Cross-Lingual Human Interest Framing: Analysing Climate Change Reporting in English and Spanish across Eight Countries

Anonymous ACL submission

Abstract

Human interest (HI) framing is a journalistic strategy that humanises a news story by adopting an individual perspective and emotional angle to present specific issues or events. Here, we advance computational methods for the study of HI framing in two ways: First, we present the first *cross-lingual* investigation of HI framing around the topic of climate change, in a study that covers four English-speaking (Australia, India, the UK, and the US) and four Spanish-speaking countries (Argentina, Colombia, Mexico, and Spain). Second, we propose a novel framework inspired by communication studies research to reliably identify HI elements in news articles, and use it to create a human-labelled gold standard data set with English and Spanish native speakers. This work offers insights into cross-cultural variations in climate change reporting and enhances our understanding of how personal narratives are integrated into climate discourse. For instance, we observe that countries with the same language generally show similar trends, and that climate-related HI news stories evoke a more negative sentiment in Spanish-speaking countries than in English-speaking countries.[1]

1 Introduction

Human Interest (HI) framing is a powerful reporting strategy that presents a story through a human angle. Research in communication studies has investigated the concept of HI framing and its effects. On the one hand, it increases audience engagement by highlighting personal narratives of struggle or success (Cho and Gower, 2006; Zhao et al., 2019; Figenschou et al., 2023), and on the other hand, it influences political understanding by making complex issues more accessible through individual stories (Jebril et al., 2013), a phenomenon particularly relevant in the context of news media.

[1]All code and data will be released upon acceptance.

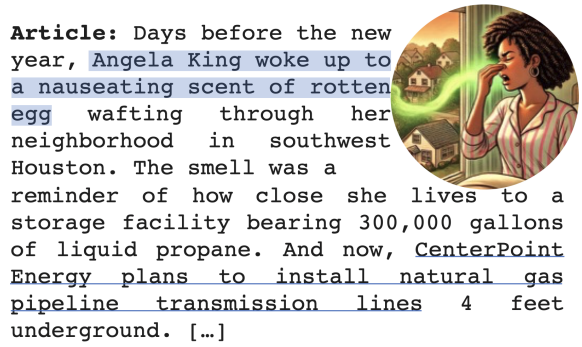

Figure 1: Example of an article containing Human Interest (HI) framing. The underlined text shows the topic of the article, while the highlighted text depicts the elements of the human interest story.

The impact of HI framing may be amplified in digital environments, where such stories are more likely to be shared on social media, provoking stronger emotional responses and fostering personal connections (Valenzuela et al., 2017). Despite its demonstrated societal impact, there is a lack of scalable methods and comprehensive computational analyses of its prevalence and distribution, particularly across different contexts. Existing research has been limited by the use of small datasets (Frermann et al., 2023) or monolingual settings (Vallejo et al., 2025). The present paper proposes an easy-to-use framework that enables researchers to identify HI framing by answering a structured set of questions. We apply this framework in a case study of articles about climate change in English and Spanish news, collected from eight countries across the global north and south.[2] This research yields three major contributions: (1) *a user- and model-friendly framework* which enables a more precise identification of HI framing in news articles, addressing challenges inherent to the ambiguity and lack of specificity; (2)

[2]EN:{Australia, India, UK, US}, ES: {Argentina, Colombia, Mexico, Spain}.

cross-lingual scalability, where our framework is applicable not only to larger datasets, but it is also effective across diverse cultural contexts, which helps mitigating the potential mismatch of concepts across cultures; and (3) a *HI annotated dataset*—we release a cross-lingual dataset with both manually and automatically annotated labels.

2 Related Work

A large body of work within social science investigates HI framing (Aalberg and Beyer, 2015; Boukes et al., 2015; Figenschou and Thorbjørnsrud, 2015; Figenschou et al., 2023). A seminal contribution is the work of Semetko and Valkenburg (2000), who introduced a five-frame typology to categorise audience frames: responsibility, conflict, economic consequences, morality, and human interest. Their work includes a detailed framework with specific sub-questions aimed at identifying these frames within media content and discourse, which we build on in this work.

In NLP, few studies have addressed HI framing. Frermann et al. (2023) made use of the sub-questions released by Semetko and Valkenburg (2000) and released a corpus of 428 English news articles annotated with five frames, including HI. While their work represents the first effort in NLP investigating the HI frame and their findings suggest HI stories are more common in right-leaning outlets, these conclusions rely on a modest sample size. Additionally, their focus on English-language articles from the US restricts the applicability of these findings to a more diverse global context.

Vallejo et al. (2025) investigated HI framing in climate change news articles across several English-language countries: Australia, India, the UK, and the US. Their findings reveal cultural differences in how climate change is reported, and demonstrate that features like topics, pronoun presence, sentiment, and entity analysis improve automated HI framing classification. They increased the sample size but simplified the methodology down to one direct question (*Is HI framing present in this article?*), and focused only on English. In this paper, we present a more rigorous framework and test its generalisability to different languages.

Inspired by the findings of Semetko and Valkenburg (2000) and Frermann et al. (2023) that splitting up the concept into several sub-questions leads to more reliable annotations, we extend the approach of Vallejo et al. (2025) to segment the concept of HI into four sub-questions. This is particularly crucial in our scenario where we work with annotators from different cultures/language backgrounds.

Q0	Can you identify the article's main issue/problem?
Q1	With respect to the main issue from Q0, does the article provide a human example or 'human face' for that issue?
Q2	With respect to the main issue from Q0, does the go into the private or personal lives of the actors?
Q3	With respect to the main issue from Q0, does the article employ adjectives or personal vignettes that generate feelings of outrage, empathy-caring, sympathy or compassion?
Q4	With respect to the main issue from Q0, does the article emphasise how individuals and groups are affected by that issue/problem?

Table 1: List of sub-questions adapted from Semetko and Valkenburg (2000).

Prior work by Alonso del Barrio and Gatica-Perez (2023) collected manual annotations for 1786 headlines from newspapers in five European countries (France, Italy, Spain, Switzerland, and the UK) and discovered that the HI frame is the most prominent frame in anti-vaccine articles. While this study includes several languages and countries, our research is the first study to conduct a cross-cultural analysis on full-length articles from countries spanning the global north and south. HI framing has been under-explored in NLP with little to no studies in a cross-lingual setting. We address this with a conceptually simple framework to determine the presence of HI framing and apply it to eight countries with English or Spanish as an official language.

In NLP, some work has focused on framing in climate change (Reimerink et al., 2024) and role narrative analysis (Frermann et al., 2023; Stammbach et al., 2024; Sinelnik and Hovy, 2024; Otmakhova and Frermann, 2025). The present paper combines both and focuses on the under-explored topic of HI framing.

3 HI Framing Framework

To systematically identify and analyse HI framing in news articles, we developed a framework based on established framing research in the communication sciences (Semetko and Valkenburg, 2000) and adapted it to our specific context. This framework consists of: (1) using sub-questions (detailed in Table 1) to identify HI framing signals; and (2) de-

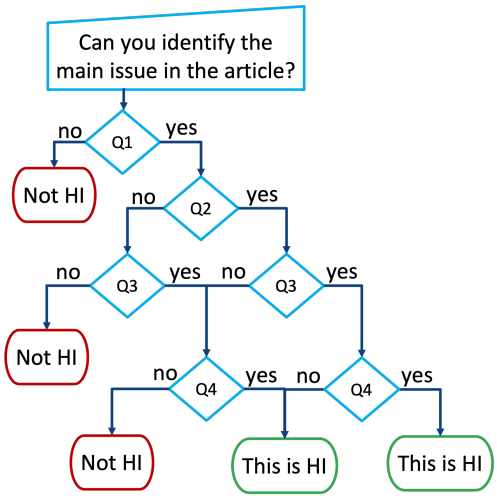

Figure 2: Our proposed framework, inspired by Semetko and Valkenburg (2000). Given an article, annotators answered a set of questions (Table 1) which were then aggregated into a binary label of the presence of HI framing. The same set of questions was used to prompt LLMs for automatic labelling.

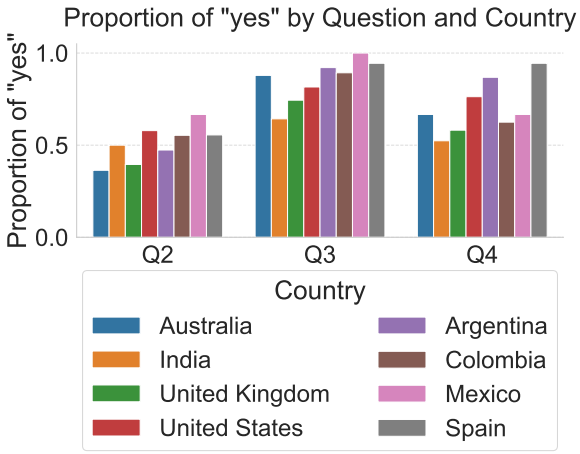

Figure 3: Proportions of sub-questions answered with 'yes', given that $Q1$ was answered with 'yes', in the manually annotated data.

veloping a function that takes as input the answers to those sub-questions, and outputs a binary label indicating whether HI framing is present (depicted as a flowchart in Figure 2).

Building upon the work of Semetko and Valkenburg (2000), our framework adapts the proposed sub-questions by adding a preliminary sub-question ($Q0$) to identify the main issue in the article. Each subsequent sub-question is answered with respect to this identified main issue. This provides focus on HI elements that are directly related to the main issue of the news article, and avoids the flagging of minor, incidental details that do not contribute to the framing of the main issue.

We collect annotations for each sub-question, and then aggregate them into a single label, as illustrated in Figure 2. Specifically, we consider an article to contain a HI story if three out of four questions are answered positively, with the additional requirement that $Q1$ must always receive a positive response. We consider this criterion essential, as the presence of a human example is fundamental to HI framing. For example, an article about climate change that focuses on how it affects a specific family would likely meet these criteria, while one that solely discusses scientific data would not (even though it could include emotional language [= $Q3$] or describe personal details without focusing on the human angle [= $Q2$]).

A key advantage of this framework is its *flexibility*, allowing researchers to tailor the annotation process to their specific definitions of HI framing. In our study, we require three out of four sub-questions to be answered positively, and this set of questions must include $Q1$. This threshold keeps a balance between capturing the nuances of HI framing and avoiding overly stringent or lenient criteria and ensures that the majority of the key aspects of HI framing are present.

To illustrate the interaction of different questions, Figure 3 shows the proportions of sub-questions being answered positive *given that* $Q1$ was answered with 'yes' (i.e., all bars for Q1 would be at 1.0) in our manually annotated data (Section 4.2). First we observe that none of questions $Q2 - Q4$ consistently 'saturates' to 1.0, in other words they all carry information on top of $Q1$. $Q2$ (private details of individuals) is the least frequently attested. By carrying out similar analyses on their own data set, researchers can empirically adjust their decision rules.

Moreover, our framework offers *scalability* across different languages and cultures, which is particularly valuable given that each culture may have its own preconceived notion of HI framing. By dissecting the complex concept of HI framing into more tangible, modular properties such as private details ($Q2$) or emotional language ($Q3$), rather than probing for the presence of HI directly, this approach helps to standardise the annotation process across diverse cultural contexts. However, it is worth noting that even with these standardised questions, some degree of cultural interpretation may still occur in the annotation process, e.g., the

Country	# Articles	# HI Articles	HI Proportion
Australia	449	79	0.18
India	1,325	87	0.07
UK	618	97	0.16
US	1,785	512	0.29
Total	4,177	775	0.19
Argentina	1,129	191	0.17
Colombia	1,660	289	0.17
Mexico	1,224	133	0.11
Spain	561	147	0.26
Total	4,574	760	0.17

Table 2: Number of articles collected and automatic labelled in our dataset.

Question	Overlap		Krippendorff's α	
	EN	ES	EN	ES
Q1	0.72	0.74	0.45	0.45
Q2	0.82	0.91	0.56	0.76
Q3	0.68	0.65	0.36	0.23
Q4	0.58	0.70	0.21	0.38
HI Label	0.74	0.79	0.43	0.51

Table 3: Inter-annotator agreement (IAA) for individual questions and the aggregated label for English (EN) and Spanish (ES). We report the absolute overlap among annotators and Krippendorff's alpha. HI Label refers to the final label of HI presence in the article aggregated from the answers to Q1–Q4 (Figure 2).

concept of emotional language might differ across annotators.

4 Dataset Construction

Our data collection pipeline consists of: (1) collecting a large corpus of climate change-related news from the eight countries in our analysis (Section 4.1); (2) manually annotating a sample of 200 articles (100 per language; Sections 4.2 and 4.3), based on the proposed framework (Figure 2); (3) automatically labelling the rest of the obtained data (Section 4.4); and (4) validating the automatically-labelled data (Section 4.5).

4.1 Data Gathering

We collected 21.7k climate change articles articles in English and 18.4k articles in Spanish articles from ProQuest with a keyword search following Vallejo et al. (2025).[3] When sampling, we ensured an even representation of left- and right-leaning outlets for each language (see Figure 6 in Appendix A). After removing duplicates, we filtered the articles to include only those with lengths between 100 and $2,000$ tokens, to ensure sufficient content for analysis while avoiding overly long documents. We then prompted DeepSeek R1 to identify articles with a primary focus on climate change.[4]

The resulting dataset consists of $4,177$ news articles in English and $4,574$ in Spanish, all published between January 1 and December 31, 2023. Table 2 shows the distribution of articles across countries and the proportions of those articles containing HI framing (based on automated labels, see Section 4.4). Additionally, we have a held-out data set of 100 articles in each language for evaluation. 62 articles of the English validation set overlap with the the labelled dataset released by Vallejo et al. (2025), allowing us to compare our extended framework with theirs (Section 4.5).[5]

4.2 Annotation Procedure

To validate our framework and assess the accuracy of the automatic labelling (Section 4.4), we first conducted a manual annotation process. We recruited 14 Spanish and 11 English native speakers, all PhD students with relevant research experience. These annotators were tasked with labelling articles to identify the presence of HI stories using the framework described in Section 3. Our instructions included examples illustrating both positive and negative instances of HI articles, along with question-level answers and explanations. Each article received three independent annotations: two from our annotators and one by one of the authors of this paper, who is fluent in English and Spanish.

Each annotator was assigned 10 articles at a time, with some annotators completing up to three rounds, but no more than that to minimise potential bias from any single annotator. We aggregated the labels for each article by applying majority vote. Annotators were compensated at an hourly rate of USD$26, exceeding the local minimum wage.

4.3 Inter-Annotator Agreement

Table 3 depicts the absolute overlap across all annotators and chance-corrected agreement based on Krippendorff's alpha (Hayes and Krippendorff, 2007). For the aggregated HI-label across three annotators and two classes (HI vs. non-HI), we ob-

[3] proquest.com

[4] We used the serverless service offered by together.ai and the model: deepseek-ai/DeepSeek-R1-Distill-Llama-70B-free.

[5] We noticed some level of noise in their data which led us to remove the remaining 38 articles which did not primarily focus on climate change.

tained a Krippendorff's alpha of 0.43 for English and 0.51 for Spanish. The average of absolute pairwise agreement is 0.74 (min=0.73, max=0.76) in English and 0.79 (min=0.78, max=0.83) in Spanish. These results indicate moderate agreement which is consistent with expectations for a complex task like HI framing annotation and comparable with prior work (Card et al., 2015; Field et al., 2018).

We compared our manual annotations against those released by Vallejo et al. (2025) using the overlapping 62 articles in both data sets. Out of the 62, there was an agreement of 66% (41 articles). Further analysis of the discrepancies revealed that all disagreements were cases where our framework indicated no HI framing, but they were originally annotated as positive. In 61% (13 articles) of these cases, our three annotators unanimously agreed on the absence of HI framing, while in the remaining 8 cases, only 1/3 of our annotators labelled the document as having HI framing. Further analysis of these articles revealed that they often contain superficial elements that could be mistaken for HI framing, such as individuals' names, first-person quotes, or discussions of how an issue affects individuals. However, our framework considers $Q1$ to be the primary indicator of HI framing, requiring that an article must provide a specific example of an individual or group directly affected by the main issue ($Q0$), sharing their personal experience or story. General statements, even if made by individuals, do not constitute a 'human face' on the issue if they do not relate a human example with the problem at hand.

This analysis highlights the nuanced nature of HI framing and the importance of clear criteria in its identification. Our framework's decomposition of the annotation process into sub-questions helps to distinguish between the presence of HI framing and articles that merely mention individuals or personal impacts in a more general sense. This comparison reinforces the value of our framework in identifying HI framing, especially because these inconsistencies are likely to increase when scaling to more languages.

4.4 Automatic Labelling

We next automated the labelling process and applied it to the entire corpus of climate-change-related articles using DeepSeek R1. We prompted DeepSeek with the same instructions provided to our human annotators, aiming to ensure consistency in the interpretation of HI framing across both manual and automatic processes. The specific prompt used for this task is included in Appendix B for reference. Following the question-wise labelling process, we aggregated the model's answers to produce a final HI label for each article again using our function in Figure 2. Table 2 shows the resulting statistics. To evaluate our automatic labelling approach, we compared the obtained labels against the manually annotated samples, as presented next.

	Spanish			English		
	F1	R	P	F1	R	P
Q1	0.79	0.90	0.71	0.83	0.91	0.76
Q2	0.78	1.00	0.64	0.60	0.72	0.52
Q3	0.60	0.97	0.43	0.72	0.64	0.83
Q4	0.66	0.82	0.55	0.63	0.67	0.60
HI Label	0.82	0.88	0.76	0.67	0.83	0.56

Table 4: Automatic labelling performance on manually annotated articles for individual questions and the aggregated HI-label, in Spanish and English.

4.5 Data Quality

Table 4 presents DeepSeek's performance on our gold standard articles broken down by individual questions ($Q1$-$Q4$) and the aggregated HI-label. For the Spanish articles, the model demonstrates strong performance in identifying HI framing, with an F1 score of 0.82. The model performs particularly well on $Q1$ and $Q2$, with F1 scores of 0.79 and 0.78 respectively. For English, the aggregated HI-label shows a lower F1 score of 0.67, with high recall (0.83) but lower precision (0.56). However, $Q1$ performs the best with an F1 score of 0.83. It is worth noting that the lower performance on $Q3$, specially in Spanish, might be due to the inherent subjectivity of the question. The interpretation of empathy-inducing or compassion-generating language can vary among individuals, making this aspect of HI framing particularly challenging to model consistently. This difficulty is also reflected in a lower annotator agreement for $Q3$ and $Q4$ (see Section 4.3). It highlights the challenge of identify HI framing signals consistently, not only for LLMs but also for humans.

With Spanish articles, the model generally achieves higher recall across all questions, while in English the model tends to have better balance between precision and recall, particularly for $Q3$. The strong performance on $Q1$ in both languages (F1

scores of 0.79 and 0.83) is particularly encouraging, given its importance in our HI framing framework, as we explain in Section 4.3. These results, along with the comparison to previous work suggest that using our proposed framework, LLMs are capable of accurately identifying HI framing in articles. Additionally, the performance is better than the reported in previous work (Vallejo et al., 2025), which achieved an accuracy of 79% compared to our value of 82%.[6] We thus apply our method to the full data set and perform an in-depth analysis of HI framing.

5 Cross-Lingual Analysis

After evaluating the quality of our automatically labelled data, we conducted a more in-depth analysis of articles containing HI framing to better understand the characteristics and nature of this phenomenon. We follow Vallejo et al. (2025) in using entities and sentiment for capturing the essence of HI framing: entities help us identify the key roles different actors play in climate change reporting, while sentiment analysis reveals the overall tone (either positive or negative) conveyed through these narratives. Our analysis focused on two aspects: first, we examined the role of entities in articles exhibiting HI framing across countries and languages, and second, we compared sentiment patterns across countries, languages, and article classes (HI vs. non-HI). The remainder of this section details these experiments and their findings, providing insights into the manifestation of HI framing in climate change reporting.

5.1 Entity Role Analysis

The primary roles that individuals can play in a conflict (such as the hero, the victim, and the villain) have been broadly explored in studies looking at news articles (Gerbner and Gross, 1976; Van Dijk, 2013; Jones et al., 2014). In the NLP community, researchers have also examined these roles (Jahan et al., 2021; Stammbach et al., 2022; Frermann et al., 2023). In this paper, we investigate the role of entities in HI frames in climate change reporting across countries and languages. Our analysis extends the work of Vallejo et al. (2025), who only modelled victims, to all three roles (hero, victim, and villain) and to a cross-lingual comparison, whereas they focused on English.

To understand the role of different entities in articles with HI framing, we followed a two-step process using the `Llama-3.3` model, which we chose due to resource constraints and because it produces outputs that are easier to parse.[7] First, we prompted the model to identify the most representative hero, victim, and villain entities in each article. Then, in a second pass, we matched these entities to stakeholder groups as defined by Frermann et al. (2023), with the addition of a 'public figures' category to address a gap in their classification. Figure 4 presents the proportion of these groups for each language (see Figure 8 in Appendix C for a detailed breakdown by country), noting that not all groups are present in both languages. For Spanish, we manually translated the prompt and the stakeholder names. After conducting the automatic entity extraction and mapping of stakeholders, we did a manual evaluation of ten samples in each language to verify that the automatic processing was reliable. Here, we conflate our results to the language-level for space reasons, but present per-country results in Appendix C.

Looking at Figure 4, we observed a major distinction between English- and Spanish-speaking countries in terms of hero portrayal. *Public figures* and *Environmental organisations and activists* are present across both languages and all countries, but the latter are proportionally more frequently shown as heroes in English-speaking countries compared to Spanish-speaking ones. Conversely, in Spanish-speaking countries, this role is often assigned to *governments and politicians*. For instance, figures such as *Luiz Inácio Lula da Silva*, *Justin Trudeau*, *Gustavo Petro*, and *Joe Biden* are frequently portrayed as heroes in Spanish articles, while *Antonio Guterres*, *President Biden*, and *John Kerry* occur more prominently in English coverage. The framing is evident in statements like: *Antonio Guterres, who has come to speak with the most force on our failure to rise to the threat of climate change* or *Even those who deny we are in a climate crisis can't deny the impact extreme heat is having on Americans, President Biden said*. Interestingly, entities from the *emissions industry & issues* group are also portrayed as heroes, especially in Spanish-speaking countries.

The *general public* and *animal and nature* groups are consistently depicted as victims across

[6] This accuracy was calculated using the 62 articles in English mentioned in Section 4.3.

[7] We use the serverless version of the model provided by `together.ai`

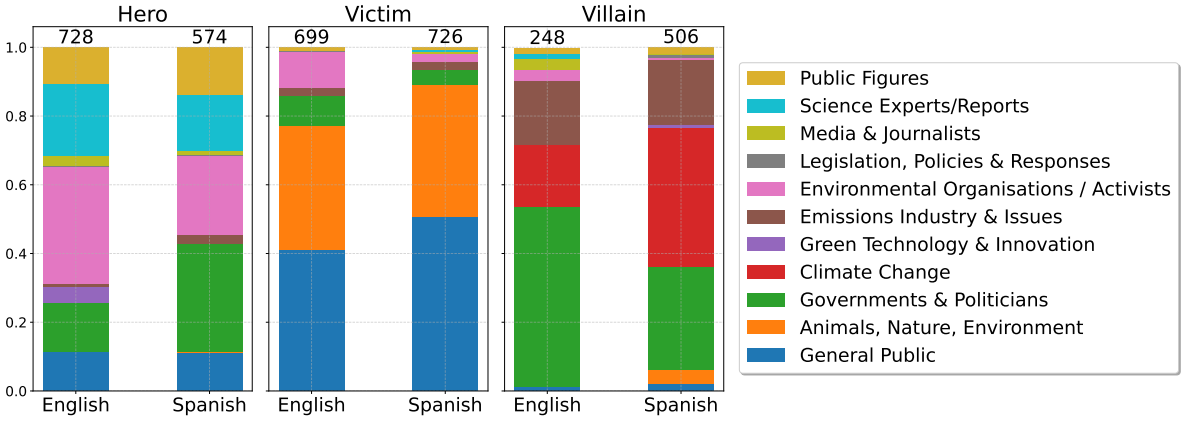

Figure 4: Distribution of stakeholder groups associated with hero, victim, and villain entities in HI-framed articles, by language.

both languages. *Government and politician entities* also appear as victims to some extent, with higher proportions in English-speaking countries; some frequent entities in these categories are *The State of California*, *Europe*, and *The Government of Australia*. *Environmental organisations and activists* are more frequently portrayed as victims in English-speaking countries compared to Spanish-speaking ones, suggesting potential differences in how these groups are perceived or represented in different cultural contexts. Sometimes these environmental organisations or activist are part of the general public who become activists demanding their rights, for instance, we found entities like *Alaska Native communities*, *The Timbisha Shoshone tribe*, or *The 16 plaintiffs*—the latter a group of young people who are suing the State of Montana for lack of action regarding climate change.

Governmental and politician entities, along with *climate change* and *emissions industry*, dominate the villain category across most countries. Interestingly, climate change entities were not present as villains in the Australian context (see Appendix Figure 8), which contrasts with all other countries in our study. A unique pattern in Spanish-speaking countries was that animals and nature are occasionally portrayed as villains, albeit in lower proportions, an example being entities representing environmental consequences of climate change, e.g. *El huracán Idalia* "Idalia hurricane", *la deforestación* "deforestation", or *la sequía* "the drought". This could reflect differences in the impact of climate change on the HI narrative framing strategies across countries.

5.2 Sentiment Analysis

We investigated sentiment in articles with and without HI framing, following Vallejo et al. (2025). We calculated an average sentiment score for each article by averaging the sentiment scores of its individual sentences. The sentiment score was calculated using a RoBERTa model, fine-tuned with Twitter data (Barbieri et al., 2020) for English, and pysentimiento for Spanish (Pérez et al., 2021) The distribution of sentiment scores for HI and non-HI articles for all countries is shown in Figure 5. We observe that the contrast between HI and non-HI sentiment was more pronounced in Spanish-speaking countries compared to English-speaking countries. In English-speaking countries, the overall sentiment for both HI and non-HI articles peaks more tightly around neutral (0), whereas Spanish-speaking countries showed a broader spread and a more negative skew for HI articles.

The graphs also reveal country-specific differences in sentiment distribution. In Australia, non-HI articles are relatively more negative, with HI articles peaking close to neutral. This trend is reversed for all other countries, where HI articles are more negative. Out of the English-speaking countries, the UK's HI articles have the broadest distribution, while the US HI articles concentrate the most sharply around the neutral point. In Spanish-speaking countries, Argentina, Colombia, and Mexico exhibited a clear difference in sentiment distribution between HI and non-HI articles, with HI articles having a more negative sentiment. Spain's sentiment distribution showed a similar trend, although with more overlap between the two curves.

More generally, these findings provide novel insights into the relationship between HI framing and

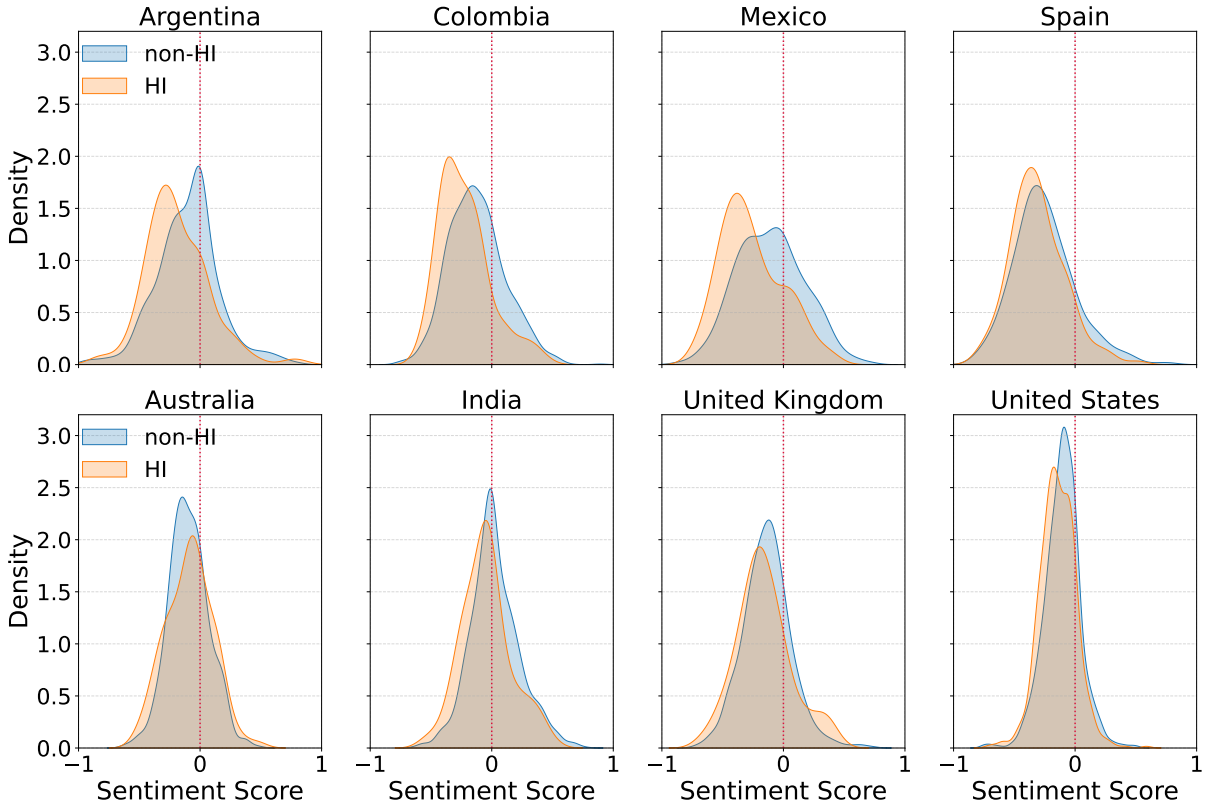

Figure 5: Sentiment analysis across both Spanish- (upper row) and English-speaking countries (lower row). The zero-line (in red) represents neutral sentiment and the sentiment spans between −1 (negative) to 1 (positive).

news article sentiment in a comparative way across countries and languages. We observe that countries with the same language by and large exhibit similar trends, and that climate-related HI stories elicit a more negative sentiment in Spanish-speaking countries compared to English-speaking countries. In line with our observation that Spanish HI stories portray *nature* as a villain more often than English HI framed articles, we posit that the 'human face' on climate impact in Spanish-speaking reporting is more dim, possibly due to a more direct impact particularly on lower-income social groups and countries in the global south.

6 Conclusion

This paper presented the first large scale cross-lingual NLP investigation of HI framing. It contributes to the analysis of HI framing in climate change reporting by proposing a novel framework for HI frame annotation and prediction. This framework offers a flexible and adaptable approach, allowing researchers to adjust the definition's strictness according to their specific research needs. We release our manual annotations and a much larger automatically generated silver-standard as a valuable resource for future research. Our analysis of features associated with HI framing, including entity roles and sentiment analysis, reveals patterns in how different entities are portrayed across various countries and languages. Particularly, our findings highlight variations in HI framing between English-speaking and Spanish-speaking countries, emphasising the importance of considering cultural and linguistic factors in framing analysis. These contributions provide a basis for further research into the role of HI framing in shaping public discourse around climate change. In summary, the insights gained from this study contribute to a deeper understanding of how HI framing operates in climate change reporting across different cultural contexts and languages.

Limitations

LLMs contain biases which arise from their training data and procedures, and the automated labels produced in this work might contain biases that we were not able to expose. We addressed this concern through rigorous manual validation and creation of a gold standard, but hidden biases might remain.

Our approach to matching entities to stakeholder

groups relied on categories derived from English data, which were translated to Spanish but may not be directly applicable to Spanish data. A more nuanced approach would be to infer these categories directly from the Spanish data itself, which is a potential avenue for future research.

References

Toril Aalberg and Audun Beyer. 2015. Human interest framing of irregular immigration: An empirical study of public preferences for personalized news stories in the united states, france, and norway. *American Behavioral Scientist*, 59:858–875.

David Alonso del Barrio and Daniel Gatica-Perez. 2023. Framing the news: From human perception to large language model inferences. In *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval*, ICMR '23, page 627–635, New York, NY, USA. Association for Computing Machinery.

Francesco Barbieri, Jose Camacho-Collados, Luis Espinosa Anke, and Leonardo Neves. 2020. TweetEval: Unified benchmark and comparative evaluation for tweet classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1644–1650, Online. Association for Computational Linguistics.

Mark Boukes, Hajo G. Boomgaarden, Marjolein Moorman, and Claes H. De Vreese. 2015. Political news with a personal touch: How human interest framing indirectly affects policy attitudes. *Journalism & Mass Communication Quarterly*, 92:121–141.

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 the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444, Beijing, China. Association for Computational Linguistics.

Seung Ho Cho and Karla K. Gower. 2006. Framing effect on the public's response to crisis: Human interest frame and crisis type influencing responsibility and blame. *Public Relations Review*, 32(4):420–422.

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.

Tine Ustad Figenschou and Kjersti Thorbjørnsrud. 2015. Faces of an invisible population: Human interest framing of irregular immigration news in the united states, france, and norway. *American Behavioral Scientist*, 59:783–801.

Tine Ustad Figenschou, Kjersti Thorbjørnsrud, and Daniel C Hallin. 2023. Whose stories are told and who is made responsible? human-interest framing in health journalism in Norway, Spain, the U.K. and the U.S. *Journalism*, 24(1):3–21.

Lea Frermann, Jiatong Li, Shima Khanehzar, and Gosia Mikolajczak. 2023. Conflicts, villains, resolutions: Towards models of narrative media framing. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8712–8732, Toronto, Canada. Association for Computational Linguistics.

George Gerbner and Larry Gross. 1976. The scary world of TV's heavy viewer. *Psychology Today*, 9(11):41–45.

Andrew F. Hayes and Klaus Krippendorff. 2007. Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, 1(1):77–89.

Labiba Jahan, Rahul Mittal, and Mark Finlayson. 2021. Inducing stereotypical character roles from plot structure. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 492–497, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Nael Jebril, Claes H. de Vreese, Arjen van Dalen, and Erik Albæk. 2013. The effects of human interest and conflict news frames on the dynamics of political knowledge gains: Evidence from a cross-national study. *Scandinavian Political Studies*, 36(3):201–226.

Michael D. Jones, Mark K. McBeth, and Elizabeth A. Shanahan. 2014. Introducing the narrative policy framework. In Michael D. Jones, Elizabeth A. Shanahan, and Mark K. McBeth, editors, *The Science of Stories*, page 1–25. Palgrave Macmillan US, New York.

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.

Juan Manuel Pérez, Juan Carlos Giudici, and Franco M. Luque. 2021. pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks. *CoRR*, arXiv:2106.09462.

Arianne Reimerink, Melania Cabezas-García, Pilar León-Araúz, and Pamela Faber. 2024. Ideological knowledge representation: Framing climate change in EcoLexicon. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 8617–8626, Torino, Italia. ELRA and ICCL.

Holli A. Semetko and Patti M. Valkenburg. 2000. Framing european politics: a content analysis of press and television news. *Journal of Communication*, 50:93–109.

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.

Dominik Stammbach, Maria Antoniak, and Elliott Ash. 2022. Heroes, villains, and victims, and GPT-3: Automated extraction of character roles without training data. In *Proceedings of the 4th Workshop of Narrative Understanding (WNU2022)*, pages 47–56, Seattle, United States. Association for Computational Linguistics.

Dominik Stammbach, Philine Widmer, Eunjung Cho, Caglar Gulcehre, and Elliott Ash. 2024. Aligning large language models with diverse political viewpoints. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 7257–7267, Miami, Florida, USA. Association for Computational Linguistics.

Sebastián Valenzuela, Martina Piña, and Josefina Ramírez. 2017. Behavioral effects of framing on social media users: How conflict, economic, human interest, and morality frames drive news sharing. *Journal of Communication*, 67(5):803–826.

Gisela Vallejo, Christine de Kock, Timothy Baldwin, and Lea Frermann. 2025. Human interest framing across cultures: A case study on climate change. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 11380–11398, Abu Dhabi, UAE. Association for Computational Linguistics.

Teun A. Van Dijk. 2013. *News Analysis*, 1st edition. Routledge.

Xinyan Zhao, Mengqi Monica Zhan, and Brooke Fisher Liu. 2019. Understanding motivated publics during disasters: Examining message functions, frames, and styles of social media influentials and followers. *Journal of Contingencies and Crisis Management*, 27(4):387–399.

A Data Distribution by Outlets

We relied on Media Bias Fact Check and Wikipedia to obtain media bias labels for each of the media outlets that we include in our analysis.[8] Below we depict the proportion of articles that we include in our analysis by outlet and the outlet's media bias categorisation as well as the number of those articles that contain HI framing. Figure 6 shows that our analysis includes a diverse range of outlets, spanning from right- to left-leaning, providing a comprehensive representation of the media landscape. Also it can be observed that there are some outlets with a much higher number of HI framing articles compared to their overall article count.

B Prompt for Labelling HI Framing

Our process of automatic labelling HI framing included experimenting with prompting to obtain a good performance. Figure 7 introduces the prompt that we used in the labelling task consisting of the four sub-questions from Semetko and Valkenburg (2000) and $Q0$ that we integrated for better linking the main issue of the article to the rest of the questions. The prompt also includes format instructions for the answer and two examples with actual articles and answers. Due to space constraints in this document, we only include the examples as placeholders.

C Entities Analysis by Country

The patterns of our analysis of entity roles in articles containing HI framing each country become more evident when looking closer at the proportions of each role by country. In Figure 8, we depict those distributions by country. As shown in the figure English-speaking countries, India tends to be an outlier having different proportions as the rest of the countries, for instance showing that *Environmental Organisations* are more prominent portrayed as heroes in HI articles or having a higher proportion of *General Public* entities labelled as victims. In comparison, Spanish-speaking countries also reveal individual patterns, with Spain being slightly different from the Latin-American countries.

[8] https://mediabiasfactcheck.com/

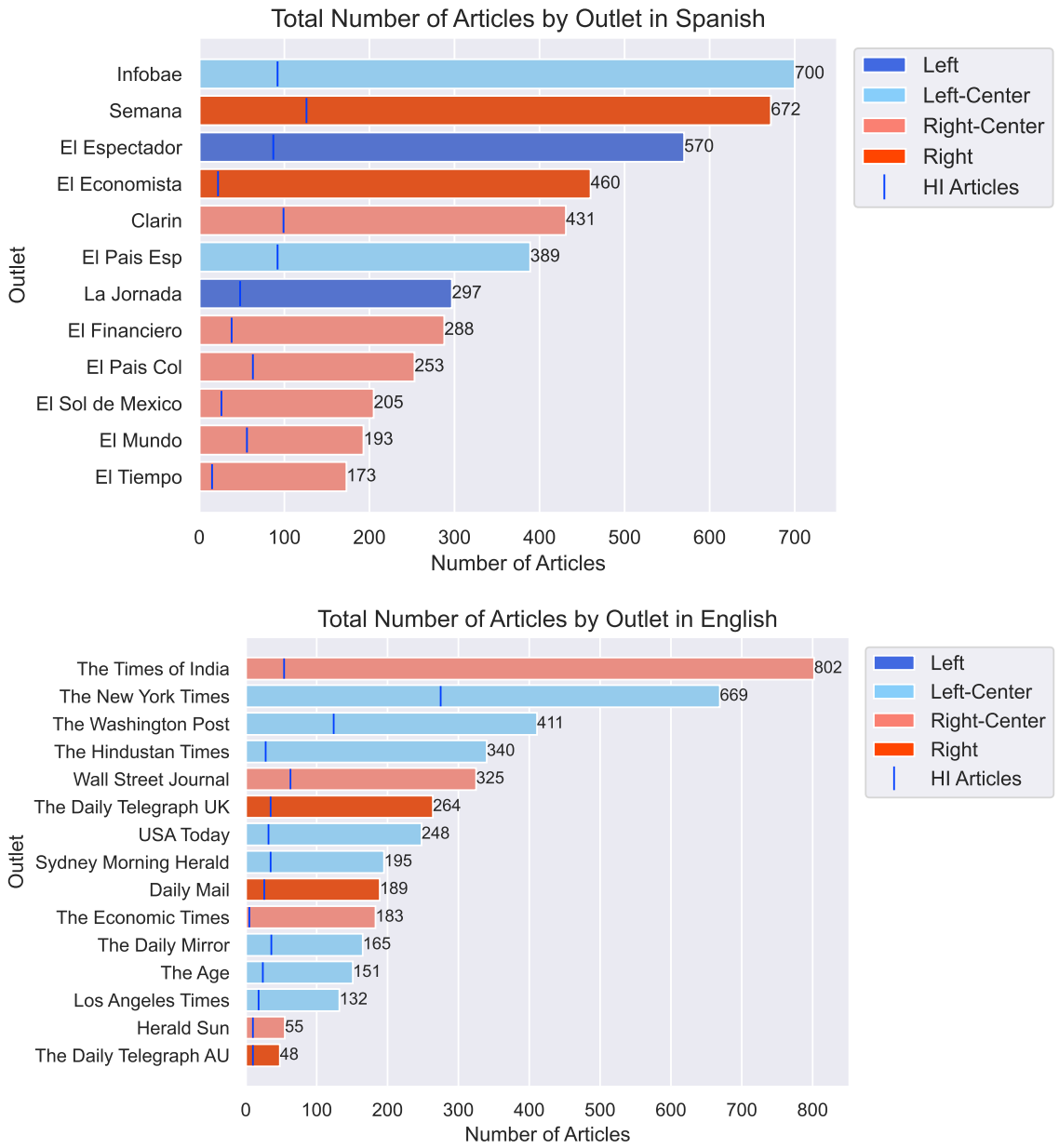

Figure 6: Number of total articles by outlet in both English and Spanish. The blue line denotes the number of articles with HI framing.

```
[INST] You are a journalist scholar, and your task is to look at this
article carefully. We are looking for examples of a *Human Interest story*
You will help us to find them by answering the questions below. You
 must provide the reasoning of why your answer is positive or negative for
questions 1 to 4:
Your answer must be in the form:
ANSWER: Q0: short sentence; Q1: yes/no; Q2: yes/no; Q3: yes/no; Q4: yes/no
REASON: Q1: reason1; Q2: reason2; Q3: reason3; Q4: reason4

Question 0: Can you identify the article's main issue/problem?; Question 1:
With respect to the main issue from Question 0, does the article provide a
human example or 'human face' on that issue?; Question 2: With respect to
the main issue from Question 0, does the article go into the private or
personal lives of the actors?; Question 3: With respect to the main issue
from Question 0, does the article employ adjectives or personal vignettes
that generate feelings of outrage, empathy-caring, sympathy, or
compassion?; Question4 With respect to the main issue from Question 0, does
the article emphasize how individuals and groups are affected by that
issue/problem?
Below, you have two examples:
[/INST]

[EX]
EXAMPLE1
{ARTICLE}

EXAMPLE2
{ARTICLE}
```

Figure 7: Prompt for identifying HI framing automatically.

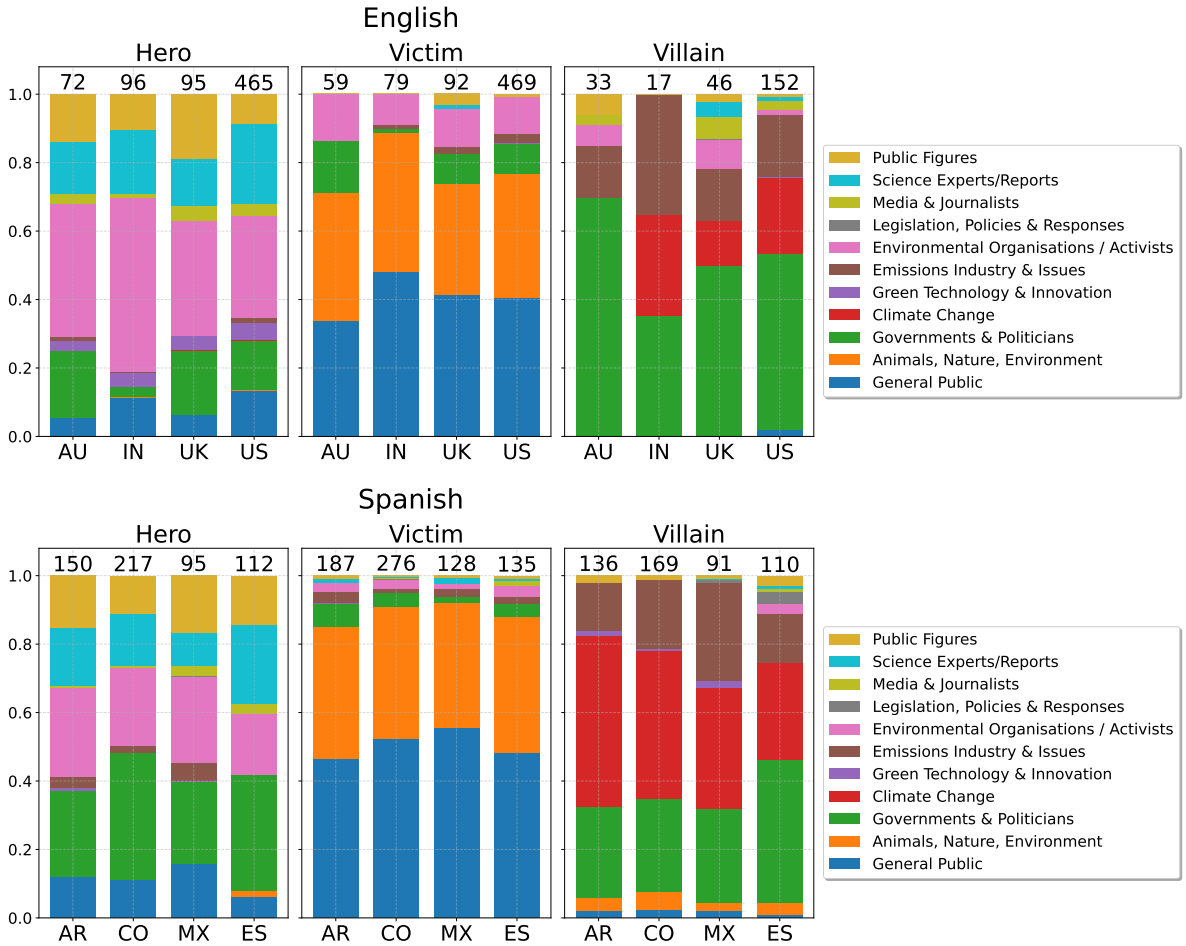

Figure 8: Distribution of stakeholder groups associated with hero, victim, and villain entities in HI-framed articles for English- (upper) and Spanish-speaking (lower) countries.