Social Inclusion
2025 • Volume 13 • Article 9317
https://doi.org/10.17645/si.9317

cogitatio

ARTICLE

Open Access Journal

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), with 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.

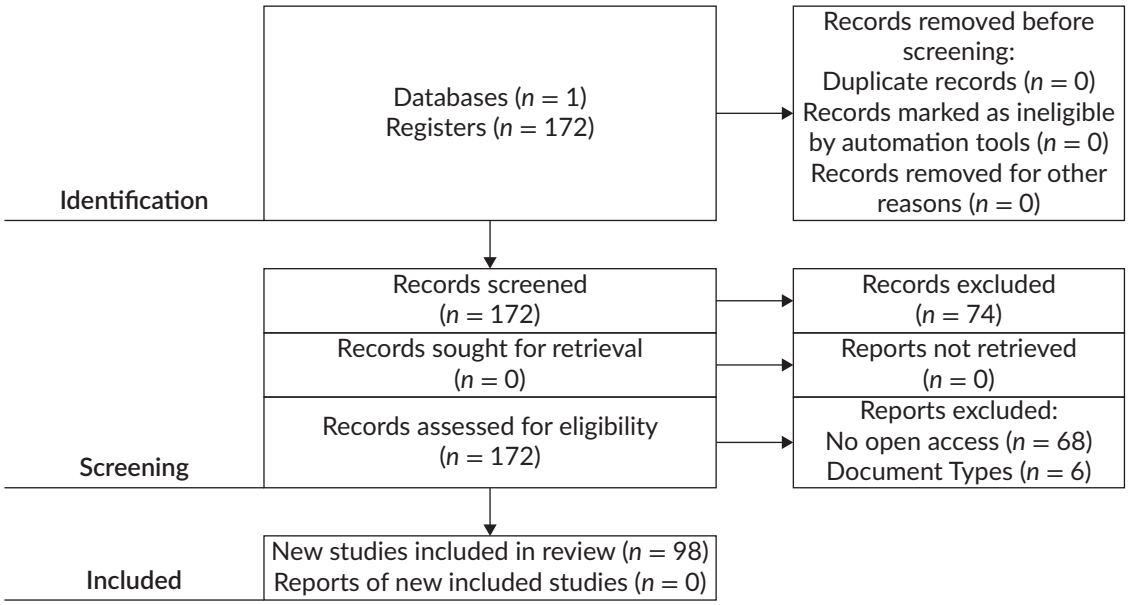

Figure 1. Identification of the new studies via databases and registers: Flow diagram. Source: Based on Haddaway et al. (2022).

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.

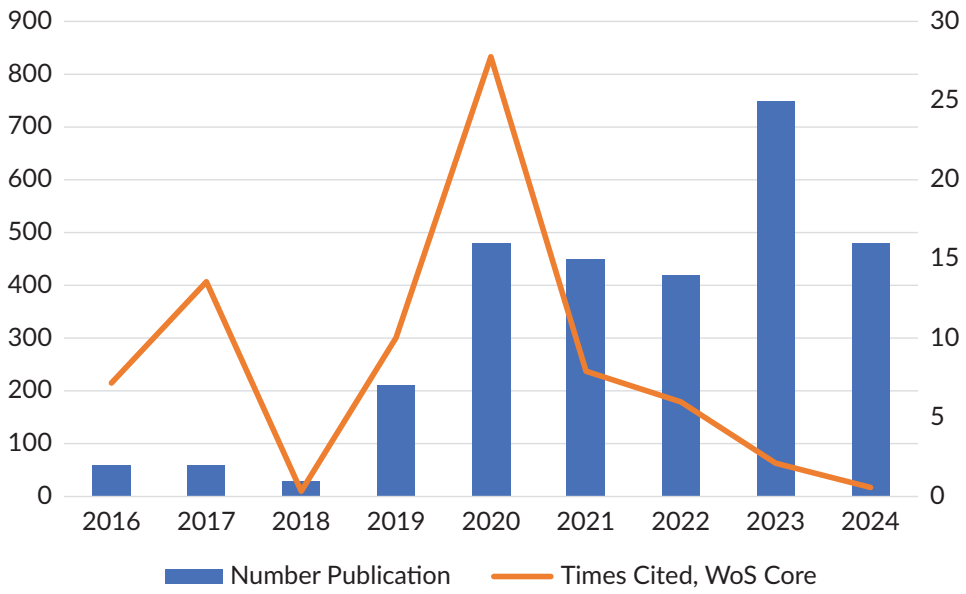

Figure 2. Evolution of the number of articles and citations on hate speech in the WoS database.

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	*Big Data & Society*	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	*Sociology*	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	*Information, Communication & Society*	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	*Epj Data Science*	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	*European Journal Of Communication*	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	*Information Communication & Society*	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	*British Journal of Criminology*	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	*Sensors*	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	*Computer Law & Security Review*	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	*Journal of Information Technology & Politics*	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 Piñeiro-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,

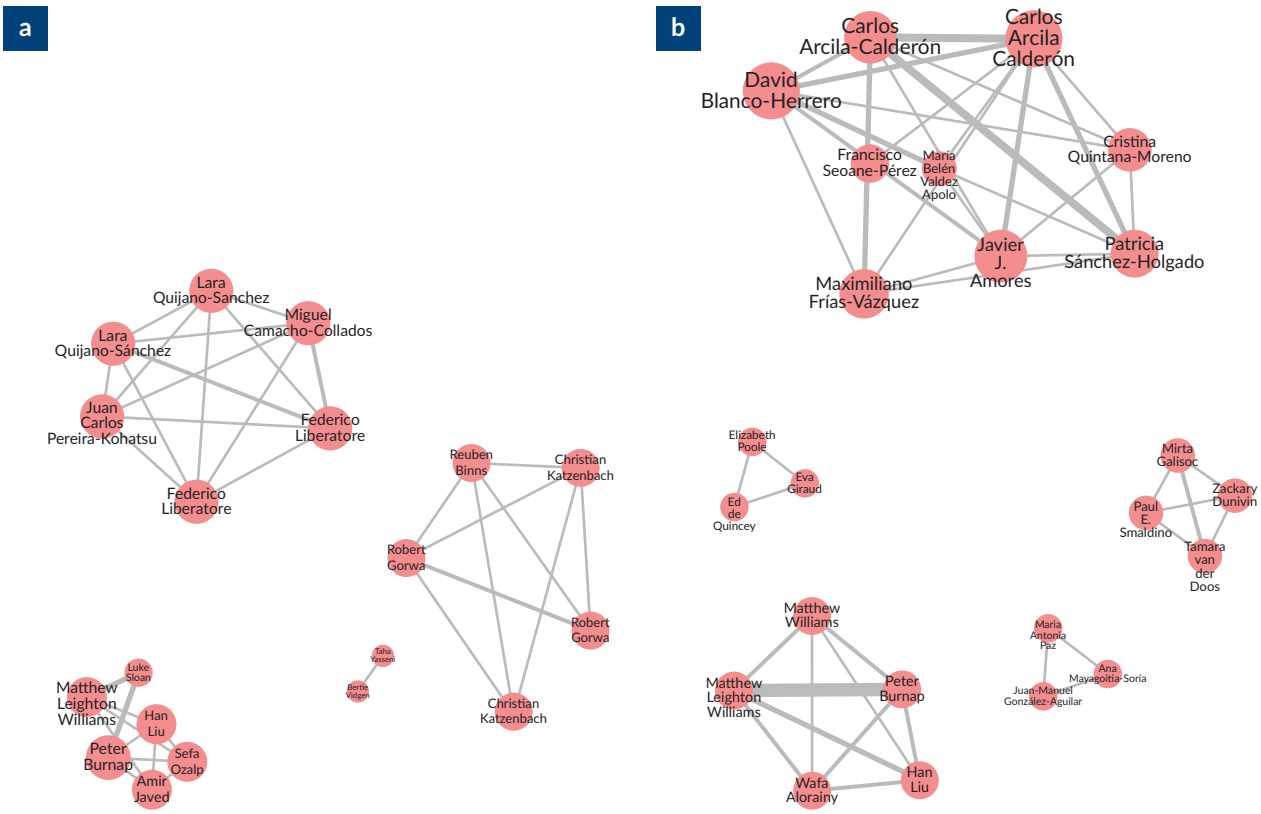

Figure 3. Co-authorship network of the most cited articles on hate speech in WoS: (a) network of nodes of the most cited authors; (b) network of nodes of the most prolific authors.

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, LGTBIphobia, 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., Bahrich, 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 LGTBIQ+: 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/10.1177/20539517211048964

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-Etxeberria, 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 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.