

MuSeD: A Multimodal Spanish Dataset for Sexism Detection in Social Media Videos



Laura De Grazia, Pol Pastells, Mauro Vázquez Chas, Desmond Elliott, Danae Sánchez Villegas, Mireia Farrús, Mariona Taulé



Sexism in Social Media

Research on sexist content in social media has mostly focused on text¹ (X, Reddit). However, sexism can also be conveyed through **multiple modalities**.

Multimodal studies have mainly examined memes² (text & image), while **sexism in videos** (text, speech, image) is still an emerging area of investigation.

Most work targets women, with little attention to sexism related to sexual orientation and gender identity.³



Motivation



It's time to admit it, men have privileges over women, both socially and legally. We have six more years of punishment for the same crime committed by a woman and twice the chances of being incarcerated.

We need **high-quality datasets** to automatically detect multimodal sexism in order to:

- Support the work of content moderators.⁴
- Identify how sexism appears across modalities.
- Recognize both explicit and implicit sexism.

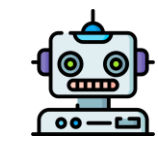
Research Questions



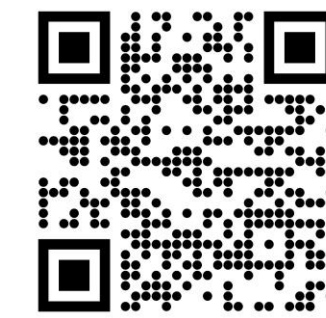
RQ1: What is the contribution of multimodal information to the identification of content as sexist or non-sexist?



RQ2: Does an extended definition of sexism improve its recognition?



RQ3: To what extent are humans and models capable of detecting explicit and implicit sexism?



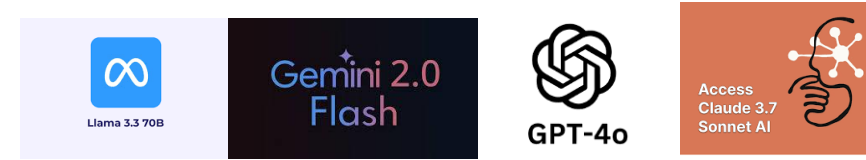
- We introduce MuSeD, the first Spanish dataset which includes sexism towards women, non-binary individuals and LGBTQ+ community.
- MuSeD is annotated across text, audio, and video, enabling a comparison of how different modalities contributes to sexism classification.
- We evaluate LLMs using both basic and detailed prompts, which incorporate the definition of sexism provided to annotators.
- Multimodal models outperform text-only models but still struggle with implicit sexism, which often relies on cultural context.

TL;DR

Evaluation of LLMs on Sexism Detection

Models Evaluated

- Textual and Multimodal LLMs



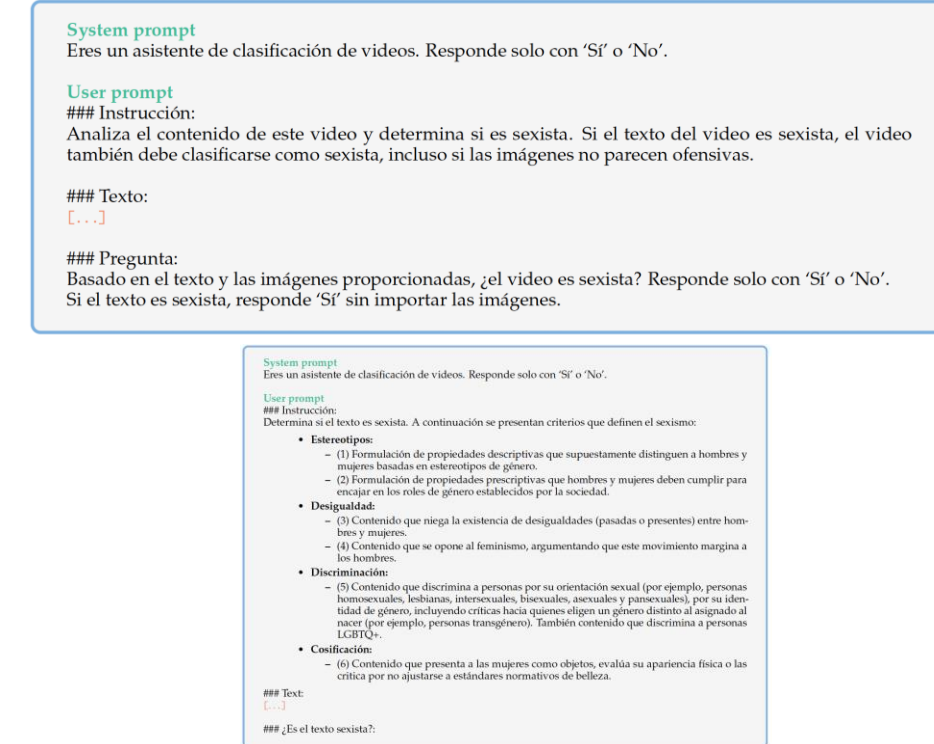
- Compare the performance of multimodal models

Prompts

Basic:
Is the video sexist?

Definition:
Is the video sexist? Criteria:

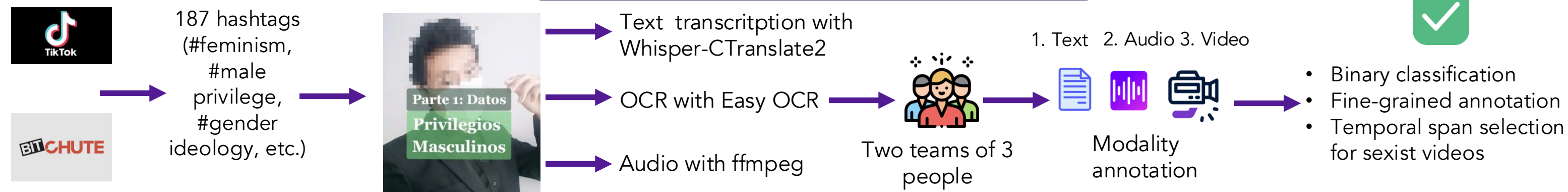
- Stereotype
- Denial of inequality
- Discrimination against LGBTQ+ community
- Objectification



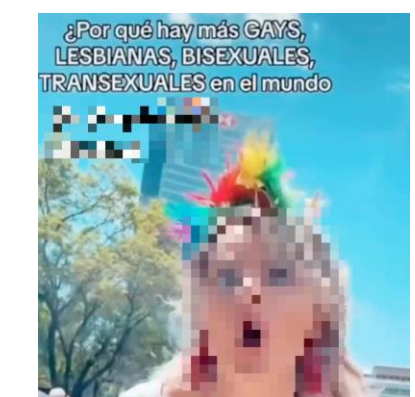
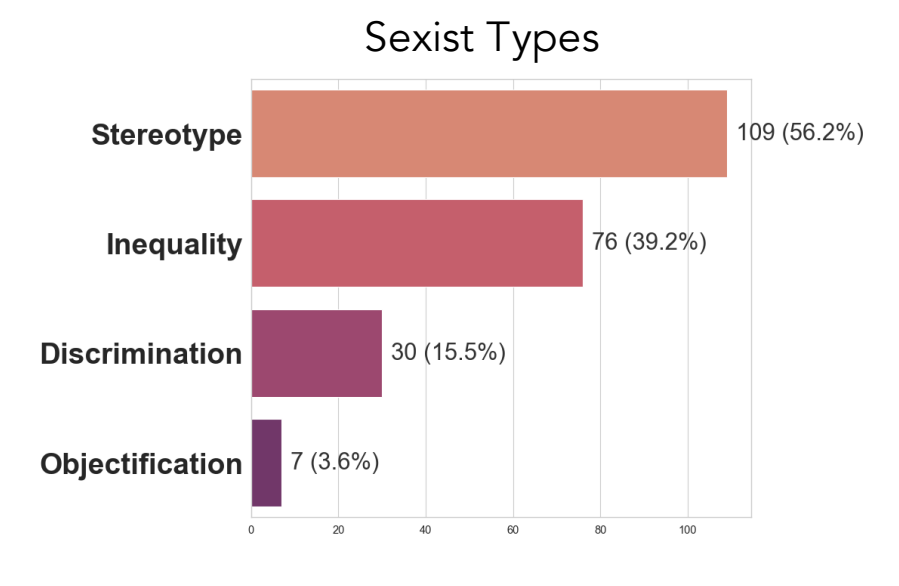
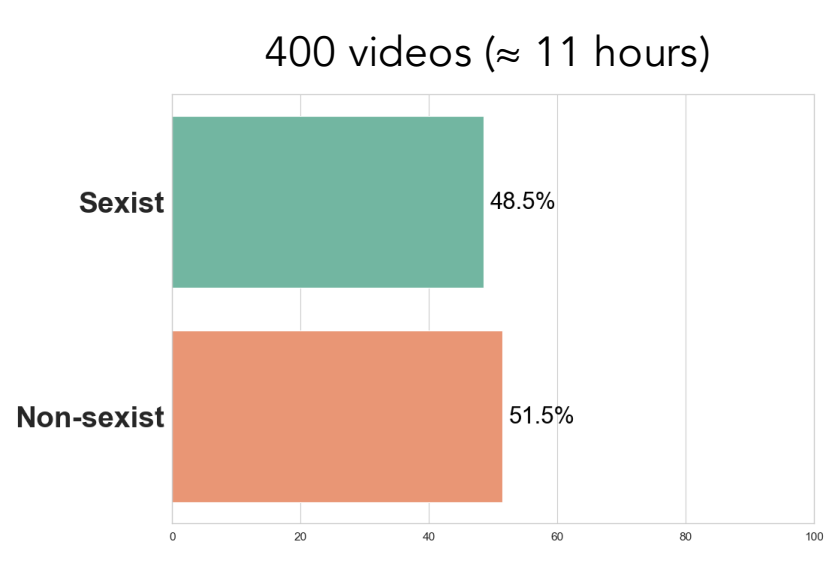
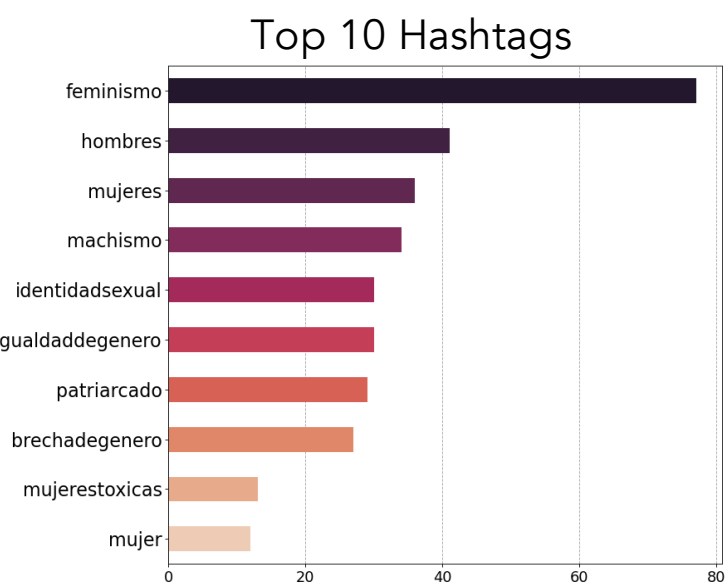
Zero-shot Experiment

- Ground truth: human labels
- Textual and Multimodal LLMs receive as inputs:

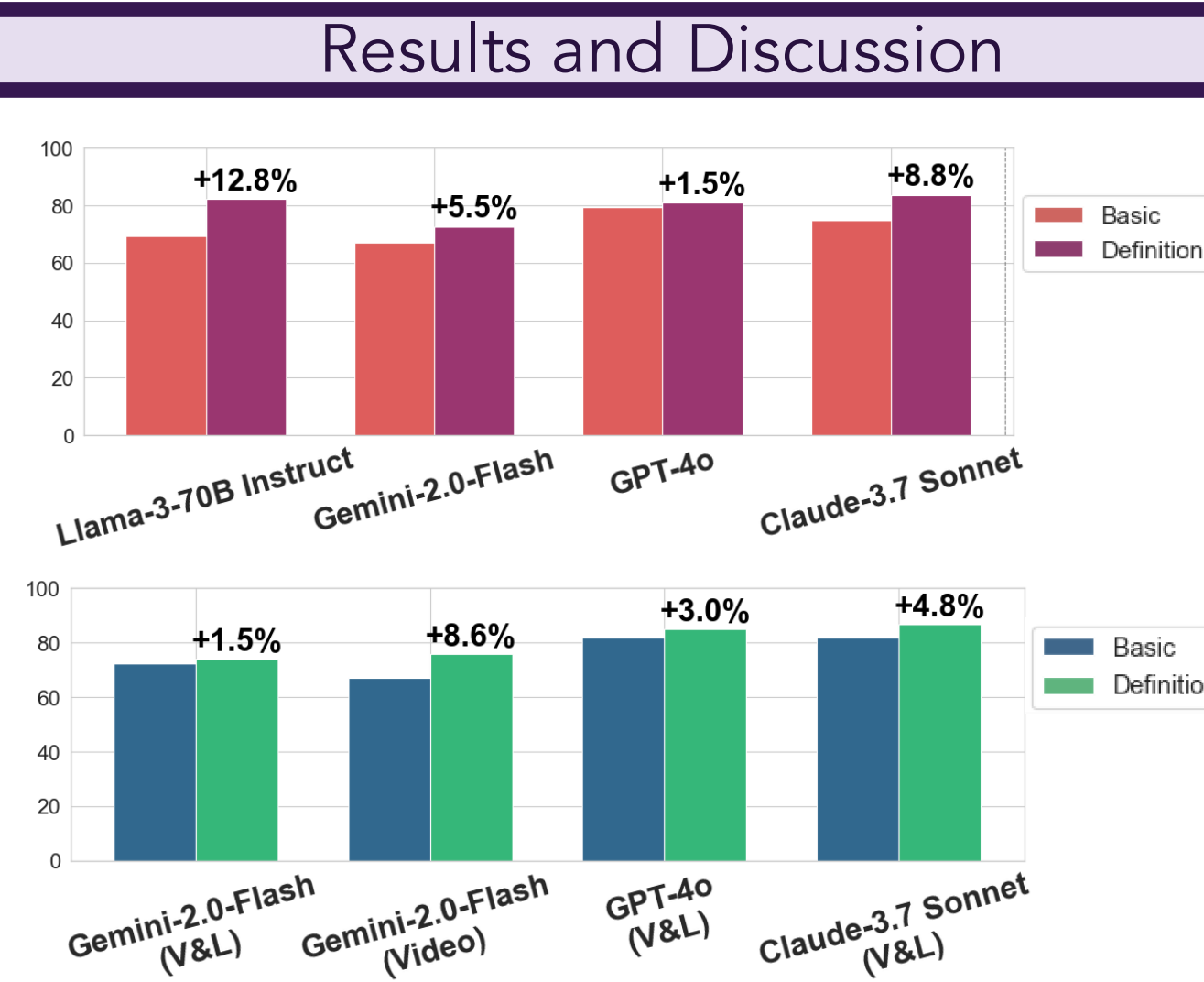
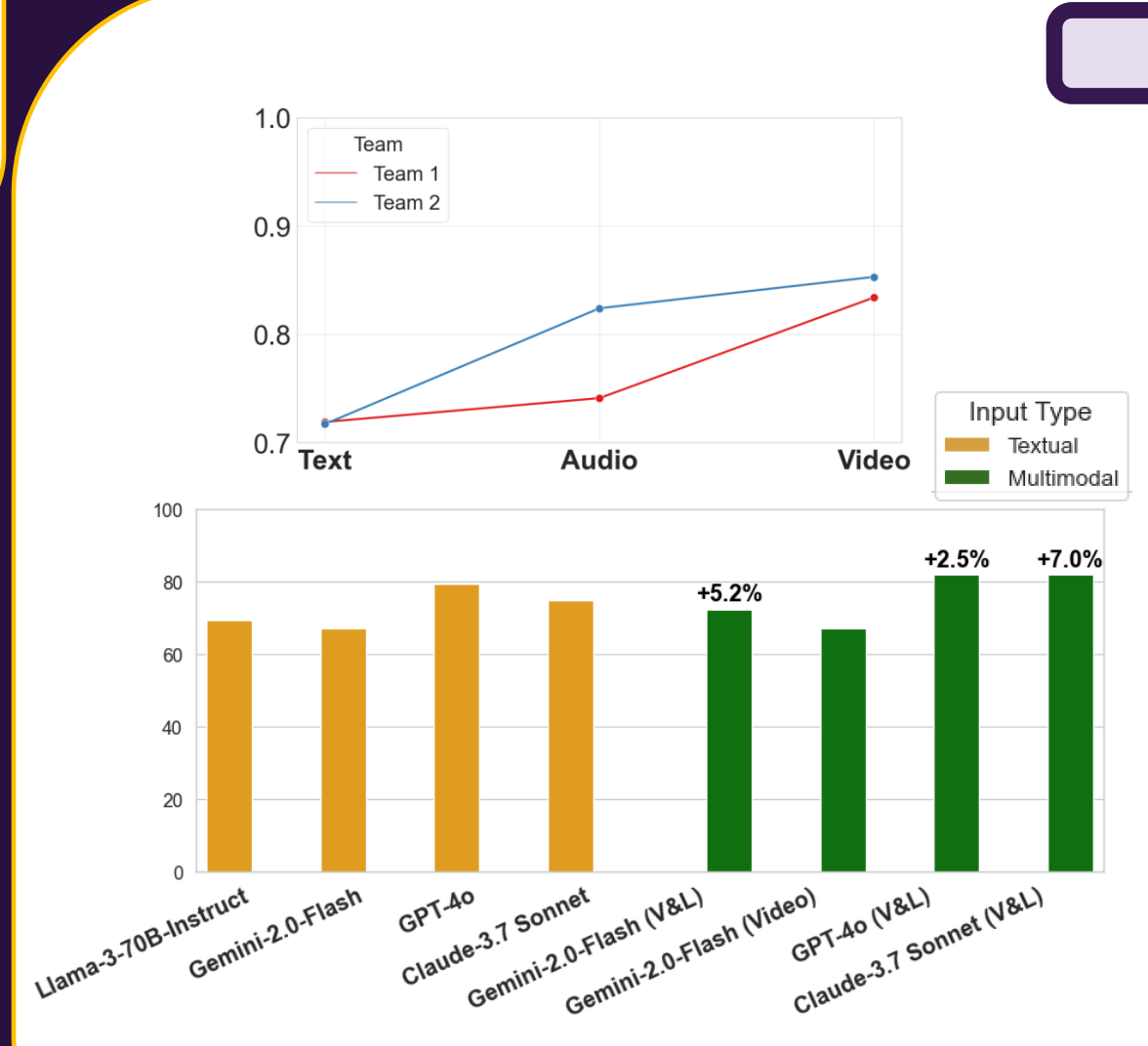
Data Collection & Annotation



MuSeD Dataset



Discrimination
Why are there more gays, lesbians, bisexuals, and transsexual in the world? Because they were sold the lie of sexual identity.



Alpha. Legs apart. Male crossing. The front. Leaning back. The head. Arms relaxed.

References
 [1] Samory, Mattia, et al. "Call me sexist, but...": Revisiting sexism detection using psychological scales and adversarial samples." Proceedings of the international AAAI conference on web and social media. Vol. 15. 2021.
 [2] Rizzi, Giulia, et al. "Recognizing misogynous memes: Biased models and tricky archetypes." Information Processing & Management 60.5 (2023): 103474.
 [3] Luo, Xuan, et al. "A Literature Survey on Multimodal and Multilingual Sexism Detection." IEEE Transactions on Computational Social Systems (2025).
 [4] Casey Newton. Facebook will pay \$52 million in settlement with moderators who developed PTSD on the job. The Verge, 12, 2020.