EXIST 2026: Physiological Data for Multimodal Sexism Characterization in Social Media

Laura Plaza[1], Jorge Carrillo-de-Albornoz[1], Elena Gomis-Vicent[2], Iván Arcos[2], María Aloy-Mayo[2], Paolo Rosso[2,3], Damiano Spina[4]

[1] Universidad Nacional de Educación a Distancia (UNED), 28040 Madrid, Spain
lplaza,jcalbornoz@lsi.uned.es
[2] Universitat Politècnica de València (UPV), 46022 Valencia, Spain
iarcgab@etsinf.upv.es, egomvic@upvnet.upv.es, malomay@upvnet.upv.es, prosso@dsic.upv.es
[3] ValgrAI – Valencian Graduate School and Research Network of Artificial Intelligence, 46022 Valencia, Spain
[4] RMIT University, 3000 Melbourne, Australia
damiano.spina@rmit.edu.au

Abstract. The paper describes the EXIST 2026 Lab on sexism identification in social networks, which is planned to take place at the CLEF 2026 conference and represents the sixth edition of the EXIST challenge. The lab comprises six tasks addressing three core problems—*sexism identification*, *source intention detection*, and *sexism categorization*—across two types of multimodal data: memes (image and text) and TikTok videos (video and text). Both the meme-based and video-based tasks are multilingual, covering English and Spanish. This multimodal and multilingual setup enables the analysis of how sexist content manifests across different media formats and interaction contexts, while also providing deeper insight into the social and communicative dynamics that shape the production and interpretation of sexist discourse online.
As in previous EXIST editions, the datasets will include annotations from multiple annotators, showing different or even conflicting opinions. This helps models learn from diverse perspectives. The novelty of the 2026 edition lies in the enrichment of the dataset with physiological and neurophysiological signals – specifically, heart rate, eye tracking, and electroencephalogram (EEG) data collected from different subjects while viewing the materials, that will be provided as training material. The aim is twofold: to explore how implicit emotional and cognitive responses correlate with the perception of sexist content, and to evaluate whether these signals can improve the automatic detection and classification of sexism. By incorporating features derived from users' unconscious reactions, machine learning models may capture subtle cues of bias or discomfort that are not evident in textual or visual content alone, leading to more accurate and human-aligned systems for identifying sexism across media.

Keywords: sexism identification · sexism categorization · learning with disagreement · EEG · eye tracking · heart rate · memes · TikTok videos · human-centric AI

1 Introduction

Sexism remains a pervasive form of social discrimination, reflected across multiple dimensions such as sexual violence, economic inequality, and online harassment. Recent data show that women represent around 85–90% of sexual violence victims in the USA, Europe, Spain, and Australia. [17, 5, 10, 11].The gender pay gap continues to disadvantage women, who earn on average between 8.7% and 21.8% less than men across these same regions [21, 4, 9, 22]. In the digital sphere, women also experience disproportionate levels of harassment and discrimination, with reported rates ranging from 16% in the USA to 41% in Australia, compared to 5–26% for men [12, 6]. Detecting sexism against women is therefore a crucial step toward recognizing and addressing these systemic inequalities. The automatic detection and categorization of sexist content in online communication not only enables a deeper understanding of how prejudice manifests linguistically and behaviorally, but also contributes to developing safer and more equitable digital environments.

In this context, the development of AI systems capable of detecting sexism on social media presents a particularly relevant challenge. The perception of what constitutes sexist behavior or expression involves a certain degree of subjectivity, as it may be influenced by cultural norms, personal experiences, and emotional reactions that cannot be fully captured through linguistic data alone. Despite significant advances in computational modeling, the mechanisms underlying human decision-making remain only partially understood. Empirical evidence suggests that human judgments are shaped not only by conscious factors—such as socio-demographic background, prior experiences, and explicit beliefs—but also by unconscious cues, including emotions, physiological states, and sensory responses that subtly guide perception and evaluation. Current AI models, largely trained on textual or visual data, lack access to these deeper layers of cognitive and affective information, limiting their ability to replicate or interpret complex social phenomena. To bridge this gap, it becomes essential to explore new training paradigms that integrate human-centered and sensor-based data – such as eye tracking, heart rate, and EEG signals – to provide richer insights into how individuals consciously and unconsciously perceive sexist content. Neurophysiological signals collected through wearable devices have already proven effective in characterizing a wide range of complex information access tasks [20, 7, 8]. Therefore, incorporating this multimodal perspective could lead to fairer and more context-aware AI systems, capable of understanding not only the explicit meaning of language but also the implicit affective and cognitive processes that drive human interpretations.

EXIST 2026[5] will mark the sixth edition of the sEXism Identification in Social neTworks challenge. EXIST is a series of shared tasks and scientific events aimed at detecting and characterizing sexism in a broad sense, ranging from explicit misogyny to more subtle and implicit sexist behaviors. The last three editions of the EXIST shared task were organized as labs within CLEF [14,

[5] `https://nlp.uned.es/exist2026`

16, 13], while the first two were held in the IberLEF evaluation forum [18, 19]. Over the past five years, more than 200 teams from research institutions and companies have participated in the EXIST challenge, submitting over 1,500 runs.

The EXIST 2026 lab will continue to focus on detecting sexism in social networks while introducing a novel paradigm that integrates human-centered signals into the AI development pipeline. Specifically, we incorporate sensor-based data—EEG, heart rate, and eye-tracking—collected from human annotators while they are exposed to potentially sexist content, with the aim of capturing implicit and pre-conscious responses to sexism.[6] These physiological signals provide complementary information to explicit annotations, reflecting affective and cognitive processes that may not be accessible through self-report or conscious reasoning alone. Findings from previous editions indicate that individuals often register sexist content at an implicit level, exhibiting measurable physiological reactions even when they are unable to explicitly articulate why the content is sexist or to consciously justify their annotation decisions.

By registering these signals during the annotation process, we aim to enrich training data with human-centered cues that capture early attentional shifts, emotional arousal, and cognitive load associated with the perception of harmful content. Given their inherently multimodal nature, our analysis will primarily focus on memes and short videos—formats that combine textual and visual information and are particularly suitable for eliciting such responses. We hypothesise that incorporating physiological correlates of implicit human perception into machine learning models will enable automated systems to better approximate human sensitivity to subtle or context-dependent forms of sexism, particularly in cases where surface linguistic features are insufficient, as previous studies have shown [3]. In this way, physiological signals are not treated as direct labels, but as auxiliary supervisory signals that can guide models toward more nuanced, human-aligned representations of sexist content.

Following the success of the previous two editions, the 2025 edition of EXIST will once again implement the LeWiDi paradigm for both dataset construction and system evaluation. This paradigm can significantly contribute to achieving human-centric AI by prioritizing diverse human perspectives and incorporating disagreements into the decision-making process [2].

2 EXIST 2026 Tasks

We define six experimental tasks within this lab, corresponding to three distinct task types applied across two multimodal datasets: memes and videos. Subjects are asked to identify the presence of sexism and to characterize it according to the intention of the source and the type of sexism. Each instance (meme or TikTok video) was independently evaluated by multiple annotators, whose physiological and behavioral responses were captured through eye tracking, heart rate

[6] Data collection was approved by the Research Ethics Committee of the Universitat Politècnica de València (Project ID: P01_31-10-2025).

monitoring, and electroencephalography. The resulting multimodal physiological signals, along with the corresponding annotation labels, are linked to each dataset instance.

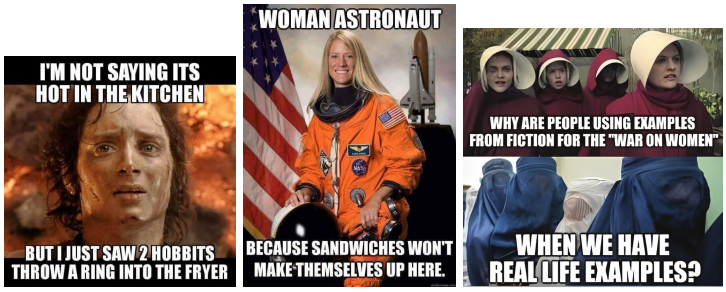
2.1 Sexism Identification

The first task is a binary classification task where systems must decide whether or not a given meme or video contains sexist expressions or behaviors (i.e., it is sexist itself, describes a sexist situation or criticizes a sexist behavior). Figure 1 show examples of sexist and not sexist memes, respectively.

2.2 Source Intention Detection

This task aims to categorize a meme or TikTok video according to the intention of the author. We propose a binary classification task: (i) direct sexist, and (ii) judgmental. The following categories are defined:

- **Direct** sexist: the intention was to create content that is sexist by itself or incites to be sexist. Figure 1 shows an example of a direct meme.
- **Judgemental** message: the intention is judgmental, since the meme or video describes sexist situations or behaviors with the aim of condemning them. Figure 1 shows an example of a judgemental meme.



(a) Not sexist (b) Sexist (direct) (c) Sexist (judgemental)

Fig. 1: Examples of direct and judgemental memes.

2.3 Sexism Categorization

In this task, each sexist meme/video must be categorized in one or more of the following categories:

- **Ideological and inequality**: this category includes memes/videos that discredit the feminist movement in order to defame the struggle of women in any aspect of their lives. It also includes contents that reject inequality between men and women, or present men as victims of gender-based oppression.
- **Stereotyping and dominance**: this category includes memes/videos that express false ideas about women that suggest they are more suitable or inappropriate for certain tasks. It also includes any claim that implies that men are somehow superior to women.
- **Objectification**: It includes content that objectifies women or reinforces stereotypes about their physical appearance, such as the expectation to meet traditional beauty standards or criticisms of their bodies.
- **Sexual violence**: this category includes memes/videos where sexual suggestions or harassment of a sexual nature (rape or sexual assault) are made.
- **Misogyny and non sexual violence**: this category includes expressions of hatred and violence towards women.

Examples of sexist memes from the previously described categories are shown in Figure 2.

(a) Ideological & inequality

(b) Objectification

(c) Stereotyping & dominance

(d) Sexual violence

(e) Misogyny & non-sexual violence

Fig. 2: Examples of memes from the different sexist categories.

3 The EXIST 2026 Dataset

The EXIST 2026 dataset comprises 8,294 multimedia instances, including memes and short videos, in both English and Spanish. This collection integrates and extends the datasets developed for EXIST 2024 (memes) and EXIST 2025 (TikTok videos), which have been reused within a novel experimental setting designed to explore how individuals perceive and interpret sexism in online media. In this new dataset, subjects' conscious judgments are reflected in the labels they assigned to each instance, while their unconscious or implicit reactions are captured through sensor data such as eye tracking, heart rate, and EEG activity. The sensor data captured during the annotation process will be provided to the participant to use (if desired) in the training process.

Detailed descriptions of the annotation methodologies for the EXIST 2024 and 2025 datasets are available in [16, 13]. Moreover, Table 1 summarizes the number of instances in the dataset per language, partition and content type.

Table 1: Number of instances in the EXIST 2026 dataset per language, partition and content type.

	Memes		TikTok	
	Spanish	English	Spanish	English
Training	1,979	2,005	1,524	1,000
Test	540	513	304	370
Total	2,519	2,518	1,828	1,370

In the new experimental framework introduced in EXIST 2026, selected subjects were exposed to the multimedia content while their physiological and behavioral responses were continuously recorded. Each instance was viewed and labeled by up to two annotators, with every participant exposed to approximately 1,000 different memes or videos. The dataset was annotated during ten annotation sessions, each comprising around 130 instances and lasting approximately 35–45 minutes.

Each session followed a structured protocol comprising several stages. First, subjects were fitted with physiological sensors for electrocardiography (ECG), eye tracking, and electroencephalography (EEG), specifically using Pupil Labs Neon glasses (binocular gaze recording at 200 Hz), a Garmin Venu 3 smartwatch (continuous heart-rate and inter-beat interval monitoring), and an OpenBCI Cyton 16-channel EEG headset (10–20 system, 250 Hz). A two-minute resting-state baseline was recorded before exposure to the stimuli. Second, subjects completed a demographic questionnaire, administered only once, that collected information such as age, gender, education level, country of residence, and occupation. Third, they reported their average daily social media usage (in hours) and the percentage of time spent on platforms such as TikTok, Instagram, X/Twitter, and Facebook.

The next stage involved completing a cognitive thinking style questionnaire composed of 24 items rated on a six-point Likert scale (from "totally disagree" to "totally agree"). The items were grouped into four cognitive dimensions: open thought, closed thought, intuitive thought, and effortful thought, capturing different aspects of individual reasoning and information processing tendencies. Finally, before each experimental session, subjects completed a visual analogue scale (0–100) to self-assess their current emotional state, indicating the degree to which they felt happy, sad, calm, tense, energetic, or sleepy.

4 Evaluation Methodology and Metrics

As in previous EXIST editions, we will carry out two types of evaluations:

1. **Soft-soft evaluation**: intended for systems that provide probabilities (soft outputs) for each category, rather than a single label. We will use ICM-Soft (see details in [15]), as the official evaluation metric. We will also report Cross Entropy.
2. **Hard-hard evaluation**: intended for systems that provide a hard, conventional output. To derive the hard labels in the ground truth from the different annotators' labels, we will use a probabilistic threshold computed for each task. The official metric for this task will be ICM [1]. We will also report F1.

Acknowledgments This work was supported by the Spanish Ministry of Science, Innovation and Universities (project ANNOTATE (PID2024-156022OB-C31 and PID2024-156022OB-C32)) funded by MICIU/AEI/10.13039/501100011033 and the European Social Fund Plus (ESF+), and by the Australian Research Council Centre of Excellence for Automated Decision-Making and Society (ADM+S, CE200100005).

Disclosure of Interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Amigó, E., Delgado, A.: Evaluating Extreme Hierarchical Multi-Label Classification. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. pp. 5809–5819 (2022)
2. Basile, V., Fell, M., Fornaciari, T., Hovy, D., Paun, S., Plank, B., Poesio, M., Uma, A.: We Need to Consider Disagreement in Evaluation. In: Proceedings of the 1st Workshop on Benchmarking: Past, Present and Future. pp. 15–21. Association for Computational Linguistics, Online (2021)
3. Divya Venkatesh, J., Jaiswal, A., Nanda, G.: Comparing Human Text Classification Performance and Explainability with Large Language and Machine Learning Models Using Eye-Tracking. Scientific Reports **14**(1), 14295 (Jun 2024). https://doi.org/10.1038/s41598-024-65080-7

4. Eurostat: Gender Pay Gap Statistics. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Gender_pay_gap_statistics (2025), Last accessed 22 Dec 2025
5. Eurostat: Violence Experienced by Total Population. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Violence_experienced_by_total_population (2025), Last accessed 22 Dec 2025
6. Gobierno de España: Brecha Digital de Género. https://www.inmujeres.gob.es/publicacioneselectronicas/documentacion/Documentos/DE2110.pdf (2021), Last accessed 22 Dec 2025
7. Ji, K., Hettiachchi, D., Salim, F.D., Scholer, F., Spina, D.: Characterizing information seeking processes with multiple physiological signals. In: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 1006–1017. SIGIR '24, Association for Computing Machinery, New York, NY, USA (2024). https://doi.org/10.1145/3626772.3657793
8. Ji, K., Hettiachchi, D., Scholer, F., Salim, F.D., Spina, D.: Senseseek dataset: Multimodal sensing to study information seeking behaviors. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. **9**(3) (Sep 2025). https://doi.org/10.1145/3749501
9. Ministerio de Igualdad, Gobierno de España: Declaración Institucional con Motivo del Día para la Igualdad Salarial. https://www.igualdad.gob.es/comunicacion/sala-de-prensa/declaracion-institucional-con-motivo-del-dia-para-la-igualdad-salarial/ (2025), Last accessed 22 Dec 2025
10. Ministerio del Interior: Informe Delitos Contra la Libertad Sexual 2023. https://estadisticasdecriminalidad.ses.mir.es/publico/portalestadistico/dam/jcr:0a219f1b-f2f5-4e95-9553-aa8cf9ee5b3f (2023), Last accessed 22 Dec 2025
11. National Association of Services Against Sexual Violence: Data on Sexual Violence Services in Australia. https://www.nasasv.org.au/data (2025), Last accessed 22 Dec 2025
12. Pew Research Center: The State of Online Harassment. https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/ (2021), Last accessed 22 Dec 2025
13. Plaza, L., Carrillo-de Albornoz, J., Arcos, I., Rosso, P., Spina, D., Amigó, E., Gonzalo, J., Morante, R.: Overview of EXIST 2025: Learning with Disagreement for Sexism Identification and Characterization in Tweets, Memes, and TikTok Videos. In: Carrillo-de Albornoz, J., García Seco de Herrera, A., Gonzalo, J., Plaza, L., Mothe, J., Piroi, F., Rosso, P., Spina, D., Faggioli, G., Ferro, N. (eds.) Proceedings of the 16th International Conference of the CLEF Association (CLEF 2025). Madrid, Spain (2025)
14. Plaza, L., Carrillo-de Albornoz, J., Morante, R., Amigó, E., Gonzalo, J., Spina, D., Rosso, P.: Overview of EXIST 2023 – Learning with Disagreement for Sexism Identification and Characterization. In: Arampatzis, A., Kanoulas, E., Tsikrika, T., Vrochidis, S., Giachanou, A., Li, D., Aliannejadi, M., Vlachos, M., Faggioli, G., Ferro, N. (eds.) Proceedings of the Fourteenth International Conference of the CLEF Association (CLEF 2023). Thessaloniki, Greece (2023)
15. Plaza, L., Carrillo-de Albornoz, J., Morante, R., Amigó, E., Gonzalo, J., Spina, D., Rosso, P.: Overview of EXIST 2023 – Learning with Disagreement for Sexism Identification and Characterization (Extended Overview). In: Aliannejadi, M., Faggioli, G., Ferro, N., Vlachos, M. (eds.) Working Notes of CLEF 2023 - Conference and Labs of the Evaluation Forum (2023)

16. Plaza, L., Carrillo-de Albornoz, J., Ruiz, V., Maeso, A., Chulvi, B., Rosso, P., Amigó, E., Gonzalo, J., Morante, R., Spina, D.: Overview of EXIST 2024 — Learning with Disagreement for Sexism Identification and Characterization in Tweets and Memes. In: Goeuriot, L., Mulhem, P., Quénot, G., Schwab, D., Di Nunzio, G.M., Soulier, L., Galuščáková, P., García Seco de Herrera, A., Faggioli, G., Ferro, N. (eds.) Proceedings of the 15th International Conference of the CLEF Association (CLEF 2024). Grenoble, France (2024)
17. RAINN: Statistics: Victims of Sexual Violence. https://rainn.org/facts-statistics-the-scope-of-the-problem/statistics-victims-of-sexual-violence/ (2025), Last accessed 22 Dec 2025
18. Rodríguez-Sánchez, F., Carrillo-de Albornoz, J., Plaza, L., Gonzalo, J., Rosso, P., Comet, M., Donoso, T.: Overview of EXIST 2021: Sexism Identification in Social Networks. Procesamiento del Lenguaje Natural **67**, 195–207 (2021)
19. Rodríguez-Sánchez, F., Carrillo-de Albornoz, J., Plaza, L., Mendieta-Aragón, A., Marco-Remón, G., Makeienko, M., Plaza, M., Spina, D., Gonzalo, J., Rosso, P.: Overview of EXIST 2022: Sexism Identification in Social Networks. Procesamiento del Lenguaje Natural **69**, 229–240 (2022)
20. Spina, D., Gwizdka, J., Ji, K., Moshfeghi, Y., Mostafa, J., Ruotsalo, T., Zhang, M., Ahmad, A., Lawati, S.F.D.A., Boonprakong, N., Fernando, N., He, J., Hoeber, O., Jayawardena, G., Lee, B.G., Liu, H., Pike, M., Pirmoradi, A., Nakisa, B., Rastgoo, M.N., Salim, F.D., Scott, F., Sun, S., Tang, H., Towey, D., Wilson, M.L.: Report on the 3rd Workshop on NeuroPhysiological Approaches for Interactive Information Retrieval (NeuroPhysIIR 2025) at SIGIR CHIIR 2025. SIGIR Forum **59**(1), 1–43 (Oct 2025). https://doi.org/10.1145/3769733.3769740
21. U.S. Bureau of Labor Statistics: Women's Earnings Were 83.6 Percent of Men's in 2023. https://www.bls.gov/opub/ted/2024/womens-earnings-were-83-6-percent-of-mens-in-2023.htm (2024), Last accessed 22 Dec 2025
22. Workplace Gender Equality Agency: Gender Pay Gap Data. https://www.wgea.gov.au/pay-and-gender/gender-pay-gap-data (2025), Last accessed 22 Dec 2025