Actes de CORIA-TALN 2026
Nantes, France, 29 juin au 3 juillet 2026

MuSeD: un dataset multimodal en espagnol pour la détection du sexisme dans les vidéos sur les réseaux sociaux

Laura De Grazia[1] Pol Pastells [1] Mauro Vázquez Chas[1] Desmond Elliott[2]
Danae Sánchez Villegas[2] Mireia Farrús[1] Mariona Taulé[1]
(1) University of Barcelona, CLiC-Language and Computing Center
(2) University of Copenhagen, Department of Computer Science
lauradegrazia@ub.edu

RÉSUMÉ

Le sexisme est une forme de discrimination systémique dont la diffusion est amplifiée par les réseaux sociaux, notamment à travers les vidéos. Sa détection automatique est complexe car elle nécessite d'intégrer des signaux verbaux, sonores et visuels. Dans cette étude : (1) nous introduisons MuSeD, un corpus multimodal en espagnol pour la détection du sexisme (≈11 heures de vidéos collectées sur TikTok et BitChute) ; (2) nous proposons un cadre d'annotation analysant l'apport des modalités textuelles, vocales et visuelles à la classification du sexisme ; (3) nous évaluons plusieurs grands modèles de langage (LLMs), unimodaux et multimodaux, sur cette tâche. Nos résultats indiquent que la modalité visuelle est déterminante pour identifier le sexisme. Si les modèles obtiennent de bonnes performances pour les formes explicites, leurs performances diminuent pour les cas implicites (les stéréotypes) où l'accord inter-annotateurs est plus faible. Cela souligne la complexité de la tâche, dépendante du contexte socioculturel.

ABSTRACT

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

Sexism is a systemic form of discrimination amplified by social media, where it increasingly spreads through videos. Its automatic detection is complex as it requires integrating verbal, audio, and visual cues. In this study : (1) we introduce MuSeD, a Multimodal Spanish dataset for sexism detection (≈11 hours of videos from TikTok and BitChute) ; (2) we propose an annotation framework analyzing the contributions of textual, vocal, and visual modalities to sexism classification ; (3) we evaluate several LLMs and multimodal LLMs on this task. Our results show that visual information is crucial for identifying sexism. While models effectively detect explicit sexism, they struggle with implicit cases (stereotypes), where inter-annotator agreement is lower. This underscores the complexity of the task, which remains dependent on sociocultural context.

MOTS-CLÉS : détection du sexisme, analyse multimodale, jeu de données, vidéos des réseaux sociaux, annotation, accord inter-annotateurs, sexisme implicite, modèles de langage, modèles de langage multimodaux.

KEYWORDS: sexism detection, multimodal analysis, dataset, social media videos, annotation, inter-annotator agreement, implicit sexism, LLMs, multimodal LLMs.

ARTICLE ACCEPTÉ À : Second Conference of Language Modeling (COLM 2025).
URL : https://openreview.net/forum?id=eSAv7GKVFt

Actes des 33ème Conférence sur le Traitement Automatique des Langues Naturelles.
Volume 2 : articles déjà publiés, pages 37–37.
Cette œuvre est mise à disposition sous licence Attribution 4.0 International.