

Textual Analysis of Political Trust in Spanish

Tomás Bernal-Beltrán

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

Abstract

The widespread use of Information and Communication Technologies has introduced new challenges in the fight against disinformation; that is, false content spread intentionally to mislead, and misinformation spread unintentionally. Moreover, hyper-partisanship, an extreme form of political loyalty that prioritizes identity over substance, fosters a highly polarized environment where political trust depends more on who says something than what is said. These dynamics, often amplified by disinformation and major political events, highlight the urgent need for advanced solutions to mitigate their harmful effects. This thesis addresses the textual analysis of political trust in Spain by exploring the use of natural language technologies to detect and analyze misleading content in political news and discourse. It also addresses related issues such as satire detection, political hate speech and biased propaganda. To this end, we are developing linguistic resources to identify misinformation, promote fair political dialogue and analyze the profiles of different media outlets reporting on political issues.

Keywords

Misinformation and Disinformation Detection, Political Impartiality Analysis, Ideological Profiling by Clustering, Textual Analysis, Natural Language Processing

1. Introduction

The rise of Information and Communication Technologies (ICTs) has a significant impact on the political sphere, creating new challenges. One of the most pressing is the fight against misinformation and disinformation [1]. Misinformation refers to false or inaccurate information, such as rumors, insults, or jokes, shared without intent to deceive, while disinformation is a deliberate and malicious action, such as hoaxes, targeted phishing attacks and propaganda. Misinformation and disinformation increase distrust in society [2]. In recent years, deceptive content has played a key role in triggering major international events, such as the assault on the U.S. Capitol and the Brazilian National Congress, fueled by populism, sensationalism, fake news and denialism [3]. In Spain, growing public discontent and hyper-partisanship have increasingly hindered citizens' access to accurate information about laws concerning euthanasia, sexual freedom law, and real and effective equality for trans people and the protection of LGTBI rights. This context undermines citizens' right to fair and impartial information, weakening democracy and the welfare state. As a result, there is a growing social need to develop tools that can detect misinformation and disinformation in the political domain.

Despite the remarkable advances in Natural Language Processing (NLP) driven by Large Language Models [4] (LLMs), the application of these technologies still has important limitations, especially in sensitive domains such as political speech. Most LLMs are primarily trained on English-language data and general-purpose texts, which limits their effectiveness in multilingual contexts or in monolingual settings involving languages other than English, such as Spanish [5]. In addition, they often struggle to capture the cultural and contextual nuances specific to non-English speaking regions. Identifying malicious disinformation and misinformation, as well as other types of content such as satirical or hopeful discourse, is a major challenge. Making this distinction correctly requires a more nuanced contextual interpretation. Therefore, existing solutions for detecting ideological bias or hate speech in political speeches are still limited and do not have adequate linguistic resources in Spanish [6]. In addition, the rise of generative models has facilitated the rapid creation and dissemination of false information and hoaxes online, further complicating efforts to combat misinformation and disinformation.

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

✉ tomas.bernalb@um.es (T. Bernal-Beltrán)

id 0009-0006-6971-1435 (T. Bernal-Beltrán)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

This thesis addresses the challenge of political trust. Given the lack of adequate resources and mechanisms in Spanish to combat misinformation and disinformation in the political domain, thus mitigating their impact on public discourse through the design and implementation of innovative, efficient and open language technologies. Two main research hypotheses will be formulated. On the one hand, the detection of misinformation, disinformation, bias, propaganda, and toxic discourse can help foster hope and transparency in political communication; and on the other hand, the analysis of the profiles of different segments of the population can provide insights into how different groups perceive political actions (laws, speeches, policies) on key issues such as the economy, gender equality, the environment and employment. In order to validate these hypotheses, the thesis sets the following objectives: (1) the development of linguistic resources and language technologies to detect disinformation and misinformation in Spanish political discourse, (2) the creation of tools to identify fair, transparent and impartial political content in Spanish, and (3) the development of clustering-based profiling methods in the Spanish political landscape.

The remainder of this document is divided into the following sections. Section 2 describes what political trust is and why it is important, and explores the state of the art of techniques and resources available in Spanish to promote high quality content for citizens, guaranteeing their right to truthful, unbiased and impartial information, as well as citing relevant works within this topic. Section 3 details the proposed methodology, with special emphasis on the activities being carried out to achieve each of the proposed objectives to validate the research hypotheses. Finally, section 4 presents the final conclusions of the work presented in this document, as well as the further work to be developed in the course of this doctoral thesis.

2. Background Information

Although trust is a controversial term, there is limited consensus on its definition. Trust implies that a person voluntarily exposes himself or herself to the risk of being harmed or betrayed by another, be it an individual, a group or an institution. It is rarely unconditional, as it is usually placed on specific individuals or entities in specific contexts [7]. For example, citizens may trust their government to protect their lives in times of war, but distrust the bureaucratic management of resources in times of peace. Trust is a judgment that can be understood in binary terms (trust or distrust) or in degrees (more or less trust). In particular, political trust refers to the degree of credibility that citizens give to government institutions, political leaders and the democratic system as a whole. Trust is, therefore, fundamental for the proper functioning of a democracy, as it promotes social cooperation, institutional legitimacy and voluntary compliance. Yet in many democratic countries, with some exceptions, trust in key institutions and public figures is lagging or deteriorating. This trend reveals a growing dissatisfaction with the perceived effectiveness of policies and, more worryingly, with the overall functioning of democratic politics [8, 9].

LLMs are large-scale deep learning architectures trained on large text corpora and designed to perform a wide range of NLP tasks, including text generation, automatic document classification, summarization and translation [4]. state of the art LLMs are based on two fundamental principles: attention mechanisms and transfer learning capabilities [10]. On the one hand, attention allows the model to be context-aware by using embeddings to represent words, thus solving linguistic challenges such as polysemy and ambiguity. On the other hand, transfer learning allows pre-trained models to be adapted to solve specific tasks in other domains, improving their adaptability and performance across domains.

Early examples of LLMs include BERT, RoBERTa and ALBERT, whose initial versions were only available in English. These were followed by multilingual models and models specifically tailored to the Spanish language [11, 12]. More recently, generative LLMs such as GPT-4, Bloom, PaLM, XLNet and DeepSeek have been introduced, demonstrating the ability to produce long, coherent and contextually appropriate text, with performance approaching or exceeding human-level benchmarks in certain tasks [13]. Despite these advances, LLMs face two significant limitations. First, their high computational

requirements make them difficult to use in resource-constrained environments without dedicated hardware (e.g., GPUs or TPUs). Second, their black-box nature poses interpretability challenges. To mitigate this problem, researchers have proposed integrating LLMs with more transparent and interpretable feature representations [14, 15, 16].

As mentioned above, this thesis addresses the dual challenges of misinformation and disinformation in Spanish-language political communication by creating novel linguistic resources, mechanisms and analytical tools. This goal will be achieved through three sub-goals:

OB1. To develop and deploy linguistic resources and language technology tools capable of detecting both misinformation and disinformation in Spanish political discourse. A key challenge is to distinguish legitimate satire and parody from content deliberately designed to deceive. Although Spanish-specific resources remain scarce, our group has already demonstrated expertise in this area [17, 18, 19].

At the same time, misleading content will be addressed from two complementary perspectives. First, we will focus on the detection of hate speech, using state of the art methods such as those proposed in the following studies [20, 21, 22, 23, 24, 25, 26, 27]. Second, we will target the identification of hopeful or positive speech, following the conceptual frameworks and methods proposed in recent studies [28, 29].

OB2. To develop and deploy advanced NLP tools to assess the fairness, transparency and impartiality of Spanish political texts. In today’s society, social media has become the primary channel for information dissemination, democratizing access to knowledge but also bringing with it significant negative effects, such as information bubbles, the rapid spread of disinformation and misinformation, and the proliferation of toxic and hateful discourse, all of this fueled by algorithms optimized for engagement and the relative anonymity of online platforms. Identifying and mitigating these biases is essential to ensuring that citizens have access to balanced and trustworthy information. To address this challenge, we will apply state of the art methods, as outlined in recent work [30, 31, 32, 33].

In addition, we will create the first Spanish corpus to detect hyper-partisan content in political propaganda, examining biases such as fanaticism, religious extremism, discrimination and violent threats, and investigate techniques to automatically identify texts generated by large-scale linguistic models that pose an emerging risk of synthetic disinformation.

OB3. To perform clustering-based profiling methods in the Spanish political landscape. One of the main innovations of this thesis is the integration of cluster-based analysis. This approach will allow a comprehensive study of political discourse, from individual speakers to collective groups and the communication media. Our group has already demonstrated expertise in this area, as the most widely used resources for author profiling in Spanish include the datasets created by our research group as part of the PAN and PoliticES 2022 shared task [34], as well as PoliCorpus 2020 [15], which have been extensively used in recent studies, and PoliticES 2023 shared task [35] as part of the IberLEF 2023. In addition, we will adopt state of the art methodologies, as outlined in [36] and review the approaches adopted by the authors featured in this survey [37], which provides a comprehensive review of the techniques used to study political polarization, including author profiling.

To conclude this review of the state of the art, we highlight what is, to the best of our knowledge, the most comprehensive study of misinformation and misleading content in the Spanish political domain, which can be found at [38]. In this paper, the authors identify several key datasets that support research in this area. For fake news and rumor detection, notable datasets include *The Spanish Fake News Corpus*, *Verification Corpus*, *FTR-18* and *PAN-AP 2020 Fake News Spreaders Detection*. For stance classification, relevant resources include *TW-10*, *MultiStanceCat* and the *Catalonia Independence Corpus (CIC)*. These datasets provide valuable benchmarks for advancing research on misinformation and stance detection in Spanish-language political discourse.

3. Methodology and Experiments

This section describes the methodology and experiments developed to validate the research hypotheses of this thesis. As three specific objectives have been defined for this purpose, the section is organized into three subsections, each detailing the methodology and experiments related to one of these objectives.

3.1. OB1. Disinformation and Misinformation Detection

This objective consists in the development and deployment of linguistic resources and language technology tools capable of detecting both misinformation and disinformation in Spanish political discourse. To this end, a large corpus of Spanish political news articles is being created using a custom-built web crawler. This crawler systematically collects news articles from a variety of online media sources, focusing specifically on content related to politics and public discourse. Media sources have been selected with the aim of ensuring diversity in ideological orientation, geographical location and frequency of publication, thereby capturing a representative and heterogeneous set of political narratives. These selection of media sources include newspapers with a strong political focus, such as *ABC*, *El País* and *El Mundo*; verification (fact-checking) and fake news detection sites, such as *Maldita*; international news sites, including *BBC*, *Huffington post* and *Europa Press*; and Latin American news sites, such as *El Herald*, *Excelsior* and *Jornada*.

The dataset currently contains approximately one million articles from 32 different online media sources. These sources include regional, Spanish and international (mainly Latin American) newspapers, digital-only news portals and international media with dedicated political sections. The crawling process is designed to be continuous, allowing the corpus to grow over time and remain up to date with current political developments. This is particularly relevant in the context of disinformation and misinformation, as such phenomena are highly dynamic and often tied to specific news cycles, events, or electoral periods.

The content of each article is extracted using the JSON-LD metadata embedded in the web pages. This structured data format enables reliable and consistent retrieval of key fields such as article title, publication date, author (when available), and most importantly, the full text content of the article body. The use of JSON-LD enables scalable and automated data extraction without relying on site-specific scraping rules.

The main goal of this activity is to create a large, high-quality corpus of Spanish political news that can be made publicly available to the research community. This resource is intended to facilitate a wide range of NLP tasks, including but not limited to political stance detection, claim verification, source reliability classification and ultimately disinformation and misinformation detection.

While the current work focuses on the creation of the corpus, future research will address the use of this corpus for the design and evaluation of mechanisms to enable the detection of misinformation and disinformation in Spanish political news.

3.2. OB2. Identification of Fair, Transparent and Impartial Political Information

This objective consists in the development and use of advanced NLP tools to assess fairness, transparency and impartiality content in Spanish political texts. To achieve this goal, several complementary activities will be carried out, each of which aims to explore different dimensions of fairness, transparency and impartiality in political information.

The first activity contributes directly to the goal described in the previous subsection by providing a rich and diverse dataset of political news articles. The heterogeneity of sources and perspectives included in the corpus serves as a basis for analyzing potential ideological biases in media coverage and allows for the identification of content that may be more balanced or more biased towards certain political ideologies.

The second activity, which is currently underway, focuses on evaluating the behavior of the LLM when generating the content of a news article with content related to politics and public discourse. Specifically, the experiment consists of prompting various state of the art LLMs with the headline of a political news item and asking them to generate the body of the corresponding article. The generated texts are then analyzed to determine whether their tone, framing and content more closely resemble the narratives typically associated with left or right-wing media sources. This analysis allows researchers to assess the presence and degree of political bias that may have been inherited by the models during pre-training, thus revealing the extent to which the models produce content that can be considered

politically unbiased. The results of this experiment can provide valuable information about the latent ideological leanings of different LLMs, with implications for their responsible use in politically sensitive contexts.

The third activity, which is also ongoing, involves participation in collaborative tasks designed to advance research in the areas of information credibility, political stance detection and claim verification, areas closely related to the broader goal of ensuring fair, transparent and impartial dissemination of political information. We are currently participating in two major competitions: the TA1C task at IberLEF 2025 [39] and the CheckThat! task at CLEF 2025 [40]. The former consists in the detection and spoiling of a set of news items including clickbaits, the latter presents a set of diverse challenges aimed at advancing technology to support and improve the journalistic verification process, introducing subtasks such as subjectivity identification, claim normalization and fact-checking of numerical claims, with a particular focus on scientific web discourse. These tasks provide structured challenges and annotated datasets that serve as both benchmarking opportunities and sources of methodological inspiration. The knowledge and skills gained through participation in these initiatives are directly applicable to identifying and promoting politically fair, transparent and impartial content.

3.3. OB3. Detecting Profiles by Clustering

This objective consists of the development and application of clustering-based profiling methods in the Spanish political landscape. To achieve this goal, we explore the use of clustering techniques to analyze ideological tendencies in news media. Rather than focusing on profiling individuals or user groups, the goal is to study how different news sources convey political information, with the broader aim of understanding how media framing and editorial lines influence the transmission of socio-political narratives. For example, by grouping articles from different media sources and analyzing their content, it may be possible to discover latent ideological orientations or stylistic tendencies in their reporting. This would allow the development of methods capable of identifying political profiles at the institutional level, such as distinguishing whether a given narrative is more closely associated with left- or right-wing media. This approach maintains the focus on clustering and profiling while avoiding direct classification of individual beliefs or identities. It is also consistent with broader research goals in media bias detection and transparency in information ecosystems.

This goal is inspired by the shared task PoliticES, organized in IberLEF 2023 [35], which aims to explore the extraction of socio-political profiles from collections of texts. Specifically, the goal of the task is to profile clusters of texts, rather than individual users, to identify characteristics such as gender, profession and political ideology. This approach is intended to address the ethical and legal issues often associated with user-level profiling, such as invasion of privacy or the possibility of discriminatory applications.

However, even in aggregated or anonymized contexts, research on profiling sensitive attributes; such as political ideology or emotional state, raises significant ethical concerns. These issues are particularly relevant in light of emerging AI regulations that call for the responsible and transparent use of machine learning models. There is a risk that such technologies could be misused to overgeneralize, reinforce biases, or infer information that individuals have not chosen to disclose.

In light of these concerns, this thesis deliberately frames its objective in a way that prioritizes ethical considerations. Rather than developing models to classify or infer personal beliefs, the focus is on understanding how ideological narratives emerge and propagate through media content. The intention is to help identify media bias and to increase transparency in information ecosystems, not to support surveillance or profiling of individuals. The ultimate goal is educational and analytical: to help illuminate the role of media institutions in shaping political discourse.

4. Conclusions and Further Work

This doctoral thesis aims to investigate, on the one hand, the detection of disinformation and misinformation and, on the other hand, the detection of fair, transparent and unbiased information with the

objective of applying it in the political domain to promote high quality content for citizens, guaranteeing their right to truthful, impartial and unbiased information and increasing political trust.

To achieve this main goal, first, the current state of the art in this field was reviewed. Second, a large corpus of Spanish political news articles will be created and made publicly available to the research community. This resource is intended to facilitate a wide range of NLP tasks, including the detection of disinformation and misinformation in Spanish political news. Third, we evaluate the behavior of LLMs when generating the content of a news article with content related to politics and public discourse. Analyzing the model's response to determine whether its tone, framing and content more closely resemble narratives typically associated with left- or right-wing media sources. Attempt to assess the presence and degree of political bias that models may have inherited during pre-training, thereby revealing the extent to which models produce content that can be considered politically unbiased. Fourth, we participate in different shared tasks of different evaluation forums such as CLEF and IberLEF, in shared tasks designed to advance research in the areas of information credibility, political stance detection and claim verification, areas closely related to the broader objective of ensuring fair, transparent and unbiased dissemination of political information, acquiring key knowledge and skills for the realization of the objectives of this thesis. Fifth, we plan to study the ideological framing and linguistic patterns of different news sources through clustering techniques that group articles from different media sources and analyze their content, which will allow us to discover latent ideological orientations or stylistic tendencies in their information. This would make it possible to develop methods capable of detecting political profiles at the institutional level, such as distinguishing whether a given narrative is more aligned with left-wing or right-wing media.

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. Mr. Tomás Bernal-Beltrán is supported by University of Murcia through the predoctoral programme.

Declaration on Generative AI

During the preparation of this work, the author used DeepL for grammatical and spelling correction. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

- [1] D. Colins, Disinformation and “fake news”: Interim report, UK House of Commons Digital (2017).
- [2] P. N. Petratos, Misinformation, disinformation, and fake news: Cyber risks to business, *Business Horizons* 64 (2021) 763–774.
- [3] A. Nai, F. Martínez i Coma, J. Maier, Donald trump, populism, and the age of extremes: Comparing the personality traits and campaigning styles of trump and other leaders worldwide, *Presidential Studies Quarterly* 49 (2019) 609–643.
- [4] C. Veres, Large language models are not models of natural language: they are corpus models, *IEEE Access* 10 (2022) 61970–61979.
- [5] Y. Jiang, J. Hao, M. Fauss, C. Li, Detecting chatgpt-generated essays in a large-scale writing assessment: Is there a bias against non-native english speakers?, *Computers & Education* 217 (2024) 105070.
- [6] A. Fernández-Roldán, C. Elías, C. Santiago-Caballero, D. Teira, Can we detect bias in political fact-checking? evidence from a spanish case study, *Journalism practice* (2023) 1–19.

- [7] M. Levi, L. Stoker, Political trust and trustworthiness, *Annual Review of Political Science* 3 (2000) 475–507. URL: <https://www.annualreviews.org/content/journals/10.1146/annurev.polisci.3.1.475>. doi:<https://doi.org/10.1146/annurev.polisci.3.1.475>.
- [8] D. Devine, Does political trust matter? a meta-analysis on the consequences of trust, *Political Behavior* 46 (2024) 2241–2262.
- [9] E. OUATTARA, T. VAN DER MEER, Distrusting democrats: A panel study into the effects of structurally low and declining political trust on citizens’ support for democratic reform, *European Journal of Political Research* 62 (2023) 1101–1121. URL: <https://ejpr.onlinelibrary.wiley.com/doi/abs/10.1111/1475-6765.12561>. doi:<https://doi.org/10.1111/1475-6765.12561>. arXiv:<https://ejpr.onlinelibrary.wiley.com/doi/pdf/10.1111/1475-6765.12561>.
- [10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, *Advances in neural information processing systems* 30 (2017).
- [11] A. Gutiérrez-Fandiño, J. Armengol-Estapé, M. Pàmies, J. Llop-Palao, J. Silveira-Ocampo, C. P. Carrino, A. Gonzalez-Agirre, C. Armentano-Oller, C. Rodriguez-Penagos, M. Villegas, Maria: Spanish language models, arXiv preprint arXiv:2107.07253 (2021).
- [12] J. Cañete, G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, J. Pérez, Spanish pre-trained bert model and evaluation data, arXiv preprint arXiv:2308.02976 (2023).
- [13] K. Singhal, S. Azizi, T. Tu, S. S. Mahdavi, J. Wei, H. W. Chung, N. Scales, A. Tanwani, H. Cole-Lewis, S. Pfohl, et al., Large language models encode clinical knowledge, *Nature* 620 (2023) 172–180.
- [14] J. A. García-Díaz, M. Cánovas-García, R. Valencia-García, Ontology-driven aspect-based sentiment analysis classification: An infodemiological case study regarding infectious diseases in latin america, *Future Generation Computer Systems* 112 (2020) 641–657.
- [15] J. A. García-Díaz, R. Colomo-Palacios, R. Valencia-García, Psychographic traits identification based on political ideology: An author analysis study on spanish politicians’ tweets posted in 2020, *Future Generation Computer Systems* 130 (2022) 59–74.
- [16] N. Du, Y. Huang, A. M. Dai, S. Tong, D. Lepikhin, Y. Xu, M. Krikun, Y. Zhou, A. W. Yu, O. Firat, et al., Glam: Efficient scaling of language models with mixture-of-experts, in: *International conference on machine learning*, PMLR, 2022, pp. 5547–5569.
- [17] J. A. García-Díaz, R. Valencia-García, Compilation and evaluation of the spanish satcorpus 2021 for satire identification using linguistic features and transformers, *Complex & Intelligent Systems* 8 (2022) 1723–1736.
- [18] J. A. García-Díaz, R. Valencia-García, Umuteam at semeval-2021 task 7: Detecting and rating humor and offense with linguistic features and word embeddings, in: *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, 2021, pp. 1096–1101.
- [19] J. A. Garcia-Díaz, R. Valencia-Garcia, Umuteam at haha 2021: Linguistic features and transformers for analysing spanish humor. the what, the how, and to whom, in: *Proceedings of the Iberian Languages Evaluation Forum (Iber-LEF 2021)*, CEUR Workshop Proceedings, Málaga, Spain, volume 9, 2021, pp. –.
- [20] V. Basile, C. Bosco, E. Fersini, D. Nozza, V. Patti, F. M. R. Pardo, P. Rosso, M. Sanguinetti, Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in twitter, in: *Proceedings of the 13th international workshop on semantic evaluation*, 2019, pp. 54–63.
- [21] F. M. Plaza-del Arco, M. Casavantes, H. J. Escalante, M. T. Martín-Valdivia, A. Montejo-Ráez, M. Montes, H. Jarquín-Vásquez, L. Villaseñor-Pineda, et al., Overview of meoffendes at iberlef 2021: Offensive language detection in spanish variants, *Procesamiento del Lenguaje Natural* 67 (2021) 183–194.
- [22] M. E. Aragón, H. J. Jarquín-Vásquez, M. Montes-y Gómez, H. J. Escalante, L. V. Pineda, H. Gómez-Adorno, J. P. Posadas-Durán, G. Bel-Enguix, Overview of mex-a3t at iberlef 2020: Fake news and aggressiveness analysis in mexican spanish., in: *IberLEF SEPLN*, 2020, pp. 222–235.
- [23] E. Fersini, P. Rosso, M. Anzovino, et al., Overview of the task on automatic misogyny identification at ibereval 2018., *Ibereval sepln* 2150 (2018) 214–228.
- [24] F. Rodríguez-Sánchez, J. Carrillo-de Albornoz, L. Plaza, J. Gonzalo, P. Rosso, M. Comet, T. Donoso, Overview of exist 2021: sexism identification in social networks, *Procesamiento del Lenguaje*

Natural 67 (2021) 195–207.

- [25] L. Arellano, H. J. Escalante, L. V. Pineda, M. M. Gomez, F. S. Vega, Overview of da-vincis at iberlef 2022:: Detection of aggressive and violent incidents from social media in spanish, *Procesamiento del lenguaje natural* (2022) 207–215.
- [26] A. Ariza-Casabona, W. S. Schmeisser-Nieto, M. Nofre, M. Taulé, E. Amigó, B. Chulvi, P. Rosso, Overview of detests at iberlef 2022: Detection and classification of racial stereotypes in spanish, *Procesamiento del lenguaje natural* 69 (2022) 217–228.
- [27] J. A. García-Díaz, S. M. Jiménez-Zafra, M. A. García-Cumbreras, R. Valencia-García, Evaluating feature combination strategies for hate-speech detection in spanish using linguistic features and transformers, *Complex & Intelligent Systems* 9 (2023) 2893–2914.
- [28] B. R. Chakravarthi, V. Muralidaran, R. Priyadharshini, S. Cn, J. P. McCrae, M. Á. García, S. M. Jiménez-Zafra, R. Valencia-García, P. Kumaresan, R. Ponnusamy, et al., Overview of the shared task on hope speech detection for equality, diversity, and inclusion, in: *Proceedings of the second workshop on language technology for equality, diversity and inclusion*, 2022, pp. 378–388.
- [29] D. García-Baena, M. Á. García-Cumbreras, S. M. Jiménez-Zafra, J. A. García-Díaz, R. Valencia-García, Hope speech detection in spanish: The lgbt case, *Language Resources and Evaluation* 57 (2023) 1487–1514.
- [30] F.-J. Rodrigo-Ginés, Automated media bias detection: Challenges and opportunities., *PLN-DS@SEPLN* (2023) 86–94.
- [31] J. Sánchez-Junquera, On the detection of political and social bias (2021).
- [32] C. Bosco, V. Patti, S. Frenda, A. T. Cignarella, M. Paciello, F. D’Errico, Detecting racial stereotypes: An italian social media corpus where psychology meets nlp, *Information Processing & Management* 60 (2023) 103118.
- [33] J. Sánchez-Junquera, P. Rosso, M. Montes, S. P. Ponzetto, Masking and transformer-based models for hyperpartisanship detection in news, in: *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, 2021, pp. 1244–1251.
- [34] J. A. García-Díaz, S. M. Jiménez-Zafra, M.-T. M. Valdivia, F. García-Sánchez, L. A. Ureña-López, R. Valencia-García, Overview of politices 2022: Spanish author profiling for political ideology, *Procesamiento del Lenguaje Natural* 69 (2022) 265.
- [35] J. A. G.-D. y Salud María Jiménez-Zafra y María-Teresa Martín-Valdivia y Francisco García-Sánchez y Luis Alfonso Ureña-López y Rafael Valencia-García, Overview of politices at iberlef 2023: Political ideology detection in spanish texts, *Procesamiento del Lenguaje Natural* 71 (2023) 409–416. URL: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6570>.
- [36] I. U. Khan, M. U. Khan, Social media profiling for political affiliation detection, *Human-Centric Intelligent Systems* 4 (2024) 437–446.
- [37] R. Németh, A scoping review on the use of natural language processing in research on political polarization: trends and research prospects, *Journal of computational social science* 6 (2023) 289–313.
- [38] E. Providel, M. Mendoza, Misleading information in spanish: a survey, *Social Network Analysis and Mining* 11 (2021) 1–26.
- [39] G. Mordecki, G. Moncecchi, J. Couto, Te ahorré un click: A revised definition of clickbait and detection in spanish news, in: *Ibero-American Conference on Artificial Intelligence*, Springer, 2024, pp. 387–399.
- [40] F. Alam, J. M. Struß, T. Chakraborty, S. Dietze, S. Hafid, K. Korre, A. Muti, P. Nakov, F. Ruggeri, S. Schellhammer, et al., The clef-2025 checkthat! lab: Subjectivity, fact-checking, claim normalization, and retrieval, in: *European Conference on Information Retrieval*, Springer, 2025, pp. 467–478.