the prompt template. Curly-bracketed content is dynamically replaced based on the experimental setting (see Appendix D for exact phrasing). The template is designed to minimize variability, ensuring that observed differences arise solely from targeted prompt modifications.

4 Prompting for Framing Classification

To systematically evaluate the impact of different prompting strategies on framing classification, we adopt a structured, template-based approach (Figure 2). LLM decision-making is often opaque. By using a fixed template, we ensure any variation in outputs comes from our controlled prompt modifications rather than randomness. The template consists of modular components that adapt to different experimental conditions.

Our approach is structured around three key components: *(1) Input Context* which determines the textual context provided for classification, specified in `input_context`. *(2) Prompt Design*, which examines prompting strategies, such as persona prompting, structured within `task_definition`, `task_instructions`, and `output_format_with_example`. *(3) Inference Strategy*, which defines whether classification is performed as a single task or decomposed into two steps, reflecting the hierarchical nature of framing. The following sections provide a detailed discussion of each component.

4.1 Input Context Variation

Framing involves selectively emphasizing certain details while omitting others to highlight specific aspects of perceived reality (Entman, 1993). Thus, optimizing context is crucial for maximizing narrative signals. To assess the impact of context granularity on framing classification, we define five input settings encompassing both extractive and summarization-based approaches,[3] each varying in the amount of contextual information available for classification.

- *Full Text* (FT): The entire article.

---

[3]See Appendix C for details on summarization.

- *Entity-Sentences* (Ent-Sent): Only sentences mentioning the entity.
- *Entity-Sentences Neighbors* (Ent-Neigh): Entity-mentioning sentences plus one preceding and one following sentence.
- *Neutral Summary* (Neutral-Sum): A summary generated by an LLM prompted to focus neutrally on the entity's involvement, actions, and framing.
- *Framing-Preserved Summary* (FP-Sum): A summary generated by an LLM prompted to preserve the article's original framing, emphasizing positive or negative actions.

4.2 Prompt Design

While various prompting strategies have been proposed (Gu et al., 2023), we focus on the following:

(1) *Role/Persona Prompting:* Assigning a role or persona in an LLM prompt may shape its performance. While one study finds that role-based prompting improves task performance (Kong et al., 2024), another suggests its effectiveness is highly task-dependent and varies across settings (Zheng et al., 2024). To assess its impact on entity framing, we compare a neutral prompt with one explicitly assigning an expert persona.

(2) *Task-Related Information:* We test two configurations: one providing only the task definition and another including both task and label definitions.

(3) *Output Rationale:* We compare predictions with and without model-generated explanations to assess whether requiring justifications improves or degrades classification accuracy.

4.3 Inference Strategy

We compare two inference strategies that determine whether classification is performed as a single task or decomposed into two stages, reflecting the hierarchical structure of framing. *Single-Step Prediction* jointly predicts both the main and fine-grained roles within a single inference step, whereas *Multi-Step Prediction* first infers the main role, which is then incorporated into the prompt to predict the fine-grained role.

Having established our prompting strategies, we now present the experimental setup used to systematically evaluate their effectiveness.

5 Experimental Setup

We first discuss a key assumption in our study, followed by evaluation metrics, and LLM setup.

Single-Label Approximation for Fine-Grained Role Classification   Although the SemEval task defines Entity Framing as a multi-label classification problem, we observe that entity mentions typically receive only one fine-grained role, averaging 1.08 fine-grained roles per mention in the English training split, a trend consistent across languages and main roles (Appendix B). Additionally, in our experiments, we observed that the model in our study consistently predicts two main roles even when allowed to predict only one. Given these patterns, we adopt a single-label approximation, assigning each mention a single fine-grained role.

Evaluation Metrics   We assess performance using two metrics: *Main Role Accuracy (MRA)*, the proportion of instances where the predicted main role matches the gold label, and *Exact Match Ratio (EMR)*, the proportion of instances where both the main and fine-grained roles match exactly, with no partial credit.

Model and API   We use GPT-4o (gpt-4o-2024-08-06) via the OpenAI API.[4]

We next examine the results obtained using different prompting strategies and input contexts across both inference setups.

6 Results

We present results from our prompting experiments on the development set, followed by the official SemEval results and post-SemEval analyses. Key findings are discussed in Section 7.

6.1 Development

We systematically assess the impact of individual prompt modifications across different input contexts in both single-step and multi-step settings, using the baseline prompt detailed in Appendix D. To ensure precise attribution of effects, we isolate prompt modifications rather than evaluating combined strategies. While one combination setting is included in Section 6.2, systematic evaluation of prompt strategy combinations is beyond the scope of this paper.

Metric	Baseline	+ EP	+ LD	+ RA
Main Role Accuracy				
FT	0.93	0.92	0.89	0.91
Ent-Sent	0.90	0.93	0.92	0.91
Ent-Neigh	0.92	0.93	0.92	0.93
Neutral-Sum	0.69	0.73	0.69	0.71
FP-Sum	**0.95**	**0.95**	0.93	0.93
Exact Match Ratio				
FT	0.29	**0.35**	0.30	0.29
Ent-Sent	0.32	0.33	**0.35**	0.31
Ent-Neigh	0.30	0.34	0.32	0.31
Neutral-Sum	0.22	0.21	0.21	0.24
FP-Sum	0.33	0.32	0.34	0.33

Table 1: Performance comparison of different input contexts and prompt engineering strategies when jointly prompting for main role and fine-grained roles. EP = Expert Persona, LD = Label Definitions, RA = Rationale. For main roles, Framing-Preserved summaries are the strongest input context, achieving the highest accuracy (0.95). For fine-grained roles, full text, and only sentences containing entity are strongest when used with a prompting strategy.

Single-Step Approach   Table 1 shows the performance of different input contexts and prompt engineering strategies in the Single Step Setup. Main role accuracy is highest with Framing-Preserved Summaries (0.95), where additional prompt strategies yield no further gains. In contrast, fine-grained role classification benefits from prompting strategies, with Full Text + Expert Persona and Entity-Only Sentences + Label Definitions achieving the highest Exact Match Ratio (0.35).

These findings suggest that a single input strategy may not optimally support both tasks, motivating further exploration of a multi-step approach, where main and fine-grained roles are predicted separately. Additionally, due to consistently lower performance, the Neutral Summary setting is excluded from subsequent experiments.

Multi-Step Approach   In the multi-step setting, the model first predicts the main role, which is then used to guide fine-grained classification.
Main Role Prediction: Table 2 reports main role accuracy when predicted independently. Framing-Preserved Summaries remain the strongest setting, achieving 0.96 accuracy.
Fine-grained roles: Main role predictions from the best-performing setup are used to inform the prompt for fine-grained role classification. Additionally, we restrict the set of possible output labels to only those fine-grained roles valid for the predicted main role. This reduces ambiguity and

Metric	Baseline	+ EP	+ LD	+ RA
FT	0.89	0.91	0.91	0.87
Ent-Sent	0.84	0.86	0.87	0.82
Ent-Neigh	0.91	0.92	0.91	0.88
FP-Sum	0.92	0.93	**0.96**	0.93

Table 2: Performance comparison of different input contexts and prompt engineering strategies in the Multi-Step Setup for Main Role Prediction, reported with accuracy. EP = Expert Persona, LD = Label Definitions, RA = Rationale. Framing-Preserved Summaries remain the strongest input context, achieving the highest accuracy (0.96), now benefiting from the Label Definitions (LD) strategy, unlike in the single-step setting.

Metric	Baseline	+ EP	+ LD	+ RA
FT	0.33	0.36	0.31	0.35
Ent-Sent	0.36	**0.44**	0.35	0.37
Ent-Neigh	0.36	0.38	0.34	0.37
FP-Sum	0.29	0.29	0.35	0.31

Table 3: Performance comparison of different input contexts and prompt engineering strategies in the Multi-Step Setup for fine-grained role classification using Exact Match Ratio metric. EP = Expert Persona, LD = Label Definitions, RA = Rationale. The multi-step approach improves EMR across most settings compared to single-step prediction, with Entity-Sentences + Expert Persona (EP) achieving the highest EMR (0.44).

Rank	Team	EMR (Δ)	MRA (Δ)
1	DUTIR	0.41	0.95
2	PATeam	0.38 (-0.03)	0.89 (-0.06)
3	DEMON	0.37 (-0.04)	0.92 (-0.03)
4	gowithnlp	0.37 (-0.04)	0.94 (-0.01)
5	TartanTritons	0.36 (-0.05)	0.72 (-0.23)
6	**Ours**	**0.34 (-0.07)**	**0.89** (-0.06)
27	Baseline	0.04 (-0.37)	0.29 (-0.66)

Table 4: Official SemEval test set results. Teams are ranked by Exact Match Ratio (EMR), with Main Role Accuracy (MRA) also reported. Our system ranked 6th, achieving an EMR of 0.34, just 0.07 behind the top system (DUTIR). Δ values indicate the difference from the top-ranked system.

Approach	EMR	MRA	Model	Price/1M Tokens	
				Input	Output
SemEval	0.345	0.894	O1	$15.00	$60.00
Improved	**0.349**	**0.894**	GPT-4o	$2.50	$10.00

Table 5: Comparison of our official SemEval submission and the refined approach. Although the performance is nearly identical, the refined approach is notable because it achieves these results using GPT-4o, a significantly smaller and more cost-effective model. This highlights the importance of optimized prompt design, which allows a more compact model to match or exceed the performance of larger, more expensive alternatives.

improves classification accuracy.

As shown in Table 3, the multi-step approach outperforms single-step prediction across most settings. The best-performing strategy is Entity-Sentences + Expert Persona (EP), which achieves an Exact Match Ratio (EMR) of 0.44, a substantial improvement over the best single-step result (0.35). Entity-Sentences remains the strongest input context for fine-grained roles(0.44 EMR), whereas full text falls behind.

These findings further support the effectiveness of separating main and fine-grained role prediction and constraining valid role labels to enhance classification accuracy.

6.2 Official SemEval Results

The findings from our development experiments highlight the effectiveness of different input contexts and prompting strategies. However, for our official SemEval submission, we employed a Full-Text + Expert Persona + Multi-Step setup using O1(o1-2024-12-17) as our model, as this configuration was selected based on our pre-competition experiments.

As shown in Table 4, our system ranked 6th overall, achieving an Exact Match Ratio (EMR) of 0.34, placing 0.07 behind the top-performing system (DUTIR). While competitive, our post-SemEval ablation studies revealed a more effective strategy, which we discuss in the next section.

6.3 Post-SemEval

After the SemEval submission, we conducted structured ablation studies to refine our approach, as discussed in Section 6.1. As shown in Table 5, the new approach achieves an EMR of 0.349, compared to 0.345 in our official submission, with MRA remaining unchanged at 0.894. This result is significant because it was achieved using GPT-4o, a smaller and significantly cheaper model.

7 Insights

We present key findings from our experiments, examining the impact of different approaches.

Framing and emphasis: selective highlighting shapes interpretation As shown in Table 1, neutral summaries which present entities in a factual, impartial manner, consistently underperformed

compared to framing-preserved summaries. This aligns with existing framing theory, reinforcing that framing is not only about fact selection but also about selective emphasis to shape interpretation (Entman, 1993).

Justifications do not improve classification performance   Our experiments show that requiring models to justify their predictions does not improve classification accuracy. Performance is primarily determined by the information in the prompt.

More information isn't always better   Longer contexts aren't always useful. Main role classification benefits from broad context that establishes overarching narratives, whereas fine-grained roles require focused, entity-specific details. Excessive input, such as full-text context, can dilute key framing signals, whereas condensed, framing-preserved summaries improve accuracy. This highlights the importance of tailoring context granularity to the specific classification task.

Multi-Step Approach Improves Performance We find that a structured, multi-step classification approach substantially improves performance. By first predicting the main role and then refining fine-grained classification based on that prediction, the model benefits from clearer context at each stage. This approach reduces ambiguity and ensures that each classification step is optimized for its specific level of framing.

Prompt Design Enables Small Models to Rival Larger Ones   Finally, our experiments highlight the effectiveness of carefully engineered prompts. Carefully crafted prompts allow smaller models to rival larger ones. A well-optimized prompt for GPT-4o matches or exceeds a less-tuned prompt on the larger, costlier o1 model. These results suggest that before scaling to larger architectures, improving prompt design for smaller models can yield substantial gains in both efficiency and accuracy.

8 Conclusion

In this paper, we explore the effectiveness of large language models (LLMs) for Entity Framing Classification in a zero-shot setting. While LLMs perform well in broad role classification, fine-grained classification remains challenging. Our multi-step approach, incorporating distinct input contexts and prompt strategies at each stage, significantly improves overall performance.

References

Shubham Atreja, Joshua Ashkinaze, Lingyao Li, Julia Mendelsohn, and Libby Hemphill. 2024. Prompt design matters for computational social science tasks but in unpredictable ways. *arXiv preprint arXiv:2406.11980*.

Kfir Eliaz and Ran Spiegler. 2024. News media as suppliers of narratives (and information). *Preprint*, arXiv:2403.09155.

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

Jindong Gu, Zhen Han, Shuo Chen, Ahmad Beirami, Bailan He, Gengyuan Zhang, Ruotong Liao, Yao Qin, Volker Tresp, and Philip Torr. 2023. A systematic survey of prompt engineering on vision-language foundation models. *arXiv preprint arXiv:2307.12980*.

Shanto Iyengar. 1990. Framing responsibility for political issues: The case of poverty. *Political behavior*, 12:19–40.

Aobo Kong, Shiwan Zhao, Hao Chen, Qicheng Li, Yong Qin, Ruiqi Sun, Xin Zhou, Enzhi Wang, and Xiaohang Dong. 2024. Better zero-shot reasoning with role-play prompting. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 4099–4113, Mexico City, Mexico. Association for Computational Linguistics.

Arina Kostina, Marios D. Dikaiakos, Dimosthenis Stefanidis, and George Pallis. 2025. Large language models for text classification: Case study and comprehensive review. *Preprint*, arXiv:2501.08457.

Swaroop Mishra, Daniel Khashabi, Chitta Baral, Yejin Choi, and Hannaneh Hajishirzi. 2022. Reframing instructional prompts to GPTk's language. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 589–612, Dublin, Ireland. Association for Computational Linguistics.

Jakub Piskorski, Tarek Mahmoud, Nikolaos Nikolaidis, Ricardo Campos, Alípio Jorge, Dimitar Dimitrov, Purificação Silvano, Roman Yangarber, Shivam Sharma, Tanmoy Chakraborty, Nuno Guimarães, Elisa Sartori, Nicolas Stefanovitch, Zhuohan Xie, Preslav Nakov, and Giovanni Da San Martino. 2025. SemEval-2025 task 10: Multilingual characterization and extraction of narratives from online news. In *Proceedings of the 19th International Workshop on Semantic Evaluation*, SemEval 2025, Vienna, Austria.

Michael Schudson. 2001. The objectivity norm in american journalism. *Journalism*, 2(2):149–170.

Shivam Sharma, Tharun Suresh, Atharva Kulkarni, Himanshi Mathur, Preslav Nakov, Md. Shad Akhtar, and Tanmoy Chakraborty. 2022. Findings of the CONSTRAINT 2022 shared task on detecting the hero, the villain, and the victim in memes. In *Proceedings of the Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situations*, pages 1–11, Dublin, Ireland. Association for Computational Linguistics.

Nicolas Stefanovitch, Tarek Mahmoud, Nikolaos Nikolaidis, Jorge Alípio, Ricardo Campos, Dimitar Dimitrov, Purificação Silvano, Shivam Sharma, Roman Yangarber, Nuno Guimarães, Elisa Sartori, Ana Filipa Pacheco, Cecília Ortiz, Cláudia Couto, Glória Reis de Oliveira, Ari Gonçalves, Ivan Koychev, Ivo Moravski, Nicolo Faggiani, and 19 others. 2025. Multilingual Characterization and Extraction of Narratives from Online News: Annotation Guidelines. Technical Report JRC141322, European Commission Joint Research Centre, Ispra (Italy).

Esther van den Berg, Katharina Korfhage, Josef Ruppenhofer, Michael Wiegand, and Katja Markert. 2019. Not my president: How names and titles frame political figures. In *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science*, pages 1–6, Minneapolis, Minnesota. Association for Computational Linguistics.

Esther van den Berg, Katharina Korfhage, Josef Ruppenhofer, Michael Wiegand, and Katja Markert. 2020. Doctor who? framing through names and titles in German. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4924–4932, Marseille, France. European Language Resources Association.

Mingqian Zheng, Jiaxin Pei, Lajanugen Logeswaran, Moontae Lee, and David Jurgens. 2024. When "a helpful assistant" is not really helpful: Personas in system prompts do not improve performances of large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 15126–15154, Miami, Florida, USA. Association for Computational Linguistics.

Jingming Zhuo, Songyang Zhang, Xinyu Fang, Haodong Duan, Dahua Lin, and Kai Chen. 2024. ProSA: Assessing and understanding the prompt sensitivity of LLMs. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 1950–1976, Miami, Florida, USA. Association for Computational Linguistics.

Caleb Ziems and Diyi Yang. 2021. To protect and to serve? analyzing entity-centric framing of police violence. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 957–976, Punta Cana, Dominican Republic. Association for Computational Linguistics.

A Entity Framing Labels

The entity framing taxonomy consists of three main roles and 22 fine-grained roles. For detailed definitions and annotation guidelines, see (Piskorski et al., 2025; Stefanovitch et al., 2025).

- **Protagonist:** Guardian, Martyr, Peacemaker, Rebel, Underdog, Virtuous.

- **Antagonist:** Instigator, Conspirator, Tyrant, Foreign Adversary, Traitor, Spy, Saboteur, Corrupt, Incompetent, Terrorist, Deceiver, Bigot.

- **Innocent:** Forgotten, Exploited, Victim, Scapegoat.

B Average Sub Roles across Language

	EN	BG	PT	HI	RU
All	1.08	1.13	1.05	1.16	1.06
Protagonist	1.06	1.02	1.01	1.24	1.00
Antagonist	1.10	1.19	1.09	1.12	1.11
Innocent	1.02	1.01	1.00	1.06	1.01

Table 6: The average number of sub-roles per instance across different languages (English (EN), Bulgarian(BG), European Portuguese(PT), Hindi(HI), and Russian(RU)) and main roles in the training split. The values remain close to one, justifying our single-label approximation for fine-grained classification.

C Prompts for Summary Generation

C.1 Neutral Summary

```
Summarize the following article with a specific focus on { entity }. Write the summary as a standalone
description, ensuring that the entity and its role are clearly introduced without referring to 'the
article' or assuming prior context. Clearly state their involvement, actions, and framing within the
event. Maintain a factual and neutral tone.
```

Figure 3: Prompt used to generate neutral summaries. The instructions guide the LLM to ensure that that the entity's role is explicitly introduced while maintaining a factual and impartial tone.

C.2 Framing-Preserved Summary

```
Write a standalone summary that clearly reflects how the author of the article frames entity and their
actions. Do not present the entity neutrally—mirror the language and implicit bias of the article
itself. If the author portrays the entity favorably, highlight their positive actions, successes, and
beneficial impact. If the author is critical, emphasize the entity's negative actions, failures, or
harmful consequences.
If the framing is mixed or subtle, encode the contrast in tone and nuance.Strongly reflect the author's
framing through:
- Loaded or emotionally charged language (if present in the article)
- Emphasis on the entity's perceived intentions and motivations
- Who is affected by their actions and how the consequences are framed
- Any direct or implied judgments made by the author.
Maintain the style and tone of the article, ensuring the framing is explicit in how the entity's role
and impact are described. Do not add external information or neutralize the bias. The summary should
feel as though it was written by the original author.
```

Table 7: Prompt used to generate framing-preserved summaries. The instructions guide the model to reflect the article's original framing, emphasizing bias and tone while avoiding neutrality or external information.

D Prompt Template Details

Our system prompt follows the structured format introduced earlier. The input context is provided in the user prompt, while the rest remains part of the system prompt.

Prompt Type	Template

System Prompt	`{expert_phrasing}`
	 You will be provided with `{input_format_phrase}`.
	 Your task is to `{task_definition}`.
	 `{task_instructions}`
	 `{output_format_with_example}`

User Prompt	`{input_context}`

Table 8: Structure of the system and user prompts. The system prompt provides instructions, while the user prompt supplies contextual input for the model.

We outline the values these elements take in different settings below.

D.1 Input Format Phrase & Input Context

The input context is formatted as `DOCUMENT:{content}` Note that in the `Entity-Sentences` and `Entity-Sentences Neighbors` settings, the individual sentences or sentence groups are separated by `[...]`.

In the system prompt, the input format phrase specifies the type of input context provided to the model. The format follows this structure:`a {document_type_str} in the following format-`

DOCUMENT:{document_type_str}. The corresponding document_type_str for each input context type is provided in the table below:

Input Context Type	document_type_str
Full Text	news article
Entity-Sentences or Entity-Sentences Neighbors	excerpt of a news article
Neutral Summary or Framing-Preserved Summary	news article summary

Table 9: Mapping between input context types and corresponding input phrases.

D.2 Prompt Design

D.2.1 Expert Persona

In the context of role/persona prompting, if no specific persona is assigned, {expert_phrasing} remains empty. Otherwise, it is replaced with the following text:

> You are an expert in analyzing how a specific named entity is portrayed in a given text. Read the text carefully and focus on everything said about {entity}.

D.2.2 Output Rationale

In the template below, we incorporate prompting for justification by including three variables:

- {ask_reasoning}: "with a reasoning for your prediction"

- {reasoning_in_json}: ', "reasoning" : "your reasoning here"

- {reasoning_example}: , "reasoning" : "The article frames Kremlin propagandists as instigators and deceivers, highlighting their role in spreading falsehoods and promoting extreme measures."

When the setting does not require justification with the output label, these variables remain empty.

D.2.3 Task Related Information

Setting	task_definition	task_instructions	output_format_with_example
Single Step	classify the narrative framing of the {entity} in the document based on the taxonomy that follows in the format [(broad role,[list of valid fine grained roles/])]. The taxonomy is {taxonomy}.	Instructions: 1. Assign exactly one broad role from: Protagonist, Antagonist, or Innocent. 2. Determine one or a maximum of two corresponding fine grained role from the taxonomy. 3. Order the sub-roles by likelihood, with the most likely fine-grained role listed first.{label_definitions}	4. Finally, return your conclusion {ask_reasoning} as a single JSON object with no extra text, in this format: { "main_role": "<most_likely_main_role>", "fine_grained_roles": ["<Most likely sub-role>", "<Second most likely sub-role if relevant>"] {reasoning_in_json}} Example Output: {"main_role": "Antagonist", "fine_grained_roles": ["Conspirator"] {reasoning_example}
Multi-Step : Main Role	classify the narrative framing of the entity in the document as either Protagonist, Antagonist or Innocent.	Instructions: 1. Assign exactly one main role from: Protagonist, Antagonist, or Innocent.{label_definitions}	2.Return your conclusion {ask_reasoning} as a single JSON object with no extra text, in this JSON format: { "main_role": "<most_likely_main_role>" {reasoning_in_json}} Example Output: {"main_role": "Antagonist" {reasoning_example}}
Multi-Step : Fine-grained Role	classify the narrative framing of the {entity} in the document. The taxonomy is {taxonomy}. {label_definitions}	You have previously identified the broader narrative frame to be main_role_candidate. Instructions: 1. Determine one or a maximum of two corresponding fine grained role from the taxonomy. 2. Order the sub-roles by likelihood, with the most likely fine-grained role listed first.	3. Finally, return your conclusion {ask_reasoning} as a single JSON object with no extra text, in this format: { "fine_grained_roles": ["<Most likely sub-role>", "<Second most likely sub-role if relevant>"] {reasoning_in_json}} Example Output: {"fine_grained_roles": ["Conspirator"] {reasoning_example} }

Table 10: Mapping between different settings, task definitions, and corresponding instructions for narrative framing classification.

Taxonomy The taxonomy shared in the prompt is the same as in Appendix A. The label definitions provided are from the official paper and is shared in Table 11.

1933

Role Level	{label_definitions}
Main Role	Protagonist: The central figure or party in a news article, often portrayed as a hero or positive force driving the narrative. Antagonist: The opposing figure or force in a news article, often depicted as the source of conflict or challenge to the protagonist. Innocent: An individual or group portrayed as untainted or blameless in the context of the news, typically victimized or wronged.
Fine Grained Roles Protagonist	Guardians: Heroes or guardians who protect values or communities, ensuring safety and upholding justice. They often take on roles such as law enforcement officers, soldiers, or community leaders (e.g., climate change advocacy community leaders). Martyr: Martyrs or saviors who sacrifice their well being, or even their lives, for a greater good or cause. These individuals are often celebrated for their selflessness and dedication. This is mostly in politics, not in Climate Change. Peacemaker: Individuals who advocate for harmony, working tirelessly to resolve conflicts and bring about peace. They often engage in diplomacy, negotiations, and mediation. This is mostly in politics, not in Climate Change. Rebel: Rebels, revolutionaries, or freedom fighters who challenge the status quo and fight for significant change or liberation from oppression. They are often seen as champions of justice and freedom. Underdog: Entities who are considered unlikely to succeed due to their disadvantaged position but strive against greater forces and obstacles. Their stories often inspire others. Virtuous: Individuals portrayed as virtuous, righteous, or noble, who are seen as fair, just, and upholding high moral standards. They are often role models and figures of integrity.
Fine Grained Roles Antagonist	Conspirator: Those involved in plots and secret plans, often working behind the scenes to undermine or deceive others. They engage in covert activities to achieve their goals. Instigator: Individuals or groups initiating conflict, often seen as the primary cause of tension and discord. They may provoke violence or unrest. Deceiver: Deceivers, manipulators, or propagandists who twist the truth, spread misinformation, and manipulate public perception for their own benefit. They undermine trust and truth. Incompetent: Entities causing harm through ignorance, lack of skill, or incompetence. This includes people committing foolish acts or making poor decisions due to lack of understanding or expertise. Their actions, often unintentional, result in significant negative consequences. Corrupt: Individuals or entities that engage in unethical or illegal activities for personal gain, prioritizing profit or power over ethics. This includes corrupt politicians, business leaders, and officials. Tyrant: Tyrants and corrupt officials who abuse their power, ruling unjustly and oppressing those under their control. They are often characterized by their authoritarian rule and exploitation. Foreign Adversary: Entities from other nations or regions creating geopolitical tension and acting against the interests of another country. They are often depicted as threats to national security. This is mostly in politics, not in Climate Change. Terrorist: Terrorists, mercenaries, insurgents, fanatics, or extremists engaging in violence and terror to further ideological ends, often targeting civilians. They are viewed as significant threats to peace and security. This is mostly in politics, not in Climate Change. Bigot: Individuals accused of hostility or discrimination against specific groups. This includes entities committing acts falling under racism, sexism, homophobia, Antisemitism, Islamophobia, or any kind of hate speech. This is mostly in politics, not in Climate Change. Saboteur: Saboteurs who deliberately damage or obstruct systems, processes, or organizations to cause disruption or failure. They aim to weaken or destroy targets from within. Traitor: Individuals who betray a cause or country, often seen as disloyal and treacherous. Their actions are viewed as a significant breach of trust. This is mostly in politics, not in Climate Change. Spy: Spies or double agents accused of espionage, gathering and transmitting sensitive information to a rival or enemy. They operate in secrecy and deception. This is mostly in politics, not in Climate Change.
Fine Grained Roles Innocent	Victim: People cast as victims due to circumstances beyond their control, specifically in two categories: (1) victims of physical harm, including natural disasters, acts of war, terrorism, mugging, physical assault, etc., and (2) victims of economic harm, such as sanctions, blockades, and boycotts. Their experiences evoke sympathy and calls for justice, focusing on either physical or economic suffering. Scapegoat: Entities blamed unjustly for problems or failures, often to divert attention from the real causes or culprits. They are made to bear the brunt of criticism and punishment without just cause. Exploited: Individuals or groups used for others' gain, often without their consent and with significant detriment to their wellbeing. They are often victims of labor exploitation, trafficking, or economic manipulation. Forgotten: Marginalized or overlooked groups who are often ignored by society and do not receive the attention or support they need. This includes refugees, who face systemic neglect and exclusion.

Table 11: Label Definitions shared in the prompt. These are from the task definitions (Piskorski et al., 2025).