



cogitatio

SOCIAL INCLUSION

Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment

Edited by Max Römer Pieretti, Beatriz Esteban-Ramiro,
and Agrivalca Canelón

Volume 13

2025

Open Access Journal

ISSN: 2183-2803



Social Inclusion, 2025, Volume 13
Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment

Published by Cogitatio Press
Rua Fialho de Almeida 14, 2º Esq.,
1070-129 Lisbon
Portugal

Design by Typografia®
<http://www.typografia.pt/en/>

Cover image: © doble-d from iStock

Academic Editors

Max Römer Pieretti (Camilo José Cela University)
Beatriz Esteban-Ramiro (University of Castilla-La Mancha)
Agrivalca Canelón (Andrés Bello Catholic University)

Available online at: www.cogitatiopress.com/socialinclusion

This issue is licensed under a Creative Commons Attribution 4.0 International License (CC BY). Articles may be reproduced provided that credit is given to the original and *Social Inclusion* is acknowledged as the original venue of publication.

Table of Contents

Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment

Max Römer-Pieretti, Beatriz Esteban-Ramiro, and Agrivalca Canelón Silva

Hate Speech Directed at Spanish Female Actors: Penélope Cruz—A Case Study

Lucía Tello Díaz and Lizette Martínez-Valerio

Intergenerational Evolution of Gender Bias in Spain: Analysis of Values Surveys

Pilar Antolínez-Merchán, Ángel Rivero Recuenco, and Elvira Carmen Cabrera-Rodríguez

Unveiling Hate Speech Dynamics: An Examination of Discourse Targeting the Spanish Meteorological Agency (AEMET)

Sergio Arce-García, Virginia Martín-Jiménez, and Leticia Rodríguez-Fernández

Semiotic Analysis of Hate Discourse in Spanish Digital News Media: Biden's Inauguration Case Study

Max Römer-Pieretti, Elías Said-Hung, and Julio Montero-Díaz

Propagation of Hate Speech on Social Network X: Trends and Approaches

Eva Matarín Rodríguez-Peral, Tomás Gómez Franco, and Daniel Rodríguez-Peral Bustos

Frames and Triggers of Extreme Speech: The Case of Transphobia

Fabienne Baider

The Italian Manosphere: Composition, Structure, and Functions of a Digital Network

Elisa Ignazzi, Mara Maretti, and Lara Fontanella

Violence, Hate Speech, and Discrimination in Video Games: A Systematic Review

Roberto Moreno-López and Catalina Argüello-Gutiérrez

Androcentrism and Violence in Online Video Games: Perpetuation of Gender Inequality

Beatriz Esteban-Ramiro, Patricia Fernández de Castro, and Orlanda Díaz-García

Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment

Max Römer-Pieretti ¹, Beatriz Esteban-Ramiro ², and Agrivalca Canelón Silva ³

¹ Communications and Design Faculty, Camilo José Cela University, Spain

² Department of Labour Law and Social Work, University of Castilla-La Mancha, Spain

³ Institute for Communication and Information Research, Andrés Bello Catholic University, Venezuela

Correspondence: Beatriz Esteban-Ramiro (beatriz.esteban@uclm.es)

Submitted: 27 January 2025 **Published:** 18 February 2025

Issue: This editorial is part of the issue “Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment” edited by Max Römer Pieretti (Universidad Camilo José Cela), Beatriz Esteban-Ramiro (Universidad de Castilla-La Mancha), and Agrivalca Canelón (Universidad Católica Andrés Bello), fully open access at <https://doi.org/10.17645/si.i415>

Abstract

This thematic issue examines the complex and current intersection between violence, hate speech, and gender bias in digital environments. It explores how digital spaces (encompassing social networks, news platforms, and online multiplayer games) can both perpetuate and challenge systemic inequalities. Contributions present diverse methodologies, including longitudinal analyses, qualitative studies, and systematic reviews, to uncover the mechanisms underlying digital violence and exclusion. This issue highlights the urgency of addressing digital exclusion to safeguard democratic values and social cohesion. It aims to inspire new research and inform policies that build more inclusive and equitable digital environments, laying the groundwork for future social science research and practice.

Keywords

digital environment; equality; gender; hate speech; inclusion; misogyny; multiplayer online games; social media; videogames; violence

1. Introduction

Coordinating this issue has been an extraordinary and rewarding challenge. Why a challenge? Because this issue is born out of the synergy between two projects. In addition, the inclusion of research from other parts of the world and the international peer-review process have reinforced the validity and relevance of our collective research efforts.

Research articles such as those that underpin this issue are validated by their ability to engage in dialogue with scholars from around the world who advance the social sciences. This dialogue enriches our understanding and strengthens the impact of the research presented. The thematic focus of this issue owes much to the following two research initiatives:

1. HATEMEDIA: Taxonomy, Presence, and Intensity of Hate Speech in Digital Environments Linked to the Spanish Professional News Media (PID2020-114584GB-I00), funded by the Ministry of Science and Innovation. This project aims to analyze how hate speech spreads within digital environments associated with professional news media and to enhance the detection and monitoring of such expressions in Spain.
2. 32 BITS: Androcentrism, Hate Speech, and Gender Bias through Online Video Games in Castilla-La Mancha (SBPLY/21/180501/000262), funded by the Junta de Comunidades de Castilla-La Mancha and co-financed by the European Union-Feder. This project explores manifestations of hate, violence, and sexism within multiplayer online gaming environments, as well as the personal and social consequences of these phenomena in Spain.

The digital transformation of society has reshaped human interaction, offering unprecedented opportunities for connection, expression, and collaboration. However, these same online spaces have also become arenas where violence, hate speech, and gender bias proliferate, undermining the potential for inclusivity and equality in the digital world. This thematic issue of *Social Inclusion* seeks to examine the dynamics of these challenges, their societal impacts, and potential pathways toward more inclusive digital environments.

The urgency of addressing digital violence and exclusion stems from their pervasive and growing nature. High-profile cases of online harassment, coordinated hate campaigns, and algorithmic bias have sparked public and scholarly debates about the ethical responsibilities of digital platforms and the adequacy of existing regulatory frameworks. In a world where digital tools increasingly mediate social interactions, such phenomena not only harm individuals but also deepen systemic inequalities, marginalizing already vulnerable populations. By exploring these issues, this thematic issue contributes to the mission of *Social Inclusion* by interrogating the barriers to equitable participation in the digital age.

The proliferation of the Internet, ICTs, and social media has dramatically transformed communication in contemporary societies. As Piñeiro-Otero and Martínez-Rolán (2021) suggest, the collective perception of the public sphere is evolving, influencing how relationships are established and reshaping processes of social inclusion and exclusion. Similarly, Paz et al. (2021) highlight key factors driving the spread of hate speech, including the mediums used, the topics addressed, and the settings of online discussions. These changes carry significant social repercussions, endangering values essential to social cohesion by fostering hate speech that exacerbates tensions and incites violence.

Digital violence, harassment, and hate speech have permeated various online spaces, including social media, digital platforms, and massively multiplayer online games (MMOs). As this thematic issue reveals, online interactions in these environments reflect broader societal dynamics of inclusion and exclusion. The thematic focus is particularly concerned with the impacts of digital environments on adolescents, and young adults, but also in the general population, emphasizing the prevalence of hate speech, harassment,

and violence shaped by identity-based and gender biases. At the same time, the issue seeks to explore countermeasures that can help mitigate these phenomena, fostering a more inclusive society informed by gender-sensitive and intersectional perspectives.

As highlighted by recent literature, combating hate is a multifaceted challenge. It requires the promotion of narratives that support democratic values and respect for diversity. The integration of these principles is vital to advancing towards a more inclusive digital society. Authors such as Castells (2006) and Fuchs (2021) emphasize the role of communication networks and the political economy of platforms in shaping online behaviors and interactions. These insights underline the need for a critical examination of the structural and cultural conditions that enable digital exclusion and violence.

2. Contributions

The articles in this thematic issue provide a comprehensive exploration of the challenges posed by violence, hate speech, and gender bias in digital environments. Tello Díaz and Martínez-Valerio (2025) begin by analyzing hate speech directed at Penélope Cruz on the social media platform X. Their study highlights the predominance of ideological hate and misogyny, revealing how stereotypes and prejudices question her intellect and professional success.

Building on this analysis, Antolínez-Merchán et al. (2025) investigate gender biases among Spanish youth using data from the World Values Survey: The study identifies media exposure and sociocultural values as key predictors of gender prejudice.

Arce-García et al. (2025) examine nearly half a million messages targeting the Spanish Meteorological Agency (AEMET), uncovering how conspiracy theories and anti-intellectualism fuel hate speech that undermines trust in scientific institutions. Römer-Pieretti et al. (2025) complement these findings by employing a semiotic approach to explore hate discourse in user comments on Spanish news media platforms. Their work highlights the influence of political radicalisation and international events, such as Joe Biden's inauguration, on digital debates.

Rodríguez-Peral et al. (2025) contribute by reviewing academic research on hate speech on Twitter, identifying theoretical frameworks and methodological approaches while pointing out gaps for future inquiry. On the other hand, Baider (2025) adopts a critical discourse perspective to analyze transphobic argumentation and counterspeech. By identifying frames such as the medical and misfit frames, the author provides actionable insights into addressing extreme speech and fostering respectful digital dialogue.

Ignazzi et al. (2025) map the prevalence of misogynistic discourse within the manosphere. Using advanced natural language processing, the authors uncover how this network influences offline behaviors and reinforces harmful gender stereotypes. Moreno-López and Argüello-Gutiérrez (2025) expand on this by conducting a systematic review of online hate speech in online video games, and their synthesis underscores the diverse forms of cyberhate present in these environments and the pressing need for inclusive interventions.

Finally, Esteban-Ramiro et al. (2025) zoom in on MMOs, where they document how androcentric norms perpetuate gender-based violence. Their qualitative analysis reveals the harassment faced by female gamers and the broader implications for societal norms.

Acknowledgments

We are grateful to the editorial team at *Social Inclusion* for their careful guidance throughout this thematic issue's publication.

Funding

Projects: Taxonomy, Presence, and Intensity of Hate Speech in Digital Environments Linked to the Spanish Professional News Media-HATE MEDIA (PID2020–114584GB-I00), funded by the Ministry of Science and Innovation and Androcentrism, Hate Speech, and Gender Bias through Online Video Games in Castilla-La Mancha—32 BITS—(SBPLY/21/180501/000262), funded by the Junta de Comunidades de Castilla-La Mancha and co-financed by the European Union–Feder.

Conflict of Interests

The authors declare no conflict of interests.

References

- Antolínez-Merchán, P., Rivero Recuenco, Á., & Cabrera-Rodríguez, E. C. (2025). Intergenerational evolution of gender bias in Spain: Analysis of values surveys. *Social Inclusion*, 13, Article 9288. <https://doi.org/10.17645/si.9288>
- Arce-García, S., Martín-Jiménez, V., & Rodríguez-Fernández, L. (2025). Unveiling hate speech dynamics: An examination of discourse targeting the Spanish meteorological agency (AEMET). *Social Inclusion*, 13, Article 9291. <https://doi.org/10.17645/si.9291>
- Baider, F. (2025). Frames and triggers of extreme speech: The case of transphobia. *Social Inclusion*, 13, Article 9379. <https://doi.org/10.17645/si.9379>
- Castells, M. (2006). *La sociedad red: una visión global*. Alianza editorial.
- Esteban-Ramiro, B., Fernández de Castro, P., & Díaz-García, O. (2025). Androcentrism and violence in online video games: Perpetuation of gender inequality. *Social Inclusion*, 13, Article 9425. <https://doi.org/10.17645/si.9425>
- Fuchs, C. (2021). The digital commons and the digital public sphere: How to advance digital democracy today. *Westminster Papers in Communication and Culture*, 16(1). <https://doi.org/10.16997/wpcc.917>
- Ignazzi, E., Maretti, M., & Fontanella, L. (2025). The Italian manosphere: Composition, structure, and functions of a digital network. *Social Inclusion*, 13, Article 9341. <https://doi.org/10.17645/si.9341>
- Moreno-López, R., & Argüello-Gutiérrez, C. (2025). Violence, hate speech, and discrimination in video games: A systematic review. *Social Inclusion*, 13, Article 9401. <https://doi.org/10.17645/si.9401>
- Paz, M. A., Mayagoitia-Soria, A., & González-Aguilar, J. M. (2021). From polarization to hate: Portrait of the Spanish political meme. *Social Media + Society*, 7(4). <https://doi.org/10.1177/20563051211062920>
- Piñeiro-Otero, T., & Martínez-Rolán, X. (2021). Say it to my face: Analysing hate speech against women on Twitter. *Profesional de la información*, 30(5). <https://doi.org/10.3145/epi.2021.sep.02>
- Rodríguez-Peral, E. M., Gómez Franco, T., & Rodríguez-Peral Bustos, D. (2025). Propagation of Hate Speech on Social Network X: Trends and Approaches. *Social Inclusion*, 13, Article 9317. <https://doi.org/10.17645/si.9317>
- Römer-Pieretti, M., Said-Hung, E., & Montero-Díaz, J. (2025). Semiotic analysis of hate discourse in Spanish digital news media: Biden's inauguration case study. *Social Inclusion*, 13, Article 9295. <https://doi.org/10.17645/si.9295>
- Tello Díaz, L., & Martínez-Valerio, L. (2025). Hate speech directed at Spanish female actors: Penélope Cruz—A case study. *Social Inclusion*, 13, Article 9250. <https://doi.org/10.17645/si.9250>

About the Authors



Max Römer-Pieretti is a full professor and researcher at Universidad Camilo José Cela, where he is the research coordinator of the Faculty of Communication and Design. His research interests include semiotics, politics, and hate studies. With more than 35 years of teaching and research experience, his teaching background includes undergraduate, graduate, and doctoral classes.



Beatriz Esteban-Ramiro holds a PhD in law from the University of Castilla-La Mancha. She is part of the teaching and research staff at the University of Castilla-La Mancha, the Department of Labour Law and Social Work, and is assistant professor PHD in social work at the Faculty of Social Sciences and Information Technologies. She is a member of the GIEPS: Inclusion, Equity and Social Policy research group. Her lines of research focus on online violence, gender, and hate speech, as well as social inequalities.



Agrivalca Canelón Silva holds a PhD in organizational communication by the Universidad de Málaga and is an associate researcher at the Institute for Communication and Information Research of the Andrés Bello Catholic University (IDICI-UCAB). Agrivalc's research interests include corporate communications and public relations, government communication and public diplomacy, sustainability and social innovation, and migration and diasporas.

Hate Speech Directed at Spanish Female Actors: Penélope Cruz—A Case Study

Lucía Tello Díaz ¹  and Lizette Martínez-Valerio ² 

¹ Arts and Social Sciences Department, Universidad Internacional de La Rioja, Spain

² Journalism and Corporate Communication Department, Rey Juan Carlos University, Spain

Correspondence: Lizette Martínez-Valerio (lizette.martinez@urjc.es)

Submitted: 5 September 2024 **Accepted:** 17 December 2024 **Published:** 18 February 2025

Issue: This article is part of the issue “Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment” edited by Max Römer Pieretti (Universidad Camilo José Cela), Beatriz Esteban-Ramiro (Universidad de Castilla-La Mancha), and Agrivalca Canelón (Universidad Católica Andrés Bello), fully open access at <https://doi.org/10.17645/si.i415>

Abstract

Penélope Cruz is one of the most internationally acclaimed Spanish performers. However, despite her successful career, she is also one of the celebrities subject to most controversy on social media and the most frequent target of hate speech. Although she does not manage her own profile on X (previously Twitter), her name and criticism of her are constant on this platform. The objective of this study is to detect possible hate speech, as well as to categorise it by its intensity and typology. The study analyses the unrestricted comments on X containing the name Penélope Cruz posted during the period between January and June 2023. The methodology utilised is that of quantitative and qualitative content analysis of the comments registered ($N = 6,620$). One of the chief results is the fact that the majority of the comments classified as hate speech refer to ideological issues (70.9%) and/or are misogynistic (8.9%), among which are the specific allusions to her acting skills and her physical characteristics. The results coincide with other studies in which hate messages directed at actresses are related to their physical appearance. However, they differ in that in this case the main type of hatred is not misogyny, but ideological hatred. The actor is not only accused of being a “communist” and a “hypocrite” for her lifestyle, but also for having used her body to succeed in her profession. Both her physique and intellectual capacity are also subject to hate speech. Most of these messages are based on conjecture, prejudice, and stereotypes.

Keywords

hate speech; ideological hatred; misogyny; Penélope Cruz; sexism; Spanish cinema; Twitter

1. Introduction

The bias of sexual objectification and reductionism towards a merely objectual perspective has taken root in the feminine conception of the acting profession. The association between acting and prostitution is not a new phenomenon since actors have long been related to a dissolute lifestyle. The equating of acting and prostitution has become so common that a phrase has been coined: “What mother, however dishonest she may be, would not rather see her daughter in the grave than in the theatre or on stage” (Garcías, 1998, p. 39).

This vision has been transferred to social media, where actresses are frequently the subject of comments in physical terms, judging their appearance, their clothing, and even their normativity based on their matching up to the traditional feminine ideal. In contrast to the notion of masculinity, a clear distinction is made between what it means and implies to be a woman. Sexist language, also observed on these platforms, highlights the social differences between sexes and establishes the way in which the construction of gender is materialised: The presence of sexism at the linguistic level did not seem to be a priority in the fight for respect for women’s rights. However, over the last decade, due to social movements that seek to revindicate women’s position in society, questions have arisen concerning different areas where they have been discriminated against, abused, and ostracized. Language is one of those spaces, through which the reality of the speakers is manifested where discrimination, stereotypes, and gender violence clearly prevail, and therefore it was decided to term this issue “linguistic sexism” (Corse et al., 2014, p. 141).

An example of the use of this type of language on social media is found in Spain during the 2021 Goya Awards ceremony. During the gala, despite the limitations imposed by the pandemic, a television presenter recorded a colleague’s sexist comments (neither his identity nor his profession were disclosed), a colleague who not only defended his right to judge and classify female actors by their appearance, but also criticized their morality when referring to one of them as a prostitute. The response was immediate and actresses like Letizia Dolera expressed their discontent on social media with the idea that a woman was a prostitute solely for wishing to act: “And once again, blurting out ‘whore’” (Riaño, 2021). However, there were no further repercussions and only TVE (the Spanish public broadcaster) issued a statement disassociating itself from the comments. Appearing on the program *Hora 25*, the journalist and screenwriter Eduardo Galán affirmed that social media treat men and women differently, which is a reality: “While they insult me by talking about how ‘dumb’ I am, they insult my wife, Marta Flich, by saying she’s a ‘whore.’ And this always happens. There you can see the machismo on social media” (Bretos, 2023).

Beyond gender issues, expressions of hatred towards audiovisual professions in Spain are mainly due to ideological factors. Film workers, particularly actors, are criticised due to a belief that they are affiliated to a certain political faction, which influences the conceptualisation of their work in certain areas. It is evident that not all actors share progressive positions; however, the prevailing national mentality considers them members of left-wing parties. In fact, one of the topics that generates the greatest contempt on social media is when an actor reveals his or her ideology:

There is a question that goes further and provides a lot of information about what these characters are like and what they really think: Who do they vote for?...Voting is secret and, like everything unknown, excites curiosity. In many cases, people in the public eye try to conceal their ideology. The main reasons for this may be to avoid being judged or not wanting to influence others with their ideas. (Yaben, 2021)

Although it is commonly believed that the world of cinema leans to the political left, there are many artists who have expressed their support for other groupings, such as actors and comedians related with Vox, or presenters and actors who support PP (People's Party). In recent times, there have been a large number of artists who have taken on a political role in the public sphere, including actors in UPyD (Union, Progress and Democracy), and presenters in *Compromís* or *Ciudadanos* (Campos, 2016)—although Spain is marked by a two-party system between PSOE (Spanish Socialist Workers' Party—progressives) and PP (liberals), it also has a wide political spectrum that includes national and regional parties. Among them, there are far-right (Vox), liberal (*Ciudadanos*), centre (UPyD), and left-wing regional groups (*Compromís*).

Although many actors have felt the need to express their disagreement with politics interfering in the artistic life of performers, the preconceived idea still prevails: "Culture should be outside of political comings and goings, it needs freedom and independence to develop itself, it is essential that we can talk about an autonomous culture" (Campos, 2016).

The conceptualisation of big screen actors as subversive "puppeteers" is one of the most obvious examples of the ideologization of the profession. The term is used to refer to actors in a derogatory and insulting way. As the actor Rosana Pastor mentions, "We are all tired of reading about 'the puppeteers or comedians' with a highly pejorative tone" (Campos, 2016). The fact of receiving subsidies has also been used against the film sector. In reality, most industries in Spain receive subsidies from the government. These include the automotive, mining, insurance, electrical, food, and construction industries (Gabinete de Estudios Económicos de Axesor, 2013). Even political parties receive funding. Such terms as "subsidised," "puppeteers," or "reds" have persisted in anti-cinema discourse in Spain over the decades, despite the fact that countries like France allocate six times more funds to their film industry and Italy offers greater tax incentives to theirs (Álvarez, 2020).

Despite the frequency with which these and other terms are used to generate hate speech against actors and actresses on social media, the academic world has shown little interest in their study. Sexism has been identified as a problem in video games, blogs, online forums, and social media (Fox & Tang, 2015; Marwick, 2013; Pedersen & Macafee, 2007; Penny, 2014; Shaw, 2014), but not towards actresses.

Women as targets of hate speech have been the subject of numerous studies. However, some female authors argue that production of research in this area is scarce, possibly due to the somewhat limited definition of hate speech in Europe (Kopytowska & Baider, 2017). Hanash (2018) compared cyberbullying of women to the witch hunts of the 16th and 17th centuries. This case study concludes that self-censorship is one of the consequences of the cyberbullying of women. Others have investigated the causes of online misogyny. The two main components that contribute to the disinhibition of sexist attacks on the internet are anonymity and the virtualisation of relationships (Fox et al., 2015; Lapidot-Lefler & Barak, 2012).

Much of the research on social media hatred toward women is based on automation. This is common in research into hate speech and misogyny (Burnap et al., 2015; Escalante et al., 2017; Fersini et al., 2018; Justo et al., 2014; Nobata et al., 2016). Some recognise that tools of this type are insufficient due to factors largely related to language and lack of context (Frenda et al., 2019; Sood et al., 2012). Even rarer are studies focused on the hate targeted at female celebrities. Some research has focused on the case of the most important influencers in Spain (Martínez-Valerio & Mayagoitia-Soria, 2021). The researchers conclude that these individuals with their millions of followers have chosen to self-censor the content they create in order

to stop receiving hate messages. Most of such messages were related to their ideology or physical appearance. There are some case studies from other countries of women involved in films. Harassment campaigns orchestrated from social media against actresses such as the American Leslie Jones (Brojakowski & Cruz, 2022) or the Indian Parvathy (Karthika, 2022) have been studied, while others have analysed the messages that the actor, screenwriter, and director Lena Dunham receives for displaying her body and her illness on Instagram (Ghaffari, 2020). All three received misogynistic messages; Jones also received racist comments, and Dunham was subject to attacks regarding her body.

As can be seen, in general terms, hate speech is a type of language that aims to be harmful (Davidson et al., 2017) and constitutes a deliberate, voluntary, and detrimental attack against identity and human dignity. Its offensive and/or obscene discourse motivates potentially damaging activities and could also lead to illegal actions such as defamation or fraud (Lamson et al., 2022). This speech diminishes the quality of democracy, spreading prejudice and intolerance (Said-Hung et al., 2023). Anonymity, mass communication, and the use of false identities on social networks are key factors that allow the rise of illicit behaviour, discriminatory comments, and increasing verbal violence against women (Terrón, 2020). The social networks sphere is used for verbalizing sexual insinuations, insults, or ridicule expressions led by profiles (mostly male) that amplify gender inequalities by reproducing stereotypes and spreading misinformation, disinformation, and hate speech against women (Tarullo et al., 2024).

1.1. Objectives

The general objective of this study is to explore the presence of hate speech directed against the actor Penélope Cruz on the X platform (formerly Twitter). As specific objectives, the work proposes to:

- Detect hate speech that contains the words “Penélope Cruz.”
- Classify the hate speech detected based on the level of hate and its typology.
- Study the content of the hate messages directed against the actor.

2. Methodology

The research works with the case study of the actor Penélope Cruz through the empirical content analysis of the posts that contain her name on the X social media platform. The selection of Penélope Cruz as a case study is due, on the one hand, to the fact that public figures, such as politicians (Piñeiro-Otero & Martínez-Rolán, 2021; Sánchez-Meza et al., 2023; Zamora-Martínez et al., 2024), actors (Karthika, 2022; Piñeiro-Otero & Martínez-Rolán, 2021), content creators (Martínez-Valerio & Mayagoitia-Soria, 2021), etc., are often common targets of hate messages on social media. On the other hand, previous studies indicate that this particular actor stands out as a target of hate messages within the Spanish film industry (Martínez-Valerio & Tello-Díaz, 2024). Furthermore, Penélope Cruz is the most internationally recognized Spanish actor. She is a renowned pioneer in Hollywood (Mao, 2019), being the first Spanish woman who won the Academy Award for Best Supporting Actress (2010). Cruz has received all the European film awards (the BAFTA Award, the Donostia Award, the Best Female Performance Award at the Cannes Film Festival, the David de Donatello Award, the Cesar Award, and the Volpi Cup), and has also achieved the Spanish Gold Medal for Merit in Fine Arts and has been nominated for the Golden Globe Awards and the European Film Awards more than five times. However, she is also known for her commitment and closeness to progressive policies, which has implied some conflicts

with Spanish media (Fernández, 2012). This is a key factor for choosing the actor as representative of hate speech in social networks.

The messages analysed can originate from various parts of the world and are not limited to Spanish users. The period of analysis consists of the first half of 2023 (January–June), months during which the most important film awards ceremonies for this case study take place (Goyas, Oscars, and the Italian David awards), as well as other notable events attended by the actor (the Met Gala). The final sample included an analysis of the comments generated in Spanish by all X users during the period explained above using the keywords “Penelope Cruz” ($N = 6,620$). Quantitative content analysis was chosen, said technique allowing for the objective and systematic study of the content of the tweets (X posts) analysed (Igartua, 2006; Piñuel, 2002), following the compilation of information by automatic download of the comments using the Export Comments tool.

The choice of X as the social media platform for the study is due to several factors. It is a social medium on which the influence of micronarratives has increased significantly since its inception and on which users follow current events (García-de-Torres et al., 2011). Moreover, not only has the capacity of the media to disseminate content increased, but also their ability to perform journalistic tasks. Furthermore, compared to other social media such as Facebook, it has more active user participation (Aruguete, 2015; Mayo-Cubero, 2019). Finally, it is one of the social media with the greatest presence of users linked to both traditional and digital media in Europe (S. González & Ramos, 2013).

Labelling was carried out separately by the two authors, coinciding in 97.7% of cases. Based on this process, the content was labelled into the following categories:

- Presence or absence of expressions of hate.
- Type of hate detected based on the vulnerable groups targeted by hate speech (González-Aguilar et al., 2023): religious, xenophobic, racist, misogynistic, sexual, ideological, and other.
- Level of intensity of hatred observed on a scale from 0 to 5 following the proposal of De Lucas et al. (2022):
 - Level 0: Messages that use expressions with socially negative connotations. There is no clear incitement to hatred, but it may be a first step toward encouraging it. E.g., “extreme left” or “fascist.”
 - Level 1: Messages in which there is no verbal violence, but something is presented factually in order to stigmatise a specific social group.
 - Level 2: Messages containing abusive expressions without being insulting; however, they clearly attribute actions with the intention of spreading a negative image. E.g., “young immigrants commit crimes.”
 - Level 3: Messages with overt verbal violence. E.g., “femi-nazi,” “fascist,” “moron.”
 - Level 4: Messages with veiled or implicit threats. E.g., “I’d better not run into one of those puppeteers on the street...”
 - Level 5: Messages that include a call for physical violence or that express a desire for another to die or suffer physical harm. E.g., “I hope Covid wipes out all those good-for-nothing actors.”
- Presence of a humorous tone within the messages containing expressions of hatred.

3. Results

Of the 6,620 tweets analysed, 213 were classified as hate speech. That is 3.2% of the total. The majority of them, 70.9%, have been categorised as hate speech for ideological reasons. The second most common type is Other, 18.8% of the total. As will be seen later in the qualitative analysis, these messages are mainly related to her performances as an actor. Next are misogynistic messages, 8.9%. Only three racist messages were found (0.9%). No messages were found that could be categorised as religious hatred, xenophobia, or discrimination against a sexual orientation (Figure 1).

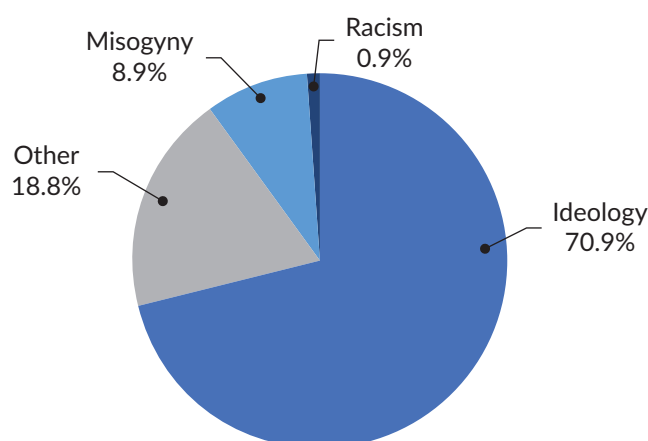


Figure 1. Categories found within hate speech.

3.1. Hate Speech Based on Ideology

The labelling of ideological messages showed that over half of the tweets are at level 0, which means content with socially negative connotations (74 tweets, 54.41%). There is also a significant portion corresponding to level 1 (62 tweets, 45.5%), in which the actor and part of her family are presented in a stigmatising way. Likewise, a single tweet (0.7%) reaches level 2, that is, it tries to spread a negative image of her and her family, as well as two tweets (1.4%) reaching level 3 (tweets with overt verbal violence) as broken down below. Humour is only observed in tweets in 8.08% of cases (11 messages).

Regarding the tweets' impact, the vast majority of them (91.4%) have fewer than 10 retweets, including the five most intense messages (those ranging from 2 to 4); 5.9% of the tweets had between 10 and 100 retweets; and 2.6% have between 981 and 2,969. The figures are similar in the number of comments, since 94% of the tweets analysed got fewer than 10 comments. Also, among them are the tweets with the most intense hatred: 3.3% have between 10 and 100 comments and 2.6% between 199 and 650 comments. These last four tweets coincide with the most retweeted.

3.1.1. Political Positioning

The core of the criticism towards Penélope Cruz on social media focuses on her political positioning (136 tweets), something that is constantly criticised through ideologically tinged messages. A large number of tweets in this category criticise areas as diverse as her political affiliation (71 tweets), her participation in an advertising campaign for the airline Fly Emirates (42 tweets), the choice of the hospital where she gave

birth (39 tweets), her friendship with Pedro Almodóvar and the subsidies for Spanish cinema (27 tweets), or her position on international politics (11 tweets). All of this, apparently disparate, is integrated into a generalised criticism of her explicit adherence to progressive actions, substantiated by her support for the LGBTBI community, defence of public healthcare, or opposition to the violence of the Iranian regime; as well as the continuous recrimination by certain sectors of the X community of actions they consider to be incoherent. Such criticism supposes a mix of animosity towards her personally, and towards Spanish cinema in general, as happens with tweets about subsidies and production. This criticism assumes a moderate tone (“Do Penélope Cruz or Almodóvar share the millions we pay for their films with our taxes?”) but is sometimes also hostile (“How ridiculous the woman is! Penélope Cruz and Javier Bardem. That scum only knows how to make money from the nobodies who go to see their mediocre, subsidised films. Penélope Cruz has no morals!”).

Another aspect of hate speech towards the actor is determined by the widespread assumption that she supports communist ideology. There are frequent tweets such as: “Javier Bardem is a communist and so is Penélope Cruz. They didn’t win the Oscar for being good actors, but because the Academy is infested with communists: the downfall of the United States”; “For mental health reasons, I’ve gone years without seeing anything with the shitty communists the Bardems and Penélope Cruz”; “I refuse to watch any Bardem or Penélope Cruz movies. Bargain-basement commies”; “Penelope Cruz from Alcobendas? Or is she a communist? Well, it’s all the same. A commie from Santa Barbara, LA.” This animosity stems from the idea of hypocrisy that some X users have of the Penélope Cruz who has such an apparent ideology and lives in the world of cinema, this being one of the commonest criticisms: “Like Bardem and Penélope Cruz who live in Calcutta in Mother Teresa’s hospice helping selflessly, right? No. But long live the farce.” As can be seen, this last tweet is highly ironic, despite which, its level of aggression is not tempered, but rather, even more evident. This is a constant in all the tweets with a certain degree of humour: “Of course, Javier Bardem and Penélope Cruz live in a slum, right? Yeah, the mask is slipping.”

3.1.2. Public Healthcare and Cedars-Sinai Medical Center

The actor and her husband, Javier Bardem, became parents with the birth of their first child in January 2011, the birth taking place in the Cedars-Sinai Medical Center private clinic in Los Angeles. This centre, founded by Jews, is frequently confused by X users with the Mount Sinai Hospital in Manhattan, founded by the New York Jewish community. X users reported the birth, criticising the actor for defending public healthcare when she herself has chosen a private hospital to give birth. It is worth mentioning that both actors, working in Hollywood studios, “belong by law to the union and their insurance covers medical expenses one hundred percent” (R. González, 2011).

Criticism also points to the hospital where she gave birth to her second child, the Ruber International Clinic in Madrid; the majority of tweets criticising the actor and her husband on this particular point are especially polarised in two senses: firstly, because both actors defend Spanish public healthcare, and, secondly, because the American hospital is Jewish-funded, something that also exacerbates anti-Semitic comments: “Penelope Cruz and her husband? The woman who gave birth at Mount Sinai, private, Jewish and in the United States.” Other tweets hammer away about this: “Penelope Cruz closed a whole floor at Mount Sinai to give birth at ease. Follow her for more posh-progressiveness”; “Almodóvar, Bardem, Luis Tosar, Penélope Cruz...and 85% of left-wing officials and politicians, who choose private healthcare”; “I guess Javier Barden will go to the

demonstration, Mount Sinai Hospital, €2,500 a day, where Penélope Cruz gave birth”; or “This socialist who rented an entire floor of a private hospital so that his wife could have her baby (Penélope Cruz) and they live in the US, where they don’t pay for the public healthcare that they defend.”

The criticism is extended to the 2023 Goyas’ Gala, where the importance of public healthcare was revindicated, despite the fact that the private company Quirón Salud Sevilla was the official medical service of the awards. That fact forced the organisation itself to justify the hiring of two ambulances, since “in such a long a ceremony, public healthcare personnel could not be used” (Salinas, 2023), although its impact also affected Penélope Cruz, this being clear in tweets with a patently ironic air: “Thank you for this gala with so many calls for public healthcare. Our protest event has been sponsored by Salud Quirón, Ruber Internacional, Mount Sinai Hospital, Pedro Almodóvar and Penélope Cruz.”

3.1.3. Iran and the Fly Emirates Controversy

Another of the axes of ideological criticism of Penélope Cruz is due to the actor’s participation in an advertising campaign for the Qatari airline Fly Emirates. The ads, directed by Robert Stromberg and produced by Ridley Scott Associates, feature Penélope Cruz as the face and ambassador of the airline (Fly Emirates, 2023). This collaboration led to considerable ideological criticism, due to the supposed incoherence of her position: “Penélope Cruz is very outspoken about feminism and is now an ambassador for Fly Emirates. Our social elites are so obscenely hypocritical that only a masterpiece of social engineering explains why there are no massive torchlit marches.”

Qatar’s human rights policies, especially with regard to women’s rights and the LGTBI community, are a constant in the tweets directed at the actor on this matter: “For those who are still dazzled by Spain’s progressive artistry, Penélope Cruz advertises Emirates airlines. She’s so feminist, so LGTBI, so communist. They are all just fakes”; “Fuck the communists of the new false neo-left. How shameless they are, they’re pitiful, all for the money”; “How much do you think they paid Penélope Cruz to make that advert?”; “She and her husband are communists, they curse capitalism, they call big businessmen like Ortega exploiters, but when push comes to shove, they don’t put anything they preach into practice.”

The actor is also criticised for speaking out against climate change, while she advertises one of the most polluting modes of transport:

This lady, who boasts of being progressive, is just another elite opportunist. Those who abhor the extreme right like Penélope Cruz does, but then advertise Qatar airlines. And the poor can’t even drive a delivery van because they emit CO2. Another “storming of the Bastille”!

This advertisement was also the occasion to rekindle another controversy involving the actor, who appeared in a video broadcast on social media in October 2022 cutting a lock of her hair to condemn the killing of Mahsa Amini, the young Kurdish woman arrested for wearing her hijab incorrectly and killed in police custody (Sahuquillo, 2022). Although the video featured 70 other Spanish and French artists (“Protestas en Irán,” 2022), only Penélope Cruz became the target of criticism for the amount of hair cut from her fringe, becoming a meme for her gesture of solidarity (“Penélope Cruz se convierte en un meme,” 2022). This was highlighted in May 2023 in numerous messages pointing to the inconsistency of supporting Qatar, while

criticising Iran. A large portion of them take a clearly humorous tone, which does not detract from their scathing criticism: “Penélope Cruz, who cut her fringe in solidarity with women who suffer repression in the Middle East, is now the new image of Emirates Airlines”; “One day Penelope Cruz is cutting her bangs in ‘solidarity’ with Iranian women, and the next she’s an Islamic religious police officer, screaming at you because you’re revealing some flesh”; and even “Penelope Cruz has agreed to cut off some fringe each time the ad is broadcast.” The criticism becomes denigrating when it moves into religious matters, especially antisemitic: “Waiting for Penélope Cruz to fix her fringe again....Ah, no...the whole of post-progressive showbiz is on all fours offering its ass to Zionism.” And even topics unrelated to the Fly Emirates issue slide into antisemitism when talking about the couple: “They did the same thing to Penélope Cruz and Javier Bardem, when they mentioned the invasion of Palestine, the owners of Hollywood, from the self-chosen people, made them retract or their careers were over.”

3.2. Misogynistic Hate Speech

The labelling of tweets with misogynistic content (40 tweets) follows the trend of ideological content, with 18.4% of the messages located at level 0 (socially negative connotations), followed by 17 tweets (42%) corresponding to level 1 (hate speech that stigmatises the actor); four tweets (10%) reach level 2 (abusive expressions) and one (2.5%) even reaches level 4 (hate speech with veiled or implicit threats), which, given the nature of this section, implies a considerable degree of risk to the actor. That is, 55% of misogynistic tweets have stigmatising, abusive, or threatening content. It should be added that all of the inappropriate tweets are posted by male profiles (as they show in their own messages). Only three of the 40 tweets of a misogynistic nature offered a humorous-ridiculing tone, this being the lowest percentage of ironic content of all the categories analysed (7.5%). Misogynistic tweets have little impact on the number of retweets and comments. Only 15.8% received retweets (fewer than five retweets each), while 31.6% received comments (also fewer than five each).

3.2.1. Sexuality, Desire, and Animosity

The most insulting tweets (55%) are aimed at highlighting Penélope Cruz’s physical attributes, not to indicate the desire she arouses in the commenters, but to recriminate her for the dislike they feel; though they do feel attracted to the actor: “Penélope Cruz is really daft, but she’s very hot.” There are also those who reproach her for making them feel aroused: “How difficult it is for me to watch movies with Penélope Cruz because I can’t stop looking at her tits” or “My 140th hand job was dedicated to Penélope Cruz, the Galician left me dry as a bone.” And there are even those who express themselves, in sexual terms, contrary to the actor’s attractiveness: “I wouldn’t even masturbate to Penélope Cruz.”

3.2.2. Talent, Accusations, and Objectification of Her Body

The tone rises when they talk about the actor’s ability to act (10.25%), directly relating her talent to the means used to reach stardom: “The most frequently seen tits in Spain: Penélope Cruz.” Or extending the criticism to other actors, such as Aitana Sánchez Gijón: “Crap films that put me to sleep. She turns up in *Volavérunt*, some pseudo-historical bullshit by Bigas Luna. Even though there were some who usually flash the gash, like Penélope Cruz or Aitana Sánchez Gijón.”

The insults turn into accusations (five tweets, 12%) alleging the actor lacks talent and has instrumentalised her body to further her film career: “The thing about Penélope Cruz with all that American crap, lots of people said at the time it was a stratagem”; “Penelope Cruz. They even gave her an Oscar, she fucked the whole of Hollywood”; “Penelope Cruz talent? Other things more like”; “This is like Penélope Cruz going to bed with all the actors she made a movie with when she went to Hollywood.”

Finally, the remaining tweets (three, 7.5%) strive to position the actor as someone volatile and unintelligent: “She’ll be like Penélope Cruz, given the absence of thoughts, ideas and opinions of her own, she borrows those of her boyfriends.” As well as utilising casually vulgar expressions of a sexual nature and a certain tone of humour: “Penelope Cruz, a close friend of mine who I affectionately call ‘penis’” or rather “being called Penélope Cruz is like being called Vagina Martínez Cara.”

3.3. *Miscellany of Hatred*

The last category of hate speech towards Penélope Cruz (“Other”) is a miscellany of slurs, insults, and denigrating comments concerning her ability as an actor (33 tweets, 75%), followed by her physical (eight tweets, 18.2%) and intellectual (five tweets, 11.36%) characteristics. There are also comments of an ideological-misogynistic nature regarding her physique (six tweets, 13.6%). Two other messages refer to her home in the United States and her tax abode (4.5%). Although the majority of comments are level 0, that is, with socially negative connotations (33), nine tweets are level 1 (stigmatising her and her family) and two are level 2, aimed at spreading a clear negative image of the actor or, directly, the tweeter’s hatred towards her, as illustrated by this tweet: “Those of us who hate Penélope Cruz are few, but resistant and we know we’re right. We’re like brothers and as such we love and support each other.” Regarding the humorous slant of the tweets, this category is the one with the greatest use of humour, found in six of the 44 tweets (13.6%) and based, essentially, on a criticism of her ability as an actor. The impact of tweets in this category is much lower than that of ideological messages. Only one had retweets, two of them. And none of them had more than 10 comments. Those with greater intensity (levels 2 and 3) had no retweets and only one comment each.

3.3.1. *Criticism of Her Acting*

The category that comprises criticisms of her ability as a performer is made up of 33 tweets, which attack the Madrilenian from various perspectives, the most frequent being her lack of acting skill, her poor choice of scripts, or her voice. Tweets such as “I try not to hate...but Penélope Cruz is a really bad actor,” “Penelope Cruz is the worst actress in the world,” or “Penelope Cruz is bad, terrible, a horror as an actor” stand out. Comparisons are also made between actors: “Eduardo Casanova makes the worst Spanish actor good. Even Penélope Cruz”; “I wouldn’t hire Antonio Banderas and Penélope Cruz to do an advert on Canal Cádiz”; or “Anne Hathaway is not worth a damn (like Penélope Cruz), no matter what they say.”

In addition to these comments, there are others that attack her nominations for the Goya Awards, like “Penelope Cruz is overrated!!! And you know it!!!!” addressed to the Spanish Film Academy, as well as “How happy I am every time Penélope Cruz doesn’t win an award” and “Penelope Cruz, despite making, I dunno, an ad for asparagus, I get the feeling that she gets nominated by decree” also referring to the 2023 Goya Awards. This last comment goes further:

Do we have to give an award to all the thick women? Penélope Cruz, Goya for wearing a fucking fake ass in an Almodóvar film and talking like in all her films, like a redneck. Sofia, wife of a thief just like her, almost sanctified for wearing horns and swallowing. S.O.B.

This message, in addition to insulting the actor for her acting skills, not only includes both a derogatory comment towards her husband (calling him a thief) and a misogynistic one, but also ends with an insult veiled behind the acronym S.O.B., for “son of a bitch.” In fact, in this comment Penélope Cruz is referred to as “thick,” an insult that is repeated countless times in other comments and to which are added “dumbass” and “useless.”

3.3.2. Awards, Insults, and Being Over-Rated

Apart from the above tweets, there is one type of tweet that combines two types of insults. On the one hand, that of a lack of intellect, and, on the other, that of being a bad actor who wins awards: “And useless Penélope Cruz winning awards!” or “If a waste of space like Penélope Cruz has won it, Brendan deserves to win it” in reference to actor Brendan Fraser’s Oscar nomination for *The Whale*. Equally insulting is the tweet suggesting that the actor’s Oscar was due to a bribe: “Didn’t Penélope Cruz win the Oscar for best actress for *Vicky Cristina Barcelona*? You can buy anything.”

These tweets are completed with a series of comments about the idea of her work being over-rated: “Over-rated Spanish actor. I’ll start: Penélope Cruz”; “Top 3 scandalously over-rated things: *The Beach*. *The Little Prince*. Penélope Cruz.” To this last comment a thread was added in which, invariably, Penélope Cruz appears in all the lists: “11 overrated things: 1. Alcohol (all drinks that contain it). 2. Coffee. 3. Cow’s milk. 4. Penelope Cruz. 5. Tobacco...” There are also those who place the actor on a scale of acting talent: “I’m currently a firm believer that we send the worst actors we have to Hollywood, because how the fuck can Ana de Armas and Penélope Cruz [be] there, while Blanca Suárez and Úrsula Corberó are still here.”

3.3.3. Ugliness, Scripts, Screen Actors Guild, and Narration

The third branch of criticism of the actor is based on aesthetic criteria. In fact, 5% of the tweets refer to her supposed ugliness: “If she wasn’t pretty, she wouldn’t catch a cold. Now, Penélope Cruz is worse, for example, even after the plastic surgery she doesn’t do it for me” or “I’m watching a movie where Penélope Cruz appears. Christ, she’s ugly.” In addition to being unattractive, she is also criticised for her supposed vulgarity: “How can he kiss that cheap bitch?—Penelope Cruz” or “I don’t like her. I can’t stand Penélope Cruz. I don’t think she’s pretty. Ana Belén’s pretty.” Another criticism of her beauty and elegance is to be found in the tweets related to her presence on the red carpet at the Goyas, going so far as to say “Penelope Cruz must use up all her glamour in Hollywood, because in the interview on RTVE at the #Goya2023 she looks like a blinged-up DIA cashier (nothing against supermarket cashiers, lots of them are glamorous university students).”

Finally, the actor is constantly insulted as a representative of the union, as portrayed in tweets such as the one below, which combines animosity towards the profession, towards Spanish cinema, towards the marriage of Cruz and Bardem, and towards left-wing ideology: “Nobody goes to the movies to see actors as crummy as those in Spain today. Javier Bardem and Penélope Cruz are the best examples, a long, long way from the golden age of Spanish and North American cinema. Vulgar lefties.”

There are also tweets that allude to the poor quality of the actor's films: "Has Penélope Cruz made a single good film?" Or her limited skills at narrating films, this in reference to her dubbing of the documentary *Nuestro Planeta*: "If you have trouble getting to sleep, I strongly recommend you watch the documentary series *Nuestro Planeta*, narrated by Penélope Cruz. You could probably be more anodyne, but it'd be difficult" or "With the number of good voice actors there are in this country...whoever had the disastrous idea of putting Penélope Cruz as the narrator...it couldn't be done worse."

4. Discussion and Conclusions

It has been clarified that hate speech is not only a harmful manner of communication, but also the expression of discriminative ideals referred to a person (or group) based on the conviction of superiority. Comments against Penélope Cruz on the X social media platform represent hate speech because they offend Cruz while unfolding ideological misogynistic content against her. The tweets are not based on specific events or proven facts, but rather on conjecture, prejudice, and stereotypes, calling into question the actor's talent and the decisions she makes regarding her way of life or her own body. The results show that the comments directed towards her go beyond mere opinion, entering the realm of insult and contempt. Despite the scarcity and satirical intention of some tweets of a humorous or ironic nature, these are every bit as humiliating as the more direct comments.

This study coincides with others (Martínez-Valerio, 2021) in pointing out the difficulty of classifying content, especially if it is done exclusively by means of software. Although there are automatic language detection tools for the study of hate speech (Amores et al., 2021; Blanco, 2021; Felipe, 2020), on many occasions machines may have difficulty interpreting the comments made; an example would be the use of emoticons. Additionally, some users write ironically, so a comment that is actually favourable may be classified as unfavourable and vice versa. Therefore, it is essential to highlight the importance of continuing to work on methodologies that employ technological advances in linguistic analysis hand-in-hand with human knowledge.

The objective of this study was to explore the presence of hate speech directed at the actor Penélope Cruz on the social media platform X. After analysing 6,620 comments, it is concluded that hate speech is only present in 3.2% of the posts in which the actor is mentioned and mostly appears in its least harmful form on the scale of possible manifestations of hate speech (levels 0 and 1). The rest of the analysed comments could be considered criticism according to the definition of hate speech cited in the introductory part of this work. The percentage is very similar to that found in other studies focused on individuals related to the film industry, regardless of their gender (Martínez-Valerio & Tello-Díaz, 2024). Observing the classification of hate speech detected based on typology, the commonest kind of hate was ideological (70.9%), within which different aspects were detected, as was the case in the category "Other" (18.8%). Misogynistic messages were the third most common type (8.9%), in which she is accused of using her body to get ahead and of buying votes in the awards she has received. In addition to the illicit means attributed to her in her rise to stardom, her intelligence, beauty, acting talent, and way of thinking are also called into question. These results coincide with the messages analysed in the work of Brojakowski and Cruz (2022) and Karthika (2022). The first study examines the case of hate messages directed at the American actress Leslie Jones. Most of these messages were misogynistic or racist in nature. In the second case, that of the Indian actress Parvathy, the analysed messages were predominantly misogynistic. In the case of Penélope Cruz, some even question whether her ideology is her own, arguing that her opinions have been acquired by living with Javier Bardem, doubting her

intellectual capacity to make political decisions. This infantilisation or need for guardianship by her husband is also supposedly shown when judging the place and the way in which the actor gave birth to her children. Users of X claim Javier Bardem made the choice of and paid for the American hospital where their eldest child was born, denying the mother's agency to decide, contract, or pay for herself.

Apart from her personal qualities, the figure of Penélope Cruz is a focal point for the criticism that extends to the entire acting profession, her name appearing even when the news, comment, or criticism had another protagonist. This also happens when the actors' union, or any of its members, takes a political stance, citing the actor in critical comments, despite her neither being present nor participating in the relevant act or event. Moreover, all of this is taken to reveal supposed ideological incoherence on the part of the actor who, without publicly expressing herself in the way that some of her colleagues have done, is systematically accused of communism, reproaching her for the paradox of having progressive ideas and a comfortable lifestyle. The intersection between ideological hate and misogyny is a particularity of this Spanish case study. Other research has found different types of intersections, with the most frequent being between misogyny and racism (Brojakowski & Cruz, 2022; López, 2022). However, considering the politicization present in various areas of Spanish society, from sports (Bueno, 2013) to the Covid-19 pandemic (Herrero et al., 2024), it is not surprising that ideological hate intersects with misogyny and even extends to other actors beyond gender.

In short, the confusion between freedom of expression on one hand and defamatory and humiliating excesses on the other should be a priority for a platform which encourages public debate and the expression of opinions. Hence the importance of detecting and analysing comments in the absence of an authority that protects against insults and accusations, ensures the well-being of users, monitors the rise of hate speech, and applies a certain sense of proportion. It would be of interest to carry out future research in which hate speech against female Spanish actors is compared with that received by their peers from other countries, chiefly in order to analyse the ideological component detected in this case study.

Funding

This article is part of the project Taxonomy, Presence and Intensity of Hate Expressions in Digital Environments Linked to Spanish Professional News Media (PID2020-114584GB-I00, Ministry of Science and Innovation).

Conflict of Interests

The authors declare no conflict of interests.

References

- Álvarez, R. (2020, February 2). ¿Qué ayudas recibe el cine español y por qué está subvencionado? *Newtral*. <https://www.newtral.es/que-ayudas-recibe-el-cine-espanol-por-que-esta-subvencionado/20200202>
- Amores, J., Blanco-Herrero, D., Sánchez-Holgado, P., & Frías-Vázquez, M. (2021). Detectando el odio ideológico en Twitter. Desarrollo y evaluación de un detector de discurso de odio por ideología política en tuits en español. *Cuadernos.info*, 49, 98–124. <https://dx.doi.org/10.7764/cdi.49.27817>
- Aruguete, N. (2015). *El poder de la agenda. Política, medios y público*. Biblos.
- Blanco, R. (2021). *Identificación de lenguaje misógino a partir de minería de textos en redes sociales* [Unpublished master's thesis]. Universidad Nacional de Educación a Distancia. <https://hdl.handle.net/20.500.14468/14234>

- Brojakowski, B., & Cruz, G. A. (2022). Busting trolls: Examining the hate campaign against actress Leslie Jones. In L. Cuklanz (Ed.), *Gender violence, social media, and online environments* (pp. 76–93). Routledge.
- Bueno, J. G. (2013). La politización del deporte en los medios escritos especializados. *Razón y Palabra*, 17(3_84), 602–617.
- Burnap, P., Rana, O. F., Avis, N., Williams, M., Housley, W., Edwards, A., Morgan, J., & Sloan, L. (2015). Detecting tension in online communities with computational Twitter analysis. *Technological Forecasting and Social Change*, 95, 96–108. <https://doi.org/10.1016/j.techfore.2013.04.013>
- Campos, P. (2016, December 14). ¿Es tan de izquierdas el cine español? Cara a cara entre los diputados actores. *El Confidencial*. https://www.elconfidencial.com/cultura/2016-12-14/rosana-pastor-felix-alvarez-felisuco-congreso-actores-diputados_1299677
- Corse, E. R., León, N. N., Patiño, B. X., Rocabado, C. E., Sosa, M. A., & Ventura, E. A. (2014). Sexismo lingüístico y estereotipos de género manifestados en el uso del lenguaje en redes sociales por la comunidad estudiantil de LAEL. *Kipus Científicus*, 1(1), 139–165.
- Davidson, T., Warmley, D., Macy, M., & Weber, I. (2017). Automated hate speech detection and the problem of offensive language. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 512–515.
- De Lucas, V. A., Römer, M., Izquierdo-Alegría, D., Montero-Díaz, J., & Said-Hung, E. (2022). *Manual para el etiquetado de mensaje de odio en medios digitales en España*. Gobierno de España, Ministerio de Ciencia e Innovación. <https://doi.org/10.6084/m9.figshare.18316313>
- Escalante, H. J., Villatoro-Tello, E., Garza, S. E., López-Monroy, A. P., Montes-y-Gómez, M., & Villaseñor-Pineda, L. (2017). Early detection of deception and aggressiveness using profile-based representations. *Expert Systems With Applications*, 89, 99–111. <https://doi.org/10.1016/j.eswa.2017.07.040>
- Felipe, P. (2020). *Detección de discurso de odio en redes sociales* [Unpublished master's thesis]. Universidad del País Vasco. <http://hdl.handle.net/10810/48827>
- Fernández, L. (2012, May 17). Javier Bardem se reconcilia con sus 'enemigos' de la prensa. *Vanitatis*. *El Confidencial*. https://www.vanitatis.elconfidencial.com/noticias/2012-05-17/javier-bardem-se-reconcilia-con-sus-enemigos-de-la-prensa_398849
- Fersini, F., Rosso, P., & Anzovino, M. (2018). Overview of the task on automatic misogyny identification at IberEval 2018. In P. Rosso, J. Gonzalo, R. Martínez, S. Montalvo, & J. Carrillo de Albornoz (Eds.), *Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages* (pp. 214–228). CEUR-WS.
- Fly Emirates. (2023, May 25). *Emirates nombra a Penélope Cruz como su nueva embajadora de marca* [Press release]. <https://www.emirates.com/media-centre/es/emirates-nombra-a-penelope-cruz-como-su-nueva-embajadora-de-marca>
- Fox, J., Cruz, C., & Lee, J. Y. (2015). Perpetuating online sexism offline: Anonymity, interactivity, and the effects of sexist hashtags on social media. *Computers in Human Behavior*, 52, 436–442. <https://doi.org/10.1016/j.chb.2015.06.024>
- Fox, J., & Tang, W. Y. (2015, May 21–25). *Women's experiences with harassment in online video games: Rumination, organizational responsiveness, withdrawal, and coping strategies* [Paper presentation]. 65th Annual Conference of the International Communication Association, San Juan, PR, United States.
- Frenda, S., Ghanemb, B., Montes-y-Gómez, M., & Rosso, P. (2019). Online hate speech against women: Automatic identification of misogyny and sexism on Twitter. *Journal of Intelligent & Fuzzy Systems*, 36(5), 4743–4752. <https://doi.org/10.3233/JIFS-179023>

- Gabinete de Estudios Económicos de Axesor. (2013, July 30). Las empresas más subvencionadas en España. Axesor. <https://www.axesor.es/blog/post/2013/07/30/Las-empresas-mas-subvencionadas-de-Espana.aspx>
- Bretos, A. (Host). (2023, February 3). Aimar Bretos entrevista al escritor y guionista, fundador de la revista 'Mongolia.' [Podcast Episode]. In *Hora veinticinco*. <https://cadenaser.com/audio/1675454909235>
- García-de-Torres, E., Yezers' Ka, L., Rost, A., Calderín, M., Edo, C., Rojano, M., & Corredoira, L. (2011). Uso de Twitter y Facebook por los medios iberoamericanos. *Profesional de la Información*, 20(6), 611–620.
- Garcías, D. (1998). *Antonio Jordada y su Tratado contra las comedias (1788)*. *Gramma Y Cal: Revista Insular de Filología*, 2, 35-48. <https://raco.cat/index.php/GrammaYCal/article/view/6586>
- Ghaffari, S. (2020). Discourses of celebrities on Instagram: Digital femininity, self-representation and hate speech. *Critical Discourse Studies*, 19(2), 161–178. <https://doi.org/10.1080/17405904.2020.1839923>
- González, R. (2011, January 26). El parto millonario de Penélope Cruz. *La Razón*. https://www.larazon.es/historico/2240-penelope-y-javier-padres-NLLA_RAZON_355127
- González, S., & Ramos, F. (2013). El uso periodístico de Facebook y Twitter: Un análisis comparativo de la experiencia europea. *Historia y Comunicación Social*, 18, 419–433. https://doi.org/10.5209/rev_HICS.2013.v18.44253
- González-Aguilar, J., Segado-Boj, F., & Makhortykh, M. (2023). Populist right parties on TikTok: Spectacularization, personalization, and hate speech. *Media and Communication*, 11(2), 232–240. <https://doi.org/10.17645/mac.v11i2.6358>
- Hanash, M. (2018). Disciplinamiento sexual: Cazando brujas y ciberfeministas. Investigación y género. Reflexiones desde la investigación para avanzar en igualdad. In I. Vázquez, M. J. Cala, A. Guil, C. García-Gil, R. Martínez, & C. Flecha (Eds.), *VII Congreso Universitario Internacional Investigación y Género* (pp. 339–350). SIEMUS.
- Herrero, D. B., Calderón, C. A., & Torrealba, M. T. (2024). Pandemia, politización y odio: Características de la desinformación en España. *Estudios Sobre el Mensaje Periodístico*, 30(3), 503–515. <https://doi.org/10.5209/emp.96593>
- Igartua, J. J. (2006). *Métodos cuantitativos de investigación en comunicación*. Bosch.
- Justo, R., Corcoran, T., Lukin, S. M., Walker, M., & Torres, M. I. (2014). Extracting relevant knowledge for the detection of sarcasm and nastiness in the social web. *Knowledge-Based Systems*, 69, 124–133. <https://doi.org/10.1016/j.knosys.2014.05.021>
- Karthika, C. (2022). Cyberbullying on celebrities: A case study on actress Parvathy. *AIP Conference Proceedings*, 2463(1), Article 020036.
- Kopytowska, M., & Baider, F. (2017). From stereotypes and prejudice to verbal and physical violence: Hate speech in context. *Lodz Papers in Pragmatics*, 13(2), 133–152. <https://doi.org/10.1515/lpp-2017-0008>
- Lamson, A., De Souza, L., & Martinez, A. (2022). Freedom of speech and hate speech: An American perspective. *Revista de Direitos e Garantias Fundamentais*, 23(2), 31–56. <https://doi.org/10.18759/rdgf.v23i2.2029>
- Lapidot-Lefler, N., & Barak, A. (2012). Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in Human Behavior*, 28(2), 434–443. <https://doi.org/10.1016/j.chb.2011.10.014>
- López, Y. (2022). Exploring narratives of hate and solidarity on social media: The case of Yalitza Aparicio. *Interculture Journal: Online-Zeitschrift für interkulturelle Studien*, 21(36), 107–127. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-80117-9>
- Mao, L. (2019). *Rhetoric, cultural identity and intercultural communication in the film: China and Spain in the comparison between the Chinese director Zhang Yimou and the Spanish director Pedro Almodóvar* [Unpublished doctoral dissertation]. Universidad de León.

- Martínez-Valerio, L. (2021). Mensajes de odio hacia la comunidad LGTBQ+: Análisis de los perfiles de Instagram de la prensa española durante la “Semana del Orgullo.” *Revista Latina de Comunicación Social*, 80, 363–388. <https://doi.org/10.4185/RLCS-2022-1749>
- Martínez-Valerio, L., & Mayagoitia-Soria, A. M. (2021). Influencers y mensajes de odio: Jóvenes y consumo de contenidos autocensurados. *Revista Prisma Social*, 34, 4–39.
- Martínez-Valerio, L., & Tello-Díaz, L. (2024). Discursos de odio dirigidos a la industria cinematográfica española. In A. Moreno-Delgado, E. Said-Hung, & M. Römer (Eds.), *Expresiones de odio en entornos digitales españoles* (pp. 169–192). Tirant Humanidades.
- Marwick, A. (2013). Gender, sexuality, and social media. In T. Senft & J. Hunsinger (Eds.), *The social media handbook* (pp. 59–75). Routledge.
- Mayo-Cubero, M. (2019). Uso de las redes sociales en la cobertura periodística de crisis, desastres y emergencias en España. *Revista Española de Comunicación en Salud*, 2019(Suppl. 1), 43–54. <https://doi.org/10.20318/recs.2019.4428>
- Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016). Abusive language detection in online user content. In J. Bourdeau, J. A. Hendler, & R. Nkambou (Eds.), *Proceedings of the 25th International Conference on World Wide Web* (pp. 145–153). International World Wide Web Conferences Steering Committee of the Republic and Canton of Geneva.
- Pedersen, S., & Macafee, C. (2007). Gender differences in British blogging. *Journal of Computer-Mediated Communication*, 12(4), 1472–1492. <http://doi.org/10.1111/j.1083-6101.2007.00382.x>
- Penélope Cruz se convierte en un meme tras cortarse el flequillo por las mujeres iraníes. (2022, October 9). *Faro de Vigo*. <https://www.farodevigo.es/vida-y-estilo/gente/2022/10/09/penelope-cruz-convierte-meme-cortarse-77014153.html>
- Penny, L. (2014). *Unspeakable things: Sex, lies, and revolution*. Bloomsbury.
- Piñeiro-Otero, T., & Martínez-Rolán, X. (2021). Eso no me lo dices en la calle. Análisis del discurso del odio contra las mujeres en Twitter. *Profesional de la Información*, 30(5), Article e300502. <https://doi.org/10.3145/epi.2021.sep.02>
- Piñuel, J. L. (2002). Epistemología, metodología y técnicas del análisis de contenido. *Estudios de Sociolingüística: Lenguas, Sociedades e Culturas*, 3(1), 1–42.
- Protestas en Irán. Las actrices españolas se cortan el pelo para sumarse a la ola de solidaridad con las mujeres iraníes. (2022, October 7). RTVE. <https://www.rtve.es/noticias/20221007/iran-actrices-espanolas-suman-solidaridad-mujeres-iranies/2405486.shtml>
- Riaño, R. (2021, March 7). Los insultos machistas que recibieron las actrices en los Goya: Comunicado de TVE. *Vanitatis. El Confidencial*. https://www.vanitatis.elconfidencial.com/famosos/2021-03-07/insultos-actrices-premios-goya-2021_2981331
- Sahuquillo, M. R. (2022, October 19). Mahsa Amini, cuya muerte bajo custodia policial en Irán prendió las protestas por la libertad, premio Sájarov del Parlamento Europeo. *El País*. <https://elpais.com/internacional/2023-10-19/mahsa-amini-cuya-muerte-bajo-custodia-policial-en-iran-prendio-las-protestas-por-la-libertad-premio-sajarov-del-parlamento-europeo.html>
- Said-Hung, E. M., Merino-Arribas, M., & Sánchez-Esparza, M. (2023). Polarización, desinformación y expresiones de odio en Twitter. Caso grupos políticos nacionalistas e independentistas en España. *Observatorio (OBS*)*, 17(2), 52–80. <https://doi.org/10.15847/obsOBS17220232179>
- Salinas, N. (2023, February 14). La Academia de Cine elude la polémica y justifica que solo contrató a dos ambulancias privadas para los Goya. *El periódico de España*. <https://www.epe.es/es/sanidad/20230214/academia-cine-elude-polemica-justifica-82945241>

- Sánchez-Meza, M., Schlesier-Corrales, L., Visa-Barbosa, M., & Carnicé-Mur, M. (2023). ¿De redes sociales a redes del odio? Análisis de la conversación digital en Twitter sobre la ministra de Igualdad española Irene Montero. *Estudios Sobre el Mensaje Periodístico*, 29(3), 717–736. <https://doi.org/10.5209/esmp.87271>
- Shaw, A. (2014). The internet is full of jerks, because the world is full of jerks: What feminist theory teaches us about the internet. *Communication and Critical/Cultural Studies*, 11(3), 273–277. <http://doi.org/10.1080/14791420.2014.926245>
- Sood, S., Antin, J., & Churchill, E. (2012). Profanity use in online communities. In J. Konstan, E. Chi, & K. Höök (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1481–1490). Association for Computing Machinery.
- Tarullo, R., Calvo, D., & Iranzo-Cabrera, M. (2024). #Feminazis on Instagram: Hate speech and misinformation, the ongoing discourse for enlarging the manosphere. *KOME*, 12(2), 91–110. <https://doi.org/10.17646/KOME.of.18>
- Terrón, A. (2020). Putas, históricas y maricones: Violencia de género en reseñas cinematográficas publicadas online por españoles. *Ambigua: Revista de Investigaciones Sobre Género y Estudios Culturales*, 7, 333–353.
- Yaben, M. (2021, April 22). De Rosalía a Cristina Pedroche: Los famosos españoles que 'se mojan' en política. *El Independiente*. <https://www.elindependiente.com/gente/2021/04/22/de-rosalia-a-cristina-pedroche-los-famosos-espanoles-que-se-mojan-en-politica>
- Zamora-Martínez, P., Gascón-Vera, P., & Gómez García, S. (2024). El discurso de odio hacia las candidatas a la Comunidad de Madrid 2023, sesgo de género y virulencia. *ICONO 14*, 22(1), Article e2079. <https://doi.org/10.7195/ri14.v22i1.2079>

About the Authors



Lucía Tello Díaz (PhD in information sciences from the UCM) is the coordinator of the MA in advanced film studies at the UNIR and lecturer in audiovisual communication. She teaches courses in scriptwriting for television and films at the MA in scriptwriting (UNIR), and film history at the ECAM. She is a director and a scriptwriter of documentary and feature film and has premiered her film *Endless Cinema* (2019) at the SEMINCI, and at the Buenos Aires International Film Festival. She collaborates with national and international television channels.



Lizette Martínez-Valerio is a graduate in journalism from the Complutense University of Madrid (UCM) with a cum laude European doctorate in information sciences from the UCM. She has been deemed a specialist in applied social research and data analysis by the CIS, and has been a professor (February 2016–February 2020) and director of the master in investigative journalism, data and visualization (December 2018–February 2020) at UNIR. She is now a full professor at Rey Juan Carlos University (since September 2019). Her research areas are (a) social networks and young people and (b) the history of television in Spain.

Intergenerational Evolution of Gender Bias in Spain: Analysis of Values Surveys

Pilar Antolínez-Merchán ¹ , Ángel Rivero Recuenco ² ,
and Elvira Carmen Cabrera-Rodríguez ³ 

¹ Faculty of Communication and Design, University Camilo José Cela, Spain

² Economy Department, University of Alcalá, Spain

³ Faculty of Entrepreneurship, Business and Legal Sciences, University Camilo José Cela, Spain

Correspondence: Pilar Antolínez-Merchán (pantolinez@ucjc.edu)

Submitted: 13 September 2024 **Accepted:** 20 November 2024 **Published:** 23 January 2025

Issue: This article is part of the issue “Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment” edited by Max Römer Pieretti (Universidad Camilo José Cela), Beatriz Esteban-Ramiro (Universidad de Castilla-La Mancha), and Agrivalca Canelón (Universidad Católica Andrés Bello), fully open access at <https://doi.org/10.17645/si.i415>

Abstract

This article uses data from the World Values Survey (WVS) and the European Values Study (EVS) for Spain, covering the years between 2005 and 2022 (waves 5, 6, and 7) to analyse the evolution of gender bias in different dimensions: politics, education, economics, and family. The results indicate a positive trend towards the reduction of gender bias, especially in areas of political leadership and education. However, certain biases remain, particularly among older generations. The analysis reveals that variables such as sex, education level, religion, political orientation, and materialistic values have a statistically significant influence on gender bias. Young people demonstrate a higher acceptance of gender equality compared to older adults. However, the younger generations are exposed to ideological and moral influences that cause changes in their perception of politics and democracy. One-fifth of the sample surveyed did not consider gender equality relevant as a constitutive element of democratic regimes, which seems to indicate a relative fading of the political and moral significance of gender equality as an issue for a significant proportion of young Spaniards. Religion is the only variable linked with a higher probability of maintaining gender biases, and even accentuating them among young people, which would be indicative of a correlation between religion and the adoption of ideologically conservative positions, in line with the socio-political dynamics of polarisation and the growing influence of neo-conservative movements in Spanish society.

Keywords

gender bias; gender equality; gender stereotypes; values survey

1. Introduction

The Office of the United Nations High Commissioner for Human Rights defines gender stereotypes as “a generalised view or preconception about attributes or characteristics, or the roles that are or ought to be possessed by, or performed by, women and men” (OHCHR, n.d.). The OHCHR also points out that gender stereotypes are harmful to both women and men as they affect their personal and professional spheres, perpetuate inequalities, and become unlawful when they violate human rights and fundamental freedoms. According to Amurrio et al. (2012, p. 228), “gender stereotypes are the characteristics, features and qualities that are attributed to a person based on their sex.” These authors point out that these stereotypes characterise and control the behaviour patterns that are expected both of men and women and which determine the models of masculinity and femininity, condemning behaviour that falls outside accepted gender norms.

The advance towards gender equality is a complex task for the development of society as a whole. In this sense, in the latest work carried out by the United Nations Development Programme (UNDP, 2023), entitled *Gender Social Norms Index (GSNI): Breaking Down Gender Biases: Shifting Social Norms Towards Gender Equality*, it is highlighted that gender biases are a generalised problem on a global level, and, in the case of women, no progress has been made in the last decade.

The GSNI measures prejudice against women through people’s attitudes about women’s roles along four dimensions: politics, education, economics, and physical integrity. The index is constructed from the answers to seven questions in the World Values Survey (WVS). Regarding this index, the application of the UNDP (2023) survey reveals that nearly nine out of 10 men and women have fundamental prejudices against women. For example, among the data, it is highlighted that 49% of the world’s population considers that men are better political leaders than women, and 43% state that men are better business executives than women.

In the face of this inequality, the report highlights how governments play an important role in changing social norms and, therefore, it is recommended that their approach be carried out through education and implementation of policies that promote gender equality and legislative changes, which address women’s rights in all spheres of life.

It is precisely in the field of education that UNESCO speaks out on gender bias and stereotypes. In their report *#Hereducationourfuture #Breakthebias: Challenging Gender Bias and Stereotypes in and Through Education* (UNESCO, 2022), and concerning children and young people, it is reported that these stereotypes and gender biases are established in people’s minds from the earliest stage of their lives and limit their future, given that these stereotypes can influence the choice of toys they play with, which subjects they study, their subsequent educational journey and, in general, throughout their lives. According to this report, if education doesn’t question the discriminatory gender norms and proposes opportunities and skills for critical thinking, nothing will change. Understanding gender stereotypes in the context of adolescence involves analysing the socialisation processes and the socio-cultural context in which the education of these young people has taken place, given that the beliefs they express stem from their past and may contribute to their behaviour as adults (Perry & Pauletti, 2011).

In this line, we will consider the study by González-Gijón et al. (2024), in which the average age of participants oscillates between 12 and 19 years old. According to this study, gender stereotypes are more

deeply rooted in boys than girls, although both have a high internalisation of prejudices. On the other hand, cultural background plays an important role in reinforcing gender beliefs derived from the upbringing styles and cultural values that have been part of their socialisation. Castillo and Montes (2007), in whose study the average age of participants was between 17 and 72 years old, conclude that educational level and political orientation emerge as the variables that have the greatest influence on gender stereotypes along with age, highlighting the importance of generational context.

Different international studies have used the WVS, a survey internationally recognised for its usefulness in the study of political, economic, and gender attitudes (Inglehart et al., 2004), to analyse issues related to gender equality (Brandt, 2011; Hussain & Haj-Salem, 2023; Inglehart & Norris, 2003; Norris & Inglehart, 2002; Rebrey, 2023; Rizzo et al., 2007; Tesch-Römer et al., 2008; Wernet, 2016).

Inglehart and Norris (2003) focus on gender equality and cultural change as a pairing that needs to be complemented by economic and political changes in any society to ensure equality between men and women. This allows for the consolidation and reinforcement of social change with legal and structural reforms that produce substantial advances for women.

Tesch-Römer et al. (2008) indicate that gender inequalities are related to gender differences. The extent of gender differences varies according to the degree of social inequality between men and women and cultural attitudes towards gender equality in the different countries under study. Thus, Egypt and Iraq show optimal development in this area in contrast to other countries such as Jordan and Morocco. Moreover, it is in line with the argument previously put forward by Inglehart and Norris (2003), who highlight the need to promote a change in the social positions that women hold in different spheres (employment, politics, etc.) to ensure greater social inclusion.

These gender inequalities are exemplified in another study by Rebrey (2023) in Russia. This research shows that although women actively participate in different areas of work with a high level of education, this does not guarantee their professional development and success, mainly due to low salaries. This makes it necessary to include a gender perspective in addressing the gender inequalities that are still prevalent in the country.

Several studies have also focused on the context of Islamic culture, democracy, and gender equality. Norris and Inglehart (2002) highlight that gender equality is a crucial aspect in understanding the cultural and political dynamics in Islamic cultures, suggesting that perceptions of gender are central to the analysis of democracy in these contexts. Rizzo et al. (2007) focus on the relationship between gender and democracy, while Hussain and Haj-Salem (2023) investigate how perceptions of women's emancipation have changed in the aftermath of the Arab Spring. All the studies challenge the idea that Islamic culture is inherently incompatible with democracy and gender equality, suggesting that there is significant variation and potential for change.

Lastly, it is important to highlight the studies by Brandt (2011) and Wernet (2016). Brandt (2011) indicates that sexism not only legitimises gender inequality, but actively worsens it by increasing it. Wernet (2016) reveals the links between social structure, the intermediate domains of religiosity and education, and gender equality attitudes.

Concerning social structure, Wernet (2016) argues that the history of communism has a different impact on attitudes towards gender equality compared to attitudes towards homosexuality. While tolerance towards

the former is increasing, tolerance towards the latter is decreasing. The same structural factor influences the two attitudes differently.

In terms of secularity, Wernet (2016) notes that the less important religion is in an individual's life, the more likely he or she is to support gender equality. The research findings show that being educated, being a woman, and being less religious significantly increases this probability. Finally, Wernet (2016) references the fact that educational level predicts gender equality attitudes and that institutional support for education, especially in less developed countries, can shape attitudes and lead to more equal opportunities for women in society.

The evolution of gender prejudices in Spain is a topic of growing relevance in the current sociopolitical context. In recent decades, Spain has undergone significant transformations in its social and political structures, which have influenced perceptions of gender equality. The transition to democracy in the 1970s allowed for the inclusion of equality policies, while feminist movements have pushed for the recognition of persistent inequalities (Bustelo, 2014; Lombardo & Alonso, 2020). Furthermore, the economic crisis of 2008 created both setbacks and new opportunities for equality activism (Lombardo & Alonso, 2020). These dynamics have shaped social awareness in order to highlight the importance of gender equality in Spain.

This article contributes to the understanding of the roots of gender biases in Spanish society. The general objective focuses on analysing the evolution of gender biases by age groups across four dimensions (political, educational, economic, and familial), identifying the sociodemographic and ideological factors that explain these biases. The analysis is centred on the youth group (ages 18–29). The specific objectives are:

1. To compare attitudes towards gender equality between different age groups.
2. To identify trends and factors that have an influence on gender bias.
3. To analyse the socio-demographic, political-ideological, and evaluative factors that explain gender prejudice in Spain.
4. To determine the explanatory power of the selected variables on gender bias.

2. Method

In this study, data available for Spain from the time series of the WVS and the European Values Study (EVS) for the period 2005–2022 have been used (Inglehart et al., 2022). This dataset combines the surveys completed in waves 5 (2005–2009), 6 (2010–2014), and 7 (2017–2022). Table 1 presents the samples arranged by waves, indicating the study (WVS or EVS).

Table 1. Number of interviews carried out in Spain per wave in the value surveys, EVS and WVS, 2005–2022.

	Wave	Frequency
WVS (2007)	2005–2009	1.200
EVS (2008)	2005–2009	1.497
WVS (2011)	2010–2014	1.189
EVS (2017)	2017–2022	1.210

Source: EVS-WVS_TimeSeries_InternalUse_spss_v3_0 (Inglehart et al., 2022).

The aggregated data file used for the analyses is the EVS-WVS_TimeSeries_InternalUse_spss_v3_0, provided by the WVS data archive (Inglehart et al., 2022). Details on documentation concerning the sample design, the method of data collection, the institution and researcher responsible for each survey, and other relevant documentation can be found on the WVS website as well (<https://www.worldvaluessurvey.org/WSDocumentationWVL.jsp>)

2.1. Measures

2.1.1. Primary Outcome Variables: Gender Bias

In order to observe the gender bias of young people in Spain, six indicators have been selected based on questions from the Values Survey questionnaire that measure egalitarian and non-egalitarian attitudes. These six indicators which function as dependent variables in our research, have been grouped into four dimensions: politics, education, economics, and family. The first three dimensions and their indicators coincide with those used by UNDP for the elaboration of the gender social norms index mentioned above (the fourth dimension that is part of the UNDP index, the physical dimension, could not be taken into account as the corresponding indicators have not been used in the latest wave of the survey).

It is essential to recognise the limitations associated with the questions selected. First, the number of questions, which is limited to six, may not be sufficient to capture the complexity of gender attitudes in various cultural and social contexts. Furthermore, by focusing only on four dimensions, important aspects of the issue may be omitted. It is also relevant to note that since gender norms are dynamic, questions can quickly become outdated. However, despite these limitations, these questions provide valuable initial insights into perceptions of gender equality and allow for the analysis of changes in attitudes towards gender over time.

Additionally, by focusing solely on four dimensions, we might be overlooking important aspects of this topic. It is also relevant to note that, since gender norms are dynamic, the questions may quickly become outdated. However, despite these limitations, these questions provide valuable initial insight into perceptions of gender equality and allow for the analysis of changes in gender attitudes over time.

Once the variables were selected, they were standardised so that they were all on the same scale and could be compared. For this purpose, each indicator takes a value of 1 when a respondent's answers imply a bias and 0 when they do not (Table 2).

Table 2. Gender bias indicators and normalisation.

Dimension	Indicator	Description	Normalisation
Politics	E233 Democracy: Women have the same rights as men	Please tell me for each of the following things how essential you think it is as a characteristic of democracy. Use this scale where 1 means <i>not at all an essential characteristic of democracy</i> and 10 means <i>it definitely is an essential characteristic of democracy; women have the same rights as men.</i>	1 = values from 0 to 7; 0 = rest

Table 2. (Cont.) Gender bias indicators and normalisation.

Dimension	Indicator	Description	Normalisation
Politics	D059 Men make better political leaders than women do	For each of the following statements I read out, can you tell me how much you agree with each? Do you strongly agree, agree, disagree, or strongly disagree? "On the whole, men make better political leaders than women do."	1 = <i>strongly agree</i> and <i>agree</i> ; 0 = rest
Education	D060 University is more important for a boy than for a girl	For each of the following statements I read out, can you tell me how much you agree with each? Do you strongly agree, agree, disagree, or strongly disagree? "A university education is more important for a boy than for a girl."	1 = <i>strongly agree</i> and <i>agree</i> ; 0 = rest
Economics	C001_01 Scarce jobs: Men should have more right to a job than women	Do you agree, disagree, or neither agree nor disagree with the following statements? "When jobs are scarce, men should have more right to a job than women."	1 = <i>agree</i> ; 0 = rest
Economics	D078 Men make better business executives than women do.	For each of the following statements I read out, can you tell me how strongly you agree or disagree with each? Do you strongly agree, agree, disagree, or strongly disagree? "On the whole, men make better business executives than women do."	1 = <i>strongly agree</i> and <i>agree</i> ; 0 = rest
Family	D061 Pre-school child suffers with working mother.	For each of the following statements I read out, can you tell me how much you agree with each? Do you strongly agree, agree, disagree, or strongly disagree? "When a mother works for pay, the children suffer."	1 = <i>strongly agree</i> and <i>agree</i> ; 0 = rest

Note: The sample is made up of the subjects who were asked about the above indicators (3,595 respondents).

To extract an index that gives all dimensions equal weighting, the Alkire and Foster (2011) methodology was used. This is the same counting methodology used for the development of the UNDP Gender Social Standards Index (Mukhopadhyay et al., 2019). The weighted average of the normalised variables is calculated by assigning equal weights to each indicator (Table 3).

Table 3. Gender bias index calculation.

Dimension	Indicator	Normalisation	Weight	Final Weight
Politics	E233 Democracy: Women have the same rights as men.	1 (values from 0 to 7)	1/2	$2/6 \times 1/2 = 1/6$
	D059 Men make better political leaders than women do.	1 (<i>agree</i> and <i>strongly agree</i>)	1/2	$2/6 \times 1/2 = 1/6$
	Global		2/6	
Education	D060 University is more important for a boy than for a girl.	1 (<i>agree</i> and <i>strongly agree</i>)	1	$1/6 \times 1 = 1/6$
	Global		1/6	

Table 3. (Cont.) Gender bias index calculation.

Dimension	Indicator	Normalisation	Weight	Final Weight
Economic	C001_01 Scarce jobs: Men should have more right to a job than women.	1 (<i>agree</i>)	1/2	$2/6 \times 1/2 = 1/6$
	D078 Men make better business executives than women do.	1 (<i>agree</i> and <i>strongly agree</i>)	1/2	$2/6 \times 1/2 = 1/6$
	Global		2/6	
Family	D061 Pre-school child suffers with working mother.	1 (<i>agree</i> and <i>strongly agree</i>)	1	$1/6 \times 1 = 1/6$
	Global		1/6	

Notes: Gender bias index = $1/6 \times (\text{women democracy}) + 1/6 \times (\text{men political leaders}) + 1/6 \times (\text{university}) + 1/6 \times (\text{jobs for men}) + 1/6 \times (\text{business executives}) + 1/6 \times (\text{working mother})$.

Regarding the interpretation of the gender bias index, the value in the index closest to 1 is interpreted as evidence of gender bias.

2.1.2. Predictor or Independent Variables: Socio-Demographic and Political-Ideological Variables

As for the predictor or independent variables, based on evidence from previous studies (Bissell & Parrott, 2013; Santoniccolo et al., 2023), two blocks of variables were included:

1. Socio-demographic variables: sex, education level, and income level.
2. Political-ideological variables: political-ideological self-placement, religion, and materialist/post-materialist index (Inglehart's 4-item scale corresponding to question Y002 of all survey waves).

Table 4 shows the variables used and their description, as well as the recordings made.

Table 4. Independent variables: Description and recoding.

Indicator	Description
X001 (sex)	Respondent's sex (Code respondent's sex by observation, don't ask about it!): (1) men; (2) women
X025 (highest educational level)	What is the highest educational level that you have attained: (0) less than primary, (1) primary, (2) lower secondary, (3) upper secondary, (4) post-secondary non-tertiary, (5) short-cycle tertiary, (6) bachelor or equivalent, (7) master or equivalent, (8) doctoral or equivalent
Recoding: X025R (highest educational level recoded in three groups)	0–2 (X025A-1) → 1 Lower (X025R); 3–4 (X025A-1) → 2 Middle (X025R); 5–8 (X025A-1) → 3 Upper (X025R)
X047 (Scale of incomes)	On this card is an income scale on which 1 indicates the <i>lowest income group</i> and 10 the <i>highest income group</i> in your country. We would like to know in what group your household is. Please specify the appropriate number, counting all wages, salaries, pensions, and other incomes that come in.

Table 4. (Cont.) Independent variables: Description and recoding.

Indicator	Description
Recoding: X047R (subjective income level recoded in three groups)	1–3 (X047) → 1 Low (X047_R3); 4–7 (X047) → 2 Middle (X047_R3); 8–10 (X047) → 3 High (X047_R3)
F034 (religious person)	Independently of whether you go to church or not, would you say you are: (1) a religious person; (2) not a religious person; (3) an atheist
E033 (left-right political scale)	In political matters, people talk of “the left” and “the right.” How would you place your views on this scale, generally speaking: (1) left...(10) right.
Recoding: E033 (left-right political scale recoded in three groups)	1–4 (E033) → 1 Left (E033_R3); 5–6 (E033) → 2 Center (E033_R3); 7–10 (E033) → 3 Right (E033_R3)
Y002 (post-materialist index; 4-item)	4-item scale: (1) materialist; (2) mixed; (3) post-materialist

In the materialist/post-materialist scale, respondents are asked to select two of their country’s most important objectives from the following four options:

1. Maintain the nation’s order.
2. Increase citizen participation in important government decisions.
3. Combat price increases.
4. Protect freedom of expression.

Respondents who selected “maintain order” and “combat rising prices” were classified as materialists, and those who chose “increase participation” and “freedom of expression” were classified as post-materialists. The remaining combinations (one materialist response and another post-materialist) are considered “mixed.”

2.1.3. Control Variable: Age

Variable X003 (age) has been selected where it asks: “This means you are [blank space] years old (write age in two digits).” Its recoding X003R2 (age recoded in three groups) was: (1) 18–29 years; (2) 30–49 years; (3) 50 years and over.

2.2. Statistical Analysis

In order to carry out our analysis, we used the software IBM SPSS Statistics (version 20). First, a descriptive analysis was carried out for the sample as a whole and for the different age groups to find out the evolution of the gender bias index by age.

Subsequently, a logistic regression analysis was developed to account for factors influencing gender bias in terms of age. The gender bias index was taken as the dependent variable. Two regression models have been run, taking into account two age groups (18–29 years and 30 years and over). The independent variables

included in the two models are those mentioned above transformed into binary variables, except for the ideological self-placement variable which, being a scale variable, did not need to be transformed. Table 5 shows the construction of these binary variables for the analysis.

Table 5. Independent variables included in the logistic regression models.

Original variables	Binary variables	Values
X001 (sex)	Sex (binary)	1 = man; 0 = woman
X003 (age)	Age (binary)	1 = 18–29 years; 0 = 30 years and over
X025 (highest educational level)	Low educational level	1 = low educational level; 0 = the rest
	Medium educational level	1 = medium educational level; 0 = the rest
	High educational level	1 = high educational level; 0 = the rest
X047 (scale of incomes)	Low income	1 = low income; 0 = the rest
	Medium income	1 = medium income; 0 = the rest
	High income	1 = high income; 0 = the rest
F034 (religious person)	Religious person	1 = religious person; 0 = not religious person
Y002 (post-materialist index; 4-item)	Post-materialist	1 = post-materialist; 0 = the rest
	Mixed	1 = mixed; 0 = the rest
	Materialist	1 = materialist; 0 = the rest

3. Results and Analysis

Table 6 provides a detailed view of how gender biases have evolved in Spain across the three studied waves: 2005–2009, 2010–2014, and 2017–2022. The indicators are grouped into four key dimensions: political, educational, economic, and family. Each indicator shows the percentage of unbiased and biased people in each wave, providing a measure of the changes in attitude towards gender equality in different years.

In general, the data shows a positive trend towards the reduction of gender bias in Spain in most of the key dimensions considered; the exception is in the political dimension (perception of equal rights for democracy), which shows a slightly increasing trend over the waves considered.

3.1. Breakdown

3.1.1. Political Dimension

In the category of essential equal rights for democracy: The view that women should have the same rights as men as it is essential for democracy increased slightly from 19.2% to 19.8%.

In the category of political leadership: The belief that men are better political leaders than women has decreased significantly, with the percentage of biased people falling from 20.9% to 9.45%.

Table 6. Gender biases by wave and indicator (Spain 2005–2022).

Dimension	Indicator		WAVE (2005–2009)	WAVE (2010–2014)	WAVE (2017–2022)
Political	Women having the same rights as men is essential for democracy.	Share of people with no bias	80.8%	80.4%	80.2%
		Share of biased people	19.2%	19.6%	19.8%
	Men make better political leaders than women do.	Share of people with no bias	79.1%	84%	90.6%
		Share of biased people	20.9%	16.0%	9.4%
Educational	University is more important for men than for women.	Share of people with no bias	86.7%	88.3%	95.5%
		Share of biased people	13.3%	11.7%	4.5%
Economic	Men should have more right to a job than women.	Share of people with no bias	82.2%	87.7%	88.7%
		Share of biased people	17.8%	12.3%	11.3%
	Men make better business executives than women do.	Share of people with no bias	81.9%	85.5%	92.7%
		Share of biased people	18.1%	14.5%	7.1%
Family	Pre-school children suffer with a working mother.	Share of people with no bias	Not asked	71.5%	73.7%
		Share of biased people	Not asked	28.5%	26.3%

Source: EVS-WVS_TimeSeries_InternalUse_spss_v3_0 (Inglehart et al., 2022).

3.1.2. Educational Dimension

Considering the importance of university: The idea that university is more important for men than for women decreased from 13.3% to 4.5%.

3.1.3. Economic Dimension

In the category of right to work: The belief that men should have more right to a job than women decreased from 17.8% to 11.3%.

In the category of business executives: The perception that men are better business executives than women has also decreased from 18.1% to 7.1%.

3.1.4. Family Dimension

Concerning the impact of working mothers on pre-school children: The belief that a pre-school child suffers if their mother works has decreased from 28.5% to 26.3% between 2010–2014 and 2017–2022.

3.2. Comparison of Attitudes Towards Gender Equality Between Different Age Groups

Table 7 shows the evolution of gender bias in Spain, broken down by ages (18–29 years, 30–49 years and 50 years or over), allowing a detailed analysis of how attitudes towards gender equality vary between different generations.

Table 7. Gender biases in Spain (2005–2022) by wave, age group, and indicator.

Dimension	Indicator	Age group								
		18–29			30–49			50 years and over		
		2005–2009	2010–2014	2017–2022	2005–2009	2010–2014	2017–2022	2005–2009	2010–2014	2017–2022
Political	Women having the same rights as men is essential for democracy.	16.3%	20.1%	22.3%	18.0%	17.5%	19.8%	21.9%	21.4%	19.2%
	Men make better political leaders than women do.	13.2%	11.0%	3.0%	17.5%	14.8%	6.5%	28.4%	19.6%	13.6%
Educational	University is more important for men than for women.	10.70%	10.7%	1.7%	11.9%	9.9%	2.7%	16.1%	14.0%	6.6%
Economic	Men should have more right to a job than women.	9.7%	5.2%	5.3%	17.2%	8.0%	7.5%	22.8%	20.1%	15.8%
	Men make better business executives than women do.	13.6%	13.2%	2.4%	14.5%	9.9%	5.1%	24.0%	19.7%	10.0%
Family	Pre-school children suffer with a working mother.	Not asked	21.8%	21.9%	Not asked	24.0%	24.0%	Not asked	36.4%	29.3%

Source: EVS-WVS_TimeSeries_InternalUse_spss_v3_0 (Inglehart et al., 2022).

In general, the data show a positive trend towards the reduction of gender bias in Spain in different key dimensions and age groups, with some differences in the different dimensions and age groups considered:

3.2.1. Breakdown

3.2.1.1. Political Dimension

In the category of equal rights are essential for democracy: Although there was a slight increase in the percentage of people with biases in the groups 18–29 years old and 30–49 years old, the group of 50 years or over demonstrated a decrease in these biases. This suggests that younger generations may be exposed to ideological influences that cause shifts in their perceptions of politics and democracy.

In the category of political leadership: The belief that men are better political leaders than women has decreased significantly in all age groups, which clearly indicates positive changes in women's acceptance of taking on political leadership roles.

3.2.1.2. Educational Dimension

Concerning the importance of university: The perception that university is more important for men than for women has decreased drastically in all age groups, especially among young people from 18–29 years old. This reflects a cultural shift towards equal educational opportunities for both genders.

3.2.1.3. Economic Dimension

In the category of right to work: The belief that men should have a greater right to work than women has decreased in all age groups, with a significant reduction in the 30–49 years age group. This indicates a greater acceptance of gender equality in the workplace.

Concerning business executives: The perception that men are better business executives than women also decreased in all age groups, especially among young people from 18–29 years, which shows positive changes in the assessment of gender equality in terms of taking on business leadership roles.

3.2.1.4. Family Dimension

On the impact of mothers' work on pre-school children: The belief that a pre-school child suffers if their mother works has decreased slightly in the group of 30–49 years old and 50 years or over, with a notable reduction in the group 50 years or over. This suggests a growing acceptance of working mothers and a decrease in gender stereotypes in the family sphere.

3.3. Analysis of Sociodemographic, Political-Ideological, and Evaluative Factors: Gender bias index

The gender bias index in Spain was elaborated using six indicators from the previously mentioned, both the WVS and the EVS. The index is calculated as the average of the normalised values of these six indicators, providing a quantitative measure of the level of gender bias across different age groups and periods.

The gender bias index could only be calculated for the waves 2010–2014 and 2017–2022 because the previous surveys did not ask about the indicator related to the family dimension.

Table 8 presents the gender bias index in Spain, broken down into waves of surveys and age groups for the periods 2010–2014 and 2017–2022.

Table 8. Gender bias index by wave and age groups (Spain 2010–2022).

Age groups	WAVE (2010–2014)	WAVE (2017–2022)
18–29 years	0.14	0.09
30–49 years	0.14	0.11
50 years and over	0.22	0.18
Total	0.17	0.13

Source: EVS-WVS_TimeSeries_InternalUse_spss_v3_0. (Inglehart et al., 2022).

The main findings according to age group are presented here:

1. Age group 18–29 years: A significant decrease is observed in the gender bias index in this age group, indicating positive attitudinal shifts towards the acceptance of gender equality among young people.
2. Age group 30–49 years: Although the decrease isn't as pronounced as in the youngest group, we can also see a positive attitudinal shift towards gender equality in this age group.
3. Age group 50 years and over: This age group presents the highest gender bias in both periods, although there is also a downward trend. This indicates that, although attitudes are changing, gender bias is more persistent among older people.

On a general level, the gender bias index decreased by 0.17 in the 2010–2014 wave and by 0.13 in the 2017–2022 wave. This confirms the general trend towards the reduction of gender bias and stereotypes in the Spanish population.

3.4. Explanatory Factors for Gender Bias in Young People: The Explanatory Power of the Selected Variables

The variables that could explain gender bias among Spaniards and whether these variables are the same for young people are discussed below. For this purpose, we tested whether there is a significant correlation between the dependent variable (gender bias index) and the independent variables. For the regression analysis, it was decided to discard the low- and middle-income scale variable and, in the post-materialist/materialist variable, the mixed category, given that the level of significance in all of them, as can be seen in Table 9, is greater than 0.05. For the rest of the variables, the correlation between the dependent variable and the independent variables is significant (p value below 0.05) and, therefore, they are kept in the model.

By applying a first regression model with the selected variables, we eliminated age and high income because their significance is higher than 0.05 (0.658 and 0.336 respectively) but kept the rest of the variables. The model that we obtained is presented in Table 10.

The data in Table 10 show the results of the regression model applied to explain the gender bias index by socio-demographic, political-ideological, and value variables in Spain between 2010 and 2022.

Table 9. Correlation between the gender bias index and potentially explanatory variables.

		Gender bias index
Sex (binary)	Pearson Correlation	.070**
	Sig. (bilateral)	0.001
Age (binary)	Pearson Correlation	.183**
	Sig. (bilateral)	0.000
Low educational level (binary)	Pearson Correlation	.172**
	Sig. (bilateral)	0.000
Medium educational level (binary)	Pearson Correlation	-.074**
	Sig. (bilateral)	0.001
High educational level (binary)	Pearson Correlation	-.132**
	Sig. (bilateral)	0.000
Low income (binary)	Pearson Correlation	0.043
	Sig. (bilateral)	0.073
Medium income (binary)	Pearson Correlation	-0.005
	Sig. (bilateral)	0.820
High income (binary)	Pearson Correlation	-.051*
	Sig. (bilateral)	0.031
Religious person (binary)	Pearson Correlation	.204**
	Sig. (bilateral)	0.000
Self-positioning in political scale	Pearson Correlation	.201**
	Sig. (bilateral)	0.000
Materialist (binary)	Pearson Correlation	.165**
	Sig. (bilateral)	0.000
Mixed (binary)	Pearson Correlation	-0.025
	Sig. (bilateral)	0.245
Post-materialist (binary)	Pearson Correlation	-.151**
	Sig. (bilateral)	0.000

Notes: ** The correlation is significant in the level 0.01 (bilateral); * the correlation is significant in the level 0.05 (bilateral). Source: EVS-WVS_TimeSeries_InternalUse_spss_v3_0 (Inglehart et al., 2022).

Table 10. Explanatory regression model of the gender bias index using socio-demographic, political-ideological and value variables. Spain 2010–2022.

Adjusted R2 = 0.100	Unstandardised coefficients	Unstandardised coefficients	Standardised coefficients	T	Sig.
	B	Deviation Error	Beta		
(Constant)	.040	.010		4.073	<.001
Sex	.050	.010	.116	5.037	<.001
Low educational level	.055	.010	.127	5.504	<.001
Religious person	.063	.011	.146	5.993	<.001
Self-positioning in the political scale (right-wingers)	.069	.013	.131	5.500	<.001
Materialist	.054	.011	.112	4.855	<.001

Note: Dependent variable: Gender bias index. Source: EVS-WVS_TimeSeries_InternalUse_spss_v3_0 (Inglehart et al., 2022).

Adjusted $R^2 = 0.100$ indicates that the model explains 10% of the variability in the gender bias index. Although it's not a very high value, it suggests that the variables included have a certain explanatory capacity.

Unstandardised coefficients (B) indicate the change in the gender bias index for each unit of change in the independent variable while maintaining the rest constant:

- Sex: The coefficient of 0.050 means that, on average, the gender bias index is 0.050 units higher for one sex compared to the other, and this effect is significant ($p < 0.001$).
- Low educational level: The coefficient of 0.055 indicates that people with a low level of education have a higher gender bias index of 0.055 units, which is also significant ($p < 0.001$).
- Religious person: The coefficient of 0.063 suggests that religious people have a higher gender bias index of 0.063 units, which is significant ($p < 0.001$).
- Self-positioning in political scale (right-wingers): The coefficient of 0.069 indicates that right-wingers have a higher gender bias index of 0.069 units and is therefore significant ($p < 0.001$).
- Materialist: The coefficient of 0.053 shows that people with materialistic values have a higher gender bias index of 0.054 units, which is equally significant ($p < 0.001$).

Standardised coefficients (Beta) allow the relative importance of each independent variable in the model to be compared.

- Sex (Beta = 0.116): Has a moderate effect on the gender bias index.
- Low educational level (Beta = 0.127): Has a slightly greater effect than sex.
- Religious person (Beta = 0.146): Has a greater effect between the variables included.
- Self-positioning in political scale of right-wingers (Beta = 0.131): Also has a considerable effect.
- Materialist (Beta = 0.112): Has a lesser effect between significant variables but is still relevant.

All the coefficients, including the constant, are highly significant (Sig.; $p < 0.0001$), indicating that there is strong evidence that these variables are associated with the gender bias index.

In summary, the regression model shows that being a man, having a low educational level, being religious, identifying politically as right-wing, and having materialist values are factors significantly associated with the gender bias index. Although the model does not explain a large part of the variability (adjusted $R^2 = 0.100$), these factors have a significant influence on gender bias. In order to apply the model by age groups we have considered two categories (18–29 years and 30 years and over) to simplify the analysis.

The results of the regression analysis by age group are presented in Table 11.

As shown in Table 11, all independent variables included in the model have an impact on gender bias in the 30-year-old or over age group. However, only religion is statistically significant for the youth group. It should be noted that the reduction of the samples, when disaggregated by age group, affects the explanatory power of the model in the case of the group of young people, whose sample N is reduced to 406 cases.

Religion is the only variable that can potentially explain gender bias among young people. In this sense, a specific analysis of the evolution of the religion variable, based on WVS and EVS data in its latest waves, shows

Table 11. Explanatory regression model of the gender bias index by age group using socio-demographic and political-ideological variables (Spain 2010–2022).

18–29 years					
Adjusted R2 = 0.041	Unstandardised coefficients	Unstandardised coefficients	Standardised coefficients	T	Sig.
	B	Deviation Error	Beta		
(Constant)	.059	.021		2.813	.005
Sex	.032	.023	.082	1.401	.162
Low educational level	.030	.024	.072	1.250	.212
Religious person	.071	.025	.167	2.842	.005
Self-positioning in the political scale (right-wingers)	.040	.033	.071	1.212	.227
Materialist	.037	.027	.079	1.367	.173
30 years or more					
Adjusted R2 = 0.106	Unstandardised coefficients	Unstandardised coefficients	Standardised coefficients	T	Sig.
	B	Deviation Error	Beta		
(Constant)	.037	.011		3.277	.001
Sex	.054	.011	.123	4.894	<.001
Low educational level	.059	.011	.135	5.332	<.001
Religious person	.060	.012	.138	5.148	<.001
Self-positioning in political scale (right-wingers)	.073	.014	.140	5.365	<.001
Materialist	.057	.012	.116	4.578	<.001

Note: Dependent variable: Gender bias index. Source: EVS-WVS_TimeSeries_InternalUse_spss_v3_0 (Inglehart et al., 2022).

that the weight of this variable has increased in a higher proportion among young people: Of all young people (18–29 years old) surveyed in the 2005–2009 wave, a total of 27.7% indicated that they were religious. In the 2017–2022 wave, the percentage increases to 41.5%. In the other age groups, an increase is also observed, but to a lesser extent: The 30–49 age group went from 39.2% to 42.4%, and the 50 years and over age group from 60.7% to 59.3%. These data would confirm the growing trend towards the adoption of neo-conservative positions in the new generations, with likely effects on the continuation of traditional gender stereotypes.

4. Conclusion

Overall, the results of the analysis show a positive trend towards the reduction of gender bias in Spain in most of the key dimensions considered. The exception would be the political dimension (perception of equal rights as a key element of democracy), which not only doesn't decrease but also shows a slight trend towards an increase in the waves analysed.

In terms of age groups, the result is similar: the analysis indicates a decrease in the reduction of gender bias in Spain, with some differences according to dimensions and age groups, which would be along the same lines as the aggregate data: The appreciation that equal rights for men and women are a key element of democracy

escapes the downward trend in gender bias. The analysis shows a slight increase in the percentage of people with a bias towards the variable cited in the 18–29 age group (higher in this group) and the 30–49 age group, but a steady decrease in the 50 years and over age group.

This suggests that younger generations may be exposed to ideological influences that cause shifts in their perceptions of politics and democracy. The fact that, in the case of young people, more than a fifth of the sample surveyed do not consider (formal) gender equality to be a relevant constituent element of democratic regimes, and that the trend is also increasing, seems to indicate not only an unfavourable stance on inequality, although this obviously cannot be inferred from the data, but also a relative loss of the political and moral significance of gender equality as an issue for a significant proportion of young Spaniards.

For its part, the gender bias index decreased from 0.17 in the 2010–2014 wave to 0.13 in the 2017–2022 wave, confirming the general trend towards the reduction of gender bias and stereotyping in the Spanish population. This reduction is greater in the 18–29 years age group, which shows a decrease in the gender bias index (from 0.14 to 0.09), indicative of positive attitudinal changes towards the acceptance of gender equality among young people. On the other hand, the age group 50 years and over has the highest rate of gender prejudice in the periods considered, although it also shows a downward trend. This indicates that, although attitudes are changing, gender bias is more persistent among older people.

Regarding the factors likely to explain the existence of gender bias, the regression model applied shows that being a man, having a low educational level, being religious, identifying politically as right-wing, and having a preference towards materialist values are factors significantly associated with the gender bias index. The most significant values are, firstly, religion (considering oneself to be a religious person) and, secondly, political-ideological positioning (on a left-right scale). On the other hand, positioning on the materialist/post-materialist value scale is the variable with the least significance (although it also has some significance).

The reduction of the samples, when disaggregated by age group, affects the explanatory power of the model, especially in the case of the group of young people, whose sample *N* is reduced to 406 cases. In any case, socio-demographic, political-ideological, and value variables are significant in gender bias in the 30-year-old and over age group. Only religion is statistically significant for the 18–29 years age group.

In any case, with the caveats derived from the aforementioned reduction in age samples, the data seem to suggest that the factors influencing gender bias may vary significantly between the general population and young people and that other factors not considered in the model may be more relevant for the latter. This invites further analysis that integrates other variables into the model.

Even so, the fact that religion is the factor with the greatest statistical significance for the sample as a whole, remaining the only explanatory variable for young people despite the reduction in its size in the breakdown by age group, calls for reflection.

Religion is certainly linked to a greater likelihood of maintaining the existence of gender bias, and even accentuating it among young people and, in general, in the population as a whole. This could be indicative of a correlation between religion and the adoption of ideologically conservative positions, in line with the socio-political dynamics of polarisation and the growing influence of neo-conservative movements in

Spanish society (as in Europe and the United States in general). This would be consistent with other studies, such as the aforementioned study by Wernet (2016), which indicates the existence of a positive correlation between religiosity and gender bias. Likewise, specifically in the case of Spain, a tendency towards dualization has been observed in the axiological positions of Spaniards, who tend to place themselves either in traditional and conservative positions (linked to the presence of religious attitudes), with a slightly greater weight in the overall group, or in more modern and secularised positions (whose weighting tends to be less; Rivero Recuenco & Antolínez Merchán, 2023).

In any case, it is important to reiterate that religion is the only variable that can potentially explain gender bias among young people. In this sense, a specific analysis of the evolution of the religion variable, based on WVS and EVS data in its latest waves, shows that the weight of this variable has increased in a higher proportion among young people, confirming the tendency towards the adoption of neo-conservative positions in new generations, affecting the continuation of the traditional gender stereotype (the percentage of young people positioning themselves on the right of the ideological scale has increased from 12.3% in 2005–2009 to 19.8% in the latest wave).

These assessments, which are still hypothetical, pave the way for further research.

5. Discussion and Prospects

The main contribution of this work is that it highlights the relationship between religiosity and gender prejudice. It calls for a deeper analysis of this relationship, which, in any case, requires further and specific research, given the limitations of the length of this article. Some progress is certainly made in this respect, by incorporating data on the increase in religiosity among young people in recent years, which are incorporated to complement the analysis of gender bias and its possible explanatory variables (religiosity in this case). However, a detailed analysis of the possible relationships between religiosity and gender bias, which may undoubtedly be related to other socio-demographic and attitudinal variables, is, as indicated above, a matter for further studies.

This article has not attempted to enter into policy recommendations. The study focuses on the socio-statistical determination of the issue. It would imply going beyond the framework of the research objectives to enter into other aspects that can be approached from positions of a more evaluative and political nature, although this reflection can also form part of subsequent research. In any case, it is understood, on a preliminary basis, that the social and ideological forces that seem to be driving the neo-conservative movements affecting Western societies are not easy to counteract. It is a complex problem that requires specific evaluative and political research.

It is certainly worth reflecting on the effectiveness of education and awareness-raising policies in a social context affected by global transformations that affect the value and ideological positions of citizens. These are transformations whose socially perceived risks do not seem to be adequately addressed by traditional political parties. One might even ask, as a hypothesis or conjecture, whether gender equality awareness campaigns (apart from the limitations of conventional advertising in today's information society, coupled with the phenomena of silos and information/communication bubbles, etc.) might be generating effects contrary to those intended, especially in the younger generations. However, given that it is a hypothesis, it can only be tested by further research.

Funding

This study is part of the TSH-Discourse project of the 10th Call for Research Grants of the Camilo José Cela University. The article is part of the activities of the Center for Research on Social Values at UCJC.

Conflict of Interests

The authors declare no conflict of interests.

References

- Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. *Journal of Public Economics*, 95 (7/8), 476–487. <https://doi.org/10.1016/j.jpubeco.2010.11.006>
- Amurrio, M., Larrinaga, A., Usategui, E., & Del Valle, A. I. (2012). Los estereotipos de género en los/las jóvenes y adolescentes. In Eusko Ikaskuntza (Eds.), *XVII Congreso de Estudios Vascos: Gizarte aurrerapen iraunkorrerako berrikuntza = Innovación para el progreso social sostenible* (pp. 227–248). <https://www.eusko-ikaskuntza.eus/PDFAnlt/congresos/17/02270248.pdf>
- Bissell, K., & Parrott, S. (2013). Prejudice: The role of the media in the development of social bias. *Journalism & Communication Monographs*, 15(4), 219–270. <https://doi.org/10.1177/1522637913504401>
- Brandt, M. J. (2011). Sexism and gender inequality across 57 societies. *Psychological Science*, 22(11), 1413–1418. <http://www.jstor.org/stable/41320046>
- Bustelo, M. (2014). Three decades of state feminism and gender equality policies in multi-governed Spain. *Sex Roles*, 74, 107–120. <https://doi.org/10.1007/s11199-014-0381-9>
- Castillo, M. R., & Montes, B. (2007). Escala de estereotipos de género actuales. *Iniciación a La Investigación*, 2, 1–21. <https://revistaselectronicas.ujaen.es/index.php/ininv/article/view/198>
- González-Gijón, G., Alemany-Arrebola, I., Ruiz-Garzón, F., & Ortiz-Gómez, M. (2024). Los estereotipos de género en adolescentes: Análisis en un contexto multicultural. *Revista Colombiana de Educación*, 90, 164–184. <https://doi.org/10.17227/rce.num90-1464>
- Hussain, M. A., & Haj-Salem, N. (2023). Did the Arab Spring change female emancipation perceptions? *Gender in Management*, 38(4), 433–453. <https://doi.org/10.1108/GM-10-2021-0299>
- Inglehart, R., Basáñez, M., Díez-Medrano, J., Halman, L. C. J. M., & Luijkx, R. (Ed.). (2004). *Human beliefs and values: A cross-cultural sourcebook based on the 1999–2002 value surveys*. Siglo XXI.
- Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Díez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., & Puranen, B. (Ed.). (2022). *World Values Survey: All rounds—Country-pooled datafile* (Dataset version 3.0.0) [Data set]. JD Systems Institute & WVSA Secretariat. <https://doi.org/10.14281/18241.17>
- Inglehart, R., & Norris, P. (2003). *Rising tide: Gender equality and cultural change around the world*. Cambridge University Press.
- Lombardo, E., & Alonso, A. (2020). Gender regime change in decentralized states: The case of Spain. *Social Politics: International Studies in Gender, State & Society*, 27 (3), 449–466. <https://doi.org/10.1093/sp/jxaa016>
- Mukhopadhyay, T., Rivera, C., & Tapia, H. (2019). *Gender inequality and multidimensional social norms*. United Nations Development Programme, Human Development Report Office. https://www.academia.edu/66607003/Gender_Inequality_and_Multidimensional_Social_Norms
- Norris, P., & Inglehart, R. (2002). Islamic culture and democracy: Testing the 'clash of civilizations' thesis. *Comparative Sociology*, 1(3), 235–263. <https://doi.org/10.1163/156913302100418592>
- Office of the United Nations High Commissioner for Human Rights. (n.d.). *Gender stereotyping: OHCHR and women's human rights and gender equality*. <https://www.ohchr.org/en/women/gender-stereotyping>

- Perry, D. G., & Pauletti, R. E. (2011). Gender and adolescent development. *Journal of Research on Adolescence*, 21(1), 61–74. <https://doi.org/10.1111/j.1532-7795.2010.00715.x>
- Rebrey, S. M. (2023). Gender inequality in Russia: Axial institutions and agency. *Russian Journal of Economics*, 9(1), 71–92. <https://doi.org/10.32609/j.ruje.9.94459>
- Rivero Recuenco, A., & Antolínez Merchán, P. (2023). Valores, creencias e ideología: Factores culturales explicativos de las preferencias sobre cualidades a transmitir en educación. In P. Antolínez, J. Díez, J. García, J.M. León, F. J. Llera, M. Martínez, M. Reynés, & A. Rivero (Eds.), *La globalización de los valores en el mundo y en España: Cualidades que deberían enseñarse a los niños* (pp. 147–193). Centro de Investigaciones Sociológicas.
- Rizzo, H., Abdel-Latif, A. H., & Meyer, K. (2007). The relationship between gender equality and democracy: A comparison of Arab versus non-Arab Muslim societies. *Sociology*, 41(6), 1151–1170.
- Santonico, F., Trombetta, T., Paradiso, M. N., & Rollè, L. (2023). Gender and media representations: A review of the literature on gender stereotypes, objectification and sexualization. *International Journal of Environmental Research and Public Health*, 20(10), 1–15. <https://doi.org/10.3390/ijerph20105770>
- Tesch-Römer, C., Motel-Klingebiel, A., & Tomasik, M. J. (2008). Gender differences in subjective well-being: Comparing societies with respect to gender equality. *Social Indicators Research*, 85(2), 329–349. <https://doi.org/10.1007/s11205-007-9133-3>
- UNESCO. (2022). *#HerEducationOurFuture #BreakTheBias: Challenging gender bias and stereotypes in and through education; the latest facts on gender equality in education*. <https://unesdoc.unesco.org/ark:/48223/pf0000380827?2=null&queryId=0b05c5d1-1855-4c01-9dbd-824d7b2b83da>
- United Nations Development Programme. (2023). *Gender Social Norms Index (GSNI): Breaking down gender biases: Shifting social norms towards gender equality*. <https://hdr.undp.org/content/2023-gender-social-norms-index-gsni#/indicies/GSNI>
- Wernet, C. A. (2016). The impact of secularity on attitudes of gender equality: A multi-level analysis of 49 nations. *Comparative Sociology*, 15(4), 508–530. <https://doi.org/10.1163/15691330-12341397>

About the Authors



Pilar Antolínez-Merchán (PhD in sociology) is a professor in the Faculty of Communication and Design at Camilo José Cela University. She is a member of the UCJC Social Sciences Research Group (INCISO-UCJC). She is the representative of the sociology area of the Centre for Research in Social Values (UCJC-CIVAS) and the principal investigator (PI) of the project Politics and Society: Discourses on Human Trafficking (TSH-Discourse) of the 10th Call for Research Grants of the Camilo José Cela University.



Ángel Rivero Recuenco (PhD in political sciences and sociology) has more than twenty years of experience in social consultancy in the public sector. A specialist in research, planning, and evaluation, he has worked with various companies and administrations in the areas of social inclusion and programme evaluation. He has taught at the European University and Carlos III University in Madrid. He is currently an associate professor at EAE Business School and at the University of Alcalá.



Elvira Carmen Cabrera-Rodríguez (PhD in law) is a criminologist and specialist in victimology, and professor in the Faculty of Entrepreneurship, Business and Legal Sciences at Camilo José Cela University. She is a member of the UCJC Social Sciences Research Group (INCISO-UCJC) and of the research team of the project Politics and Society: Discourses on Human Trafficking (TSH-Discourse) of the 10th Call for Research Grants of the Camilo José Cela University.

Unveiling Hate Speech Dynamics: An Examination of Discourse Targeting the Spanish Meteorological Agency (AEMET)

Sergio Arce-García ¹ , Virginia Martín-Jiménez ² , and Leticia Rodríguez-Fernández ³ 

¹ Universidad Internacional de La Rioja, Spain

² Universidad de Valladolid, Spain

³ Universidad de Cádiz, Spain

Correspondence: Sergio Arce-García (sergio.arce@unir.net)

Submitted: 13 September 2024 **Accepted:** 2 December 2024 **Published:** 27 January 2025

Issue: This article is part of the issue “Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment” edited by Max Römer Pieretti (Universidad Camilo José Cela), Beatriz Esteban-Ramiro (Universidad de Castilla-La Mancha), and Agrivalca Canelón (Universidad Católica Andrés Bello), fully open access at <https://doi.org/10.17645/si.i415>

Abstract

This article examines hate speech directed at AEMET, the Spanish meteorological state agency, on the social media platform X. We analysed nearly half a million messages posted between 31 December 2021 and 19 April 2023, using hate speech detection algorithms, text mining techniques, and qualitative analysis to identify patterns and themes in the discourse. Our research reveals a troubling reality, with around 25% of the messages collected displaying some degree of hostility towards AEMET, its staff, and its scientific work. A considerable amount of hate speech was expressed through derogatory comments and insults aimed at meteorologists, which is indicative of a wider trend of anti-intellectualism and scepticism of scientific expertise. Furthermore, the spread of conspiracy theories, particularly those related to geoengineering and chemtrails, highlights the spread of misinformation within online communities. This study emphasises the importance of acknowledging and addressing the spread of hate speech in meteorology and scientific communication. By emphasising the negative effects of such language on public perception and trust in scientific institutions, this article advocates for collaborative efforts to promote a culture of informed dialogue and evidence-based discourse. The results highlight the importance of combating hate speech and misinformation to protect the integrity and credibility of scientific institutions such as AEMET.

Keywords

AEMET; anti-intellectualism; disinformation; hate speech; meteorology; X/Twitter

1. Introduction

The last decade has been marked by a severe crisis of confidence that has affected the credibility and reputation of most organisations. The effects of the 2008 economic crisis have recently been compounded by the Covid-19 pandemic, creating a scenario of erosion and disaffection in which citizens demand more attention and solutions from institutions. In this context, disinformation, whose main objective is to create divisions and polarise society, emerges as another threat that exacerbates the weaknesses of democracies. According to the World Economic Forum (2024), disinformation will become the most relevant global risk in the next two years, as digital disinformation campaigns aim to undermine public trust in institutions and sow doubts about democratic processes (García-Orosa, 2021).

Recently, UNESCO (2023) presented an action plan for combating disinformation. The work was based on a survey ($N = 8,000$) conducted in 16 countries that were to hold elections in 2024. The survey found that 85% of citizens were concerned about the impact of disinformation and 87% believed that it had already had a major impact on political life in their country. However, disinformation is also a major problem in other areas, such as security and public health. This fact was demonstrated globally during the Covid-19 pandemic, making the flow of information from institutional sources and international organisations even more necessary (Ferreira Caceres et al., 2022). Thus, the WHO has taken various actions and entered into various collaborations, both through its own media and with third parties such as Wikipedia (Fidalgo, 2020) or, more recently, TikTok (WHO, 2024), to provide reliable health information and limit misinformation.

In the face of this new framework, the European Union has developed several measures to improve the resilience of the population and mitigate the impact of disinformation, including legislative measures such as the Digital Services Directive, an early warning system, and a new compendium on cybercrime (European Commission, 2024).

The latest Winter Eurobarometer Survey (2022) found that 69% of European citizens surveyed often came across information and news that they believed misrepresented reality or was even false, a percentage that rises to 78% in the case of Spain (European Union, 2024). Similarly, 76% of European respondents believed that the existence of news or information that distorts reality or is even false was a problem in their country, a view shared by 83% of Spanish respondents. This situation is likely to worsen as artificial intelligence becomes more widespread. On average, 74% of respondents in 29 countries believed that AI will make it even easier to create fake news or images in a very realistic way (IPSOS, 2023).

The public seems to feel disconnected from institutions, which is reflected in many works dealing with trust. This crisis also seems to be global. The 2024 Edelman Trust Barometer found that 63% of citizens surveyed in 28 countries ($N = 32,000$) did not trust political leaders, 61% did not trust business leaders and 64% felt the same way about journalists and the media (Edelman Trust, 2024). In all cases, distrust had risen by between one and three points since the same survey had been conducted the previous year. However, scientists enjoy high prestige and a good reputation when it comes to trust in authority figures: 74% of the respondents believed scientists tell them the truth about innovations and technologies. This puts scientists ahead of technical business experts, CEOs, and journalists as the most representative figures in terms of credibility.

This sentiment is also reflected in the specific data for Spain, which has one of the highest levels of mistrust of the 28 countries surveyed by the Edelman Trust: 81% of Spanish citizens in the survey did not believe political leaders, 79% did not believe journalists, and 73% did not believe business leaders. There was also a fear of misinformation among 61% of the respondents. There is a perception that these leaders manipulate the truth by making claims that they know are false or exaggerated. According to the same report, 51% of the population think that science is too politicised and 61% feel that government and research organisations have too much influence on the scientific community.

From the point of view of organisational communication, disinformation particularly affects institutions, especially the most visible ones (the Spanish government or the royal family), followed by local and regional institutions (Rodríguez-Fernández, 2019). This trend can also be seen in the aftermath of the Covid-19 pandemic, with almost 23% of an analysed set of fact-checking verifications in Spain over one month being related to these organisations (Almansa-Martínez et al., 2022).

2. Disinformation and Hate Speech: Science and Climate Change and Its Impact on Society

Climate denial is “the use of rhetorical arguments to create the appearance of a legitimate debate where none exists, with the ultimate aim of rejecting a proposal or claim on which there is widespread consensus” (Diethelm & McKee, 2009, p. 2). Through this type of denial, narratives are disseminated that influence popular beliefs about issues of real importance to society. These narratives include the denial of climate change or, for example, the existence of so-called “chemtrails,” condensation trails from aeroplanes that are said to be a weather modification system (Asmelash, 2024).

Social networks, including YouTube, are currently the main sources for the dissemination of denialist discourse (Vicente Torrico & González Puente, 2023). In addition to the highly viral nature of the content and the difficulty of content curation, there is also the generation of new discourses that are less obvious but are in line with the denial of climate change. In an analysis of 12,058 videos from 96 climate denial YouTube channels, 70% were found to contain “new” narratives about climate impacts, solutions, and current climate science (Center for Countering Digital Hate, 2024). Here the arguments are less overt and more subtle, and include those made by climate delayers calling for a delay in policy action to mitigate climate change. This type of discourse is more dangerous because it appears more reasonable and can therefore generate more persuasion and influence than more aggressive and blunt discourse (Dionis, 2022). As Jiménez-Gómez and Martín-Sosa (2022, pp. 533–534) point out, two basic frameworks can be observed: the ideological one, in which the idea of freedom of expression functions as the main vector, and the economic one, with “a sceptical attitude towards the measures to be adopted in the face of climate change.”

Influencers also play an important role as potential disseminators of disinformation. Nieto-Sandoval and Ferré-Pavia (2023) identify influencers as the actors who publish most videos on climate change on TikTok, more than media or institutions, and also point out that the content posted by influencers does not identify the sources of their information. Among the conclusions of this study is the observation that the sample of users analysed accepted climate change as a fact, but did not propose ways to change this, being merely passive subjects.

On the other hand, it should be noted that organisations can also act as disseminators of disinformation to benefit their activities or to clean up their image (Olivares-Delgado et al., 2023). Exxon, for example, promoted disinformation campaigns on climate change by funding 43 organisations between 1998 and 2005 to challenge the scientific consensus on the phenomenon. The economic contribution amounted to almost \$16 million (Union of Concerned Scientists, 2007). At the same time, the company was processing relevant scientific information on the harmful effects of its activities, while publicly issuing communiqués that contradicted this internal data. Similar information laundering has been carried out by the tobacco industry for decades (Supran et al., 2023).

This “corporate disinformation” is often issued by companies whose products and services pose a threat to society, and entails them generating alternative or denialist narratives to manipulate and reduce negative public perceptions. Some researchers call these practices “dark PR” (Ennis, 2023), although the term is also used, along with its synonym “black PR,” to describe “negative influence campaigns aimed at damaging the reputation of a competitor or political opponent” (Rodríguez-Fernández, 2023, p. 122). To this, we can add practices such as greenwashing, which violates the expectations of “consumers” by deliberately misleading them about the company’s environmental practices or the benefits of its products/services (Santos et al., 2024).

Work carried out in Spain shows that misinformation on climate issues has an important belief niche: 85% percent of respondents ($N = 8,000$) believed it to be true that human activity causes extreme weather events, and half of the respondents believed that patent holders restrict the supply of cancer drugs to increase their profits or that genetically modified foods are unsafe (Wiesehomeier & Flynn, 2020).

In the fields of health, science, and climate change, these disinformation campaigns on social networks go hand in hand with the generation of hate speech aimed at attacking the social actors involved (Rodrigo Cano & Del Río Álvarez, 2021) to discredit their discourse, undermine the credibility of their research and influence the adoption of measures that entail economic or social changes. In the case of climate change, for example, and as reflected in the volume of reviews, institutions appear to be the most affected actors, followed by activists, the media, and scientists (Vicente Torrico et al., 2024).

An example of this reality, and of how hate speech is one more element in disinformation strategies, can be found in the digital conversation directed at AEMET, the Spanish meteorological state agency. AEMET is a Spanish institution that began its activities in the nineteenth century under a different name (Palomares Calderón, 2015). It is dedicated to “the development, implementation and provision of meteorological services of national competence and support for the exercise of other public policies and private activities, contributing to the safety of people and goods and the welfare and sustainable development of Spanish society” (Agencia Estatal de Meteorología, 2024). Recently, AEMET has suffered attacks on its work and has received insults and threats that have been publicly denounced by the institution itself (El Rastreador, 2023). Conspiracy theorists accuse the agency of “climate manipulation,” “provoking drought as one of the objectives of the satanic 2030 agenda,” “fumigation, sterilisation,” or “population control.” These attacks have increased sharply since the Covid-19 pandemic, to the point that in 2023 a message had to be published on social networks with the following text:

AEMET in #SocialNetworks. We understand it as a space for meeting and closeness with society, amateurs and professionals. We share our work and our knowledge. We respect freedom of expression and welcome interaction. But NOT everything goes. (“La AEMET responde,” 2023)

It should be noted that at the end of October 2024, during the final revision of this article, Spain faced one of the greatest weather disasters in its history, as a result of isolated depression at high levels (DANA), with the peninsula being devastated. The Autonomous Community of Valencia suffered the worst consequences, with more than 200 deaths. In the context of this crisis, disinformation became an additional problem, with AEMET being one of the most affected institutions and the target of numerous pieces of false content. Among the countless hoaxes that circulated, it was said that AEMET manipulated the weather, that its parameters were modified to simulate climate change, and that historical records showed that it manipulated its categorisations (“Hoaxes, disinformation,” 2024).

3. Objectives

Taking into account this socio-communicative context, this research has the main general objective to explore the presence of hate speech in communications on the social network X against AEMET.

Our specific objectives are:

1. To identify the presence and characteristics of hate speech against AEMET on X, by analysing the topics, the intensity and the target of the hostile messages.
2. To study the diffusion of conspiracy and denial theories and their links with hate towards AEMET, assessing how concepts such as geoengineering and chemtrails fuel hostility towards this institution on social networks.
3. To evaluate the structure and nature of the groups responsible for this hate speech, using network analysis to understand the links between the senders of this type of message.
4. To reflect on the need to propose effective strategies to reduce hate speech and misinformation in meteorology and science communication, encouraging evidence-based dialogue and respect for science.

4. Methodology

The study was carried out on hate speech in messages on the social network X/Twitter that mention one of the accounts of AEMET, namely its main account or the delegated accounts it has in each of the Spanish Autonomous Communities. The collection was carried out on 21 April 2023, using the history of messages posted between 31 December 2021 and 19 April 2023 (one year and four months) to obtain a continuum of discourse that is independent of isolated events or campaigns. The choice of X was made because it is a social network with high penetration in Spain (43% of the Spanish population searches for information or content of interest on social networks, and 48% of these individuals choose X; see We are Social, 2023). As well as being one of the most widely used social networks in the country, X has been identified as one of the biggest distributors of misinformation, hoaxes, and hate speech (Arce-García & Díaz-Campo, 2024; DeVerna et al., 2024). It has also been shown that hate speech on X leads to an increase in offline hate crimes in Spain (Arcila Calderon et al., 2024). The network has been singled out for its failure to remove disinformation, although, like other networks, it has a very low level of removal even after complaints (Bergsmanis-Korats & Haiduchyk, 2024). Its ability to provide a wide range of data for academics made it stand out from other platforms until mid-2023, the time of the arrival of its new owner Elon Musk. Its primary use in disseminating information made it a unique source.

As some of the messages were recovered at a later date, they could have been deleted by their authors at that time or even removed by the social network X itself for violating its rules. However, the European Union (“Twitter and other social media,” 2022) and the Center for Countering Digital Hate (2023) have reported that more than 80% of hate messages are not deleted.

The messages were collected using the `academicwtwiteR` library on the X/Twitter API 2.0, under the academic account (Barrie & Chun-ting, 2021) using R software. The following analyses were then carried out to understand the structure of the groups participating in the conversation, the identification of hate, its main topics, and the structure and nature of the hate speech emitters.

Network analysis was carried out using graph theory (Barabasi, 2016) to determine the structure of the relationships between users based on the retweets issued. The data were transformed using R software and then processed using Gephi software version 10.1. They were represented graphically using the ForceAtlas2 algorithm (Jacomy et al., 2014), and grouped by cluster or modularity (Chen et al., 2020) using the Louvain algorithm (Blondel et al., 2008). The nature and behaviour of each group was also analysed by identifying aspects such as average edge length, average degree, network diameter, eigenvector, mediation, and the number of edges.

To identify the presence of hate speech in the set of selected messages, the `Syuzhet` library in R was used. This has pre-established lexicons for sentiment levels and polarity, among other things (Jockers, 2017). The algorithm detects hate speech when it finds words previously classified as such and determines their intensity when these words are combined with related words that can modulate their value. For this purpose, we used the Spanish hate lexicon developed by Said-Hung et al. (2023), which contains 7,210 words or lemmas, both simple and compound, identified as hate speech in informational contexts. The use of this Spanish-specific lexicon avoids problems associated with the detection of hate speech in other languages and communicative contexts, as it does not require the translation of lemmas for automatic classification by the algorithm, which increases the reliability of the results. This lexicon-based detection method is one of the established approaches in machine learning systems for hate speech detection, achieving a reliability of around 80%, as reported by Alkomah and Ma (2022).

Using text mining, and after removing words or “stopwords” that do not add sense or meaning to a message (such as articles, adverbs, etc.), messages identified as hateful were plotted in two dimensions using a multidimensional scaling (MDS) graph. In addition, a graph was created showing the frequency of words that occur in the messages identified as hateful. The words are represented by colours that indicate their grouping into clusters, which are formed according to the frequency with which these words occur together. This grouping was done using the *k*-means algorithm with a value of $k = 7$, to identify a sufficient number of themes. With this technique, we tried to identify the main themes of hate speech within each of the clusters identified in the first point.

5. Results

A total of 480,559 messages were obtained in the data collection, of which 148,345 (30.9%) were direct messages, 5,715 (1.2%) were quotes, and 326,499 (67.9%) were retweets. This means that the majority of messages were in support of others through a simple retweet.

The temporal distribution of the broadcasts over the period of analysis of a year and four months is more or less uniform, although certain periods show increases in the number of messages. The network analysis formed from the retweets, shown in Figure 1, shows a very compact network around the different official AEMET accounts (general, Valencia, Canary Islands, Madrid, etc.) as well as the emergency broadcasting accounts of 112 (the emergency telephone number in Spain). This network has a modularity of 0.661, an average path length of 4.849, and a network diameter of 13. It has 61,415 nodes (accounts) and 114,109 edges (forwards).

This is a fairly cohesive network, except for one cluster, shown in blue in Figure 1, which the algorithm distances from the other clusters and which is formed from 8.84% of the total network. This cluster has 5,431 nodes (8.84%) and 11,615 edges (10.18%), so it has a greater number of connections between the nodes than the rest of the clusters detected. Although it is a cluster far from the centre of the network, it has a wide interconnection with the main accounts in its eigenvector values: @mOrtiz_RT, @purereason1972, @A_Moon73, and @JMurilloSanchez. Among the main accounts in the above cluster, only one account of a public figure was identified: @ldpsincomplejos (eigenvector 0.01), the account of radio journalist Luis del Pino, who moved from esRadio to the Intereconomía group and elToroTV in 2024 (“Luis del Pino cambia,” 2024). All of the above-mentioned accounts published between 200 and almost 800 messages citing the AEMET accounts during the period analysed. Although the accounts do not represent clear influencers, they do represent accounts with high levels of mediation between accounts: @GeoingenieriaMu (centrality 499,754.29), @Empecinado101 (406,134.14), @877rfm (386,721.02), and @JMurilloSanchez (288,162.43). All these accounts are linked to conspiracy theories and accusations against AEMET. These conspiracy theories are not found in the rest of the identified clusters.

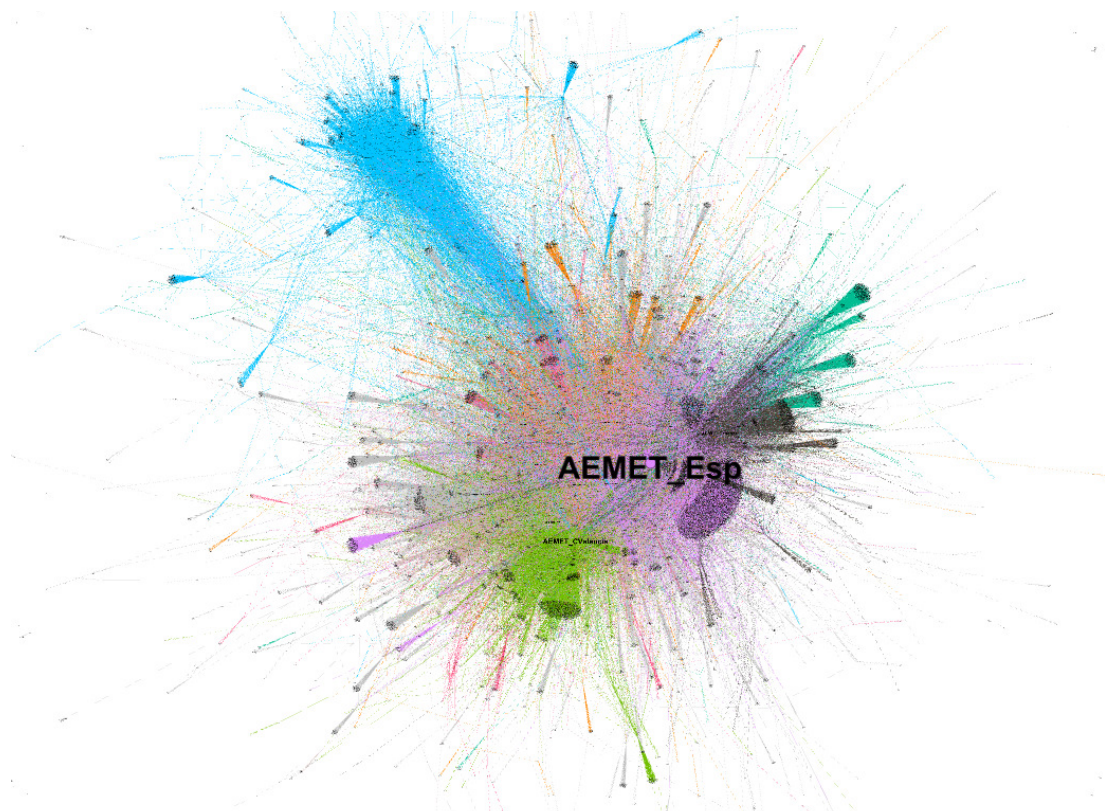


Figure 1. Retweet network around AEMET accounts.

5.1. Hate Speech Analysis

The hate detection algorithm identified a total of 119,084 hate messages (24.78% of the total number of messages) posted by 36,599 accounts. Of the messages, 26,122 (21.9%) were direct messages, 845 (0.7%) were quoted, and 92,117 (77.4%) were retweets. Analysis of the retweets shows that 10.85% of the hate messages came from the group previously identified as conspiracy theorists, but there are a large number of messages with hate content that were directly posted. Thus, 31.8% of the hate messages were sent against the general account of @AEMET_Esp, 9.46% against @AEMET_CValenciana, 6.66% against @AEMET_Andalucia, 6.03% against @AEMET_Canarias, 4.42% against @AEMET_Madrid, 3.33% against @AEMET_Murcia, 2.66% @BBFFLuchaCyl and 2.52% against journalists from El País, RTVE, Cope or Maldita (@vbenayas, @Eltiempo_tve, @Maldito_clima, @Divulgameteo).

The most commonly used words in the hate speech, as shown in Figure 2, falling among the messages in the third quartile with the highest average hate intensity (> 0.25), reflect clear insults such as “sons of bitches,” “criminals,” or “shit.” There are also expressions such as “fuck” and “scare,” as well as the names of the official accounts of AEMET and some of its offices in Valencia, the Canary Islands, the Balearic Islands, Aragon, Madrid, and Andalusia. In other words, the hatred is mainly generated around personalised insults, with accusations of trying to ‘scare’ the population with their announcements and weather warnings.

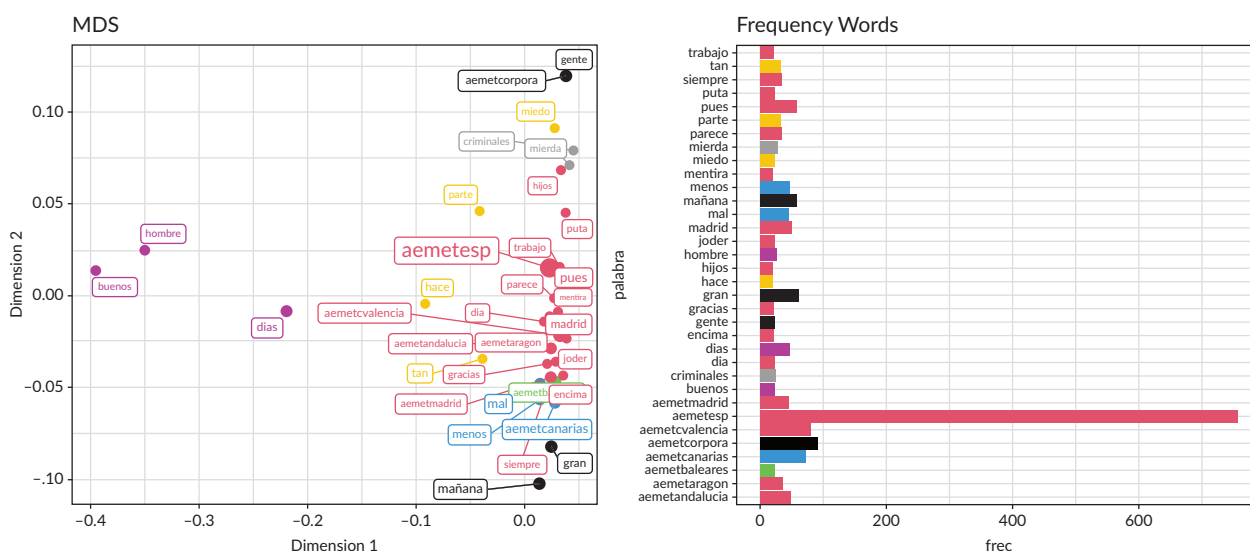


Figure 2. MDS and frequency analysis of the words most frequently used in hate speech against AEMET.

A selection of the most hate-filled messages is shown in Table 1. It can be seen that the hatred mainly revolves around climate change denial, conspiratorial aspects (climate manipulation, Illuminati, elites, viruses, and NATO), or the 2030 Agenda (the UN’s sustainable development goals).

Hate messages have a median of seven retweets and an average of 79.87, while they receive hardly any likes, with a median of 0 and an average of 0.68. The accounts spreading hate also show certain characteristics:

1. They have a high level of activity on the network: The median number of messages they have written is 20,217 (6.92 messages per day) and the average is 51,253 (17.55 per day) in their history.

2. They are accounts of a nano-influencer nature, as they often appear in astroturfing-type campaigns (Arce-García et al., 2022; Ong et al., 2019), with a median of 631 followers and 588 accounts followed.
3. The level of hate is high enough to generate emotions in the listener, but it is not at such a high level that it could be detected by algorithms or even constitute a crime, as other research has shown (Arce-García et al., 2024). Thus, they have an average level of hate per message, with a median of 0.2105 and a mean of 0.2234 (on an average scale from 0 to 1).

Table 1. Examples of climate-themed hate speech.

	Example
1	@MADRID @AEMET_Esp This is the freedom that the PP uses as a flag, they are brainless this City Council all dictatorship and populism and liar, who would think of closing the parks with shadows to avoid the heat wave, PP you give stumbles of organisation you are only commissioners card sharks
2	@AndaluciaJunta @AEMET_Esp @AEMET_Andalucia We have everything against us, damn it: the weather, which doesn't rain, snow or get cold; politics, a pot of crickets, each one more shrill, inept and disloyal; the virus and, as a consequence, the crisis; the Ukrainian conflict, which they want to get us into a war by the nose, etc.
3	@HbCecilia @MADRID @AEMET_Esp So? Agenda 2030 has nothing to do with ideologies. It is a dictatorial imposition of globalist elites and big corporations. Politicians are just their puppets. It doesn't matter if they are left or right wing.
4	@vbenayas @Rub_dc @AEMET_Esp Another LIE. Since there is data there have been other heat waves with higher temperatures. for now no station has exceeded the maximum for the month of June. less lying. justifying a farce with lies is shameful. So there is no climate change in Asturias? https://t.co/ct7iTaKZNP
5	AEMET_Andalucia MURDERERS HDLPS ILLUMINATIS, YOU WILL PAY, EVERY TIME PEOPLE KNOW MORE ABOUT THE PLANNED NEW WORLD ORDER AND ITS CRIMINAL AGENDA, THE SAME AS THE STONES OF GEORGIA WILL HAPPEN TO ALL THE SATANISTIC AND CRIMINAL STUPIDITY, STOP UN-NATO, NO TO THE ARTIFICIAL MODIFICATION OF THE GENOCIDE CLIMATE https://t.co/Gc8k9Nyppa
6	@Carolina_adh19 @Davicin93 @MADRID @AEMET_Esp Shut up damn bot, you're replying to everyone, don't you have a life? It's absurd that you close it and that's it. It's a beautiful morning and it's not hot at all. I work every day inside the Retiro and this is absurd so shut the fuck up.
7	@ledblues72 @AEMET_Esp That's the problem, they play God, No! They are true Satanist psychopaths, they control all levels of society, 95% of the money is in their power, and they can sink the global economy with a simple war, or by creating a health problem. Ag2030
8	@greenpeace_eng @MonicaParrill @AEMET_Esp Anomalous is that a few years ago you were saying that by this time Valencia would be under the waters of the Mediterranean. Your predictions and scaremongering are rubbish that only aims to impoverish the middle class to subject them to the dictatorship of the "ecohappysustainable."
9	@AEMET_Esp MURDERERS, HDLPS, NO MORE ARTIFICIAL CHEMICAL SUMMERS, GENOCIDES...WHO ARE YOU GOING TO FOOL AT THIS POINT, THE CLIMATE IS MANIPULATED AND PUTS IT IN THE BOE OF SPANISH MAFIA STATE AND PROGRAMMED AGENDAS WE WANT HEADS TO ROLL AND THIS ARTIFICIAL AERIAL CLIMATE GENOCIDE https://t.co/OGxV4nlWYZ
10	@MADRID @AEMET_Esp All to put fear into people...what a load of bullshit measures. Buy the Globalist Agenda 2030 pin because it is what is causing this among many other things, and it is the only thing we have to fight against, and not against these little shit politicians who are just errands.

Notes: PP stands for Partido Popular, the main centre-right party in Spain; HDLPS is a son of a bitch; BOE is the Official State Bulletin of Spain.

As for the appearance of other conspiratorial elements in the hate speech, “chemtrails” appear in 804 messages, “geoengineering” in 1,670 messages, “Covid-19” in 97, “vaccines” in 62, “the HAARP project” (the high atmosphere research antennas that are said to cause catastrophes; see Arce-García & Díaz-Campo, 2024) in 251, and “Agenda 2030” in 688 messages. Elements such as “fumigation” (204) or the “spread of poison” (57) also appear. Thus, in many hate messages, the attack on AEMET is linked to the dissemination of well-known conspiracy theories.

The analysis of the self-descriptions of the accounts in the third quartile with the highest average hate associated with the messages is shown in Figure 3. It can be seen that these users define themselves mainly under the following associated themes: “family” together with “freedom” and “agenda”; “truth” and “freedom”; “Spanish” or “from Spain” together with “father,” “life,” “lover,” “nature,” “death,” and “account”; and “meteorology enthusiast.” There is therefore a clear trend of accounts that define themselves as Spanish, for freedom, family, life, and nature, as well as meteorology amateurs.

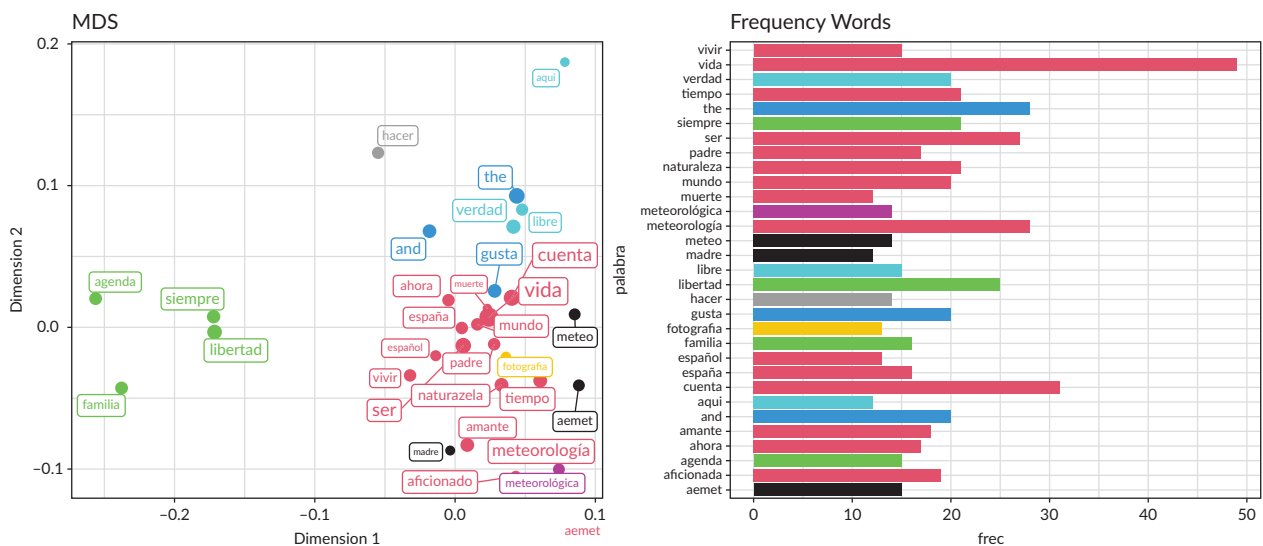


Figure 3. MDS and frequency analysis of the most commonly used words in the description of the accounts with the highest average hate load.

6. Conclusion

This research analyses hate speech against AEMET) on the social network X. Using hate speech detection algorithms, text mining techniques, and qualitative analysis, patterns of hostility against AEMET were studied for messages posted between 31 December 2021 and 19 April 2023. The results show a significant percentage of messages with hate content, driven by conspiracy theories and climate change denial and related to scientific scepticism.

On the extent of hate speech and its impact on public trust (specific objectives 1 and 2), one of the most striking findings of the study is the significant proportion of messages containing hate speech towards AEMET: Around 25% of the texts analysed show some degree of hostility. This suggests that a significant part of the online public conversation about the Spanish state agency is characterised by negative emotions, ranging from direct insults to questions about the professional integrity of scientists and meteorologists.

The impact of this type of message goes beyond the mere emotional or personal damage to AEMET and its professionals, as this hate speech creates a climate of misinformation that contributes to the erosion of public trust in AEMET and its staff, leading to scepticism about science in general (which is the attitude of denialists and conspiracy theorists). This kind of mistrust can have serious consequences, such as a lack of preparedness for climate emergencies or an indifference to critical weather warnings, putting both public safety and environmental sustainability at risk.

Concerning conspiracy theories as a catalyst for hate speech (specific objective 2), our analysis of the messages shows that much of the hatred towards AEMET is linked to the spread of conspiracy theories, particularly those related to geoengineering and chemtrails. These theories, which claim that AEMET and other scientific institutions are manipulating the weather for dark purposes, have been promoted by certain online communities driven by very specific interests. Messages citing these theories not only misinform but also create an environment of hatred, the radicalisation of opinion, and rejection of scientific evidence.

The spread of these theories reflects a wider phenomenon of anti-intellectualism and the rejection of science, fuelled by a distrust of elites, governments, and international institutions. This study highlights how conspiracy theories, based on false or distorted information, can act as a catalyst for hatred and misinformation, exacerbating social polarisation and fostering a hostile attitude towards scientific knowledge and professionals.

On the structure of the hate network in social networks (specific objective 3), another important aspect of the study was the identification of the actors and network structure behind the hate speech on X. Using network analysis techniques, we were able to visualise how certain groups and user accounts were clustered around the dissemination of hate messages. In particular, we identified a cluster of highly interconnected users who specialise in spreading conspiracy theories related to AEMET and meteorology in general.

Concerning impact, this particular group, although relatively small compared to the total number of users, is highly influential because of the number of retweets and quotes they generate. Messages from this cluster tend to be disproportionately shared, amplifying the reach of hate speech and conspiracy theories. These aspects are consistent with Granovetter's (1973) sociological theory that the weak links are the ones who end up contributing the main ideas to the group. Similarly, the most influential accounts within this network tend to be those that promote a discourse of rejection of public policies related to climate change, linking them to global political agendas such as the UN's 2030 Agenda.

On the need for strategic responses to mitigate hate and misinformation (specific objective 4): The purpose of this study is not only to highlight the problems caused by hate speech but also to suggest possible solutions to mitigate its negative effects. First, institutions such as AEMET must implement effective communication strategies that not only provide clear and accurate information but also directly address myths and conspiracy theories circulating on social media. Greater transparency in science communication can help reduce uncertainty and mistrust.

It is also important to foster collaboration between scientific institutions, social media platforms, and governments to develop strategies to limit the spread of hate speech and misinformation. This could include implementing more effective algorithms to detect and remove harmful content, as well as supporting media literacy initiatives that help the public identify and challenge false information.

Finally, it is essential to promote a culture of respectful and evidence-based dialogue on social media. The study shows that hatred towards scientific institutions such as AEMET is motivated not only by manipulation or ignorance but also by increasing social and political polarisation. To counter this phenomenon, it is important to foster an environment in which the exchange of opinions is encouraged and discussed in a constructive and reasoned manner, without resorting to hostility or misinformation.

In the field of meteorology, hate speech not only affects public perception but can also have a direct impact on scientific work. Social pressure and smear campaigns can discourage scientists from interacting with the public or even communicating their research openly. This chilling effect can limit the advance of scientific knowledge and restrict public access to accurate and high-quality information.

In addition, the spread of conspiracy theories about climate control and weather manipulation directly affects the perception of the current climate alert. Denial that extreme weather events are a result of global warming finds support in these theories, undermining global efforts to mitigate and respond to climate change.

6.1. Future Implications and Recommendations for Future Research

This study lays the groundwork for future research in the area of hate speech and its relationship to science. Given that the phenomenon of hatred towards scientific institutions is not unique to AEMET, future research could focus on other scientific fields where hate campaigns, misinformation, and scepticism similarly affect public credibility and trust. For example, vaccination research, renewable energy research, or even medicine in general are areas in which hate speech and conspiracy theories have had a significant impact.

In addition, it would be beneficial to extend the research to allow comparisons between different countries or cultural contexts to identify similarities and differences in the spread of conspiracy theories and hate speech towards science. By exploring how other contexts deal with these issues, new strategies and models may emerge that can be applied to a wider setting.

Hate speech towards AEMET, as demonstrated in this study, is not just an isolated manifestation of online hostility, but part of a wider problem related to distrust of science and the spread of conspiracy theories generated by a clear interest in misinformation. Addressing these issues requires a combination of efforts: improving scientific communication, strengthening content moderation on social media platforms, and promoting a culture of respectful, evidence-based dialogue. Only a collaborative and informed response can mitigate the impact of hate and misinformation on scientific institutions and protect their integrity and credibility in the future.

Funding

This research has been carried out within the framework of the research project Mapping of Hate Speeches in Spain From Communication (PID2019-105613GB-C31), funded by the Ministry of Economy and Competitiveness (Spain), and the Politainment Facing Media Fragmentation: Disintermediation, Engagement by Polarization (PID2020-114193RB-I00), funded by the Ministry of Science and Innovation (Spain). The publication of this work was funded by the Call for Grants for the translation of scientific articles of the Universidad Internacional de La Rioja.

Conflict of Interests

The authors declare no conflict of interests.

Data Availability

Tweets mentioning AEMET 2021–2023 can be found here: <https://doi.org/10.6084/m9.figshare.27961212>

References

- Agencia Estatal de Meteorología. (2024). *Quiénes somos*. https://www.aemet.es/es/conocenos/quienes_somos
- Alkomah, F., & Ma, X. (2022). A literature review of textual hate speech detection methods and datasets. *Information*, 13(6), Article 273. <https://doi.org/10.3390/info13060273>
- Almansa-Martínez, A., Fernández-Torres, M. J., & Rodríguez-Fernández, L. (2022). Desinformación en España un año después de la COVID-19. Análisis de las verificaciones de Newtral y Maldita. *Revista Latina de Comunicación Social*, 80, 183–200. <https://www.doi.org/10.4185/RLCS-2022-1538>
- Arce-García, S., & Díaz-Campo, J. (2024). HAARP conspiracy: Analysis of its role in the 2023 Turkey & Syria earthquakes on Twitter. *Estudios sobre el Mensaje Periodístico*, 30(2), 323–333. <https://doi.org/10.5209/esmp.95257>
- Arce-García, S., Said-Hung, E., & Montero-Díaz, J. (2024). Unmasking coordinated hate: Analysing hate speech on Spanish digital news media. *New Media & Society*. Advance online publication. <https://doi.org/10.1177/14614448241259715>
- Arce-García, S., Said-Hung, E., & Mottareale-Calvanese, D. (2022). Tipos de campaña Astroturfing de contenidos desinformativos y polarizados en tiempos de pandemia en España. *Revista Científica De Comunicación Y Tecnologías Emergentes*, 21(1). <https://doi.org/10.7195/ri14.v21i1.1890>
- Arcila Calderon, C., Sánchez Holgado, P., Gómez, J., Barbosa, M., Qi, H., Matilla, A., Amado, P., Guzmán, A., López-Matías, D., & Fernández-Villazala, T. (2024). From online hate speech to offline hate crime: The role of inflammatory language in forecasting violence against migrant and LGBT communities. *Humanities & Social Sciences Communications*, 11, Article 1369. <https://doi.org/10.1057/s41599-024-03899-1>
- Asmelash, L. (2024, March 12). ¿Qué es la teoría de la conspiración de los chemtrails, o estelas químicas, y por qué es popular? *CNN Español*. <https://cnnespanol.cnn.com/2024/03/12/teoria-conspirativa-chemtrails-estelas-quimicas-trax>
- Barabasi, A. L. (2016). *Network science*. Cambridge University Press.
- Barrie, C., & Chun-ting, H. J. (2021). AcademictwitterR: An R package to access the Twitter Academic Research Product Track v2 API endpoint. *Journal of Open Source Software*, 6(62), Article 3272.
- Bergsmanis-Korats, G., & Haiduchyk, T. (2024). *Social media manipulation for sale. 2024 experiment on platform capabilities to detect and counter inauthentic social media engagement*. Nato Strategic Communications Centre of Excellence. <https://stratcomcoe.org/publications/social-media-manipulation-for-sale-experiment-on-platform-capabilities-to-detect-and-counter-inauthentic-social-media-engagement/311>
- Blondel, V., Guillaume, J., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), Article 10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Center for Countering Digital Hate. (2023). *How Twitter/X continues to host posts reported for extreme hate speech*. https://counterhate.com/wp-content/uploads/2023/09/230907-X-Content-Moderation-Report_final_CCDH.pdf
- Center for Countering Digital Hate. (2024). *The new climate denial*. https://counterhate.com/wp-content/uploads/2024/01/CCDH-The-New-Climate-Denial_FINAL.pdf

- Chen, J., Hossain, S., & Zhang, H. (2020). Analyzing the sentiment correlation between regular tweets and retweets. *Social Network Analysis and Mining*, 10, Article 13. <https://doi.org/10.1007/s13278-020-0624-4>
- DeVerna, M. R., Aiyappa, R., Pacheco, D., Bryden, J., & Menczer, F. (2024). Identifying and characterizing superspreaders of low-credibility content on Twitter. *PLoS ONE*, 19(5), Article 0302201. <https://doi.org/10.1371/journal.pone.0302201>
- Diethelm, P., & McKee, M. (2009). Denialism: What is it and how should scientists respond? *European Journal of Public Health*, 19(1), 2–4.
- Dionis, M. (2022). Del retardismo climático al tecnooptimismo: Los nuevos discursos de desinformación sobre el cambio climático. *Newtral*. <https://www.newtral.es/desinformacion-cambio-climatico/20221108>
- Edelman Trust. (2024). *Edelman Trust Barometer 2024*. <https://www.edelman.com/es/trust/2024/spain-trust-barometer>
- El Rastreador. (2023, April 20). “Asesinos,” “lo pagaréis”: Insultos y amenazas contra la Agencia Estatal de Meteorología por contar la crisis climática. *El Diario*. https://www.eldiario.es/rastreador/asesinos-pagareis-insultos-amenazas-agencia-estatal-meteorologia-contar-tesis-climatica_132_10138006.html
- Ennis, G. (2023). *Dark PR: How corporate disinformation harms our health and the environment*. Daraja Press.
- European Commission. (2024). *Strategic communication & tackling disinformation*. https://commission.europa.eu/topics/strategic-communication-and-tackling-disinformation_en
- European Union. (2024). *Standard Eurobarometer 98–Winter 2022–2023*. <https://europa.eu/eurobarometer/surveys/detail/2872>
- Ferreira Caceres, M. M., Sosa, J. P., Lawrence, J. A., Sestacovschi, C., Tidd-Johnson, A., Rasool, M. H. U., Gadamidí, V. K., Ozair, S., Pandav, K., Cuevas-Lou, C., Parrish, M., Rodríguez, I., & Fernandez, J. P. (2022). The impact of misinformation on the COVID-19 pandemic. *AIMS Public Health*, 9(2), 262–277. <https://doi.org/10.3934/publichealth.2022018>
- Fidalgo, P. (2020, November 11). WHO and Wikipedia team up against ‘fake news’ about coronaviruses. *El Plural*. https://www.elplural.com/leequid/voces/oms-wikipedia-alian-fake-news-coronavirus_252134102
- García-Orosa, B. (2021). Disinformation, social media, bots, and astroturfing: The fourth wave of digital democracy. *Profesional de la información*, 30(6). <https://doi.org/10.3145/epi.2021.nov.03>
- Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78, 1360–1380.
- Hoaxes, disinformation, conspiracies and threats that attempt to discredit the AEMET. (2024, October 31). *Maldita*. <https://maldita.es/clima/20241031/bulos-desinformaciones-desacreditar-atacar-aemet>
- IPSOS. (2023). *Global views on AI and disinformation*. <https://www.ipsos.com/en-nz/global-views-ai-and-disinformation>
- Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLOS ONE*, 9(6), Article 98679. <https://doi.org/10.1371/journal.pone.0098679>
- Jiménez-Gómez, I., & Martín-Sosa, S. (2022). Análisis discursivo del escepticismo climático en los medios impresos y digitales españoles entre 2015 y 2021. *Estudios sobre el Mensaje Periodístico*, 28(3), 525–536. <https://doi.org/10.5209/esmp.80779>
- Jockers, M. (2017). *Syuzhet: Extracts sentiment and sentiment-derived plot arcs from text* [Software]. <https://rdrr.io/cran/syuzhet>
- La AEMET responde a las amenazas recibidas en los últimos días: “No todo vale.” (2023, April 21). *La Vanguardia*. <https://www.lavanguardia.com/natural/20230421/8913932/aemet-responde-amenazas-recibidas-ultimos-dias-vale.html>

- Luis del Pino cambia esRadio por El Toro TV. (2024, February 21). *La Región*. <https://www.laregion.es/articulo/sociedad/luis-pino-cambia-esradio-toro-tv/202402211810231275149.html>
- Nieto-Sandoval, A. G., & Ferré-Pavia, C. (2023). TikTok y cambio climático: Comunicar sin fuentes ni soluciones. *Revista de Comunicación*, 22(1), 309–331. <https://doi.org/10.26441/RC22.1-2023-2994>
- Olivares-Delgado, F., Benlloch-Osuna, M., Rodríguez-Valero, D., & Breva-Franch, E. (2023). Corporate disinformation: Concept and typology of forms of corporate disinformation. In N. Martins & D. Brandão (Eds). *Advances in design and digital communication III* (pp. 536–550). Springer. https://doi.org/10.1007/978-3-031-20364-0_46
- Ong, J.-C., Tapsell, R., & Curato, N. (2019). *Tracking digital disinformation in the 2019 Philippine midterm election*. New Mandala.
- Palomares Calderón, M. (2015). *Breve historia de la Agencia Estatal de Meteorología (AEMET)*. AEMET. https://repositorio.aemet.es/bitstream/20.500.11765/899/1/breve_historia_AEMET.pdf
- Rodrigo Cano, D., & Del Río Álvarez, M. J. (2021). La desinformación sobre salud y cambio climático en redes sociales. In J. C. Figueroa-Benítez & R. Mancinas-Chávez (Eds.), *Las redes de la comunicación: Estudios multidisciplinares actuales* (pp. 833–846). Dykinson.
- Rodríguez-Fernández, L. (2019). Desinformación y comunicación organizacional: Estudio sobre el impacto de las fake news. *Revista Latina de Comunicación Social*, 74, 1714–1728. <http://www.revistalatinacs.org/074paper/1406/89es.html>
- Rodríguez-Fernández, L. (2023). Desinformación y relaciones públicas. Aproximación a los términos Black PR y Dark PR. *ICONO 14. Revista Científica De Comunicación Y Tecnologías Emergentes*, 21(1). <https://doi.org/10.7195/ri14.v21i1.1920>
- Said-Hung, E., Römer Pieretti, M., Montero-Díaz, J., De Lucas Vicente, O., & Martínez, J. (2023). *Hate speech library in Spanish* [Data set]. Proyecto Hatemedia. <https://doi.org/10.6084/m9.figshare.22383643.v2>
- Santos, C., Coelho, A., & Marques, A. (2024). The greenwashing effects on corporate reputation and brand hate, through environmental performance and green perceived risk. *Asia-Pacific Journal of Business Administration*, 16(3), 655–676. <https://doi.org/10.1108/APJBA-05-2022-0216>
- Supran, G., Rahmstorf, S., & Oreskes, N. (2023). Assessing ExxonMobil's global warming projections. *Science*, 379(6628). <https://doi.org/10.1126/science.abk0063>
- Twitter and other social media sites slipped on removing hate speech in 2022, EU review says. (2022, November 24). CBS News. <https://www.cbsnews.com/news/twitter-other-social-media-slip-on-removing-hate-speech-european-union-review>
- UNESCO. (2023, November 6). *Online disinformation: UNESCO unveils action plan to regulate social media platforms* [Press release]. <https://www.unesco.org/en/articles/online-disinformation-unesco-unveils-action-plan-regulate-social-media-platforms>
- Union of Concerned Scientists. (2007). *Smoke, mirrors & hot air: How ExxonMobil uses Big Tobacco's tactics to manufacture uncertainty on climate change*. <https://www.jstor.org/stable/resrep00046.7?seq=2>
- Vicente Torrico, D., & González Puente, V. (2023). Negacionismo y desinformación climática en YouTube. Análisis del discurso negacionista entre los vídeos más vistos en castellano. *Miguel Hernández Communication Journal*, 14(1), 89–108. <https://doi.org/10.21134/mhjourn.v14i.181>
- Vicente Torrico, D., Hernando Lera, M., & González Puente, V. (2024). El obstruccionismo climático en redes sociales: Desinformación y ataques contra las voces de la ciencia. *Zer*, 29(56), 173–199. <https://doi.org/10.1387/zer.25929>
- We are Social. (2023). *Special report digital 2023*. <https://wearesocial.com/es/blog/2023/01/digital-2023>
- WHO. (2024, September 24). *WHO TikTok collaboration to generate more science-based information on*

health and wellness. <https://www.who.int/es/news/item/26-09-2024-who-and-tiktok-to-collaborate-on-more-science-based-information-on-health-and-well-being>

Wiesehomeier, N., & Flynn, D. J. (2020). *¿Quién se cree las 'fake news' en España?* Fundación La Caixa. <https://elobservatoriosocial.fundacionlacaixa.org/es/-/quien-se-cree-las-fake-news-en-espana>

World Economic Forum. (2024). *The global risk report 2024*. https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf

About the Authors



Sergio Arce-García is associate professor and researcher at the International University of La Rioja (UNIR). He earned a PhD in humanities and communication from the University of Burgos, with an extraordinary doctorate award, and a degree in chemistry from the University of Valladolid. He teaches at the School of Engineering and Technology at UNIR and researches communication (study of media and social networks). Recognised for a six-year research period.



Virginia Martín-Jiménez is associate professor at the University of Valladolid (UVA). She earned a PhD in journalism and history with an extraordinary doctorate award and is now part of the research projects Mapping of Hate Speeches in Spain From Communication and Politainment Facing Media Fragmentation: Disintermediation, Engagement by Polarization. She has training in several disciplines—history, journalism, and politics and sociology—which is reflected in her career and the focus of her research: television studies, political communication, and hate speech and digital communication.



Leticia Rodríguez-Fernández is a professor in the Department of Marketing and Communication and General Director of Strategic Communication at the University of Cádiz. She is an expert in the fields of strategic communication and disinformation, which she has woven into her academic work. She holds a PhD in information sciences, a master's degree in protocol, a master's degree in communication of public and political institutions, and a bachelor's degree in advertising and public relations.

Semiotic Analysis of Hate Discourse in Spanish Digital News Media: Biden's Inauguration Case Study

Max Römer-Pieretti ¹ , Elías Said-Hung ²  and Julio Montero-Díaz ^{2,3} 

¹ Communications & Design Faculty, Camilo José Cela University, Spain

² Universidad Internacional de La Rioja, Spain

³ Villanueva University, Spain

Correspondence: Max Römer-Pieretti (mwalter@ucjc.edu)

Submitted: 13 September 2024 **Accepted:** 26 November 2024 **Published:** 13 January 2025

Issue: This article is part of the issue “Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment” edited by Max Römer Pieretti (Universidad Camilo José Cela), Beatriz Esteban-Ramiro (Universidad de Castilla-La Mancha), and Agrivalca Canelón (Universidad Católica Andrés Bello), fully open access at <https://doi.org/10.17645/si.i415>

Abstract

This study analyzes hate in Spanish digital media from a semiotic standpoint by focusing on the coverage and discourse of Joe Biden's inauguration as the US President in January 2021 by *El País*, *La Vanguardia*, ABC, *El Mundo*, and 20Minutos in Spain on the X platform. The event drew significant attention from international and Spanish media. A qualitative investigation was conducted on the interactions, denotations, connotations, and semiosis related to hate in the Spanish media and their followers. The analysis, which is based on a semiotic matrix from Greimas and Courtés (1979), Greimas (1976), Barthes (1970), Kristeva (1969), and Lyotard (1979/2019), and was developed by the authors, covered 661 news items and 721 literal fragments and generated 2,074 interactions for examination. This study offers a semiotic framework for understanding how hate expressions are constructed and disseminated in digital media. It is crucial to recognize the narrative structures that promote the dissemination of hate expression in news content published by digital media on social media platforms. A scenario emerges in which fear, politically charged expressions, and terms aimed at accusing, discrediting, or undervaluing the recipients of such messages become tools for spreading content. Therefore, digital news media must review their content moderation practices to better manage the discussions generated concerning the news that they publish in the current digital landscape. This landscape is dominated not only by hostility rather than violence toward social groups represented by news protagonists but also by people who are used to promoting narratives filled with stereotypes and prejudices through dehumanization or demonization.

Keywords

digital news media; digital semiotics; discourse analysis; hate speech; social media platforms

1. Introduction

The semiotic structures in societies are complex, multifaceted, and deeply rooted in their diverse social, institutional, and communicative frameworks (Young-Jung Na, 2023). These underlying structures contribute to the dissemination and establishment of hate speech (Kahn, 2022), primarily because of the ubiquity of social media in the current digital environments in which communication occurs. In these contexts, power dynamics emerge that intertwine these structures in public opinion debates (Koval et al., 2019). Accordingly, the development of semiotic analysis of the dissemination of hate speech has become increasingly important. Hate messages are defined as messages of a public nature with the intention of humiliating, discriminating against, or even physically threatening an individual belonging to a group that is vulnerable because of their gender, race, ethnicity, ideology, religion, or other shared specific characteristics (Gómez-García et al., 2021). These analyses elucidate the semiotic strategies that integrate these terms into social discourse (Bianca, 2021). They also identify the various symbolic meanings embedded in the published digital content and facilitate an understanding of the mechanisms of propagation of these expressions and their subsequent characterization as exclusionary, threatening, or violent toward certain social groups (Erdogan-Ozturk & Isik-Guler, 2020; Gramigna, 2022; Riquelme et al., 2022).

Social media platforms contribute to the propagation of hostile and violent media environments. These platforms engender sentiments of anger, resentment, and opposition, which foster antagonistic or adversarial attitudes toward specific individuals or groups or promote the utilization of physical force or aggressive actions to inflict harm (Kim, 2022; Levin-Banchik, 2020; Walters & Espelage, 2020). The proliferation of hate speech of varying intensity in digital communication contexts likely serves as a foundation for reinforcing hostile rhetoric, which functions as a precursor and heightens the probability of social violence directed toward specific groups (Cover, 2022; Kim, 2022; Lee et al., 2022).

Users and participants in debates facilitated by digital media occasionally generate expressions that legitimize and normalize prejudices and stereotypes (Perreault, 2023). Algorithms typically govern such conversations (Kleis Nielsen & Ganter, 2018; Salonen et al., 2022). The limitations of existing moderation procedures have prompted efforts to develop mechanisms that ensure the effective management of these discussions and maintain a minimum standard of quality in the contributions of participating users (Lin & Kim, 2023).

The proliferation of hate speech primarily affects social media platforms, which are experiencing increasing polarization and the growing presence of hate speech and disinformation (Czopek, 2024). Political content is predominant in this dissemination (Falkenberg et al., 2024). Media organizations are considering withdrawing from these environments (“‘La Vanguardia’ dejará,” 2024; Soni & Singh, 2024) because of the ineffectiveness of regulatory policies (Bergmanis-Koräts & Haidechyk, 2024; Center for Countering Digital Hate, 2023). The consequence is the promotion of an increasingly toxic and polarized communication context. Social agents and media entities must redefine their roles as traditional gatekeepers and consider new responsibilities that ensure positive and enriching dynamics in news dissemination, consumption, and discourse in contemporary social media landscapes (Kobellarz et al., 2022; Salonen et al., 2022).

A correlation between current events presented by the media and the prevalence of disinformation and hate speech in the ensuing debates has been observed, which is characterized by the utilization of textual and semiotic elements that construct hostile and violent narratives toward specific groups, such as politicians,

women, immigrants, and LGBTBI+ (Erdogan-Ozturk & Isik-Guler, 2020; Rajan & Venkatraman, 2021; see also Hameleers et al., 2022). This necessitates understanding the visual and textual codes that legitimize and normalize prejudices and stereotypes toward these groups (Rajan & Venkatraman, 2021). These codes are addressed in this study, which aims to comprehend from a semiotic perspective how symbols, languages, and images are employed to transmit and disseminate expressions of hate within the debate contexts associated with information in digital environments. Specifically, this study focuses on the authors of these media platforms in Spain. The objective of this study is to advance the establishment of linguistic markers that facilitate a better understanding of hate speech (Määttä, 2023).

2. Literature Review on the Semiotic Analysis of Hate

There has been an increase in semiotic studies on digital communication, which encompass both textual and visual elements (Ghaffari, 2020). Despite this growth, the outcomes of the emerging field of “digital semiotics” remain limited (Berlanga-Fernández & Reyes, 2022) in regional studies, such as those conducted in Spain (Salaverría & Martínez-Costa, 2023) in specific domains, particularly cultural and literary topics (Castillo, 2022; Navarrete, 2019), and in contexts of particular significance, notably during the Covid-19 pandemic (Rubio-Pinilla & Candón-Mena, 2021). Hate speech propagation on social media platforms and its societal implications can be understood by analyzing these discourses’ underlying symbolic and communicative structures (Vasist et al., 2023). The findings from such analyses will facilitate not only a more comprehensive understanding of the construction, dissemination, and perception of messages containing this type of expression but also their capacity to influence public opinion (Lilleker & Pérez-Escolar, 2023; Schäfer et al., 2022).

This semiotic perspective is multifaceted (multimodal). It examines social media’s linguistic, emotional, and communicative dynamics in expressing and disseminating hate speech. These mechanisms enhance the persuasive power of the promoted narratives by combining verbal and iconic elements that involve rational and emotional cognitive components (Koltsova & Kartashkova, 2022). In Spain, this approach has focused on the study of the media framework, linguistic characteristics, and social events that have facilitated the shaping and dissemination of hatred toward immigrants (Gómez-Camacho et al., 2023; Lilleker & Pérez-Escolar, 2023). At the international level, this topic has become more comprehensive and multifaceted. Some of the manifestations of this approach include the identification of strategies, factors, and discursive elements to evade the action of media moderation mechanisms (Retta, 2023); the behavioral characterization of users and their psycholinguistic patterns (Perera et al., 2023); the exploration of mechanisms that enable multimodal, text, and image analysis (Chhabra & Vishwakarma, 2023); and the examination of media discourses employed by contemporary digital news media in favor of a more informed and balanced public opinion (Labiano et al., 2023).

This semiotic analysis of contemporary digital communication scenarios necessitates advancements in defining the interactions of users who comment. This entails a requisite review and adaptation of certain fundamental concepts in semiotics. Such an undertaking requires a process of adaptation and integration of concepts that serve as a foundation for developing the analysis process from a semiotic perspective.

This work has been progressing from a semiotic perspective in various geographical regions. This is exemplified by studies conducted by Fonseca et al. (2024) in Portugal, which focus primarily on identifying

the interaction patterns that aid in estimating the temporal occurrence of such messages within social media platform debates. Additionally, other research such as that conducted by Barth et al. (2023) has analyzed diverse communicative contexts in the United States, Finland, Great Britain, and Germany to assess the influence of these contexts on the interpretations associated with hate-expressing messages. Retta (2023) studied in Italy how insults and derogatory epithets reinforce polarization and strengthen values within social media platform user groups. Research by Phadke et al. (2018) concentrated on analyzing hate messages from the perspective of framing and propaganda in the United States. Furthermore, Määttä (2023) reported an absence of linguistic tools adapted to cultural contexts in Finland and France. This research is necessary for comprehending and identifying the discursive structures of hate-expressing messages and how they are legitimized in debates generated on social media platforms.

For these reasons, this study, in pursuit of its overarching objective, revisits the theoretical frameworks of Greimas and Courtés (1979), Kristeva (1969), Lyotard (1979/2019), and Barthes (1970) as conceptual resources that may contribute to the examination of hate expressions in contemporary digital environments. Although these scholars did not formulate their theoretical postulations from a digital perspective, the timeless nature of their semiotic principles, which focus on studying signs, symbols, and meaning-making processes, can be applied to digital communication with an efficacy comparable to that of traditional media. This allows us to adapt these theories to current digital communication scenarios where hate expressions are disseminated.

This discussion exemplifies Kristeva's (1969) concept of intertextuality, which entails conceptualizing linguistic structures as isomorphic or analogous to other systems, such as literature and politics (Alirangues, 2018). This concept applies to analyzing hate messages spread on social media platforms, particularly regarding their capacity to reference and influence one another (Barth et al., 2023).

Additional concepts pertinent to this analysis are message narratives and connotative characteristics (Barthes, 1970). The literal definition of a word or phrase (denotation) and its cultural and emotional associations (connotations) are included. Consequently, the comprehension of aesthetic and narrative elements (denotative) and cultural and emotional aspects (connotative) can contribute to the construction of meaning because of the capacity to generate a novel combination with the signifier, which subsequently becomes meaningful (Nöth, 2011).

Barthes's (1970) approach, when applied to the digital domain and focused on the study of hate speech, can facilitate the identification of messages containing hostile narratives toward specific social groups. This methodology enables the classification of various linguistic nuances (literal and explicit content, underlying implications, and emotional connotations) associated with such content. Furthermore, it contributes to understanding the impact of the interactions generated around these messages (Tontodimamma et al., 2022). This approach has been instrumental in highlighting the predominance of discursive strategies and underlying values that tend to be more connotative (Inwood & Zappavigna, 2023).

Greimas and Courtés' (1979) proposal of generative trajectories and Lyotard's (1979/2019) approach to the narrative of emancipation are of significant interest. The former, oriented toward the narrative analysis of the structural and semantic dimensions of messages, emphasizes the roles and functions of these elements (Hernández et al., 2023; Imbert, 2019). The latter, which was not originally conceived to analyze digital

communication scenarios, highlights the need to focus on the study of micronarratives. Micronarratives can elucidate transformations and issues in society through the recognition of the importance of situational perspectives that aid in assessing the legitimacy of various discourses promoted through language. This approach is particularly relevant when considering the capacity of digital spaces for user groups to share their narratives (Christian et al., 2020; Sanders et al., 2023). Both concepts can contribute to a more comprehensive understanding of hate speech dissemination on social media because they facilitate a better understanding of the associated narrative dynamics (Fan et al., 2022) and their interconnected nature and evolution (Johnson et al., 2019). This study integrates these theories and their application to the semiotic and multidimensional analysis of hate speech on social media.

3. Methodology

This study investigates the dissemination of hate through digital news media in Spain (among users on the X platform) by employing a semiotic perspective. The research focuses on a case study: the coverage in five primary digital news media with user accounts on X regarding the inauguration of Joe Biden as president on January 20, 2021. The digital news media examined include *El País* (@el_pais), *El Mundo* (@elmundoes), ABC (@abc_es), 20Minutos (@20m), and *La Vanguardia* (@LaVanguardia). According to Statista (2020a, 2020b) and Miguel (2020), these outlets were selected based on their prominence at the time.

The event was selected for its international significance, as it generated extensive media coverage, partly due to preceding events (Trump's electoral defeat and subsequent allegations of electoral fraud in certain states and the assault on the United States Congress on January 6, 2021). These circumstances resulted in high polarization and increased journalistic interest in the event.

The following specific aims were established to accomplish the objectives delineated in this study:

- OE1. Assess the prevalence of hate speech in messages associated with news content published by the digital news media under investigation.
- OE2. Identify the dimensions of hatred presented in the content published in this digital news media.
- OE3. Establish the semiotic elements associated with hatred.
- OE4. Determine the relationships among hatred, denotation, connotation, and semiosis in the comments generated in response to the news content published by the selected digital news media.

The study employs a qualitative approach by utilizing an intentional sample that initially comprised seven news items published from the accounts of the selected digital news media in Spain on X and all of the comments generated in response to them until March 2024, the date on which this sample was collected (Table 1). These news items were published on the day preceding the speech (January 19, 2021) and following the event (January 21, 2021).

The acquisition of the informative content identified and utilized as a case study and the subsequent generated comments were achieved through the application of the API and the Airflow platform in X. To extract these cases, all informative content and associated comments from the five digital news media under consideration were obtained (Table 2).

Table 1. Informative content taken as case studies of the topic.

Media	Informational content URL	Informational content publication date	Posts	Number of associated comments
<i>El Mundo</i>	https://x.com/elmundoes/status/1351980993968799749	January 20, 2021	30	34
ABC	https://x.com/abc_es/status/1351924787703144455	January 20, 2021	41	349
20Minutos	https://x.com/20m/status/1351432082262794241	January 19, 2021	8	8
<i>La Vanguardia</i>	https://x.com/LaVanguardia/status/1352197277805400065	January 20, 2021	2	1
<i>La Vanguardia</i>	https://x.com/LaVanguardia/status/1351944823855575040	January 21, 2021	2	5
<i>El País</i>	https://x.com/el_pais/status/1351934637552107522	January 21, 2021	69	104
<i>El País</i>	https://x.com/el_pais/status/1352190043905994753	January 20, 2021	4	4

Table 2. Distribution of the total messages collected and analyzed in January 2021 in the media selected for the case study.

Media analyzed on X	Total number of information content collected	Total number of information content resulting from preprocessing
20Minutos	77,796	58,629
ABC	72,893	56,632
<i>El Mundo</i>	35,242	35,123
<i>El País</i>	56,901	43,582
<i>La Vanguardia</i>	42,970	42,965

The messages were preprocessed following the procedure applied by Ruíz-Iniesta et al. (2024), which included the removal of empty or duplicate texts; eliminating URLs, emojis, and user mentions; and cleaning and homogenizing the data. This standardization process involved converting text to lowercase; removing punctuation marks and additional blank spaces; eliminating numbers, stop words, and words with fewer than two characters; and tokenizing and lemmatizing the text and words comprising each message. The total number of informative content collected and cleaned was 236,931. Manual identification was conducted on this corpus to select the content relevant to the topic analysis proposed in this study. The resulting sample comprised 661 comments, including 156 news items published within the seven news items selected as a case study, and 505 comments generated in response to these items.

Among the 661 content (study sample), 721 verbatim or literal fragments were not only analyzed textually, specifically, fragments within this content, but also extracted for semiotic analysis based on the specific objectives delineated in this study (Figure 1). This analysis generated 2,074 interactions, which represented the various combinations of each word obtained, with the rest predicated on the categories established for analysis.

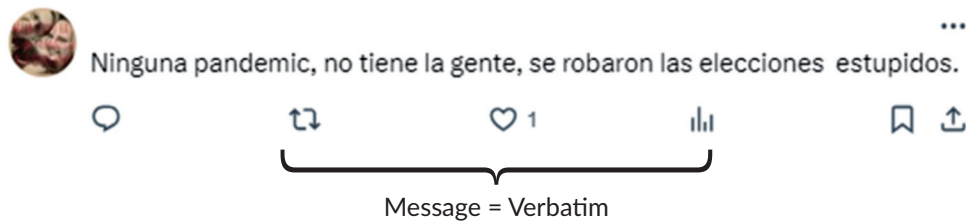


Figure 1. Identification of the verbatim content in the study sample.

An analysis matrix was developed for semiotic analysis on the basis of the theoretical frameworks presented by Greimas and Courtés (1979), Barthes (1970), Kristeva (1969), and Lyotard (1979/2019). This matrix is illustrated in Figure 2. Initially, the denotative and connotative aspects of the selected messages were examined. Furthermore, this study analyzed the emancipatory narrative applied to these messages and the speculative approaches associated with the narratives employed by users who participated in the discourse generated by the analyzed publications.

The matrix presented in Figure 2 proposes an analysis of the text and its accompanying image from both denotative and connotative perspectives, traversing the destructive/constructive domain to evaluate their capacity for semiosis generation. This analysis was conducted within the framework of the emancipation narrative, which posits that messages can be simultaneously independent and dependent. Essentially, the context in which a message expression occurs may either attempt to generate a metalanguage or respond to one. The matrix demonstrates the potential of primary and secondary messages, indicating that additional messages can be derived from a single text. This inherent complexity in the narrative necessitates thoroughly examining each internet user’s input to determine the significance of the interactions.

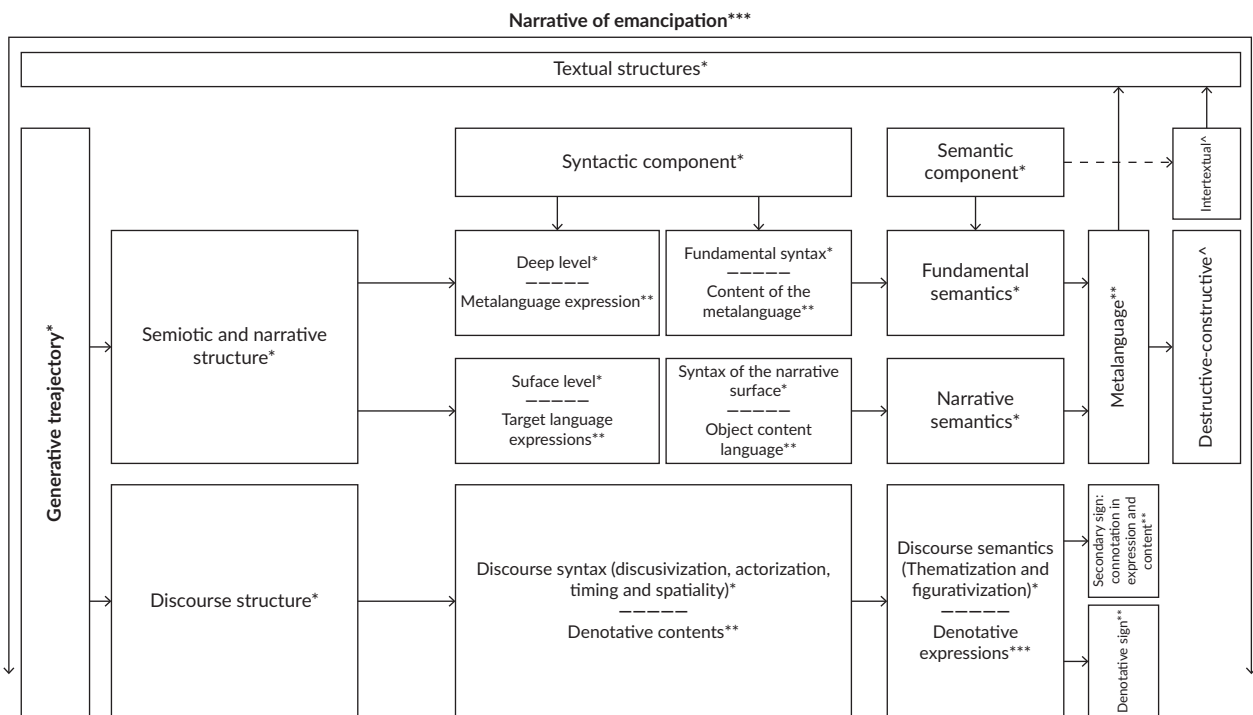


Figure 2. Semiotic analysis matrix applied to the analyzed messages. Note: Prepared by the authors based on *Greimas and Courtés (1979), **Barthes (1970), ^Kristeva (1969), and ***Lyotard (1979/2019).

Concerning the prevalence and intensity of hate speech observed in the corpus of the messages analyzed in this study, the following intensity levels were utilized as a reference framework (De Lucas et al., 2022):

- Intensity level 0: Communication exhibiting expressions of prejudice wherein a group or social collective is delineated or identified in a pejorative manner or through negative connotations.
- Intensity level 1: Communication containing expressions of prejudice utilized factually to stigmatize a specific group or social collective.
- Intensity level 2: Communication including discriminatory expressions of an abusive nature or conveyed with malicious intent to unequivocally attribute specific actions to promote an unfavorable perception of the members of a group or social collective.
- Intensity level 3: Communication involving expressions of animosity characterized by verbal aggression toward specific groups or social collectives.
- Intensity level 4: Expressions that allude to veiled or implicit threats, manifestations of satisfaction regarding violence directed toward a specific group or social collective, or the utilization of intimidating terminology.
- Intensity level 5: Statements that advocate for action or promote explicit violence against a specific individual or social group or where a clear intention is expressed for them to experience physical harm.

These six levels can be categorized into two primary groups (Kim, 2022; Levin-Banchik, 2020; Walters & Espelage, 2020). The first group with intensities of 0, 1, and 2 encompasses expressions of hate that foster a context of media hostility against vulnerable social groups. The second group of messages with expressions of hate have intensities of 3, 4, and 5 and promotes a context of media violence, with an implicit or explicit call to action, against the members of these social groups.

The Atlas.Ti application (version 23) was employed for the processing and semiotic analysis of all comments associated with the information content published by the selected media. Seventeen analysis categories were classified, including the theoretical construct of reference (Table 3).

Based on the traditional semiotic conceptual references in Table 3, this approach was applied to hate speech in contemporary digital contexts. This procedure has the potential to enhance current methodological approaches that utilize semiotics from a computational perspective, such as those that employ machine learning techniques and natural language processing (NLP) or the lexicon associated with these expressions (Arce-García & Menéndez-Menéndez, 2022; Linguardi et al., 2019). From the perspective of cultural semiotics, semiosphere models (Lotman, 1996) have also been applied to social networks (Gramigna, 2022). These models focus primarily on understanding the impact of algorithms in selecting and disseminating content that promotes polarization and incites hatred. Barth et al. (2023) drew upon systems theories and concentrated on understanding the functions and consequences of hate speech in digital ecosystems. Both approaches offer methodological frameworks complementary to the one presented here, which aims to identify the linguistic patterns in hate speech on social media platforms.

Table 3. Analysis categories applied to the sample.

Reference author	Category	Indicator
Greimas and Courtés (1979)	Generative path	Representation of meaning in the message.
	Textual structures	Messages that seek narrativity.
	Semantic component	Parts of the semantic structure of the message.
	Semiotic and narrative structure	The narrative structure of the message.
	Discursive structure	Modalities within the types of discourse.
	Syntactic component	Rules of message structure.
	Surface level	Semantic construction of the message.
	Syntax of the narrative surface	Manipulation of state and action statements in the message.
	Deep level	When the character of the message is the one who expresses himself.
	Discursive syntax	The first enunciative level is the sentence's relationship with its subject, verb and predicate (example).
	Fundamental syntax	The role that words play in the message.
	Denotative sign	It relates to something directly indicated.
	Semantic component	What it implies to the surrounding world.
	Fundamental semantics	The denotative or connotative character of the language used in the message.
	Narrative semantics	The meaning of the text.
	Discursive semantics	The meaning of the message.
	Actorization	Refers to the presence and creation of an actor within the message.
	Figurativization	Refers to the presence of a figure that may have a public character.
	Thematization	Refers to the introduction of a topic in the message.
	Barthes (1970)	Denotative expressions
Language object expressions		Refers to the objects of the message.
Language object content		What the objects mean within the language.
Metalinguage content		When the language can speak for itself.
Metalinguage expression		The spoken topic is the language itself.
Metalinguage		Language that is used to speak another language.
Secondary sign		Not necessarily intended to communicate.
Kristeva (1969)	Denotative sign	That which indicates and points out.
	Intertextuality	Discursive relationships within the messages are taken as a sample.
	Destructive	Going beyond communicative language by deconstructing the text. Seeking destruction based on the message.
	Constructive	Going beyond communicative language by deconstructing the text. Seeking construction based on the message.
Lyotard (1979/2019)	Intertextuality	Relationship between messages.
	Narrative of emancipation	Totalizing and multi-encompassing speeches are those that can close a discussion.

4. Results

4.1. Presence, Dimensions, and Semiotic Elements of Promoting Hatred

A total of 92.4% of the interactions derived from the text extracted from the 661 news items and comments that comprise the sample contained expressions of hate. Such expressions were absent in the remaining 7.6% of the sample. Consequently, it can be concluded that users who propagate hate effectively use this platform to disseminate their narrative.

Joe Biden's inauguration as President of the United States was utilized as an event for disseminating various forms of prejudice, including misogynistic, political, racial, and xenophobic sentiments. These manifestations of bias were frequently predicated on the propagation of fear (e.g., plutophobia), the employment of politically charged terminology, and the utilization of language intended to accuse, discredit, or devalue the recipients of such communications.

The results generated from the semiotic analysis facilitated the description of three fundamental elements: semiotic interactions, hatred intensity, and hatred type. Figure 3 illustrates examples of hate relation typologies, where various levels are integrated, along with diverse semiotic elements, denotations, and connotations. These examples are derived from the user interventions in discussions on news content published by digital news media. Through these examples, this study aims to demonstrate how most of the messages analyzed in this work present exclusionary language associated with intensities 0, 1, and 2 considered in the methodological section. This observation reveals an underlying semiotic structure more

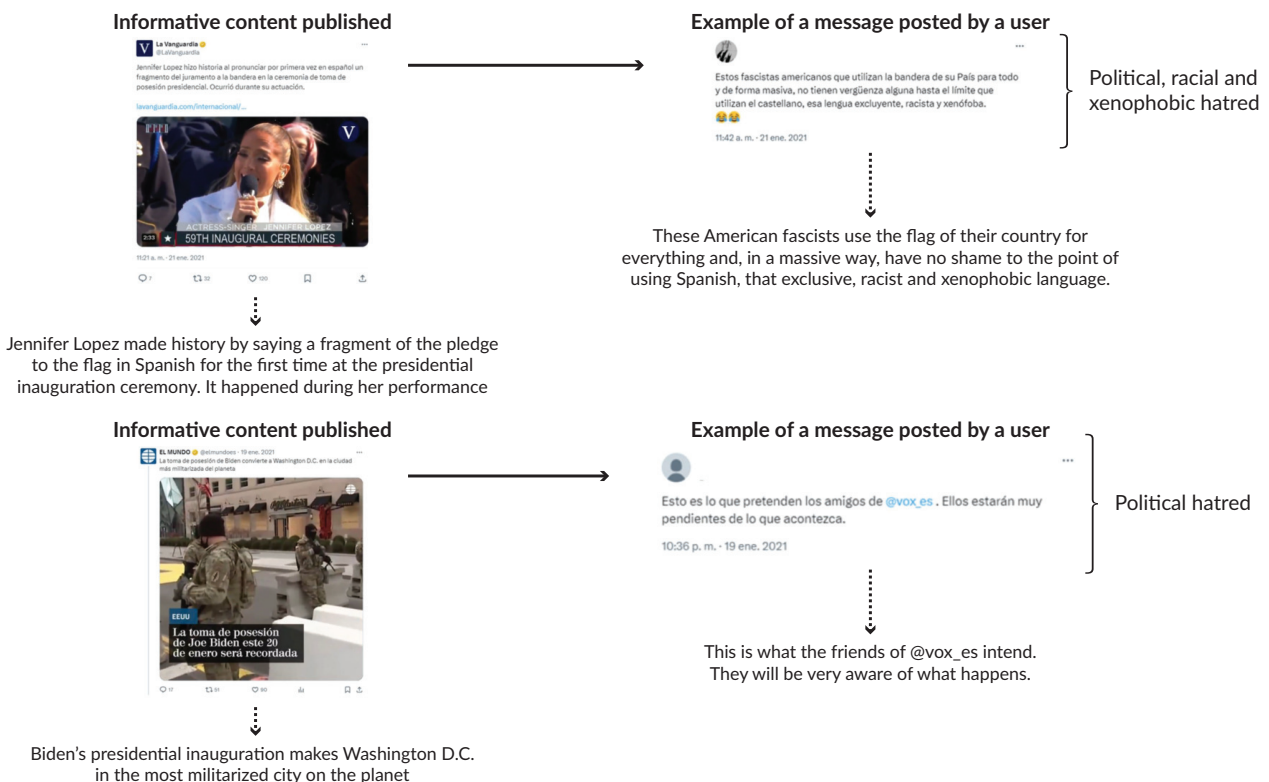


Figure 3. Examples of the types of hate promoted.

oriented toward promoting a hostile social framework aimed at legitimizing narratives laden with prejudices and stereotypes for political, racial, or other reasons against certain national social groups (in Spain) through news published by digital news media regarding Joe Biden's inauguration in the United States. The narrative from the metalanguage used in Figure 3, which seeks to produce semiosis around the connotations, can be interpreted from the image of J. Lo, suggesting that Americans, in addition to being fascists, utilize the flag indiscriminately, analogous to the actions of right-wing parties in Spain, and employ it as a symbolic element to connote affiliation toward policies more aligned with right-wing ideology.

In the case of the image of the military (Figure 3), the objective is to establish a correlation between the ideological policies of the right-wing party VOX and the behavioral affinities of the state's armed forces. Internet users seek to express this political animosity when articulating their opinions about the image of the military.

Among its many manifestations, hate speech predicated on racial, misogynistic, and ideological grounds is particularly salient. Within the semiotic relationship framework, these expressions seek to be established, engendering metalanguages that other users can adopt and utilize. This process of transmission and transformation of meanings is termed semiosis.

There are also narrative axes centered on fear and ideological positions. This is exemplified by the message in Figure 4: "These American fascists who use their country's flag for everything and massively, have no shame to the point of using Spanish, that exclusive, racist and xenophobic language." This statement demonstrates the relationship between exclusionary language and intertextuality and constructs new levels of animosity associated with parallel themes external to the news content provided by the media.

It is worth observing the concentration of hate message recipients around specific figures participating in Joe Biden's inauguration. This phenomenon is exemplified by *El País* and its coverage of Lady Gaga's participation in the event (Figure 4). Messages containing hate expressions exhibit exclusionary language and metalanguage, which are the fundamental elements of semiosis. In the case of exclusionary language, messages were used to exclude others. At the same time, metalanguage was employed using argued scientific language to render it accessible or "proximate" to other users participating in the debate generated from the news. In this instance, semiosis manifests through various representations that users generate (e.g., cultural alienation and patriotism) via messages or comments in response to a news item.

Figures 4 and 5 show several examples. The first pertains to exclusionary messages regarding Lady Gaga, which emphasize that the singer is associated with North American show business, which, from the metalanguage perspective, appears to be valid. This is contrasted with the situation in Mexico, where ceremonies incorporating folklore or Mexican artists are perceived to have diminished authority. The second example displays a reproduction of text from the ABC newspaper, which enumerates the artists participating in the inauguration ceremony. Internet users subsequently employ the images of these women to ascribe the connotations surrounding their figurativization and the factorization of their inclusion in the ceremony as part of the context.

The comment analysis also examined the factorization and figuration of hate messages within the debates generated in the digital news media accounts. Specifically, this refers not only to the capacity of language to

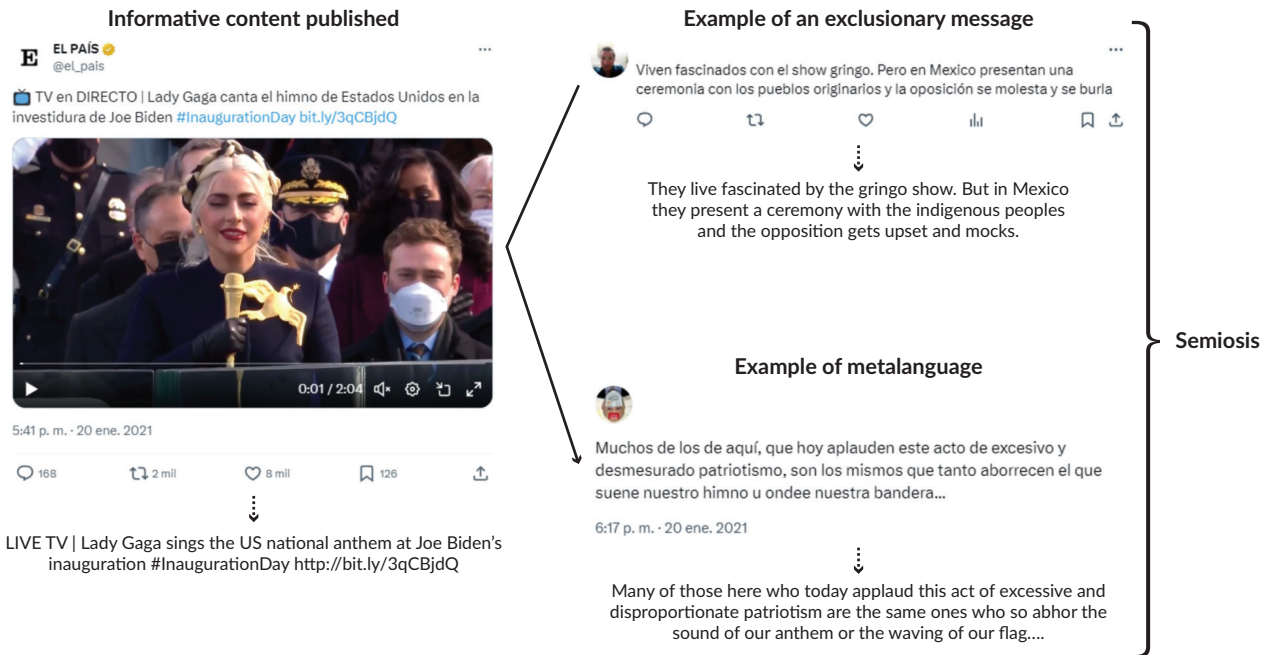


Figure 4. Examples of exclusive language, metalanguage, semiosis.

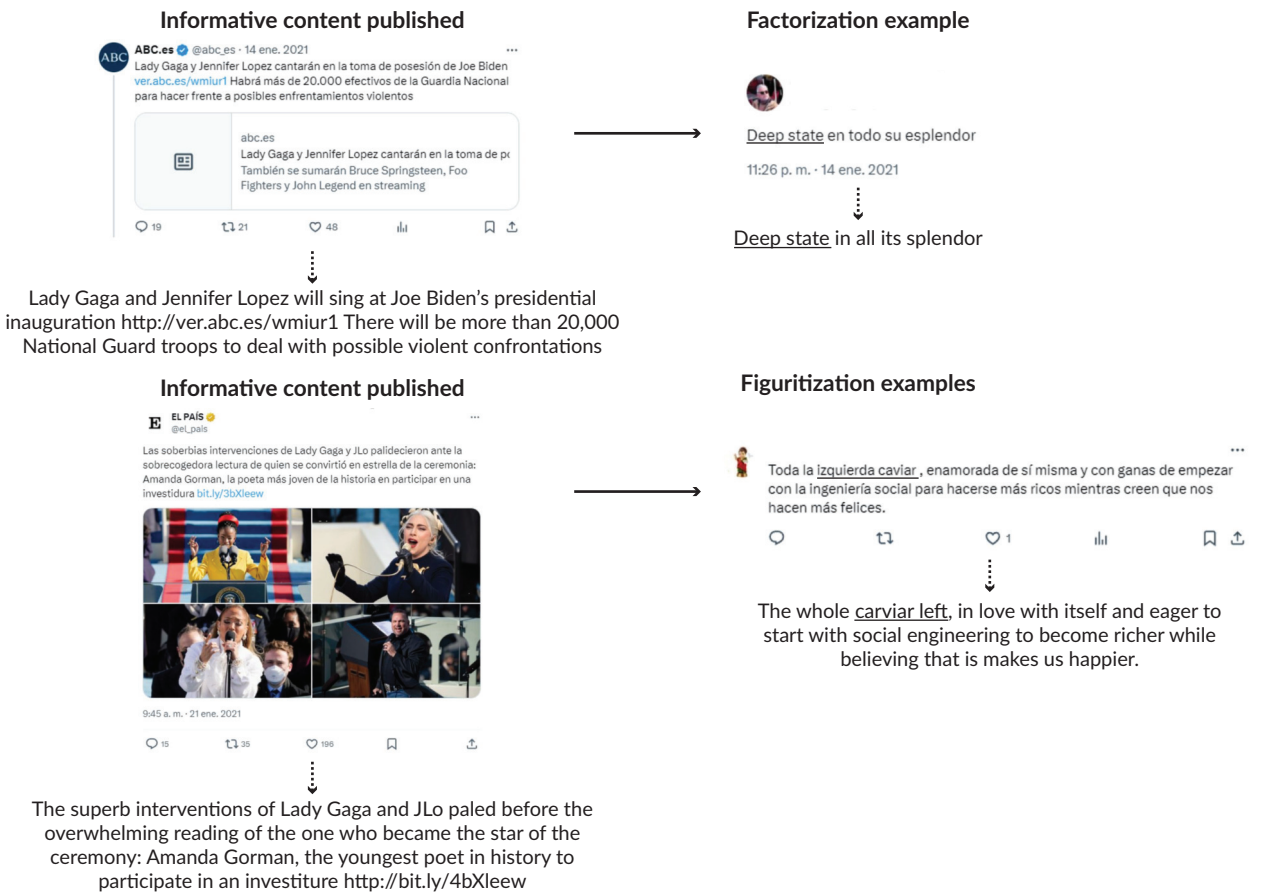


Figure 5. Examples of factorization and figurativization.

alter discourse by providing it with connotations beyond its narrative meaning but also to the attempt to attribute characteristics to individuals based on their social relevance (e.g., president of government). An illustrative example of this phenomenon can be observed in Figures 4 and 5, wherein concepts such as “deep state” and “caviar left” are employed for these purposes. The utilization of these concepts exemplifies how hatred can manifest in diverse forms, generating semiosis around a particular topic. This instance pertains to the ideological themes associated with the examined event that are employed to express hatred directed at figures or protagonists featured in Spanish digital news media content. Factorization and figuration appear to occur on a relatively broad scale, at least within the corpus of the messages analyzed in this research.

4.2. *The Intensity of Hatred, Denotation, Connotation, and Semiosis of the Messages*

Table 4 shows that most of the 2,074 interactions generated around the comments were categorized as “low intensity” hate. This classification was applied to intensity levels 0, 1, and 2 (De Lucas et al., 2022). The social consequence of this typology is, at a minimum, the establishment of a widely accepted negative prejudice against the targeted minority. Although it may not explicitly call for action, it constitutes a fundamentally anti-democratic, discriminatory cultural principle. Moreover, the promotion of a hostile environment toward these discriminated minorities not only fails to prevent but also may lead to tangible social actions that range from effective discrimination to violent acts. Indeed, detected and systematic hate speech campaigns focus primarily on disseminating low-intensity messages (Arce-García et al., 2024). This act could be interpreted as cultivating a “cultural” climate conducive to subsequent discriminatory or violent actions (Figure 6).

Table 4. Interactions generated around messages with detected hate speech.

By intensity of hate		
Exclusionary language (intensity 0, 1, and 2): 177	Violent language (intensity 3): 32	Threatening language (intensity 4 and 5): 5
By denotation		
Discursivization: 273		Actorization: 357
By connotation		
Thematization: 279		Figurativization: 371
By semiosis		
Metalinguage: 129	Destructive: 120	Intertextuality: 174

Figures 6, 7, and 8 present selected examples that illustrate the varying intensities of hatred. One instance pertains to the violent message directed at the poet Amanda Gorman, wherein she is pejoratively referred to as a “girl,” and the term “sheep” is employed to depict her as part of a collective. Another category of language is exclusionary. In one case, this exclusion stems from misogyny—exemplified by the use of purple as a feminist color—or the use of red as a communist color. Additionally, hostile or threatening language is observed concerning the photograph of Biden, where he is explicitly labeled a “genocidal maniac.” A notable example of connotative use derived from denotation is within the figure. This is evident in the out-of-context photographs of Sanders employed as a meme. In the initial instance on the left, Sanders is depicted in isolation wearing wool mittens, whereas, in the subsequent image, he is positioned at a table with a sculpture to convey a connotation of solitude. This visual manipulation augments the humor inherent in Sanders’ posture while simultaneously connoting his isolation following his defeat in the presidential elections.

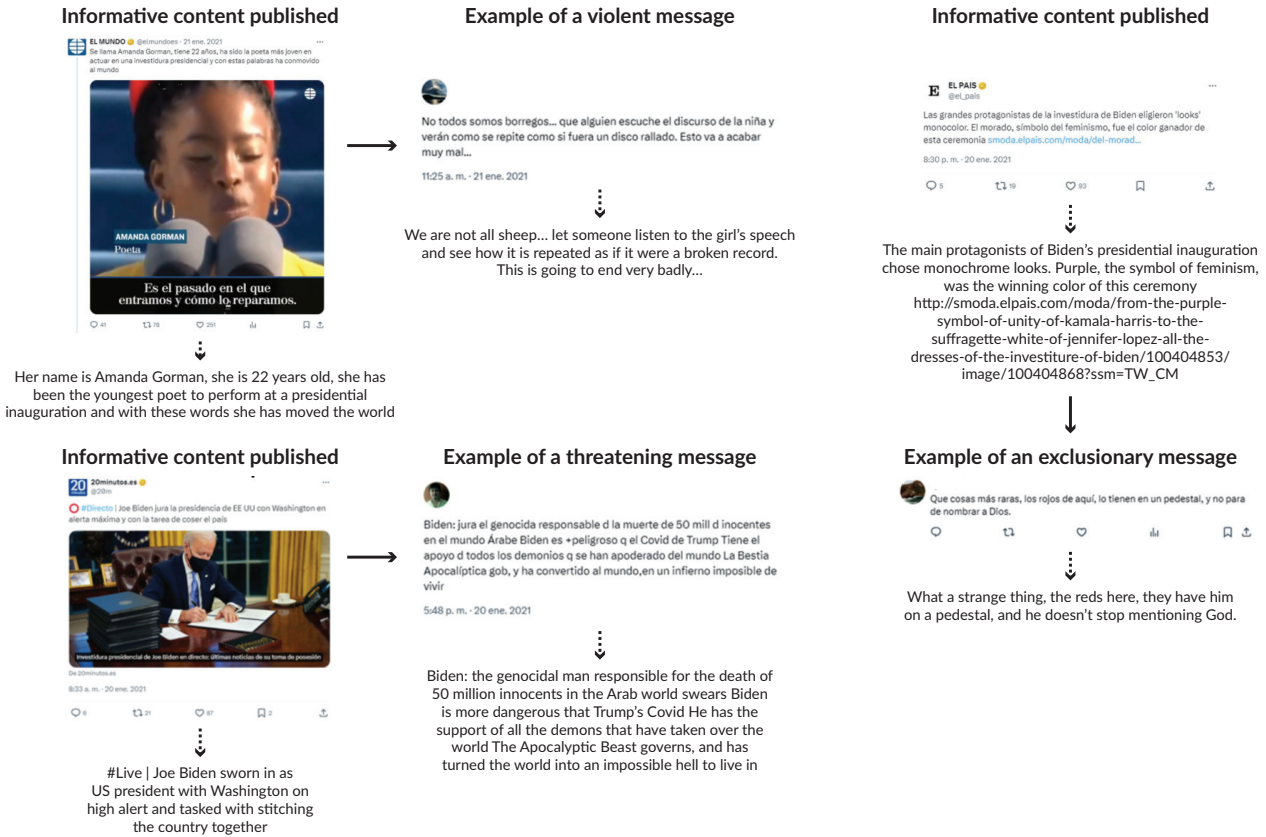


Figure 6. Examples of the interactions with hate speech by intensity.

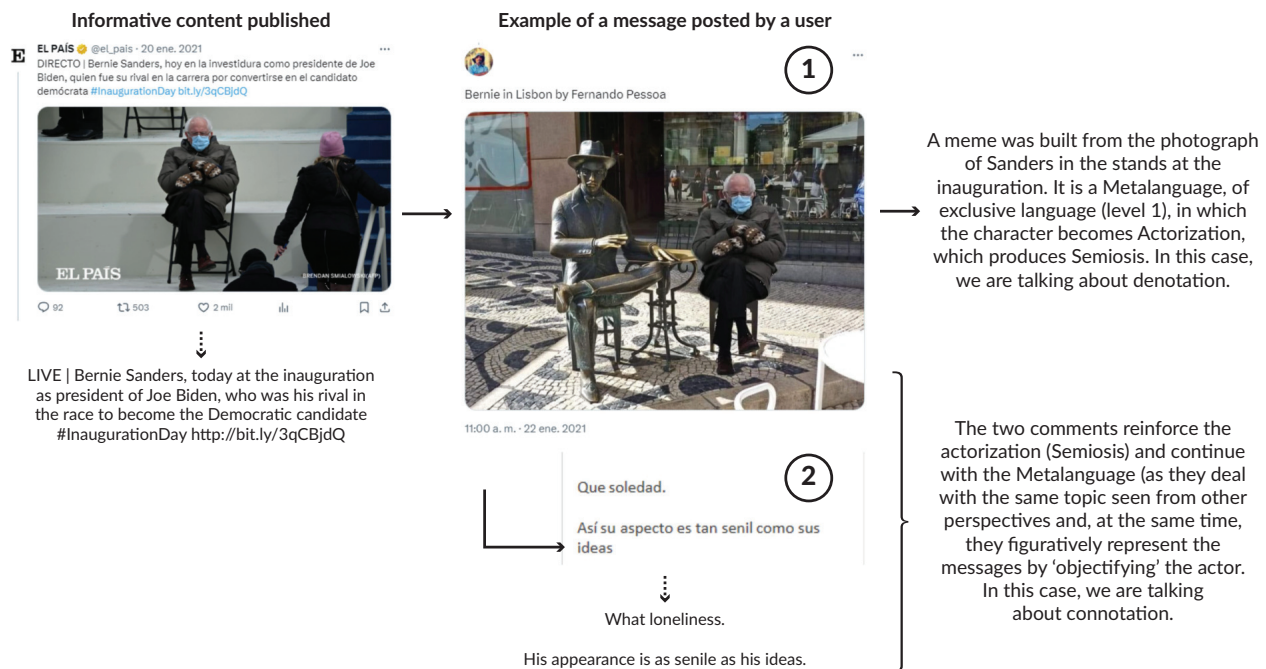


Figure 7. Examples of the interactions with denotation and connotation.



Figure 8. Examples of the interactions with metalanguage, destructiveness, and intertextuality.

Figure 8 is particularly noteworthy, as it addresses metalanguage, destructiveness, and intertextuality. The photograph is significant: It depicts President Biden with the American flag, but the internet user's interpretation elicits multiple analyses because of the metalanguage employed. The user utilizes the phrase "a country eclipsed, polarized and very soon in decline as a World Power." This concise statement elucidates how Obama's former vice president was instrumental in the conflicts during the administration, references the narrow margin of votes between Trump and Biden, and alludes to the decline of a formerly dominant nation. These elements collectively characterize the image as exemplifying metalanguage, destructive language, and intertextuality when juxtaposed with the media outlet's post.

At both the denotative and connotative levels, it is evident that most of the hate speech messages focus predominantly on the actors or protagonists of the news (action and representation), namely, Joe Biden, Jennifer Lopez, Katy Perry, Lady Gaga, or Amanda Gorman, as well as prominent political representatives of the Democratic Party such as Nancy Pelosi, Bernie Sanders, and Barack Obama (Figure 9). This approach eschews discursivization and thematization within the identified hate speech messages. In essence, most hate speech messages employ a strategy that targets the actor or protagonist of the news and their actions within that context (e.g., the role assumed in this event) rather than addressing the speech or messages conveyed during the event. The example presented in Figure 5 demonstrates that messages containing hate expressions appear to be more closely associated with the subject of the news than with the news context itself.

The data presented in Table 4 facilitate the analysis of the semiotic processes in this context. The observed semiosis is destructive and primarily aimed at discrediting the opposing party by utilizing hate-based expressions within the messages. Furthermore, the data revealed an increased presence of intertextuality, or more specifically, the interconnected relationships among messages in a dialogical manner (Figure 5).

4.3. Relationships Among Hatred, Denotation, Connotation, and Semiosis

From the perspective of semiosis, Figure 9 shows the relative coherence and relationship between the types of metalanguage and destructive semiosis proposed by Kristeva (1969). This figure indicates a context wherein, when observing semiosis, metalanguage safeguards the destructive element. Consequently, this necessitates a broader context in which semiosis is generated.

From a denotative and connotative perspective, the characters—Trump and Biden—in their roles as public figures (actorization) and their participation (figurativization) ultimately become recipients of messages containing expressions of hate. The news event (thematization) is relegated to a secondary position. The significant aspect, the inauguration, which constitutes the news event (thematization) itself, assumes a subordinate role in the interactions generated by the commenters (Table 3). In essence, the news event and the topics addressed in the news content published by Spanish digital media on X appear to serve primarily as a vehicle for users propagating hate to promote stereotypes and prejudices (political, racial, or misogynistic) while attacking the protagonists for their role in the event.

Figure 9 shows how hate messages were directed at the Obama period without going beyond the exclusive language aimed at actorization-thematization-figurativization and with a destructive-intertextual approach. The denotative elements (actorization) and connotative elements (thematization and figurativization) can be seen. They are placed in a secondary plane of interest in contrast to the relationship between the intensity of hate and semiosis; this is a result of the intertextuality that arises from the presence of Obama in his double role of former president and Nobel Prize winner. Above all, the users who promote hate speech revive the debate generated by Obama’s presidency and the Nobel Peace Prize using a discursive axis of this type of shared message through the news content published by the examined digital news media in Spain.

Figure 9 shows how hate messages were directed at the Obama period without exceeding the exclusive language oriented toward actorization-thematization-figurativization and by employing a destructive intertextual approach. The denotative elements (actorization) and connotative elements (thematization and figurativization) are evident. They are positioned in a secondary plane of interest relative to the relationship that occurs between the level of hatred intensity and semiosis resulting from the intertextuality arising from Obama’s presence in his dual capacity as former president and Nobel Prize laureate.

Another noteworthy example of these antagonistic relationships with semiosis is presented in one of the messages from users participating in news-driven debates (Figure 9). In this instance, within the narrative

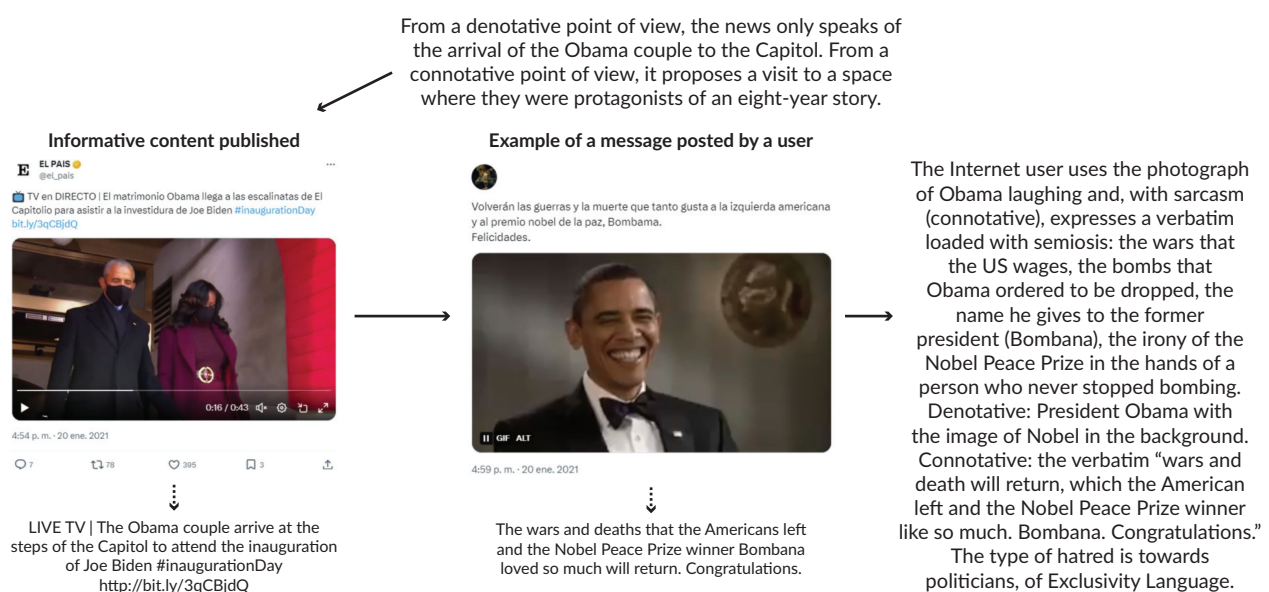


Figure 9. Examples of the relationships among hatred, denotation, connotation, and semiosis.

aspect, we observe discursivization, actorization, and the level of hostility and violent language; in the connotative aspect, there is figurativeness, and in semiosis, intertextuality appears. The message refers to Biden when he served as Vice President of the United States and ordered the bombing of Iraq in 2014, although the internet user employs the figure of Obama as the person ultimately responsible for the American Left. Notably, between the intensity of violent hostility and the semiosis “destructive,” we are directed toward the figurativeness of Obama and his Nobel Peace Prize. This demonstrates how the central figures of the news analyzed in this work become a point of support for semiosis.

A comparable relationship is observed for the news content of the studied media. The example presented, associated with Joe Biden's inauguration, shows the relationships among violent, actor-like, thematic, and destructive hostility.

Former President Obama was not exempt from the interactions, as shown in Figure 9. Furthermore, users who disseminated hate speech attempted to propagate a narrative centered on associating Democrats with the warlike character that they are claimed to possess. This narrative is constructed through a relationship centered on actorization, destructive semiosis, violent expressions of hatred, and thematization.

5. Discussion

The analysis was conducted using a semiotic matrix based on approaches proposed by Greimas and Courtés (1979), Kristeva (1969), Lyotard (1979/2019), and Barthes (1970). When applied to contemporary digital scenarios, such as the one examined in this study, it potentially facilitates an approximation of the analysis of hate expressions disseminated within debates generated by news content published in digital media on X, particularly in Spain. This approach reveals significant semiotic conceptual coexistence. Although these authors predated the emergence of digital communication, their analytical frameworks remain applicable to digital media, particularly the X platform. Their postulates remain relevant as foundational contributions to semiotics in the latter half of the twentieth century. This relevance is particularly evident when their concepts are integrated into the semiotic analysis matrix employed in this study. Concepts such as metalanguage, connotation, and emancipatory narrative facilitate an exploratory approach to messages containing expressions of hostility in environments such as X. This approach extends to the interactions that users generate in response to news content published by digital media outlets in Spain that focus on national and international public interests, which serves as the basis for developing this research.

These findings on hate speech on social media platforms, particularly regarding the discourse generated by digital news media content, have implications for the public sphere and regulatory frameworks, including ethical considerations. The case study presented here demonstrates how specific semiotic indicators can facilitate the identification of patterns that elucidate the mechanisms of toxic content dissemination on platforms such as X. This platform is increasingly recognized as a primary vector for the propagation of hate speech and misinformation. Its influence contributes to establishing narratives that reinforce dichotomous, exclusionary, and ally-enemy paradigms in political discourse (Czopek, 2024).

This digital aspect is crucial for analyzing hate speech and disinformation (Falkenberg et al., 2024). A comparative study of nine countries (Canada, France, Germany, Italy, Poland, Spain, Turkey, the United Kingdom, and the United States) elucidates the bimodal nature of the political polarization promoted in

these countries (left-right) and the significant increase in the toxicity of their messages when disseminated among ideologically opposed groups of users or content. These processes form ideal polarization frameworks centered on displaying uncivil or malicious messages that narratively establish denigration and negative sentiments toward the recipients of their attacks. These vulnerable communities are represented by the protagonists of the content who are used to promote discourses that propagate prejudices and stereotypes and erode democratic values. In this scenario, the media (Spanish in this case) plays a significant ethical and regulatory role to the extent that it can mitigate this type of message. Their intermediation necessitates a different approach from the current one (gatekeeper). This is particularly relevant considering the rise of social media platforms in the dynamics of dissemination, consumption, and debate surrounding news shared in contemporary digital communication spaces (Salonen et al., 2022; Silver et al., 2022).

Several Spanish digital news media outlets (e.g., *La Vanguardia*) and international media organizations (e.g., *The Guardian* or CNN) have opted to discontinue their presence on platforms such as X (“‘La Vanguardia’ dejará,” 2024; Soni & Singh, 2024). The information presented herein can facilitate reflection on the relevance and efficacy of the regulatory mechanisms employed in response to the proliferation of hate speech and disinformation generated in the discussions of their readers. This is particularly pertinent given the evidence of the limited effectiveness of applied regulatory policies, as corroborated by studies conducted by the Center for Countering Digital Hate (2023) and NATO (Bergmanis-Koräts & Haidechyk, 2024). Despite user complaints, both entities indicate a low or nonexistent message deletion rate containing such expressions.

Media organizations should evaluate and enhance moderation mechanisms within various social networks to mitigate the amplification of messages containing hate speech and misinformation that exploit discussions surrounding shared content in these environments. A combination of approaches may be necessary to enforce participation guidelines and restrict or block the dissemination of such messages. One potential strategy involves increasing the visibility of moderators on these platforms. Additionally, implementing artificial intelligence tools may facilitate early detection and informed decision-making in response to the proliferation of messages that exhibit semiotic characteristics similar to those analyzed in this study.

The proposed multidimensional approach to developing semiotic analysis facilitates the deconstruction of its fundamental components and examines how connotative and denotative meanings influence the dissemination of hostile narratives. The findings demonstrate a semiotic approach to analyzing hatred expressed in the content of a global news event across five Spanish digital news media outlets. Regarding OE1, the comments associated with the debate about the news contained a substantial number of verbatim expressions of hate. This situation fosters a hostile media climate. However, it did not promote violence when considering the proportion of words located at intensity levels 0, 1, and 2. These are notably more prevalent than intensities 3, 4, and 5, at least within the parameters established by other authors (Kim, 2022; Walters & Espelage, 2020).

The various dimensions of this promoted hatred (OE2) are characterized by hostile expressions stemming from misogynistic, political, racial, and xenophobic motivations. Predominant narrative strategies employ fear and arguments that accuse, discredit, or devalue the subjects of these news items. This dominant discursive strategy is generated through the debates involving news published by Spanish digital news media on X.

The narrative framework is oriented toward perpetuating stereotypes and prejudices against specific social groups (e.g., democratic politicians and women) based on the predominant semiotic elements (OE3). These elements were identified during the analysis: actorization (at the narrative level), figurativization (at the connotative level), the construction of meaning (semiosis) through metalanguage, the destructive nature of the messages, and the capacity of the messages to establish relationships between other messages in a dialogical manner (intertextuality). Accordingly, this comprehensive analysis contributes to understanding how hate messages are constructed and interpreted on social media platforms (Barth et al., 2023).

At the denotative and connotative levels, it is evident that most of the hate speech messages focus predominantly on the actor or protagonist of the news (actorization and figuration), namely, Joe Biden, Jennifer Lopez, Katy Perry, Lady Gaga, or Amanda Gorman, as well as prominent political representatives of the Democratic Party such as Nancy Pelosi, Bernie Sanders, and Barack Obama (Figure 4, 5, 6, 7, 8, and 9). This strategy eschews discursivization and thematization within identified hate speech messages. That is, most of the hate speech messages analyzed in this study are predicated on a strategy against the actor or protagonist of the news and their actions within that context (e.g., the role assumed within this act) rather than on the discourse or messages that they impart during the event. Consequently, beyond being formal recipients of messages containing expressions of hate, these news actors (the protagonists) serve as narrative resources for the representation of stereotypes and prejudices toward what each of them represents, which originate from their association with specific social groups (e.g., women, Afro-descendant communities, politicians, etc.).

These semiotic elements are significantly interrelated (OE4). This interrelation facilitates the construction of a compelling narrative from the perspective of the authors who have informed the design of the semiotic analysis matrix (Barthes, 1970; Greimas & Courtés, 1976; Kristeva, 1969; Lyotard, 1979/2019). This interrelation demonstrates the capacity of these narratives to create a persuasive context that effectively influences public perception, thereby perpetuating the dissemination of impactful expressions of hate.

This exploratory approach has several implications for communication in digital scenarios. First, it facilitates the recognition of dominant semiotic elements that can contribute to developing debate moderation mechanisms through the linguistic patterns identified in this research (Kleis Nielsen & Ganter, 2018; Perreault, 2023; Salonen et al., 2022). Identifying the underlying semiotic structures (Kahn, 2022; Young-Jung Na, 2023) that enable the propagation of messages containing expressions of hate necessitates considering cultural and social dynamics and current events to identify the linguistic markers of hate (Määttä, 2023; Rajan & Venkatraman, 2021).

The semiotic analysis indicates that it can aid in identifying the elements necessary for the effective management of discussions, both in their iterations on social media platforms and institutional websites, in digital news media (Lin & Kim, 2023). This is particularly significant when considering that the corpus of the messages containing expressions of hate, as identified and analyzed in this study, reveals an underlying semiotic structure characterized by the prevalence of implicit meanings (connotations), the transmission and transformation of meanings (semiosis), and the predominance of exclusionary language. This language aims to promote a narrative that seeks social legitimacy through prejudices and stereotypes and fosters hostility rather than promoting a constructive social framework. Furthermore, the analysis reveals the application of a dominant narrative strategy predicated on the fear, accusation, discrediting, and devaluation of new

protagonists. These elements are utilized as resources to propagate stereotypes and prejudices directed at social groups represented by the protagonists (e.g., women, the LGTBI+ community, the Afro-descendant community, and immigrants). In this content, the news's subject matter appears inconsequential for its dissemination within contemporary digital communication scenarios.

6. Conclusion

This exploratory study provides a method that facilitates the analysis of explicit and implicit content in hate messages on social media platforms. The research was conducted within the limitations of the specific case addressed: Spanish digital news media with a presence on social media platforms and the spaces for debate aimed at their readership.

By applying a semiotic approach, this study provides an approximate framework that facilitates comprehending hate speech construction and dissemination within communication environments fostered by digital news media. This approach enables the identification of the underlying narrative structures, which are addressed in the results and discussion sections presented in this study. Furthermore, the findings may contribute to the development of diverse strategies to counter the proliferation of expressions of hate on social media platforms and institutional web spaces associated with digital news media.

The analysis of messages from the perspective of the semiotic structure inherent in them (Inwood & Zappavigna, 2023), with expressions of hate associated with a news event of particular social relevance in the digital news media in Spain, is a necessary step to develop analysis tools and strategies that detect and counteract the presence of this type of terminology. The specific analyzed case demonstrates a framework in which fear, the utilization of expressions with a clear political orientation, and the use of terms intended to accuse, discredit, or devalue the recipients of such messages ultimately become resources for the dissemination of this content. In this context, the protagonists of the news assume an operational function, as narratives laden with stereotypes and prejudices are promoted through the dehumanization or demonization of what each protagonist represents. The outcome supports the justification or rationalization of hatred from a persuasive context that aims to guide public perception through the debate spaces—in this case, the digital news media on X. This analysis can assist digital news media in reviewing their content moderation practices for the more effective management of discussions, both in their versions on social media platforms and institutional websites (Lin & Kim, 2023). This is particularly significant when considering that the semiotic analysis matrix aids in better assessing the climate of hatred promoted in a context in which hostility dominates more than the call (veiled or not) to violence and discrimination. This necessarily focuses more on control and practices to counteract the messages of “low-intensity hatred” that are the most prevalent and that can potentially render expressions of intense hatred more detrimental in the medium term.

This approach involves the issue highlighted in the discussion: it is imperative to advance research in similar studies whose applicability can be translated into “hybrid” interventions. These interventions could provide resources to professionals who moderate these debate spaces on social networks and equip them with indicators to mitigate the promotion of hostile media environments. This is particularly crucial when external actors exploit the spaces provided by the news media for this purpose. Such guidelines are predicated on the level of hostility intensity exhibited in comments concerning shared news articles and on the categorization of the protagonists involved (a mechanism to propagate fears, prejudices, and stereotypes

that transcends the individual and targets the social groups that they represent based on race, gender, or political orientation).

These moderation practices should be accompanied by additional measures to elucidate the underlying semiotic mechanisms identified in this study that are implemented at various levels of intervention. For example, the establishment of media literacy programs is essential. Educational initiatives can be developed to ensure that readers comprehend the significance of their role in mitigating messages containing hate speech. Furthermore, it is imperative that the professionals responsible for moderating these messages ensure that they possess adequate competencies to recognize the underlying semiotic structures in these messages, including identifying the utilization of exclusionary language, among other skills.

NLP techniques associated with artificial intelligence can serve as a resource for this regulation. These techniques could enhance the role of guarantors in the dynamics of the dissemination, consumption, and debate of news shared by media outlets. This enhancement stems from comprehending cultural and linguistic connections that are seemingly disparate but are associated with the messages (Demuru, 2022; Gramigna, 2022). Another level of practical utility is to aid the social agents (public institutions and third-sector actors) who are invested in combating such expressions on social media platforms within the previously outlined terms. The exploratory nature of this research is evident, and semiotic analysis represents an initial proposal aimed at fostering elements of academic discourse on hate speech from a semiotic perspective. This necessitates the development of further studies and the utilization of novel conceptual approaches according to authors such as Greimas and Courtés (1979), Kristeva (1969), Lyotard (1979/2019), and Barthes (1970), whose relevance persists, given the enduring nature of their contributions to the field of semiotics.

This research addresses a specific geographical and communicative context and is used as a case study. However, its results and discussion have implications beyond the case itself (and are analogous to how knowledge of oceans can be derived from analyzing a single water droplet). The political dimension does not diminish the significance of other factors when the dissemination of hate expressions on social networks is examined. It is imperative to advance comprehensive studies and novel approaches (for example, from a semiotic perspective; see Falkenberg et al., 2024; Kupferschmidt, 2024). Such innovative methods can transcend the analysis of specific studies (from a particular country) and facilitate an understanding of the semiotic structures underlying hate messages in a more global context. One potential avenue is the promotion of studies that investigate the same news event (such as the one considered here) of global interest, which serves as an articulating axis of common or differentiating aspects for each geographical area, communicative context, or temporal period (Määttä, 2023). These efforts will enhance the global reach and potential utility of the findings.

Acknowledgments

The authors thank Possible INC for its support in the extraction of the data that served as a basis for the conducted analysis.

Funding

This article is a result of the Hatemedia project (PID2020-114584GB-I00) funded by MCIN/AEI/10.13039/501100011033 and Project 32BIT financed by Junta de Comunidades de Castilla-La Mancha—Counseling of Education, Culture and Sports and co-financed by the European Union.

Conflict of Interests

The authors declare no conflicts of interest. In this article, editorial decisions were undertaken by Ulf R. Hedetoft (University of Copenhagen, Denmark).

Data availability

The raw data used for the selection and analysis of the case studied in this work can be found at <https://doi.org/10.6084/m9.figshare.25222118.v3>

References

- Alirangues, M. (2018). La palabra precaria. Prolegómenos para una teoría negativa de la ficción. *Escritura e Imagen*, 14, 149–166. <https://doi.org/10.5209/esim.62768>
- Arce-García, S., & Menéndez-Menéndez, M.-I. (2022). Inflaming public debate: A methodology to determine origin and characteristics of hate speech about sexual and gender diversity on Twitter. *El Profesional De La información*, 32(1). <https://doi.org/10.3145/eipi.2023.ene.06>
- Arce-García, S., Said-Hung, E., & Montero-Díaz, J. (2024). Unmasking coordinated hate: Analyzing hate speech on Spanish digital news media. *New Media & Society*. Advance online publication. <https://doi.org/10.1177/14614448241259715>
- Barth, N., Wagner, E., Raab, P., & Wiegärtner, B. (2023). Contextures of hate: Towards a systems theory of hate communication on social media platforms. *The Communication Review*, 26(3), 209–252. <https://doi.org/10.1080/10714421.2023.2208513>
- Barthes, R. (1970). L'ancienne rhétorique. *Communications*, 16, 172–229.
- Bergmanis-Koräts, G., & Haiduchyuk, T. (2024). *Social media manipulation for sale: 2024 Experiment on platform capabilities to detect and counter inauthentic social media engagement*. NATO Strategic Communications Centre of Excellence. <https://stratcomcoe.org/pdfjs/?file=/publications/download/Social-Media-Manipulation-2024-DIGITAL-1eb50.pdf?zoom=page-fit>
- Berlanga-Fernández, I., & Reyes, E. (2022). The digital approach to semiotics: A systematic review. *Text & Talk*, 44(1), 119–140. <https://doi.org/10.1515/text-2021-0073>
- Bianca, E. C. (2021). Semiótica, poder e intolerância: Populismo, direitos humanos e a crise do Estado Democrático de Direito. *Estudos Semióticos*, 17(1), 59–81. <https://doi.org/10.11606/issn.1980-4016.esse.2021.173041>
- Castillo, J. R. (2022). Semiótica, pandemias, COVID-19 y teatro. *Signa: Revista de la Asociación Española de Semiótica*, 31, 27–37. <https://doi.org/10.5944/signa.vol31.2022.32184>
- Center for Countering Digital Hate. (2023). *How Twitter/X continues to host post reported for extreme hate speech*. https://counterhate.com/wp-content/uploads/2023/09/230907-X-Content-Moderation-Report_final_CCDH.pdf
- Chhabra, A., & Vishwakarma, D. K. (2023). Multimodal hate speech detection via multi-scale visual kernels and knowledge distillation architecture. *Engineering Applications of Artificial Intelligence*, 126, Article 106991. <https://doi.org/10.1016/j.engappai.2023.106991>
- Christian, A. J., Day, F., Díaz, M., & Peterson-Salahuddin, C. (2020). Platforming intersectionality: Networked solidarity and the limits of corporate social media. *Social Media + Society*, 6(3), Article 205630512093330. <https://doi.org/10.1177/2056305120933301>
- Cover, R. (2022). Digital hostility: Contemporary crisis, disrupted belonging and self-care practices. *Media International Australia*, 184(1), 79–91. <https://doi.org/10.1177/1329878x221088048>
- Czopek, M. (2024, October 25). 2 weeks, 450+ posts: How Elon Musk uses his X profile to push FEMA,

- immigration, voting falsehoods. Poynter. <https://www.poynter.org/fact-checking/2024/elon-musk-uses-x-account-spread-misinformation>
- De Lucas, A., Römer-Pieretti, M., Izquierdo, D., Montero-Díaz, J., & Said-Hung, E. (2022). *Manual para el etiquetado de mensajes de odio*. Hatemedia Project. <https://doi.org/10.6084/m9.figshare.18316313.v4>
- Demuru, P. (2022). Qanons, anti-vaxxers, and alternative health influencers: A cultural semiotic perspective on the links between conspiracy theories, spirituality, and wellness during the Covid-19 pandemic. *Social Semiotics*, 32(5), 588–605. <https://doi.org/10.1080/10350330.2022.2157170>
- Erdogan-Ozturk, Y., & Isik-Guler, H. (2020). Discourses of exclusion on Twitter in the Turkish context: #ülkemdesuriyeliistemiyorum (#idontwantsyriansinmycountry). *Discourse Context & Media*, 36, Article 100400. <https://doi.org/10.1016/j.dcm.2020.100400>
- Falkenberg, M., Zollo, F., Quattorciocchi, W., Pfeffer, J., & Baronchelli, A. (2024). Patterns of partisan toxicity and engagement reveal the common structure of online political communication across countries. *Nat Commun*, 15, Article 9596. <https://doi.org/10.1038/s41467-024-53868-0>
- Fan, L., Yin, Z., Yu, H., & Gilliland, A. J. (2022). Using machine learning to enhance archival processing of social media archives. *Journal on Computing and Cultural Heritage*, 15(3), 1–23. <https://doi.org/10.1145/3547146>
- Fonseca, A., Pontes, C., Moro, S., Batista, F., Ribeiro, R., Guerra, R., Carvalho, P., Marques, C., & Silva, C. (2024). Analyzing hate speech dynamics on Twitter/X: Insights from conversational data and the impact of user interaction patterns. *Heliyon*, 10(11), Article 32246. <https://doi.org/10.1016/j.heliyon.2024.e32246>
- Ghaffari, S. (2020). Discourses of celebrities on Instagram: Digital femininity, self-representation and hate speech. *Critical Discourse Studies*, 19(2), 161–178. <https://doi.org/10.1080/17405904.2020.1839923>
- Gómez-Camacho, A., Hunt-Gómez, C. I., Núñez-Roman, F., & Esteban, A. N. (2023). Not all motherfuckers are MENA, but most MENA are motherfuckers. *Journal of Language Aggression and Conflict*, 11(2), 256–278. <https://doi.org/10.1075/jlac.00083.gom>
- Gómez-García, S., Paz-Rebollo, M.-A., & Cabeza-San-Deogracias, J. (2021). Newsgames against hate speech in the refugee crisis. *Comunicar*, 29(68), 123–133. <https://www.revistacomunicar.com/ojs/index.php/comunicar/article/view/C67-2021-10>
- Gramigna, R. (2022). Inside Facebook’s semiosphere: How social media influence digital hate and fuel cyber-polarization. *Social Semiotics*, 32(5), 606–633. <https://doi.org/10.1080/10350330.2022.2157171>
- Greimas, A. J., & Courtés, J. (1979). *Semiotics and language*. Indiana University Press.
- Hameleers, M., van der Meer, T., & Vliegthart, R. (2022). Civilized truths, hateful lies? Incivility and hate speech in false information—Evidence from fact-checked statements in the US. *Information, Communication & Society*, 25(11), 1596–1613. <https://doi.org/10.1080/1369118X.2021.1874038>
- Hernández, G. L., Drzewiecka, J. A., & Greco, S. (2023). FortressEurope integrating through division: An actantial narrative analysis. *Journal of Contemporary European Studies*, 31(4), 1437–1451. <https://doi.org/10.1080/14782804.2023.2171968>
- Imbert, P. (2019). Concepts of narrative, founding violence, and multiculturalism in the Americas: Greimas, Girard, and Kymlicka. *Semiotica*, 2019(227), 245–259. <https://doi.org/10.1515/sem-2015-0142>
- Inwood, O., & Zappavigna, M. (2023). Conspiracy theories and white supremacy on YouTube: Exploring affiliation and legitimation strategies in YouTube comments. *Social Media + Society*, 9(1). <https://doi.org/10.1177/20563051221150410>
- Johnson, N. F., Leahy, R., Restrepo, N. J., Velasquez, N., Zheng, M., Manrique, P., Devkota, P., & Wuchty, S. (2019). Hidden resilience and adaptive dynamics of the global online hate ecology. *Nature*, 573(7773), 261–265. <https://doi.org/10.1038/s41586-019-1494-7>
- Kahn, R. (2022). COVID masks as semiotic expressions of hate. *International Journal for the Semiotics of*

- Law/Revue Internationale de Sémiotique Juridique*, 35(6), 2391–2407. <https://doi.org/10.1007/s11196-022-09885-7>
- Kim, T. (2022). Violent political rhetoric on Twitter. *Political Science Research and Methods*, 11(4), 673–695. <https://doi.org/10.1017/psrm.2022.12>
- Kleis Nielsen, R., & Ganter, S. A. (2018). Dealing with digital intermediaries: A case study of the relations between publishers and platforms. *New Media & Society*, 20(4), 1600–1617. <https://doi.org/10.1177/1461444817701318>
- Koltsova, E. A., & Kartashkova, F. I. (2022). Digital communication and multimodal features: Functioning of emoji in interpersonal communication. *RUDN Journal of Language Studies, Semiotics and Semantics*, 13(3), 769–783. <https://doi.org/10.22363/2313-2299-2022-13-3-769-783>
- Kobellarz, J. K., Bročić, M., Graeml, A. R., Silver, D., & Silva, T. H. (2022). Reaching the bubble may not be enough: news media role in online political polarization. *EPJ Data Science*, 11(1), 1–28. <https://doi.org/10.1140/epjds/s13688-022-00357-3>
- Koval, E., Sychev, A., & Zhadunova, N. (2019). Hate rhetoric vs trust rhetoric in contemporary public discourse. *SHS Web of Conferences*, 72, Article 03038. <https://doi.org/10.1051/shsconf/20197203038>
- Kristeva, J. (1969). *Le langage, cet inconnu*. Seuil.
- Kupferschmidt, K. (2024, October 31). A field's dilemmas: Misinformation research has exploded. But scientists are still grappling with fundamental challenges. *Science*. <https://www.science.org/content/article/five-biggest-challenges-facing-misinformation-researchers>
- 'La Vanguardia' dejará de publicar en X, convertida en una red de desinformación. (2024, November 14). *La Vanguardia*. <https://www.lavanguardia.com/vida/20241114/10105193/vanguardia-dejara-publicar-x-convertido-red-desinformacion.html>
- Labiano, R., Azurmendi, A., & Novoa-Jaso, M. F. (2023). The constructive role of the media in hate speech controversies: The Valtònyc and Hasél cases. *Journalism Practice*, 18(2), 452–489. <https://doi.org/10.1080/17512786.2023.2242818>
- Lee, S. Y., Kim, S. J., Lee, H., & Chock, T. M. (2022). Why people became hostile during the COVID-19 pandemic: Exploring the role of social media use, blame attribution, and collective efficacy. *Mass Communication and Society*, 26(4), 619–645. <https://doi.org/10.1080/15205436.2022.2095917>
- Levin-Banchik, L. (2020). Precrisis military hostility and escalation in international crises. *Conflict Management and Peace Science*, 38(1), 63–86. <https://doi.org/10.1177/0738894220906376>
- Lilleker, D., & Pérez-Escobar, M. (2023). Demonizing migrants in contexts of extremism: Analysis of hate speech in UK and Spain. *Politics and Governance*, 11(2). <https://doi.org/10.17645/pag.v11i2.6302>
- Lin, H., & Kim, Y. (2023). Learning from disagreement on social media: The mediating role of like-minded and cross-cutting discussion and the moderating role of fact-checking. *Computers in Human Behavior*, 139, Article 107558. <https://doi.org/10.1016/j.chb.2022.107558>
- Lingiardi, V., Carone, N., Semeraro, G., Musto, C., D'Amico, M., & Brena, S. (2019). Mapping Twitter hate speech towards social and sexual minorities: A lexicon-based approach to semantic content analysis. *Behaviour & Information Technology*, 39(7), 711–721. <https://doi.org/10.1080/0144929X.2019.1607903>
- Lotman, I. (1996). *La semiosfera. Semiótica de la cultura y del texto*. Cátedra.
- Lyotard, J.-F. (2019). *La condición postmoderna. Informe sobre el saber*. Cátedra. (Original work published 1979)
- Määttä, S. K. (2023). Linguistic and discursive properties of hate speech and speech facilitating the expression of hatred. *Internet Pragmatics*, 6(2), 156–172. <https://doi.org/10.1075/ip.00094.maa>
- Miguel, A. V. (2020, June 18). Se debilita la confianza en los medios, resisten las marcas periodísticas y emerge el periodismo local. *Digital News Report*. <https://incom.uab.cat/portacom/digital-news-report->

es-2020-los-medios-afrontan-los-retos-de-recuperar-una-confianza-debilitada-y-seguir-ampliando-
ingresos-por-suscripciones

- Navarrete, M. T. N. (2019). "Se dice poeta": Poesía española, mujer y nuevas tecnologías. *Signa: Revista de la Asociación Española de Semiótica*, 28, Article 245. <https://doi.org/10.5944/signa.vol28.2019.25048>
- Nöth, W. (2011). Self-referential postmodernity. *Semiotica*, 183(1), 199–217. <https://doi.org/10.1515/semi.2011.010>
- Perera, S., Meedin, N., Caldera, M., Perera, I., & Ahangama, S. (2023). A comparative study of the characteristics of hate speech propagators and their behaviours over Twitter social media platform. *Heliyon*, 9(8), Article 19097. <https://doi.org/10.1016/j.heliyon.2023.e19097>
- Perreault, G. (2023). Digital hate: Normalization in management of online hostility. In S. A. Eldridge, J. A. Swart, S. Banjac, & D. Cheruiyoy (Eds.), *Routledge companion to digital journalism* (pp. 406–416). Routledge.
- Phadke, S., Lloyd, J., Hawdon, J., Samory, M., & Mitra, T. (2018). *Framing hate with hate frames: Designing the codebook* [Poster presentation]. ACM Conference on Computer Supported Cooperative Work and Social Computing, New Jersey, USA. <https://doi.org/10.1145/3272973.3274055>
- Rajan, B., & Venkatraman, S. (2021). Insta-hate: An exploration of Islamophobia and right-wing nationalism on Instagram amidst the COVID-19 pandemic in India. *Journal of Arab & Muslim Media Research*, 14(1), 71–91. https://doi.org/10.1386/jammr_00025_1
- Retta, M. (2023). A pragmatic and discourse analysis of hate words on social media. *Internet Pragmatics*, 6(2), 197–218. <https://doi.org/10.1075/ip.00096.ret>
- Riquelme, F., Rivera, D., & Serrano, B. (2022). Analyzing the far-right political action on Twitter: The Chilean constituent process. *Social Network Analysis and Mining*, 12(1), Article 161. <https://doi.org/10.1007/s13278-022-00990-w>
- Rubio-Pinilla, P., & Candón-Mena, J. (2021). Comunicación digital de Médicos Sin Fronteras en España durante la crisis de la COVID-19. *Index Comunicación*, 11(2), 231–255. <https://doi.org/10.33732/ixc/11/02lacomu>
- Ruíz-Iniesta, A., Blanco-Valencia, X., Pérez, D., De Gregorio-Vicente, O., Cubillas-Mercado, J., Montero-Díaz, J., & Said-Hung, E. (2024). *Informe de limpieza de datos brutos obtenidos en medios informativos digitales españoles en X (Twitter), Facebook y portales web*. Hatemedia Project. <https://doi.org/10.6084/m9.figshare.25293226>
- Salaverría, R., & Martínez-Costa, M. (2023). Digital journalism in Spain: Technological, sociopolitical and economic factors as drivers of media evolution. *Journalism*, 25(5), 1050–1069. <https://doi.org/10.1177/14648849231170519>
- Salonen, M., Olbertz-Siitonen, M., Uskali, T., & Laaksonen, S. (2022). Conversational gatekeeping—Social interactional practices of post-publication gatekeeping on newspapers' Facebook pages. *Journalism Practice*, 17(9), 2053–2077. <https://doi.org/10.1080/17512786.2022.2034520>
- Sanders, T., Du Plessis, C., Mullens, A. B., & Brömdal, A. (2023). Navigating detransition borders: An exploration of social media narratives. *Archives of Sexual Behavior*, 52(3), 1061–1072. <https://doi.org/10.1007/s10508-023-02556-z>
- Schäfer, S., Sülflow, M., & Reiners, L. (2022). Hate speech as an indicator for the state of the society: Effects of hateful user comments on perceived social dynamics. *Journal of Media Psychology: Theories, Methods, and Applications*, 34(1), 3–15. <https://doi.org/10.1027/1864-1105/a000294>
- Silver, N., Kierstead, E., Kostygina, G., Tran, H., Briggs, J., Emery, S., & Schillo, B. (2022). The influence of provaping "gatewatchers" on the dissemination of COVID-19 misinformation on Twitter: Analysis of Twitter discourse regarding nicotine and the COVID-19 pandemic. *Journal of Medical Internet Research*, 24(9), Article 40331. <https://doi.org/10.2196/40331>
- Soni, A., & Singh, J. (2024, November 14). Bluesky attracts millions as users leave Musk's X after Trump win.

- Reuters. <https://www.reuters.com/technology/hate-speech-watchdog-ccd-h-quit-musks-x-ahead-terms-change-2024-11-14>
- Statista. (2020a). Ranking de los 10 medios de comunicación con más seguidores en Twitter en España en abril de 2020. *Statista*. <https://es.statista.com/estadisticas/518802/twitter-perfiles-de-medios-de-comunicacion-con-mas-seguidos-en-espana>
- Statista. (2020b). Ranking de las principales marcas de medios de comunicación online según el porcentaje de población que las usaba de forma semanal en España en 2020. *Statista*. <https://es.statista.com/estadisticas/670743/principales-medios-online-porcentaje-de-usuarios-semanales-espana>
- Tontodimamma, A., Fontanella, L., Anzani, S., & Basile, V. (2022). An Italian lexical resource for incivility detection in online discourses. *Quality & Quantity*, 57(4), 3019–3037. <https://doi.org/10.1007/s11135-022-01494-7>
- Vasist, P. N., Chatterjee, D., & Krishnan, S. (2023). The polarizing impact of political disinformation and hate speech: A cross-country configural narrative. *Information Systems Frontiers*, 26, 663–688. <https://doi.org/10.1007/s10796-023-10390-w>
- Walters, G. D., & Espelage, D. L. (2020). Hostility, anger, and dominance as mediators of the sibling aggression–school fighting relationship: Mechanisms of violence generalization. *Psychology of Violence*, 10(1), 48–57. <https://doi.org/10.1037/vio0000227>
- Young-Jung Na, T. (2023). Hate speech and institutionalized society. *Sookmyung Research Institute of Humanities*, 14, 67–83. <https://doi.org/10.37123/th.2023.14.67>

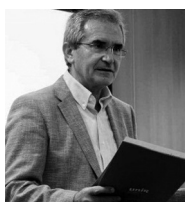
About the Authors



Max Römer-Pieretti is a full professor and researcher at Universidad Camilo José Cela, where he is the research coordinator of the Faculty of Communication and Design. His research interests include semiotics, politics, and hate studies. With more than 35 years of teaching and research experience, his teaching background includes undergraduate, graduate, and doctoral classes.



Elías Said-Hung is a professor, researcher, and the principal investigator of the research group Socio-Educational and Intercultural Inclusion, Society and Media (SIMI). He is the former director of the master's degree program in inclusive and intercultural education (2018–2024) and director of the *Spanish Journal of Pedagogy*, based at the International University of La Rioja (UNIR). He is also the co-principal investigator of the Hatemedia project and has authored over 100 scientific publications pertaining to disinformation, hate speech, and the sociology of communication and education.



Julio Montero-Díaz is a university professor in the field of communication. He directed the doctoral program of the Department of History of Communication at the Complutense University of Madrid and the UNIR Doctoral School, where he serves as vice-rector for research. He has supervised 22 doctoral theses. He is the author of approximately twenty books and one hundred academic articles and book chapters. Over the past fifteen years, he has chaired several university degree verification committees at ANECA and the Madri+d Foundation. Since July 2024, he has held the position of Chair of the Villanueva University Foundation.

Propagation of Hate Speech on Social Network X: Trends and Approaches

Eva Matarín Rodríguez-Peral ¹ , Tomás Gómez Franco ² ,
and Daniel Rodríguez-Peral Bustos ³ 

¹ Department of Communication Sciences and Sociology, Rey Juan Carlos University, Spain

² Department of Economics, Francisco Vitoria University, Spain

³ Faculty of Information Sciences, Complutense of Madrid University, Spain

Correspondence: Eva Matarín Rodríguez-Peral (eva.matarin@urjc.es)

Submitted: 16th September 2024 **Accepted:** 17th December 2024 **Published:** 23 January 2025

Issue: This article is part of the issue “Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment” edited by Max Römer Pieretti (Universidad Camilo José Cela), Beatriz Esteban-Ramiro (Universidad de Castilla-La Mancha), and Agrivalca Canelón (Universidad Católica Andrés Bello), fully open access at <https://doi.org/10.17645/si.i415>

Abstract

Digital technologies have democratized the transmission of information, enabling individuals to interact and share information instantly through social networks. However, these advancements have also brought about negative aspects such as the propagation of hate speech on social media. This research aims to address the following question: What are the predominant theoretical and methodological approaches in academic research on hate speech on X (formerly known as Twitter)? This study aims to identify and analyze the trends in existing academic research on the proliferation and dissemination of hate speech on the social network X, to provide a comprehensive overview of the current state of knowledge in this field, and to highlight areas for future research. To conduct this analysis, a mixed-methods methodology is employed and a systematic literature review is applied as the research technique. Quantitative analysis involves descriptive statistical analysis, while qualitative analysis is conducted using a deductive strategy to study the predetermined categories of research included in this study. Among the main contributions is the integration of findings from multiple studies, facilitating the understanding of this phenomenon, as well as enabling the identification of best practices and existing knowledge gaps in this field.

Keywords

academic analysis; digital interaction; hate speech; social conflict; social network X; Twitter

1. Introduction

Medium theory posits that media not only transmits information but also shapes the interactions that occur within them, fostering certain types of interaction while limiting the possibilities of others (Meyrowitz, 1994). The advent of social media as platforms enabling user interaction has transformed the way individuals communicate. In light of these new forms of interaction, the European Union Agency for Fundamental Rights (FRA, 2023a) notes that these platforms have the power to amplify the discourses produced within them, including hate speech.

An example of this is the assertion by Oboler (2008), who points out that technological changes have given rise to what he has termed “online antisemitism” or “antisemitism 2.0,” a new model of social antisemitism that emerges with the advent of online social networks, where conspiracy theories and Holocaust denial messages are shared. In this context, user profiles and behaviors vary depending on the platform used. Researcher Ott (2016, p. 60) asserts that platform X, formerly known as Twitter, influences user behavior by fostering “disdain for others, thereby promoting a mean-spirited and malicious discourse.”

Citing the right to freedom of expression, various individuals disseminate discourse targeting other social groups (Bustos Martínez et al., 2019). These discourses are commonly referred to as hate speech because they incite violence and discrimination against the identity of a group of people. Hate speech is a complex phenomenon with multiple dimensions that can trigger dangerous consequences in democratic societies, potentially increasing levels of violence and crime, and even leading to wars and genocidal persecutions linked to group identity (Committee of Ministers, 2022).

The increase in such rhetoric has been so significant that, in 2019, the United Nations introduced a strategy and plan of action on hate speech to support states in their efforts to combat these discourses while simultaneously respecting freedom of expression. Furthermore, in 2021, the United Nations proclaimed June 18th as the International Day for Countering Hate Speech (United Nations, 2024).

The concept of hate speech itself is inherently imprecise, creating difficulties in determining what constitutes hate and what does not, depending on its levels of intensity (Benesch et al., 2021). This has led to legislative challenges, as there is currently no global consensus or universal definition of hate speech (UNESCO, 2021). This is precisely due to the cultural characteristics and values associated with each society, which differentiate them from one another. A hate crime is understood as a crime motivated by prejudice against a specific group—when an individual is intentionally attacked for traits linked to their identity (OSCE, 2014). Hate speech itself can increase prejudices and stereotypes towards a group and lead to an increase in hate crimes (Schäfer et al., 2024).

The propagation of hate speech through digital channels such as social media is becoming a global phenomenon that affects all societies (United Nations, 2024). As noted by the United Nations (2019), hate affects all societies broadly, regardless of whether they are more liberal or authoritarian. Furthermore, it poses a threat and a challenge to democratic states and their peaceful coexistence (Martínez Valerio, 2022). In the first quarter of 2021 alone, YouTube removed 85.247 videos, Facebook reported a total of 25.2 million pieces of content, Instagram 6.3 million, and Twitter, between July and December 2020, deleted 1.628.281 messages that violated their hate speech policies (UNESCO, 2021).

For this reason, the United Nations has developed the following definition, which aims to have a broad reach and a social consensus: “Any type of communication, whether oral or written, or behavior, that attacks or uses pejorative or discriminatory language in reference to a person or group based on who they are, in other words, based on their religion, ethnicity, nationality, race, color, ancestry, gender, or other forms of identity” (United Nations, 2022).

Musa Gassama, representative of the OHCHR in the Human Rights Division of the United Nations Multidimensional Integrated Stabilization Mission in the Central African Republic (MINUSCA), states that “hate messages disseminated through traditional media and the Internet have the peculiarity of generating physical and psychological violence in individuals and social groups” (OHCHR, 2019). Gassama emphasizes the need to prevent the proliferation of hate speech by addressing it at the earliest signs. He notes that in 2018, MINUSCA identified 44 press articles containing hate speech, which were disseminated by 14 media outlets in the Central African Republic (OHCHR, 2019).

According to the FBI (2023), 10,840 hate crimes were reported in 2021, resulting in 12,411 victims. The majority of these crimes, 64.5%, were motivated by the victim’s ethnicity. Additionally, 15.9% of the crimes were due to sexual orientation, 14.1% to religion, 3.2% to gender identity, 1.4% to disability status, and 1% to gender (US Department of Justice, 2023).

OSCE, the Organization for Security and Co-Operation in Europe, the largest security organization comprising states from Europe, Central Asia, and North America, has compiled a report gathering information on hate crimes in 47 states. According to this report, 9,891 hate crimes were recorded in 2023 (OSCE, 2024). However, the organization warns that many hate crimes remain hidden under other crime categories. Similar to the findings of the FBI, the OSCE highlights that racially and xenophobically motivated hate crimes are the most prevalent.

Regarding the countries of the European Union, member states are required to combat hate crimes. However, not all of them explicitly define hate crimes in their penal codes. Each country can establish different sanctions. Some impose harsher penalties when a crime is aggravated by hatred towards the victim’s identity. Nevertheless, due to the significance of this phenomenon at the European level, the European Union has established a high-level group to combat these crimes and the discourses that motivate them. The FRA participates in this group, aiming to promote coexistence (FRA, 2023b).

In Spain, between 2019 and 2023, hate crimes increased by 24.71%, linked to incidents such as anti-Gypsyism, anti-Semitism, aporophobia, religious beliefs, discrimination against people with disabilities, age-related discrimination, illness, gender, ideology, sexual orientation, or gender identity (Ministerio del Interior, 2024). State authorities have promoted various projects to train relevant bodies in the detection, analysis, and evaluation of hate speech, as well as to promote counter-narrative strategies, such as the European Real-UP project in which Spain participates (Ministerio del Interior, 2024). To address the challenge of reducing these types of crimes, police initiatives have focused on preventing and tackling the propagation of hate speech in the digital space.

Preventing social conflicts is a fundamental goal of societies. In this context, entities such as the Online Hate Prevention Institute (OHPI) in Australia have emerged globally, dedicated to eradicating online hate and

empowering society to combat it. This is because online hate has the capacity to become normalized, leading individuals to perceive it as acceptable, which subsequently affects real-world coexistence (Oboler, 2022).

This study aims to contribute to the curbing of hate discourse propagation. With this goal in mind, we propose to increase knowledge about the ongoing lines of research on this topic by examining the scientific production related to hate speech on social networks, given the proliferation of its dissemination through these channels.

In April 2023, 4.8 billion user identities were detected on social media platforms. Although this figure does not correspond exactly to 4.8 billion individuals, it serves to illustrate the magnitude of digital social networks, demonstrating that the majority of internet users engage with these platforms. This figure represents 60% of the global population.

In Spain, only 8% of internet users do not use social networks (IAB Spain, 2023). These have become a common channel for communication and social interaction in daily life, with users of practically all ages, including young people, adults, and older individuals. Currently, within the age range of 12 to 74 years of age, 86% of internet users use social networks (IAB Spain, 2023), a figure that has been increasing in recent years. However, social networks have a higher penetration in the age groups between 18 to 24 years of age (94%) and between 35 to 44 years of age (91%; see IAB Spain, 2023). Additionally, young people between the ages of 18 and 34 are the most active users, as they use an average of six different social networks (IAB Spain, 2023).

Given the everyday use of social media by the global population, we consider it essential to examine platforms such as Twitter, which have a significant impact on the dissemination of discourse and social interaction. As Oboler (2008) points out, it is necessary to combat this type of discourse through knowledge. In this study we seek to identify and analyze academic trends that address the propagation of hate speech on the social network X. In 2023, Twitter.com emerged as one of the most visited websites globally, attracting 2.3 billion visitors (Kemp, 2023). This social network, which has undergone multiple changes in its name, logo, and ownership since that year, stands out not only for its large user base but also for its structure, which enables public and instantaneous interaction. This feature facilitates the virality of messages and the spread of discourse among users, even those who are not personally acquainted.

Moreover, Twitter has garnered significant attention from various media outlets, such as the BBC (Wendling, 2023) and the *New York Times* (Frenkel & Conger, 2022), which have underscored the increasing scrutiny of the platform as a medium that facilitates the propagation of hate speech. In this context, scholars such as Miller et al. (2023) have observed that since October 2022, when Elon Musk acquired Twitter, there has been a 106% average weekly increase in antisemitic hate speech in English, with 325,739 tweets identified over a nine-month period. Similarly, Amores et al. (2021) indicate that it is advisable to analyze the propagation of hate speech on X, as they consider that its increase may be related to a rise in hate crimes.

In light of the aforementioned points, this study poses the following questions: What are the predominant methodological trends in academic research on hate speech on X? What is the scope of academic articles that address these issues? What effects do hate speeches disseminated on X have? What characteristics do the most impactful articles in the academic field present?

In summary, this research is conducted to synthesize the most recent studies on hate speech and highlight the current state of knowledge on this social phenomenon, with the aim of contributing to social reflection on the development of public policies to address this issue. Additionally, it seeks to show the level of dissemination of academic research that addresses this topic.

1.1. State of the Art

Delving into the study of the first publications that became part of the WoS database in 2016, it is observed that even in those early articles, there was concern about curbing the spread of these discourses. Authors Burnap and Williams (2016) focused on enabling the automatic classification of new and diverse types of hate speech spreading online, referred to at that time as cyberhate. They considered that several types of hate were beginning to intersect simultaneously. In their work, the predictive value of the labels associated with each class within the model was significant. However, despite enabling improvements in the prediction provided by their model, further enhancements were still required for it to be more generalizable, as it performed well in specific events, but its capacity diminished outside of these.

An example of intersectionality in hate speech can be clearly observed in Seijbel et al.'s (2023) article, where the Covid-19 pandemic is addressed as a period during which anti-vaccine discourses, denialism, and insults during confinement emerged. Added to this was the high level of connection to digital platforms by the population. On the other hand, the article addresses anti-semitism and sports, a space where, despite the values promoted, hate speech is common, mainly during matches and, afterward, on social networks.

In research focused on detecting hate speech, it is crucial to situate the concern of NGOs that observe the direct impact these discourses have on the population, as well as the work of researchers like Pereira-Kohatsu et al. (2019), who have focused their research on monitoring these discourses on social networks and understanding their evolution. Other authors are interested in the role of X users themselves as protectors of their data and privacy, and how interactions on social networks are conditioned by the anonymity of their users (Williams et al., 2017). The digital space has a disinhibiting effect that influences the way we relate to each other, sometimes blending public and private spaces (Williams et al., 2017). It is at this moment that social network users lower their privacy protection barriers and leave information that belongs to the private space in a public space like social networks, without perceiving the exposure to any risk.

From another perspective, but also linked to the change in behavior and values of individuals, Matamoros-Fernández (2017) points out that this change in social interaction fostered by the digital environment has produced a new form of racism and discrimination: platform racism. Matamoros-Fernández (2017, p. 2) defines this as racist discourse amplified by platforms such as Facebook, Twitter, etc.

It is precisely the new behavioral and interaction habits inherent to social networks that have facilitated its emergence. Publications like Matamoros-Fernández's (2017), on discourses promoting racism on social networks, are linked to one of the areas of hate speech research that focuses on the victims of these discourses. Similarly, authors Rodríguez-Sánchez et al. (2020) show their interest in delving into sexist discourse and the various subtle ways it is expressed.

Additionally, Matamoros-Fernández (2017) also focuses on the instruments that facilitate these discourses. In this regard, she has studied how humor has been used to disguise biases that have contributed to

promoting racial hatred on social networks. In line with this idea, Paz et al. (2021) investigate how political memes, instead of contributing new ideas humorously, actually reproduce discriminatory, biased expressions, and disqualifications that hinder coexistence and benefit political polarization.

Determining the elements that hinder the curbing of hate speech is a relevant aspect. In this regard, some research addresses the impact on researchers of the restrictions on access to their application programming interfaces (API) that platforms like Facebook and X have imposed, as well as the social implications this has in terms of misinformation, polarization, the propagation of fake news, and the promotion of discourses that put democracies and coexistence at risk (Bruns, 2019).

Currently, another burgeoning line of research is analyzing the effectiveness of counter-speech as a tool to combat hate speech (Schäfer et al., 2024). In this regard, Baider (2023) points out the ineffectiveness of existing measures against these discourses, highlighting the fact that most of these attacks are covert, making them difficult to counteract. Similarly, in line with Baider (2023), Hangartner et al. (2021) notes that despite the apparent momentum of counter-speech, there is currently a lack of data that can provide empirical evidence of its effectiveness.

2. Methods

The objective of this research is to understand the academic trends related to the analysis of hate speech and to determine its evolution and the specialization of researchers. This study employs a systematic literature review and bibliometrics analysis (Velt et al., 2020). Following the planned phases (Dekkers et al., 2019), this systematic literature review starts from the following research question: What are the trends in academic research on the dissemination of hate speech through social networks?

For its preparation, the criteria of the PRISMA statement (Rodríguez-Izquierdo & García Bayón, 2024) were taken into account, and a Boolean search was conducted in the Web of Science (WoS) database at the end of August 2024 (in the core collection and within the Social Sciences Citation Index [SSCI]; see Römer et al., 2023). The reasons for selecting this database are manifold, but the most notable is its recognized prestige in the academic field, as well as the high-quality indices and peer review of its articles. Additionally, its international scope and the ease it offers for obtaining and downloading information were considered, enabling the necessary filtering to meet the inclusion and exclusion criteria (Vicent et al., 2020).

Furthermore, WoS is one of the oldest research publication databases in the world. Known as the Science Citation Index since its inception in 1964, it was combined in 1997 with the SSCI and the Arts & Humanities Citation Index (A&HCI) to form what we now know as WoS (Birkle et al., 2020). This extensive history grants it one of the longest and most established records compared to similar databases, endowing it with notable experience and rigor. Additionally, WoS offers comprehensive coverage in the social sciences and provides valuable information for conducting bibliometric analyses.

Another important aspect is that WoS includes journals from the Emerging Sources Citation Index (ESCI), which, although they do not possess the high impact required for indexing in the main WoS indices, or other databases such as Scopus, are nonetheless relevant and of interest to the academic community.

The descriptors used for the search in English are as follows: hate speech (all fields) and X (topic), and Twitter (topic), which resulted in the following search equation: (ALL = (hate speech) AND TS = (X)) AND (OA = = ("OPEN ACCESS") AND DT = = ("ARTICLE")) OR (ALL = (hate speech) AND TS = (Twitter)) AND (OA = = ("OPEN ACCESS") AND DT = = ("ARTICLE")).

Inclusion criteria required studies to be (a) open access articles, (b) available in any language, (c) unrestricted in time, and (d) that they address the dissemination of hate speech on the social network Twitter or X. The exclusion criteria are: (a) that the type of document is different from an article; (b) that it is not open access; (c) that the topic does not jointly address the dissemination of hate speech and the social network X; (d) that there are duplications.

In an initial search, 172 articles matched the descriptors. By filtering the search according to the inclusion and exclusion criteria: open access and article type, a total of 98 articles were obtained that meet the established inclusion and exclusion criteria. Figure 1 shows a flow diagram that visualizes the selection process of the final sample of analyzed articles.

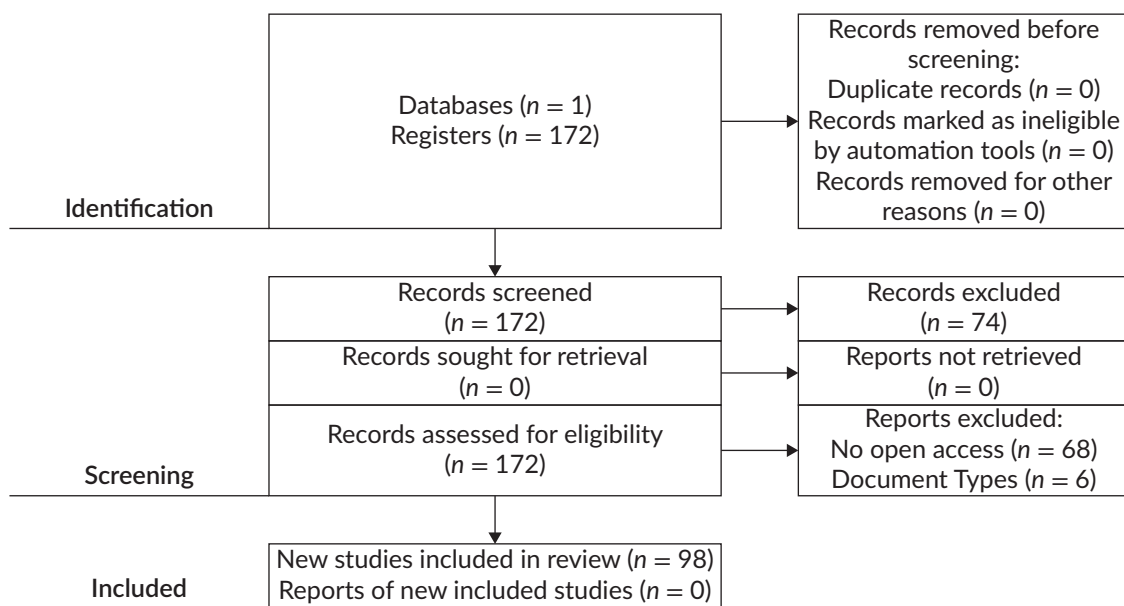


Figure 1. Identification of the new studies via databases and registers: Flow diagram. Source: Based on Haddaway et al. (2022).

3. Analysis of Results

3.1. Bibliometric Analysis

Firstly, it has been considered relevant to understand the interest this topic generates in the academic field. To this end, the evolution of both the number of articles published over time and the number of citations received each year by articles on hate speech has been studied. Thus, Figure 2 shows the number of articles published in the WoS database each year, and the secondary axis of the graph refers to the citations.

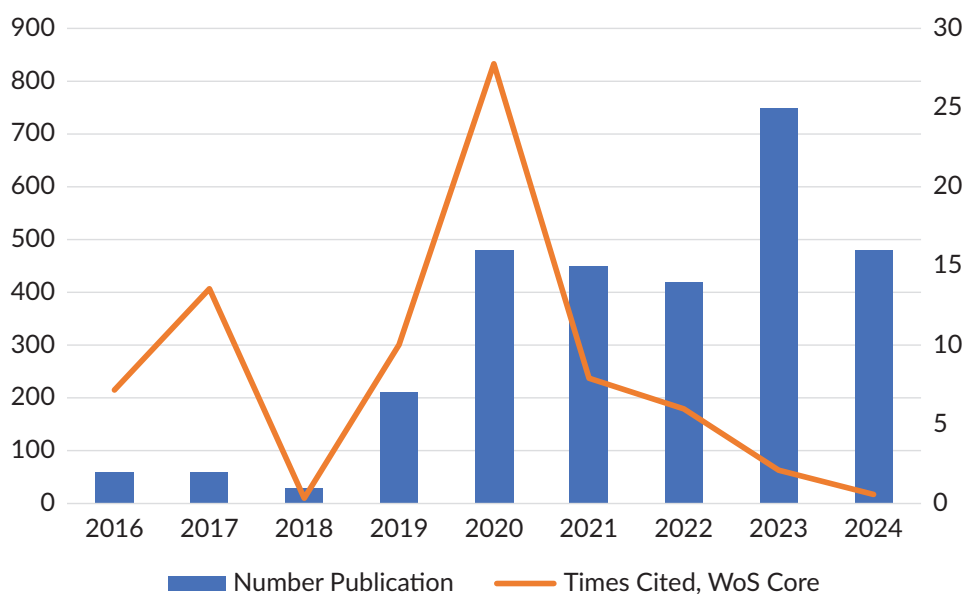


Figure 2. Evolution of the number of articles and citations on hate speech in the WoS database.

Interest in this topic has continued to grow since 2016 (Figure 2), the first year in which publications on hate speech were recorded in WoS. After a peak of interest in 2020, it seemed stable. However, during 2023 there was a new surge. Additionally, it is observed that the trend in 2024 is also increasing, as in the first nine months of the year, it has already surpassed the total number of articles published during 2022 or 2021.

To analyze the impact of the publications, Figure 2 and Table 2 have been prepared. Figure 2 reflects the impact through the number of citations of articles per year. Additionally, Table 2 lists the most cited articles, highlighting their key characteristics and confirming the interest received from the academic field in this field.

Only seven countries show a predisposition to disseminate research on hate speech, concentrating the total of the analyzed publications (98). Table 1 shows that the issue of hate speech is a topic of particular interest to scientific journals in England and the US, followed, although at a considerable distance, by Spain, Switzerland, and the Netherlands.

Table 1. Countries of the main journals addressing this topic.

Country of the journal	Number of articles	Percentage
England	32	32,65%
USA	31	31,63%
Portugal	10	10,20%
Spain	10	10,20%
Switzerland	10	10,20%
Netherlands	4	4,08%
Canada	1	1,02%
Total	98	100%

Table 2. Ranking of the most cited articles.

Authors	Article Title	Source Title	Author Keywords	Times Cited, WoS Core	Percentage	Publication Year	WoS Categories
R Gorwa R. Binns C. Katzenbach	Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance	<i>Big Data & Society</i>	platform governance; content moderation; algorithms; artificial intelligence; toxic speech; copyright	259	11,45%	2020	Social Sciences; Interdisciplinary
M. L. Williams P. Burnap L. Sloan	Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation	<i>Sociology</i>	algorithms; computational social science; context collapse; ethics; social data science; social media; Twitter	215	9,50%	2017	Sociology
A. Matamoros-Fernández	Platformed Racism: The Mediation and Circulation of an Australian Race-Based Controversy on Twitter, Facebook and YouTube	<i>Information, Communication & Society</i>	racism; platforms; digital methods; Twitter; Facebook; YouTube	192	8,49%	2017	Communication; Sociology
P. Burnap M. L. Williams	Us and Them: Identifying Cyber Hate on Twitter Across Multiple Protected Characteristics	<i>Epj Data Science</i>	cyber hate; hate speech; Twitter; NLP; machine learning	178	7,87%	2016	Mathematics; Interdisciplinary Applications; Social Sciences; Mathematical Methods
R. Rogers	Deplatforming: Following Extreme Internet Celebrities to Telegram and Alternative Social Media	<i>European Journal Of Communication</i>	deplatforming; social media; digital methods; Telegram; extreme speech	166	7,34%	2020	Communication

Table 2. (Cont.) Ranking of the most cited articles.

Authors	Article Title	Source Title	Author Keywords	Times Cited, WoS Core	Percentage	Publication Year	WoS Categories
A. Bruns	After the APIcalypse: Social Media Platforms and Their Fight Against Critical Scholarly Research	<i>Information Communication & Society</i>	Cambridge Analytica; social science one; Facebook; Twitter; application programming interface; social media	145	6,41%	2019	Communication; Sociology
M. L. Williams P. Burnap A. Javed H. Liu S. Ozalp	Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime	<i>British Journal of Criminology</i>	hate speech; hate crime; social media; predictive policing; big data; far right	110	4,86%	2020	Criminology & Penology
J. C. Pereira-Kohatsu L. Quijano-Sánchez F. Liberatore M. Camacho-Collados	Detecting and Monitoring Hate Speech in Twitter	<i>Sensors</i>	hate crime; sentiment analysis; text classification; predictive policing; social network analysis; Twitter	68	3,01%	2019	Chemistry; Analytical; Engineering, Electrical & Electronic; Instruments & Instrumentation
J. Van Dijck	Governing Digital Societies: Private Platforms, Public Values	<i>Computer Law & Security Review</i>	digital societies; private platforms; internet governance; platform values	67	2,96%	2020	Law
B. Vidgen T. Yasseri	Detecting Weak and Strong Islamophobic Hate Speech on Social Media	<i>Journal of Information Technology & Politics</i>	Hate speech; Islamophobia; prejudice; social media; natural language processing; machine learning	61	2,70%	2020	Communication; Political Science

The highest number of citations received corresponds to articles published in 2020. Among the 98 articles studied, two stand out for their relevance in the number of citations, both focusing on mathematical algorithms and their role in digital platforms. The 2020 article “Algorithmic Content Moderation: Technical and Political Challenges in the Automation of Platform Governance,” by Robert Gorwa, Reuben Binns, and Christian Katzenbach, has garnered the most citations (259 citations; 11.45%). It addresses the challenges digital platforms face in the algorithmic era when moderating their content. It is followed by an older publication (2017), published in *The Journal of the British Sociological Association*, by Matthew L. Williams, Pete Burnap, and Luke Sloan—which, although it has a more social focus, also addresses computational analysis and algorithms. The 2017 article “Platformed Racism: The Mediation and Circulation of an Australian Race-Based Controversy on Twitter, Facebook and YouTube” by Ariadna Matamoros-Fernández also has a sociological focus. It falls under the categories of Communication and Sociology in WoS. In this case, the article shifts its focus from algorithms to what the author has termed “platform racism” (see also Table 2). All of these articles have been published in English.

In terms of research areas, the interest this subject generates is most notable in the field of Communication (26.75%). This is the area that pays the most attention to this topic and therefore gathers the most articles. It is followed at a considerable distance by the area of Government Law (7.64% of publications), Computer Science (6.37%), Information Science—Library Science (6.37%), Social Sciences—Other Topics (5.73%), and Sociology (5.10%). In other areas, the number of publications does not reach 5%. Additionally, the journals most inclined to publish this type of article are *Social Media Society* with a total of seven articles (7.14%), *El Profesional de la información* with six articles (6.12%), *Media and Communication* with five articles (5.10%); and *Politics and Governance* also with five articles (5.10%).

Furthermore, hate speech as a subject of study attracts the interest of a diverse range of researchers, with 200 having contributed to the 98 articles published in WoS on this topic. However, despite the large number of authors, only 12 have published more than one article, indicating the limited specialization in this topic within the databases of academic impact journals indexed in WoS (Table 3). Additionally, most authors specialized in this field, meaning those who have published more than one article in WoS, are men.

Table 3. Authors specialized in the analysis of hate speech.

Authors	Number of Articles	Percentage of Total Articles
Burnap, Pete	5	2,24%
Williams, Matthew L.	5	2,24%
Arcila Calderón, Carlos	4	1,79%
Blanco-Herrero, David	4	1,79%
Liu, Han	3	1,35%
De Quincey, Ed	2	0,90%
Galesic, Mirta	2	0,90%
González-Aguilar, Juan Manuel	2	0,90%
Ozalp, Sefa	2	0,90%
Piñeiro-Otero, Teresa	2	0,90%
Poole, Elizabeth	2	0,90%
Sánchez-Holgado, Patricia	2	0,90%

Table 3 shows the ranking of authors with the most publications. It can be observed that authors Pete Burnap and Matthew Williams are the ones who have published the most articles, with 2.24% each, of the total articles related to this topic indexed in the WoS database.

It is pertinent to note that some of the most frequently cited publications have been authored by some of the 12 specialized authors who have the most publications in WoS on this topic (Tables 2 and 3). However, eight of them, despite being part of the group of authors with the most publications, do not have any articles among the most cited, such as Carlos Arcila Calderón, David Blanco-Herrero, Ed De Quincey, Mirta Galesic, Juan Manuel González-Aguilar, Teresa Piñero-Otero, Elizabeth Poole, or Patricia Sánchez-Holgado.

Figure 3 allows for an in-depth analysis of the work and co-authorship networks of the authors linked to the 10 most cited articles on hate speech or those of the most specialized authors (more than one publication). Regarding the networks formed by the most cited authors, four distinct and separate networks can be observed. Researcher Pete Burnap, who, as previously mentioned, is one of the authors with the most articles published in WoS on this topic, has formed a network with five other authors. Three of them are also among the most specialized authors in the field: Matthew Williams (with five articles), Han Liu (three articles), and Sefa Ozalp (two articles).

The next network is formed by the authors Lara Quijano-Sánchez, Miguel Camacho-Collados, Juan Carlos Pereira-Kohatsu, and Federico Liberatore. These are authors whose articles have a high number of citations,

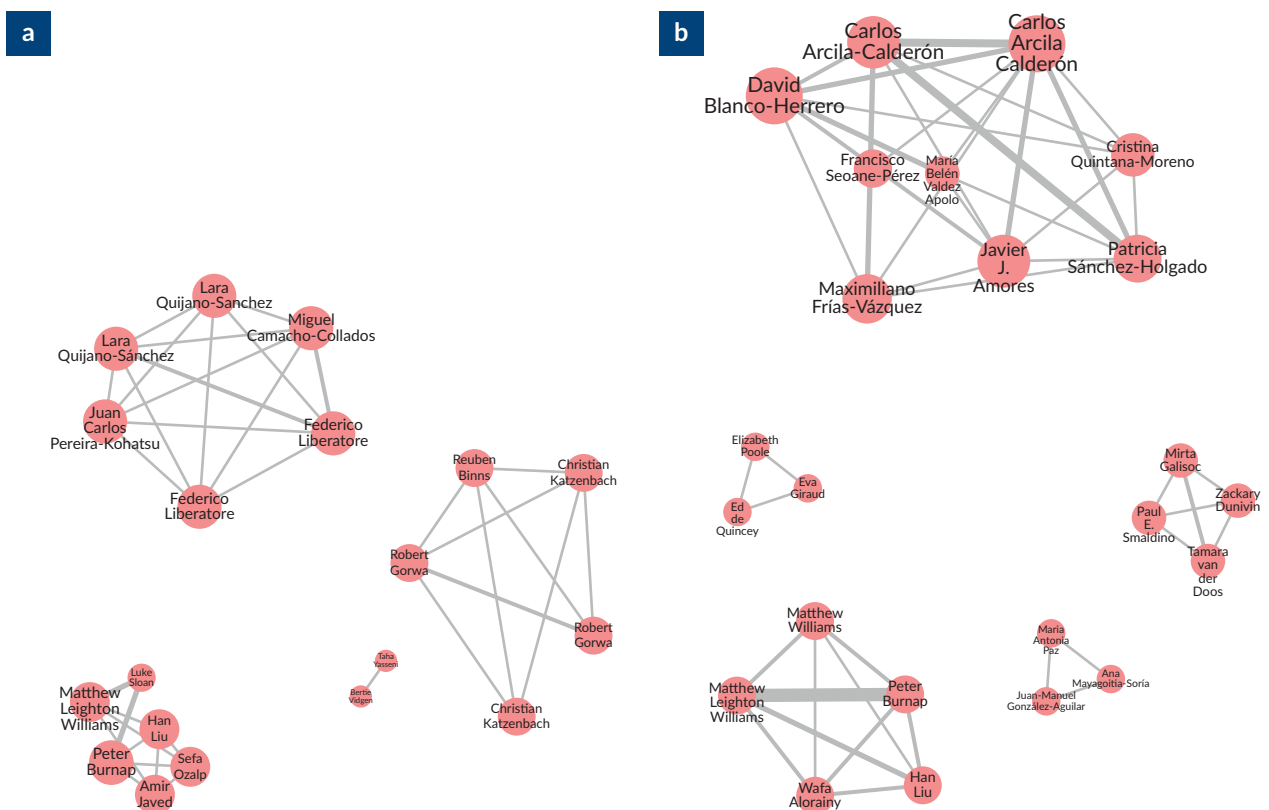


Figure 3. Co-authorship network of the most cited articles on hate speech in WoS: (a) network of nodes of the most cited authors; (b) network of nodes of the most prolific authors.

despite not being highly specialized in the topic, understanding specialized as having published more than one article on hate speech in WoS.

Among the members of the third network are Robert Gorwa, Reuben Binns, and Christian Katzenbach, who are the authors of the most cited article. However, all of them currently only have that one publication in WoS on hate speech. The last network is also formed by the authors of a highly cited work, namely Berte Vidgen and Taha Yasser.

Of the 98 articles analyzed, 86.73% were co-authored by multiple authors, while 13.27% were written individually. The majority of the most cited articles were co-authored (70%). Additionally, the primary language of publication is English (94.90%), with only 5.10% published in Spanish. It is noteworthy that the articles in Spanish are among the least cited, ranging between 10 and 28 citations.

Furthermore, the study of the methodologies employed by the analyzed publications revealed that most of the articles in the WoS database use a mixed-methods approach (53.06%) in their research. This provides the research with both qualitative and quantitative perspectives. Additionally, quantitative articles are more prevalent (27.55%) compared to those employing qualitative techniques (19.39%).

Next, we will address the trend in research techniques used for the study of hate speech on X. It is worth noting that only eight studies (out of a total of 98) have had an experimental nature. Additionally, Table 4 shows the diversity of tools used in this type of analysis, as well as their frequency of use. Noteworthy is the use of Content Analysis (49.57%), along with automated learning models (35.04%) and statistical analysis (22.22%).

Table 4. Ranking of research techniques in WoS publications addressing hate speech on X.

Research technique	Frequency	Percentage
Content analysis	58	49,57%
Machine learning model	41	35,04%
Statistical analysis	26	22,22%
Network analysis	7	5,98%
Discourse analysis	7	5,98%
Survey	6	5,13%
Interview	6	5,13%
Case study	6	5,13%
Data analysis	3	2,56%
Ethnographic study	3	2,56%
Critical literature review	3	2,56%
Comparative analysis	2	1,71%
Lexical approach (manual or automated)	2	1,71%
Critical discourse analysis	1	0,85%
Bibliometric analysis	1	0,85%
Critical analysis based on Erving Goffman's presentation of self theory	1	0,85%
Mixed engagement analysis	1	0,85%
Systematic theoretical analysis of mediated authenticity strategies	1	0,85%
Total	117	100%

Understanding the use of different research tools employed in these publications provides insight into the scope from which the issue has been approached. Multiple studies delve into this matter using algorithms aimed at preventing, monitoring, and even avoiding the dissemination of such discourse. Hence, 35.04% of the analyzed articles employed machine learning models linked to big data analysis. In this case, the handling and interest in this tool far surpass the use of other techniques that were once at their peak, such as surveys (5.13%) and interviews (5.13%).

3.2. Thematic Analysis

The following is a thematic analysis of the publications that are part of this research. This analysis aims to identify thematic trends and the evolution of research approaches in this field.

Through the study of keywords, it is possible to discern the trends in researchers' interests when conducting studies. Specifically, the predominant themes of these articles can be identified. In the analyzed publications, a total of 399 keywords were counted. Among these, there are references to the characteristics of hate speech. Thus, references to the victims of hate speech can be observed, including mentions of racism (1.75%), Islamophobia (1%), refugees (1%), immigration (0.25%), gender (2.26%), feminism (2.265%), transgender (0.75%), transphobia (0.75%), and sexism (1%), among others.

Additionally, there are references to digitalization and the channels of dissemination, such as Twitter (11.03%), Facebook (1%), platforms in general (1.75%), networks (4.51%), and the internet (1%). Furthermore, mentions of polarization of positions (1.75%), politics (4.51%), populism (1.25%), and journalism (1.25%) are also present. References to Covid-19 (1.50%) are also observed. Table 5 provides a detailed overview of the thirty most frequently occurring words.

As previously mentioned in the theoretical framework, research on hate speech has evolved and specialized in various areas. The initial publications in this database focused on the detection of hate speech, monitoring, and tools that, through machine learning, enable the prevention of hate speech.

Another set of publications addresses the types of hate speech or rather refers to some of the sectors that are victims of hate speech, such as articles discussing immigrants who are victims of such discourse (Matamoros-Fernández, 2017), religious discourse, mainly linked to Islam (Seijbel et al., 2023), or antisemitism (Ozalp et al., 2020).

Currently, when discussing hate speech, although the immigrant community is one of the most affected by such attacks, the themes of the publications have diversified along with the profiles of the victims, with new themes also emerging related to these discourses. In this regard, there remains an interest and need to study the dissemination of racist discourse on social networks (Agudelo & Olbrych, 2022; Criss et al., 2023; Nikunen, 2021). However, there are also publications that emphasize other topics, such as the propagation of a sexist discourse that subtly fosters discrimination against women (Haim & Maurus, 2023; Piñeiro-Otero & Martínez-Rolán, 2021; Rodríguez-Sánchez et al., 2020), discourse on sexual diversity (Arce-García & Menéndez-Menéndez, 2022), or those related to the Trans Law in Spain (Sánchez-Holgado et al., 2023).

Table 5. Most frequent keywords from the analyzed publications.

Number	Text	Frequency	Percentage
1	Social	58	14,54%
2	Speech	56	14,04%
3	Hate	54	13,53%
4	Twitter	44	11,03%
5	Media	41	10,28%
6	Analysis	27	6,77%
7	Online	23	5,76%
8	Politics	18	4,51%
9	Network	18	4,51%
10	Content	15	3,76%
11	Language	13	3,26%
12	Learning	10	2,51%
13	Violence	10	2,51%
14	Community	10	2,51%
15	Gender	9	2,26%
16	Digital	9	2,26%
17	Natural	8	2,01%
18	Processing	8	2,01%
19	Data	7	1,75%
20	Racism	7	1,75%
21	Platform	7	1,75%
22	Public	7	1,75%
23	Discourse	7	1,75%
24	Polarization	7	1,75%
25	Machine	6	1,50%
26	Sentiment	6	1,50%
27	Model	6	1,50%
28	Moderation	6	1,50%
29	Critical	6	1,50%
30	Covid	6	1,50%

Another topic addressed in the current analysis of the dissemination of hate speech is the manipulation of public opinion linked to the political sphere (Macagno, 2022; van der Does et al., 2022). Political representatives sometimes become a source of hate speech dissemination (Paz et al., 2021). An example is the study by Díez-Gutiérrez et al. (2022), which asserts that there is a destabilizing political intent.

Likewise, both journalists and politicians are sometimes victims of hate speech and thus become a focus of interest for researchers. Blanco-Castilla et al. (2022) have focused on how female sports journalists are victims of hate speech on social media. In the same vein, there are studies addressing the relationship between hate speech and feminism, focusing on the politician Irene Montero (Durántez-Stolle et al., 2023).

To conclude the thematic analysis, it should be noted that international relations between different countries and armed conflicts also impact the propagation of hate attitudes. Researchers Caldevilla-Domínguez et al. (2023) have investigated the emergence of possible cases of Russophobia on social media. Some of the most recent articles discuss the impact of the politicization of hate and the use of digital platforms (Ridwanullah et al., 2024).

4. Conclusion

In an increasingly globalized world, where technology plays a fundamental role, it has been observed that the growing democratization of the internet has enabled and, as Matamoros-Fernández (2017) suggests, has even driven the propagation of hate speech. This discourse aims to find and emphasize differences between people, highlighting identity markers. Hence, the prevalence of discourses related to racism, LGBTIphobia, immigration, feminism, etc. The strategies followed by these discourses are framed within theories of signaling, dehumanization, and the isolation of the victim of hate speech (Mafu, 2024).

In line with the theses of technological determinism (Ridwanullah et al., 2024), which advocates that technology is one of the main drivers of cultural changes, new technologies have led individuals to interact in different spaces and modify their behavior patterns. The change is not solely determined by a screen, but the fact that communication occurs without physical contact, can be asynchronous, and sometimes without revealing identity, has driven a change in how individuals relate to each other. From this same position, some authors (Matamoros-Fernández, 2017) have expressed concern about how this change in our way of interacting socially favors the emergence and dissemination of hate speech, which, combined with the speed of current communications through social networks, makes it spread and sometimes go viral quickly.

Furthermore, focusing on global elites and their ability to disseminate information, various authors have highlighted the disruptive or disturbing role that certain political leaders' discourses sometimes acquire concerning hate speech. Instead of being a realm that guarantees peace and coexistence, it can be repeatedly observed how various political discourses worldwide become generators of hate propagated through social networks via new technologies (Ridwanullah et al., 2024).

It should be noted that the systematic review we have conducted meets the objectives set at the beginning of the research, facilitating the understanding of the different manifestations of hate speech and its evolution. It has been observed that interest in this topic among researchers is gradually increasing. Focusing on the specialization of researchers, it should be noted that most publications have been presented by more than one researcher (86.73%), which shows the development of teamwork networks. In fact, most of the articles with the greatest impact, meaning those with the highest number of citations, have been produced by teams of several researchers (60%). This contrasts with the image of research work as an autonomous activity, despite its progress being due to advances in previous research.

Regarding thematic trends, different thematic blocks or lines of research can be identified, among which those focused on promoting machine learning models, on the one hand, aimed at automatically detecting hate speech on networks to provide a quick response at the moment it occurs stand out—these learning models also aim to segment hate content based on its nature to address it strategically, with the goal of reducing individuals' exposure to these discourses. On the other hand, there are lines of research more

linked to the analysis of hate speech according to its nature. For this reason, the topics addressed by these studies are so broad, as there are researchers who focus on sexist hate speech, others on racist hate speech, others on discrimination against sexual diversity, etc.

We would conclude with the line of research that focuses on investigating how to counteract hate speech from the discursive rhetoric itself and analyzing its scope and impact.

Among the limitations of this study are those related to the selection of search descriptors and the possible biases that may arise from deciding to use certain criteria over others. Additionally, the selection of the WoS database and the exclusion of other databases may also introduce biases in the research. However, it should be noted that this systematic review facilitates the identification of evidence-based trends that enable the development of prevention and reduction policies for these discourses. Once the reality of the dissemination of hate speech has been diagnosed, future research should analyze the impact of hate speeches on affected communities and explore countermeasures from the fields of sociology and communication.

Funding

This work is a result of the Hatemedia project (Proyecto PID2020-114584GB-I00), funded by MCIN/AEI/10.13039/501100011033.

Supplementary Material

Supplementary material for this article is available online in the format provided by the author (unedited).

References

- Agudelo, F. I., & Olbrych, N. (2022). It's not how you say it, it's what you say: Ambient digital racism and racial narratives on twitter. *Social Media + Society*, 8(3). <https://doi.org/10.1177/20563051221122441>
- Amores, J. J., Blanco-Herrero, D., Sánchez-Holgado, P., & Frías-Vázquez, M. (2021). Detecting ideological hatred on Twitter. Development and evaluation of a political ideology hate speech detector in tweets in Spanish. *Cuadernos.info*, 49, 98–124. <https://doi.org/10.7764/cdi.49.27817>
- Arce-García, S., & Menéndez-Menéndez, M.-I. (2022). Inflaming public debate: A methodology to determine origin and characteristics of hate speech about sexual and gender diversity on Twitter. *Profesional de la información*, 32(1). <https://doi.org/10.3145/epi.2023.ene.06>
- Baider, F. (2023). Accountability issues, online covert hate speech, and the efficacy of counter-speech. *Politics and Governance*, 11(2), 249–260. <https://doi.org/10.17645/pag.v11i2.6465>
- Benesch, S., Buerger, C., Clavinic, T., Manion, S., & Bateyko, D. (2021). *Dangerous speech: A practical guide*. The Dangerous Speech Project. <https://dangerousspeech.org>
- Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, 1(1), 363–376. https://doi.org/10.1162/qss_a_00018
- Blanco-Castilla, E., Fernández-Torres, M. J., & Cano-Galindo, J. (2022). Disinformation and hate speech toward female sports journalists. *Profesional de la información*, 31(6), Article 310613. <https://doi.org/10.3145/epi.2022.nov.13>
- Bruns, A. (2019). After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information Communication and Society*, 22(11), 1544–1566. <https://eprints.qut.edu.au/131676>
- Burnap, P., & Williams, M. L. (2016). Us and them: Identifying cyber hate on Twitter across multiple protected

- characteristics. *EPJ Data Science*, 5, 1–15. <https://link.springer.com/content/pdf/10.1140/epjds/s13688-016-0072-6.pdf>
- Bustos Martínez, L., De Santiago Ortega, P. P., Martínez Miró, M. Á., & Rengifo Hidalgo, M. S. (2019). Discursos de odio: Una epidemia que se propaga en la red. Estado de la cuestión sobre el racismo y la xenofobia en las redes sociales. *Mediaciones Sociales*, 18, 25–42. <https://doi.org/10.5209/meso.64527>
- Caldevilla-Domínguez, D., Barrientos-Báez, A., & Padilla-Castillo, G. (2023). Dilemmas between freedom of speech and hate speech: Russophobia on Facebook and Instagram in the Spanish media. *Politics and Governance*, 11(2), 147–159. <https://www.cogitatiopress.com/politicsandgovernance/article/view/6330>
- Committee of Ministers. (2022). [1434/4.4] *Steering Committee on Anti-Discrimination, Diversity and Inclusion (CDADI) and Steering Committee on Media and Information Society (CDMSI): Recommendation CM/Rec(2022)16 of the Committee of Ministers to member states on combating hate speech—Explanatory memorandum*. Council of Europe. <https://search.coe.int/cm/?i=0900001680a6891e>
- Criss, S., Nguyen, T. T., Michaels, E. K., Gee, G. C., Kiang, M. V., Nguyen, Q. C., Norton, S., Titherington, E., Nguyen, L., Yardi, I., Kim, M., Thai, N., Shepherd, A., & Kennedy, C. J. (2023). Solidarity and strife after the Atlanta spa shootings: A mixed methods study characterizing Twitter discussions by qualitative analysis and machine learning. *Frontiers in Public Health*, 11, Article 952069. <https://doi.org/10.3389/fpubh.2023.952069>
- Dekkers, O. M., Vandenbroucke, J. P., Cevallos, M., Renehan, A. G., Altman, D. G., & Egger, M. (2019). COSMOS-E: Guidance on conducting systematic reviews and meta-analyses of observational studies of etiology. *PLoS Med*, 16(2), Article 1002742. <https://doi.org/10.1371/journal.pmed.1002742>
- Díez-Gutiérrez, E. J., Verdeja, M., Sarrión-Andaluz, J., Buendía, L., & Macías-Tovar, J. (2022). Political hate speech of the far right on Twitter in Latin America. *Comunicar: Media Education Research Journal*, 30(72), 97–109.
- Durántez-Stolle, P., Martínez Sanz, R., Piñeiro Otero, T., & Gómez-García, S. (2023). Feminism as a polarizing axis of the political conversation on Twitter: The case of #IreneMonteroDimision. *Profesional de la información*, 32(6). <https://doi.org/10.3145/epi.2023.nov.07>
- European Union Agency for Fundamental Rights. (2023a). *Online content moderation—Current challenges in detecting hate speech*. https://fra.europa.eu/sites/default/files/fra_uploads/fra-2023-online-content-moderation_en.pdf
- European Union Agency for Fundamental Rights. (2023b). *EU high level group on combating hate speech and hate crime*. <https://fra.europa.eu/en/news/2023/eu-high-level-group-combating-hate-speech-and-hate-crime>
- FBI. (2023). *Crime data explorer*. <https://cde.ucr.cjis.gov/LATEST/webapp/#/pages/explorer/crime/hate-crime>
- Frenkel, S., & Conger, K. (2022, December 2). Hate speech's rise on Twitter is unprecedented, researchers find. *The New York Times*. <https://www.nytimes.com/2022/12/02/technology/twitter-hate-speech.html>
- Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews*, 18, Article 1230. <https://doi.org/10.1002/cl2.1230>
- Haim, M., & Maurus, K. (2023). Stereotypes and sexism? Effects of gender, topic, and user comments on journalists' credibility. *Journalism*, 24(7), 1442–1461. <https://doi.org/10.1177/14648849211063994>
- Hangartner, D., Gennaro, G., Alasiri, S., Bahrach, N., Bornhoft, A., Boucher, J., Demirci, B. B., Derksen, L., Salón, A., Jochum, M., Murias, M., Richter, M., Vogel, F., Wittwer, S., Wüthrich, F., Gilardi, F., & Donnay, K.

- (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences*, 118(50), Article 2116310118. <https://doi.org/10.1073/pnas.2116310118>
- IAB Spain. (2023). XV Edición: Estudio de redes sociales 2024. <https://iabspain.es/estudio/estudio-de-redes-sociales-2024>
- Kemp, S. (2023, April 27). What's really going on with Twitter. *Datare Portal*. <https://datareportal.com/reports/digital-2023-deep-dive-the-state-of-twitter-in-april-2023>
- Macagno, F. (2022). Argumentation profiles and the manipulation of common ground. The arguments of populist leaders on Twitter. *Journal of Pragmatics*, 191, 67–82. <https://doi.org/10.1016/j.pragma.2022.01.022>
- Mafu, L. (2024). Discourse structures, weaponization of language and Ethiopia's civil war. *Sage Open*, 14(3). <https://doi.org/10.1177/21582440241262755>
- Martínez Valerio, L. (2022). Mensajes de odio hacia la comunidad LGTBQ+: Análisis de los perfiles de Instagram de la prensa española durante la "Semana del Orgullo." *Revista Latina de Comunicación Social*, 80, 363–388. <https://www.doi.org/10.4185/RLCS-2022-1749>
- Matamoros-Fernández, A. (2017). Platformed racism: The mediation and circulation of an Australian race-based controversy on Twitter, Facebook and YouTube. *Information, Communication & Society*, 20(6), 930–946. <https://doi.org/10.1080/1369118X.2017.1293130>
- Meyrowitz, J. (1994). Medium theory. In D. Crowley & D. Mitchell (Eds.), *Communication theory today* (pp. 50–77). Stanford University Press.
- Miller, C., Weir, D., Ring, S., Marsh, O., Inskip, C., & Chavana, N. P. (2023). *Antisemitism on twitter before and after Elon Musk's acquisition*. Institute for Strategic Dialogue.
- Ministerio del Interior. (2024). *Informe sobre la evolución de los delitos de odio en España 2023*. https://www.interior.gob.es/opencms/export/sites/default/.galleries/galeria-de-prensa/documentos-y-multimedia/balances-e-informes/2023/Informe_evolucion_delitos_odio_Espana_2023.pdf
- Nikunen, K. (2021). Ghosts of white methods? The challenges of Big Data research in exploring racism in digital context. *Big Data & Society*, 8(2). <https://doi.org/10.1177/205395172111048964>
- Oboler, A. (2008). Online antisemitism 2.0. "Social antisemitism" on the "social web." *Jerusalem Center for Public Affairs*, 67. <https://jcpa.org/article/online-antisemitism-2-0-social-antisemitism-on-the-social-web>
- Oboler, A. (2022). *Online hate prevention institute submission: National anti-racism framework*. Online Hate Prevention Institute. <https://ohpi.org.au/wp-content/uploads/2022/03/Online-Hate-Prevention-Institute-Submission-re-National-Anti-Racism-Framework.pdf>
- OSCE. (2014). *What is hate crime*. <https://hatecrime.osce.org>
- OHCHR. (2019). *Centroafricana: Preventing incitement to hatred and violence in the Central African Republic*. <https://www.ohchr.org/es/stories/2019/05/preventing-incitement-hatred-and-violence-central-african-republic>
- OSCE. (2024). ODIHR's hate crime data. https://hatecrime.osce.org/sites/default/files/2024-11/2023%20Hate%20Crime%20Data%20Findings_FINAL_for%20PPT%20and%20PDF_1811%20%281%29.pdf
- Ott, B. L. (2016). The age of Twitter: Donald J. Trump and the politics of debasement. *Critical Studies in Media Communication*, 34(1), 59–68. <https://doi.org/10.1080/15295036.2016.1266686>
- Ozalp, S., Williams, M. L., Burnap, P., Liu, H., & Mostafa, M. (2020). Antisemitism on Twitter: Collective efficacy and the role of community organisations in challenging online hate speech. *Social Media + Society*, 6(2). <https://doi.org/10.1177/2056305120916850>
- Paz, M. A., Mayagoitia-Soria, A., & González-Aguilar, J. M. (2021). From polarization to hate: Portrait of the Spanish political meme. *Social Media + Society*, 7(4). <https://doi.org/10.1177/20563051211062920>

- Pereira-Kohatsu, J. C., Quijano-Sánchez, L., Liberatore, F., & Camacho-Collados, M. (2019). Detecting and monitoring hate speech in Twitter. *Sensors*, 19(21), Article 4654. <https://doi.org/10.3390/s19214654>
- Piñeiro-Otero, T., & Martínez-Rolán, X. (2021). Say it to my face: Analysing hate speech against women on Twitter. *Profesional de la información*, 30(5). <https://doi.org/10.3145/epi.2021.sep.02>
- Ridwanullah, A. O., Sule, S. Y. U., Usman, B., & Abdulsalam, L. U. (2024). Politicization of hate and weaponization of Twitter/X in a polarized digital space in Nigeria. *Journal of Asian and African Studies*. Advance online publication. <https://doi.org/10.1177/00219096241230500>
- Rodríguez-Izquierdo, R. M., & García Bayón, I. (2024). Revisión sistemática sobre educación para una ciudadanía global transformadora. *Revista Internacional de Educación para la Justicia Social*, 13(1), 171–186. <https://doi.org/10.15366/riejs2024.13.1.009>
- Rodríguez-Sánchez, F., Carrillo-de-Albornoz, J., & Plaza, L. (2020). Automatic classification of sexism in social networks: An empirical study on Twitter data. *IEEE Access*, 8, 219563–219576. <https://ieeexplore.ieee.org/abstract/document/9281090>
- Römer, M., Camilli, C., & Matarín, E. (2023). Uso y utilidad de las redes sociales para la generación de conocimiento científico en profesores no universitarios en España: Una revisión sistemática. In E. Said-Hung (Ed), *Comunicación y diseminación científica a través de las redes sociales. Aproximación desde el ámbito educativo en España* (pp. 161–188). Tirant Humanidades.
- Sánchez-Holgado, P., Arcila-Calderón, C., & Gomes-Barbosa, M. (2023). Hate speech and polarization around the “trans law” in Spain. *Politics and Governance*, 11(2), 187–197. <https://doi.org/10.17645/pag.v11i2.6374>
- Schäfer, S., Rebasso, I., Boyer, M. M., & Planitzer, A. M. (2024). Can we counteract hate? Effects of online hate speech and counter speech on the perception of social groups. *Communication Research*, 51(5), 553–579. <https://doi.org/10.1177/00936502231201091>
- Seijbel, J., van Sterkenburg, J., & Spaaij, R. (2023). Online football-related antisemitism in the context of the COVID-19 pandemic: A multi-method analysis of the Dutch Twittersphere. *American Behavioral Scientist*, 67(11), 1304–1321. <https://doi.org/10.1177/00027642221118286>
- UNESCO. (2021). *Addressing hate speech on social media: Contemporary challenges*. <https://unesdoc.unesco.org/ark:/48223/pf0000379177>
- United Nations. (2019). *La estrategia y plan de acción de las Naciones Unidas para la lucha contra el discurso de odio*. https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_ES.pdf
- United Nations. (2022). *Understand what hate speech is*. <https://www.un.org/es/hate-speech/understanding-hate-speech/what-is-hate-speech>
- United Nations. (2024). *Hate speech is rising around the world*. <https://www.un.org/en/hate-speech>
- US Department of Justice. (2023). *2021 hate crime statistics*. <https://www.justice.gov/es/hatecrimes/hate-crime-statistics-2021>
- van der Does, T., Galesic, M., Dunivin, Z. O., & Smaldino, P. E. (2022). Strategic identity signaling in heterogeneous networks. *Proceedings of the National Academy of Sciences*, 119(10). <https://doi.org/10.1073/pnas.2117898119>
- Velt, H., Torkkeli, L., & Laine, I. (2020). Entrepreneurial ecosystem research: Bibliometric mapping of the domain. *Journal of Business Ecosystems*, 1(2), 43–83. <http://doi.org/10.4018/JBE.20200701.oa1>
- Vicent, N., Castrillo, J., Ibañez-Etxebarria, A., & Albas, L. (2020). Conflictos armados y su tratamiento en educación. Análisis de la producción científica de los últimos 25 años en la Web of Science. *Panta Rei: Revista digital de Historia y didáctica de la Historia*, 14(2), 55–91. <http://doi.org/10.6018/pantarei.445721>

Wendling, M. (2023, April 13). Twitter and hate speech: What's the evidence? *BBC*. <https://www.bbc.com/news/world-us-canada-65246394>

Williams, M. L., Burnap, P., & Sloan, L. (2017). Towards an ethical framework for publishing Twitter data in social research: Taking into account users' views, online context and algorithmic estimation. *Sociology*, 51(6), 1149–1168.

About the Authors



Eva Matarín Rodríguez-Peral is a sociologist and PhD in audiovisual communication, advertising, and public relations. She is an accredited PCD by ANECA and was the deputy academic coordinator of the master's degree Inclusive and Intercultural Education at the UNIR. She is currently a lecturer in sociology at the Universidad Rey Juan Carlos. She has participated in international conferences and has published articles in various indexed scientific journals such as the *Revista Salud Colectiva* and the *Revista Latina de Comunicación Social*. Her lines of research focus mainly on the study of migration and also address issues such as hate speech and disinformation.



Tomás Gómez Franco holds a PhD in applied economics from the UNED. He has earned his graduate degree in economics and business studies from the UCM and in law from the UNED, with an official master's degree in access to the legal profession. He was accredited by ANECA as a Hired Doctor Professor and Private University Professor. As a professor at the Francisco of Vitoria University, he also holds a six-year research period. Tomás is the author of several articles, books, and chapters in prestigious publications and publishers. His lines of research are linked to economics and health.



Daniel Rodríguez-Peral Bustos is the head of PR and external communication at Sopra Steria. With a master's in digital marketing and journalism, a degree in philosophy, and currently a PhD candidate in journalism at the Faculty of Information Sciences at the Complutense University of Madrid, Daniel joined Sopra Steria in 2017 after working for large media companies such as Agencia EFE. He is currently the head of media relations and digital marketing at the company.

Frames and Triggers of Extreme Speech: The Case of Transphobia

Fabienne Baider 

Department of European Studies, University of Cyprus, Cyprus

Correspondence: Fabienne Baider (fabienne@ucy.ac.cy)

Submitted: 7 October 2024 **Accepted:** 8 January 2025 **Published:** 18 February 2025

Issue: This article is part of the issue “Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment” edited by Max Römer Pieretti (Universidad Camilo José Cela), Beatriz Esteban-Ramiro (Universidad de Castilla-La Mancha), and Agrivalca Canelón (Universidad Católica Andrés Bello), fully open access at <https://doi.org/10.17645/si.i415>

Abstract

This research explores, from a critical discourse perspective, the following questions: What frames are used to construct transphobic argumentation and its counterargumentation? What triggers the discriminatory comments? Answers to these questions will help in understanding what motivates transphobia and how to improve counterspeech. Our data comprises 1137 annotated comments discussing trans personnel in the UK army. The annotations and our analysis focus on the argumentation used in transphobic speech to construct the topos of threat and its counterspeech. We adopt both a quantitative and a qualitative approach and identify two main argumentative frames (the medical and misfit frames), their counterspeech, which is mainly based on logic and facts, and the triggers of transphobia, namely gender ideology based on binarism as well as the role played by public figures such as politicians in spreading disinformation and prejudice. Our results include suggesting an argumentation schema (argument, premise, conclusion and claim) based on the topos of threat, a schema which may be used in automatic counterspeech.

Keywords

counterspeech; misfit and medical frames; argumentation; topos of threat; triggers of transphobia; Trump

1. Introduction

On 27 January 2025, President Trump signed an executive order banning trans people from the army (“Trump signs order,” 2025). This article focuses on transphobic speech posted on social media after Donald Trump had taken similar steps in 2017. Transphobic online comments have not yet been addressed from a critical discourse analysis perspective, except for a study by Colliver et al. (2019) that adopted a critical discursive psychology framework. Indeed, in the relevant literature, we can find research targeting offline representations and perceptions of trans people (Craig et al., 2015; McInroy & Craig, 2015) and

representations of trans identities in the press (Baker, 2014; Turner et al., 2009; Zottola, 2021). In the present study, we examine the discursive strategies used to construct the othering of these specific non-heteronormative identities by using the topos of threat and the counterspeech used to deconstruct these strategies. Based on our analysis, we suggest a schema of argumentation to potentially automatically create counterspeech. Our data consists of online and annotated computational data (IMsyPP project, 2019–2022). To begin, we describe the fundamental concepts underpinning this study, i.e., by explaining what distinguishes hate speech from extreme speech, and how we define the concepts of frame, trigger, and topos. We next undertake a brief literature review, followed by a discussion of the data and quantitative results. We present our analysis of the main frames of transphobia and their counterspeech, as well as the triggers of such frames. We conclude by suggesting a schema of argumentation which could be used in automating counterspeech.

2. Fundamental Definitions

2.1. Extreme Speech and Hate Speech

In this study, we use the term “extreme speech” to describe the hurtful discourse targeting trans people. We do not use the label “hate speech” because the legal definition used in the European legal space (EU Council Framework Decision of 28 November 2008, 2008) restricts hate speech to a statement that fulfils three essential criteria: (a) the speaker/writer’s intent to incite the audience to do harm against a targeted group; (b) the speaker advocates violence and/or hatred; (c) the speaker targets a group that is historically disadvantaged and vulnerable (see Baider, 2020; Guillén-Nieto, 2023). Examples of transphobic hate speech would include comments such as “trannies ought to be shot dead” because it includes a call for violence and demonstrates hatred. Although hate speech is often accompanied by abusive language, not all abusive statements can be classified as hate speech (Ibrohim & Budi, 2023). Indeed, the above criteria defining hate speech do not fit our data, as most comments analysed in the present study constitute discriminatory and/or offensive speech that manifests in the form of prejudiced statements and judgements, such as “transgenders suffer from mental illness.” They do not call for acts of violence or hatred directed at trans individuals and are not threatening. We have therefore labelled such prejudiced speech “extreme speech” (see Baider & Gregoriou, in press).

Extreme speech can express anxiety, anger, and/or grievance and frustration (van der Vegt et al., 2021). It “pushes the boundaries of civil language,” as suggested by several authors (e.g., Udupa & Pohjonen, 2019; Udupa et al., 2021), and it is not always aimed at a disadvantaged individual or a group, which is the case in hate speech definitions. Although it may be socially reprehensible discourse, it is still protected by freedom of expression laws (Baider, 2018; Maynard & Benesch, 2016; Udupa et al., 2021; Udupa & Pohjonen, 2019). Indeed, speech that is offensive or aggressive is not generally considered a human rights violation, and most social media watchdogs do not take down extreme speech (Cohen-Almagor, 2014). The label “extreme speech” for such derogatory discourse is useful in the sense that it allows researchers to examine and analyse extreme comments and posts, as these may lead the audience to accept hate speech. Understanding extreme speech and its argumentation helps in composing convincing counterarguments that, for example, can debunk the disinformation typically found in extreme speech. Therefore, by labelling and analysing extreme speech, we hope to find a way to effectively derail a potential spiral of verbal violence.

2.2. Frame, Topos, and Trigger

The concept of framing is widely used in sociology, and it is also a concept commonly used in linguistics (Baider, 2018; Kecskes, 2006; Rácz, 2013). Among different definitions and interpretations of a frame (Entman, 1993; Fillmore, 1982; Goffman, 1974), we decided to work with Fillmore's (1982) and Entman's (1993) definitions. Both theoreticians refer to framing as the process of selecting the most salient elements with which individuals organise their understanding of a given situation (Rácz, 2013); in turn, this selection of some aspects of reality influences the interpretation and evaluation of the events, ideas, and people:

To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described. (Entman, 1993, p. 52)

In other words, framing means choosing or identifying lexical items, grammatical structures, and rhetorical devices that knowingly or unknowingly fit a specific worldview (Entman, 1993; Fillmore, 1982). Alternatives to the word "frame" include the words "schema," "script," "scenario," or "cognitive model" (Fillmore, 1982, p. 11). Fillmore (1982) explains how inferences, a most important characteristic of extreme speech, are subsequently made based on elements in the frame:

Any system of concepts related in such a way that to understand any of them you have to understand the whole structure in which it fits; when one of the things in such a structure is introduced into a text or into a conversation, *all of the others are automatically made available*. (p. 11, emphasis added)

Framing is a cognitive structure wherein all elements are interdependent; as such, when one of the elements of such a structure is found in a text, all the other elements may come to mind. For instance, framing non-heterosexuality as a social/moral threat (Baider, 2018, 2020) allows commenters to interpret the pride parade as a social/moral provocation rather than a celebration of human rights and freedom. The frames in our data will be associated with the most salient concepts in the debates, which can be identified by finding the most frequent lexical entities in the data (Müller et al., 2022).

While identifying frames is key to understanding the cognitive structures characterising a specific extreme speech comment, it is also essential to identify what triggers such frames to be able to foresee discriminatory debates. A noun or a verb *trigger* refers to the release and the cause of another event, as, for example, in a chain reaction. Triggers can be people, texts, or topics that tend to elicit or release toxic comments. It should be noted that often it is not the events themselves but their coverage by the media that triggers extreme reactions and comments (Legewie, 2013).

2.3. Transphobic Speech Studies

In our literature review, we found that the noun "transgender" was used in the press; however, the word "trans" has been recommended by some researchers (Colliver et al., 2019) and by the community itself (Stonewall, 2023). For that reason, we will use the terms "trans/trans people" in our research, but we will use the word "transgender" when quoting our data. Hill and Willoughby (2005, p. 533) described transphobia as "emotional disgust toward individuals who do not conform to society's gender expectations"; it "involves the feeling of

revulsion to masculine women, feminine men, cross-dressers, transgenderists, and/or transsexuals.” It refers to hostile responses to people who are perceived to be trans and who do not comply with the gender standards of society. The “phobia” suffix implies “an irrational fear or hatred, one that is at least partly perpetuated by cultural ideology” (Hill & Willoughby, 2005, p. 91). Homophobia (Weinberg, 1973) has been defined in the same way, i.e., as a dread of being close to homosexuals. Both transphobia and homophobia are literally based on a fear of non-heterosexuality, which is then felt and thought of as a threat. Transphobia and homophobia are, therefore, closely related to the topos of threat to heterosexuality. The topos of threat is here based on Reisigl and Wodak’s (2009, p. 102) proposal: If a political decision carries specific threatening consequences, it should be prohibited. Comments examined in the present study are deploying argumentation to sustain or resist this topos.

Discourse studies focused on homophobic speech (Baider, 2018; Brindle, 2016; Lillian, 2005; Mongie, 2016; Reddy, 2002) have identified the most common frames used to construct a homophobic discourse: a threat to national or individual security, family values, and morality, to social fairness, or as a physical threat (Lillian, 2005). In fact, Nagoshi et al. (2008) concluded that transphobia was positively correlated with homophobia, political conservatism (strong support for social conventions), gender role beliefs (concerning traditional values and roles), and religiosity. In this respect, transphobic speech implies not only sexual prejudices but also negative attitudes toward outgroups in general (Norton & Herek, 2013, p. 749). Research has also noted significantly greater negative attitudes towards transgender people than towards members of other sexual minorities (Chakraborti & Hardy, 2015; Norton & Herek, 2013, p. 749; Turner et al., 2009), hence the importance of working on transphobic speech. Speech acts that convey such a discriminatory attitude/belief manifest themselves in the form of prejudice, discrimination, harassment and acts of violence (Bandini & Maggi, 2014). Discriminatory comments (44%) and verbal abuse (27%) are the most common forms of harassment targeting trans people (Turner et al., 2009, p. 20); understanding such hurtful discourse can help us create counterspeech that can address such verbal violence. As noted in the introduction section, Baker (2014), Colliver et al. (2019), and Zottola (2021) are among the few studies devoted to representations and construction of trans individuals’ identities online or in the press. The studies by Baker (2014) and Zottola (2021) aimed to determine if the discursive strategies used to represent transgender identities reflect specific political and ideological stances. Colliver et al. (2019) is the only study that analysed online comments, albeit with a specific focus on the debate triggered by gender-neutral toilets. They found similar discursive strategies as the ones found in homophobic speech. Our study will attempt to generalise the specific frames and argumentation that construct the topos of threat regarding trans people and identify the triggers of such a topos.

Therefore, our research questions are the following:

- RQ1. What frames are discursively building transphobia and therefore the topos of threat?
- RQ2. What strategies are used to respond to these frames?
- RQ3. What schema of argumentation can summarise these strategies and frames?

3. Data, Methodology, and Identification of Frames

3.1. Data and Methodology

We worked with data from the IMsyPP EU project (2020–2022) that focused on LGBTQ+ issues: 3,000 Facebook posts annotated as “extreme speech” and counterspeech that were extracted from a total of 15,000 annotated and other comments related to a variety of topics, e.g., migration. The dataset was annotated in the EU IMsyPP project and is described in Baider (2023). More specifically, we annotated the discursive strategies used in extreme speech, i.e., distinguishing a rhetoric category, subdivided into argumentation (and further divided into logic/reasoning, statistics, examples, history, and other facts) and affective rhetoric (subdivided into insult, personal attack, empathy with acknowledging grievances, displaying positive emotions, displaying negative emotions, and sarcasm).

Among the 3,000 annotated comments focusing on LGBTQ+ issues, we noted two press articles that triggered more than 1,000 comments—one of which was the only article focusing on trans people. It comprised approximately 1,137 Facebook comments posted in 2017 in English. The comments were posted under an article on the BBC website focusing on the presence of trans people in the British army. The article was titled “UK Military Chiefs Praise Transgender Troops” and the first paragraphs read as follows:

Commanders from British armed forces have opposed any ban on transgender people serving in the military.

[This position] comes after Donald Trump said that transgender people would not be allowed in the US military due to “tremendous” medical costs and disruption. But British officials have supported people serving in the Army, Royal Navy and Royal Air Force. Commander of UK Maritime Forces Rear Admiral Alex Burton tweeted: “I am so glad we are not going this way.”

The Obama administration decided last year to allow transgender people to serve openly in the US military. But in June, Defence Secretary James Mattis agreed to a six-month delay in the recruitment of transgender people. (“UK Military Chiefs,” 2017)

We analysed our data with corpus linguistic tools (Baker et al., 2008; Brindle, 2016), including the AntConc software, as well as with a qualitative analysis, to identify the frames and triggers. For the qualitative analysis, the choices related to the lexicon (lexical paradigm) and those related to rhetoric (for instance, sarcasm and metaphors) were investigated.

3.2. Identification of Frames

Based on Love and Baker’s (2015, p. 64) view that keywords signal the most important themes in a text, we identified the most frequent words. We believe that these will reveal the most salient concepts in the thread we are examining (Baider, 2018; Giora, 1999; Müller et al., 2022) and these are the concepts that will frame the debate.

As the most frequent words in any text will include grammatical prepositions, articles, adverbs, etc., and the verbs “to be” and “to have,” we did not include these in our word count. The adjective “mental” was the only adjective to show the same frequency as a noun (69) and we discuss this later. Table 1 illustrates the most frequent nouns in our data.

Table 1. Most frequent nouns in the thread (word types = 4549; word tokens = 33,093).

Nouns	Number of occurrences
People	243
Military	158
Transgender	156
Gender	114
Country	94
Trump	75
Women	71
Army	55
Trans	52
Gay	49
Man	47
Sex	47
Illness	47
Life	46

What can these keywords tell us about the likely frames of the debate? Not surprisingly, the words “military,” “transgender,” “trans,” “gender,” “army,” “sex,” and “gay” are high on the list, given that they are the main themes of the article (focusing on trans people in the army). However, the words “Trump,” “women,” “illness,” and “life” are less expected and more revealing of the specificities of our data; therefore, we focused our attention on these keywords and the comments in which they appear. In other words, we returned to the textual context using the Key Word in Context (KWIC) extracts; as such, we were able to understand the role of these keywords in the argumentation for and against trans persons serving in the army, as well as how these arguments fit into a common frame.

The presence of the name Trump indicates, foremost, that the words of the American president have a strong influence on the debates. Trump stated that allowing trans in the army would lead to “tremendous” medical cost and disruption. Indeed, looking at comments containing the noun “Trump,” the medical burden argument is used extensively against trans. The concept of “disruption” is derived from analysis of the contextual use of the keyword “life”: Trans will disrupt the objectives of an army to serve and protect the nation because they are too physically weak—not only are they unable to save lives, but they would also endanger lives. Trans are seen as unfit to serve in an army.

The frequency of the word “women” indexes a contiguity between the concept of transgender and the female body/femininity. In numerous KWIC extracts, commenters describe women as too physically weak and “too emotional” to serve in the armed forces. This contiguity also testifies to the association of transphobic speech with the heterosexism inherent in traditional misogyny, here using the typical

stereotypes of frail and emotional women. These stereotypical female traits are incompatible with a life in the armed services. The noun “illness,” in expressions such as “mentally ill,” “mentally unstable,” or “mental illness,” explains the high frequency of the adjective “mental” and is suggestive of a psychopathological stereotype associated with trans people.

Most of the arguments supporting the claim that trans should not be in the army because they will pose a threat involve negative stereotypes such as physical and psychological weaknesses, medical costs and psychological disease. From these, we can identify two main frames: the misfit frame, where trans are physically and psychologically weak—like women—and unable to adapt to military life; the medical frame, where trans are seen as posing a medical burden and a psychopathological hazard to the army. These two frames derive from Trump’s (quoted above) view against trans people in the army—a result that confirms the important role of politicians and the media in setting the agenda of online debates and in influencing the argumentation and opinion of commenters. The president’s statement quite literally framed the debates (cf. Section 6 on triggers).

4. Qualitative Analysis of Identified Frames

4.1. Misfit Frame

The misfit frame is present in 85% of comments, which is unsurprising as the thread under study is about serving in the army. This frame is built on arguments that transpose a danger to other soldiers owing to their weak character and/or their physical weakness.

4.1.1. Physical Weakness

In our literature review, we noted that Nagoshi et al. (2008) found a strong correlation between transphobia and political conservatism (strong support for social conventions, including traditional gender roles). These correlations, and more generally, gender stereotyping and heterosexism are evident in most transphobic comments.

Comment [595] calls for banning not only trans persons but also women from combat:

[595] The military/war in general is a brutal place....Just because you can, doesn’t mean you should....I’m all for women being aircraft pilots, submarine commanders, battleship commanders, but certain jobs are not meant for everyone. That goes for transgenders. We always have to be ready to fight and win conflicts, be at our peak without jeopardizing standards....The military isn’t a social experiment, it’s a place where we train to slaughter our enemies and create warriors.

Hegemonic masculinity, evident in the comment above, is stereotypically associated with virility, evident in the vocabulary (“brutal,” “slaughter,” “warrior,” “fight,” “win”) used to describe actions that only “real” men are capable of, as in the comment below:

[1206] They need real men on the battlefield, not the ones covered with make-up and worried about their nails and mascara.

Commenters seem to believe that the participation of both women and trans people in combat would threaten the survival of a nation at war, ultimately leading to its defeat.

4.1.2. Character Weakness

We found that the argument for character weakness was even more frequent than the argument citing physical weakness.

Comment [518] makes a forceful discriminatory speech, which can here be associated with hate speech, dehumanising trans people as “mentally ill creatures,” an expression that equates them with some sort of monster:

[518] Our *brave* Christian men and women that fight to defend our Christian *nation* should not have to fight side by side unstable mentally ill creatures. I pray that Trump moves forward with banning ALL LGBT from the military. This is the only way forward for a more powerful and efficient military force.

The quotation above takes Trump’s call for a ban to the extreme, asking for a ban on *all* (emphasised) LGBTQ+, including gays, bisexuals, and lesbians, not just trans people. This broad generalisation infers that all LGBT people are “mentally ill creatures,” the expression expressing utter contempt. The comment also sets US (good) against THEM (bad) using the ideological square (Allport, 1954; van Dijk, 1993): The “good ones” are those Christian men and women who are brave and, therefore, able to defend the nation. These Christian soldiers are associated with power and efficiency, while the emphasis on men *and* women reinforces both the sexual dichotomy and heterosexism. In contrast, the “bad ones” are the entire LGBT community who are supposedly not brave, surprisingly not Christian, and who would certainly disrupt the efficiency of the army.

Distrust is core to the argumentation of commenter [1108], who opens with a sarcastic remark questioning other commenters’ knowledge of the topic in relation to specialists in the field. The distrust of trans people (“don’t trust them,” “wouldn’t feel safe”) is based on the fact that trans are, for the commenter, confused: The desire to change gender is repeatedly interpreted as “not even understanding what gender they are” and to not “even understand their own ‘gender.’” Following a rather twisted logic, this means that the trans person would not understand what to do in a war zone. The statements “most of the enlisted” and “other soldiers agree” are intended to be an argument for authority, i.e., that the majority concur with the commenter:

[1108] Glad you guys know more than medical doctors who spent their entire lives researching this....People who have no idea what gender they are have no business in a war zone. Most of the enlisted, and I was one, don’t trust them because they don’t even understand their own “gender.” I wouldn’t feel safe in a foxhole with one, and other soldiers agree.

Although comment [459] expresses the same mistrust more laconically, its brevity actually serves to heighten the absurdity of the argument, which is, in fact, a syllogistic fallacy (Riesigl & Wodak, 2009, p. 102): The desire for a gender/sex change is interpreted as holding labile beliefs in *all* matters, which in the army would result in endangering their fellow soldiers:

[459] ...if they can change their sex, they can also change the side they are fighting for.

Therefore, the “labile” trans person is even potentially a traitor to their comrades and the nation, an accusation also found in homophobic data (Baider, 2018).

In summary, this misfit frame is argued by stating that trans display inherent physical, psychological, and moral weakness, making them completely unsuitable as soldiers. Worse, they are a threat to their comrades, to the army, and to the nation.

4.2. Medical Frame

The medical/scientific discourse functions as an argument for authority to back up the claim that trans cannot enrol in the army. This medical frame is essentialist in that it dehumanises trans people either explicitly or implicitly as psychopathological individuals (cf. “mentally ill creatures” in the previous comment [518]). Quantitatively, this frame is present in 75% of the comments, and from the very outset of the thread [496] through to its end [1500]. It is constructed throughout with the repetition of such expressions as “mental illness,” “gender dysphoria,” “confused,” which express the idea that “they do not know what gender they are.” In fact, these expressions can be considered as linguistic markers of transphobic discourse.

4.2.1. Psychopathology

In our data, “mental illness” is the term that most frequently co-occurs with the term “transgender.” Although this expression can refer to a psychopathology that affects only *some* trans individuals, as in the comment below, it is generally argued that it would be risky to have *any* trans people in the army:

[1474] Also, mental illness is sky-high in transgenders (not all, but why would you run the risk). Does anybody seriously think that this makes the military stronger?

Comments [870] and [1367] consider mental illness to be intrinsic to trans individuals’ psychological make-up. This conclusion is based on disinformation related to gender dysphoria—the commenters misinterpret this dysphoria as “they do not know their gender” (when in fact they know very well what gender they are):

[870] They don’t know their gender!? That’s mental illness. If you are a girl, then you are a girl. If you are a man, then you are a man. How can you not know your gender? I call it mental illness.

[1367] Transgenderism is a proven mental disorder, and the military has more important things to worry about and deal with than people confused about who they are!

Commenters even use both expressions, “mental illness” and “gender dysphoria,” in a dehumanising way, metonymically referring directly to trans people, reducing their identity to this supposed illness, such as in the quotations below:

[518] THANK YOU President Trump for keeping mental illness out of our military!

[1161] Gender dysphoria is the same thing as transgender.

Deciding that there is a “mental illness” involved allows commenters to again draw absurd parallels, e.g., associating trans people with mediocrity, as we see in comment [1080], or a lack of decency in [789], and considering them morally and cognitively lacking because of a gender identity mismatch:

[1080] We celebrate mediocrity and mental illness now.

[789] ...both concepts [transexual and transgender] are as confusing as the individuals who practice it. No decency or sense of direction.

However, medical discourse is the most powerful way that commenters validate these transphobic assessments. Science and medical authorities, for example, are cited in comment [1100] to prove that trans people are mentally ill, and their authority is underlined through emphasis on the title of the source, University *Distinguished Service Professor*, and his lengthy study (40 years).

[1100] Let’s talk facts and real science. Genetically, there are two genders in our species: male and female. They make up 99.93% of the population as either XY or XX chromosome pairs....The science is settled. 96.2% of doctors and scientists agree....The science is settled, transgenderism is a mental disorder! For forty years as the University Distinguished Service Professor of Psychiatry at Johns Hopkins Medical School...he has been studying people who claim to be transgender.

The commenter adopts a sententious tone (“let’s talk facts and real science”), which relegates any counterarguments to fantasyland and fake science; the repetition of the phrase “the science is settled” and the extraordinary numbers given (“99.93% of the population,” “96.2% of doctors and scientists”) aim to convince forum participants that if they disagree, they belong to a negligible minority. The verb “claim” points to the writer’s conviction that such dysphoria does not exist and is only a “perception disorder” (see below in [1100]). As Colliver et al. (2019, p. 223) also observed, the scientific discourse serves not only to prove that binarism is the only “tangible” reality but at the same time “negate[s] the possibility of transgenderism as [being] real or *authentic*” (emphasis added).

The commenter also notes the consequences of being trans which may endanger the trans person (suicide or other severely dysfunctional behaviour), but more importantly, which may affect the other soldiers; trans individuals are, therefore, a medical threat to those around them:

[1100] He [the professor in the quote above] notes this is a perception disorder (gender dysphoria) that must be treated as such, or extremely detrimental consequences will likely occur. He cites a long-term study that indicates high rates of suicide and other severely dysfunctional behaviour...

The medical frame is also obvious in comments averring that accepting trans people in the army would pose an economic threat or a financial burden for both the army and society at large.

4.2.2. Medical Burden

Trans individuals who opt for a surgical sex change undergo costly procedures, which is a central transphobic argument in the medical frame. The cost of medication is central to comments [546] and [1474]:

[546] [This is] not a gender issue. It's a medical condition issue....Are they willing to give up daily medications to serve, if needed?

[1474] It's the constant medical treatments and hormone injections that transgenders need which is a costly and inconvenient procedure.

And although it was Trump who raised the issue, a few commenters went further by spreading the fake news that trans people enlist *solely* to have sex change operations paid for, as in [456] and [1313]:

[456] No, they sign up to get a free sex change and drain our tax dollars just like every other liberal.

[1313] Let the UK take over the trans take care of their medical cost surgeries.

A final central argument is the interpretation of transgenderism as being a pre-existing condition, which should disqualify people from serving, as there are other illnesses that exclude people from the army. Therefore, allowing trans people to serve would make them a privileged minority compared to people with other "illnesses":

[1170] There are hundreds of medical and mental disorders that disqualify many.

[1175] That's why people with pre-existing conditions are precluded from military service.

Here, we see a reversal of the gender equality argument: it is a question of fairness and equality to reject trans persons from enlisting in the army, just as others with certain illnesses are disqualified.

5. Countering Transphobic Comments

Qualitative analysis is essential to identify the strategies that can successfully derail verbal violence; it is also essential to generate automatic production of counterspeech (Chung et al., 2019). Counterspeech strategies include debunking mis- and disinformation by providing testimonies and facts (Benesch, 2020; Ullmann & Tomalin, 2023).

5.1. Counterspeech Strategies

It has been suggested that to stop extreme or hate speech, we should ignore it (Cohen-Almagor, 2014; Vedder, 2001). However, many activists and researchers alike have underlined the fact that ignoring such speech against a community could serve to legitimise the disinformation and the prejudiced statements, and therefore may even encourage such talk. It also shows a lack of solidarity towards the targets of this symbolic violence, and ultimately leaves the online space to the extremists, discouraging any intervention in this intimidating environment (Howard, 2021).

We might define counterspeech as comments that respond to offensive statements or trolls and strive to affect the behaviour and thinking of people who spread or may sympathise with these prejudiced statements (Benesch, 2020; Ullmann & Tomalin, 2023). Researchers have identified a few discourse

strategies (Cohen-Almagor, 2014; Maynard & Benesch, 2016): correcting misstatements or misperceptions; discrediting the transphobic author/accuser; denouncing the speech as hateful; using humour to de-escalate conflict; adopting a positive tone to appeal to the other participants; or adopting hostile language to potentially persuade a participant to delete their message.

Most studies focusing on counterspeech have concluded that the most effective counterspeech is speech that fosters critical thinking (Braddock & Horgan, 2016; Gagliardone et al., 2015; Hangartner et al., 2021; Ullmann & Tomalin, 2023; Woo & Cho, 2023). Another effective strategy is to express positive feelings such as empathy and to acknowledge grievances (Baider, 2023; Wachs et al., 2023). However, our earlier studies found that counterspeech using positive emotions on social media was rare, while anger and contempt were the most common emotions. Sadly, therefore, even if such a strategy could be effective, it is rarely used and seldom leads to a favourable change in discourse (Baider, 2023; Howard, 2021; Konikoff, 2021). In our earlier research, and for this study, we drew on these suggestions and focused the annotation on the same rhetoric categories as we did for extreme speech, i.e., argumentative rhetoric and/or affective rhetoric.

In our data, we found a high degree of counter speech: From a sample of approximately 700 comments, we recorded 191 counternarratives against 506 transphobic comments, thus 38% counternarratives. Consider this in relation to earlier research focused on LGBTQ+ data, where counternarratives represented only 10% of responses (Baider, 2023; Chung et al., 2019). The strategies used to answer the two frames primarily focused on the use of logic and arguments (55%), followed by the use of statistics and facts (17%) and, finally, personal examples (8%). The other 20% are using affect, mainly displaying negative emotions towards the commentator.

5.2. Challenging the Misfit Frame

5.2.1. Logical Argumentation

Logical argumentation was one of the main strategies we recorded for counterspeech to the misfit frame. Some commenters argued against a link between sexuality/gender and fitness for battle, as in comment [631]:

[631] A person's gender or sexuality has no bearing on their "toughness." You could be a straight man and be weak, or a transgender man/woman and be tough.

Comment [608] supports the same argument with the logical explanation that passing the army test means you are fit to serve:

[608] If someone passes the tests and meets the standard set in those tests, then they are fit to be there, regardless of their gender or anything else.

Another example includes comment [584], which discusses what issues are problematic when recruiting for the army, such as "addictions [that] impair judgment," and notes they have no connection to trans people. Qualities that should be sought are a sense of responsibility, consideration, etc., and there is no evidence to suggest that trans do not display these very qualities:

[548] I was in the military. I wouldn't have had any issue serving with transgender people. Smokers and excessive drinkers were more of a problem. Those addictions impair judgement and give away locations. I wanted to serve with responsible, considerate, hygienic people who knew, and did, their role well. Nothing except that mattered/s.

5.2.2. Providing Data, Statistics, and Facts

The other most common argumentative strategy involved providing data to discredit false information, namely that trans cannot fight. The number of trans people already serving is cited, as in comment [659], which is only one among many comments noting that important fact:

[659] There are already 15,000 trans people serving. This from John McCain: "There is no reason to force service members who are able to fight, train, and deploy to leave the military—regardless of their gender identity...any American who wants to serve our country and is able to meet the standards should have the opportunity to do so—and should be treated as the patriots they are."

The qualification "patriot" in comment [659] is especially important. On the one hand, it responds to the earlier suggestion that trans people are potential traitors to the nation, while it also counters criteria disqualifying trans people from the army on the basis of patriotism—see comments [518] and [508], where a patriot is objectively defined as "any American willing and able to serve their country." The textual contiguity of the terms *trans people* and *respect* in comment [659] is particularly significant, as hostile comments typically associate trans persons with the opposite emotion: contempt.

The tremendous influence of Trump's comment is undermined by quoting other politicians such as McCain and military authorities (Pentagon, career military officers) who supported or support the enrolment of trans as in comment [695]:

[695] The decision to allow openly transgender persons to serve was taken with the support of the Pentagon. There are thousands of transgender persons in active service now....Politicians and career military officers have also gone on the record supporting the military's first steps into the modern world.

Furthermore, the commenter above considers that accepting trans people in the military indicates being in tune with progress ("supporting the military's first steps into the modern world"), which implies that any opponents are reactionaries and, therefore, preventing the army from evolving or progressing.

Finally, the most factual and objective comment, which renders the post even more powerful, debunked the "disruption" argument put forward by Donald Trump by quoting a survey and offering scientific evidence to counter the transphobic argument:

[1428] 1) There was a study performed by RAND Corp. in 2016 to determine the effects of having transgender people in the military, and the cost was determined to be negligible. 2) That study also examined several other military forces around the world that have transgender people enlisted and found no significant instances of loss of unit cohesion or morale. 3) That same study again noted that

the number of transgender people serving was between 1320 and 6630....Here's the study, if you're interested: <https://www.rand.org/news/press/2016/06/30.html>

However, no one commented on the information given. Hence the question whether such an objective and factual comment silenced contradictors who found the arguments to be valid or in contrary has no impact and was bypassed by the commenters.

5.3. Challenging the Medical Frame

Counterarguments that challenge the medical frame also include logical arguments to discredit disinformation about trans people suffering from mental illness, but the more frequent use of personal testimonies signals a difference with the misfit frame.

5.3.1. Logical Argumentation

To fight the stereotype of trans being mentally ill, commenter [900] argues that, confused or not, trans people are aware of their gender, thus contesting a claim about trans persons being confused and, therefore, mentally unstable:

[900] They know what gender they are. It's you who is confused.

This leads to the conclusion that it is the commenter who is confused, reversing the argument and the stigma, while also hinting at the ignorance of the other commenter.

5.3.2. Providing Data, Statistics, and Facts

Scientific authorities are also invoked to counter the mentally ill label:

[607] Being transgender isn't a mental illness. A psychologist or psychiatrist cannot diagnose you with transgenerness according to the Diagnostic and Statistical Manual of Mental Illness.

One of the most frequent attacks against the commenters defending trans rights is to accuse them of never having served and, therefore, of not knowing what they are talking about: 20% of the counterspeech comments start by asserting their legitimacy and offering personal testimonies, which preemptively disarm commenters. We found that this is the most effective strategy:

[666] As a current serving member with over 21yrs service in the infantry and as a firefighter, I have served with and currently serve with women and transgender....We also have transgender firefighters at some of our firehalls. There has never been a time that their abilities, dedication, sacrifice, motivation, fitness, has ever come into question. Too many men still feel insecure. They somehow try to use the excuse...it's a distraction, they can't meet the same physical standards, etc. There's so much ignorance going around. It blows my mind.

The argument about medication and surgery costs is addressed with facts, such as the clarification that cost would be covered by the NHS, that the NSH pays also for Viagra. This argument does not apply for the American medic system, though:

[457] You mean the drugs that the NHS provide?? Yeah. What a drain on the system.

[528] You okay with \$90M worth of Viagra?

Another commentator argues that it was unlikely that someone would enlist solely to undergo a sex change, however not providing any statistics nor research to back up the “unlikeliness” of such request:

[458] NOBODY signs up TO GET KILLED in order to get a free sex change—and IN THE UK IT’S FREE.

These few examples show how important it is that a counterargument is well-prepared, applicable, and based on facts. Otherwise, there is the risk that the dialogue will spiral out of control and lead to even more destructive comments, with the counterspeech representatives’ credibility compromised. Effective arguments demand that the commenter has equally strong arguments as their opponent.

6. Triggers of Extreme Speech

In this study, we distinguished three types of triggers: those that occur at the macro, meso, and micro levels (Fairclough, 2015). The comments are triggers found at the micro level: these involve the detailed critical analysis of the language used in the online conversations; at the meso level are the articles that triggered the comments and the specific context; triggers at the macro level implicate the broader social and cultural structures that shape the comments under scrutiny, especially the social/power relations, ideologies, and institutional practices that are at stake.

6.1. Triggers at the Macro Level

The language of denigration and the topos of threat we examined in this study do not happen in a social vacuum; it is a response to a broader agenda (Colliver et al., 2019; Mongie, 2016, pp. 164–165) related to the sexual and gender order. Our study topic, transpeople, is an example of a controversial topic (insofar as it is sensitive and disturbs the societal heterosexism) that contains multiple points of view that could *trigger a debate* (Wang et al., 2024). Discursive transphobia is assimilated into a symbolic, sexual, and gendered violence that reinforces the ideals of a patriarchal state (Mongie, 2016, pp. 164–165) and involves the broader issues of language, gender, sex, and sexual orientation (Colliver et al., 2019). Indeed, binarism ideology had been advocated by Donald Trump before he signed the executive order in 2025. Therefore, specific topics such as migration, gender, sexuality, or nationalism will trigger extreme speech.

6.2. Triggers at the Meso Level

The meso level concerns the language used by public figures such as politicians and then by journalists who report what politicians have said: Both function as the main *triggers* of discriminatory speech and represent “a source of transphobic attitudes observable within society and an instrument which reinforces them

further” (Derecka, 2019, p. 107). In our data, for example, Trump’s statement, uncritically mediated by journalists, played a major role in the arguments used in transphobic speech: his reference to the high cost of trans serving in the army was used in 50% of comments in the medical frame. His position as president of the USA at the time lent his words an authority, which was called on to reinforce an argument.

6.3. Triggers at the Micro Level

At the micro level, we consider the triggers that emerged during the online thread and from the analysis of the comments. Disrespectful comments against transpeople triggered fewer than eight reactions. Only three comments led to a significant number of responses. Comment [595] elicited 12 reactions. The comment was sexist and used the most dehumanising expression (“social experiment”); comment [700] elicited 25 reactions. The comment stated that too many privileges were given to a minority. Comment [846] elicited 71 comments and was a one-line summary of the main transphobic claim that not knowing your gender should forbid you from enrolling, yet it triggered the most responses. The same comment also produced the most uncivil counter speech, 80% being ad hominem attacks against the commenter (such as “assnoodle”), whilst comment [595] was more verbally violent. Thread dynamics must be considered to explain such results (Horawalavithana et al., 2022). Counterspeech citing unverified research and statistics triggered the most hostile responses. For instance, the transphobic comment [1100], which had the most impact on the thread, referenced a medical journal to refute an earlier counterargument that had also quoted a medical source; this shows how counterspeech can easily backfire (Howard, 2021, p. 934).

7. Conclusion

In this article, we examined how the well-known topos of threat in LGBTQ+ data (Baider, 2018; Brindle, 2016; Colliver et al., 2019; Lillian, 2005) is discursively constructed in transphobic comments. We identified the two main frames—the misfit frame (trans are weak) and the medical frame (trans are sick)—that are used to sustain the conclusion that trans persons should not be allowed to enlist in the army. We further subdivided these frames to categorise the two main threats: on the one hand, “physical weakness” and “character weakness”; and on the other, “psychopathology” and “medical burden.” These frames are similar to those found in Colliver et al.’s (2019) study, despite our different data (1100 comments focused on the specific societal role of being a soldier). Similarly to our findings, a trans person is described in Donald Trump’s 2025 executive order as a person whose mental and physical health conditions are “incompatible with active duty” (“Trump signs order,” 2025). To summarise our findings, we suggest the following argumentation schemes (Table 2) that illustrate how this topos functions in our data, using the example of the medical frame (“trans are sick”).

Table 2. Argumentation schemes of the threat topos, for the medical frame.

Medical frame	Argumentation scheme 1	Argumentation scheme 2
Argument	Trans people are mentally confused	Trans people need medication and surgery
Premise	The army does not enlist sick people because it endangers the security of soldiers and the nation	Medication and surgery are expensive
Conclusion	Trans people endanger national and individual security	Trans people will inflict a tremendous cost on the army
Claim	The army should not enlist trans people	The army should not enlist trans people

Arguments 1 and 2 are backed up by a premise that is accepted by most commenters; this accepted premise leads to the conclusion; this conclusion, in turn, supports the claim which is the main proposition that transphobic speech puts forward. Working further on this suggested argumentation scheme may offer a way to build counterspeech automatically and systematically. We found that the counterspeech examined in this article attempted to undermine each of the elements of the schema: It undermines the conclusion (the army should not enlist trans) by stating a fact (trans people are already serving); it undermines argument 1 (trans people are mentally confused) by quoting a scientific report that explains gender dysphoria; it undermines the claim (trans people endanger national and individual security) by listing problematic characteristics that pose a danger to the army that are unrelated to gender. Therefore, automation of such counterspeech based on such argumentation scheme may be an avenue of research. Indeed, while counterspeech may not affect the online debate, it is effective in defending the dignity and rights of the victim, strengthening their will to fight back (McInroy & Craig, 2015), and possibly enhancing critical thinking among commenters and readers. We also identified certain triggers of transphobia, especially the role played by public figures, such as politicians, in spreading disinformation and prejudices and by the media uncritically relaying such discourse. This leads us to emphasise the importance, in parallel with an online presence of counterspeech, of learning and teaching others how to recognise and challenge prejudices and disinformation in everyday life, thus sensitising young people to the complexities of various hate-motivated social attitudes that contribute to the production of social, economic and political hierarchies of domination (Woo & Cho, 2023).

Acknowledgments

I thank the reviewers, whose insights and suggestions enhanced the clarity, accuracy, and depth of the present work.

Funding

This article is based on data collected during the EU-co-founded IMsyPP project (2019–2022), Grant No. ID 875263 (<http://imsypp.ijs.si>) and on the research work carried out within the ARENAS project (2023–2027), Grant No. 101094731 (<http://arenasproject.eu>). The European Commission's support for the production of this publication does not constitute an endorsement of its contents, which reflect the views of the author alone, and the Commission cannot be held responsible for any use that may be made of the information contained therein.

Conflict of Interests

The author declares no conflict of interest.

References

- Baider, F. (2018). 'Go to hell you fagots, may you die.' Framing the LGBTQ subject in online comments. *Lodz Paper in Pragmatics*, 14(1), 69–92.
- Baider, F. (2020). Pragmatics lost? Overview, synthesis and proposition in defining online hate speech. *Pragmatics and Society*, 11(2), 196–218.
- Baider, F. (2023). Accountability issues, online covert hate speech and the efficacy of counter-narratives. *Politics and Governance*, 11(2), 249–260.
- Baider, F., & Gregoriou, A. (in press). Annotating linguistic features of extreme narratives: A literature review and a proposal. *Studii de Linguistica*, 14(2).
- Baker, P. (2014). Bad wigs and screaming Mimis: Using corpus-assisted techniques to carry out critical

- discourse analysis of the representation of trans people in the British press. In C. Hart & P. Cap (Eds.), *Contemporary critical discourse studies* (pp. 211–235). Bloomsbury.
- Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273–306. <https://doi.org/10.1177/0957926508088962>
- Bandini, E., & Maggi, M. (2014). Transphobia. In G. Corona, E. A. Jannini, & M. Maggi (Eds.), *Emotional, physical and sexual abuse* (pp. 49–59). Springer.
- Benesch, S. (2020). But Facebook's not a country: How to interpret human rights law for social media companies. *Yale Journal on Regulation Bulletin*, 38, 86–111.
- Bettcher, T. M. (2007). Evil deceivers and make-believers: On transphobic violence and the politics of illusion. *Hypatia*, 22(3), 43–65.
- Braddock, K., & Horgan, J. (2016). Towards a guide for constructing and disseminating counternarratives to reduce support for terrorism. *Studies in Conflict & Terrorism*, 39(5), 381–404.
- Brindle, A. (2016). *The language of hate: A corpus linguistic analysis of white supremacist language*. Routledge.
- Chakraborti, N., & Hardy, S. J. (2015). *LGB&T hate crime reporting: Identifying barriers and solutions*. Equality and Human Rights Commission.
- Chung, Y. L., Kuzmenko, E., Tekiroglu, S., & Guerini, M. (2019). CONAN—COunter NArratives through Nichesourcing: A multilingual dataset of responses to fight online hate speech. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 2819–2829). Association for Computational Linguistics.
- Cohen-Almagor, R. (2014). Countering hate on the internet. *Jahrbuch Für Recht Und Ethik/Annual Review of Law and Ethics*, 22, 431–443.
- Colliver, B., Coyle, A., & Silvestri, M. (2019). The 'online othering' of transgender people in relation to 'gender neutral toilets.' In K. Lumsden & E. Harmer (Eds.), *Online othering. Exploring digital violence and discrimination on the web* (pp. 215–238). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-12633-9_9
- Craig, S. L., McInroy, L., McCready, L. T., & Alaggia, R. (2015). Media: A catalyst for resilience in lesbian, gay, bisexual, transgender, and queer youth. *Journal of LGBT Youth*, 12(3), 254–275. <https://doi.org/10.1080/19361653.2015.1040193>
- Derecka, M. (2019). Manifestations of transphobia in computer mediated communication. *Discourse Studies in Polish Linguistics*, 14 (3), 101–123. <https://doi.org/10.4467/23005920SPL.19.016.11081>
- Entman, R. (1993). Framing toward clarification of a fractured paradigm. *Journal of Communication*, 10, 155–173.
- EU Council Framework Decision 2008/913/JHA of 28 November 2008. (2008). *Official Journal of the European Union*, L 328/55.
- Fairclough, N. (2015). *Language and power* (3rd ed.). Routledge.
- Fillmore, C. (1982). Frame semantics. In The Linguistic Society of Korea (Ed.), *Linguistics in the morning calm: Selected papers from SICOL-1981* (pp. 111–137). 137Hanshin.
- Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). *Countering online hate speech*. UNESCO Publishing. <https://unesdoc.unesco.org/ark:/48223/pf0000233231>
- Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics*, 31(7), 919–929.
- Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harvard University Press.
- Guillén-Nieto, V. (2023). The wording of hate speech prohibition: “You can’t see the wood for the trees.” In

- V. Guillen Nieto, D. Stein, & A. Doval Pais (Eds.), *From fear to hate: Legal-linguistic perspectives on migration* (pp. 173–200). de Gruyter. <https://doi.org/10.1515/9783110789157-008>
- Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhoft, A., Boucher, J., & Demirci, B. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences of the United States of America*, 118(50), Article 2116310118.
- Hill, D., & Willoughby, B. (2005). The development and validation of the genderism and transphobia scale. *Sex Roles*, 53(7/8), 531–544.
- Horawalavithana, S., Choudhury, N., Skvoretz, J., & Iamnitich, A. (2022). Online discussion threads as conversation pools: Predicting the growth of discussion threads on reddit. *Computational and Mathematical Organization Theory*, 28, 112–140. <https://doi.org/10.1007/s10588-021-09340-1>
- Howard, J. W. (2021). Terror, hate and the demands of counter-speech. *British Journal of Political Science*, 51, 924–939.
- Ibrohim, M. O., & Budi, I. (2023). Hate speech and abusive language detection in Indonesian social media: Progress and challenges. *Helyon*, 9(8), Article 18647. <https://doi.org/10.1016/j.heliyon.2023.e18647>
- Kecskes, I. (2006). On my mind: Thoughts about salience, context and figurative language from a second language perspective. *Second Language Research*, 22(2), 219–237. <https://doi.org/10.1191/0267658306sr266ra>
- Konikoff, D. (2021). Gatekeepers of toxicity: Reconceptualizing Twitter’s abuse and hate speech policies. *Policy Internet*, 13, 502–521.
- Legewie, J. (2013). Terrorist events and attitudes toward immigrants: A natural experiment. *American Journal of Sociology*, 118(5), 1199–1245.
- Lillian, D. (2005). Homophobic discourse: A “popular” Canadian example. *Sky Journal of Linguistics*, 18, 119–144.
- Love, R., & Baker, P. (2015). The hate that does not dare speak its name. *Journal of Language Aggression and Conflict*, 3(1), 57–86.
- Maynard, L. J., & Benesch, S. (2016). Dangerous speech and dangerous ideology: An integrated model for monitoring and prevention. *Genocide Studies and Prevention*, 9(3), 70–95.
- McInroy, L. B., & Craig, S. L. (2015). Transgender representation in offline and online media: LGBTQ youth perspectives. *Journal of Human Behavior in the Social Environment*, 25(6), 606–617.
- Mongie, L. (2016). Critical discourse analysis as queer linguistics: Religious pro- and anti-LGBT equality framing and counter framing in two letters to the editor in the City Press. *Stellenbosch Papers in Linguistics Plus*, 49, 23–43.
- Müller, P., Harrendorf, S., & Mischler, A. (2022). Linguistic radicalisation of right-wing and Salafi jihadist groups in social media: A corpus-driven lexicometric analysis. *European Journal on Criminal Policy and Research*, 28, 203–244.
- Nagoshi, J. L., Adams, K. A., Terrell, H. K., Hill, E. D., Brzuzy, S., & Nagoshi, C. T. (2008). Gender differences in correlates of homophobia and transphobia. *Sex Roles*, 59, 521–531. <https://doi.org/10.1007/s11199-008-9458-7>
- Norton, A. T., & Herek, G. M. (2013). Heterosexuals’ attitudes toward transgender people: Findings from a national probability sample of U.S. adults. *Sex Roles*, 68, 738–753. <https://doi.org/10.1007/s11199-011-0110-6>
- Rácz, P. (2013). *Saliency in sociolinguistics: A quantitative approach*. de Gruyter.
- Reddy, V. (2002). Perverts and sodomites: Homophobia as hate speech in Africa. *Southern African Linguistics and Applied Language Studies*, 20(3), 163–175.

- Reisigl, M., & Wodak, R. (2009). The discourse-historical approach (DHA). In R. Wodak & M. Meyer (Eds.), *Methods for critical discourse analysis* (pp. 87–121). Sage.
- Stonewall. (2023). *LGBTQ+ inclusion in primary schools*. <https://www.stonewall.org.uk/resource-topic/anti-bullying>
- Trump signs order to ban transgender people from US military. (2025, January 28). *Le Monde*. https://www.lemonde.fr/en/international/article/2025/01/28/trump-signs-order-to-ban-transgender-people-from-us-military_6737504_4.html
- Turner, L., Whittle, S., & Combs, R. (2009). *Transphobic Crime in the European Union*. Press For Change. https://www.ilga-europe.org/sites/default/files/transphobic_hate_crime_in_the_european_union_0.pdf
- Udupa, S., Gagliardone, I., & Hervik, P. (2021). *Digital hate: The global conjuncture of extreme speech*. Indiana University Press.
- Udupa, S., & Pohjonen, M. (2019). Extreme speech and global digital cultures. *International Journal of Communication*, 13, 3049–3067.
- UK military chiefs praise transgender troops. (2017, July 26). *BBC*. <https://www.bbc.com/news/uk-40733701>
- Ullmann, S., & Tomalin, M. (2023). *Counterspeech: Multidisciplinary perspectives on countering dangerous speech*. Routledge.
- van der Vegt, I., Mozes, M., Kleinberg, B., & Gill, P. (2021). The grievance dictionary: Understanding threatening language use. *Behavior Research Methods*, 53, 2105–2119. <https://doi.org/10.3758/s13428-021-01536-2>
- van Dijk, T. A. (1993). Principles of critical discourse analysis. *Discourse and Society*, 4(2), 249–283.
- Vedder, A. H. (2001). Accountability of internet access and service providers: Strict liability entering ethics. *Ethics and Information Technology*, 3(1), 67–74.
- Wachs, S., Krause, N., Wright, M. F., & Gámez-Guadix, M. (2023). Effects of the prevention program “HateLess: Together against hatred” on adolescents’ empathy, self-efficacy, and countering hate speech. *Journal of Youth and Adolescence*, 52(6), 1115–1128.
- Wang, J.-C., Widi Tamtama, G. I., & Latifah, R. (2024). Civil or uncivil: Seeing the role of user actors and providing comments on social media [Paper presentation]. *21st International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. Khon Kaen.
- Weinberg, G. (1973). *Society and the Healthy Homosexual*. Anchor.
- Woo, H., & Cho, Y. Y. (2023). Fighting hate and hate speech. *Societies*, 13(11), <https://doi.org/10.3390/soc13110240>
- Zottola, A. (2021). *Transgender identities in the press: A corpus-based discourse analysis*. Bloomsbury.

About the Author

Fabienne Baidier is a professor at the University of Cyprus. She works on semantics and discourse from a socio-cognitivist and contrastive perspective. Her research includes gender and language, conceptual metaphors, and emotions in political discourse and online communication, with a focus on hate speech. Her methodology includes corpus linguistics and critical discourse analysis. She has published internationally in linguistics and other disciplines. She has coordinated or is involved in EU social justice projects, focusing on the detection of hate speech and counternarratives.

The Italian Manosphere: Composition, Structure, and Functions of a Digital Network

Elisa Ignazzi ¹ , Mara Maretti ² , and Lara Fontanella ³ 

¹ Department of Neuroscience, Imaging and Clinical Sciences, G.d'Annunzio University of Chieti-Pescara, Italy

² Department of Legal and Social Sciences, G.d'Annunzio University of Chieti-Pescara, Italy

³ Department of Socio-Economic, Managerial, and Statistical Studies, G.d'Annunzio University of Chieti-Pescara, Italy

Correspondence: Mara Maretti (mara.maretti@unich.it)

Submitted: 9 September 2024 **Accepted:** 2 December 2024 **Published:** 9 January 2025

Issue: This article is part of the issue “Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment” edited by Max Römer Pieretti (Universidad Camilo José Cela), Beatriz Esteban-Ramiro (Universidad de Castilla-La Mancha), and Agrivalca Canelón (Universidad Católica Andrés Bello), fully open access at <https://doi.org/10.17645/si.i415>

Abstract

The digital sphere is pivotal in shaping social norms, and the Italian “manosphere” is a key player in this process. This study examines the composition and structure of the Italian manosphere, an intricate online ecosystem characterised by antifeminist and often misogynistic ideologies. Through a comprehensive analysis of Facebook networks and blog presentations from various groups, we mapped and classified the main actors within this ecosystem, shedding light on their connections and functions. The analysis focuses on two main aspects by employing natural language processing techniques and social network analysis. First, we investigated the functions of different groups within the network—Men’s Rights Activists, Men Going Their Own Way, Involuntary Celibates, and Pick-Up Artists—identifying their roles, how they interconnect and their ties to the international manosphere. Second, we analysed the blog presentations of members to explore the motivations driving individuals to join these communities, revealing the key themes emerging from their narratives. Our findings highlight the manosphere as a complex and interconnected phenomenon that not only reflects global neosexist trends but also integrates unique socio-cultural elements specific to the Italian context. This study underscores the significance of understanding the manosphere’s influence on public discourse and its far-reaching implications for the socio-political landscape in Italy, particularly concerning gender relations.

Keywords

digital sphere; Incels; misogyny; men’s rights activists; men going their own way; natural language processing; pick-up artists; sexism; social movements

1. Introduction

The growing proliferation of digital technologies has dramatically reshaped the public sphere, turning the internet into a fertile ground for the spread of both established and emerging ideologies, the shaping of social norms, and the polarisation of public opinion (Arora et al., 2022; Bernholz et al., 2021). Within this framework, the manosphere emerges as an online subculture that connects diverse social groups around a traditionalist gender ideology, one that is often characterised by sexism and anti-feminism, and advocating a form of hegemonic masculinity (Kennedy-Kollar, 2024). It has become a significant force in contemporary discourse on gender relations. The manosphere has been the focus of numerous studies exploring online misogyny (for a comprehensive review see Fontanella et al., 2024; O’Hanlon et al., 2024).

The manosphere is not monolithic but can be represented as a diverse archipelago (Ging, 2019) comprising various groups with differing perspectives and levels of violence, often linked to far-right, homophobic, and racist ideologies (Dickel & Evolvi, 2022). Despite their differences, all groups share the belief that feminism is inherently discriminatory and poses a threat to men (Farci & Righetti, 2019). Specifically, the manosphere upholds the concept of a “gynocentric order” and the Red Pill ideology, a metaphor from the film *The Matrix* that awakens the protagonist to reality upon taking the “red pill.” This metaphor not only reflects but also amplifies the ideological dynamics of the manosphere, becoming a powerful rhetorical tool that transcends the boundaries of individual groups to permeate broader cultural discourse (Nurminen, 2022). While these groups hold distinct beliefs, many employ the concept of “misandry,” a term denoting hatred of men, which serves ideological and community-building purposes. It reinforces a misogynistic worldview that frames feminism as a movement hostile to men and boys (Marwick & Caplan, 2018). The concept of misandry resonates across extremist misogynistic subcultures as well as more moderate men’s-rights groups. It allows these groups to co-opt the language of identity politics, positioning men as victims of “reverse discrimination” in political, economic, and social life, thereby justifying their sense of entitlement (Farci & Righetti, 2019). Discursive strategies are employed by the manosphere to position men as victims, portray women as monstrous others, and reassert traditional gender hierarchies, often with continuous references to rape in their rhetoric (Hopton & Langer, 2022).

The manosphere uses the internet not only as a content repository but also as a means to recruit and politically and socially mobilise its members, thereby directly influencing social and political dynamics (Banet-Weiser & Miltner, 2015; Hawley, 2017). Some scholars perceive the diverse realm of the manosphere as akin to a fully-fledged organised political movement capable of influencing public discourse and political elections (Nagle, 2017), whereas others emphasise the network’s dynamics characteristic of online communities of practice, where individuals congregate to exchange experiences and strategies rather than pursue explicit political objectives.

In English-speaking nations, specific segments of the manosphere overlap with far-right movements, such as the alt-right, leading to the significant politicisation and radicalisation of public discourse regarding gender issues (Hawley, 2017). This is especially apparent in the United States and the United Kingdom, where anti-feminist discourse frequently intersects with nationalist and xenophobic ideologies (Nagle, 2017).

In Europe, interest in these movements has emerged more recently, and, while a comparative, cross-cultural analysis of the aforementioned groups is intricate, it is possible to identify distinctions in their themes,

objectives and functions. The German manosphere, akin to its US and UK equivalents, seems more closely associated with far-right movements and nationalism, merging anti-feminist discourse with anti-immigration and xenophobic ideologies. In Germany, manospheric groups exhibit greater politicisation and are closely associated with populist movements (Ebner, 2017). In France, these groups appear to concentrate on the power dynamics within male–female relationships, emphasising the culture of seduction and the preservation of national cultural identity. French organisations recontextualise the prevailing masculinity in reaction to feminism (Gourarier & Benveniste, 2017). The discourse in Spain seems to centre on a robust critique of gender-equality policies in support of traditional family values. Spanish manospheric factions openly challenge legislation addressing gender-based violence and other governmental efforts to foster equal opportunities in a regime of male victimhood (García-Mingo et al., 2022; García-Mingo & Diaz-Fernandez, 2023; Núñez Puente et al., 2021). In Italy, the manosphere has discovered an atmosphere characterised by the emergence of groups and digital platforms that advocate narratives centred around the crisis of masculinity and an alleged feminist hegemony that threatens the conventional social structure (Vingelli, 2019).

Within this framework, this article presents research findings based on the following questions: What are the primary groups in the Italian manosphere that are active in the digital realm? What is the network’s structure, and what defines the connections among the different groups within this archipelago? What are the ideologically predominant groups? What is the demographic profile of their adherents? What themes are prevalent in the Italian manosphere? Ultimately, what functions do these groups serve: are they genuinely organised political pressure entities, or should they be regarded as online communities of practice that arise and develop in response to social discontent?

The following research activities were carried out to address these inquiries: (a) ethnographic observation of the online manosphere, with a particular focus on blogs, forums, and social-media platforms; (b) selection, classification, and analysis of the network arising from Facebook pages; (c) examination of the network structure and characteristics of the emerging groups; (d) typological analysis of participants in forums and blogs; and (e) identification of themes emerging from participant self-presentations in these blogs and forums.

The article is structured as follows. Sections 2 and 3 present definitions, concepts, and the context of the manosphere at both international and Italian levels. This is followed by a discussion of the materials and methodology used for the analysis of the network and profiles of members across the different subgroups. Next, the research findings are discussed, focusing on the roles played by the various groups and the emerging themes. Finally, the conclusions offer reflections on the social and cultural implications of the phenomenon. Graphical representations of the main findings are available in the online Supplementary Material.

2. The Manosphere Archipelago

Structurally, the manosphere can be categorised into four primary groups: Men’s Rights Activists (MRAs), Men Going Their Own Way (MGTOW), Involuntary Celibates (Incels), and Pick-Up Artists (PUAs).

MRAs champion men’s rights, asserting that feminism has resulted in systemic discrimination against men, especially in child custody, divorce, and workplace policies. MRAs use a personal action framework to create

a credible, albeit fictitious, narrative of male subjugation (Carian, 2022). Central themes of this narrative encompass the disparagement of women, the trivialisation or rejection of rape allegations, characterising #MeToo as a feminist conspiracy, representing men as victims, and promoting the reinstatement of patriarchal values (Dickel & Evolvi, 2022). This ideological project normalises, trivialises, and legitimises diverse manifestations of sexual violence against women (García-Mingo et al., 2022). MRAs' rhetoric regarding sexual violence contests gendered viewpoints on rape and sexual assault, alleging that feminists overlook male victimisation and characterising the feminist notion of rape culture as a moral panic (Gotell & Dutton, 2016). The MRA depiction of rape culture exemplifies a collective application of D.A.R.V.O. (deny, attack, and reverse victim and offender), a strategy that negates the issue, assails its proponents, and inverts the roles of victims and offenders (Scotto di Carlo, 2024).

“Incels” are men who perceive themselves as incapable of forming romantic or sexual relationships, blaming this predicament on a societal preference for attractive or high-status males. The term was first introduced in 1997 by “Alana,” the creator of a website aimed at single people of all genders and sexual orientations (Taylor, 2018). Two decades later, however, it was co-opted by a community of heterosexual men who self-identify as unattractive and who attribute their involuntary celibacy to women through dark humour, victimisation (Capelos et al., 2024), and the glorification of violence (Hopton & Langer, 2022; Scotto di Carlo, 2023). The group is notorious in the United States for its violent and misogynistic rhetoric, often devolving into hate speech directed at women (Nagle, 2017; Pilla & Dolce, 2019). While most Incels do not advocate violence, those who hold more extreme misogynistic beliefs are more likely to endorse violent actions as a way to gain attention, seek revenge, or achieve political change (O'Donnell & Shor, 2022). Participation in Incel forums tends to intensify their misogynistic views (Speckhard et al., 2021). The term “femoid” in Incel discourse exemplifies the dehumanisation of women, depicting them as monstrous entities, thus rationalising violence against them (Chang, 2020). Incels believe in a hetero-patriarchal racial hierarchy and justify their lack of sexual activity in ideas rooted in biological determinism and victimisation by women and feminism (Lindsay, 2022). “Femmephobia” reinforces this heteropatriarchal view of femininity and what Incels see as societal forces that push them into celibacy (Menzie, 2022). Their hierarchical view of men places attractive males at the top, excluding women from similar hierarchies and often judging them negatively (Heritage & Koller, 2020). Incel discourse employs symbolic, gendered actors (Menzie, 2022), such as “Stacy” for a highly attractive woman, “Becky” for a moderately attractive woman, and “Chad” for a dominant alpha male. Incels describe a sexual deficit affecting most men, criticise hyper-feminine women like “Stacy” for not meeting heteropatriarchal standards, and express jealousy and resentment towards more sexually successful men like “Chad.” Interestingly, Incels portray themselves in a self-deprecating manner, deviating from the typical “us vs them” dynamic (van Dijk, 1998), which fosters a cycle of self-pity and self-contempt that reinforces a sense of brotherhood within the community. Their perceived subordinate status is used to justify their misogyny and legitimise their degradation of women (Halpin, 2022).

MGTOW promote a separatist, anti-feminist agenda focused on self-sufficiency and individual empowerment, distancing themselves from women (Lin, 2017). The attitudes within this subculture are diverse, spanning from a total repudiation of sexual or romantic relationships with women to a commitment to monogamy as a safeguard against possible allegations of sexual misconduct. MGTOW perceive marriage and relationships as instruments of female subjugation of men (De Gasperis, 2021). Ricci and Farci (2021) assert that this group embodies a radical response to modernity and women's liberation. A faction of MGTOW adheres rigidly to the “Pence Principle,” a notion ascribed to former US Vice President Mike Pence by Randall Bentwick in his

2018 bestseller, *The Pence Principle: Lessons All Men Must Learn from Ford-Kavanaugh* (Lacalle et al., 2023). This principle encompasses directives, such as refraining from meeting a woman one-on-one and abstaining from consuming alcohol at social events without the presence of one's spouse. The MGTOW ideology is associated with hegemonic and heteronormative toxic masculinity, and its rhetoric spreads and legitimises misogynistic beliefs through online harassment. While the misogyny and violence endorsed by MGTOW are not overtly extreme, their appeals to rationality make these views appear commonsensical (Jones et al., 2020). Hunte (2019) identifies the key traits of MGTOW, including the use of linguistic practices that foster a shared identity and community, a narrative of victimisation, and the exchange of commonly held beliefs. The internal dynamics of MGTOW forums demonstrate that, while their discourse is highly misogynistic, it also promotes a strong sense of community, strengthening group bonds and participation (Wright et al., 2020). MGTOW discussions predominantly revolve around two themes: women—often in a deeply misogynistic context—and defining MGTOW, both as a collective identity and a personal journey (Wright et al., 2020).

PUAs consider sexual relationships a conquest game (Scotto di Carlo, 2023). This community disseminates techniques to assist men in attracting multiple women at their discretion. It has faced academic examination due to its ostensibly complex group-building lexicon and use of neurolinguistic programming theories (Hopton & Langer, 2022). PUAs concentrate on methods for enticing women, frequently using contentious tactics to augment their perceived appeal and emphasising the notion of “game,” aimed at enhancing male attractiveness. These factions are notably active on YouTube and other video-sharing platforms, providing “guidance” on how to coerce women into sexual relationships (Kennedy-Kollar, 2024). Men's online PUA advice platforms and forums depict casual sex with numerous women as a crucial element of enhancing social status and esteem. In their analysis of online PUA content, Cosma and Gurevich (2020) identify three primary interpretive frameworks. “Embattled masculinity” employs themes of defensiveness and conflict to protect male privilege through the covert pursuit and sexual dominance of women. “Feminine commodities” views women's bodies as symbols of masculine achievement. “Pressured pursuit and consent as control” portrays men as authorities in sexual affairs, possessing the responsibility and power to navigate and surmount female consent as a hindrance. The quest for sexual gratification is paramount, yet allusions to the intrinsic worth of women are conspicuously lacking or entirely disregarded.

3. The Manosphere in Italy

The manosphere in Italy is a recent phenomenon that has rapidly expanded over the past decade, primarily due to the emergence of online and social-media platforms that enable the creation and consolidation of virtual communities (Cannito et al., 2021). Recent studies indicate that the Italian manosphere is interconnected with global trends while simultaneously exhibiting distinct characteristics influenced by the local sociocultural context (Ricci & Farci, 2021; Vingelli, 2021). The expansion of the manosphere in Italy originated in the early 2010s, characterised by the rise of blogs, forums, and Facebook groups advocating anti-feminist and traditionalist ideologies. The Italian manosphere groups mirror significant communities initially established in North America (the US and Canada), each possessing a distinct agenda yet interconnected, all unified by a pervasive anti-feminist and sexist ideology. Though these four groups serve different functions within the manosphere, they are united by the philosophy of the Red Pill, sharing the belief that women are manipulative and that men must “dominate” them to succeed in relationships (Longo, 2020).

Research on Italian MRAs suggests that these groups maintain a robust online presence, advocating a narrative of “male oppression” that often deteriorates into misogyny (Vingelli, 2021). Farci and Righetti (2019), in their examination of the Italian MRA network on Facebook, determined that, although men’s-rights groups comprise a variety of organisations, two predominant trends prevail within the network. The first trend pertains to pages dedicated to topics like violence against men and the rights of fathers and children, primarily appealing to those who have undergone separation and divorce. These groups typically promote conservative ideologies and exhibit a more militant form of antifeminism, frequently directing reactionary and antagonistic rhetoric towards women. The second trend encompasses antisexist platforms that highlight the challenges of masculinity, primarily attracting younger men who perceive themselves as disempowered by recent changes in gender dynamics. These groups exhibit a seemingly progressive position, directing their antifeminism primarily towards feminists rather than women in general. Cannito and Mercuri (2022) conducted an exploratory content analysis of an Italian non-resident fathers’ online forum’s Facebook group and page, revealing that fatherhood operates as an “empty box” wherein fathers’ rights are strategically utilised to bolster hegemonic masculinity, rationalise gender-based violence and disseminate antifeminist and anti-women ideologies. Moreover, these organisations employ the notion of fathers’ rights as a mechanism for political advocacy, frequently aligning with right-wing parties.

Dordoni and Magaraggia (2021), in their examination of the interactions, representations, and discourses concerning gender identities and masculinity paradigms within the Italian incel community, discerned two interrelated themes: (a) aesthetics, frustration, and a narrative of self-victimization and (b) the objectification of, the dehumanisation of, and violence perpetrated against women. Their findings indicate that violence is fundamentally embedded in the language and imagery of this virtual community, constituting a core element of the new masculinity models that this manosphere group establishes and sustains. Cannito and Ferrero Camoletto (2022) analysed the fundamental principles of the Red Pill and LMS (look, money, status) theories as presented in Italian Incel forums, demonstrating that these theories are intrinsically contradictory. They endeavour to integrate socio-economic and biological explanations for sexual relations while maintaining a rigid interpretation of gender rooted in alleged natural inclinations. Furthermore, their findings suggest that these theories propagate misogynistic narratives, attributing men’s dissatisfaction to women while simultaneously revealing intragender dynamics. This engenders tensions within the manosphere, as these theories establish hierarchies of masculinities and provoke enquiries regarding the possibility for individuals to go beyond the biological determinism linked to the Incel identity.

Within the Italian context, MGTOW and PUAs have garnered relatively limited scholarly focus. The manosphere is frequently analysed alongside the more notable factions of MRAs and Incels, yet the ideologies and practices of MGTOW and PUAs merit additional scrutiny. Understanding MGTOW and PUAs is crucial for acquiring a thorough perspective of the intricate and varied landscape of the manosphere in Italy.

4. Methods and Materials

4.1. Digital Ethnography

The reconstruction of the digital landscape of the manosphere in Italy initially uses observational techniques within the framework of digital ethnography. Digital observation facilitates the systematic and structured

monitoring and analysis of interactions and behaviours within digital environments, including social media, online forums, and virtual communities. This methodology enables the acquisition of qualitative data regarding the relational, cultural, and social dynamics that arise online by immersing oneself in digital environments (Hine, 2015). In contrast to conventional observation, digital ethnography requires the capacity to examine texts, images, and various forms of technology-mediated communication (Kozinets, 2010; Postill & Pink, 2012).

4.2. Reconstructing the Italian Manosphere Networks on Facebook

We focused on Facebook pages because the “like” mechanism enables the reconstruction of interaction networks among pages. Furthermore, the public nature of Facebook pages often makes their content accessible for content analysis.

The research employed a snowball sampling strategy to systematically identify and map Facebook pages within the manosphere community. This approach allowed for the gradual expansion of the network by starting with a set of seed pages and identifying additional related pages through interactions between them. The process began with the selection of an initial collection of public Facebook pages (seed nodes), determined by their relevance to the study and their established connections to manosphere communities. We meticulously selected initial seed nodes from established manosphere communities, focusing on prominent Facebook pages that represented MRAs, PUAs, Incels, and MGTOW. The sampling grew by recursively identifying additional connected pages from these seeds, which we accomplished by monitoring interactions, such as likes, comments, mentions, and page follows, which facilitated the mapping of the extensive network of associated pages. We repeatedly conducted the iterative process until the network’s expansion reached a state of equilibrium. This recursive identification facilitated the network’s organic expansion, uncovering broader connections within the manosphere. The collection process was carried out in August 2024.

Upon the establishment of the network, the Facebook pages were categorised into the four groups that make up the manosphere. The categorisation process involved a detailed review of each page’s content to ensure their accurate alignment with specific manosphere groups. This review included examining page descriptions, posts, discussions, and community guidelines. If a page’s description and the tone or themes of its posts and discussions were consistent with the beliefs or rhetoric typical of a specific group, it was categorised accordingly. Any pages that could not be reliably assigned to a specific category using these criteria were grouped under the “other” category to preserve the clarity and coherence of the classification. Additionally, pages related to equally shared parenting (ESP) movements were clustered into a separate group to distinguish their unique focus within the broader categorisation framework.

We generated a graph to illustrate the connections and relationships among the Facebook pages. This visual depiction elucidated the interactions among pages, illustrating the links between different groups of the manosphere and offering a comprehensive overview of the structure and relationships within the manosphere on Facebook.

To analyse the structure of the manosphere network on Facebook, various centrality metrics were calculated that assessed the significance and influence of each page. *Indegree centrality* measures the number of incoming connections a page has, indicating its authority within the network. Pages with high indegree are

often referenced or followed by others, signifying their importance. *Outdegree centrality* measures outgoing connections, reflecting a page's activity or engagement with others. Pages with high outdegree serve as hubs, facilitating the flow of information. PageRank is a more comprehensive metric that considers both the quantity and quality of incoming links, identifying pages with significant influence by their connection to other influential pages. These metrics help identify key actors, content disseminators, and authorities within the manosphere community.

4.3. Profiling Members of Manosphere Forums

We conducted an analysis of the profiles of members from four distinct manosphere forums linked to MRAs, MGTOW, Incels, and PUAs. The textual content was manually extracted in August 2024. Specifically, we gathered all self-presentations that were present on the forums and blogs at that time, regardless of when they were originally produced. This analysis focused on the characteristics expressed in the members' self-presentations, such as personal background, motivations for joining the forum, and views on gender relations. Additionally, we identified key thematic patterns by examining the characteristic words and phrases commonly used in these self-descriptions. This approach allowed us to capture both the individual traits of the members and the linguistic markers that distinguish each subgroup within the manosphere. The study of linguistic features used keyness analysis (Gabrielatos, 2018), a statistical method designed to identify words or phrases, referred to as "keywords," that appear significantly more often in one sub-corpus compared to the other sub-corpora. A keyness score was determined through the chi-square test. This score evaluates whether the actual frequency of a word in the target corpus significantly deviates from the expected frequency derived from the reference corpus. We used a statistical significance threshold (p -value for the chi-square test below 0.01) to identify significantly over-represented words in each subgroup, thereby elucidating distinct linguistic patterns within forum members' presentations.

5. Results

5.1. Digital Ethnography Findings

The ethnographic investigation across various social-media platforms (Facebook pages, blogs, forums and YouTube channels, TikTok, Discord, and Instagram) revealed that the manosphere is present on all major social networks. Facebook, however, despite being a platform with a predominantly older user base, is particularly well suited for network reconstruction due to the presence of public pages and more extensive narrative discussions. Moreover, Facebook's structural design makes it easier to trace connections between pages, allowing for a more comprehensive mapping of relationships within the manosphere. The structure of the network is highly interconnected, with a significant degree of cross-posting and hyperlinking between the different nodes, which strengthens the digital ecosystem of the manosphere. From this phase of ethnographic observation, the Italian manosphere network was reconstructed through snowball sampling.

From the observation and annotation of posts and comments found particularly on blogs, forums, and Facebook, we can confirm what has already emerged from several studies conducted at the national level (De Gasperis, 2021; Pilla & Dolce, 2019; Ricci & Farci, 2021; Semenzin & Bainotti, 2020), which highlight the main themes characterising this phenomenon. Among these, the following stand out: anti-feminism, the "crisis of masculinity," men's rights, sexism, gender-based violence, and male victimhood—all elements that

reflect the central concerns of the various groups, serving as an ideological background. Anti-feminism is a prevalent theme and manifests as a violent reaction against the progress of the feminist movement, which is perceived as a direct threat to men's position in society (Pilla & Dolce, 2019). The narrative of the "crisis of masculinity" (Ciccione, 2017) accompanies this theme, portraying men as victims of a "feminisation" process that diminishes traditionally masculine qualities. Sexism and gender-based violence are equally central, with groups like Incels and MGTOW portraying women as manipulative and inferior, thus justifying aggressive behaviour towards them (De Gasperis, 2021; Pilla & Dolce, 2019). Finally, male victimhood is a recurring narrative in which men perceive themselves as oppressed by a social system that favours women, a view particularly upheld by Italian MRAs (Ricci & Farci, 2021; Semenzin & Bainotti, 2020).

Ethnographic observation of blogs, forums, and social networks allowed us to identify Facebook as a useful platform for analysing nodes (like pages) and the connections among them to reconstruct the structural characteristics of the Italian manosphere.

5.2. Italian Manosphere Networks on Facebook

Ethnographic analysis allowed for the initial identification of seed pages: that is, initial nodes known to belong to the manosphere, such as "Vita da Brutto—The Red Pill Never Dies" for Incels, "MGTOW Italia" and "Coscienza Maschile" for MGTOW, "Diritti Maschili—Equità e Umanità" for MRAs, and "Essere Uomo" for PUAs. Through interactions like mentions, likes, comments, and follows, we identified other connected pages from these initial nodes. This methodology allowed us to map the connections among the various pages that belong to different categories of the manosphere.

The result is a complex and interconnected network composed of 158 nodes whose classification highlighted a significant presence of pages related to MRAs (39%) and MGTOW (27%). Incels and PUAs both have a lesser presence (5% and 3%, respectively). Various pages (18%) were identified as mixed pages, which share themes related to the rights of separated fathers (ESP). This constellation, while focusing attention on legal issues related to child custody and the conditions of separated or divorced fathers, conveys topics from the manosphere. Furthermore, we classified 8% of the retrieved pages as "others."

The network analysis reveals a complex structure (see Figure 1) in which different pages play specific roles, contributing to the cohesion and dissemination of these groups' ideologies. Additional graphical representations are included in the Supplementary Material, along with the centrality measures for the most significant pages within the network.

MRAs' Facebook pages seem to be the most central, exhibiting elevated indegree, authority, and PageRank values (see centrality measures graph and tables in the Supplementary Material). This signifies that these pages not only acquire numerous connections but also attract links from reputable and substantial pages. Websites like *Il Maschicidio*, *Diritti Maschili—Equità e Umanità*, and *Smascheriamo il Femminismo* serve as authoritative entities within the network, indicating their role as primary sources of information and ideology in the Italian manosphere. Their prominence underscores that the MRA movement both is influential and serves as a benchmark for other pages and groups within the manosphere network.

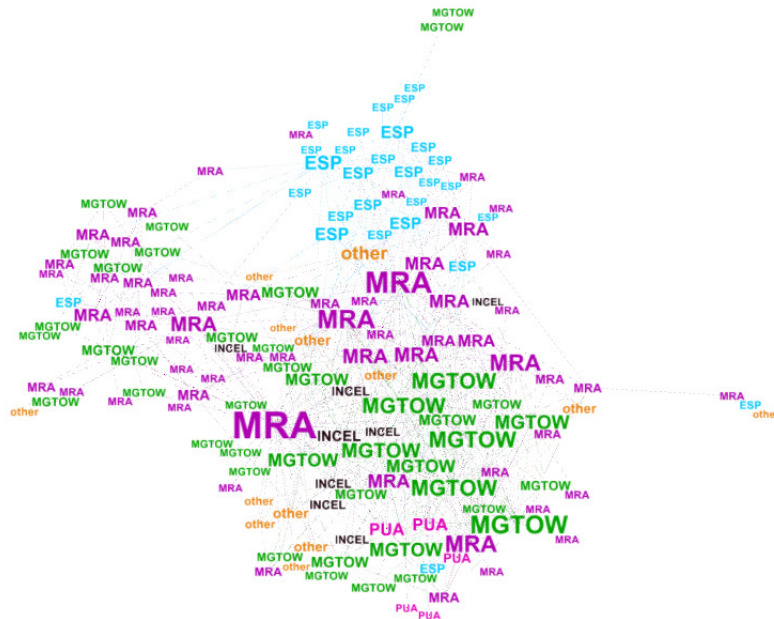


Figure 1. Network of Facebook pages related to the manosphere. Node labels depict manosphere groups, and their size is proportional to node degree. The network is visualised through Gephi (<https://gephi.org>) with MultiGravity Force Atlas 2 algorithm (LinLog mode).

MGTOW pages function as gateways within the network, as evidenced by their elevated outdegree values: 12 out of 20 pages with the highest hub scores belong to this category. Therefore, MGTOW pages are notably effective in linking various segments of the network, serving as hubs that promote the distribution of content among otherwise isolated groups. These pages are essential in consolidating the manosphere, facilitating the exchange of ideas among diverse factions, and strengthening the sense of unity within the movement.

Incel pages hold a distinctive status within the Italian manosphere. Despite their lack of prominence in centrality measures, these pages are critical in user radicalisation. Incels frequently function on the periphery of the network, yet their discourse serves as a catalyst for individuals already inclined towards radical ideologies. This group, although less interconnected than MRAs or MGTOW, is crucial for comprehending the internal dynamics of the manosphere, as it appears to serve as an entry point for individuals. We should not interpret its marginal position as a sign of irrelevance but rather as a manifestation of its unique role within the network: It attracts individuals who may later engage with the movement's more central or radical discourses. The Incel community frequently comprises young men who find it challenging to connect with women, and, upon entering the manosphere, they may exacerbate their feelings of frustration, thereby deepening their involvement in the network. The Incel network functions as a “filter” that facilitates individuals’ entry into the manosphere and, in certain instances, propels them towards more extreme viewpoints. This phenomenon of radicalisation, although less pronounced in Italy, has been noted in the United States and Canada. The insular and self-perpetuating dynamics of these groups enable a milieu where radical ideas can proliferate unchallenged, due to weak external connections, as indicated by low indegree.

PUA pages exhibit relatively low centrality measures. This suggests that, while they are part of the manosphere network, PUA pages have limited influence and do not serve as central nodes within the Italian manosphere.

Their presence, however, is not irrelevant, as they act as specialised niches within the network, focusing on specific themes related to seduction and the manipulation of interpersonal relationships.

The Italian network exhibits a robust affiliation with content and pages from global entities: Of the 158 pages, 36 are not based in Italy. Furthermore, Figure 2 illustrates that a substantial segment of the Italian manosphere is linked to international Facebook pages that predominantly disseminate content from the American manosphere.



Figure 2. Network of Facebook pages related to the manosphere. Node label colour indicates the origin of the page: red labels represent Italian pages, while blue labels represent international pages not based in Italy. Node label size is proportional to node degree. The network is visualised through Gephi (<https://gephi.org>) with MultiGravity Force Atlas 2 algorithm (LinLog mode).

5.3. Profiles of Italian Manosphere Forum Participants

The examination of blogs and forums within the Italian manosphere has facilitated the categorisation of participants across diverse groups and corroborated the insights derived from ethnographic analysis and network analysis on Facebook concerning the roles of the various groups. The sample consists of 695 self-presentations made on blogs and forums. Among these, 309 (44%) originate from Incel groups, 187 (27%) from MRAs, 152 (22%) from PUAs, and 47 (7%) from MGTOW.

We identified demographic characteristics, including gender and age, from the self-presentations. Among the participants, 9.9% are female, with the lowest proportion of women (4.3%) noted among MGTOW individuals. Women occasionally demonstrate interest in the movements, frequently influenced by familial experiences involving sons, husbands, or brothers who encountered instances of “discrimination.” Women’s participation is frequently motivated by curiosity and academic pursuits, primarily in disciplines like anthropology and psychology, or aimed at providing or soliciting counsel and sharing personal narratives rather than partaking in a formal ideological debate, as is typical among men. On the other hand, female proponents of manosphere ideologies articulate dissatisfaction regarding the absence of a dedicated platform on which women can exchange experiences and narratives with others who comprehend or have

endured analogous circumstances. Analysis of the self-presentations by women reveals that they generally adopt a more cautious and less confrontational demeanour than men. There is a prevalent inclination to apologise or request permission prior to speaking, suggesting a perceived minority status or a belief that their contributions may not conform to the prevailing tone of the discussions.

The Incel group has the lowest average age of 27 years, whereas MGTOW groups, which typically consist of separated or divorced men who have withdrawn from stable relationships, have the highest average age of 36 years. The mean age of participants in PUA groups is approximately 32 years, whereas, for MRAs, it is about 33 years.

The qualitative analysis of self-presentations, coupled with the findings from the keyness analysis that facilitated the reconstruction of the vocabulary employed by members of diverse groups (see Figure 3), uncovers notable disparities among the subgroups regarding the topics addressed, the manner in which users identify with and engage in these communities and the interactions and discourse between members of different groups and users from other groups.

5.3.1. MRAs' Self-Presentations

MRAs set themselves apart from other manosphere groups through their recognition as a social and cultural movement, their use of formal and sophisticated language, their frequent integration of Latin quotations, and their generally elevated rhetoric, all of which lend legitimacy and authority to their assertions. MRAs primarily concentrate on perceived disparities detrimental to men, particularly in domains like family law, the labour market, and criminal justice. For instance, the self-presentations include the following statements:

Dear friends, let's form an "alliance." We resist because we are the axis mundi, the beacon, the reference point, the compass. We men.

Many men's movements do excellent work, but they accept the rules of a futile and gynocentric society built on foundations such as divorce and mass abortion (both subject to female discretion), and empty, commodified human relationships. To improve the condition of men, it is not enough to reform a few laws (which are useless in a feminist system, as seen with shared custody) but it is necessary to change the rules of an entirely anti-male system.

Today, Western society seems determined to throw away, in a very short time, all or nearly all of the progress, the achievements of civilisation, rights, freedom, and humanity gained over the last two thousand years; not by chance, these achievements are due to men; they are masculine achievements. It is no coincidence that men and everything masculine are now despised, blamed, denigrated, and ridiculed.

MRAs keywords encompass: feminism (*femminismo*), man (*uomo*), woman (*donna*), equality (*uguaglianza*), masculine (*maschile*), rights (*diritti*), privileges (*privilegi*) and violence (*violenza*; refer to Figure 3). Although they harbour animosity towards feminism, their methodology seems less confrontational and more organised than that of MGTOW, focusing on legal and institutional challenges rather than a complete repudiation of relationships and social interactions.

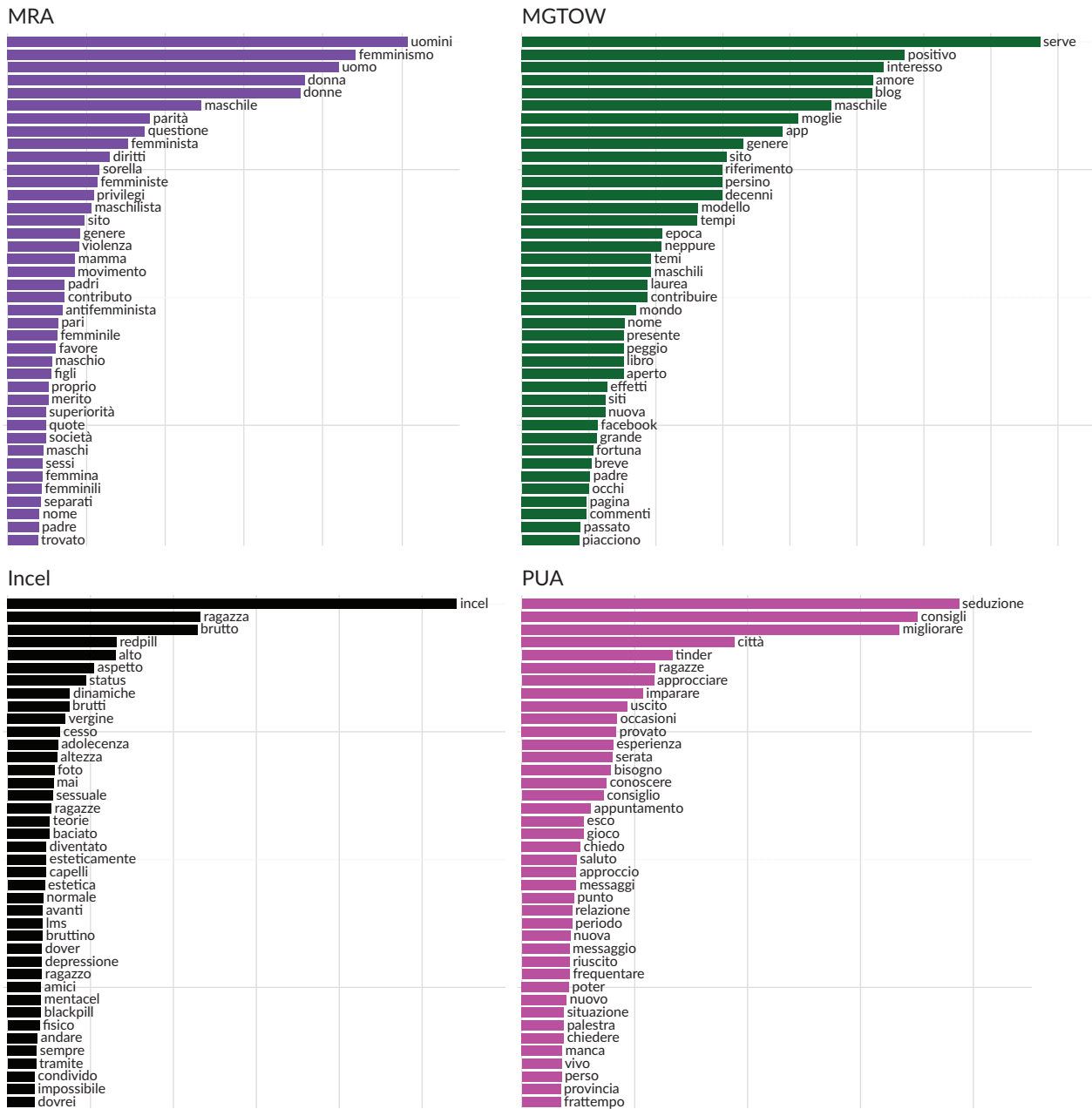


Figure 3. Keywords used by the members of manosphere forums.

5.3.2. MGTOW's Self-Presentations

MGTOW do not convey frustration regarding limited access to relationships; instead, they voluntarily reject conventional romantic dynamics. Their rhetoric, characterised by strong autonomy, is imbued with contempt for interpersonal connections and emotional engagement, perceived as a snare for men. The keyness analysis (see Figure 3) substantiates this interpretation, as the emerging keywords pertain to the realm of relationships between men and women (*amore*—love, *maschile*—masculine, *genere*—gender, *moglie*—wife).

This community exhibits a diverse array of experiences: Some individuals identify as single by choice, while others are engaged, married, or divorced, yet they all adhere to the principles of emotional detachment and self-sufficiency:

I have been “battling” with the fair ladies for many years, thankfully not in an extreme form. However, I can’t help but notice that my personal freedom is always called into question by them. Fortunately, I don’t have any draining family histories behind me, and honestly, I hope to never have any in the future.

In my life, I have had a few long-term relationships (LTR) and some brief flings, and at the moment, I am single. After the last and, as always, disastrous LTR that happened a few years ago, I decided to stop, execute the command “format C:” in my brain, and reinstall a new, updated, and “clean” operating system.

The result was only to understand how I ended up in the trap of marriage (I am now divorced). The greatest disappointment was discovering that all the abuses I suffered were not random, but rather the result of a narrative that sees men as disposable objects.

My luck was having a truly womanly wife, intelligent, who sacrificed her job for family and children and who always shared with me the male issue while dissociating herself from feminism in all its forms. On this occasion, she encouraged me to support the gender issue by recognising the now-commonplace misandry.

MGTOW forum members often employ misogynistic language in their presentations, belittling women and reinforcing negative stereotypes. This language not only reflects their frustrations but also fosters a sense of camaraderie among members, as they bond over shared grievances and reinforce their ideological beliefs. The effect of this rhetoric can be significant, as it contributes to a toxic environment that discourages healthy relationships and promotes a divisive view of gender dynamics:

When I pointed out their impulsiveness (not emotional as the common narrative would suggest, but impulsive) and frivolity, I was told that I hated women and that I couldn’t understand them. Alas, it seems to me that understanding women is quite simple: They are beings driven by convenience, often exhibiting distorted behaviours and acting as social climbers.

I found myself immersed in an environment filled with cuckolds of various kinds who worshiped women and reacted violently to anyone who dared to show that they also have (and perhaps only) flaws.

One must be truly foolish to fall in love with women in today’s world!...And anyway, if someone absolutely can’t do without them, they should realise that a dog can give more pure and sincere affection than a woman ever could!

5.3.3. Incels' Self-Presentations

The examination of Incel self-representations reveals that Incels frequently act as the principal gateway into the manosphere. These individuals present themselves as victims of an unfair social system focused on perceived “discrimination” against less attractive or economically and socially disadvantaged men:

When I message girls on Instagram or Tinder, they ghost me before we even meet up, or they disappear after the first date where nothing ever happens. I would give everything I don't have to be a GigaChad and get tons of them pregnant.

They suddenly became cold and hostile, started making up excuses and commitments that never existed, and a few days later I'd find out they were hooking up with the latest Chad in line.

Themes like the “friendzone” (perceived as an injustice committed by women who fail to reciprocate these men's romantic affection) and casual relationships (often viewed as failures or elusive pursuits) centre their personal narratives:

Not ugly, but completely inadequate. And girls could sense this; after a while, they would either run away or friendzone me with no chance of anything more.

These users frequently encounter bullying and perceive societal rejection or neglect by women, exacerbating their frustration:

I am a guy affected by nervous tics, without relationships, experiencing maladaptive daydreaming, and continuously haunted by images of my past life as a bullied boy.

In short: a life destroyed by bullying experienced in adolescence. All my problems stem from that.

I managed to get just one kiss at the age of 17 with a classmate of mine, and then when I tried to ask her to hang out alone, she rejected me like I was the lowest of the low.

Notably, numerous individuals self-identify as Incels despite being in stable relationships or married with children. This highlights the idea that involuntary celibacy may be more of a psychological condition than a reality—a mindset that influences even those who have attained what Western society considers significant success, i.e., establishing a family:

Although I am married (with a son), I consider myself an Incel.

I am 55 years old, have been married for 22 years, but in reality, it feels like I am alone; I have always been alone and always will be.

The Incels' vocabulary (see Figure 3) reveals that aesthetic beauty (*alto*—tall, *brutto*—ugly, *cesso*—piece of trash, *bruttino*—unattractive, *fisico*—physique, *estheticamente*—aesthetically), socioeconomic status (*amici*—friends), sexuality and relationships with girls (*baciato*—kissed, *sessuale*—sexual, *vergine*—virgin),

antifeminism (*femminismo*—feminism, redpill, blackpill), and personal and social distress (*depressione*—depression) are the main themes emerging from the presentations.

5.3.4. PUAs' Self-Presentations

PUAs advocate techniques and offer advice for “seducing” women, grounded in a heavily stereotypical perspective of interpersonal dynamics. Users portray themselves as “hunters,” using overtly misogynistic language that reduces women to mere targets for manipulation. In contrast to Incels, who perceive relationships as inaccessible, PUAs assert that they can master them through acquired skills and deliberate strategies. Nonetheless, their writings indicate a desire for community validation and social endorsement, as numerous users appear to pursue affirmation of their masculine value through relational and sexual achievements:

On Tinder, I usually give 25/30 likes a day on days when there's a lot of activity.

Before showing my true self, my character, I have this damn need for other people's approval, so I need to be sure that a girl likes me before I make a move and show my best self. If I'm not sure, I freeze.

I joined this forum hoping to regain a winning mindset and improve my seduction skills with women.

I'm planning to dive deep into the texts that could help me the most—like six months of in-depth study—to at least have a better theoretical understanding of what needs to be done.

So, I hope to find in this forum a place where I can sharpen my seduction skills.

You helped me get back on my feet by giving me some foundations that are now part of my daily life, and I can't thank you enough.

The prominence of the theme of seduction and manipulation for seductive objectives is apparent from the keyness analysis findings. Terms like advice (*consigli*), learn (*imparare*), approach (*approcciare*), improve (*migliorare*), and especially seduction (*seduzione*) frequently emerge in introductions on forums and blogs associated with Italian PUAs.

6. Discussions: Social Movements, Political Movements, or Communities of Practice?

Based on the results, we can delineate the characteristics of the different groups within the Italian manosphere. Like the global arena, it presents characteristics that define both social and political movements. There is, however, a broad consensus that it is mainly perceived as a social movement with significant political implications. The factions constituting the manosphere typically coalesce around narratives of opposition and resistance to societal transformations, especially regarding women's rights and feminism. This positioning portrays them as social agents engaging in the dynamics of conflict and identity-driven assertions, typical of social movements. The Italian manosphere exhibits a political bias, particularly when its members use language that reflects radical right-wing rhetoric or aligns with populist movements. Thus, it functions as an “echo chamber” for far-right political ideologies and, in the Italian context, incorporates aspects of the radical left, thereby exacerbating the polarisation of public discourse.

An alternative approach to understanding the manosphere is to view it as a constellation of “communities of practice,” a term that denotes groups of individuals who share a common interest and engage in regular interactions to enhance and refine skills relevant to that interest.

The digital realms of the manosphere enable the formation of homosocial communities, wherein users exchange experiences, language, and practices that bolster their shared identity. The use of specific language and practices, including the generation and dissemination of sexist memes as instruments of internal communication, strengthens the group’s cohesion. In many instances, engagement within the manosphere prioritises the cultivation of a collective culture that reinforces and magnifies the prevailing beliefs of its constituents over explicit political objectives.

Virtual Communities of Practice (VCoPs) are online communities that emerge either organically or through deliberate design, as distinguished by the exchange of shared interests, experiences, or challenges among participants. These virtual groups offer digital environments for individuals to collaboratively learn, exchange knowledge, and acquire new skills. In contrast to conventional communities of practice, VCoPs are not constrained by geographical proximity; rather, they use digital platforms to connect their members, fostering networks of support and collaboration in the digital realm (Maretti et al., 2022). The notion of communities of practice was initially formulated by Lave and Wenger (1991) and has subsequently been modified for analogous structures established in digital environments. These virtual communities coalesce around shared issues, enabling participants to exchange experiences and resources and thereby enhancing collective knowledge (Wenger, 2011). VCoPs stand out for their ability to foster informal learning and the development of social capital, functioning beyond the confines of formal, institutionalised organisations (Jiménez-Zarco et al., 2015). A fundamental characteristic of VCoPs is their capacity to adapt and address the evolving needs of their members. These communities frequently serve as instruments for resilience, particularly in crisis or emergencies, where they facilitate collective well-being and information management (Amaratunga, 2014; Lucini, 2014).

We can identify several essential traits and characteristics of VCoPs. Initially, these communities arise from common issues, often stemming from a collective need among the members. This cultivates an initial rapport among participants, encouraging a cooperative atmosphere. Secondly, VCoPs operate as informal mutual-support entities, offering emotional and practical assistance as members exchange experiences and solutions to shared challenges. Third, they emphasise the exchange of practices and knowledge, encouraging the collective enhancement of social capital within the group. Moreover, members possess a common value system that directs their interactions, enhancing group cohesion and collective identity through a shared culture and language. VCoPs foster the development of social capital by cultivating trust-based relationships that may transcend the virtual realm and spread to real-world contexts. Ultimately, they possess the capacity to generate opinion bubbles, reinforcing homogeneous viewpoints while potentially constraining exposure to alternative perspectives. These characteristics define VCoPs as dynamic settings for collective development and cooperation.

Despite the absence of specific literature, research suggests that most nodes within the manosphere can be characterized as VCoPs, sharing common ideologies and methodologies to reinforce their identities. These virtual communities function analogously to the established VCoPs in the educational, social, and health domains, emphasising knowledge dissemination, collaborative learning and the development of a shared

culture through consistent online engagement (Ging, 2019; Horta Ribeiro et al., 2021; Van Valkenburgh, 2021). The VCoPs in the manosphere serve as venues in which members convene, articulate, and sustain particular narratives of masculinity, frequently in contrast to conventional values. For example, participants in groups like MGTOW, Incels, PUAs, and communities of separated or divorced fathers convene online to exchange experiences, adopt a shared vocabulary and formulate internal guidelines for interaction that bolster their perspective on the male role in society. Consequently, the manosphere operates as a collection of online communities of practice that enhances the resonance of shared ideologies through the community-reinforcement mechanisms characteristic of VCoPs. Using Papacharissi's (2015) definition of "affective publics" as story structures linked by emotions and personal experiences, the manosphere's nodes appear and disappear through emotional connections, using and solidifying stories of individual pain to create an affective consensus about a claimed shared gender experience—specifically, men's perceived diminished status in society due to feminism (Ging, 2019).

It would be reductive, however, to limit the manosphere phenomenon exclusively to VCoPs, as it can be regarded as a political movement with global affiliations associated with the conservative political realm: the alt-right in the United States and various European contexts, as evidenced by numerous studies on conservative right-wing factions, or, in the Italian context, even with a segment of the conservative left. An underlying ideology, extensively documented in the literature and this research, crystallises into a central tenet of the antifeminist narrative, specifically within MRA groups. These groups clearly define specific objectives to achieve full parity between men and women. These objectives include organising campaigns against violence towards women and men, protecting paternal rights, abolishing alimony for ex-wives, imposing stricter penalties for sexual crimes, eliminating compulsory military service for men, and supporting male school dropouts. Additional goals include raising awareness of male victims of trafficking and exploitation, ensuring media coverage for men's rights issues, and campaigning for equal representation in employment and public spaces. There is also a focus on improving men's health and promoting psychological support for victims of workplace discrimination.

The assertions and organisational framework of the manosphere, along with the requisite distinctions among these diverse groups, lead to its being accurately characterised as an authentic social movement. A "social movement" is defined as a collective of individuals or groups united by a shared objective to effect social, economic, or political change, collaborating to influence or alter public policy, society, or established institutions. Diani (1992) defines "social movements" as networks of informal interactions among diverse individuals, groups, and/or organisations engaged in a political or cultural conflict rooted in a common collective identity. This concept is applicable to the manosphere, a diverse network of online groups (including MGTOW, Incels, and MRAs) that propagate a shared antifeminist narrative and advocate male interests. They establish a communal collective identity. According to Diani, a shared collective identity unites members through common beliefs, values, and objectives. Within the manosphere, this collective identity is frequently formulated around the concept of "masculinity in crisis" and opposition to feminism and perceived societal transformations as threats. Furthermore, social movements participate in political and cultural conflicts that seek to advocate or resist cultural transformations. The manosphere comprises groups that are variably and intensely involved in a cultural conflict, aiming to influence gender norms and public discourse concerning the roles of men and women in society. Social movements function independently of formal institutions. The manosphere, operating mainly on digital platforms and social media, eschews conventional institutional methods for enacting social and cultural change, opting instead to shape public

opinion and construct counter-narratives. Collective identities are formed through meaning-making processes and symbolic representations. This is evident in the formulation of novel terminology and concepts within the manosphere (such as “red pill” or “beta male”) that define the identities and experiences of its members.

7. Conclusion

Our parallel analysis of user profiles on blogs and forums, as well as the network of groups on Facebook, highlights the diversity of user profiles while simultaneously revealing the structure and functions of the different groups within the Italian manosphere, offering a perspective that goes beyond the structural insights provided by Facebook. Blogs and forums in particular serve as privileged spaces for the construction of identity narratives and the consolidation of anti-feminist ideologies, functioning as tools closely tied to personal reflection and radicalisation. These platforms enable a deeper exploration of subjective experiences and emotional dynamics, making them an essential complementary element in the analysis of the manosphere.

Our examination of the composition, structure, and functions of the Italian manosphere underscores its diverse yet interrelated characteristics. The reconstruction of the Facebook network and analysis of self-presentations on blogs and forums have yielded a reliable understanding of the internal dynamics of this digital ecosystem. Despite thematic and objective distinctions within the Italian manosphere, its factions collectively espouse a fundamental anti-feminist and misogynistic ideology, as articulated through discourses on the “crisis of masculinity” and “male victimisation.”

MRA pages function as pivotal nodes within the network, acting as ideological authorities for other groups. MGTOW pages serve as “bridges” among different segments of the manosphere, enabling content exchange among otherwise segregated groups. The role of Incels, while peripheral, is pivotal: They act as gatekeepers and are critical to the radicalisation of users, exacerbating anti-feminist ideologies. Ultimately, PUAs constitute a niche subgroup, yet they are noteworthy for their contributions to discussions on seduction and dominance in interpersonal relationships.

Although the Italian network’s international connections are less pronounced than those in Anglophone contexts, they confirm the influence of global trends on the Italian manosphere while maintaining distinctive elements tied to the national socio-cultural context. Italian manosphere groups seem to be less radical than those found in the US context and are more connected to a conservative view of gender roles and the traditional family. Nevertheless, the Italian manosphere represents an archipelago of groups that, collectively, significantly shape public discourse on gender relations, especially among members of the younger generations.

Given the functions of these groups, they can be termed “online communities of practice” in which members share experiences and strategies, thereby strengthening their collective identity. Moreover, they may be perceived as authentic social movements with possible political ramifications in shaping gender relations. Understanding this phenomenon, particularly concerning younger generations, is crucial due to its increasing influence on the polarisation of public discourse and on gender representations in Italian society.

It is important to acknowledge some limitations of the present research. The decision to focus exclusively on Facebook as the platform for analysis, although justified by the availability of public data and the ability to map connections among pages, potentially excludes significant dynamics present on other platforms such as Reddit, TikTok, or Discord, where some subgroups of the manosphere, as well as younger generations in general, are more active. The data collected reflect a specific temporal snapshot, providing a picture of the Italian network on Facebook and offering reflections that will undoubtedly necessitate further exploration on other platforms.

Acknowledgments

During the preparation of this work, the authors used ChatGPT 4o to improve the language and readability of their article. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Funding

This research was undertaken as part of the ICOMIC (Identifying and Counteracting Online Misogyny in Cyberspace) project funded by EU Next Generation, MUR-Fondo Promozione e Sviluppo-DM 737/2021.

Conflict of Interests

The authors declare no conflict of interests.

Data Availability

Data are available from the authors upon request.

Supplementary Material

Supplementary material for this article is available online in the format provided by the author (unedited).

References

- Amaratunga, D. (2014). Typhoon Haiyan in the Philippines and the RAY plan. *International Journal of Disaster Resilience in the Built Environment*, 5(1). <https://doi.org/10.1108/IJDRBE-01-2014-0006>
- Arora, S. D., Singh, G. P., Chakraborty, A., & Maity, M. (2022). Polarization and social media: A systematic review and research agenda. *Technological Forecasting and Social Change*, 183, Article 121942. <https://doi.org/10.1016/j.techfore.2022.121942>
- Banet-Weiser, S., & Miltner, K. M. (2015). #MasculinitySoFragile: Culture, structure, and networked misogyny. *Feminist Media Studies*, 16(1), 171–174. <https://doi.org/10.1080/14680777.2016.1120490>
- Bernholz, L., Landemore, H., & Reich, R. (Eds.). (2021). *Digital technology and democratic theory*. University of Chicago Press.
- Cannito, M., Crowhurst, I., Ferrero Camoletto, R., Mercuri, E., & Quaglia, V. (2021). Doing masculinities online: Defining and studying the manosphere. *AG-ABOUT GENDER*, 10(19), 1–34. <https://doi.org/10.15167/2279-5057/AG2021.10.19.1326>
- Cannito, M., & Ferrero Camoletto, R. (2022). The rules of attraction: An empirical critique of pseudoscientific theories about sex in the manosphere. *Sexes*, 3, 593–607. <https://doi.org/10.3390/sexes3040043>
- Cannito, M., & Mercuri, E. (2022). Fatherhood and gender relations in the manosphere: Exploring an Italian non-resident fathers' online forum. *European Journal of Cultural Studies*, 25(4), 1010–1029. <https://doi.org/10.1177/13675494211036967>

- Capelos, T., Salmela, M., Talalakina, A., Cotena, O., & Candiotta, L. (Ed.). (2024). Resentment in the manosphere: Conceptions of morality and avenues for resistance in the incel hatred pipeline. *Philosophies*, 9(2), Article 36. <https://doi.org/10.3390/philosophies9020036>
- Carian, E. (2022). "No seat at the party": Mobilizing white masculinity in the men's rights movement. *Sociological Focus*, 55(1), 27–47. <https://doi.org/10.1080/00380237.2021.2009075>
- Chang, W. (2020). The monstrous-feminine in the incel imagination: Investigating the representation of women as "femoids" on/r/Brincels. *Feminist Media Studies*, 22(2), 254–270. <https://doi.org/10.1080/14680777.2020.1804976>
- Ciccione, S. (2017). *Essere maschi: tra potere e libertà*. Rosenberg & Sellier.
- Cosma, S., & Gurevich, M. (2020). Securing sex: Embattled masculinity and the pressured pursuit of women's bodies in men's online sex advice. *Feminism & Psychology*, 30(1), 42–62. <https://doi.org/10.1177/0959353519857754>
- De Gasperis, A. (2021). "Giacomino uno di noi." Letteratura italiana e pratiche di maschilità nel Forum dei Brutti. *AG ABOUT GENDER*, 10(19). <https://doi.org/10.15167/2279-5057/ag2021.10.19.1250>
- Diani, M. (1992). The concept of social movement. *The Sociological Review*, 40(1), 1–25. <https://doi.org/10.1111/j.1467-954X.1992.tb02943.x>
- Dickel, V., & Evolvi, G. (2022). "Victims of feminism": Exploring networked misogyny and #MeToo in the manosphere. *Feminist Media Studies*, 23(4), 1392–1408. <https://doi.org/10.1080/14680777.2022.2029925>
- Dordoni, A., & Magaraggia, S. (2021). Modelli di mascolinità nei gruppi online Incel e Red Pill: Narrazione vittimistica di sé, deumanizzazione e violenza contro le donne. *AG-ABOUT GENDER*, 10(19), 35–67. <https://doi.org/10.15167/2279-5057/AG2021.10.19.1268>
- Ebner, J. (2017). *The rage: The vicious circle of Islamist and far-right extremism*. Bloomsbury Publishing.
- Farci, M., & Righetti, N. (2019). Italian men's rights activism and online backlash against feminism. *Rassegna Italiana di Sociologia*, 60(4), 765–781. <https://doi.org/10.1423/96115>
- Fontanella, L., Chulvi, B., Ignazzi, E., Sarra, A., & Tontodimamma, A. (2024). How do we study misogyny in the digital age? A systematic literature review using a computational linguistic approach. *Humanities and Social Sciences Communications*, 11, Article 478. <https://doi.org/10.1057/s41599-024-02978-7>
- Gabrielatos, C. (2018). Keyness Analysis: Nature, metrics and techniques. In C. Taylor, & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 225–258). Routledge.
- García-Mingo, E., Díaz-Fernández, S., & Tomás Forte, S. (2022). (Re)configurando el imaginario sobre la violencia sexual desde el antifeminismo: El trabajo ideológico de la manófera española. *Política y sociedad*, 59(1). <https://doi.org/10.5209/POSO.80369>
- García-Mingo, E., & Diaz-Fernandez, S. (2023). Mapping research on online misogyny and manosphere in Spain: The way ahead. *Masculinities & Social Change*, 12(3), 293–309. <https://doi.org/10.17583/mcs.11882>
- Ging, D. (2019). Alphas, betas, and incels: Theorizing the masculinities of the manosphere. *Men and Masculinities*, 22(4), 638–657. <https://doi.org/10.1177/1097184X17706401>
- Gotell, L., & Dutton, E. (2016). Sexual violence in the 'manosphere': Antifeminist men's rights discourses on rape. *International Journal for Crime, Justice and Social Democracy*, 5(2), 65–80. <https://doi.org/10.5204/ijcjsd.v5i2.310>
- Gourarier, M., & Benveniste, A. (2017). Alpha mâle, Séduire les femmes pour s'apprécier entre hommes? *Chimères*, 2(92), 175–178. <https://doi.org/10.3917/chime.092.0175>
- Halpin, M. (2022). Weaponized subordination: How incels discredit themselves to degrade women. *Gender & Society*, 36(6), 813–837. <https://doi.org/10.1177/08912432221128545>

- Hawley, G. (2017). *Making sense of the alt-right*. Columbia University Press. <https://doi.org/10.7312/hawl18512>
- Heritage, F., & Koller, V. (2020). Incels, in-groups, and ideologies: The representation of gendered social actors in a sexuality-based online community. *Journal of Language and Sexuality*, 9(2), 152–178. <https://doi.org/10.1075/jls.19014.her>
- Hine, C. (2015). *Ethnography for the internet: Embedded, embodied and everyday*. Bloomsbury.
- Hopton, K., & Langer, S. (2022). “Kick the XX out of your life”: An analysis of the manosphere’s discursive constructions of gender on Twitter. *Feminism & Psychology*, 32(1), 3–22. <https://doi.org/10.1177/09593535211033461>
- Horta Ribeiro, M., Blackburn, J., Bradlyn, B., De Cristofaro, E., Stringhini, G., Long, S., Greenberg, S., & Zannettou, S. (2021). The Evolution of the manosphere across the web. *Proceedings of the International AAAI Conference on Web and Social Media*, 15(1), 196–207. <https://doi.org/10.1609/icwsm.v15i1.18053>
- Hunte, Z. (2019). “Female nature, cucks, and simps”: *Understanding men going their own way as part of the manosphere* [Unpublished master’s dissertation]. Uppsala University.
- Jiménez-Zarco, A. I., González-González, I., Saigí-Rubió, F., & Torrent-Sellens, J. (2015). The co-learning process in healthcare professionals: Assessing user satisfaction in virtual communities of practice. *Computers in Human Behavior*, 51, 1303–1313. <https://doi.org/10.1016/j.chb.2014.11.057>
- Jones, C., Trott, V., & Wright, S. (2020). Sluts and soyboys: MGTOW and the production of misogynistic online harassment. *New Media & Society*, 22(10), 1903–1921. <https://doi.org/10.1177/1461444819887141>
- Kennedy-Kollar, D. (2024). *Extremism and radicalization in the manosphere: Beta uprising*. Taylor & Francis. <https://doi.org/10.4324/9781032631080>
- Kozinets, R. V. (2010). *Netnography: Doing ethnographic research online*. Sage.
- Lacalle, C., Gómez-Morales, B., & Vicent-Ibáñez, M. (2023). Misogyny and the construction of toxic masculinity in the Spanish Manosphere (Burbuja.info). *Profesional De La Información*, 32(2). <https://doi.org/10.3145/epi.2023.mar.15>
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815355>
- Lin, J. L. (2017). Antifeminism online. MGTOW (men going their own way). In U. U. Frömming, S. Köhn, S. Fox, & M. Terry (Eds.), *Digital environments: Ethnographic perspectives across global online and offline spaces* (pp. 77–96). transcript. <https://doi.org/10.25595/503>
- Lindsay, A. (2022). Swallowing the black pill: Involuntary celibates’ (incels) anti feminism within digital society. *International Journal for Crime, Justice and Social Democracy*, 11(1), 210–224. <https://doi.org/10.5204/ijcjsd.2138>
- Longo, T. (2020). Misoginia online: Le nuove forme di radicalizzazione all’interno del terrorismo Incel. *Sicurezza, terrorismo e società*, 1, 85–99.
- Lucini, B. (2014). *Disaster resilience from a sociological perspective. Exploring three Italian earthquakes as models for disaster resilience planning*. Springer. <https://doi.org/10.1007/978-3-319-04738-6>
- Maretti, M., Russo, V., & Lucini, B. (2022). Resilience in online communities of practice during the COVID-19 pandemic: An Italian case study. *International Review of Sociology*, 32(2), 332–351. <https://doi.org/10.1080/03906701.2022.2114871>
- Marwick, A. E., & Caplan, R. (2018). Drinking male tears: Language, the manosphere, and networked harassment. *Feminist Media Studies*, 18(4), 543–559. <https://doi.org/10.1080/14680777.2018.1450568>
- Menzie, L. (2022). Stacys, Beckys, and Chads: The construction of femininity and hegemonic masculinity within incel rhetoric. *Psychology & Sexuality*, 13(1), 69–85. <https://doi.org/10.1080/19419899.2020.1806915>

- Nagle, A. (2017). *Kill all normies: Online culture wars from 4chan and Tumblr to Trump and the alt-right*. Zero Books.
- Núñez Puente, S., D'Antonio Maceiras, S., & Romero, D. (2021). New discourses of masculinity in the context of online misogyny in Spain: The use of the “feminazi” and “gender ideology” concepts on Twitter. *Socialni Studia*, 2, 49–66. <https://doi.org/10.5817/SOC2021-2-49>
- Nurminen, M. (2022). From swallowing the red pill to failing to build the wall: Allusive cognitive metaphors in advocating political and extremist views. *Poetics Today*, 43(2), 309–334.
- O'Donnell, C., & Shor, E. (2022). “This is a political movement, friend”: Why “incels” support violence. *British Journal of Sociology*, 73(2), 336–351. <https://doi.org/10.1111/1468-4446.12923>
- O'Hanlon, R., Altice, F. L., Lee, R. K.-W., LaViolette, J., Mark, G., Papakyriakopoulos, O., Saha, K., De Choudhury, M., & Kumar, N. (2024). Misogynistic extremism: A scoping review. *Trauma, Violence, & Abuse*, 25(2), 1219–1234. <https://doi.org/10.1177/15248380231176062>
- Papacharissi, Z. (2015). Affective publics and structures of storytelling: Sentiment, events, and mediality. *Information, Communication & Society*, 19(3), 307–324. <https://doi.org/10.1080/1369118X.2015.1109697>
- Pilla, F., & Dolce, R. (2019). *Il Web che odia le donne*. Ledizioni.
- Postill, J., & Pink, S. (2012). Social media ethnography: The digital researcher in a messy web. *Media International Australia*, 145(1), 123–134. <https://doi.org/10.1177/1329878X1214500114>
- Ricci, O., & Farci, M. (2021). “Io onestamente oggi non vorrei nascere maschio.” Il caso Marco Crepaldi. *AG-ABOUT GENDER*, 10(19), 167–198. <https://doi.org/10.15167/2279-5057/AG2021.10.19.1269>
- Scotto di Carlo, G. (2023). An analysis of self-other representations in the incelosphere: Between online misogyny and self-contempt. *Discourse & Society*, 34(1), 3–21. <https://doi.org/10.1177/09579265221099380>
- Scotto di Carlo, G. (2024). Deconstructing men's rights activism: A discourse analysis of D.A.R.V.O. in MRA discourse on rape culture. *Journal of Language and Discrimination*, 8(1), 74–99. <https://doi.org/10.1558/jld.25747>
- Semenzin, S., & Bainotti, L. (2020). The use of Telegram for non-consensual dissemination of intimate images: gendered affordances and the construction of masculinities. *Social Media + Society*, 6(4). <https://doi.org/10.1177/2056305120984453>
- Speckhard, A., Ellenberg, M., Morton, J., & Ash, A. (2021). Involuntary celibates' experiences of and grievance over sexual exclusion and the potential threat of violence among those active in an online incel forum. *Journal of Strategic Security*, 14(2), 89–121. <https://doi.org/10.5038/1944-0472.14.2.1910>
- Taylor, J. (2018, August 30). The woman who founded the ‘incel’ movement. *BBC News*. <https://www.bbc.com/news/world-us-canada-45284455>
- van Dijk, T. A. (1998). Editorial: Discourse and ideology. *Discourse & Society*, 9(3), 307–308. <https://doi.org/10.1177/0957926598009003001>
- Van Valkenburgh, S. P. (2021). Digesting the red pill: Masculinity and neoliberalism in the manosphere. *Men and Masculinities*, 24(1), 84–103. <https://doi.org/10.1177/1097184X18816118>
- Vingelli, G. (2019). Antifemminismo online. I Men's rights activists in Italia. *Im@go: A Journal of the Social Imaginary*, 14, 219–247. <https://doi.org/10.7413/22818138157>
- Vingelli, G. (2021). Miss Hitler. Destra radicale e ideologie di genere. In A. Cammarota & M. Meo (Eds.), *Populismo e questioni di genere. Rappresentazioni, politiche, movimenti* (pp. 31–46). Franco Angeli.
- Wenger, E. (2011). *Communities of practice: A brief introduction*. National Science Foundation. <http://hdl.handle.net/1794/11736>

Wright, S., Trott, V., & Jones, C. (2020). "The pussy ain't worth it, bro": Assessing the discourse and structure of MGTOW. *Information, Communication & Society*, 23(6), 908–925. <https://doi.org/10.1080/1369118X.2020.1751867>

About the Authors



Elisa Ignazzi is a PhD student in business and behavioural science at the G.d'Annunzio University of Chieti-Pescara, Italy. Her research interests are primarily focused on human behaviour drivers, hate speech, and online misogyny.



Mara Maretti (PhD) is a full professor of sociology at the G.d'Annunzio University of Chieti-Pescara. She coordinates the Computational Social Research Laboratory within the Department of Legal and Social Sciences. Her work focuses on social research methodology, with particular attention to the study of social phenomena in digital spaces, social and inclusion policies, and sustainability.



Lara Fontanella is a full professor of social statistics at the G.d'Annunzio University of Chieti-Pescara, Italy, specialising in social media, Big Data, text mining, and social network analysis. Her research primarily addresses the diffusion of hate speech, online misogyny, political debates, and conspiracy theories. Methodologically, she focusses on graph theory, latent variable models, Bayesian item response theory, Bayesian structural equation models, functional data analysis, and statistical shape analysis.

Violence, Hate Speech, and Discrimination in Video Games: A Systematic Review

Roberto Moreno-López ¹  and Catalina Argüello-Gutiérrez ² 

¹ University of Castilla-La Mancha, Spain

² Universidad Internacional de La Rioja, Spain

Correspondence: Roberto Moreno-López (roberto.moreno@uclm.es)

Submitted: 14 October 2024 **Accepted:** 2 December 2024 **Published:** 9 January 2025

Issue: This article is part of the issue “Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment” edited by Max Römer Pieretti (Universidad Camilo José Cela), Beatriz Esteban-Ramiro (Universidad de Castilla-La Mancha), and Agrivalca Canelón (Universidad Católica Andrés Bello), fully open access at <https://doi.org/10.17645/si.i415>

Abstract

This systematic review analyses the relationships between violence, hate speech, discrimination, and video games. A comprehensive search of the Web of Science and Scopus databases identified 47 relevant studies published between 2018 and 2023. The review examines how video games may provide fertile ground for online violence, hate speech, and discrimination, while also exploring their potential as educational tools. Key findings suggest that exposure to violent video game content can increase aggressive cognitions and behaviours, particularly when combined with competitive gameplay. However, prosocial aspects of gaming may promote positive intergroup attitudes and reduce prejudice. Hate speech and discriminatory behaviours remain prevalent issues in online gaming communities, disproportionately affecting marginalised groups. The article highlights the complex interactions between game content, individual factors, and sociocultural contexts in shaping player experiences and behaviours. While video games pose risks, they also offer opportunities for fostering empathy, cultural understanding, and critical thinking, if they are thoughtfully designed. The findings underscore the need for evidence-based interventions to mitigate online hate and maximise the educational potential of video games.

Keywords

cyberhate; hate speech; inclusion; video games; violence

1. Introduction

The rapid growth of the video game industry as a global entertainment cornerstone (Ripoll & Muñoz, 2023) has raised concerns about cyberbullying, a form of online violence that compromises safety by the

intimidation of individuals through text, audio, or video communication (Rančić, 2018). In addition to being a form of entertainment, video games have been used as propaganda tools, not only by political and social actors (Aouragh, 2016) but also by terrorist organisations and extremist groups (Hartgers & Leidig, 2023; Moreno, 2022).

Although many social networks have implemented mechanisms to detect and block such content, these measures are often superficial and largely ineffective (Pohjonen & Udupa, 2017). Research has shown that the use of online media drives aggressive behaviour across cultures, regardless of social norms (Anderson et al., 2017), and facilitates the spillover of bullying behaviour into real life (Barter et al., 2017). In this vein, the Anti-Defamation League's (ADL) annual report on online gaming experiences (ADL, 2022) reveals that while many gamers appreciate the social connectivity offered by these spaces, a troubling majority still faces instances of hate and harassment. The survey, conducted in the United States with a representative sample of multiplayer online gamers, reveals disturbing reports of hate-related experiences. As many as 83% of adults aged 18–45 have experienced harassment while playing these games, representing more than 80 million adult gamers. In addition, 60% of young people aged 13–17 have experienced harassment in these environments, which equates to almost 14 million young gamers. Meanwhile, 8% of adults and 10% of young people have been exposed to online conversations involving white supremacist ideology, which promotes a belief in the superiority of white people over other races. In addition, 7% of adult gamers in online multiplayer environments have been exposed to Holocaust denial during gaming sessions (ADL, 2022).

Video games also provide an immersive environment in which players actively engage in interactive narratives, increasing their persuasive power. The rules of the game determine what is permissible, creating a space that can serve as a metaphor for real-world norms and values (Bogost, 2008). It is this ability to simulate complex social and political contexts that makes video games an effective medium for internalising extremist ideologies. Interaction in virtual communities has enabled radical groups to exploit the empathic potential of video games to recruit new members. Such narratives reinforce the idea that violence is a legitimate means of resolving social and political conflicts, in line with the beliefs of the target audience. Unlike other forms of propaganda, video games are interactive, meaning that players can make moral choices within the game. As Sicart (2011) argues, video games are a powerful tool for ethical reflection because players experience the consequences of their actions in the virtual world. Research has highlighted the long-term impact of video games on attitude and behavioural change. Immersive interventions have been shown to reduce fearful attitudes and encourage more cautious decision-making in ambiguous situations, which is essential in contexts of violence (Hasson et al., 2019).

In addition, video games not only change attitudes towards other groups but also help players internalise prosocial values that are transferable to the real world. Despite the limitations outlined above, the research demonstrates solid evidence supporting the use of video games as an educational and preventive tool to combat violence (Breves, 2020; Hasson et al., 2019; Olson & Harrell, 2020). Video games have been shown to be effective in reducing prejudice and promoting social inclusion by providing a platform for interactive learning and the development of interpersonal skills (Lippe et al., 2022; Pech & Caspar, 2022). Studies by Pech and Caspar (2022) have shown how video games improve cultural understanding and encourage helping behaviour towards out-groups. The portrayal of ethnically and culturally diverse characters in video games has proven effective in reducing discriminatory attitudes and promoting tolerance. Breves (2020) found that interactions with characters from different ethnic groups significantly reduced prejudice, suggesting that these

interactions can be transferred to the real world. This phenomenon is akin to the “Proteus effect,” in which players take on the traits of the characters they control, thereby influencing their actual behaviour (Olson & Harrell, 2020). Immersive virtual reality environments have been shown to be more effective in reducing explicit bias than 2D video games (Breves, 2020).

Video games can be valuable tools in education and prevention because they can be adapted to different educational settings (Lippe et al., 2022) and provide controlled scenarios where norms and stereotypes can be safely challenged, helping to counter extremist narratives. However, they must be designed to be inclusive and promote peaceful solutions (Lippe et al., 2022; Pech & Caspar, 2022). The effect of virtual experiences on reducing fear and increasing empathy is crucial. Hasson et al. (2019) found that participants who assumed the perspective of an out-group in virtual reality exhibited greater empathy towards that group, even months later. This suggests that such interventions using video games can have a lasting impact on intergroup relations, which is essential for preventing violent extremism.

In a digital society, characteristics such as idealism, self-centredness, and a propensity to engage in risky behaviours are exacerbated by widespread access to technology and social networks, which bring with them risks such as privacy issues, internet addiction, and the misuse of social media platforms (Marín-Díaz & Cabero-Almenara, 2019). Such characteristics are inherent to adolescence (Martín-Martín et al., 2021), which makes young people a specific target of many video games.

Due to the rapid increase in research on this topic, an updated systematic review is essential to inform future decision-making. The aim of this systematic review is therefore to analyse the relationships between online violence, hate speech, discrimination, and video games among young people. It examines how video games facilitate these behaviours, identifies key individual and sociocultural factors that influence these experiences, explores their potential as educational tools, and assesses the methodological quality of existing research in this area.

2. Method

2.1. Eligibility and Study Selection Criteria

This systematic review followed the updated PRISMA guidelines (Page et al., 2021). The scientific literature search was conducted in the Web of Science and Scopus databases. Titles, abstracts, and subject headings were searched using the following terms: hate speech OR hate crime OR violent behaviour OR cyber hate OR online hate OR racism OR discrimination OR violence against women AND video games OR MMOs OR online games OR games OR games AND adolescents OR adolescent OR youth OR young people OR young adults OR children OR child. The search was conducted in both English and Spanish. Relevant articles and their corresponding abstracts were selected. Subsequently, the studies were screened to ensure that they met the inclusion criteria.

The inclusion criteria for the selection of studies were: (a) studies focusing on violence and hate speech in video games, (b) studies focusing on adolescents and young adults (under 35 years old), (c) studies with a publication date between 2018 and 2023, and (d) articles written in English or Spanish. The following were excluded from the review: single case studies ($n = 4$), studies in which the sample was over 35 years of age

($n = 15$), doctoral theses ($n = 8$), articles on extremism and violent radicalisation not related to potential hate crimes ($n = 24$), and cyberbullying or other forms of harassment not related to discrimination or hate speech ($n = 64$), as these were beyond the scope of the systematic review. The identification and selection process were carried out by blind peer-review, with two independent reviewers using the search strategies to identify scientific articles only, excluding books, book chapters, conference proceedings, and reviews. After performing this procedure in the two databases, 47 studies were identified (WOS, $n = 31$; Scopus, $n = 16$). The details of the study selection process are outlined in the PRISMA flowchart (Figure 1).

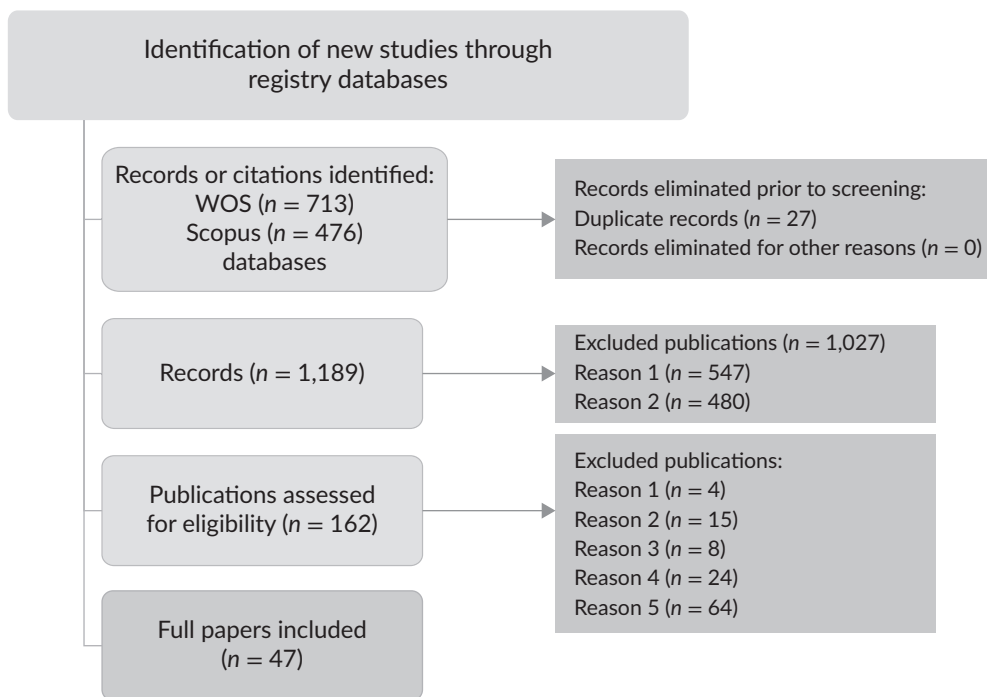


Figure 1. PRISMA. Selection process of articles that met the criteria for this review.

2.2. Data Analysis

A double-entry table was designed to facilitate the collection and categorisation of the characteristics and interventions of each study. To this end, the following data were extracted from each study, where available: study objectives; type of intervention; evaluation design; country; sample and sample characteristics; instruments and/or variables; data analysis; results; and limitations.

The characteristics of each study were extracted, taking into account both methodological strengths and limitations and overall quality. To assess the quality of the studies, we used the classification system proposed by León and Montero (2020), which allowed the documents to be organised by type. This classification allowed the subsequent use of a quality assessment tool.

2.3. Quality of the Studies

The methodological quality of the studies was assessed independently by two reviewers using a 27-item checklist (Downs & Black, 1998) organised into five dimensions:

1. General information: assessing the clarity and accuracy of the description of key aspects such as objectives and hypotheses, study measures, procedures, interventions, confounding factors, main results, appropriate statistical estimates, probability values, sample characteristics, and limitations (ten items).
2. External validity: measuring the extent to which the findings can be generalised to the population from which the studies were drawn (three items).
3. Internal validity and bias: assessing potential biases in the measurement of the intervention and in the interpretation of the results (seven items).
4. Internal validity and confounding variables: examining bias in the selection and assignment of participants (six items).
5. Power: assessed by a single item evaluating the capacity of the study to detect significant differences.

Responses to the items were scored as either 0 or 1, except for item 5 in the general information dimension, which was scored from 0 to 2, and item 27 in the power dimension, which was scored from 0 to 5. The highest possible total score was 32 points.

A checklist was used to assess the studies according to their level of compliance with the reviewed characteristics. Higher scores reflected a higher quality of study. The two reviewers agreed on the rating of most of the studies, and where there was disagreement, the articles were reviewed to reach a consensus.

3. Results

Table 1 summarises the studies that analysed the relationships among violence, hate speech, discrimination, and video games, organised based on the scores obtained during the review process. As can be observed, study scores ranged from 16 to 30. Most studies were clustered around scores 19 and 20, while lower scores 16–18 had fewer studies, suggesting a possible minimum quality threshold. More than half of the studies (61.70%) have a score of 20 or more. The distribution of studies across the higher scores highlights their relevance in the context of the current review.

Table 1. Selected studies organised by obtained quality score.

Study	Score	Main Content
Costa et al. (2021)	30	Discusses gamification and video games as tools to combat online hate speech, promoting media literacy and digital citizenship.
da Silva and Ifa (2022)	30	Examines hate speech in online video games, highlighting adolescents' experiences and the need to tackle tolerance of hate in digital spaces.
Obermaier and Schmuck (2022)	30	Investigates why young people become victims of online hate speech using criminology's routine activity theory.
Vicentini et al. (2023)	30	Develops a theory on hate speech learning among e-sports players through a grounded theory-based retrospective study.

Table 1. (Cont.) Selected studies organised by obtained quality score.

Study	Score	Main Content
Guggisberg (2020)	29	Stresses the need for awareness, education, and early intervention to prevent harm and violence.
Aguerri et al. (2023)	28	Analyses the toxicity of League of Legends, focusing on content moderation and managing digital gaming communities.
Li (2022)	28	Finds a link between exposure to violent video games and moderate levels of cyberaggression in university students.
Zhu et al. (2020)	27	Shows that beliefs about aggression mediate the link between exposure to violent games and cyberbullying, particularly in aggressive adolescents.
Shahghasemi (2018)	27	Reports no significant link between violent online games and perceptions of the world as unsafe, with minimal impact on social perceptions.
Bowman et al. (2022)	27	Analyses Reddit posts showing the potential of video games to prevent violence.
Waddell (2020)	26	Finds that gender most strongly influences expectations about the effects of violent games.
Zhang et al. (2019)	26	Shows that players of high-violence games react more negatively to review volume, while social game players react negatively to review readability and negativity.
Koehler et al. (2022)	25	Reviews cases of minors radicalised through gaming platforms, leading to extreme-right criminal activities.
Mason and Turner (2018)	25	Describes an educational game aimed at addressing gaps in healthcare providers' knowledge about domestic violence.
Gilbert et al. (2018)	25	Links playing sports video games to increased alcohol consumption, drug use, and criminal behavior, influenced by masculinity ideology.
Dickmeis and Roe (2019)	25	Shows a positive association between violent/competitive online games and self-reported aggression.
Yuldasheva and Mukhopadhyay (2022)	22	Finds that violent content influences aggressive behaviour in young people.
Gong and Piller (2018)	21	Studies parental involvement in gaming among immigrant and native-born parents in the US.
Adelhardt and Eberle (2019)	21	Examines how a six-month break from gaming affects adolescents' media consumption habits.
TaeHyuk Keum and Hearn (2022)	21	Analyses the impact of racism in online gaming on gamers' psychological well-being.
Zhao and Liu (2023)	21	Identifies shifts in media portrayal of video games from commercial to social concerns between 2010 and 2020.
Hodge et al. (2020)	21	Explores the relationship between moral development and video game play, showing higher moral scores in male university students.
Cabras et al. (2019)	20	Reveals gender and age differences in anxiety, self-esteem, and aggression among gamers but no direct link to violent games.
Hodge et al. (2019)	20	Finds higher moral reasoning in boys than girls in the context of video gaming and notes shifts in moral development.
Wu et al. (2023)	20	Examines how online gaming influences unethical decision-making in young adults.

Table 1. (Cont.) Selected studies organised by obtained quality score.

Study	Score	Main Content
Halbrook et al. (2019)	20	Suggests that social engagement within games may promote prosocial behavior and mitigate aggression.
Paz et al. (2020)	20	Analyses media messages and how they mobilise followers.
Ferguson and Colwell (2020)	20	Finds no link between exposure to sexualised content and sexist attitudes or reduced empathy.
Albaker et al. (2021)	19	Assesses the impact of video game addiction on the mental and physical health of university students.
Charmaraman et al. (2020)	19	Finds greater depressive symptoms and problematic internet behaviours in adolescents who play high-risk games.
M. Salter (2018)	19	Analyses Gamergate, focusing on the role of platforms in facilitating gender-based harassment.
Wearing et al. (2022)	19	Critiques reductionist views of video games and suggests they be studied as leisure experiences in consumer culture.
Ferguson et al. (2020)	19	Argues that the link between aggressive video games and real-world aggression is still inconclusive.
Ybarra et al. (2022)	19	Shows a significant link between violent media consumption and serious violent behaviour.
Fleet and Nurmikko-Fuller (2019)	19	Suggests massively multiplayer online (MMO) video games could serve as research environments for organised crime.
Makarova and Makarova (2019)	19	Highlights cybervictimisation as a pressing issue for device-dependent adolescents.
Bacovsky (2021)	19	Finds that adolescents who play more video games are less engaged in sociopolitical issues and prosocial behaviour.
Irmak and Erdoğan (2019)	18	Shows different influences on gaming behaviour for adolescent boys and girls, including family environment and self-efficacy.
Shortland et al. (2022)	18	Examines how personality influences extremist cognitions following exposure to extremist content.
Li (2022)	17	Finds a positive link between violent video games and cyberaggression in university students.
Madden et al. (2021)	27	Identifies gender role biases in video games, including the belief that girls dislike violence.
Kaakinen et al. (2020)	17	Identifies links between online hate and both personal and group behaviour from a social-psychological perspective.
Reer and Krämer (2018)	17	Suggests that research on first-person shooters should consider the positive aspects of clan membership alongside violence.
Liby et al. (2023)	17	Identifies key themes in young people's experiences of online racism, including types of aggression and coping strategies.
Clarice et al. (2021)	17	Explores how female gamers perceive violence in online gaming, with psychological abuse and sexual harassment being common.
Imran et al. (2023)	16	Discusses the ongoing debate on whether violent video games lead to violent behaviour.
Keipi et al. (2018)	16	Finds that online hate content negatively impacts young people's well-being in both Finland and the US.

3.1. Violence, Perception of Violence, and Violent Effects

The growing popularity of video games, especially online, has created a need to understand their influence on human behaviour. According to Guggisberg (2020), the need to understand the effects of sexual and violent content in video games is of vital importance today. In this context, Li (2022) found a positive correlation between exposure to violent video games and cyberaggressive behaviour among university students, with anger acting as a significant mediator. Similarly, Zhu et al. (2020) found that normative beliefs about aggression may partially mediate the relationship between exposure to video game violence and cyberbullying in adolescents, with aggressiveness being an important moderating factor.

The combination of competition and violence in video games may increase the likelihood of physical aggression, as shown by Dickmeis and Roe (2019), highlighting the complexity of the effects of video games on behaviour. Yuldasheva and Mukhopadhyay (2022) found that media violence exacerbated the risks for young people during the pandemic and suggested strategies to mitigate these negative effects.

Exposure to violent media during childhood correlates with an increased risk of violent behaviour in adolescence and young adulthood, suggesting that such exposure is a modifiable factor in the development of violent behaviour (Ybarra et al., 2022). Contrary to popular belief, Shahghasemi (2018) argued that there is no significant relationship between playing violent video games and perceptions of safety in society. This suggests that other factors, possibly sociocultural or contextual, may have a greater influence on perceptions of safety. Meanwhile, Bowman et al. (2022) suggested that certain levels in video games, such as “No Russian” from Call of Duty, can elicit eudaimonic responses and potentially act as reflective spaces for violence prevention. Media perceptions also play a crucial role in shaping perceptions of violent video games.

3.2. Personal Factors and Well-Being

Research on video games reveals the diverse effects they have on human behaviour. While some studies focus on differences in anxiety and aggression depending on the type of game, others explore their influence on ethics and well-being. In addition, addiction and sexualisation in video games are emerging as major concerns. Cabras et al. (2019) found significant differences in levels of anxiety, self-esteem, and aggression between players of violent and non-violent video games. However, preference for violent games was not the sole determining factor. Wu et al. (2023) explored how online gaming can influence unethical decision-making, highlighting the role of cheating and moral disengagement.

Halbrook et al. (2019) argued that the effects of video games on well-being depend on moderating variables such as motivations for gaming, suggesting that there may be an “optimal gaming profile” that maximises benefits. Bos (2023) explored how games such as This War of Mine challenge popular understandings of geopolitics and encourage reflection on the consequences of urban conflict. Sexualisation in video games and its impact on sexist attitudes is another important consideration. Ferguson and Colwell (2020) concluded that exposure to sexualised content does not necessarily increase sexist attitudes, especially in individuals with higher levels of trait aggression.

In addition, a study by Adelhardt and Eberle (2019) showed that separating adolescents from video games during a trip led to a significant decrease in video game use. Addiction, especially among university students,

has an impact on sleep habits, as observed by Albaker et al. (2021). Finally, Charmaraman et al. (2020) highlighted the negative social consequences of high-risk gaming, including depressive symptoms and problematic internet behaviour.

3.3. Sociocultural Factors: Social Influence and Extremism

Research on video games has revealed their impact on public perceptions and sociopolitical attitudes. Factors such as player gender, family background, and game features influence these dynamics. In addition, video games can influence radicalisation and provide a framework for studying complex behaviours. Waddell (2020) noted that media communications can shape public perception, with the participant's gender being a significant factor in how expected effects are perceived. At the same time, Zhang et al. (2019) demonstrated that the level of violence and social orientation of video games affects how consumers respond to online reviews, with significant differences observed depending on individual traits and product features.

Gaming can have a negative impact on the sociopolitical and prosocial attitudes of adolescents, posing a challenge to the development of democratic attitudes (Bacovsky, 2021). Irmak and Erdoğan (2019) found that factors such as family environment and school performance influence adolescent gaming behaviour, with computer self-efficacy and impulse control being key factors. Parental perception also plays a significant role in mediating video game use. Gong and Piller (2018) highlighted differences in perceptions between immigrant and native-born parents in the US, with implications for parental mediation of exposure to violent games. Studies by Adelhardt and Eberle (2019) and TaeHyuk Keum and Hearn (2022) have addressed both the negative and positive effects of video game use, including addiction and social well-being.

In terms of radicalisation, Koehler et al. (2022) examined right-wing extremism on video gaming platforms and highlighted the importance of online and offline interactions. However, they found no evidence of organised recruitment campaigns. This suggests that radicalisation may be a more subtle process, influenced by multiple everyday interactions within the gaming environment. Moreover, MMO video games provide a controlled environment in which to study behaviours such as those of organised criminal networks, providing valuable data for research and the prevention of real-world problems (Fleet & Nurmikko-Fuller, 2019).

3.4. Hate Speech

Analysis of hate speech in video games reveals that these platforms may have an impact on discriminatory behaviour and discourse. The Play Your Role: Gamification Against Hate Speech project has highlighted the importance of harnessing gamification to create spaces for dialogue and awareness raising around hate speech (Costa et al., 2021). In this context, tools such as pervasive and serious games have been developed to transform video game culture into an engine for media literacy and digital citizenship. Another study, through a nationally representative survey, identified six profiles of young victims, highlighting factors such as sex, migratory origin, religion, and age (Obermaier & Schmuck, 2022).

Research by Costa et al. (2024) showed that young people have been victims and witnesses of hate speech in games such as Free Fire. This study highlights the lack of awareness regarding the seriousness of hate speech among gamers and developers, and underlines the need to combat the tolerance of hate speech in online environments. Violence and toxicity in video games are ongoing concerns, particularly in popular titles

such as League of Legends. Aguerri et al. (2023) found that a high percentage of matches in this game are affected by disruptive behaviour. This phenomenon not only affects the gaming experience but also has implications for content moderation and community policies, as only a fraction of these games involve extremely harmful behaviour.

Furthermore, research into e-sports by Costa et al. (2024) showed that hate speech is often learned and reinforced through gameplay behaviours and outcomes. This highlights the need for e-sport administrators to implement effective measures to minimise these behaviours and promote a more positive environment for players.

Mechanisms to prevent hate speech and violence in video games have focused on education and the development of advanced technological tools. The Interactive Narratives Propose Pluralist Speech project focused on the use of interactive narratives to engage educators, trainers, and young people in curbing hate speech online (da Silva et al., 2020). This initiative highlighted the importance of collaboration between researchers, public institutions, and the educational community in the development of transmedia educational tools.

3.5. Discriminatory Behaviours

Research on video games and online platforms has revealed complex interactions between content, perception, and behaviour. Several studies acknowledged the importance of recognising that while sexualised content may not always promote sexist attitudes, it is nonetheless essential to understand and address the ways in which such portrayals can influence perceptions and behaviour. It is also worth noting that online racism can have a significant impact on the mental health of people from racial minority groups. Media coverage reflects concerns about addiction and social responsibility. Furthermore, personality may have a greater impact on radicalisation than propaganda, suggesting the need for personalised approaches to intervention.

Ferguson et al. (2020) concluded that exposure to sexualised content does not necessarily lead to an increase in sexist attitudes. Meanwhile, TaeHyuk Keum and Hearn (2022) examined the impact of online racism, which adversely affects the mental health of gamers from racial minority backgrounds. Zhao and Liu (2023) analysed media coverage in Chinese newspapers and found a shift towards social concerns and video game addiction, which is reflected in greater responsibility being attributed to multiple stakeholders. The Gamergate phenomenon illustrates how online platforms can facilitate abuse and highlights the need to address technological rationality to mitigate these issues (M. Salter, 2018).

Advergaming with prosocial narratives can positively influence attitudes towards gaming and game companies, with the interaction between game narratives and brand placement playing a key role (Sung & Lee, 2020). Personality, and aggression in particular, may be more influential than exposure to extremist propaganda in the development of extremist cognitions (Shortland et al., 2022), suggesting that interventions should take personality factors into account.

3.6. Educational Potential

Serious video games have also proven to be valuable educational tools. Mason and Turner (2018) developed a video game to train healthcare providers on domestic violence, filling an important knowledge gap. Gilbert

et al. (2018) observed that certain genres of video games, such as sports games, are associated with risk-taking behaviours influenced by masculinity ideologies, while online games are negatively associated with substance use. This suggests a complex relationship between video games and risk-taking behaviours.

Khalid and El-Maliki (2020) highlighted the role of educational videos in raising awareness and encouraging behavioural change. In terms of moral development, Hodge et al. (2019, 2020) show that different genres of video games can influence moral reasoning, although other factors such as game content are also determining factors. Wearing et al. (2022) considered gaming to be a leisure experience that contributes to the formation of self-identity in adolescents.

Finally, domestic legislation, such as the new youth protection law in Switzerland, is beginning to address the protection of young players from the negative consequences of video games, including hate speech and violent content (Lischer et al., 2022; Shinohara, 2024). These laws impose age and content restrictions and are an important step towards creating a safer and more responsible gaming environment.

4. Discussion

Overall, the reviewed research demonstrates a clear link between exposure to violent video games and aggressive behaviour, with anger and beliefs about aggression playing mediating roles. However, it is important to note some contradictory findings that highlight certain nuances. While some studies suggest that violent video games increase aggression (e.g., findings from the General Aggression Model), others argue that factors such as competitiveness or moral disengagement mediate these effects rather than violence itself (Aleissa et al., 2022; Olejarnik & Romano, 2023). In this sense, personal factors such as individual traits, motivations, and game type influence the effects of gaming on well-being. Addiction and mental health are key concerns, especially for university students. Sociocultural factors such as family background and gender shape gaming behaviour. While video games can affect sociopolitical attitudes and extremism, there is no evidence of organised recruitment through gaming. Hate speech is common in games such as Free Fire, highlighting a need for better moderation and awareness. Additionally, online racism has a significant impact on the mental health of minority groups, and careful attention should be given to sexualised content in games due to its potential influence on behaviour.

Research also suggests that interventions using video games can be effective in preventing hate violence, not only by reducing support for violent narratives, but also by promoting changes in gender dynamics and female representation in contexts of violence. Often these representations are subject to violence or are based on stereotypes. For video games to become an effective tool against hate and violence, they need to be designed from a perspective that reflects the experiences of other marginalised groups. It is important to consider that technologically mediated education seeks to understand how exposure to information in symbolic social spaces, such as social networks, affects individuals' learning processes (Moreno-López et al., 2021).

Despite the risks associated with violent video games, several studies suggest that they can also serve as an educational and preventive tool (Cipagauta & Gómez, 2019; Miranda-Palma et al., 2023). Video games provide a safe space where young people can explore identities, learn conflict resolution skills, and develop social and emotional skills (Gurlesin et al., 2020). Furthermore, their ability to promote empathy and cooperation can help to reduce prejudice and foster prosocial behaviour. Breves (2020) showed that

interactions with characters from different cultural backgrounds in video games can improve attitudes towards other social groups, promoting inclusion and diversity. Another important aspect is the potential of video games for building resilience and enhancing psychological well-being. Video games that cultivate skills such as decision-making, empathy, and conflict resolution can serve as preventive tools. Recent research indicates that educational video games improve digital resilience and critical thinking (Lippe et al., 2022). Moreover, Pusey et al. (2022) suggest that video games contribute to affective well-being by fulfilling the need for autonomy, competence, and interpersonal relationships.

The impact of these and other sexist actions by gamers has sparked numerous discussions about the hostility of this environment towards women (A. Salter & Blodgett, 2012). For many women, this hostility has caused them to feel less confident and secure in their gaming abilities compared to men (Kaye & Pennington, 2016). Madden et al. (2021) identified gender bias in e-sports that stems from stereotypical gender roles, such as the idea that girls dislike violence and boys are naturally competitive. While gender segregation in e-sports can help female players find role models and increase their confidence, it also perpetuates these stereotypes. In this sense, it would be interesting to explore this area in greater depth using tools such as the Sexism Against Women Gamers Scale (SAWGS), an eight-item instrument that measures sexism in the gaming community. Findings suggest that the SAWGS can help develop programmes to eradicate sexism and mitigate its negative effects on female gamers (Bustos-Ortega et al., 2023).

To maximise the potential of video games as a preventive tool against violence, there is a need for a systematic approach that integrates the representation of diversity. This will make a more effective contribution to combating hatred and promoting a fairer and prejudice-free society. Further research is needed on how to design video games that maximise this potential and effectively address gender issues and the power dynamics that perpetuate hate violence.

5. Conclusions

Research on video games brings to light both risks, such as addiction, racism, and their influence on moral development, as well as potential benefits, such as improvements in social and personal well-being. Individual motivations and sociocultural contexts are essential when assessing the effects of gaming. While video games can have a negative impact on behaviour, they also offer opportunities for education and self-reflection. Overall, the analysed studies were of good methodological quality. More research is needed to fill knowledge gaps and to develop educational strategies that maximise benefits while minimising risks, especially through interventions tailored to age, gender, and culture. Effective tools, such as content warnings and educational programmes, are crucial to address the risks associated with violent and antisocial media.

Educational games are often hampered by barriers such as a lack of inclusive design, which limits accessibility for diverse learners. It is also a challenge to strike a balance between entertainment and educational objectives, as overly entertaining elements may distract from learning, while an overly didactic approach risks disengaging players. This study suggests implementing better control strategies, in the form of stricter age verification for violent games and parental controls to mitigate the risks for younger players. Game developers should also incorporate prosocial narratives and diverse character representations to combat hate speech and discrimination. There should also be greater collaboration between researchers, educators, and policymakers when designing inclusive educational games.

The study has several limitations that need to be considered. Firstly, while it establishes a link between violent video games and aggressive behaviour, it does not thoroughly explore how individual traits and sociocultural factors, such as family background and gender, may influence this relationship. This limits the generalisability of its findings. In addition, the research recognises the potential of video games as educational tools but lacks a comprehensive analysis of how these games can be effectively designed to maximise their potential to prevent violence. Furthermore, the study highlights the prevalence of hate speech in games but does not provide empirical evidence on the effectiveness of moderation strategies in mitigating such content. Lastly, while the methodological quality of the studies analysed is considered to be good, there is a need for further research to address existing knowledge gaps and to develop tailored interventions that take into account age, gender, and cultural contexts. Future studies should explore the long-term impact of serious games, particularly their effectiveness in fostering empathy and reducing prejudice. Research could also investigate how specific game design elements contribute to sustained behavioural and attitudinal changes over time and influence gaming behaviours and outcomes.

In conclusion, video games pose undeniable risks, such as promoting addiction, perpetuating hate speech, and reinforcing discriminatory behaviours. However, they also hold immense potential as tools for education, social inclusion, and reducing prejudice. To fully leverage this potential, game developers must prioritise inclusive design that reflects diverse perspectives and promotes empathy, cooperation, and critical thinking. Policymakers and educators should work together to design interventions that harness the prosocial aspects of gaming while mitigating its harmful effects. Future research should focus on developing evidence-based strategies to address gender dynamics, reduce online toxicity, and explore the long-term effects of video games on intergroup relations. By taking a holistic approach that balances risks and opportunities, video games can become a powerful medium for building a more equitable and inclusive digital society.

Funding

This article is based on the Hatemedia project (PID2020-114584GB-I00), funded by MCIN/AEI/10.13039/501100011033, and the 32BITS project (SBPLY/21/180225/000108), funded by EU FEDER and Castilla-La Mancha Government.

Conflict of Interests

The authors declare no conflict of interest.

References

- Adelhardt, Z., & Eberle, T. (2019). Effects of a long-term separation from online gaming on teenagers' gaming habits. In W. Popma & S. Francis (Eds.), *ECSM 2019 6th European Conference on social media* (p. 1). Academic Conferences and Publishing International.
- Aguerri, J. C., Santisteban, M., & Miró-Llinares, F. (2023). The enemy hates best? Toxicity in League of Legends and its content moderation implications. *European Journal on Criminal Policy and Research*, 29(3), 437–456. <https://doi.org/10.1007/s10610-023-09541-1>
- Albaker, A. B., Abdalla, S. M., Almuhideb, A. S., Alobaid, O. A., Alanazi, F. M., alfadhel, T. O., & Alsakran, B. S. (2021). Impact of video games on physical and psychological wellbeing among medical students. *Journal of Research in Medical and Dental Science*, 9(5), 327–333. <https://www.jrmds.in/articles/impact-of-video-games-on-physical-and-psychological-wellbeing-among-medical-students-78003.html>
- Aleissa, M. A., Alenezi, S., Saleheen, H. N., Bin Talib, S. R., Khan, A. H., Altassan, S. A., & Alyahya, A. S. (2022).

- The association between video game type and aggressive behaviors in Saudi youth: A pilot study. *Behavioral sciences*, 12(8), Article 289. <https://doi.org/10.3390/bs12080289>
- Anderson, C. A., Suzuki, K., Swing, E. L., Groves, C. L., Gentile, D. A., Prot, S., Lam, C. P., Sakamoto, A., Horiuchi, Y., Krahe, B., Jelic, M., Liuqing, W., Toma, R., Warburton, W. A., Zhang, X. M., Tajima, S., Qing, F., & Petrescu, P. (2017). Media violence and other aggression risk factors in seven nations. *Personality & Social Psychology Bulletin*, 43(7), 986–998. <https://doi.org/10.1177/0146167217703064>
- Anti-Defamation League. (2022). *Hate is no game: Harassment and positive social experiences in online games 2021*. <https://www.adl.org/hateisnogame>
- Aouragh, M. (2016). Hasbara 2.0: Israel's public diplomacy in the digital age. *Middle East Critique*, 25(3), 271–297. <https://doi.org/10.1080/19436149.2016.1179432>
- Bacovsky, P. (2021). Gaming alone: Videogaming and sociopolitical attitudes. *New Media & Society*, 23(5), 1133–1156. <http://doi.org/10.1177/1461444820910418>
- Barter, C., Stanley, N., Wood, M., Lanau, A., Aghtaie, N., Larkins, C., & Øverlien, C. (2017). Young people's online and face-to-face experiences of interpersonal violence and abuse and their subjective impact across five European countries. *Psychology of Violence*, 7(3), 375–384. <https://doi.org/10.1037/vio0000096>
- Bogost, I. (2008). *Persuasive games: The expressive power of videogames*. MIT Press.
- Bos, D. (2023). Playful encounters: Games for geopolitical change. *Geopolitics*, 28(3), 1210–1234. <http://doi.org/10.1080/14650045.2021.2002846>
- Bowman, N. D., Kowert, R., & Cohen, E. (2022). *Video games and well-being: Press start*. Palgrave Macmillan.
- Breves, P. (2020). Reducing outgroup bias through intergroup contact with non-playable video game characters in VR. *Presence: Virtual and Augmented Reality*, 27(3), 257–273. https://doi.org/10.1162/pres_a_00330
- Bustos-Ortega, M., Carretero-Dios, H., Megías, J. L., & Romero-Sánchez, M. (2023). Sexist attitudes in online video gaming: Development and validation of the sexism against women gamers scale in Spanish and English. *Psychology of Women Quarterly*, 47(3), 320–342. <https://doi.org/10.1177/03616843231162837>
- Cabras, C., Cubadda, M. L., & Sechi, C. (2019). Relationships among violent and non-violent video games, anxiety, self-esteem, and aggression in female and male gamers. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 11(3), 15–37. <http://doi.org/10.4018/IJGCMS.2019070102>
- Charmaraman, L., Richer, A. M., & Moreno, M. A. (2020). Social and behavioral health factors associated with violent and mature gaming in early adolescence. *International Journal of Environmental Research and Public Health*, 17(14), Article 4996. <http://doi.org/10.3390/ijerph17144996>
- Cipagauta, C. A., & Gómez, M. B. (2019). Application of video games in education. *Revista Vínculos*, 16(1), 104–109. <https://doi.org/10.14483/2322939X.15465>
- Clarice, R., Callou, M., Jorgeanne, S., Bezerra, B., Lira, F. T., Moreira, S., Moreira Belém, J., & Albuquerque, G. A. (2021). Cyberbullying e violência de gênero em jogos online. *Saúde e Pesquisa*, 14(3), 543–554. <https://doi.org/10.17765/2176-9206.2021V14N3E7920>
- Costa, S., da Silva, B. M., & Tavares, M. (2021). Video games and gamification against online hate speech? In A. Fernandes-Marcos, P. Bernardino Bastos, M. Manuela Lopes, A. Araújo, & L. Fabian Olivero (Eds.), *ACM International Conference Proceeding Series* (Article No. 35). Association for Computing Machinery. <https://doi.org/10.1145/3483529.3483679>
- Costa, S., Martins, A., Martins, A. F., & Gavina, A. (2024). Online hate speech in video games: Concepts, prevalence, and prevention strategies. In B. M. da Silva & A. F. Martins (Eds.), *Interactive narratives propose pluralist speech: Results of a project aimed at researching and countering online hate speech in video games* (pp. 21–60). CIAC Editions. <https://doi.org/10.34623/ggqf-Og28>

- da Silva, R. C., & Ifa, S. (2022). Percepção de discursos de ódio em jogos eletrônicos online por adolescentes de Alagoas. *SOLETRAS*, 2022(43), 30–54. <https://doi.org/10.12957/soletras.2022.63256>
- da Silva, B. M., Tavares, M., Cerol, F., da Silva, S. M., Falcão, P., & Alves, B. I. (2020). Playing against hate speech—How teens see hate speech in video games and online gaming communities. *Journal of Digital Media & Interaction*, 3(6), 34–52. <https://doi.org/10.34624/jdmi.v3i6.15064>
- Dickmeis, A., & Roe, K. (2019). Genres matter: Video games as predictors of physical aggression among adolescents. *Communications*, 44(1), 105–129. <https://doi.org/10.1515/COMMUN-2018-2011/MACHINEREADABLECITATION/RIS>
- Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology & Community Health*, 52(6), 377–384. <https://doi.org/10.1136/jech.52.6.377>
- Ferguson, C. J., & Colwell, J. (2020). Sexualised video games, sexist attitudes and empathy towards victims of rape: Correlational evidence for a relationship is minimal in an online study. *Criminal Behaviour and Mental Health*, 30(1), 16–27. <http://doi.org/10.1002/cbm.2138>
- Ferguson, C. J., Copenhaver, A., & Markey, P. (2020). Reexamining the findings of the American Psychological Association’s 2015 Task Force on Violent Media: A meta-analysis. *Perspectives on Psychological Science*, 15(6), 1423–1443. <https://doi.org/10.1177/1745691620927666>
- Fleet, R., & Nurmikko-Fuller, T. (2019, June). The potential for serious spaceships to make a serious difference. In *Proceedings of the 10th ACM conference on Web Science* (pp. 97–104). Association for Computing Machinery. <http://doi.org/10.1145/3292522.3326017>
- Gilbert, M. A., Giaccardi, S., & Ward, L. M. (2018). Contributions of game genre and masculinity ideologies to associations between video game play and men’s risk-taking behavior. *Media Psychology*, 21(3), 437–456. <https://doi.org/10.1080/15213269.2018.1450153>
- Gong, H., & Piller, Y. (2018). Differences in parental involvement and perception of video games: A pilot study on American-born and immigrant parents. *Eurasia Journal of Mathematics, Science and Technology Education*, 14(3), 785–796. <http://doi.org/10.12973/ejmste/80913>
- Guggisberg, M. (2020). Sexually explicit video games and online pornography—The promotion of sexual violence: A critical commentary. *Aggression and Violent Behavior*, 53, Article 101432. <https://doi.org/10.1016/J.AVB.2020.101432>
- Gurlesin, O., Akdag, M., Alasag, A., & Avest, I. T. (2020). Playful religion: An innovative approach to prevent radicalization of Muslim youth in Europe. *Religions*, 11(2), Article 67. <https://doi.org/10.3390/rel11020067>
- Halbrook, Y. J., O’Donnell, A. T., & Msetfi, R. M. (2019). When and how video games can be good: A review of the positive effects of video games on well-being. *Perspectives on Psychological Science*, 14(6), 1096–1104. <http://doi.org/10.1177/1745691619863807>
- Hartgers, M., & Leidig, E. (2023, June 1). Fighting extremism in gaming platforms: A set of design principles to develop comprehensive P/CVE strategies. *International Centre for Counter-Terrorism*. <https://www.icct.nl/publication/fighting-extremism-gaming-platforms-set-design-principles-develop-comprehensive-pcve>
- Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2019). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, 16(2), 114–121. <https://doi.org/10.1016/j.tics.2011.12.007>
- Hodge, S. E., Taylor, J., & McAlaney, J. (2019). It’s double edged: The positive and negative relationships between the development of moral reasoning and video game play among adolescents. *Frontiers in Psychology*, 10, Article 28. <http://doi.org/10.3389/fpsyg.2019.00028>

- Hodge, S. E., Taylor, J., & McAlaney, J. (2020). Is it still double edged? Not for university students' development of moral reasoning and video game play. *Frontiers in Psychology*, 11, Article 1313. <http://doi.org/10.3389/fpsyg.2020.01313>
- Imran, A., Khanom, N., & Rahman, A. (2023). Parental perception of children's online behaviour: A study on ethnic communities in Australia. *International Journal of Environmental Research and Public Health*, 20(7), Article 5342. <https://doi.org/10.3390/ijerph20075342>
- Irmak, A. Y., & Erdoğan, S. (2019). Predictors for digital game addiction among Turkish adolescents: A Cox's interaction model-based study. *Journal of Addictions Nursing*, 30(1), 49–56. <http://doi.org/10.1097/JAN.0000000000000265>
- Kaakinen, M., Sirola, A., Savolainen, I., & Oksanen, A. (2020). Impulsivity, internalizing symptoms, and online group behavior as determinants of online hate. *PloS ONE*, 15(4), Article 0231052. <http://doi.org/10.1371/journal.pone.0231052>
- Kaye, L. K., & Pennington, C. R. (2016). "Girls can't play": The effects of stereotype threat on females' gaming performance. *Computers in Human Behavior*, 59, 202–209. <https://doi.org/10.1016/j.chb.2016.02.020>
- Keipi, T., Räsänen, P., Oksanen, A., Hawdon, J., & Näsi, M. (2018). Exposure to online hate material and subjective well-being: A comparative study of American and Finnish youth. *Online Information Review*, 42(1), 2–15. <http://doi.org/10.1108/OIR-05-2016-0133>
- Khalid, F., & El-Maliki, T. (2020). Teachers' experiences in the development of digital storytelling for cyber risk awareness. *International Journal of Advanced Computer Science and Applications*, 11(2). <http://doi.org/10.14569/IJACSA.2020.0110225>
- Koehler, D., Fiebig, V., & Jugl, I. (2022). From gaming to hating: Extreme-right ideological indoctrination and mobilization for violence of children on online gaming platforms. *Political Psychology*, 44(2), 419–434. <https://doi.org/10.1111/POPS.12855>
- León, O. G., & Montero, I. (2020). *Métodos de investigación en psicología y educación*. McGraw-Hill.
- Li, Y. (2022). Linking violent video games to cyberaggression among college students: A cross-sectional study. *Aggressive Behavior*, 48(2), 241–252. <https://doi.org/10.1002/AB.22002>
- Liby, C., Doty, J. L., Mehari, K. R., Abbas, I., & Su, Y.-W. (2023). Adolescent experiences with online racial discrimination: Implications for prevention and coping. *Journal of Research on Adolescence* 33(4), 1281–1294. <https://doi.org/10.1111/jora.12875>
- Lippe, F., Walter, R., & Hofinger, V. (2022). Evaluating an online-game intervention to prevent violent extremism. *Journal for Deradicalization*, 32, 1–34. <https://journals.sfu.ca/jd/index.php/jd/article/view/641>
- Lischer, S., Jeannot, E., Brülisauer, L., Weber, N., Khazaal, Y., Bendahan, S., & Simon, O. (2022). Response to the regulation of video games under the Youth Media Protection Act: A public health perspective. *International Journal of Environmental Research and Public Health*, 19(15), Article 9320. <https://doi.org/10.3390/ijerph19159320>
- Madden, D., Liu, Y., Yu, H., Sonbudak, M. F., Troiano, G. M., & Hartevelt, C. (2021, May). "Why are you playing games? You are a girl!": Exploring gender biases in e-sports. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (Article 323). Association for Computing Machinery. <http://doi.org/10.1145/3411764.3445248>
- Makarova, E. A., & Makarova, E. L. (2019). Aggressive behavior in online games and cybervictimization of teenagers and adolescents. *International Electronic Journal of Elementary Education*, 12(2), 157–165. <https://doi.org/10.26822/iejee.2019257663>
- Marín-Díaz, V., & Cabero-Almenara, J. (2019). Las redes sociales en educación: Desde la innovación a la investigación educativa. *Revista iberoamericana en educación a distancia*, 22(2), 25–33. <https://doi.org/10.5944/ried.22.2.24248>

- Martín-Martín, M., Asensio-Muñoz, I., & Bueno-Álvarez, J. (2021). Uso de las redes sociales en estudiantes de secundaria: Análisis de perfiles para la intervención educativa. *Revista complutense de educación*, 32(3), 303–314. <http://doi.org/10.5209/rced.57189>
- Mason, R., & Turner, L. (2018). Serious gaming: A tool to educate health care providers about domestic violence. *Health Care for Women International*, 39(8), 859–871. <https://doi.org/10.1080/07399332.2018.1464572>
- Miranda-Palma, C. A., Canche, M., Narváez, L. E., Chi, V. M., & Llanes-Castro, E. R. (2023). Methodology that combines entertainment and learning to develop educational videogames. *Revista Espacios*, 44(2), 30–44. <https://doi.org/10.48082/espacios-a23v44n02p03>
- Moreno, A. C. (2022). Terrorism, war of ideas and video games: Theory and practice. *Araucaria*, 24(50), 381–405. <https://doi.org/10.12795/araucaria.2022.i50.16>
- Moreno-López, R., López-Pérez, M., & Sandulescu-Budea, A. (2021). Hábitos de uso de las redes sociales entre los jóvenes universitarios españoles. *Bibliotecas. Anales de investigación*, 17(4), 179–194. <https://dialnet.unirioja.es/servlet/articulo?codigo=8320368>
- Obermaier, M., & Schmuck, D. D. S. (2022). Youths as targets: Factors of online hate speech victimization among adolescents and young adults. *Journal of Computer-Mediated Communication*, 27(4). <https://doi.org/10.1093/JCMC/ZMAC012>
- Olejarnik, J., & Romano, E. (2023). Is playing violent video games a risk factor for aggressive behaviour? Adding narcissism, self-esteem and PEGI ratings to the debate. *Frontiers in Psychology*, 14, Article 1155807. <https://doi.org/10.3389/fpsyg.2023.1155807>
- Olson, D. M., & Harrell, D. F. (2020). “I don’t see color”: Characterizing players’ racial attitudes and experiences via an anti-bias simulation videogame. In G. N. Yannakakis, A. Liapis, P. Kyburz, V. Volz, F. Khosmood, & P. Lopes (Eds.), *FDG ’20: Proceedings of the 15th International Conference on the Foundations of Digital Games* (Article 89). Association for Computing Machinery. <https://doi.org/10.1145/3402942.3409783>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, Article 71. <https://doi.org/10.1136/bmj.n71>
- Paz, M. A., Montero-Díaz, J., & Moreno-Delgado, A. (2020). Hate speech: A systematized review. *Sage Open*, 10(4). <https://doi.org/10.1177/2158244020973022>
- Pech, G. P., & Caspar, E. A. (2022). Can a video game with a fictional minority group decrease intergroup biases towards non-fictional minorities? A social neuroscience study. *International Journal of Human-Computer Interaction*, 40(2), 482–496. <https://doi.org/10.1080/10447318.2022.2121052>
- Pohjonen, M., & Udupa, S. (2017). Extreme speech online: An anthropological critique of hate speech debates. *International Journal of Communication*, 11, 1173–1191. <https://ijoc.org/index.php/ijoc/article/view/5843/1965>
- Pusey, M., Wong, K. W., & Rappa, N. A. (2022). Using case studies to explore need satisfaction and frustration in puzzle video games. *Games and Culture*, 17(5), 752–772. <https://doi.org/10.1177/15554120211056126>
- Rančić, J. (2018). Vršnjačko nasilje na društvenim mrežama u Republici Srbiji. *CM: Communication and Media*, 13, 95–124. <https://doi.org/10.5937/comman13-18786>
- Reer, F., & Krämer, N. C. (2018). Psychological need satisfaction and well-being in first-person shooter clans: Investigating underlying factors. *Computers in Human Behavior*, 84, 383–391. <http://doi.org/10.1016/j.chb.2018.03.010>
- Ripoll, C. O., & Muñoz, K. A. (2023). Discrimination in video games: The Fortnite case. *Communication Papers: Media Literacy and Gender Studies*, 12(24), 121–137. https://doi.org/10.33115/udg_bib/cp.v12i24.22906

- Salter, A., & Blodgett, B. (2012). Hypermasculinity & Dickwolves: The contentious role of women in the new gaming public. *Journal of Broadcasting and Electronic Media*, 56(3), 401–416. <https://doi.org/10.1080/08838151.2012.705199>
- Salter, M. (2018). From geek masculinity to Gamergate: The technological rationality of online abuse. *Crime, Media, Culture*, 14(2), 247–264. <http://doi.org/10.1177/1741659017690893>
- Shahghasemi, E. (2018). How online game players perceive pervasiveness of violence in the society. In 2018 2nd National and 1st International Digital Games Research Conference: Trends, Technologies, and Applications (DGRC) (pp. 134–144). IEEE.
- Shinohara, T. (2024). Brief overview of the new youth protection law and e-sports in Switzerland. *Gaming Law Review*, 28(1), 2–7. <https://doi.org/10.1089/GLR2.2022.0043>
- Shortland, N., Nader, E., Thompson, L., & Palasinski, M. (2022). Is extreme in the eye of the beholder? An experimental assessment of extremist cognitions. *Journal of interpersonal violence*, 37(7/8), NP4865–NP4888. <http://doi.org/10.1177/0886260520958645>
- Sicart, M. (2011). *The ethics of computer games*. MIT Press.
- Sung, Y. H., & Lee, W. N. (2020). Doing good while playing: The impact of prosocial advergames on consumer response. *Computers in Human Behavior*, 106, Article 106244. <http://doi.org/10.1016/j.chb.2020.106244>
- TaeHyuk Keum, B., & Hearn, M. (2022). Online gaming and racism: Impact on psychological distress among Black, Asian, and Latinx emerging adults, 17(3), 445–460. <https://doi.org/10.1177/15554120211039082>
- Vicentini, L., Kirst, F., De-Bortoli, A., & De-Bortoli, R. (2023). Learning hate speech in e-sports. *Psicologia, Saúde & Doença*, 24(1), 148–160. <https://doi.org/10.15309/23PSD240113>
- Waddell, T. F. (2020). Do press releases about digital game research influence presumed effects? How details about methodology and references to societal violence affect the anticipated influence of violent video games. *Mass Communication and Society*, 23(3), 400–420. <https://doi.org/10.1080/15205436.2019.1672751>
- Wearing, S. L., Porter, D., Wearing, J., & McDonald, M. (2022). Exploring adolescent computer gaming as leisure experience and consumption: Some insights on deviance and resistance. *Leisure Studies*, 41(1), 28–41. <http://doi.org/10.1080/02614367.2021.1942525>
- Wu, Y., Hu, J., & Li, W. (2023). The link between online gaming behaviour and unethical decision-making in emerging adults: The mediating roles of game cheating and moral disengagement. *Behaviour & Information Technology*, 42(10), 1534–1547. <http://doi.org/10.1080/0144929X.2022.2087539>
- Ybarra, M. L., Mitchell, K. J., & Oppenheim, J. K. (2022). Violent media in childhood and seriously violent behavior in adolescence and young adulthood. *Journal of Adolescent Health*, 71(3), 285–292. <http://doi.org/10.1016/j.jadohealth.2022.03.003>
- Yuldasheva, R., & Mukhopadhyay, B. R. (2022). You are what you watch? Evaluating the impact of media violence on youth behaviour during the COVID-19. *International Journal of Applied Behavioral Economics*, 11(1). <https://doi.org/10.4018/IJABE.308784>
- Zhang, P., Lee, H. M., Zhao, K., & Shah, V. (2019). An empirical investigation of eWOM and used video game trading: The moderation effects of product features. *Decision Support Systems*, 123, Article 113076. <https://doi.org/10.1016/J.DSS.2019.113076>
- Zhao, S., & Liu, Y. (2023). From industry development to social influence: Video games in Chinese newspaper coverage, 2010–2020. *Chinese Journal of Communication*, 16(1), 35–52. <http://doi.org/10.1080/17544750.2022.2108860>
- Zhu, X. W., Chu, X. W., Zhang, Y. H., & Li, Z. H. (2020). Exposure to online game violence and cyberbullying among Chinese adolescents: Normative beliefs about aggression as a mediator and trait aggressiveness

as a moderator. *Journal of Aggression, Maltreatment & Trauma*, 29(2), 148–166. <https://doi.org/10.1080/10926771.2018.1550830>

About the Authors



Roberto Moreno-López has a PhD in education from the University of Castilla-La Mancha (Spain) and is a tenure-track lecturer at the Department of Pedagogy of the same university. His main lines of research currently focus on social and educational inclusion and hate crimes and hate speech and their consequences for society, especially for young people. Accredited by ANECA and with a six-year research period from ANECA-CNEAI, he has also been granted with a positive evaluation as tenure-track lecturer.



Catalina Argüello-Gutiérrez has a PhD in social psychology from the University of Granada (UGR) and is a tenure-track lecturer at the Department of Education of Universidad Internacional de La Rioja (Spain). Her main research interests are intergroup relations, social cognition, intercultural education, and mental health. She is an expert in group psychology and stress and trauma management. Accredited by ANECA and with a six-year research period from ANECA-CNEAI (Spain).

Androcentrism and Violence in Online Video Games: Perpetuation of Gender Inequality

Beatriz Esteban-Ramiro , Patricia Fernández de Castro , and Orlanda Díaz-García 

Department of Labour Law and Social Work, University of Castilla-La Mancha, Spain

Correspondence: Beatriz Esteban-Ramiro (beatriz.esteban@uclm.es)

Submitted: 15 October 2024 **Accepted:** 14 January 2025 **Published:** 18 February 2025

Issue: This article is part of the issue “Violence, Hate Speech, and Gender Bias: Challenges to an Inclusive Digital Environment” edited by Max Römer Pieretti (Universidad Camilo José Cela), Beatriz Esteban-Ramiro (Universidad de Castilla-La Mancha), and Agrivalca Canelón (Universidad Católica Andrés Bello), fully open access at <https://doi.org/10.17645/si.i415>

Abstract

This article aims to offer an analysis of the presence of hate speech, harassment, and cyber-violence towards women in the interaction within massively multiplayer online video games (MMOs), an act of gender-based violence perpetrated through new technologies in an environment generally described as androcentric. Filling a gap in the experience in Spain, a qualitative methodology was applied that allowed us to analyse the forms of harassment towards female video gamers, as well as some of the consequences of experiencing these situations for them. A systematic analysis of the discourses was carried out after five focus groups and 15 in-depth interviews with young people between 18 and 29. Amongst the results, it is worth highlighting that, as occurs in offline reality, in MMOs gender schemas and asymmetric relationships are reproduced that perpetuate gender-based violence. On the one hand, female gamers affirm that they suffer sexist violence in gaming environments through the use of language focused on physical appearance, hypersexualisation, and undervaluation, and are often considered intruders. On the other hand, they also report experiencing overprotective and paternalistic behaviours. Female gamers sometimes decide to leave their games, or avoid certain types of games a priori, because of toxic behaviours perpetrated by male gamers, which generates feelings of discrimination in their experiences as players, perpetuating androcentric patterns in this sort of environment. Applying a gender approach, this article will examine the social implications of these behaviours in order to propose appropriate social and educational responses.

Keywords

cybermisogyny; cybersexism; cyberviolence; gender; hate speech; MMOs; videogames; young people

1. Introduction

For young people, online environments have become important spaces for interaction and socialisation. As well as providing entertainment, these spaces allow young people to build relationships, develop social skills, and shape their identities (B. Braun et al., 2016; de Hesselle et al., 2021; Romo et al., 2023). Online video games, in particular, have become embedded in contemporary culture, serving as one of the main forms of entertainment, communication, and socialisation for young people. As such, they are not only technological but also sociocultural tools (Martínez-Verdú, 2023; Voorhees, 2018). This is reflected not only in increasing turnover and business growth figures, but also in the number of users (Calderón-Gómez & Gómez-Miguel, 2023). In 2023, the number of Spanish gamers reached 20 million, of which 51% are male and 49% female, with an average weekly playing time of more than seven hours. It is estimated that 86% of young people aged 15–24 play video games (Asociación Española de Videojuegos, 2024). The penetration of video games into society via the younger generations is very much a reality, with around three out of four adolescents and young people in Spain playing video games (Calderón-Gómez & Gómez-Miguel, 2022).

The ability to play online with people from all over the world is increasingly appealing to many young people. Massively multiplayer online video games (MMOs) allow them to play with people from across the globe, creating virtual worlds where they can cooperate, compete, and communicate (Raith et al., 2021). Authors like Kafai et al. (2008) or Taylor (2008) discuss how these virtual worlds and casual games have attracted a wider female audience. These spaces build communities and social networks that influence the development of values and behaviours, but they are also vulnerable to cyberhate (Arroyo-López & Esteban-Ramiro, 2022; Costa et al., 2021). However, they lack adequate safeguards, exposing young people to risks such as privacy violations and direct exposure to online violence and bullying (Aghazadeh et al., 2018; Marín-Díaz & Cabero-Almenara, 2019; Tang et al., 2020).

1.1. Video Games as Androcentric Spaces

Despite almost half of gamers being female, the stereotype that video games are a male pastime persists (McLean & Griffiths, 2019; Mendick et al., 2021; Paaßen et al., 2017). Online video games are seen as a “man’s world” where interactions reinforce masculine norms (Blackburn & Scharrer, 2019; Cote, 2020). Within this androcentric gamer culture, masculinity is associated with competitiveness, dedication, and superior skill (Shaw, 2012). As a result, many women do not identify with “the gamer” stereotype and are excluded from this category (Kuss et al., 2022; Vermeulen et al., 2017). Studies like Taylor’s (2003) specified how women interact with games that are not explicitly designed for them, broadening the debate on gender and technology.

The increasing presence of women in online video games is perceived as a threat by many male gamers, leading to feelings of discrimination and rejection (Amores, 2023; Bustos-Ortega et al., 2024; Tang et al., 2020). This marginalisation also affects women in fields such as design, programming, and content creation (Aghazadeh et al., 2018; Isaaman & Tolaine-Sage, 2022). Vossen (2018) introduces the concept of “cultural inaccessibility” to describe how women are perceived as outsiders in gaming culture, facing obstacles that make them feel unwelcome. In this way, the gaming industry and its internal culture maintain contradictory narratives about inclusion and the persistence of sexism (Cote, 2020). Video games can be toxic spaces for women, with misogyny and harassment deeply ingrained in the culture (Jenson & de Castell, 2021).

Stereotypes and expectations towards women have a negative impact on their performance. Facing comments such as “You play like a girl” affects their gameplay (Kaye & Pennington, 2016; Vermeulen et al., 2017), influencing their self-identification and participation in video games (Taylor, 2008). Furthermore, a gender hierarchy has been normalised in these environments, presenting women as a threat to the male status quo (Jenson & de Castell, 2021) and relegating them to a subordinate position within the gaming domain (Bustos-Ortega, 2024; Gray et al., 2017).

1.2. Toxicity in Gaming: Online Violence, Hate Speech, and Sexism Towards Women Gamers

In MMOs, players often comment on the varying levels of toxicity between communities. Research has looked at how tension between players and antisocial behaviours manifest through insults and offensive language in online gaming (Kwak & Blackburn, 2015; Tang et al., 2020). They include hate speech motivated by gender, race, or sexual orientation, which further exacerbates the risks of harassment in these environments for people who belong to these groups. Additionally, ethnographic studies like Shaw’s (2014) highlight how the social representation of gender—like other characteristics, like being LGBTQ, influences experiences and generates inequality and exclusion among players. According to the Anti-Defamation League (2022), three in five young people aged 13–17 have experienced harassment in online video games, affecting approximately 14 million in the United States. Anonymity and the normalisation of hate speech in these spaces is a growing social and academic concern (Blaya & Audrin, 2019), with studies showing high levels of hate in these environments (Reichelmann et al., 2020; Wachs et al., 2022). It is essential to consider the unique characteristics of the online context—such as anonymity, competition, and social disinhibition—that distinguish it from the offline gaming environment (Ruvalcaba et al., 2018). Feelings of insecurity, particularly among players from vulnerable groups, are heightened by the lack of effective mechanisms to curb such behaviour.

There has been a growing interest in defining the forms of sexism experienced by women gamers, with studies highlighting the various ways in which it manifests (Fox & Tang, 2017; McLean & Griffiths, 2019; Tang et al., 2020). These studies are informed by ambivalent sexism theory (Glick & Fiske, 1996, 2011), which describes the coexistence of both hostile and benevolent sexism as forms of female subordination. Numerous cases of harassment and hostility towards women gamers have been reported in the US, accounting for 25% of the worldwide total (Entertainment Software Association, n.d.). This situation is described as “cyber gender violence” or “cyber misogyny,” where male chauvinist stereotypes and roles are played out in online spaces through visual representations and power dynamics (Donoso-Vázquez, 2018; Jane, 2018). To analyse this kind of cyber misogyny, Bustos-Ortega et al. (2023) have developed the Sexism Against Women Gamers Scale that takes into account the particular characteristics of an online gaming environment. This includes underestimating aggression towards women gamers, considering them too sensitive, and belittling their interest and ability.

The rise of online video games has raised concerns about discrimination against women gamers, manifested in forms such as sexual harassment and sexist comments (Amores, 2023). The repercussions of these incidents go beyond simply abandoning a game; women gamers experience anxiety, depressive symptoms, stress, and negative effects on their identity as gamers (Fox et al., 2018; McLean & Griffiths, 2019). This creates a sense of helplessness that leads women to conceal their gender, avoid communication, or even drop out of gaming altogether (Cote, 2020; Fox & Tang, 2017). This hostility has become normalised as part of the gaming experience, downplaying the impact it has on women gamers (Vergel et al., 2024).

Most of these studies have been conducted from a quantitative perspective, focusing on the percentage of women who have experienced such situations or on the relationship of these experiences with various study variables (Fox & Tang, 2017; McLean & Griffiths, 2019). The most relevant findings have generally been developed in English-speaking contexts, leaving gaps in research for Spanish-speaking cultural contexts. Regarding young people in Spain, there is limited literature on the subject, with some notable publications exploring hostile behaviours and sexism against women in online video games (Bustos-Ortega, 2024; Calderón-Gómez & Gómez-Miguel, 2023; Santana-Rodríguez, 2020). These studies conclude that experiencing violence and hate in these spaces often has a strong gender component, pointing to the sexism inherent in these contexts.

Moreover, a significant corpus of research in English-speaking countries has used qualitative methodologies, such as content analysis, ethnography, and interviews, to delve into women's meanings and interactions in online video games and MMOs. These studies explore how gender stereotypes and misogyny affect women's participation and enjoyment of gaming, as well as power dynamics within the industry (Cote, 2020; Gray & Leonard, 2018). They also address key events such as the 2014 Gamergate movement, which intensified harassment against women journalists, developers, and academics speaking out against sexism in gaming culture (Vossen, 2018). Furthermore, these studies examine how underrepresentation of identities based on gender, race, and sexuality creates exclusion and cultural inaccessibility for certain groups (Gray & Leonard, 2018; Shaw, 2014; Taylor, 2008). They adopt a critical and comprehensive approach, interpreting reality through the meanings attributed by individuals and analysing how gender influences the creation, consumption, and perception of video games, as well as how feminism can drive a more inclusive and equitable industry (Jenkins & Cassell, 2008).

In Spain, there is little research that captures women's direct experiences of harassment and sexism in video games using qualitative methodologies. Some studies have examined gaming culture broadly in terms of gamer identity, but these do not specifically focus on understanding women's experiences (Muriel, 2018). Moldes (2019) analysed the working conditions of female esports players, while Afonso Noda and Aguilera Ávila (2021) explored feelings, experiences of inequality, and opinions about exclusive spaces for women in gaming, using only six semi-structured interviews. More recently, Mihura-López et al. (2023) investigated the experiences of female gamers, showing how violence and toxicity affect their participation, often leading to self-exclusion or rejection of the gamer label due to its toxic connotations. However, their findings rely on a single focus group.

The limited transferability of findings from previous research underscores the need for further exploration, as proposed in this study. Calderón-Gómez and Gómez-Miguel (2023, p. 11) emphasise the importance of studying the experiences of women gamers in Spain, focusing on violence and sexual harassment in online multiplayer games through quantitative methodologies.

The absence of qualitative research in Spanish limits a deeper understanding of how gender dynamics in gaming affect Spanish-speaking women. Qualitative studies are essential to gather direct testimonies and explore personal narratives, offering a more comprehensive perspective on the impact of sexism and gender representation in this community.

In order to explore interactions in online games in greater depth, we propose a feminist approach that examines gender bias. This article has two aims: (a) to analyse the presence, typology, and manifestations of

hate speech, online violence, and sexism against women in multiplayer video games; and (b) to determine personal and social effects of these behaviours on women through their own experiences as gamers.

2. Methodology

2.1. Design

The study used a qualitative methodological design to explore phenomena such as violence, hate, and sexism in video games through the experiences of women gamers. The aim is to ask questions that help to reconstruct reality as observed by subjects within a defined social system (Sampieri, 2018) and to know how this reality acts specifically in the Spanish context since there is no previous solid evidence. Semi-structured interview ($n = 15$) and focus group ($n = 5$) techniques were used. This study is part of a larger research project approved by the Ethics Committee of the University of Castilla-La Mancha in 2023. Ethical principles, confidentiality, and protection of personal data were guaranteed throughout the process.

This type of methodology captures the complexity of the problem and attempts to understand the coping strategies, impact, and social and cultural implications of sexism in online gaming spaces. It also addresses the power structures and gender norms at play in these environments.

2.2. Participants and Sample

A total of 49 participants were involved, 38 of whom were women. Fifteen of the women participated in semi-structured interviews, while the others took part in focus groups, as shown in Table 1.

Table 1. Participants and techniques.

Technique	Total no. of participants	No. of women	No. of men
Focus groups	34	23	11
Semi-structured interviews	15	15	0
Total	49	38	11

The participants, aged 18–29 years ($M = 25.2$), were mostly students (87.7%, $n = 43$), while 12.3% ($n = 6$) were employed. Ninety-four percent were Spanish nationals and six percent ($n = 3$) had dual nationality (Spanish and Romanian or Chilean). All participants lived or studied in Castilla-La Mancha, the geographical area covered by the project.

Recruitment was voluntary and targeted at “young gamers,” with a particular focus on attracting *women* gamers. Three of the five focus groups were mixed in order to analyse whether men recognised the same misogynistic situations during gameplay or if—on the contrary—they had other experiences different from the girls. A non-probabilistic snowball sampling method was therefore used, with different sample selection strategies:

1. Using an earlier study with surveys in educational centres, where interested participants left their data;
2. Online dissemination via Instagram profiles: one from a collaborating organisation (@INEEYS) and another from an influential gamer (@feminismoen8bits), which encouraged women who were initially interested in participating to recommend other women.

The focus groups were made up entirely of people who had heard about the study at their educational establishments, while the interviews were proportionally split between the two recruitment methods.

2.3. Procedure

Fifteen in-depth interviews and five focus groups were conducted between May and September 2024. Participants who provided their contact details were invited to the interviews after giving informed consent and agreeing on their preferred method (in-person or Microsoft Teams). All interviews were conducted in private, with informed consent to record the session and assurances of confidentiality of participants' personal data. The interviews and focus groups were conducted by three members of the research team.

The interviews were structured into eight thematic blocks with 23 open-ended questions aligned with the objectives of the study. At the end of each interview, participants were invited to share any additional thoughts on the situation of women gamers.

The focus groups, consisting of six or seven participants and a researcher, were designed to verify and supplement the narratives from the interviews. For this reason, the groups were mixed to capture the perceptions of some men about behaviours towards women in gaming as a contrast method, since it was not one of the central objectives of the study. A series of categories were established to guide the group discussion of the issues (identification of video games; identity; online violence, harassment, and sexism; personal and social impacts; measures and codes of conduct; and proposals for effective and awareness-raising initiatives). Members of the research team were also provided with a list of possible questions to help generate topics of interest.

2.4. Analysis

On completion of the interviews and focus groups, the data were anonymised and fully transcribed by two researchers, who then familiarised themselves thoroughly with the information collected. Initial coding was then carried out to ensure that data saturation had been achieved (Guest et al., 2020).

Different semantic codes were used to generate independent themes, which were then compared. This process identified recurring patterns that were grouped into thematic lines and sub-themes. This process followed the guidelines of V. Braun and Clarke (2006). As a quantitative study had been carried out beforehand, it was found that most of the categories corresponded to the original proposals. However, additional sub-themes were included to allow for a more detailed and in-depth analysis of the situations under investigation.

A structural discourse analysis was carried out, following the approach proposed by Conde (2009). The focus is on the interest in discourse for the knowledge of social reality. Considering the context in which they are produced, it is about knowing how the group constructs social reality—how social discourses shape and give identity to these subjects. On the basis of this analysis, a classification was established that identified discursive patterns, analysed semantic spaces, and determined discursive positions. This enabled the identification of narrative configurations that show how experiences affect the identities of women gamers and their participation in the community, as detailed in the results section.

To ensure internal homogeneity and external heterogeneity, a committee of research team members reviewed, grouped, and reorganised the themes. This process ensured internal consistency and the inclusion of all issues raised by respondents. The codes used for the narratives were ME (*mujer entrevista*) followed by a number for the women interviewed (e.g., ME 7). For the narratives of women and men in groups we use M or HG (*mujeres/hombres en grupos*) followed by a letter (identifying the group) and a number (identifying the individual). It should be noted that the interviews were conducted and transcribed in Spanish, an English translation of this text has been made.

It is important to note certain limitations of the study. It applies a qualitative methodology, which limits its generalisability, although it provides important knowledge for understanding the phenomenon. The research is limited to the geographical area of Castilla-La Mancha in Spain and involves a sample of young people, collecting the direct experiences of female players only up to the age of 29. To avoid response bias, participants were informed that the aim was to explore their experiences as video gamers. However, beforehand they signed informed consent documentation with all the information about the study. Despite this, all participants were free to express their opinions.

3. Results and Discussion

Five main themes emerged from the initial analysis. Following the coding and categorisation exercise, sub-themes were added to these main themes (which aligned with the overarching categories of the wider research) to provide greater insight into the narratives, as shown in Table 2.

The findings are thus presented bearing in mind how gender mandates cut across these new relational spaces in the context of online video games. The following sections outline each of the main themes and their development within the structural analysis of the narratives concerned. Finally, we will map the narrative configurations by identifying the discursive positions that emerged from the initial discourse analysis.

Table 2. Identified themes and sub-themes.

Gaming habits and experiences	Manifestations of violence: hatred and sexism towards women gamers	Contributing factors	Consequences of and responses to violence	Countermeasures and actions
Video games as spaces for socialisation	Insults, harassment, and humiliation; sexual objectification, belittling, intrusion	Regarding women as outsiders and threats	Personal and social impacts (psycho-social impacts)	Ineffectiveness and inadequacy of measures
Identity concealment	Overprotective behaviour, paternalism, and possessiveness.	Anonymity	Downplayed/ minimisation and failure to call out subjugation	Lack of adaptation to real-life gaming contexts
Gender roles and stereotypes in characters and avatars		Impunity	Building support networks: community	Identification of mediators

3.1. Gaming Habits and Experiences

The most widely played video games include League of Legends (LOL), World of Warcraft (WOW), Host, Valorant, and Fortnite. Most participants have many years of experience playing video games, making them a fundamental space for entertainment and social interaction.

3.1.1. Video Games as Spaces for Socialisation

Some participants started gaming in childhood (at the of 6, 8, or 10), but most report having begun as teenagers. The Covid-19 pandemic and lockdown are mentioned as pivotal moments that increased the amount of time spent playing video games. For them, gaming has become one of their main leisure activities, where they can unwind and meet up with friends:

To be honest, when I started, you end up joining a gaming community and you don't know anyone there, but over time, you meet people and that's cool. For example, I'm still in touch with the people I first started talking to back in 2020 and we do organise meet-ups, I see them every month or every other month and in the end they've become very close friends. (MGB-3)

Some of the participants reported having entered into romantic relationships:

I began a romantic relationship and everything, believe it or not! We just started chatting, exchanged Instas [referring to the social network Instagram] and after a while we met up and all that and it took off. (ME 7)

The importance of this online space as a vehicle for social interaction is noted, and it is seen as a sociocultural as well as a technological environment (Martínez-Verdú, 2023), where young people use that to build communities, relationships, and identities (Tang et al., 2020) which positions them not only as recreational spaces but as significant cultural forms, what Voorhees (2018) calls the "ludic Renaissance."

3.1.2. Identity Concealment

Participants reported direct risks and experiences of privacy violations, where other players exposed their identities or social media profiles by discovering names or initials, especially in MMOs. Gender bias emerges as many women hide their gender or know others who do, using neutral nicknames, avoiding specifying their gender, or using voice chat. The women see these as preventive measures, as they often face harassment and bullying when they are identified as women in the game. Most take such measures to avoid what they consider to be toxic situations. There are many examples of this phenomenon:

They're always trying to hit on you, or give you gifts, or just say inappropriate things to you. So, based on the advice of people who were already playing, I decided to go down that route, I don't identify myself as a girl, and because of that, I can play without any hassles. (ME 4)

In games where you can hide your username, I keep it hidden, and if there's an option not to use voice chat, I do that so I don't have to deal with insults and hate....I pass on that. (ME 12)

3.1.3. Gender Roles and Stereotypes in Characters and Avatars

Stereotyping and sexual segregation have been studied extensively, with particular interest in game content (Gestos et al., 2018; Mccullough et al., 2020; Robinson, 2023). Recent studies have analysed the importance of the link between stereotypes in characters and games and their impact on gaming skills (Lynch et al., 2024). Thus, qualitative research such as Shaw's (2014) applies an intersectional approach, highlighting how gender, race, class, and sexual orientation do not operate in isolation as they generate inequalities and exclusions across the video game ecosystem. It is important to draw attention to this, as some research shows a link between the sexualisation of video games and online sexual harassment of women (Burnay et al., 2019).

Interviewees highlight the hypersexualization of female characters in video games, promoting unrealistic images that influence behavior. Male avatars are linked to power and combat, while female avatars reinforce sexualized archetypes like the femme fatale, Lolita, or powerful fighter.

They suggest that the characterisation of the character and/or avatar ultimately influences in-game roles, in line with recent research showing how these roles influence behaviour towards women gamers (Chappetta & Barth, 2020).

3.2. Manifestations of Violence: Hatred and Sexism Towards Women Gamers

Gender, race, and sexual orientation are direct triggers for hate speech during gameplay, with gender being the main reason for insults against women gamers. All interviewees acknowledged the presence of violence, albeit to varying degrees. There is a consistent pattern of violence towards women gamers simply because they are women, particularly in MMOs, where insults and harassment are clearly gender-based. Such attacks conflate their female identity with their role as gamers, with sexist behaviour manifesting itself in two main ways:

1. Hostility, aggression, and sexual harassment (hostile sexism);
2. Overprotectiveness and possessiveness (benevolent sexism).

This is reflected in the extract from this mixed focus group:

They can go one of two ways, they're either more into insulting and putting women down or they try to hit on them [everyone laughs in unison and say "yeah, yeah, exactly," nodding in agreement]. (HGe-1)

[Finishing off what a man has just said] It's either I'll help you and you make it up to me...or you're just a piece of shit. (MGe-5)

The presence of ambivalent sexism (Glick & Fiske, 1996) in relationships between boys and girls in online video games is therefore clearly evident, as demonstrated by the interviewees. This is consistent with the findings of Fox and Tang (2017).

3.2.1. Insults, Harassment, and Hatred: Hostile Sexism

Hostile attitudes, aggression, and sexual harassment towards women are clearly evident during video game interactions. It is noticeable that the insulting behaviour directed at women gamers during gameplay is usually

linked to gender biases resulting from the androcentric power structures of the space. Many women gamers accept open expressions of hatred and hostility as being part of the gaming experience. In most cases, this is expressed in the following ways:

1. By belittling their attributes and playing ability (simply because they are women);
2. By unreasonable demands and accusations of being intruders in the game, often accompanied by insults or taunts rooted in traditional male chauvinism, suggesting that they should stick to domestic tasks. This is reflected in the following examples:

You hear all sorts of things, like, “go back to the kitchen,” “you’re of no use here.” (ME 6)

“Go and do the dishes,” “make me a sandwich; that’ll do you good.” (MGc-7)

There’s always that one guy who starts it off first....I mean, they don’t all jump in at the same time, there’s always one guy who goes: “Look, we’ve got a bitch here, what the hell!...” or suddenly one guy comes right out and says: “Woman, go do the dishes!” and then everyone else joins in and plays along. (ME 11)

The boys participating in the groups pointed out the same devaluations and insults towards the girls:

The usual, they tell them to go back to the kitchen or call them a whore, which always comes up, and of course, they throw sexual insults at them. (HGc-2)

One of the interviewees, who played competitively at a semi-professional level, had her account hacked immediately after winning a tournament when it was discovered that she was a woman:

I’ve played competitively before, and when you get into a professional-level tournament, that’s when they really come after you. That’s where you feel the difference, they even went as far as deleting my account when we won, they found out from the other team that I was a woman so they hacked me, and I can’t access that account anymore. (MGB-2)

Finally, there is a range of sexual harassment and abuse that reduces these players to mere sexual objects:

In everyday terms, when I know, for example, in games like LOL or Overwatch...the second you hear over the chat, because you can hear the voices over the equipment, the second a girl’s voice comes on, all hell breaks loose...either they’re all blaming you or they’re asking for your number to send a photo, or whatever. (ME 8)

It happens all the time that as soon as they hear that there is a woman there, they start asking for a pic or your Insta, they hit on you....It’s gross, they go all out. (MGB-2)

Even harassment...suddenly they find out that you are a woman and they start talking to you every day, asking you for your phone number and chatting to you....It’s like they start to assume that you’re up for a date, that you want something more....It’s like you go from being just...like....I don’t know how to put it...like because you play video games and he’s in that game you must be into him. (ME 1)

Sexual comments and attempts to initiate private relationships with women players are common, making them vulnerable to harassment or abuse. Studies like McLean and Griffiths (2019) report that 54% of women gamers receive requests for sexual favors. Research by Fox and Tang (2017), Fox et al. (2018), and Bustos et al. (2023) indicate that aggression towards women in gaming follows patriarchal patterns: Male gamers treat them as outsiders, belittling their skills and sexualizing them.

3.2.2. Protective Paternalism: Overprotection

In addition to explicit expressions of hatred and violence towards women, there is another set of more subtle sexist attitudes that manifest themselves in overprotective behaviours during gameplay. In such cases, some boys take on the stereotypical role of protector, in the belief that women need to be saved or protected. These players, coming from a position of power and skill, engage with women gamers in a possessive manner:

And what I've experienced the most is like an attempt at positive discrimination—"don't shoot her, poor thing, she's a girl," "oh, no, sorry, this girl just took you out!" (ME 13)

They say silly things...or give [girls] gifts that make no sense. I mean...you always see paternalistic attitudes or rude comments or things that would never be said if it was a boy [involved]. (ME 4)

Male participants also acknowledge these overprotective attitudes towards women. They mention that many guys act like "heroes" and see themselves as "saviours" when playing with them.

We can therefore conclude that sexism in its various forms is deeply rooted in the online gaming environment and can be seen in men's interactions with women gamers, following the existing literature (Tang et al., 2020; Vergel et al., 2024). The growth of video games to wider, more diverse audiences has not removed sexism in gaming. Women still face gender stereotypes and misogyny, affecting their participation and enjoyment (Cote, 2020).

3.3. Contributing Factors

Among the factors that many of the women highlight as contributing to misogyny and sexist behaviour during gameplay, there are certain recurring narratives, like the way that women are regarded as intruders (they are "in a man's world"), or the anonymity afforded by gaming and a sense of impunity in the face of their actions. In many cases, a distinction is made between communities that are labelled as either more toxic or safer. It seems, therefore, that MMOs (where you can play with people from all over the world), are more prone to these kinds of situations. Participants also point out that this tends to be linked to games where there is more competition and stress. In these cases, more violent behaviour is observed, both in general and specifically against women. Examples include games such as LOL, WOW, or Fortnite:

It really depends on the games because some communities are super toxic and others are much friendlier...In their own team, maybe someone messes up and is very nervous and they start badmouthing you. They'll start taking it out on you for whatever reason and instead of playing to win, they play to lose and screw you over. (MGc-1)

It's mainly because of the competitive aspect. The more competitive the game, the more stressful it's going to be. I see it in terms of sport, like in physical sports, for example, you're physically relieving yourself of the stress of wanting to win. But of course, in gaming, people's way of relieving themselves of everything from that desire to win is to insult one other. (HGc-2)

The women gamers themselves argue that they feel like a minority and that men who act violently typically do so in the belief that the women players are intruding into a space that does not belong to them. This reinforces their male dominance and leads to harassment and hostility towards those women who are present, as they are perceived as a threat.

I don't want to interact with...because you feel a little bit...in online games, as most of them are guys, it's like you're alone in a nightclub. (ME 9)

So at the slightest glimpse of a girl they already assume that you're no good, that you're not up to it, it's internalised that this is a man's world. (MGb-4)

A common and widely recognised issue is the role that anonymity plays in these expressions of violence. All participants agree that this is a fundamental aspect and condition that reinforces such behaviour, as is the case in other online environments. In addition, a sense of impunity in relation to the consequences of insulting or harassing others in these environments is identified as the second main factor behind these situations:

The screen makes them think they are protected from everything and that there are no consequences for anything. (ME 4)

In the end, social networks or games are supposed to bring us closer to people, but there's a barrier of anonymity that makes it seem as though there are no consequences for whatever you're doing at any given time. (ME 5)

Furthermore, much of the narrative shows that this problem is linked to a wider system of values. We live in a sexist society and this is reflected in gaming as well as in other areas of life. Not everyone explicitly articulates this discursive line of thought, which clearly underlies the analysis of narrative configurations.

These findings should be interpreted bearing in mind that hostility has been normalised as part of the gaming experience, downplaying the potential consequences for women gamers (Vergel et al., 2024). These factors play an important role in fostering hostile attitudes, consistent with studies suggesting that the intrinsic features of online video games, such as competitiveness and anonymity, encourage negative reactions towards women (Breuer et al., 2015; Ruvalcaba et al., 2018). The online disinhibition effect, so-called toxic disinhibition, facilitates sexism in online interactions (Wachs & Wright, 2018).

3.4. Responses to and Consequences of Violence

Women gamers respond to virtual violence by blocking, muting, reporting, or ignoring aggressors, and often by leaving the game to avoid harassment. The normalization and downplaying of such behaviors are common

coping mechanisms. Online communities play a key role as protectors and buffers, with women emphasizing the need to build their own community' for support and protection against violence:

I usually just stop talking and I've even dropped out of a game, to be honest. But normally I try not to take it so personally, for example, and I don't talk and I just keep quiet and that's that, and you avoid talking or you stay silent. (ME 7)

What I did was ignore them, not give them any attention at all....I act like they're not there...but yeah, at times I've ended up leaving the game because they get really pushy. (ME 1)

Such situations have a number of direct consequences for women gamers, which are sometimes explicitly recognised as psychological and social effects, while at other times downplayed as being "part of the game."

In terms of personal and social effects, women gamers experience increased pressure, as well as feelings of anger, rage, or fear during gameplay. In some cases, they choose to quit the game to avoid further repercussions. However, they do not always recognise this as a negative outcome, even though it directly affects their ability to play on an equal footing and limits their personal freedom. In many cases, there is a tendency to downplay these consequences:

I didn't really care because I'm very laid back....I mean, I used to get so mad before, I even got sad because they ruined the game for me. Look, now, I prefer to leave the game, I'll start doing other stuff and then maybe I'll be back in an hour, but honestly, in my case, it doesn't bother me because I know my worth. These days it doesn't matter what you say to me. (ME 13)

To me it's obviously a negative vibe. I know of women who have tried to get into LOL, for example, and...they've given up gaming and don't want to play online games anymore because they don't want to be judged or insulted or treated badly, people aren't made of stone even if they are in a video game. (ME 3)

I've often stayed out of games because I don't want to have people insulting me. I've got my life, I've got my own shit to deal with without putting up with insults from strangers, it really pisses me off. (ME 9)

Their preferred response, however, is to play and build communities of fellow gamers to avoid such situations and, above all, to confront and address them:

In World of Warcraft...your group is usually your guild. Once you find a guild that you feel comfortable in, then the guild itself ends up protecting you, so to speak. And before then you do the vast majority of things in the game with them. (ME 3)

Of course, I'm telling you that I play with my community, we're all in the WhatsApp group, [and] there, the first guy that goes too far, people go after him and we don't let it go...of course. (ME 13)

I'm lucky, I play with my community, so to say. With friends, they have my back, or I don't need to use voice chat outside my group...and in the end, that's a safe environment for me. (ME 1)

Seeking community alternatives and choosing to stop playing are the most effective responses for women gamers. However, both options limit their ability to explore the games and improve their skills, and affect their self-esteem as gamers.

Research shows that women gamers often face anxiety, stress, loneliness, and negative impacts on their gamer identity (Fox et al., 2018; McLean & Griffiths, 2019). Victims are left to find solutions, limiting their freedom of choice, while support networks in gaming are crucial. Studies like Vergel et al. (2024) confirm that cybersexism causes women to play less or leave gaming, perpetuating inequality, reducing digital participation, and widening the gender digital divide. Persistent gender bias in gaming culture leads to segregated safe spaces that offer social support for women gamers (Zhang et al., 2023). These findings align with the proposals of Cote (2020) and Vossen (2018), documenting how women and other marginalised groups are creating alternative communities and projects that challenge the dominant norms of gaming culture.

3.5. Measures Taken by the Platforms and Actions to Combat the Problem

Women gamers criticize the ineffectiveness of platform measures, which fail to align with current game dynamics. While initiatives like banning certain words in chats and adding mediators in smaller games exist, they are insufficient. Reporting is the primary tool but is seen as ineffective due to delays and strict conditions for action. Bans for sexist behavior are rare, with harsher penalties typically targeting grieving. Additionally, the lack of an instant reporting option for voice chat further limits responses to inappropriate behavior:

It's that they know that nothing happens, at most that they leave you without playing for a few minutes, an hour or so, which doesn't usually happen. If there were other legal consequences, of knowing who you are and so on...but they don't care. If they ban you or take away your game account, well, let's see! In a minute they make another one for free. (ME 8)

3.6. Analysis of the Narrative Configurations

The significance of the results lies in their ability to establish connections through the narratives and experiences shared by players themselves. By constructing a narrative framework, the structural analysis of discourse structural discourse analysis reveals that online video games, particularly MMOs, function as androcentric power spaces, which is evident in their design, gameplay dynamics, and asymmetrical relationships. Women are perceived as intruders in a masculinised environment that becomes toxic and competitive within certain games and communities (competitive, stressful, high-tension multiplayer games).

Female players specifically describe experiences of hatred and derogatory comments rooted in sexism, highlighting the use of crude and often simplistic language when referencing "feminised roles." Figure 1 illustrates an analytical framework comprising two closely interrelated axes: the axis of violence and the axis of manifestations of effects. This framework seeks to connect women's narratives, offering more than a descriptive approach by exploring how women interrelate the primary scenarios they encounter in the context of online gaming.

On the one hand, there is a clear relationship between violence and its effects: Hostile and overprotective attitudes directly impact female players, restricting their ability to participate freely and affecting their

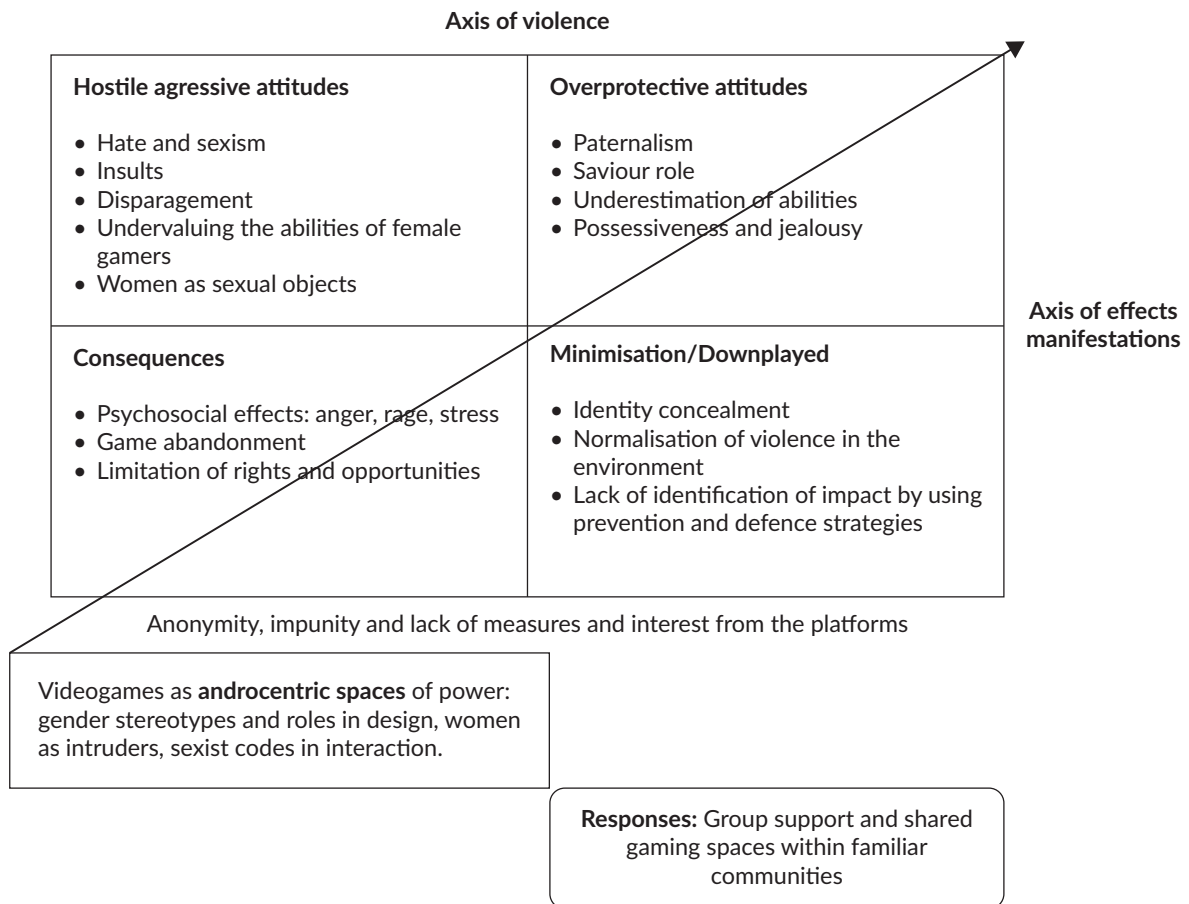


Figure 1. Narrative configurations.

personal well-being. For example, insults and the devaluation of their abilities frequently lead to concealment strategies or even withdrawal from gaming.

On the other hand, there is a feedback loop in which the effects reinforce the perpetuation of violence: individual responses, such as the normalisation and minimisation of these behaviours, bolster existing power dynamics by failing to challenge the androcentric system that dominates video games. Furthermore, the failure of gaming platforms to recognise or address these impacts further perpetuates this reality. Group support and safe communities act as mediators, showing that this violence is not only reflected in specific acts but has a cumulative impact on female players' freedom and participation of the players. This is illustrated in Figure 1.

4. Conclusion

Interviews and focus groups with women gamers reveal key insights into their experiences, shedding light on the manifestations of hate speech and sexism, the dynamics of social interaction, and the consequences they endure. This study highlights the persistence of unequal gender structures in online video games, where female players face dynamics of violence, exclusion, and sexism. These findings underscore how video games, far from being mere spaces of leisure, also function as environments that reflect and reinforce existing social inequalities, as noted by Gray and Leonard (2018) and Jenson and de Castell (2021).

Although the qualitative nature of this study necessitates cautious interpretation, it provides valuable contributions by reconstructing social realities through the subjective experiences of Spanish-speaking gamers, shaped by the role of language in constructing social reality (Van Dijck, 2019). The complexity of sexist interactions in video games is effectively captured through the methodological approach, enabling the identification of relationships between manifestations of violence, their consequences, and women's reactions, all of which are related to gender conditioning factors, as shown in the mapping in narrative configurations. Despite players' attempts to avoid violence through various strategies, such as selectively choosing games, toxicity remains a significant issue, adversely affecting their participation and well-being. Notably, these avoidance strategies, while protective, also limit players' empowerment and freedom within gaming spaces, a critical issue warranting further exploration, a central issue that must continue to be explored in greater depth.

Through narrated experiences, this research confirms that individual responses, such as normalising or downplaying harmful behaviours, contribute to reinforcing the androcentric systems that dominate video games. These behaviours, combined with structural conditions—such as stereotypical game design and inadequate platform responses—perpetuate gender inequality in gaming environments. Additionally, the heightened violence observed specifically in MMO-type games, often normalised by the stressful nature of these games, aligns with studies indicating a higher incidence of violence against intensive female gamers (Calderón-Gómez & Gómez-Miguel, 2023). This hostility, compounded by other cultural factors, influences women's preference for single-player games and their tendency to conceal their identity in online settings (Kuss et al., 2022; Paaßen et al., 2017; Vermeulen et al., 2017).

The women interviewed avoid denigrating situations by carefully selecting the games they play, aligning with research on the exclusion faced by women and marginalised groups, particularly experienced in genres such as shooters and massive multiplayer online role-playing games (MMMORG; see Cote, 2020). An intersectional approach is essential to address the interplay between social structures underpinning video game design and the lived experiences of female players, considering axes of oppression such as race, gender, sexuality, and class (Shaw, 2014). Our findings underscore the responsibility of gaming platforms and developers to implement technological measures and foster cultural shifts to discourage hate speech and promote equality.

Our research indicates that violence against women in video games manifests both as hostile behaviours, such as insults and sexual harassment, and as overprotective actions that reinforce stereotypes of men as “saviours.” Both contribute to gender-based exclusion by portraying women as either intruders or dependents, as demonstrated by Amores (2023), Bustos-Ortega (2024), and Fox and Tang (2017). Furthermore, anonymity and impunity normalise this violence, enabling hate speech to flourish (Fox & Tang, 2017; Vermeulen et al., 2017; Wachs et al., 2022). These harmful dynamics result in significant consequences for women, including negative emotions, restricted participation, and, in some cases, withdrawal from gaming spaces.

Women expressed dissatisfaction with gaming platforms' inadequate measures against hate speech and sexism, criticising companies for prioritising sales over player protection. They called for stronger actions and greater awareness to create safer and more equitable gaming environments. Supportive communities, however, can mitigate the impact of sexism, as observed by Zhang et al. (2023).

The main value of this research lies in amplifying the voices of female gamers within the cultural context of Spain, providing a solid foundation for future educational, political, and social interventions aimed at addressing inequality in video games. By examining the sociocultural framework shaping women's experiences, this study offers critical insights for designing preventive and socio-educational actions targeting young people. As cultural products, video games hold transformative potential; however, this potential can only be realised if their underlying structures are critically examined. In line with studies by Fox and Tang (2017), Gray et al. (2018), and Shaw (2020), this work argues that video games must be reimaged as inclusive spaces that celebrate diversity rather than perpetuate exclusion.

Acknowledgments

Acknowledgment is due to all participants for sharing their experiences and to @feminismoen8bits for their support.

Funding

This article has been produced in the context of the 32 BITS Project: Androcentrism, Hate Speech and Gender Bias Through Online Video Games in Castilla-La Mancha (SBPLY/21/180225/000108), funded by Junta de Comunidades de Castilla-La Mancha and EU FEDER and by the grant UCLM-GRIN-34480 for the research group GIEPS—GRUPO INCLUSIÓN, EQUIDAD Y POLÍTICAS SOCIALES, and co-financed by the EU through the European Regional Development Fund.

Conflict of Interests

The authors declare no conflict of interests.

References

- Afonso Noda, S., & Aguilera Ávila, L. (2021). Desigualdades en el mundo de los videojuegos desde la perspectiva de los jugadores y las jugadoras. *Investigaciones Feministas*, 12(2), 677–689. <https://doi.org/10.5209/infe.60947>
- Aghazadeh, S. A., Burns, A., Chu, J., Feigenblatt, H., Laribee, E., Maynard, L., Meyers, A. L. M., O'Brien, J. L., & Rufus, L. (2018). GamerGate: A case study in online harassment. In J. Golbeck (Ed.), *Online harassment* (pp. 179–207). Springer. https://doi.org/10.1007/978-3-319-78583-7_8
- Amores, M. (2023). *Play like a girl: Desafíos de las mujeres en la industria del videojuego y la tecnología*. Libros Cúpula.
- Anti-Defamation League. (2022). *Hate is no game: Harassment and positive social experiences in online games 2021*. <https://www.adl.org/hateisnogame>
- Arroyo-López, C., & Esteban-Ramiro, B. (2022). Online videogames: Scenarios of interaction and experiences of violence in women gamers. *International Humanities Review/Revista Internacional de Humanidades*, 14(3), 1–16. <https://doi.org/10.37467/revhuman.v11.4116>
- Asociación Española de Videojuegos. (2024). *The industry of the video game in Spain in 2023*. https://www.aevi.org.es/web/wp-content/uploads/2024/05/02_DEF_AEVI_Anuario-2023_com.pdf
- Blackburn, G., & Scharrer, E. (2019). Video game playing and beliefs about masculinity among male and female emerging adults. *Sex Roles*, 80(5/6), 310–324. <https://doi.org/10.1007/s11199-018-0934-4>
- Blaya, C., & Audrin, C. (2019). Toward an understanding of the characteristics of secondary school cyberhate perpetrators. *Frontiers in Education*, 4. <https://doi.org/10.3389/feduc.2019.00046>
- Braun, B., Stopfer, J. M., Müller, K. W., Beutel, M. E., & Egloff, B. (2016). Personality and video gaming:

- Comparing regular gamers, non-gamers, and gaming addicts and differentiating between game genres. *Computers in Human Behavior*, 55, 406–412. <https://doi.org/10.1016/j.chb.2015.09.041>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Breuer, J., Kowert, R., Festl, R., & Quandt, T. (2015). Sexist games = sexist gamers? A longitudinal study on the relationship between video game use and sexist attitudes. *Cyberpsychology, Behavior, and Social Networking*, 18(4), 197–202. <https://doi.org/10.1089/cyber.2014.0492>
- Burnay, J., Bushman, B. J., & Larøi, F. (2019). Effects of sexualized video games on online sexual harassment. *Aggressive Behavior*, 45(2), 214–223. <http://doi.org/10.1002/ab.21811>
- Bustos-Ortega, M. (2024). *Actitudes sexistas hacia las mujeres en el ámbito de los videojuegos: Evaluación y análisis psicosocial de su impacto* [Unpublished Doctoral tesis]. Universidad de Granada. <https://hdl.handle.net/10481/89849>
- Bustos-Ortega, M., Carretero-Dios, H., Megías, J. L., & Romero-Sánchez, M. (2023). Sexist attitudes in online video gaming: Development and validation of the sexism against women gamers scale in Spanish and English. *Psychology of Women Quarterly*, 47(3), 320–342. <https://doi.org/10.1177/0361684323116283>
- Bustos-Ortega, M., Romero-Sánchez, M., Megías, J. L., & Carretero-Dios, H. (2024). Gaming with a feminist: Sexism and perception of sexist incidents in online video games. *Sex Roles*, 90, 1200–1217. <https://doi.org/10.1007/s11199-024-01506-w>
- Calderón-Gómez, D., & Gómez-Miguel, A. (2022). *Consumir, crear, jugar. Panorámica del ocio digital de la juventud. Centro Reina Sofía sobre Adolescencia y Juventud*. Fundación FAD Juventud. <https://doi.org/10.5281/zenodo.6338126>
- Calderón-Gómez, D., & Gómez-Miguel, A. (2023). Desigualdad de género y violencia en los videojuegos: Un análisis desde las experiencias videolúdicas de la juventud. *Methaodos. Revista de ciencias sociales*, 11(1). <http://doi.org/10.17502/mrcs.v11i1.648>
- Chappetta, K. C., & Barth, J. M. (2020). Gaming roles versus gender roles in online gameplay. *Information, Communication & Society*, 25(2), 162–183. <https://doi.org/10.1080/1369118X.2020.1764608>
- Conde, A. F. (2009). *Análisis sociológico del sistema de discursos*. Centro de Investigaciones Sociológicas (CIS). https://www.scielosp.org/article/ssm/content/raw/?resource_ssm_path=/media/assets/resp/v84n1/resena.pdf
- Costa, S., da Silva, B. M., & Tavares, M. (2021). Online hate speech in video games communities: A counter project. In S. Ullmann & M. Tomalin (Eds.), *Counterspeech* (pp. 123–143). Routledge. <http://dx.doi.org/10.4324/9781003377078-9>
- Cote, A. C. (2020). *Gaming sexism: Gender and identity in the era of casual video games*. In A. Cote (Ed.), *Gaming sexism* (pp. 86–147). New York University Press. <https://nyupress.org/9781479802203/gaming-sexism>
- Donoso-Vázquez, T. (2018). Las ciberviolencias de género, nuevas manifestaciones de la violencia machista. In T. Donoso Vázquez & A. Rebollo Catalán (Eds.), *Violencia de género en entornos virtuales* (pp. 15–29). Octaedro.
- de Hesselle, L. C., Rozgonjuk, D., Sindermann, C., Pontes, H. M., & Montag, C. (2021). The associations between Big Five personality traits, gaming motives, and self-reported time spent gaming. *Personality and Individual Differences*, 171, Article 110483. <https://doi.org/10.1016/j.paid.2020.110483>
- Entertainment Software Association. (n.d.). *2023 Essential facts about the U.S. video game industry*. Circana for ESA. <https://www.theesa.com/2023-essential-facts>
- Fox, J., Gilbert, M., & Tang, W. Y. (2018). Player experiences in a massively multiplayer online game: A diary study of performance, motivation, and social interaction. *New Media & Society*, 20(11), 4056–4073. <http://doi.org/10.1177/1461444818767102>

- Fox, J., & Tang, W. Y. (2017). Sexism in video games and the gaming community. In R. Kowert & T. Quandt (Eds.), *New perspectives on the social aspects of digital gaming* (1st ed., pp. 115–135). Routledge.
- Gestos, M., Smith-Merry, J., & Campbell, A. (2018). Representation of women in video games: A systematic review of literature in consideration of adult female wellbeing. *Cyberpsychology, Behavior, and Social Networking*, 21(9), 535–541. <https://doi.org/10.1089/cyber.2017.0376>
- Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70(3), 491–512. <https://doi.org/10.1037/0022-3514.70.3.491>
- Glick, P., & Fiske, S. T. (2011). Ambivalent sexism revisited. *Psychology of Women Quarterly*, 35(3), 530–535. <https://doi.org/10.1177/0361684311414832>
- Gray, K. L., Buyukozturk, B., & Hill, Z. G. (2017). Blurring the boundaries: Using Gamergate to examine real and symbolic violence against women in contemporary gaming culture. *Sociology Compass*, 11(3), Article 12458. <https://doi.org/10.1111/soc4.12458>
- Gray, K. L., & Leonard, D. J. (2018). Introduction: Not just a game. En K. L. Gray & D. J. Leonard (Eds.), *Woke gaming: Digital challenges to oppression and social injustice* (pp. 3–15). University of Washington Press. <https://doi.org/10.6069/9780295744191.c001>
- Gray, K. L., Voorhees, G., & Vossen, E. (2018). Introduction: Reconsidering play with feminist games studies. In K. L. Gray, G. Voorhees, & E. Vossen (Eds.), *Feminism in play* (pp. 1–17). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-90539-6_1
- Guest, G., Namey, E., & Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. *PLoS ONE*, 15(5), Article 0232076. <https://doi.org/10.1371/journal.pone.0232076>
- Isaaman, M.-C., & Tolaine-Sage, S. (2022). *Construyendo un sector del videojuego en igualdad*. Asociación Española de Videojuegos. <https://bit.ly/3ZzOX30>
- Jane, E. A. (2018). Gendered cyberhate as workplace harassment and economic vandalism. *Feminist Media Studies*, 18(4), 575–591. <https://doi.org/10.1080/14680777.2018.1447344>
- Jenkins, H., & Cassell, J. (2008). From Quake Grrls to Desperate Housewives: A decade of gender and computer games. In Y. B. Kafai, C. Heeter, J. Denner, & J. Y. Sun (Eds.), *Beyond Barbie and Mortal Kombat: New perspectives on gender and gaming* (pp. 5–20). MIT Press. <https://doi.org/10.7551/mitpress/7477.003.0004>
- Jenson, J., & de Castell, S. (2021). Patriarchy in play: Video games as gendered media ecologies. *Explorations in Media Ecology*, 20(2), 195–212. https://doi.org/10.1386/eme_00084_1
- Kafai, Y. B., Heeter, C., Denner, J., & Sun, J. Y. (2008). Introduction: A new look at women and gaming. In Y. B. Kafai, C. Heeter, J. Denner, & J. Y. Sun (Eds.), *Beyond Barbie and Mortal Kombat: New perspectives on gender and gaming* (pp. 1–12). MIT Press. <https://doi.org/10.7551/mitpress/9780262113120.003.0001>
- Kaye, L. K., & Pennington, C. R. (2016). Girls can't play: The effects of stereotype threat on females' gaming performance. *Computers in Human Behavior*, 59, 202–209. <https://doi.org/10.1016/j.chb.2016.02.020>
- Kuss, D. J., Kristensen, A. M., Williams, A. J., & Lopez-Fernandez, O. (2022). To be or not to be a female gamer: A qualitative exploration of female gamer identity. *International Journal of Environmental Research and Public Health*, 19(3), Article 1169. <https://doi.org/10.3390/ijerph19031169>
- Kwak, H., & Blackburn, J. (2015). Linguistic analysis of toxic behavior in an online video game. In L. Aiello & D. McFarland (Eds.), *Social informatics* (Vol. 8852, pp. 256–272). Springer. https://doi.org/10.1007/978-3-319-15168-7_26
- Lynch, T., Tompkins, J. E., Gilbert, M., & BurrIDGE, S. (2024). Evidence of ambivalent sexism in female video game character designs. *Mass Communication and Society*, 27(6), 1529–1554. <https://doi.org/10.1080/15205436.2024.2311229>

- Marín-Díaz, V., & Cabero-Almenara, J. (2019). Las redes sociales en educación: Desde la innovación a la investigación educativa. *RIED*, 22(2), 25–33. <https://doi.org/10.5944/ried.22.2.24248>
- Martínez-Verdú, R. (2023). University students, gender violence, the digital divide and video games. *VISUAL Review*, 15(4). <https://doi.org/10.37467/revvisual.v10.4637>
- Mccullough, K. M., Wong, Y. J., & Stevenson, N. J. (2020). Female video game players and the protective effect of feminist identity against internalized misogyny. *Sex Roles*, 82, 266–276. <https://doi.org/10.1007/s11199-019-01055-7>
- McLean, L., & Griffiths, M. D. (2019). Female gamers' experience of online harassment and social support in online gaming: A qualitative study. *International Journal of Mental Health and Addiction*, 17, 970–994. <https://doi.org/10.1007/s11469-018-9962-0>
- Mendick, H., Ottemo, A., Berge, M., & Silfver, E. (2021). Geek entrepreneurs: The social network, Iron Man and the reconfiguration of hegemonic masculinity. *Journal of Gender Studies*, 32(3), 283–295. <https://doi.org/10.1080/09589236.2021.1981836>
- Mihura-López, R., Piñeiro-Otero, T., & Seoane-Nolasco, A. (2023). 'No soy una gamer': Sexismo, misoginia y toxicidad como moduladores de la experiencia de las mujeres videojugadoras. *Revista de Investigaciones Feministas*, 14(2), 215–227. <https://doi.org/10.5209/infe.87614>
- Moldes, R. (2019). E-sports y brecha de género: La percepción de las profesionales españolas. *Arxius de Ciències Socials*, 40, 139–148. <https://dialnet.unirioja.es/servlet/articulo?codigo=7038148>
- Muriel, D. (2018). *Identidad gamer: Videojuegos y construcción de sentido en la sociedad contemporánea*. AnaitGames.
- Paaßen, B., Morgenroth, T., & Stratemeyer, M. (2017). What is a true gamer? The male gamer stereotype and the marginalization of women in video game culture. *Sex Roles*, 76(7/8), 421–435. <https://doi.org/10.1007/S11199-016-0678-Y>
- Raith, L., Bignill, J., Stavropoulos, V., Milllear, P., Allen, A., Stallman, H. M., Mason, J., De Regt, T., Wood, A., & Kannis-Dymand, L. (2021). Massively multiplayer online games and well-being: A systematic literature review. *Frontiers in Psychology*, 12, Article 698799. <https://doi.org/10.3389/fpsyg.2021.698799>
- Reichelmann, A., Hawdon, J., Costello, M., Ryan, J., Blaya, C., Llorent, V., Oksanen, A., Rasanen, P., & Zych, I. (2020). Hate knows no boundaries: Online hate in six nations. *Deviant Behavior*, 42(9), 1100–1111. <https://doi.org/10.1080/01639625.2020.1722337>
- Robinson, J. A. (2023). "I ain't no girl": Exploring gender stereotypes in the video game community. *Western Journal of Communication*, 87(5), 857–878. <https://doi.org/10.1080/10570314.2022.2130004>
- Romo, C., Sell, L., Vera, T., & Delgado, J. (2023). Identidades y exposición a las violencias online. Aproximación a una clasificación temática de los mensajes de odio. *Revista Latina de Comunicación Social*, 81, 538–553. <https://www.doi.org/10.4185/RLCS-2023-1998>
- Ruvalcaba, O., Shulze, J., Kim, A., Berzenski, S., & Otten, M. (2018). Women's experiences in eSports: Gendered differences in peer and spectator feedback during competitive video game play. *Journal of Sport and Social Issues*, 42(4), 295–311. <http://doi.org/10.1177/0193723518773287>
- Sampieri, R. H. (2018). *Metodología de la investigación: Las rutas cuantitativa, cualitativa y mixta*. McGraw Hill México.
- Santana-Rodríguez, N. (2020). *Género, gamers y videojuegos. Una aproximación desde el enfoque de género, al consumo de videojuegos y la situación de las jugadoras en el sector*. Universidad de las Palmas de Gran Canaria; Cátedra Fundación Telefónica. https://catedratelefonica.ulpgc.es/wpcontent/uploads/estudio_genero_gamers_videojuegos.pdf
- Shaw, A. (2012). Do you identify as a gamer? Gender, race, sexuality, and gamer identity. *New Media & Society*, 14(1), 28–44. <https://doi.org/10.1177/1461444811410394>

- Shaw, A. (2014). *Gaming at the edge: Sexuality and gender at the margins of gamer culture*. University of Minnesota Press. <https://doi.org/10.5749/minnesota/9780816693153.001.0001>
- Tang, W. Y., Reer, F., & Quandt, T. (2020). Investigating sexual harassment in online video games: How personality and context factors are related to toxic sexual behaviors against fellow players. *Aggressive Behavior*, 46(1), 127–135. <https://doi.org/10.1002/ab.21873>
- Taylor, T. L. (2003). Multiple pleasures: Women and online gaming. *Convergence*, 9(1), 21–46. <https://doi.org/10.1177/135485650300900103>
- Taylor, T. L. (2008). Becoming a player: Networks, structure, and imagined futures. In Y. B. Kafai, C. Heeter, J. Denner, & J. Y. Sun (Eds.), *Beyond Barbie and Mortal Kombat: New perspectives on gender and gaming* (pp. 51–65). MIT Press. <https://doi.org/10.7551/mitpress/9780262113120.003.0008>
- Van Dijck, J. (2019). *La cultura de la conectividad: una historia crítica de las redes sociales*. Siglo XXI editores.
- Vergel, P., La Parra-Casado, D., & Vives-Cases, C. (2024). Examining cybersexism in online gaming communities: A scoping review. *Trauma, Violence, & Abuse*, 25(2), 1201–1218. <https://doi.org/10.1177/15248380231176059>
- Vermeulen, L., Van Bauwel, S., & Van Looy, J. (2017). Tracing female gamer identity. An empirical study into gender and stereotype threat perceptions. *Computers in Human Behavior*, 71, 90–98. <https://doi.org/10.1016/j.chb.2017.01.054>
- Voorhees, G. (2018). Replaying Janet Murray: Feminist game studies and the ludic renaissance. In K. L. Gray, G. Voorhees, & E. Vossen (Eds.), *Feminism in play* (pp. 295–312). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-90539-6_15
- Vossen, E. (2018). *On the cultural inaccessibility of gaming: Invading, creating, and reclaiming the cultural clubhouse*. University of Waterloo. <https://uwspace.uwaterloo.ca/handle/10012/13649>
- Wachs, S., Wettstein, A., Bilz, L., & Gámez-Guadix, M. (2022). Adolescents' motivations to perpetrate hate speech and links with social norms. *Comunicar*, 30(71), 9–20. <http://doi.org/10.3916/C71-2022-01>
- Wachs, S., & Wright, M. F. (2018). Associations between bystanders and perpetrators of online hate: The moderating role of toxic online disinhibition. *International Journal of Environmental Research and Public Health*, 15(9), Article 2030. <https://doi.org/10.3390/ijerph15092030>
- Zhang, Z., Mu, H., & Huang, S. (2023). Playing to save sisters: How female gaming communities foster social support within different cultural contexts. *Journal of Broadcasting & Electronic Media*, 67(5), 693–713. <https://doi.org/10.1080/08838151.2023.2254432>

About the Authors



Beatriz Esteban-Ramiro holds a PhD in law from the University of Castilla-La Mancha. She is part of the teaching and research staff at the University of Castilla-La Mancha, the Department of Labour Law and Social Work, and is assistant professor PHD in social work at the Faculty of Social Sciences and Information Technologies. She is a member of the GIEPS: Inclusion, Equity and Social Policy research group. Her lines of research focus on online violence, gender, and hate speech, as well as social inequalities.



Patricia Fernández de Castro has been a professor of social work at the Faculty of Social Sciences and Information Technologies in Talavera de la Reina (UCLM) since 2008. She holds a law degree from the Complutense University of Madrid and a Diploma in Social Work from UCLM. She earned her PhD in fundamental rights and public freedoms from the same university. As a researcher involved in various national and European projects, her main research areas focus on citizenship, social and political participation, empowerment, and community social intervention.



Orlanda Díaz-García holds a PhD in fundamental rights and public freedoms, a master's degree in occupational risk prevention, and a degree in social work. She is a lecturer in social work at the Faculty of Social Sciences and Information Technology in Talavera de le Reina (University of Castilla-La Mancha). Her main lines of research are psychological harassment at work, family and childhood and youth, welfare state, and sustainable development. She is a member of the GIEPS: Inclusion, Equity and Social Policy research group.



SOCIAL INCLUSION
ISSN: 2183-2803

Social Inclusion is a peer-reviewed open access journal which provides academics and policymakers with a forum to discuss and promote a more socially inclusive society.

The journal encourages researchers to publish their results on topics concerning social and cultural cohesiveness, marginalized social groups, social stratification, minority-majority interaction, cultural diversity, national identity, and core-periphery relations, while making significant contributions to the understanding and enhancement of social inclusion worldwide.



www.cogitatiopress.com/socialinclusion