

Analyzing Women’s Representation and Gender Bias in AI Media Coverage

Jorge Garcelán
Signal Theory and
Communications Department
Universidad Carlos III de Madrid
Madrid, Spain
jgarcela@pa.uc3m.es

Carmen Peláez-Moreno
Signal Theory and
Communications Department
Universidad Carlos III de Madrid
Madrid, Spain
carmen@tsc.uc3m.es

Teresa Onorati
Computer Science Department
Universidad Carlos III de Madrid
Madrid, Spain
tonorati@inf.uc3m.es

ABSTRACT

In a context where Artificial Intelligence (AI) is gaining social, economic, and political relevance, it is essential to examine how media narratives around this technology are constructed, particularly with regard to the role of women as experts, protagonists, or information sources. This work aims to analyze the representation of women and the potential gender biases in AI coverage in the Spanish mainstream press.

The analysis combines quantitative and qualitative methodologies, supported by advanced natural language processing (NLP) techniques and the use of state-of-the-art large language models (LLMs).

KEYWORDS

AI, gender bias, media, natural language processing

1 INTRODUCTION

AI now features in almost every aspect of daily life, and how it is shown in the press shape how people understand it. Yet women’s voices are still too often missing in its coverage. [1] found that popular language models link female names to caregiving roles three to six times more often than male names. Moreover, [2] showed that over 75 percent of quoted experts in major newswire articles are men. This imbalance not only sidelines women’s contributions but also reinforces stereotypical ideas about who “belongs” in the AI field.

As noted in [3], fully automated solutions can introduce safety, ethical, and legal risks, while fully manual approaches may be both error-prone and time-consuming. Therefore, this work proposes AI as a helpful tool that augments rather than replaces human judgement. To put this into practice, we are developing an interactive visualization tool that will allow to explore the analyzed data, facilitating journalists and researchers in examining AI news from a gender perspective.

2 METHODOLOGY

2.1 Data Collection

The study’s core dataset comprises articles published throughout 2024 in major Spanish national newspapers: *El País*, *El Mundo*, *Infolibre*, and *La Vanguardia*. To gather and verify AI-related content across these outlets, three specific search terms were defined: “AI”, “artificial intelligence”, and “the AI”. These keywords were used to perform targeted queries on each newspaper’s website via a custom Python script. Each retrieved record includes a verification field: titles, article bodies, and subtitles were extracted and checked to ensure the presence of at least one keyword, thus confirming that the articles genuinely belong to artificial intelligence.

2.2 Data Annotation

A dedicated team of experts will then manually annotate this corpus to validate our findings and enhance the performance of our AI models. Since the 2024 collection is still in its final stages of aggregation and preprocessing, we are initially working with an analogous, fully annotated 2023 dataset to develop our methods.

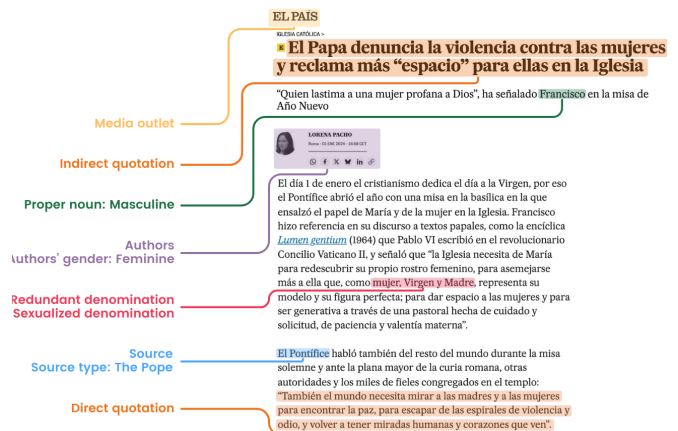


Figure 1. Example of annotated article

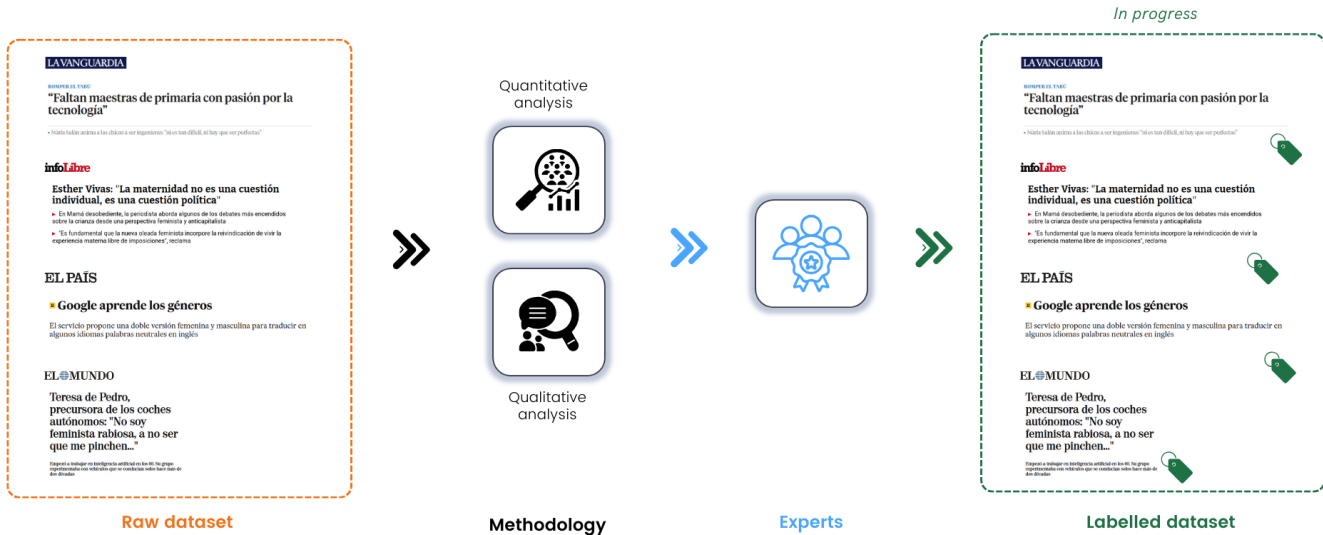


Figure 2. Pipeline of the creation of the labelled dataset

In these datasets, the variables under study have been grouped into several blocks:

- **Block I and II - Basic Information**, cover attributes such as the media outlet, headline, authors, topic and the journalist's gender.
- **Block III - Discourse and Language Analysis**, identifies instances of sexist language, the use of visible dual forms, androcentric framing, infantilization of women and dependent denominations, among other linguistic markers.
- **Block IV - Photographs**, examines whether an article contains images, the balance of their composition, planar and color asymmetries, and the gender of the subjects.
- **Block V - Sources**, records the presence of quoted statements, the name and biographical details of each person cited, the type of source, and the source's gender.

At present, our focus and emphasis is on language; consequently, we are initially exploring Blocks I and II, III, and V.

2.3 Analysis

Following [4] two approaches for the analysis will be conducted. In the quantitative phase, we will examine the presence and roles of women in relation to variables such as media type, news topic, and the roles ascribed to them. The qualitative phase will focus on identifying and interpreting specific fragments where biases, narrative patterns, or gender stereotypes emerge, offering a deeper understanding of the discourse.

3 PRELIMINARY RESULTS

As mentioned, this is a work in progress and final results are forthcoming. In the meantime, we have been stress-testing our pipeline on the fully annotated 2023 dataset. In particular, we are working with a hybrid methodology that combines traditional NLP tools with state-of-the-art LLMs: we use Flair's **ner-spanish-large** [5] for named-entity recognition, **genderize.io** for gender inference, and **OpenAI's models** to perform zero-shot validation and consistency checks. Thanks to this approach,

Blocks I and II are now almost fully automated. Blocks III and V, however, demand deeper semantic understanding, so our models still struggle with those. We are therefore using the aforementioned models for preliminary processing and await full annotation of the 2024 corpus. Once it is available, we will fine-tune specialized models to tackle these more complex tasks.

4 INTERACTIVE WEB APPLICATION

To maximize the impact and accessibility of our research, we are developing an interactive web application that allows users to explore and analyze texts in real time. This application is structured on a three-tier architecture (web, API, and database) and is being built using the Flask framework.

In particular, AI models trained on the described dataset will produce an initial analysis, which will then be refined through a human-in-the-loop workflow [3]. By doing so, researchers retain full control by reviewing, editing, and approving the AI's output before it is finalized. Moreover, the platform will be designed to implement automated report generation so that insights can be delivered in a clear, easily accessible format.

REFERENCES

- [1] H. Kotek, R. Dockum, and D. Q. Sun. 2023. *Gender bias and stereotypes in large language models*. *arXiv preprint arXiv:2308.14921* [cs.CL].
- [2] S. Zhao, H. Zhao, and H. Schütze. 2020. Analyzing gender bias in newswire corpora. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL '20)*. Association for Computational Linguistics, 5816–5828. DOI:https://doi.org/10.18653/v1/2020.acl-main.519.
- [3] V. Lai, C. Chen, Q. V. Liao, A. Smith-Renner, and C. Tan. 2021. *Towards a science of human-AI decision making: a survey of empirical studies*. *arXiv preprint arXiv:2112.11471*.
- [4] C. Sainz de Baranda Andújar, M. P. Diezhandino Nieto, Á. Bahamonde Magro, and UC3M Department of Journalism and Audiovisual Communication. 2013. *Mujeres y deporte en los medios de comunicación: estudio de la prensa deportiva española (1979–2010)*.
- [5] S. Schweter and A. Akbik. 2020. *FLERT: Document-Level Features for Named Entity Recognition*. *arXiv:2011.06993* [cs.CL].