Multimodal Analysis of Emotion and Harmful Content in Online Communication

Ronghao Pan

Facultad de Informática, Universidad de Murcia, Campus de Espinardo, 30100, Spain

Abstract

The increasing prevalence of emotional and harmful content in online communication raises important challenges for understanding and promoting healthy public discourse. Emotional expression and harmful language play a key role in shaping narratives and influencing audiences in various contexts, from news media and political debates to popular culture, including emotionally charged references. This research proposes a multimodal approach to analyzing online communication, integrating features that detect emotion, harmful language, and their interaction across text, audio, and images. We explore strategies that combine multimodal embeddings and Large Language Models (LLMs), aiming to provide insights into how emotional and harmful content contribute to the dynamics of online communication. To support this analysis, we have compiled and published several datasets such as the (*Spanish MEACorpus 2023*) and the (*Spanish MTLHateCorpus 2023*) for multimodal emotion recognition and for hate speech detection, respectively. In addition, we have participated in several shared tasks, achieving competitive results in international workshops such as IberLEF, CLEF, and SemEval. For example, we obtained a top-10 rank in Task 4 of SemEval-2024 (64.77% Hierarchical F1 in persuasion detection), 1st place in IberLEF-2024's FLARES task (65.82% F1 in reliability assessment), and 1st place in the caption prediction subtask of ImageCLEFmed Caption 2025. We also ranked 1st in category detection in several languages in SemEval-2023 Task 3. Additionally, we ranked in the top 10 in multiple subtasks of EXIST 2025, which focused on multimodal sexism detection. These results highlight the potential and effectiveness of our multimodal and classification methods for addressing emotional and harmful content in complex discourse scenarios.

Keywords

Large Language Models, Multimodal Analysis, Multimodal Emotion Recognition Hate Speech Detection, Natural Language Processing

1. Introduction

In recent years, online communication has become the primary space for the dissemination and consumption of information. Social networks, news websites, and digital platforms have transformed public discourse, providing unprecedented access to diverse sources while also facilitating the rapid spread of emotional, harmful, and misleading content [1]. This dual dynamic poses significant challenges to fostering informed, healthy, and democratic conversations in digital environments.

A growing body of research has demonstrated the critical role of emotional appeals and harmful and polarizing language in shaping narratives, influencing cognition and driving engagement in the online networks [2, 3]. In particular, emotional appeals have been linked to the perceived veracity and amplification of misleading or manipulative messages, making it more difficult for the public to critically evaluate information. In addition, harmful and offensive language (hate speech) often accompanies emotionally charged content, reinforcing polarizing debates and fostering toxic communication climates. These dynamics are not limited to political contexts but extend to other areas of public communication, including sports, entertainment, and popular culture, where emotionally charged references, such as football quotes, mobilize strong affective responses among audiences.

Traditional approaches to analyzing online communication and identifying problematic content have largely relied on textual analysis, fact checking, or source verification. However, these methods are limited when faced with the multimodal and increasingly sophisticated nature of online content, which includes text, audio, and images. The rise of multimodal communication, particularly in politics, news

Doctoral Symposium on Natural Language Processing, 25 September 2025, Zaragoza, Spain.

ronghao.pan@um.es (R. Pan)

0009-0008-7317-7145 (R. Pan)

content, and social media, requires analytical frameworks capable of processing and interpreting signals across multiple modalities.

At the same time, advances in artificial intelligence have opened new possibilities for the analysis of complex discourse. Large Language Models (LLMs), such as GPT-4 [4], LLaMA-3 [5], Gemma-2 [6], and Gemini [7], have demonstrated remarkable abilities to understand, generate, and reason over natural language. Beyond text, the development of multimodal LLMs, such as Phi-4 [8] and Canary[1], enables models to process and integrate information across text, images, and other modalities, providing new ways to analyze the complexity of online communication.

Despite these technological advances, distinguishing between authentic content from misleading, harmful, or manipulative content remains a fundamental challenge. Previous research has shown that users' emotions and affective responses influence their likelihood to share content, often without verifying its accuracy [3, 9]. Recognizing and analyzing the emotional framing of messages is therefore critical to understanding how content spreads and influences audiences. Emotion recognition (ER), defined as the task of automatically identifying emotional cues expressed through text, speech, facial expressions, or gestures, has emerged as a promising tool in this area [10, 3]. ER systems can classify emotions such as anger, fear, happiness, sadness, or disgust, providing insight not only into what is being communicated but also into how it is emotionally framed.

In addition to emotional cues, recent research has shown that hate speech and offensive language are closely associated with fake news, often co-occurring to amplify polarizing or manipulative narratives. Detecting hate speech alongside emotional signals provides complementary information that can enhance the analysis of online discourse, particularly in multimodal settings that include both textual and visual content [10].

The objective of this Ph.D. thesis is to develop a multimodal approach to analyzing online communication by integrating features that detect emotional expressions, harmful or polarizing language, and their interaction across different modalities, including text, audio, and images. Building on recent advances in LLMs and multimodal embeddings, this research explores strategies for using these tools to provide a more comprehensive understanding of how emotional and harmful content shapes the dynamics of online discourse. The study incorporates diverse data sources, including news articles, websites known to spread misinformation, and transcripts of political debates such as parliamentary sessions, capturing a wide range of communicative contexts.

The insights generated by this thesis aim to support future applications, such as improving content analysis tools or informing interventions to mitigate problematic discourse in digital environments. For example, the findings may contribute to improving disinformation detection pipelines by adding complementary layers of analysis based on emotional and harmful content signals, especially in complex, multimodal, and multilingual contexts such as the Spanish-speaking digital space.

2. Research Hypotheses

The research hypotheses in this thesis focus on the analysis of emotional and harmful content in online communication through multimodal approaches that integrate emotional features, hate speech detection, and leverage LLMs. While not exclusively aimed at disinformation detection, the analysis seeks to uncover patterns and mechanisms by which emotional and harmful content influences online discourse, with the potential to support downstream applications such as improving disinformation detection systems or content moderation tools. The core hypotheses are:

- **(H1)**: Multimodal approaches (text, audio, images) provide richer and more accurate insights into online discourse than unimodal methods.
- **(H2)**: Incorporating emotional cues and hate speech features enhances the detection of polarizing or harmful content.

[1] https://huggingface.co/nvidia/canary-1b

- **(H3)**: Leveraging LLMs improves generalization and interpretability, particularly in Spanish multimodal contexts.

While multimodality is a well-established area, this thesis focuses on exploring modality interactions within specific domains (e.g., politics, sport, among others) and under what conditions multimodality yields measurable gains in performance and interpretability.

To achieve this, the following objectives have been defined: (OB1) Creation of a multimodal corpus in Spanish for training models of emotion recognition and hate speech detection. This dataset will include text, audio, and image data collected from different online sources to support model development and evaluation; (OB2) Evaluation of different multimodal approaches that combine and fuse features from different modalities (e.g., text-audio, text-image, fully multimodal) for emotion recognition and hate speech detection; (OB3) Investigation of LLM-based approaches for emotion and hate speech detection, including fine-tuning and prompt-based techniques, with special attention to their performance on Spanish language content and their ability to handle multimodal input; (OB4) Construction of a multimodal corpus covering different domains of online communication (e.g., political, sports, entertainment) from different sources such as Twitter, official websites, YouTube and fact-checking platforms such as Newtral[2] and Maldita[3]. This corpus will be used to analyze how emotional content and hate speech contribute to the spread of information and misinformation in different communicative contexts.

3. Methodology and Experiments

In this section, we describe the methodology and experiments designed to address the research objectives and hypotheses outlined in this thesis.

As a first step, we developed a web crawler to systematically collect data from multiple online sources. This crawler retrieves content from various domains of online communication, including political news websites, official government portals, sports and entertainment platforms, social media (e.g., Twitter), and video platforms such as YouTube. In addition, we include data from fact-checking platforms such as Maldita and Newtral, which serve as repositories of verified misinformation claims. Thus, the main objective of this data collection and corpus creation process is twofold: first, to provide high-quality resources for training and evaluating models capable of detecting emotional expressions and harmful language across different modalities; second, to lay the foundation for building a broader corpus spanning different online communication domains (e.g., politics, sports, entertainment), which will enable the analysis of how emotional content and hate speech interact and influence the spread of information and misinformation in online environments.

Using the data collected by this crawler, we have created and published two corpora: a multimodal corpus to support the development of models for emotion recognition and a text-based corpus for hate speech detection. The first corpus is *Spanish MEACorpus 2023* [11], a multimodal corpus for emotion recognition in Spanish. This dataset contains 13.16 hours of speech, divided into 5,129 labeled segments, annotated by three members of our research group according to Ekman's six basic emotions (disgust, anger, happiness, sadness, fear, and neutral). We have evaluated several multimodal approaches that combine speech representation techniques and linguistic models for emotion classification. These approaches range from simple text-based emotion detection to fusion methods that integrate automatic speech recognition models such as Wav2Vec2-BERT [12] with pre-trained language models such as BETO. Among the fusion strategies, we have explored late fusion (concatenating or averaging model outputs), multi-head cross-attention (integrating cross-attention mechanisms to better capture audio-text relationships), and ensemble learning (combining predictions by averaging or selecting the maximum probability). Looking at the results obtained by the multimodal model, there are a total of 102 cases of errors. We have grouped these errors into 4 categories: (1) both the text model and the audio model failed in prediction; (2) the multimodal learns from the text model; (3) the multimodal

[2] https://www.newtral.es/
[3] https://maldita.es/

learns from the audio model; and (4) the multimodal and both the text model and the audio model failed in prediction. The Table 1 shows some examples of these three categories.

Table 1
Error analysis grouped by category error. The analysis includes the text and ground truth (Truth) as well as the individual results for multimodal model prediction (M), W2V-BERT model prediction (A), and MarIA model prediction (T). The emotions are: Anger (A), Joy (J), Neutral (N), Sadness (S), and Disgust (D).

Text	Truth	M	A	T
Error 1: Multimodal is successful, but audio and text are not.				
Hoy quedan 34 días también, porque esto es un 3 en 1.	J	J	A	N
Error 2: Multimodal and text are successful, but audio is not.				
que diga hablan con mala leche porque se nota que como ya no están como que parece que la escena no se ha usado el sitio pues no haberte ido pues haberlo hecho mejor anda que capello que juega con diarrea y emerson de doble pivote que era cemento armado	A	A	N	A
Error 3: Multimodal and audio are successful, but text is not.				
¿Cómo valora los informes de la AIREF y del Banco de España sobre la economía española? Pues mire, señor Sánchez, le quedan 24 horas.	A	A	A	D
Error 4: No model is successful.				
que acabe ya esa jaula de grillos, esa casa de los líos ¿Espanya merece un gobierno con ese lío?	A	D	D	D

This dataset has been used as a basis for the organization of the EmoSPeech task [13] in IberLEF 2024, which consists of two subtasks: text-based automatic ER and multimodal automatic ER. The novelty of this task lies in its multimodal approach to ER, analyzing the performance of language models on the *Spanish MEACorpus 2023*.

The second corpus is *Spanish MTLHateCorpus 2023* [14], which was created to address the growing challenge of harmful content online through the detection of hate speech. This corpus was built using tweets and online content collected by the crawler, and includes annotations for several subtasks: identifying the intensity of hate speech, determining the target groups, and distinguishing whether the target is an individual or a collective entity. We evaluated a multi-task learning approach using mBART and T5, comparing its performance against LLMs in zero-shot learning as a baseline, and against an ensemble of fine-tuned models as an upper bound. The results showed that multi-task learning improves versatility by allowing a single model to effectively handle multiple tasks, achieving competitive results, especially in group target identification, although ensemble learning achieved slightly better performance.

As a second step, we have evaluated different multimodal approaches that combine and fuse features from different modalities (e.g., text-audio, text-image, full multimodal) for the tasks of emotion recognition and hate speech detection. This evaluation has been carried out through participation in shared tasks and competitive benchmarks organized by leading evaluation forums, including IberLEF, CLEF, and SemEval. Participation in these shared tasks provides a practical framework to validate our approaches under real-world conditions and allows comparison with other state-of-the-art methods. Each competition addresses different challenges related to multimodal analysis, providing valuable insights into the strengths and limitations of different modeling strategies. The tasks involved are as follows:

- **EXIST 2025**: We developed multimodal systems for binary sexism detection, source intention classification, and sexism categorization on text, image, and video inputs. Using XLM-RoBERTa, ViT, and VideoMAE, our models handled both soft and hard evaluation settings. Our systems ranked in the top 10 in multiple subtasks among 244 teams.

- **EXIST 2O24** [15]: In this task, we used the CLIP model to extract the embedded text and image, and then combined them by diagonal multiplication to obtain the classification models. We ranked 33rd in sexism identification and 18th in both source intent and sexism categorization [16].
- **SemEval-2024 Task 4: Multilingual Detection of Persuasion Techniques in Memes** [17]: For this task, we evaluated LlaVa to extract image descriptions and combine them with the meme text. Our system performed well in all subtasks, achieving the tenth-best result with a Hierarchical F1 of 64.774%, the fourth best in Subtask 2a with a Hierarchical F1 of 69.003%, and the eighth best in Subtask 2b with a Macro F1 of 78.660% [18].
- **SemEval-2024 Task 10: Emotion Discovery and Reasoning its Flip in Conversation** [19]: It focused on recognizing and reasoning about emotional changes in conversation. This task included different languages like English and Hindi. Our best result was the 6th place in Subtask 2 with an F1 score of 26% [20].
- **SemEval-2025 Task 1: AdMIRe**: We participated in Subtask A, adopting a multimodal approach combining textual and visual features with pre-trained language models and vision transformers, and achieve 17th place in English and 9th place in Portuguese in the official evaluation.

As a third step, we have investigated LLM-based approaches to misinformation detection, including both fine-tuning strategies and prompt-based techniques. These approaches focus on evaluating the ability of LLMs to process and analyze Spanish language content, as well as their potential for handling multimodal input.

Similar to the evaluation strategy described in the second step, we have tested these LLM-based approaches by participating in various shared tasks and competitive benchmarks, allowing us to validate their performance under different scenarios and datasets.

The tasks involved are as follows:

- **SemEval-2025 Task 3: Mu-SHROOM**: We addressed multilingual hallucination detection using a token classification approach based on XLM-RoBERTa-large, enhanced with contextual information from Llama-3.1-70B, achieving superior performance over baselines in detecting token-level hallucinations.
- **SemEval-2025 Task 7: Multilingual and Crosslingual Fact-Checked Claim Retrieval**: We implemented a multilingual retrieval system combining XLM-RoBERTa with metric learning and Multi-Similarity Loss, ranking 25th overall while achieving over 50% hit rate in most languages.
- **IberLEF 2024: FLARES** [21]: We developed a NER-based approach integrating BETO, POS, and Dependency features for 5W1H identification, ranking 2nd in Task 1 (56.778%) and achieving 1st place in Task 2 (65.820%) for 5W1H-based reliability assessment.
- **SemEval-2024 Task 6: SHROOM** [22]: We used prompt-based zero-shot classification with LLaMa-2, Tulu, and Mistral, ranking 18th in the model-aware setup (78.4% accuracy) and 29th in the model-agnostic setup (76.93% accuracy) [23].
- **SemEval-2023 Task 3: Detecting the Category, the Framing, and the Persuasion Techniques in Online News in a Multi-lingual Setup** [24]: We applied a multilingual transformer fine-tuned across languages, ranking 1st in some languages for category detection and achieving top 10 positions across all languages in framing detection [25].
- **SemEval-2023 Task 5: Clickbait** [26]: We fine-tuned pre-trained models for spoiler classification, achieving top 10 results across most measures and reaching top 3 in F1 score for passage spoiler detection [27].
- **ImageCLEF 2025 : ImageCLEFmed Caption**: We participated in both caption prediction and concept detection subtasks using a two-stage vision-language system. For captioning, we fine-tuned the BLIP model on a radiology dataset with a composite relevance loss (BERTScore, ROUGE-1, and exact match), ranking 1st with a score of 0.3771. For concept detection, we combined SciSpacy NER, SapBERT retrieval, and BERT-based reranking, reaching an F1-score of 0.2398 despite challenges from class imbalance and entity ambiguity.

In addition to these evaluations, we have studied the detection of hate speech in Spanish through an intelligent example selection system for Few-Shot Learning (FSL) based on diversity and uncertainty metrics, which improved recognition over Zero-Shot Learning and Random FSL approaches across multiple datasets, with Gemma-2 models achieving the best results [28]. Similarly, in [29], we investigated the use of LLMs for detecting sexist and hateful content online, comparing zero-shot, few-shot, and fine-tuning strategies, and showed that the Zephyr model outperformed previous benchmarks in hate speech detection tasks. These publications provide valuable empirical evidence on the capabilities of LLMs for NLP tasks such as classification.

Finally, once the entire multimodal corpus has been compiled, we plan to conduct an analysis of how emotional expression and hate speech influence the spread of information in different online communication domains. This analysis will explore the role that emotional and harmful content plays in the dynamics of content spread in contexts such as politics, sports, and entertainment. In addition, we aim to evaluate the potential benefits of incorporating emotion recognition and hate speech detection signals as complementary features in misinformation detection systems, and to assess whether these affective and harmful cues can enhance the detection of misleading or manipulative content across modalities and domains.

4. Conclusions and Future Work

This Ph.D. thesis focuses on the analysis of emotional and harmful content in online communication through multimodal approaches that integrate emotion recognition, hate speech detection, and the application of LLMs, with special attention to Spanish-language content. While not exclusively aimed at misinformation detection, the research explores how emotional expressions and hate speech interact and contribute to the spread of content in different online communication domains, including politics, sports, and entertainment.

Throughout this research, we have developed key resources and performed evaluations in line with our objectives. Specifically, we have created a multimodal corpus for emotion recognition (*Spanish MEACorpus 2023*), a corpus for hate speech detection (*Spanish MTLHateCorpus 2023*), and a web crawler for collecting multimodal content from different sources such as political news, social media, official websites, and fact-checking platforms. We have also evaluated different multimodal and LLM-based approaches by participating in shared tasks and competitive benchmarks organized by international forums such as SemEval and IberLEF.

We are currently in the final stages of compiling a broader multimodal corpus that spans different communication domains. Once completed, the next step will be to conduct an in-depth analysis of how emotional content and hate speech influence the spread of information and misinformation across domains and modalities. This analysis will also explore the potential benefits of incorporating emotion recognition and hate speech detection signals as complementary features in misinformation detection systems, and evaluate their contribution to improving detection performance across contexts.

Future work will focus on analyzing how emotional expression and hate speech influence the spread of information in different online communication domains. We also plan to evaluate the integration of these features into downstream applications, such as misinformation detection or content moderation, and assess their added value. In addition, we will explore the scalability of the developed models to other languages and domains, investigate the use of Retrieval Augmented Generation (RAG) to improve the analysis, and extend the corpora with new modalities and annotations to support broader research.

Acknowledgments

This work is part of the research project LaTe4PoliticES (PID2022-138099OB-I00) funded by MCIN/AEI/10.13039/501100011033 and the European Fund for Regional Development (ERDF)-a way to make Europe.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT, Grammarly in order to: Grammar and spelling check, Paraphrase, translate and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, science 359 (2018) 1146–1151.

[2] W. J. Brady, J. A. Wills, J. T. Jost, J. A. Tucker, J. J. V. Bavel, Emotion shapes the diffusion of moralized content in social networks, Proceedings of the National Academy of Sciences 114 (2017) 7313–7318. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1618923114. doi:10.1073/pnas.1618923114. arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1618923114.

[3] Z. Liu, T. Zhang, K. Yang, P. Thompson, Z. Yu, S. Ananiadou, Emotion detection for misinformation: A review, Information Fusion 107 (2024) 102300. URL: https://www.sciencedirect.com/science/article/pii/S1566253524000782. doi:https://doi.org/10.1016/j.inffus.2024.102300.

[4] O. J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, R. Avila, I. Babuschkin, S. Balaji, V. Balcom, P. Baltescu, H. ing Bao, M. Bavarian, J. Belgum, I. Bello, J. Berdine, G. Bernadett-Shapiro, C. Berner, L. Bogdonoff, O. Boiko, M. laine Boyd, A.-L. Brakman, G. Brockman, T. Brooks, M. Brundage, K. Button, T. Cai, R. Campbell, A. Cann, B. Carey, C. Carlson, R. Carmichael, B. Chan, C. Chang, F. Chantzis, D. Chen, S. Chen, R. Chen, J. Chen, M. Chen, B. Chess, C. Cho, C. Chu, H. W. Chung, D. Cummings, J. Currier, Y. Dai, C. Decareaux, T. Degry, N. Deutsch, D. Deville, A. Dhar, D. Dohan, S. Dowling, S. Dunning, A. Ecoffet, A. Eleti, T. Eloundou, D. Farhi, L. Fedus, N. Felix, S. P. Fishman, J. Forte, I. abella Fulford, L. Gao, E. Georges, C. Gibson, V. Goel, T. Gogineni, G. Goh, R. Gontijo-Lopes, J. Gordon, M. Grafstein, S. Gray, R. Greene, J. Gross, S. S. Gu, Y. Guo, C. Hallacy, J. Han, J. Harris, Y. He, M. Heaton, J. hannes Heidecke, C. Hesse, A. Hickey, W. Hickey, P. Hoeschele, B. Houghton, K. Hsu, S. Hu, X. Hu, J. Huizinga, S. Jain, S. Jain, J. Jang, A. Jiang, R. Jiang, H. Jin, D. Jin, S. Jomoto, B. Jonn, H. Jun, T. Kaftan, L. Kaiser, A. Kamali, I. Kanitscheider, N. S. Keskar, T. Khan, L. Kilpatrick, J. W. Kim, C. Kim, Y. Kim, H. Kirchner, J. R. Kiros, M. Knight, D. Kokotajlo, L. Kondraciuk, A. Kondrich, A. Konstantinidis, K. Kosic, G. Krueger, V. Kuo, M. Lampe, I. Lan, T. Lee, J. Leike, J. Leung, D. Levy, C. Li, R. Lim, M. Lin, S. Lin, M. teusz Litwin, T. Lopez, R. Lowe, P. Lue, A. Makanju, K. Malfacini, S. Manning, T. Markov, Y. Markovski, B. Martin, K. Mayer, A. Mayne, B. McGrew, S. M. McKinney, C. McLeavey, P. McMillan, J. McNeil, D. Medina, A. Mehta, J. Menick, L. Metz, A. Mishchenko, P. Mishkin, V. Monaco, E. Morikawa, D. P. Mossing, T. Mu, M. Murati, O. Murk, D. M'ely, A. Nair, R. Nakano, R. Nayak, A. Neelakantan, R. Ngo, H. Noh, O. Long, C. O'Keefe, J. W. Pachocki, A. Paino, J. Palermo, A. Pantuliano, G. Parascandolo, J. Parish, E. Parparita, A. Passos, M. Pavlov, A. Peng, A. Perelman, F. de Avila Belbute Peres, M. Petrov, H. P. de Oliveira Pinto, M. Pokorny, M. Pokrass, V. H. Pong, T. Powell, A. Power, B. Power, E. Proehl, R. Puri, A. Radford, J. W. Rae, A. Ramesh, C. Raymond, F. Real, K. Rimbach, C. Ross, B. Rotsted, H. Roussez, N. Ryder, M. D. Saltarelli, T. Sanders, S. Santurkar, G. Sastry, H. Schmidt, D. Schnurr, J. Schulman, D. Selsam, K. Sheppard, T. Sherbakov, J. Shieh, S. Shoker, P. Shyam, S. Sidor, E. Sigler, M. Simens, J. Sitkin, K. Slama, I. Sohl, B. Sokolowsky, Y. Song, N. Staudacher, F. P. Such, N. Summers, I. Sutskever, J. Tang, N. A. Tezak, M. Thompson, P. Tillet, A. Tootoonchian, E. Tseng, P. Tuggle, N. Turley, J. Tworek, J. F. C. Uribe, A. Vallone, A. Vijayvergiya, C. Voss, C. L. Wainwright, J. J. Wang, A. Wang, B. Wang, J. Ward, J. Wei, C. Weinmann, A. Welihinda, P. Welinder, J. Weng, L. Weng, M. Wiethoff, D. Willner, C. Winter, S. Wolrich, H. Wong, L. Workman, S. Wu, J. Wu, M. Wu, K. Xiao, T. Xu, S. Yoo, K. Yu, Q. ing Yuan, W. Zaremba, R. Zellers, C. Zhang, M. Zhang, S. Zhao, T. Zheng, J. Zhuang, W. Zhuk, B. Zoph, Gpt-4 technical report, 2023. URL: https://api.semanticscholar.org/CorpusID:257532815.

[5] A. Grattafiori, A. Dubey, A. Jauhri, A. Pandey, A. Kadian, A. Al-Dahle, A. Letman, A. Mathur, A. Schelten, A. Vaughan, et al., The llama 3 herd of models, arXiv preprint arXiv:2407.21783 (2024).

[6] G. Team, M. Riviere, S. Pathak, P. G. Sessa, C. Hardin, S. Bhupatiraju, L. Hussenot, T. Mesnard, B. Shahriari, A. Ramé, et al., Gemma 2: Improving open language models at a practical size, arXiv preprint arXiv:2408.00118 (2024).
[7] G. Team, R. Anil, S. Borgeaud, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth, K. Millican, et al., Gemini: a family of highly capable multimodal models, arXiv preprint arXiv:2312.11805 (2023).
[8] M. Abdin, J. Aneja, H. Behl, S. Bubeck, R. Eldan, S. Gunasekar, M. Harrison, R. J. Hewett, M. Javaheripi, P. Kauffmann, et al., Phi-4 technical report, arXiv preprint arXiv:2412.08905 (2024).
[9] C. Martel, G. Pennycook, D. G. Rand, Reliance on emotion promotes belief in fake news, Cognitive research: principles and implications 5 (2020) 1–20.
[10] E. F. Ayetiran, Özlem Özgöbek, An inter-modal attention-based deep learning framework using unified modality for multimodal fake news, hate speech and offensive language detection, Information Systems 123 (2024) 102378. URL: https://www.sciencedirect.com/science/article/pii/S030643792400036X. doi:https://doi.org/10.1016/j.is.2024.102378.
[11] R. Pan, J. A. García-Díaz, M. Ángel Rodríguez-García, R. Valencia-García, Spanish meacorpus 2023: A multimodal speech–text corpus for emotion analysis in spanish from natural environments, Computer Standards & Interfaces 90 (2024) 103856. URL: https://www.sciencedirect.com/science/article/pii/S0920548924000254. doi:https://doi.org/10.1016/j.csi.2024.103856.
[12] L. Barrault, Y.-A. Chung, M. C. Meglioli, D. Dale, N. Dong, M. Duppenthaler, P.-A. Duquenne, B. Ellis, H. Elsahar, J. Haaheim, et al., Seamless: Multilingual expressive and streaming speech translation, arXiv preprint arXiv:2312.05187 (2023).
[13] R. Pan, J. A. García-Díaz, M. Á. Rodríguez-García, F. García-Sanchez, R. Valencia-García, Overview of EmoSPeech 2024@IberLEF: Multimodal Speech-text Emotion Recognition in Spanish, Procesamiento del Lenguaje Natural 73 (2024).
[14] R. Pan, J. A. García-Díaz, R. Valencia-García, Spanish mtlhatecorpus 2023: Multi-task learning for hate speech detection to identify speech type, target, target group and intensity, Computer Standards & Interfaces 94 (2025) 103990. URL: https://www.sciencedirect.com/science/article/pii/S0920548925000194. doi:https://doi.org/10.1016/j.csi.2025.103990.
[15] L. Plaza, J. Carrillo-de Albornoz, V. Ruiz, A. Maeso, B. Chulvi, P. Rosso, E. Amigó, J. Gonzalo, R. Morante, D. Spina, Overview of exist 2024 – learning with disagreement for sexism identification and characterization in social networks and memes (extended overview, in: Working Notes of CLEF 2024- Conference and Labs of the Evaluation Forum, Springer, 2024.
[16] R. Pan, J. A. García-Díaz, T. Bernal-Beltrán, R. Valencia-García, Umuteam at exist 2024: Multimodal identification and categorization of sexism by feature integration, in: CEUR Workshop Proceedings, volume 3740, CEUR-WS, 2024, pp. 1135–1147.
[17] D. Dimitrov, F. Alam, M. Hasanain, A. Hasnat, F. Silvestri, P. Nakov, G. Da San Martino, Semeval-2024 task 4: Multilingual detection of persuasion techniques in memes, in: Proceedings of the 18th International Workshop on Semantic Evaluation, SemEval 2024, Mexico City, Mexico, 2024.
[18] R. Pan, J. A. García-Díaz, R. Valencia-García, Umuteam at semeval-2024 task 4: Multimodal identification of persuasive techniques in memes through large language models, in: SemEval 2024 - 18th International Workshop on Semantic Evaluation, Proceedings of the Workshop, Association for Computational Linguistics (ACL), 2024, pp. 655–666.
[19] S. Kumar, M. S. Akhtar, E. Cambria, T. Chakraborty, Semeval 2024 – task 10: Emotion discovery and reasoning its flip in conversation (ediref), in: Proceedings of the 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 2024. URL: https://arxiv.org/abs/2402.18944.
[20] R. Pan, J. A. García-Díaz, D. Roldán, R. Valencia-García, Umuteam at semeval-2024 task 10: Discovering and reasoning about emotions in conversation using transformers, in: SemEval 2024 - 18th International Workshop on Semantic Evaluation, Proceedings of the Workshop, Association for Computational Linguistics (ACL), 2024, pp. 703–709.
[21] R. S.-T. y Alba Bonet-Jover y Isam Diab y Ibai Guillén-Pacho y Isabel Cabrera-de Castro y Carlos Badenes-Olmedo y Estela Saquete y M. Teresa Martín-Valdivia y Patricio Martínez-Barco y

L. Alfonso Ureña-López, Overview of flares at iberlef 2024: Fine-grained language-based reliability detection in spanish news, Procesamiento del Lenguaje Natural 73 (2024) 369–379. URL: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6624.

[22] T. Mickus, E. Zosa, R. Vazquez, T. Vahtola, J. Tiedemann, V. Segonne, A. Raganato, M. Apidianaki, SemEval-2024 task 6: SHROOM, a shared-task on hallucinations and related observable overgeneration mistakes, in: A. K. Ojha, A. S. Doğruöz, H. Tayyar Madabushi, G. Da San Martino, S. Rosenthal, A. Rosá (Eds.), Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), Association for Computational Linguistics, Mexico City, Mexico, 2024, pp. 1979–1993. URL: https://aclanthology.org/2024.semeval-1.273/. doi:10.18653/v1/2024.semeval-1.273.

[23] R. Pan, J. A. García-Díaz, F. García-Sánchez, R. Valencia-García, Umuteam at flares@iberlef 2024: Enhancing disinformation detection with 5w1h techniques and transformer models, in: CEUR Workshop Proceedings, volume 3756, CEUR-WS, 2024.

[24] J. Piskorski, N. Stefanovitch, G. Da San Martino, P. Nakov, SemEval-2023 task 3: Detecting the category, the framing, and the persuasion techniques in online news in a multi-lingual setup, in: A. K. Ojha, A. S. Doğruöz, G. Da San Martino, H. Tayyar Madabushi, R. Kumar, E. Sartori (Eds.), Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 2343–2361. URL: https://aclanthology.org/2023.semeval-1.317/. doi:10.18653/v1/2023.semeval-1.317.

[25] R. Pan, J. A. García-Díaz, M. A. Rodríguez-García, R. Valencia-García, Umuteam at semeval-2023 task 3: Multilingual transformer-based model for detecting the genre, the framing, and the persuasion techniques in online news, in: 17th International Workshop on Semantic Evaluation, SemEval 2023 - Proceedings of the Workshop, Association for Computational Linguistics, 2023, pp. 609–615.

[26] M. Fröbe, B. Stein, T. Gollub, M. Hagen, M. Potthast, SemEval-2023 task 5: Clickbait spoiling, in: A. K. Ojha, A. S. Doğruöz, G. Da San Martino, H. Tayyar Madabushi, R. Kumar, E. Sartori (Eds.), Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023), Association for Computational Linguistics, Toronto, Canada, 2023, pp. 2275–2286. URL: https://aclanthology.org/2023.semeval-1.312/. doi:10.18653/v1/2023.semeval-1.312.

[27] R. Pan, J. A. García-Díaz, F. García-Sánchez, R. Valencia-García, Chick adams at semeval-2023 task 5: Using roberta and deberta to extract post and document-based features for clickbait spoiling, in: 17th International Workshop on Semantic Evaluation, SemEval 2023 - Proceedings of the Workshop, Association for Computational Linguistics, 2023, pp. 624–628.

[28] R. P. y José Antonio García-Díaz y Rafael Valencia-García, Optimizing few-shot learning through a consistent retrieval extraction system for hate speech detection, Procesamiento del Lenguaje Natural 74 (2025) 241–252. URL: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6678.

[29] R. Pan, J. Antonio García-Díaz, R. Valencia-García, Comparing fine-tuning, zero and few-shot strategies with large language models in hate speech detection in english, CMES - Computer Modeling in Engineering and Sciences 140 (2024) 2849–2868. URL: https://www.sciencedirect.com/science/article/pii/S1526149224000493. doi:https://doi.org/10.32604/cmes.2024.049631.