Scientific production of open access articles on hate speech: a scoping review

Alessandra Farné
Departamento de Traducción y Comunicación, Universitat Jaume I (España) ✉ iD
Tamer Al-Naijar Trujillo
Departamento de Ciencias de la Comunicación, Universitat Jaume I (España) ✉ iD
Pablo Parra Valero
Departamento de Biblioteconomía y Documentación, Universidad Complutense de Madrid (España) ✉ iD
Aurora Cuevas Cerveró
Departamento de Biblioteconomía y Documentación, Universidad Complutense de Madrid (España) ✉ iD

ENG Abstract. This work presents a study of scientific publications on hate speech with the objective to map the trends in the field. This study follows a scoping review using the PRISMA-ScR methodology. The search was conducted in four databases (Web of Science, Scopus, Social Science Data Base, and Social Science Research Network) using the terms 'hate speech' and 'communication'. The research analysed a sample of 102 open-access scientific journal articles from the period 2016-2021. The analysis focused on production aspects, topics and relevance of the papers for hate speech studies. The study shows a significant increase in production since 2020. However, a low level of depth and specificity is detected when addressing hate speech. Many articles are loosely or barely related to the topic or do not provide specific definitions. There is a prevalence of studies linked to online cases, though not strictly hate speech but rather general topics of cyberbullying or online harassment and an increase in computer-based methodological approaches. Despite a growing scientific production on the subject, the overall quality of most articles does not offer a significant contribution to the understanding, debate, or prevention of hate speech.

Keywords. Hate speech, scoping review, journal articles, scientific databases, open access, definitions, methodological approaches

ES Producción científica de artículos en acceso abierto sobre discursos de odio: una revisión sistemática exploratoria

ES Resumen. Este trabajo presenta un estudio de publicaciones científicas sobre discurso de odio con el objetivo de mapear las tendencias en el campo. A partir de la aplicación de la metodología PRISMA-ScR, se realizó la búsqueda en cuatro bases de datos (Web of Science, Scopus, Social Science Data Base y Social Science Research Network) con los términos en inglés "hate speech" y "communication". La investigación analiza una muestra de 102 artículos de revistas científicas de acceso abierto del periodo 2016-2021. El análisis se centra en aspectos de producción, temáticas y relevancia de los artículos para los estudios del discurso de odio. El estudio muestra un aumento significativo de la producción desde 2020. Sin embargo, se detecta un bajo nivel de profundidad y especificidad a la hora de abordar el discurso de odio. Muchos artículos están vagamente o apenas relacionados con el tema o no proporcionan definiciones específicas. Por otra parte, predominan los estudios vinculados a casos en línea, aunque no estrictamente discursos de odio sino más bien temas generales de ciberbullying o acoso en línea y un aumento de enfoques metodológicos basados en computadoras. En conclusión, existe una creciente producción científica sobre el tema, pero la calidad general de la mayoría de los artículos no ofrece una contribución significativa a la comprensión, el debate o la prevención del discurso de odio.

Palabras clave. Discursos del odio, revisión sistemática exploratoria, artículos de revista, bases de datos científicas, acceso abierto, definiciones, enfoques metodológicos.

Sumario. 1. Introduction. 2. Literature review. 3. Methods. 4. Results. 5. Discussion. 6. Conclusions. 7. References.

1.Introduction[1]

Academic interest in the analysis and conceptualization of hate speech has grown in the last decade, as evidenced by the analysis of scientific publications on this research topic (Tontodimamma, et al., 2021; Ramírez García, et al., 2022). This has led to the consolidation of an interdisciplinary scientific field concerned with the analysis of issues related to hate speech, its definitions, and possible solutions.

From this starting point, the central objective of this study is to examine open access scientific publications on hate speech from the period 2016-2021 in four databases: Web of Science, Scopus, Social Science Data Base and Social Science Research Network. The focus on open access publications is relevant because accessible articles have a greater impact on the scientific community, and focusing on these allows for a more specific understanding of the literature that is consulted and cited (UNESCO, 2021). The period chosen encompasses a timeframe in which several significant international events have occurred, which may have had an impact on the scientific production regarding hate speech, such as migratory processes, a backlash against feminism, the rise of political or scientific denialism (e.g., in response to events like the Syrian war refugees, the #metoo movement and the anti-feminist reaction, the election of Donald Trump, Brexit and the COVID-19 pandemic, among others). It is during this period that there has been an increase in the quantity of scientific publications on hate speech, which justifies the focus on the dates selected (Izquierdo Montero; Laforgue Bullido; Abril Hervás, 2022; Adiprasetio; Rahmawan; Wibowo, 2021).

Therefore, this research aims to analyse open access scientific production regarding hate speech based on the following questions:

RQ1. What is the overall distribution of articles by year, author countries, journal countries, scientific fields, and databases?

RQ2. What are the trends in definitions, topics, and analytical approaches regarding hate speech?

RQ3. Is there a significant change following the pandemic with articles related to COVID-19?

RQ4. What is the prevalence of articles that make a significant contribution to the field of hate speech studies?

RQ5. What is the prevalence of relevant publications on hate speech towards migrants or related to gender issues?

2. Literature review

Concern about hate speech was consolidated after the horrors generated by the Second World War (Paz; Montero Díaz; Moreno Delgado, 2020) through two interconnected trends. On one hand, the legal and institutional trend focuses on the implementation of various legal norms with the aim of demarcating which types of speech can be penalised in relation to hate crimes, and on the ongoing debate about the limits of freedom of expression (Howard, 2019). On the other hand, the academic perspective has provided a broader view that goes beyond purely legal or juridical aspects, focusing on the characterization of hate speech, the comparison of different regulations, and cultural-discursive analyses.

From academia, prior to the popularisation of the Internet and social media, warnings were issued about the rapid expansion of hate speech targeting individuals and groups based on their characteristics. Matsuda (1989) states, "The hate speech flaring up in our midst includes insulting nouns for racial groups, degrading caricatures, threats of violence, and literature portraying Jews and people of color as animal-like and requiring extermination" (2333). Brown (2017) points out that while the legal concept of hate speech refers to speech that is illegal or unlawful under the law, there is also an ordinary or general concept of hate speech influenced by both the law and public discourse in the social context, making it a broad and expansive concept. Hate speech is not simply any discourse that offends someone or attacks or derogates someone, nor is it something that some people hate. In this regard, Ibarra's (2017) conception can be useful for its simplicity and accuracy in characterising hate speech as messages that promote actions against the dignity and equal rights of individuals, minority or majority groups, and vulnerable populations through different channels and formats. Díaz Soto (2015) emphasises that hate speech presents a moral dilemma to society, which tends to be resolved according to the cultural and judicial particularities of each territory, often prioritising pragmatism over reflection.

Brown and Sinclair (2020: 16) state that hate speech is characterised by the following: 1) the use of language that is dehumanising, threatening, abusive, or derogatory, including images; 2) targeting individuals or groups based on their ethnicity, sexual orientation, gender identity, citizenship, disability, or other protected characteristics; 3) having the intention to humiliate, intimidate, or promote discrimination or violence against these individuals or groups; and 4) resulting in psychological or dignitary harm, undermining social cohesion, or exacerbating inequality. Therefore, hate speech refers to derogatory speech directed at vulnerable people based on their characteristics, with the aim of causing harm and inciting hatred.

In fact, hate speech can be directed towards various marginalised groups or minorities, such as women (Nadim; Fladmoe, 2021), Muslims (Obermaier; Schmuck; Saleem 2021), migrants (Lilleker; Pérez Escolar, 2023), or the LGBTQ+ community (de Casas Moreno; Parejo Cuéllar; Vizcaíno Verdú, 2023), among others. However, hate speech can be understood from an intersectional perspective. This means that it is not only focused on dehumanising or degrading based on a specific or isolated aspect, but rather as a combination of different elements of oppression, such as misogyny and racism (Saresma; Karkulehto; Varis, 2020: 238).

In the legislative-institutional field, the most significant contributions at the international level to advancing the fight against hate speech come from the Council of Europe (1997) and the European Commission against Racism and Intolerance (2016). It is precisely in their definitions that it is emphasised that hate speech takes the form of racist, xenophobic, or stereotypical and stigmatising narratives, insults, or threats:

> Hate speech for the purpose of the Recommendation entails the use of one or more particular forms of expression – namely, the advocacy, promotion or incitement of the denigration, hatred or vilification of a person or group of persons, as well any harassment, insult, negative stereotyping, stigmatization or threat of such person or persons and any justification of all these forms of expression – that is based on a non-exhaustive list of personal characteristics or status that includes "race", colour, language, religion or belief, nationality or national or ethnic origin, as well as descent, age, disability, sex, gender, gender identity and sexual orientation (European Commission against Racism and Intolerance, 2016: 16).

Therefore, from this institutional perspective, there has been an ongoing emphasis in recent years on governments implementing a legislative framework to combat this type of speech (Council of Europe, 2010).

From a global perspective, the United Nations (UN) warns that there is no internationally agreed-upon legal definition, as it is an ongoing discussion subject to change and controversy. The consideration of hate speech as a crime and its levels of appreciation differ among countries, so an excessively precise definition may not encompass all possible legal violations (UN, 2019). The UN emphasises that hate speech can manifest in various forms, including images, memes, symbols, or gestures, both online and offline. These types of narratives are discriminatory and derogatory towards different cultural and social factors, with group and individual effects.

Currently, hate speech is a complex issue that extends across multiple domains, encompassing individual, group and societal dimensions. It is not only a legal problem but also a technological, political and international one (Brown; Sinclair, 2020: 19-28). In fact, both academia and institutions express concerns about the expansion of this type of speech in the digital environment. Several studies highlight that the advent of the Internet and social media platforms has exacerbated the problem due to their amplification capabilities (Arcila Calderón, et al., 2022). There has been an "exponential increase in scientific interest in the relationship between hate speech and social media" (Ramírez García, et al., 2022: 137) as well as the development of automated detection tools for hate speech, cyberbullying, and misogynistic hate speech on platforms such as Twitter and YouTube (Tontodimamma, et al., 2021).

With this regard, information and media literacy plays an essential role for its potential in strengthening educational strategies to prevent the dissemination of disinformation and hatred online (Essalhi Rakrak; Pinedo González, 2023). This field even includes the development of counter-narratives to counteract hate speech in social media, promoting discursive frames based on values of human rights and intersectional perspectives instead of limiting to denouncing the perils of hate speech (Poole; Giraud; de Quincey, 2021). Similarly, there is an emerging area focused on developing models of automatic generation of counter-narratives, given the social situation of alt-right movements reaching institutional power with discursive strategies infused with hatred (Vallecillo Rodríguez; Montejo Ráez; Martín-Valdivia, 2023).

3. **Methods**

To address the study of scientific publications on hate speech, the scoping review strategy or exploratory systematic review approach has been followed. This method differs from other systematic literature reviews based on the research question and objectives, which leads to distinct procedures in the selection, evaluation, analysis and synthesis of documents, focusing on the characteristics of a specific knowledge domain (Castells Fos, et al., 2023).

As a validation protocol, the checklist proposed in the 2018 PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension) statement has been used as a reference (Tricco, et al., 2018). PRISMA-ScR is an extension of the general PRISMA framework designed to help authors improve the reporting of systematic reviews (Moher, et al., 2009). Although this tool was initially designed primarily for systematic reviews of studies that evaluate the effects of health interventions, regardless of the design of the included studies, the checklist items are applicable to publications of systematic reviews that evaluate other areas not related to health, such as social or educational issues (Page, et al., 2021), as in the case of the present study. The PRISMA-ScR checklist consists of twenty-two items (two of which are optional) and was developed by a group of twenty-six experts following the guidelines of the EQUATOR Network (Enhancing the QUAlity and Transparency Of health Research), which includes specialists in methodology, communication, and publication. The PRISMA guidelines for publishing specific to scoping reviews state that this methodological framework follows a systematic approach to map the evidence on the topic at hand and identify key concepts, theories, and sources of scientific knowledge (Tricco, et al., 2018).

In this study, we refer to the twenty essential elements of the PRISMA-ScR checklist, based on the contributions by Tricco et al. (2018). These elements range from the first one, which is the inclusion of the term in the title, to the last one, which pertains to research funding. The two optional items that were excluded (Critical appraisal of individual sources of evidence, and Critical appraisal within sources of evidence) are more applicable to systematic reviews focused on interventions or treatments in the field of health, where this research method has gained significant recognition in recent years (Codina; Lopezosa; Freixa Font, 2022).

For the development of the scoping review, the following stages were conducted based on the PRISMA-ScR protocol:

1. identification of the research question and objective, with the establishment of inclusion and exclusion criteria for the evidence,
2. formulation of the evidence search plan in the selected databases,
3. extraction and evaluation of the evidence according to the established criteria,
4. interpretation and presentation of results.

For article selection, the following four databases were considered: Web of Science (Clarivate Analytics), Scopus (Elsevier), Social Science Data Base (ProQuest), and Social Science Research Network (SSRN), an open-access preprint repository for academic research by Elsevier.

The search was conducted on June 11, 2022, and the final parameters used for all databases were to include the following formula in the title, abstract, and/or keywords: communication AND 'hate speech'. When searching only for 'hate speech', the results obtained were close to 10,000 documents and included aspects that were not related to the research objective. Therefore, the decision was made to analyse the discursive aspect of the phenomenon by adding the term 'communication'. Additionally, the following inclusion criteria were determined:

- scientific articles published in the selected databases between January 1, 2016, and December 31, 2021,
- full text available in Open Access,
- of the 307 papers identified in the search strategy, a total of 102 were fully analysed and included in the study. As shown in the figure below, the different phases of the scoping review proposed by the PRISMA-ScR guidelines were carried out.
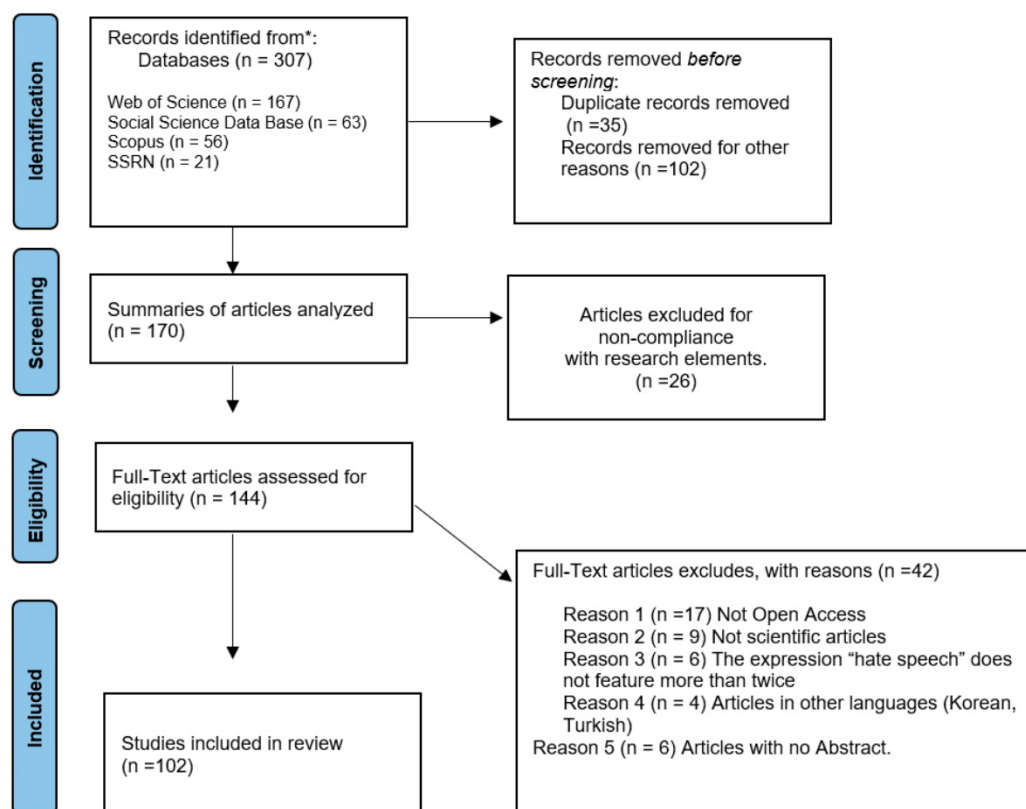
(figure.1) Flowchart of the PRISMA study. Source: own elaboration.

It is worth noting that in the Screening phase, approximately one-third of the documents were excluded, mainly due to not meeting the criterion of full-text access in Open Access. Numerous errors were observed in the results returned by the four databases and in how they categorise the works.

To enhance the consistency of the work, guidelines for data extraction were established before starting the article review, and any doubts or disagreements were resolved through consensus and discussion, although the protocol was not registered. Data entry was done using a form where each of the four reviewers recorded the data independently. The dataset generated and analysed in this research is available at this link: https://osf.io/7qnkf.

All data used for this research were publicly available aggregated data. This article does not involve any studies with human participants; therefore, ethical approval is not applicable.

4. Results

4.1. Temporality, locations and other production details

Based on the analysis conducted on the 102 articles, several results can be extracted regarding the general characteristics of scientific production on hate speech, addressing RQ1. Firstly, there is a clear temporal evolution in the published articles (Figure 2). The first two years of the analysed period (2016-2017) accounted for 10 articles (9.8% of the total), while the last two years (2020-2021) show a significant increase in academic interest in this field, with 68 articles (66.7% of the total).

Regarding the languages of scientific production, English is the predominant language (78.4%), followed by Spanish (7.8%), Polish (5.9%), while other languages account for 7.8% of the total. The production in other languages includes Portuguese, French, Russian, and Serbian, with varying proportions.

(figure.2) Year of publication of the articles. Source: own elaboration

Regarding production by countries, both the country of the journal and author affiliations have been analysed. The journals with the most published articles are from the United Kingdom (18) and the United States (15), which together account for one-third of the total production (32.4%), followed by Spain (8), Poland (7), Switzerland (6), Brazil (5), South Africa and Germany (4 each), Romania, Malaysia and France (3 each), others with 2 each (Hungary, Indonesia, Netherlands, Russia, Slovakia, Sweden) and one each (Bulgaria, Canada, Chile, China, Cyprus, Czech Republic, Estonia, New Zealand, Portugal, Singapore, Serbia, Slovenia, Thailand, Turkey). This geographical distribution of journals differs from the countries of the authors of the articles. Figure 3 shows the countries of affiliation, with Spain, Germany, and Indonesia being the top three. Additionally, there are a whole series of countries that contribute to global scientific production, despite not having many national journals that publish these publications, reflecting the effort of science internationalisation carried out by many countries, especially non-Anglo-Saxon ones.

With regard to journal variety, the sample of 102 articles are published in 87 journals. Sixty-eight per cent of the journals are included in Sherpa Romeo (Rights Metadata for Open Archiving), an online tool that aggregates and analyses publisher open access policies. In this resource, experts analyse copyright and open access archiving policies of each journal and, within the journals of the sample that are included in Sherpa Romeo, in the field OA Fee it can be observed that in more than a third (34%) authors have paid an additional cost to publish their article open access and archive the published version in a repository.
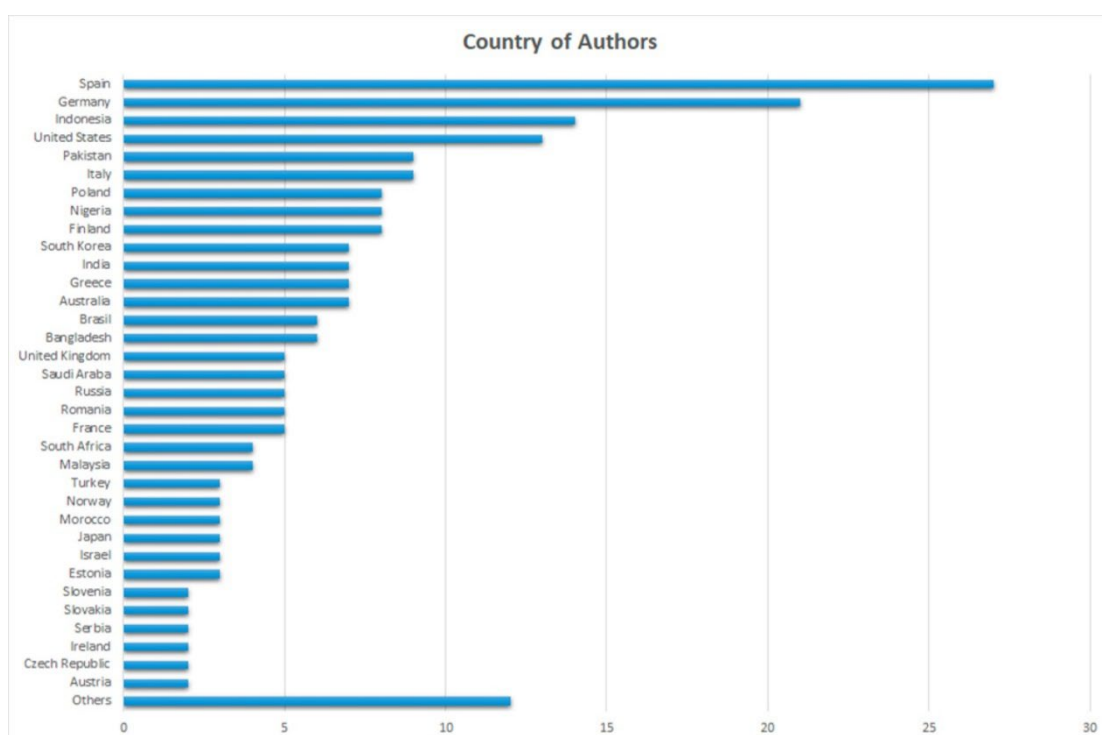
(figure.3) Countries of affiliation of the authors. Source: own elaboration

In terms of the field of the journal where the articles are published (Figure 4), social sciences publications predominate, specifically Communication, Sociology, Political Science, Social Science (general/miscellaneous) and Education, together accounting for almost half of the articles (48%). Additionally, there is also some heterogeneity due to the presence of other different fields, particularly Law (16.7%), Computer Science (12.7%) and Humanities, including Linguistics and Language, Philosophy and Humanities (11.8%).
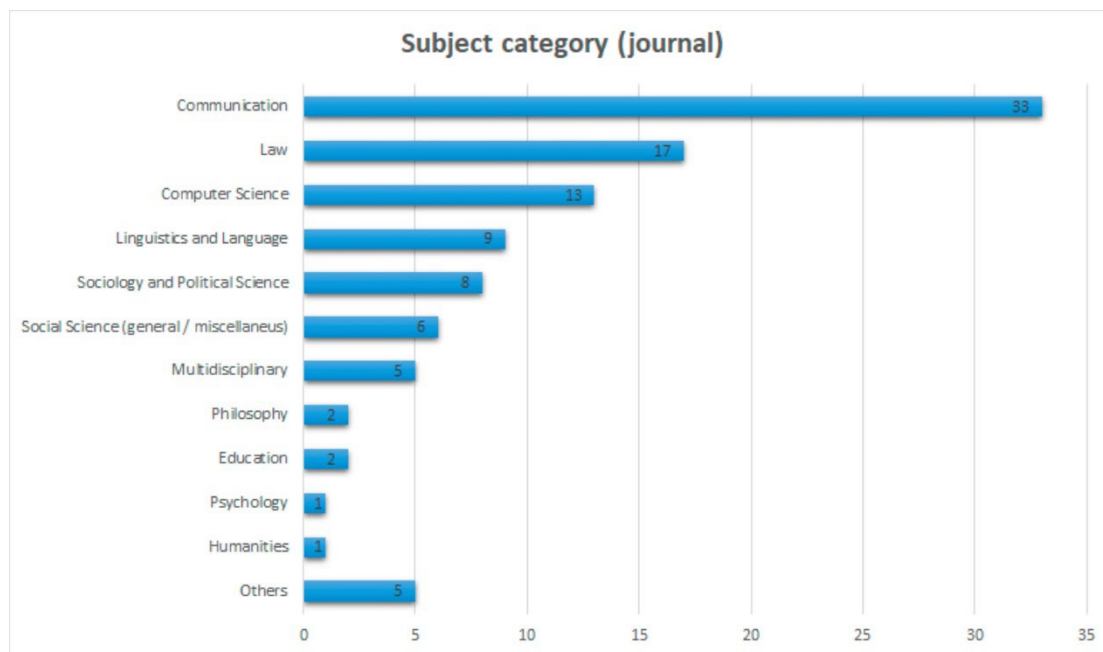
(figure.4) Thematic areas of the journals. Source: own elaboration

Regarding the journals, it is worth mentioning that there is a great variety, and very few journals publish more than one article, only a total of 13. In this regard, the journal *IEEE Access*, from the United States and in the field of Computer Science, stands out, as it features 4 articles on hate speech. There are 12 journals with 2 articles, and the rest have only one.

Finally, it is interesting to analyse the distribution of articles across different databases where they are available: only present in Web of Science (33), only in Scopus (18), only in SSDB (13), only in SSRN (6), and articles that are present in two or more databases (32), with the most common combination being Web of Science and Scopus together (22). Therefore, Web of Science and Scopus stand out both in terms of the number of articles and the heterogeneity of the journal fields, while SSDB and SSRN have a much smaller production.

4.2. **Approaches and Themes**

To address RQ2, regarding the content of the articles, the analysis has focused on the themes of hate speech addressed, the definitions and the type of analysis primarily used.

In terms of the themes, Figure 5 presents the overall distribution, which shows a low level of specificity. It is worth noting that 20 articles (19.6%) do not even address the issue of hate speech, despite having this term in the title, abstract, or keywords. Of those that do address hate speech, 23 take a general approach (theoretical or legal). For articles that address specific types of hate speech, the motives have been grouped based on the categories provided by the European Commission against Racism and Intolerance (2016) and the Attorney General of Spain (2019), who issued a clarification circular in this regard. The most frequently addressed themes, in order, are: political and ideology, origin and provenance, religion and faith, racism, gender issues (including sexual orientation and identity) as well as sexism and misogyny. Other themes include cyberbullying and online harassment, among others.

In the case of articles that address political and ideology, it is noteworthy that they focus on cases of political tension and polarisation, particularly online, rather than hate speech itself. Meanwhile, those related to religion and faith are linked to various issues such as minorities, Islamophobia, or anti-Semitism.
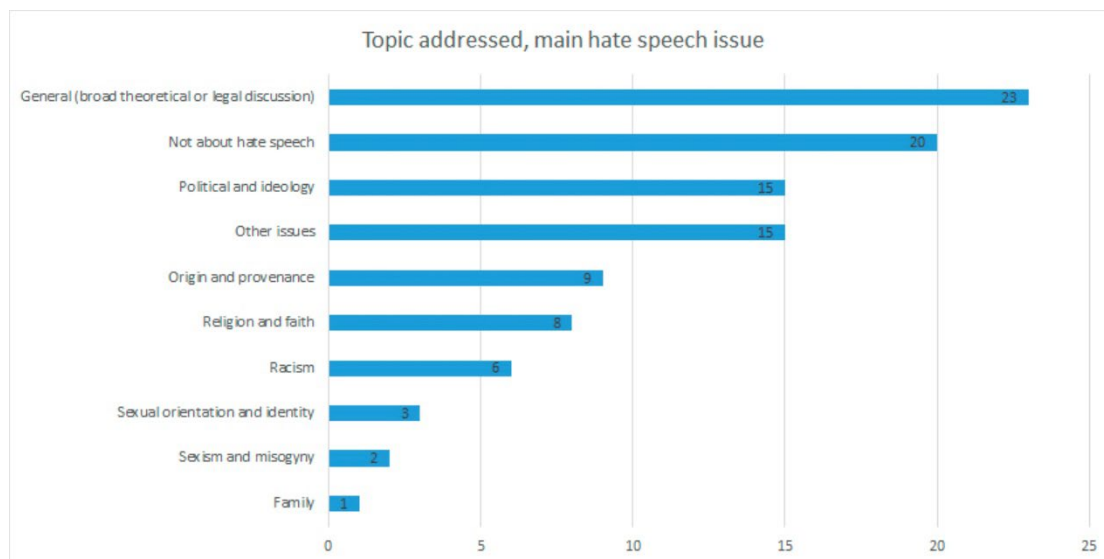
(figure.5) Primary type of hate speech addressed. Source: own elaboration

Regarding the definitions of hate speech addressed in the articles (Figure 6), the results once again show a low level of specificity, as 50% of the articles either lack a clear definition or do not specify what they understand as hate speech. Of the other half, where definitions were found, both the main definition and a possible second or third definition, if available, were taken into account. Overall, if there are definitions, the majority come from the legal and institutional sphere (43 in total), followed by the academic sphere. Notable definitions from the legal sphere come from the rules and regulations of various European institutions, particularly the European Commission against Racism and Intolerance's definition (2016). In the academic sphere, there are a variety of definitions and authors, without a clear common or predominant perspective. Some of the cited authors are George (2015), Fortuna and Nunez (2018) and Erjavec and Kovačič (2012), who define hate speech as follows:

> There is no standard definition of hate speech, but the term is normally used to cover forms of expression aimed at persecuting people by vilifying their racial, ethnic, or other identities. While the immediate target may be a single person or small group, the harm caused by hate speech can extend to entire communities by promoting discrimination and intolerance (George, 2015: 1).
> Hate speech is language that attacks or diminishes, that incites violence or hate against groups, based on specific characteristics such as physical appearance, religion, descent, national or ethnic origin, sexual orientation, gender identity or other, and it can occur with different linguistic styles, even in subtle forms or when humor is used (Fortuna; Nunez, 2018: 5).
> Expression that is abusive, insulting, intimidating, harassing, and/or incites to violence, hatred, or discrimination. It is directed against people on the basis of their race, ethnic origin, religion, gender, age, physical condition, disability, sexual orientation, political conviction, and so forth (Erjavec; Kovačič, 2012: 900).

Among the articles that use other definitions, a reference was found to The Guardian newspaper or Facebook.
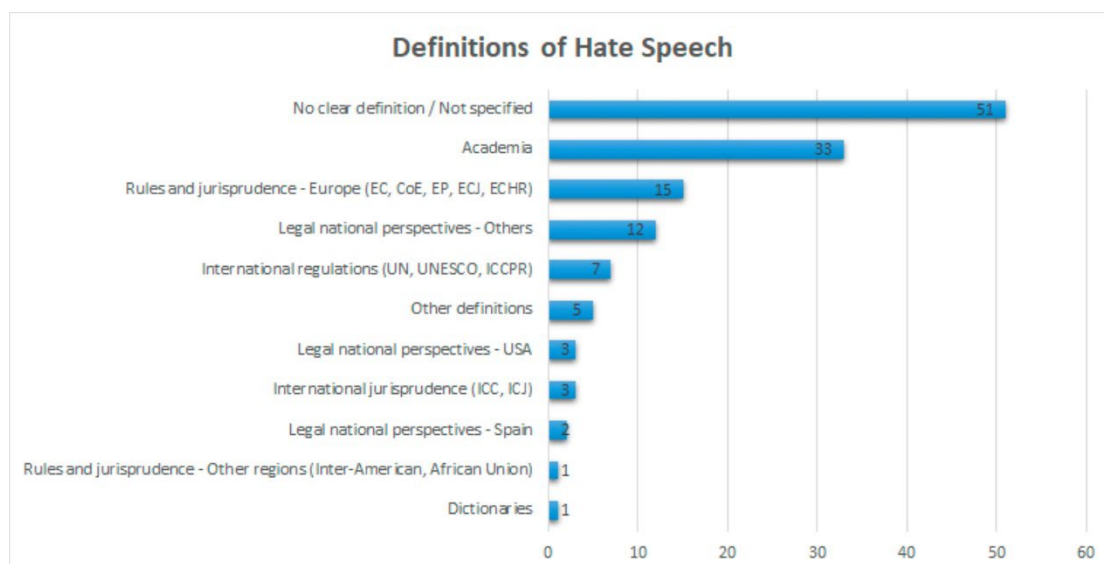
(figure.6) Definitions of hate speech Source: own elaboration

Regarding the type of analysis and methodological approaches (Figure 7), the results reveal a high level of heterogeneity. Quantitative research predominates, with more traditional techniques such as bibliometric analysis or content analysis, as well as more recent computer-based analysis, accounting for 34.3% of the total. Theoretical approaches (reviews or general discussions), mixed methods and, to a lesser extent, qualitative analysis follow. Articles with legal analysis (9.8%) focus on studying laws or court rulings.
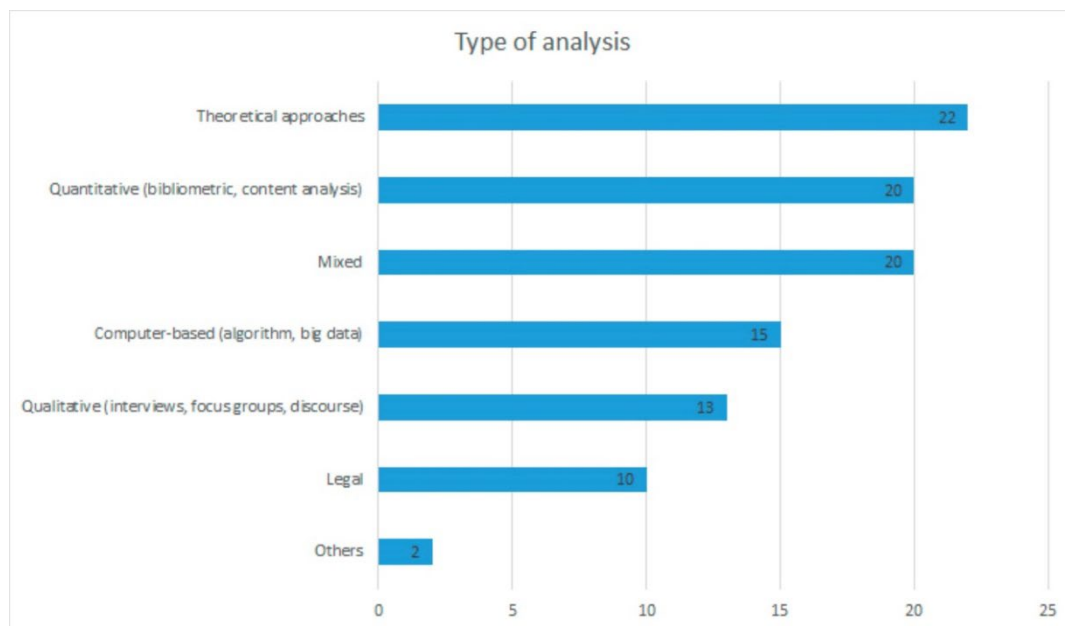
(figure.7) Type of analysis. Source: own elaboration

In addition to the type of analysis and methodology employed, it is worth noting that over half the articles focus on studying cases of online communication (53.9%), with some articles specifically addressing hate speech, such as in cases involving migrants. However, there are also many of these cases that do not actually address hate speech but rather delve into broader issues of online polarisation, cyberbullying, and so on. These results on the themes also align with the analysis of keywords, where, after 'hate speech' (44), the ten most common terms are: 'social media' (18), 'Twitter' (9), 'political communication' (8), 'freedom of expression' (7), 'internet' (7), 'fake news' (6), 'machine learning' (6), 'communication' (6), 'democracy (5) and 'journalism' (5).

4.3. Themes relating to Covid-19
To address RQ3, the articles and their possible relation to COVID-19 have been analysed. It was found that only 7 articles (6.9%) mention any issues related to COVID-19, indicating that there is no significant increase in the potential connection between hate speech and a pandemic context. Furthermore, it should be noted that these few articles are not highly relevant to the topic of hate speech, as they are minimally or barely focused on this subject.

4.4. Depth and relevance
To address RQ4 and RQ5, the overall level of depth regarding hate speech in the articles and their relevance to gender or migration issues has been analysed. Four groups of publications have been identified:

• Barely related (23%): these are articles that are minimally or barely related to hate speech. These are articles in which, in many cases, the term 'hate speech' is only found in the abstract or among the keywords but is not developed in the text. In other instances, the word appears one or two times in the text as a mere generic reference but is not addressed in the article.
• Low relevance (61%): these are articles that are not focused on hate speech and do not offer a significant contribution to this field of study. For example, the term 'hate speech' may appear more frequently in the text, but it is not the central topic of the publications. Instead, they address related subjects such as cyberbullying, discussions on freedom of expression, or provide very vague approaches.
• Medium 1 (8%): these are articles that make a focused contribution to hate speech, where the topic is defined or there is a contribution to the field. Among these articles, meta-analyses of publications, theoretical reviews, or articles that include specific sections to develop the topic of hate speech can be found. They provide academic or legal definitions to address the subject.
• Medium 2 (6%): these are articles that address hate speech in relation to gender or migration issues. All of these articles primarily focus on hate speech related to origin and provenance. In some cases, elements of racism or sexual orientation are also included, but none of them primarily address hate based on gender issues.
• High relevance (2%): these are articles that meet the criteria of both Medium 1 and Medium 2, making them publications with a highly significant contribution in the field of hate speech related to gender or migration. Only 2 articles have been registered as making a significant contribution to the field of hate speech, with a certain intersectional perspective. One focuses on hate speech related to sexual orientation and identity, as well as origin and provenance (Vehovar; Jontes, 2021), while the other links racial hatred with origin and provenance (Rollnert Liern, 2020).

5. Discussion

The results show a clear temporal evolution in scientific activity regarding hate speech. The increase in articles published during the period 2020-2021 (66.7% of the total) compared to the first two years of the analysed period demonstrates a growing academic interest in this field. This finding aligns with the trends reflected in previous literature on the scientific evolution of this topic (Tontodimamma, et al., 2021; Ramírez-García, et al., 2022).

However, the production demonstrates a low level of specificity. It is surprising to observe that 19.6% of the articles do not even address the issue of hate speech, despite having this term in the title, abstract, or keywords; 84% of the publications are minimally or barely related to hate speech. It is also notable that half of the articles either lack definitions or are unclear about hate speech. This issue was already highlighted in Adiprasetio, Rahmawan and Wibowo (2021), where an epistemological problem was emphasised in the field of study due to the absence of clear definitions in academic discussions. Among the articles that do address hate speech, 23 provide a general approach. Regarding articles that address some form of hate speech, it is worth emphasising that migration issues are addressed more than gender issues. Articles related to migration focus on the proliferation of hateful digital content, ranging from studies on anti-immigration or xenophobic discourse (Krobová; Zàpotocký, 2021; Alberdi Urquizu, 2019; Carlson; Kennedy, 2021; da Silva, 2020) to strategies to identify and monitor these types of discourse in Southern Europe (Vrysis, et al., 2021).

As for articles on gender issues, surprisingly there is none that directly addresses hatred towards women, despite the growing scientific production on hate speech in general and the increasing concern at international level for misogyny, as reflected by the United Nations in the Plan of Action for Women in Communities to Counter Hate Speech and Prevent Incitement to Violence that Could Lead to Genocide and related Atrocity Crimes (The Napoli Women in Communities Plan of Action) launched in 2023.

Only twenty-four per cent of the articles analysed present a theoretical treatment primarily focused on legal and communication aspects, mainly addressing the consideration of certain behaviour as crimes in different countries (according to specific and national regulations) and international regulation analyses. The limits of freedom of expression in relation to hate speech are also discussed from a constitutional and criminal code perspective. It is important to note that all of these articles primarily focus on Western cases, especially European and North American contexts.

In terms of the methodological approach, a significant imbalance is observed. Quantitative research (34.3%) with bibliometric analysis and computational analysis techniques prevails over theoretical and qualitative treatments. Furthermore, more than half of the articles tackle cases of online communication (53.9%), confirming a trend found in other studies (Ramírez García, et al., 2022). For instance, there are case studies related to political communication (Losada Díaz, 2021; Bajari, 2021; Oparaugo, 2021; Nick, 2018), LGTB topics (Silva and Silva, 2021) or issues regarding social media (Gallardo; Selva, 2021; Lev-On, 2018; Park, 2021), among others. These results on the academic interest in the online environment align with the social and institutional concern regarding the presence of hate speech in the digital sphere. For instance, in 2016, the European Commission reached an agreement with various social media platforms to establish an action plan for the regulation of hate speech, and several digital platforms are starting to include different recommendations or rules regarding this type of narrative in their protocols (YouTube, n.d.). However, these measures have not always been implemented or proven effective, leading other organisations to promote a preventive approach from the very conception of digital platforms, known as safety by design (Center for Countering Digital Hate, 2023).

In line with the preventive perspective, it is surprising to observe that only two articles directly address the topic of education (Saputra; Al Siddiq, 2020; Jääskeläinen, 2020). Although there are four additional articles that touch upon educational issues, they do so focusing on aspects related to the field of Communication. In fact, the two articles that tackle the concept of 'information disorder' do not delve into hate speech (Hansson, et al. 2021; Aznar, 2019).

6. Conclusions

The scientific production of open access articles on hate speech has been studied for the period 2016-2021 using the PRISMA-ScR methodology. Almost one-third of the originally detected documents were discarded, primarily due to not meeting the Open Access requirements. Additionally, numerous errors have been observed in the results returned by the databases examined and how they categorise the works, ultimately reducing the sample to 102 articles.

In the analysed publications, a quantitative evolution has been detected in the number of published articles, which demonstrates significant progress in academic interest in this field. This evolution indicates the increasing interest of researchers in the study of hate speech and the recognition of its importance by the scientific community, aligning with the social concern regarding the prevalence of hate crimes worldwide.

Publications in English-language journals in the fields of communication and legal sciences from the United Kingdom and the United States prevail, accounting for one-third of the total analysed production. These results highlight the lack of linguistic diversity and the need to promote plurality in scientific publications.

Regarding the content, quantitative research prevails over theoretical approaches. More than half of the articles study cases of online communication, although many of these cases do not specifically address hate speech but rather more general aspects of online polarisation or cyberbullying.

The topics covered in the articles highlight a low level of specificity. The majority of publications are loosely or barely related to hate speech, or they take a general approach, with a focus on online cases and computer-based methodologies. The most addressed topics in articles that tackle hate speech are related to politics or ideology, followed by those addressing origin and racism, and to a lesser extent, gender-related issues. Other topics that appear prominently include cyberbullying and online harassment. In the case of articles addressing ideological themes, cases of tension and political polarisation, especially online, stand out more than explicit hate speech.

It is revealing to observe that migration issues are addressed more than gender issues. There is not a single article that specifically tackles hatred towards women. The limited treatment of a group that has traditionally been a target of hate speech highlights the need to increase studies focused on gender issues. Similarly, it is surprising to unveil the limited articles related to education, despite the importance of education and information and media literacy in the context of hate speech. Prevention, especially at a young age, plays a crucial role in promoting values such as respect, tolerance, and acceptance of diversity. Hate speech can undermine these values and foster discriminatory, xenophobic, or racist attitudes among youth.

The set of results obtained demonstrates the relevance and significance of the topic and indicates that we are facing a serious situation that demands both preventive and proactive measures, as well as significant attention from the political and scientific spheres.

Possible avenues for further research could involve extending the study period until 2023, including non-open access articles and incorporating the analysis of other sources outside the traditional realm of scientific journals, allowing for a better understanding of the social impact of hate speech.

7. References

Adiprasetio, J.; Rahmawan, D.; Wibowo, K. (2021). A meta-analysis of hate speech in Indonesia: The yielding of an academic discourse to the discourse of authority. *Pacific Journalism Review*, 1-2 (27), 251-267. https://doi.org/10.24135/pjr.v27i1and2.1167

Alberdi Urquizu, C. (2019). Anti-immigration Discourse on Twitter: Discourse on Others and Hate Speech. *Caietele Echinox*, 36, 133-150. https://doi.org/10.24193/cechinox.2019.36.11

Attorney General. Circular 7/2019, de 14 de mayo, de la Fiscalía General del Estado, sobre pautas para interpretar los delitos de odio tipificados en el artículo 510 del Código Penal [Circular 7/2019, of 14 May, of the State Attorney General's Office, on guidelines for interpreting hate crimes under article 510 of the Criminal Code]. <https://www.boe.es/diario_boe/txt.php?id=BOE-A-2019-7771> [Consulta: 03/02/2024).

Arcila Calderón, C.; Sánchez Holgado, P.; Quintana Moreno, C.; Amores, J.; Blanco Herrero, D. (2022). Hate speech and social acceptance of migrants in Europe: analysis of tweets with geolocation. *Comunicar*, 71 (30), 21-35. https://doi.org/10.3916/C71-2022-02

Bajari, A.; Koswara, I.; Erlandia, D. (2021). Hatenography: An analysis of hate speech on facebook in 2019 Indonesian presidential campaign. *Jurnal Komunikasi: Malaysian Journal of Communication*, 4 (37), 122-141. https://doi.org/10.17576/JKMJC-2021-3704-08

Brown, A. (2017). What is hate speech? Part 1: The myth of hate. *Law and Philosophy*, 4 (36), 419-468. http://www.jstor.org/stable/44980883

Brown, A.; Sinclair, A. (2020). *The Politics of Hate Speech Laws*. Routledge.

Carlson, B.; Kennedy, T. (2021). Us mob online: The perils of identifying as indigenous on social media, *Genealogy*, 5 (52). https://doi.org/10.3390/genealogy5020052

Castells Fos, L.; Lopezosa, C.; Pont Sorribes, C.; Codina, L. (2023). Reputación, visibilidad y fidelización de audiencias de medios: una propuesta operativa desde la academia [How to put news media reputation, visibility and audience loyalty to work: a proposal from an academic point of view]. *Estudios sobre el mensaje periodístico*, 2 (29), 419-430. https://dx.doi.org/10.5209/esmp.85655

Center for Countering Digital Hate (2023). Star Framework: Safety by Design. <https://counterhate.com/blog/star-framework-safety-by-design/> [Consulta: 02/03/2024]

Codina, L., Lopezosa, C.; Freixa Font, P. (2022). Scoping reviews en trabajos académicos en comunicación: frameworks y fuentes [Scoping reviews in academic papers in communication: frameworks and sources], in *Información y Big Data en el sistema híbrido de medios-XIII Congreso Internacional de Ciberperiodismo.* Bilbao: Universidad del País Vasco, 67-85. http://hdl.handle.net/10810/56473

Council of Europe (1997). Recommendation No. R (97)20. <https://rm.coe.int/1680505d5b> [Consulta: 12/02/2024]

Council of Europe (2010). Recommendation CM/Rec(2010)5 of the Committee of Ministers to member states on measures to combat discrimination on grounds of sexual orientation or gender identity. <https://www.coe.int/en/web/sogi/rec-2010-5> [Consulta: 11/02/2024]

Da Silva, Izabel (2021). Bota fogo nesses vagabundos: Entextualizations of xenophobia in the textual trajectory of fake news. *Trabalhos em Linguística Aplicada*, 3 (*59)*, 2123-2161.http://dx.doi.org/10.1590/01031813829331620201106

Da Silva, M. P.; da Silva, L. S. (2021). Disseminação de discursos de ódio em comentários de notícias: uma análise a partir de notícias sobre o universo LGBT em cibermeios sul-mato-grossenses no Facebook [Hate speech dissemination in news comments: analysis of news about LGBT universe on Facebook cybermedia from Mato Grosso do Sul]. *Intercom: Revista Brasileira de Ciências da Comunicação*, 2 (44), 137-155. https://doi.org/10.1590/1809-5844202127

de Casas Moreno, P.; Parejo Cuéllar, M.; Vizcaíno Verdú, A. (2023). Hate speech on Twitter: The LGTBIQ community in Spain, in *Hate Speech on Social Media: A Global Approach.* Quito: PUCE Publications Centre, 143-159 https://labcomca.ubi.pt/wp-content/uploads/2023/05/Hate-Speech-on-Social-Media.pdf

Díaz Soto, J. M. (2015). Una aproximación al concepto de discurso del odio [An approach to the concept of hate speech]. *Revista Derecho del Estado,* 34, 77-101. https://doi.org/10.18601/01229893.n34.05

Essalhi Rakrak, A.; PinedoGonzález, R. (2023). #EspañaInvadida. Disinformation and hate speech towards refugees on Twitter: A challenge for critical thinking. *Profesional de la información*, 3 (32). https://doi.org/10.3145/epi.2023.may.10

European Commission against Racism and Intolerance (2016). ECRI General Policy Recommendation N°15 on combating Hate Speech. Council of Europe. <https://rm.coe.int/ecri-general-policy-recommendation-no-15-on-combating-hate-speech/16808b5b01> [Consulta: 11/02/2024]

European Commission (2016). The EU Code of conduct on countering illegal hate speech online. <https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en> [Consulta: 11/02/2024]

Gallardo Hurtado, L.; Selva-Ruiz, D. (2021). El fenómeno fan de las microcelebridades: el caso de AuronPlay [The fandom of microcelebrities: the case of AuronPlay]. *Revista Mediterránea De Comunicación*, 2 (12), 131144. https://doi.org/10.14198/MEDCOM.18894

Hansson, S.; Orru, K.; Torpan, S.; Bäck, A.; Kazemekaityte, A.; Meyer, S.; Ludvigsen, J.; Savadori, L.; Galvagni, A.; Pigrée, A. (2021). COVID-19 information disorder: six types of harmful information during the pandemic in Europe. *Journal of Risk Research*, 3-4 (24), 380-393. https://doi.org/10.1080/13669877.2020.1871058

Howard, J (2019). Free speech and hate speech. *Annual Review of Political Science*, 22, 93-109. https://doi.org/10.1146/annurev-polisci-051517-012343

Izquierdo Montero, A.; Laforgue Bullido, N.; Abril Hervás, D. (2022). Hate speech: a systematic review of scientific production and educational considerations. *Revista Fuentes*, 2 (24), 222-233. https://doi.org/10.12795/revistafuentes.2022.20240

Jääskeläinen, T. (2020). Countering hate speech through arts and arts education: Addressing intersections and policy implications. *Policy Futures in Education*, 3 (18), 344-357. https://doi.org/10.1177/1478210319848953

Krobová, T.; Zàpotocký, J. (2021). 'I Am Not Racist, But...': Rhetorical Fallacies in Arguments about the Refugee Crisis on Czech Facebook. *Journal of Intercultural Communication*, 2 (*21)*, 58-69. https://doi.org/10.36923/jicc.v21i2.14

Lev On, A. (2018). The anti-social network? Framing social media in wartime. *Social Media+ Society*, 4 (4). https://doi.org/10.1177/2056305118800311

Lilleker, D.; Pérez Escolar, M. (2023). Demonising Migrants in Contexts of Extremism: Analysis of Hate Speech in UK and Spain. *Politics and Governance*, 2 (11), 127-137. https://doi.org/10.17645/pag.v11i2.6302

Losada Díaz, J. C.; Zamora Medina, R.; Martinez Martínez, H. (2021). El discurso del odio en Instagram durante las Elecciones Generales 2019 en España [Hate speech on Instagram during 2019 General Election in Spain]. *Mediterranean Journal of Communication*, 2 (12), 195-208. https://doi.org/10.14198/MEDCOM.19142

Matsuda, M. (1989). Public Response to Racist Speech: Considering the Victim's Story. *Michigan Law Review*, nº. 8 (87), 2320–2381. https://doi.org/10.2307/1289306

Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D. (2009)*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, 4 (151), 264-269. https://doi.org/10.1016/j.ijsu.2010.02.007

Nadim, Marjan; Fladmoe, A. (2021). Silencing Women? Gender and Online Harassment. *Social Science Computer Review*, 2 (39), 245-258. https://doi.org/10.1177/0894439319865518

Nick, I. (2018). In the wake of hate: A mixed-method analysis of anonymous threatening communications sent during the 2016 US presidential election. *Nordic Journal of Linguistics*, 2 (41), 183-203. https://doi.org/10.1017/S0332586518000148

Obermaier, M.; Schmuck, D.; Saleem, M. (2021). I'll be there for you? Effects of Islamophobic online hate speech and counter speech on Muslim in-group bystanders' intention to intervene. *New Media & Society*, nº 9 (25), 2339-2358. https://doi.org/10.1177/14614448211017527

Oparaugo, B. (2021). Media and Politics: Political Communication in The Digital Age. *International Journal of Multidisciplinary Research and Explorer*, 2 (1), 14-23. http://dx.doi.org/10.2139/ssrn.3780554

Page, M.; McKenzie, M.; Bossuyt, P.; *et al*. (2021). PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Revista Española de Cardiología*, 9 (74), 790-799. https://doi.org/10.1016/j.recesp.2021.10.020

Park, S.; Han, S.; Kim, J.; Molaie, M. M.; Vu, Hoang D.; Singh, K.; Han, J. Lee, W.; Cha, M. (2021). COVID-19 discourse on Twitter in four asian countries: case study of risk communication. *Journal of medical Internet research*, 3, *23* (3). https://doi.org/10.2196/23272

Paz, M. A.; Montero Díaz, J.; Moreno Delgado, A. (2020). Hate Speech: A Systematized Review. *SAGE Open*, 4 (10). https://doi.org/10.1177/2158244020973022

Poole, E.; Giraud, Eva H.; de Quincey, E. (2021). Tactical interventions in online hate speech: The case of #stopIslam. *New Media & Society*, 6 (*23)*, 1415-1442. https://doi.org/10.1177/1461444820903319

Ramírez García, A.; González Molina, A.; Gutiérrez Arenas, M. del P.; Moyano Pacheco, M. (2022). Interdisciplinarity of scientific production on hate speech and social media: A bibliometric analysis. *Comunicar*, 72 (30), 129-140. https://doi.org/10.3916/C72-2022-10

Rollnert Liern, G. (2020). Redes sociales y discurso del odio: Perspectiva internacional [Social networks and hate speech: an international perspective]. *IDP. Revista de Internet, Derecho y Política*, 31, 1-14. http://dx.doi.org/10.7238/idp.v0i31.3233

Saputra, M.; Al Siddiq, I. H. (2020). Social Media and Digital Citizenship: The Urgency of Digital Literacy in the Middle of a Disrupted Society Era. *International Journal of Emerging Technologies in Learning*, 7 (15), 156–161. https://doi.org/10.3991/ijet.v15i07.13239

Saresma, T.; Karkulehto, S.; Varis, P. (2021). Gendered violence online: Hate speech as an intersection of misogyny and racism, in *Violence, gender and affect: Interpersonal, institutional and ideological practices.* Switzerland: Palgrave Macmillan, 221-243.

Tontodimamma, A.; Nissi, E.; Sarra, Annalina; Fontanella, L. (2021). Thirty years of research into hate speech: topics of interest and their evolution. *Scientometrics*, 126, 157-179. https://doi.org/10.1007/s11192-020-03737-6

Tricco, A.; Lillie, E.; Zarin, We; *et al*. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine*, 7 (169), 467-473. https://doi.org/10.7326/M18-0850

United Nations (2019). United Nations Strategy and Plan of Action on Hate Speech. <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf> [Consulta: 15/02/2024]

United Nations (2023). The Plan of Action for Women in Communities to Counter Hate Speech and Prevent Incitement to Violence that Could Lead to Atrocity Crimes. <https://www.un.org/en/genocideprevention/documents/publications-and resources/Napoli_PoA_final_1_Nov_23.pdf > [Consulta: 15/02/2024]

UNESCO (2021). UNESCO Recommendation on Open Science. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en> [Consulta: 15/02/2024]

Vallecillo Rodríguez, E.; Montejo Ráez, A.; Martín Valdivia, T. (2023). Automatic counter-narrative generation for hate speech in Spanish. *Procesamiento del Lenguaje Natural*, 71. https://rua.ua.es/dspace/bitstream/10045/137174/1/PLN_71_18.pdf

Vehovar, V.; Dejan J. (2021). Hateful and Other Negative Communication in Online Commenting Environments: Content, Structure and Targets. *Acta Informatica Pragensia*, 3 (10), 257-274. https://doi.org/10.18267/j.aip.165

Vrysis, L.; Vryzas, N.; Kotsakis, R.; Saridou, T.; Matsiola, M.; Veglis, A.; Arcila-Calderón, C.; Dimoulas, C. (2021). A Web Interface for Analyzing Hate Speech. *Future Internet*, 3 (13). https://doi.org/10.3390/fi13030080

YouTube (n. d.). Community Guidelines <www.youtube.com/intl/ALL_ca/howyoutubeworks/policies/community-guidelines/> [Consulta: 06/02/2024]