Information, Communication & Society

ISSN: 1369-118X (Print) 1468-4462 (Online) Journal homepage: www.tandfonline.com/journals/rics20

Social media platforms and the spreading of hate speech targeting sexual and gender minorities: a scoping review

Anne Clausen , Lena Frischlich & Peter Mayer

To cite this article: Anne Clausen , Lena Frischlich & Peter Mayer (11 Apr 2026): Social media platforms and the spreading of hate speech targeting sexual and gender minorities: a scoping review, Information, Communication & Society, DOI: 10.1080/1369118X.2026.2636138

To link to this article: https://doi.org/10.1080/1369118X.2026.2636138

Published online: 11 Apr 2026.

INFORMATION, COMMUNICATION & SOCIETY
https://doi.org/10.1080/1369118X.2026.2636138

Routledge
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

Social media platforms and the spreading of hate speech targeting sexual and gender minorities: a scoping review

Anne Clausen[a], Lena Frischlich[a] and Peter Mayer[b,c]

[a]Digital Democracy Centre, University of Southern Denmark, Odense, Denmark; [b]Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark; [c]Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, Karlsruhe, Germany

ABSTRACT

Today, social media users can access various social media platforms with different characteristics that offer distinct opportunity structures for the spread of hate speech. Yet little is known about the relationship between social media platforms with their unique characteristics, including content modalities and affordances, and the spreading of hate speech towards marginalized groups, including sexual and gender minorities (LGBTQIA+). This scoping review contributes to closing this gap. We conducted a scoping review according to the PRISMA Extension for Scoping Reviews (PRISMA-ScR). To account for the interdisciplinary nature of hate speech research, we searched communication-specific and broad academic databases (Communication Source, Scopus, Web of Science and Academic Search Premier). The full-text database includes $n = 145$ full-text papers. The analysis encompassed a mixed-methods approach: Texts were manually coded for social media platforms, content modalities and affordances, and a BERTopic model was trained to identify key topics across literature. Findings show that Twitter/X was the most researched platform, followed by YouTube, Facebook, and Instagram. Textual content was the most researched content modality, compared to visual and audiovisual content. Few publications explicitly mentioned platform affordances, while some pointed to anonymity and visibility as facilitators of the spreading of hate speech. Key topics revolved around the development of models and datasets for hate speech detection. Our findings call for future studies to investigate alternative platforms, visual and audiovisual content, platform affordances, and apply a research focus beyond the development of detection models to understand how platform dynamics contribute to the spreading of hate speech.

ARTICLE HISTORY
Received 3 November 2025
Accepted 18 February 2026

KEYWORDS
Social media platforms; hate speech; sexual and gender minorities; scoping review; affordances; BERTopic

CONTACT Anne Clausen aclausen@sam.sdu.dk Digital Democracy Centre, University of Southern Denmark, 5230 Odense, Denmark

Introduction

Social media users today have access to a large variety of different social media platforms with distinct characteristics (Bayer et al., 2020). The current social media environment not only includes different types of social media, such as microblogs, video streaming platforms, etc., but also offers various alternative platforms targeting extreme or fringe communities (Zeng & Schäfer, 2021). The average user nowadays uses 6.8 different platforms each month (Kemp, 2025).

Social media platforms differ from each other in terms of content modality (text, image, video), technical features (e.g., 'like' button), and affordances (e.g., anonymity). These differences in platform characteristics create distinct opportunity structures for the spreading of intolerance and hate speech. Intolerance is the use of discriminatory or exclusionary speech towards others (Rossini, 2020), while hate speech is a specific subtype of intolerance aimed at people because of their perceived collective identity (e.g., sexual orientation, gender identity) (Rossini, 2020; United Nations, 2019). Hate speech may be harmful and lead to the discrimination of specific individuals or groups (Siegel, 2020; United Nations, 2023) and trigger trauma responses in the targets (Leets, 2002). Hate speech can further harm democratic participation as exposure may scare people away from engaging in online public debate to avoid attacks (Amnesty International Danmark, 2025; Geschke et al., 2019). This review concentrates on hate speech targeting sexual and gender minorities, who face disproportionate levels of targeting relative to other marginalized groups (ADL Center for Technology & Society, 2024). This focus allows for an in-depth account of the interplay between platform characteristics and the spreading of hate speech towards a marginalized group, central to current societal debate. To our knowledge, no review has systematically synthesized the literature to report what is known about platform characteristics and the spreading of hate speech towards sexual and gender minorities. The current study employs a scoping review to address this gap. The article is structured as follows: First, we describe the conceptual background and scope of our work and present our research questions. We then explain our methodology, present results of our analysis, discuss findings and methodological limitations of our study, and lastly, we provide conclusions and distill recommendations for future research.

Conceptual background

Social media platforms

Social media platforms are interactive websites and apps where social media users can create and share content online (Bayer et al., 2020; Carr & Hayes, 2015). Examples of mainstream social media platforms include Facebook, Instagram, YouTube, TikTok and Twitter/X among many others. Examples of alternative social media platforms include Telegram, 8kun, Gab and 4chan.

Social media platforms can be characterized by their specific content modalities, features and affordances. Social media content can take different forms of modalities – text (e.g., posts, messages, comments), visual (e.g., images, memes) or audiovisual (e.g., videos, reels). Multimodality is understood as the use of different modes of communication to express meaning (Adami, 2017), and in the context of social media platforms, content can be multimodal if it combines text, images, audio and video, such as a TikTok video with audio and text overlays. In an experimental study, different effects were

reported for text-based versus multimodal hate speech (Menini et al., 2020), indicating the need to account for different modalities. User interaction with social media content is based on the features of a platform. Features are technically embedded in the platform design and refer to the opportunities that platforms offer to their users to engage with the content, for example, the possibility to follow, share, comment (Schulze et al., 2024), like, repost, retweet, upvote, etc. (Bimber & Gil de Zúñiga, 2020). Affordances (often referred to as technological affordances (Kakavand, 2024; Schulze et al., 2024)) occur when users interact with the features of a platform and can be broadly described as 'possibilities for action' (S. K. Evans et al., 2017). An example of a feature and an affordance on the platform Twitter/X is the feature of a freely selectable handle and account name, which may afford anonymity for its users (Schulze et al., 2024). Each platform offers its own set of features and affordances (S. K. Evans et al., 2017; Schulze et al., 2024).

Specific affordances, including anonymity, visibility, collectivity and interactivity have been linked to online-radicalization processes (Schulze et al., 2024), a particularly extreme form of intolerance. Plausibly, the amount of hate speech also depends on the affordances of a platform (Ben-David & Matamoros Fernández, 2016). The affordance of anonymity eases hate speech by allowing users to disconnect online behavior from offline identity, which can lead to toxic disinhibition and the expression of hatred unlikely to occur in face-to-face interactions (see online disinhibition effect, Suler, 2004). Visibility impacts the reach of content, facilitated by digital means (external visibility) or social presence online (internal visibility) (Schulze et al., 2024). Hate speech achieves higher external visibility than non-hateful content via social media sharing mechanisms, evidenced by retweet dynamics on Twitter/X (Maarouf et al., 2024). For sexual and gender minorities, it has been demonstrated that content creators advocating for minority rights enact internal visibility by creating social media content claiming representativity (Badaoui, 2024); however, this can also attract hateful attacks. Collectivity arises from platform design enabling group formation (Schulze et al., 2024). Group norms and dynamics can legitimize hate speech, facilitating coordination and escalation (Schmitz et al., 2022). This can ease the spread of 'us versus them' narratives, portraying minorities as threatening out-groups. The multilevel affordance interactivity is rooted in other affordances. For example, interactivity facilitates exchange and coordination within collectives through sharing and commenting (Schulze et al., 2024), thereby contributing to processes where hate speech can spread.

Intolerance and hate speech

Intolerance has been defined as the use of discriminatory or exclusionary speech towards others (Rossini, 2020). This scope is relatively broad and does not necessarily include the expression of hatred. Hate speech is a specific form of intolerance, as suggested by Rossini (2020), that can be understood as "any kind of communication in speech, writing or behavior, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, color, descent, gender or other identity factor" (United Nations, 2019). More specifically, hate speech is when discriminatory language or hateful expressions are targeted towards individuals because of their (ascribed) membership in a social group or category (Frischlich, 2023; Rieger et al., 2021). Explicit hate speech includes, for example, transphobic slurs, while implicit hate speech can include the transmission of negative stereotypes, subtle discrimination, or derogations disguised as humor (e.g., homophobic jokes).

Sexual and gender minorities

Sexual and gender minorities (non-heterosexual orientations and gender identities outside of the traditional gender-binary) are important groups to study in the context of social media and hate speech as such hostility intersects with broader issues of stigma and minority stress.

Sexual and gender minorities are frequent targets of hate speech on social media (ADL Center for Technology & Society, 2023, 2024; Analyse & Tal & Os & Data, 2025; Luke, 2021; Paterson et al., 2018), and social media users perceive homophobic hate speech as 'less uncivil' and accordingly less worthy of counter speech than racist or misogynistic hate speech (Obermaier et al., 2023). This perception may intensify the harm experienced by victims, with hate speech targeting sexual minorities capable of triggering trauma-like responses similar to antisemitic hatred (Leets, 2002). Further, victims of transphobic hate speech are even overrepresented among those reporting suicidal intentions because of hateful attacks (Williams & Tregidga, 2014). Denying the severity of such hate speech can lead to secondary victimization and likely contribute to underreporting of hate speech and discrimination (Vergani & Navarro, 2023; Weise et al., 2021; Williams & Tregidga, 2014). While this hostility occurs across platforms, platform policies can too exacerbate the problem, as seen in the sharp rise in transphobic and homophobic speech on X following Elon Musk's acquisition of Twitter (Hickey et al., 2025).

Conceptual scope

Other studies have investigated literature on social media platforms and hate speech targeting sexual and gender minorities. One review provided a bibliometric overview (Sánchez-Sánchez et al., 2024), another review reported on psychological consequences in victims (Stefanita & Buf, 2021), and a third review examined the impact on mental health and social inclusion in the setting of South Africa (Adekola, 2025). Accordingly, to the best of our knowledge, no review has systematically synthesized the literature to report what is known about platform characteristics (which platforms, content modalities, and affordances), and the spreading of hate speech towards sexual and gender minorities, nor has it reported on the key topics addressed in this body of literature. The current study employs a scoping review to address these gaps.

Objectives and research questions

The objectives of the scoping review are to provide a comprehensive overview of what is known from the scholarly literature about platform characteristics and the spreading of hate speech towards sexual and gender minorities, uncover topics discussed so far, identify gaps in the literature and distill recommendations for future research.

Research questions

Guided by the objectives of the scoping review, we aim to answer the following Research Questions (RQs):

RQ1: Which social media platforms have been the focus of empirical investigations into hate speech targeting sexual and gender minorities?

RQ2: What do we know about the spreading of hate speech towards sexual and gender minorities in different social media modalities (text, visual or audiovisual content)?

RQ3: What do we know about the association between platform affordances (anonymity, visibility, collectivity, interactivity) and the spread of hate speech towards sexual and gender minorities?

RQ4: What are the key topics addressed in research on hate speech directed at sexual and gender minorities on social media platforms?

Methods

To answer the research questions, we conducted a scoping review according to the PRISMA Extension for Scoping Reviews (PRISMA-ScR) (Tricco et al., 2018). The PRISMA-ScR Checklist is provided in: https://osf.io/sma4f/

The study was pre-registered at: https://osf.io/rfe4s/[1]

As this review was intended to map the state of evidence, we focused on the inclusion of empirical studies. Figure 1 provides an overview of the literature selection process using a PRISMA flow diagram. Overall, $n = 2213$ references were identified from the literature search. After duplicate removal, $n = 1766$ references were manually screened for eligibility for inclusion in the review, which ultimately included $n = 145$ research publications in the full-text database for further analysis. Citations for the full text database are provided in the OSF repository linked above.

Search strategy

We conducted a primary database search in relevant broad and communication-specific peer-reviewed databases to identify scientific literature. We performed a supplemental search in Google Scholar to identify complementary sources to get a broad understanding of the phenomenon. In the first step, we screened literature on the title/abstract level and included/excluded material for full text screening based on a set of predefined criteria (See Table 1). Afterwards, we assessed the material in full text for final inclusion/exclusion in the review. Finally, we conducted a snowball search/citation search by searching through the bibliographies of included literature.

Primary database search

The database search was performed as a Boolean search in the databases *Scopus*, *Web of Science, Academic Search Premier,* and *Communication Source (includes Communication and Mass Media Complete* and *Communication Abstracts)*. The primary database search was performed in July 2025. The search string included three search blocks: the first block included terms for social media platforms, the second block included terms for hate speech, and the third block included terms for sexual and

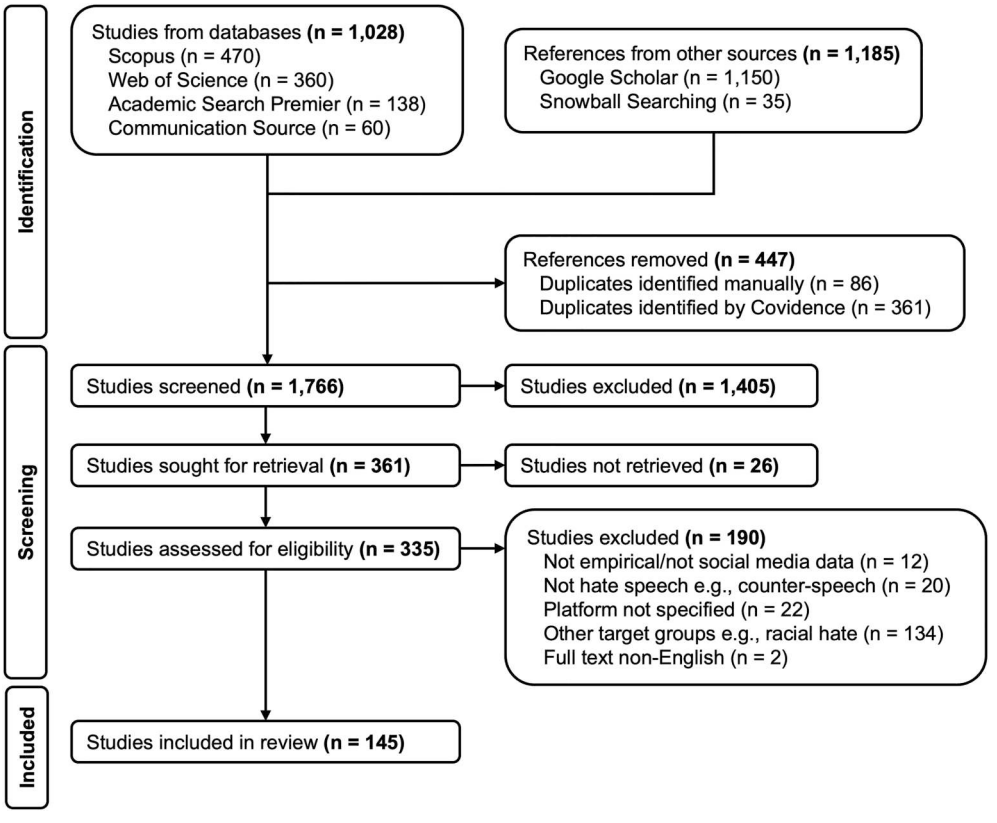

Figure 1. PRISMA flow diagram.

gender minorities. Search terms for blocks 1 and 2 drew inspiration from a search string by Strippel and colleagues (Strippel et al., 2023), and search terms for block 3 drew inspiration from a search string by Vergani and colleagues (Vergani et al.,

Table 1. Inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
• Empirical investigations using actual social media data for analyses	• Non-empirical material, including theoretical papers, conceptual discussions, opinion pieces, and studies using synthetic/simulated data.
• Written in English language	• Non-English publications
• Full text available	• Material that has no available full text (e.g., abstracts from a conference without proceedings, unpublished B.Sc. theses)
• Peer-reviewed published articles, articles in conference proceedings, editorials, books and book chapters	
• Grey literature that is not necessarily peer-reviewed, e.g., essays, dissertations, etc.	
• Material that addresses all 3 search blocks from the Boolean search strategy. Material was included if it addressed social media platforms, hate speech and sexual and gender minorities according to the definitions provided in the 'Conceptual background' section.	• Material that addresses only 1 or 2 of the search blocks from the Boolean search strategy. For example, material that investigates online hate speech that is not directed at sexual or gender minorities or hate speech that has not occurred on a social media platform.

2024). Search terms were adapted to fit the scope of this review. The search string used the Boolean search modifiers asterisk* for word truncations, and quotation marks "" for exact phrases of more than one word, e.g., "social media". The search string with blocks 1–3 was further refined in collaboration with an information specialist from the library at the first author's university and is presented in Table 2. We did not apply any limits for publication year. The database search resulted in a total of $n = 1028$ publications for further processing (see Figure 1).

Supplementary Google Scholar search

Google Scholar has limited search capacities compared to the academic databases (Vergani et al., 2024), we thus applied an alternative strategy using the scholarly module in Python (pypi.org, 2023). We searched for each individual query from the search string ($N = 7840$ query combinations) and obtained a maximum of 30 hits per query. The Google Scholar search was finished in September 2025. Duplicate records were excluded, and it was ensured that the remaining records were relevant for screening by application of regular expressions to titles and abstracts to check if they contained the search terms from the original Boolean search string. This resulted in a total of $n = 1150$ references for further processing (See Figure 1).

Title/abstract screening

Search results were uploaded to the online tool Covidence, which is used to perform literature reviews (covidence.org, 2025). The material was manually screened by the first author on the title/abstract level using Covidence and assessed for eligibility for inclusion, based on the set of pre-defined criteria outlined in Table 1. If it was unclear from the title and abstract whether the material addressed all

Table 2. Boolean search string for database search.

BLOCK 1: **Social media platforms** (OR between search terms)	AND	**BLOCK 2:** **Hate speech** (OR between search terms)	AND	**BLOCK 3:** **Sexual and gender minorities** (OR between search terms)
Twitter, X, "Twitter/X", "X/Twitter", "Twitter (X)", "X (Twitter)", Twitter-X, X-Twitter, Facebook, Instagram, Youtube, Snapchat, TikTok, Twitch, Discord, WhatsApp, Signal, Threema, Telegram, Gab, Bitchute, Parler, "Truth Social", 4Chan, 8kun, DLink, Rumble, Gettr, Linkedin, Tumblr, Pinterest, 9gag, "across platform*", cross-platform*, "social platform", "social media", "online social media", "online social network*", "online discussion", "social network*", "digital social network*", "social web", "social network* site*", "dark social", "alternative social", "fringe communit*", "technological affordance*", multimodal, "platform feature*"		"hate speech", cyberhate, "harmful speech", "violent speech", "extreme speech", "toxic speech", incivil*, uncivil*		LGBT*, lesbian, gay, homosexual, bisexual, transgender, queer, asexual, intersex, nonbinary, non-binary, homophobi*, gender*, "sexual minorit*", "gender minorit*", lesbophobi*, gayphobi*, biphobi*, transphobi*, queerphobi*

three search blocks from the search strategy, a precautionary principle was applied, meaning the material was included for full-text screening to finish the assessment. Decisions were validated by the second author. A total of $n = 1766$ references were screened on the title/abstract level, of which $n = 361$ were included for further processing.

Full-text screening

Inclusions from the title/abstract screening were sought for retrieval in full text. Full texts could not be retrieved for $n = 26$ references, which left $n = 335$ full texts for screening. The inclusion criteria still applied; however, in the full text screening, it was reassured that the material addressed all three search blocks (in case of doubt, e.g., if the abstract was not fully informative). The first author performed the full-text screening manually. After exclusion of publications that did not meet the inclusion criteria a total of $n = 145$ studies were included in the full text database for further analysis (See Figure 1).

Analytical approach

The analysis comprised a mixed-methods approach. A combination of manual and computational methods allowed for detailed categorization and broader topical insights. First, each text was manually coded to extract key descriptive information, such as studied language(s), and examined social media platforms, content modalities, and affordances. For affordance coding, papers were keyword-searched for mentions of relevant terms: 'affordance', 'anonymity', 'visibility', 'collectivity' and 'interactivity'. If mentioned, it was manually assessed if the affordance was referred to as an action possibility occurring due to platform design/features (aligned with the definition of affordances presented in *Conceptual background*). This manual coding enabled a structured overview of the scope and characteristics of included studies, necessary for answering RQs 1–3. Then we trained a BERTopic model (Grootendorst, 2022) to identify key topics discussed across the literature, answering RQ4.

Corpus preprocessing and topic modeling with BERTopic

All $n = 145$ research publications were converted into text files (.txt) for topic modeling analyses with BERTopic (Grootendorst, 2022). Each text file included all text from the publication, excluding author keywords and reference lists, to avoid bias in training the topic model. Text preprocessing included lowercasing, lemmatization, tokenization, removal of non-alphabetic tokens, years, punctuation, and standard as well as custom stop words. Removal of custom stop words and phrases included standalone mention of et, al., and al (pertains to academic referencing), as well as hate, hate speech and hs, as these were search criteria and thus non-informative for the topic modeling. Next, we trained a BERTopic model to identify topics in our corpus. BERTopic leverages transformers and class-based TF-IDF to create dense clusters, thereby providing readily interpretable topics (Grootendorst, 2022). Each text file's semantic embedding was created using the sentence transformer model all-MiniLM-L6-v2. Dimensionality reduction

was applied to embeddings with UMAP and clustered with HDBSCAN, applying a minimum cluster size of 3. The latter meant that a cluster needed a minimum of 3 research publications to form a topic. This relatively small cluster size was decided upon because texts were long (research publications), and manual inspection of initial results confirmed that it produced a fair number of topics while keeping the outlier cluster at a tolerable size. For each topic, 10 keyword representations were extracted, and topics were then manually labeled by the first author upon inspection of the representations.

Results

Descriptive analyses

Figure 2 provides an overview of the descriptive characteristics of the analyzed publications. The first paper in our database was published in 2016. Since then, there has been an overall yearly increase in the number of publications within the subject and steep increases from 2021 to 2022, and from 2023 to 2024, respectively. The literature search was conducted mid-year, which explains the apparent 2025 decline as publications from the second half of the year are not included. English was the most researched language (54% of publications; $n = 79$), followed by Spanish (26%; $n = 37$) and Tamil (24%; $n = 35$). Percentages reflect non-mutually exclusive inclusion of languages, meaning that one paper could have investigated hate speech in multiple languages. Most publications (74%; $n = 108$) used automated methods for the detection/classification of hate speech in data from social media platforms. Publications focusing solely on sexual and gender minorities as hate targets constituted 63% ($n = 91$) of our database, while 37% ($n = 54$) covered multiclass targets, meaning that besides sexual and gender minorities, other hate targets were also covered, e.g., hate targeted at race, ethnicity, religion, etc.

Main analyses

Figure 3 provides an overview of the platform characteristics identified in the analyzed publications. Percentages reflect non-mutually exclusive inclusion for platform coverage, content modalities and affordance coverage, meaning that one paper could, for instance, have investigated data from multiple platforms.

Social media platforms

RQ1 asked for the social media platforms studied so far. In our database, the microblogging service Twitter/X was the most researched platform, with 50% ($n = 72$) of publications analyzing data from this platform, followed by YouTube with 37% ($n = 53$), Facebook with 13% ($n = 19$), and Instagram with 7% ($n = 10$). All other platforms were studied in less than 5% of publications. Notably, alternative social media platforms such as Gab or 4Chan, which have been described as particularly hateful contexts (e.g., Rieger et al., 2021; Zannettou et al., 2018), are understudied compared to the larger mainstream platforms.

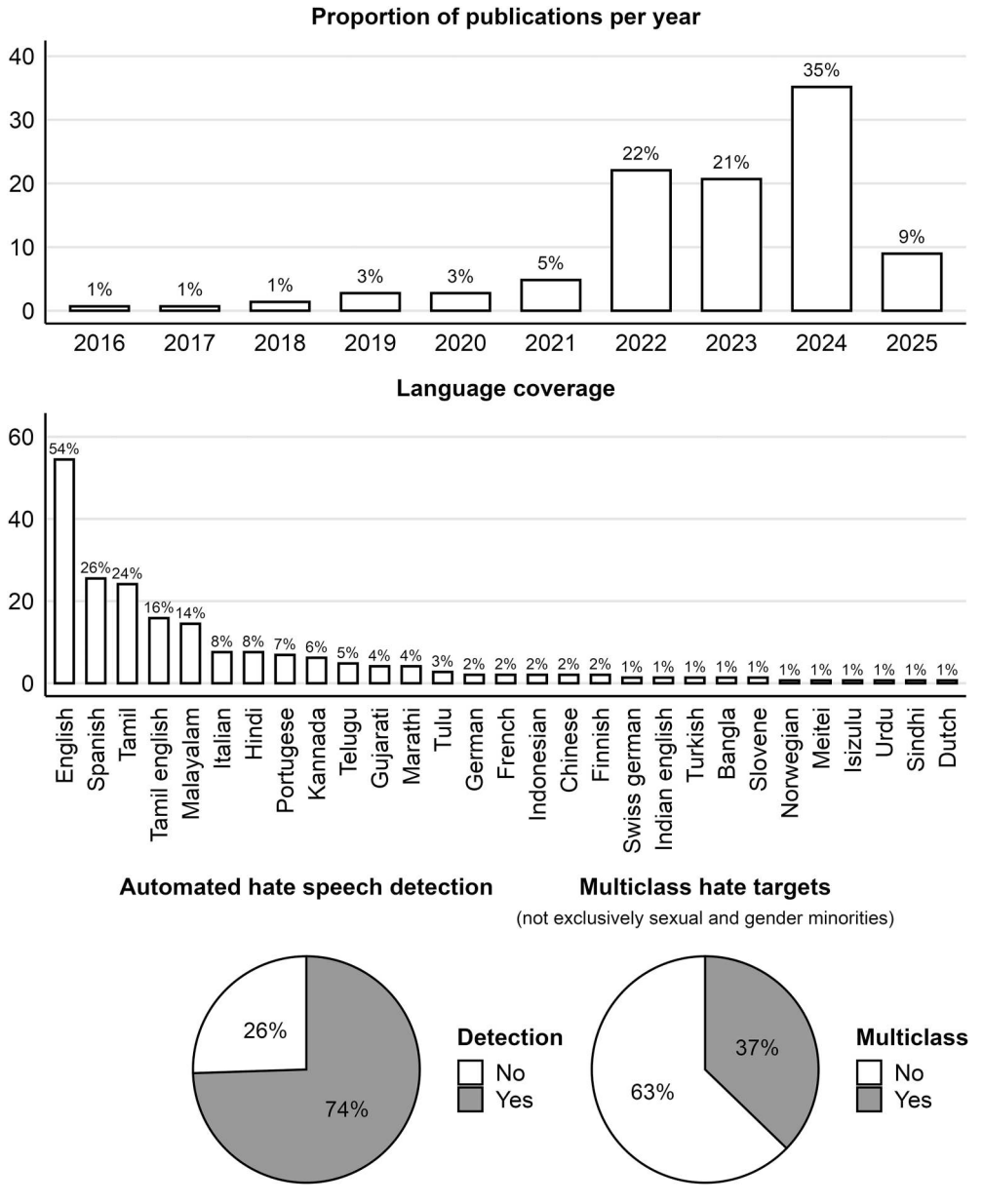

Figure 2. Descriptive characteristics of publications (*n*=145).

Content modalities

RQ2 asked which content modalities were investigated. Textual content, typically including comments, was analyzed by 96% ($n = 139$) of publications, whereas visual content was investigated in 8% ($n = 11$) of publications and audiovisual content in 2% ($n = 3$) of publications.

Visual content included text-embedded images such as memes (G. Kumar et al., 2021; R. Kumar et al., 2024; Mkhize et al., 2020; Oehmer-Pedrazzi & Pedrazzi, 2024; Shah et al.,

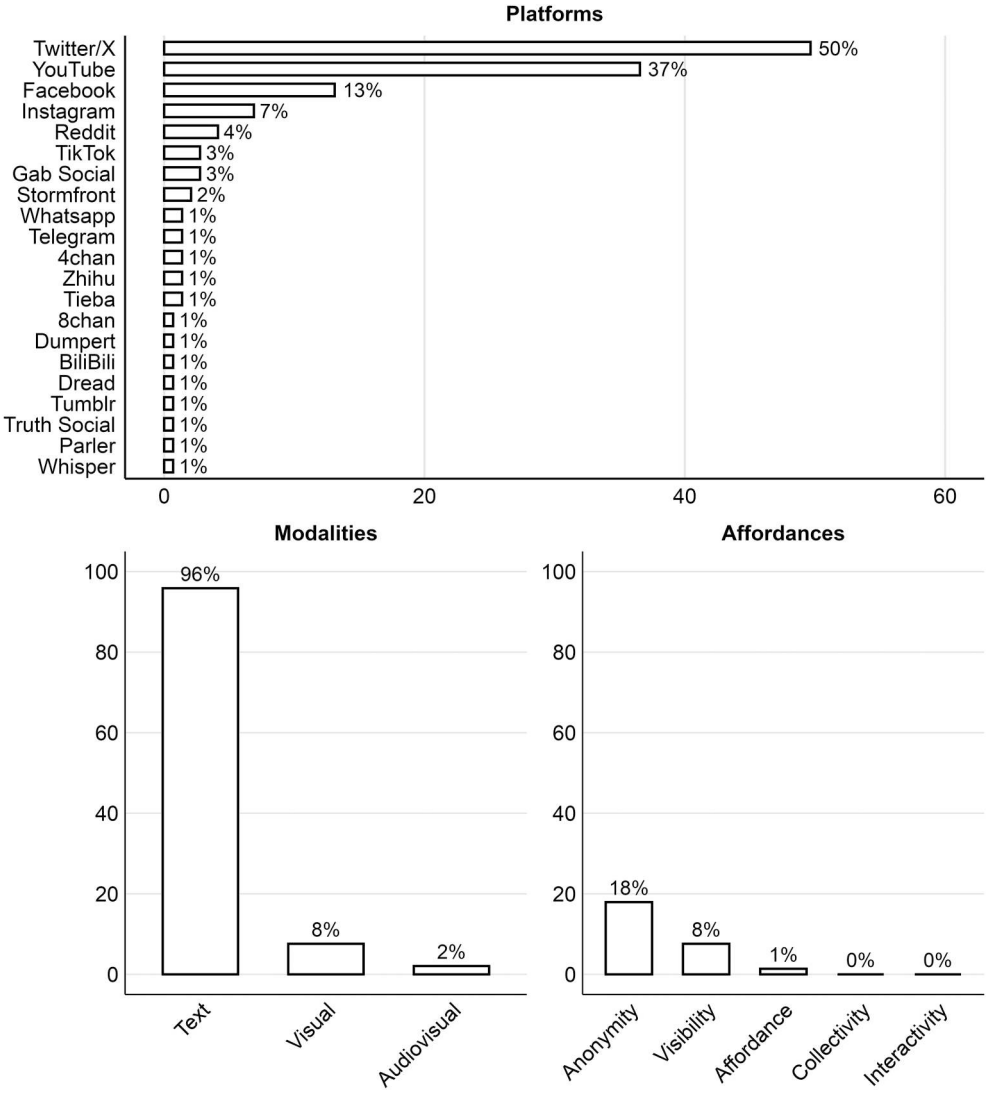

Figure 3. Platform characteristics in publications (*n*=145).

2024; Venegas Carrasco et al., 2024; Yang et al., 2022), emojis and GIFs (Oliveira et al., 2024; Venegas Carrasco et al., 2024), cartoons (Oehmer-Pedrazzi & Pedrazzi, 2024) and images (Brody et al., 2024; G. Kumar et al., 2021; Mkhize et al., 2020; Nguyen et al., 2025; Oehmer-Pedrazzi & Pedrazzi, 2024; Oliveira et al., 2024; Sívori & Zilli, 2022).

Audiovisual content included videos (W. Evans, 2024; Wang et al., 2024), and audio extracted from videos (R. Kumar et al., 2024).

Affordances

RQ3 tapped into platform design and the spreading of hate speech, focusing specifically on four affordances (anonymity, visibility, collectivity and interactivity) linked to online

radicalization in prior research (Schulze et al., 2024). Affordances were mentioned explicitly in 1% ($n = 2$) of publications. Anonymity and visibility were mentioned as results of platform design and as elements essential to the spread of hate speech on platforms in 18% ($n = 26$) respectively, 8% ($n = 11$) of publications. This finding indicates that publications referring to anonymity or visibility did not necessarily label them as affordances. Notably, publications mentioning affordances, anonymity or visibility typically referred to the terms in the background or in the discussion but did not empirically examine the link between affordances and the spreading of hate speech. Exceptions include a few publications ($n = 2$) measuring hate speech visibility through digital means, including likes and hashtags, respectively (Hickey et al., 2025; Santos Fernández, 2024). Engagement metrics (likes and reposts) of hate tweets were used as proxies of visibility on Twitter/X, revealing a statistically significant rise in hate tweet visibility, as measured by weekly likes, before and after Elon Musk's acquisition of Twitter (Hickey et al., 2025). Specific hashtags on Twitter were identified as amplification tools for the spreading of homophobic hate speech and the links between the top hashtags were analyzed in a co-occurrence network analysis (Santos Fernández, 2024). In the remainder of publications referring to visibility, the mentions were not supported by empirical analyses. They covered different aspects, for example, that sexual and gender minorities are easy targets of online hate speech, because they have a growing internal visibility on social media platforms (Arce-García & Menéndez-Menéndez, 2022; Hara et al., 2024; Klutse et al., 2023; Venegas Carrasco et al., 2024). Conversely, one Telegram study argued that gender minorities were difficult to study as hate targets in their data, possibly due to low internal visibility of this group in Telegram channels or because coded language was used to target them (Alvisi et al., 2025).

Anonymity was mentioned as a facilitator of the spread of hate speech because perpetrators can stay anonymous when they create hateful content and avoid facing consequences, however these mentions were not backed by empirical analyses (examples: Chakravarthi et al., 2021; Rieger et al., 2021; Unlu et al., 2025). Collectivity and interactivity were not covered by any publications in the database. In general, our database revealed a lack of empirical data to demonstrate a causal relationship between affordances and the spreading of hate speech.

Topics discussed in research literature

RQ4 asked for the key topics addressed in research on social media platforms and hate speech directed at sexual and gender minorities. We discuss the results of the BERTopic model analysis in the following. In our corpus of $n = 145$ research publications, BERTopic identified 14 coherent topic clusters, and an outlier topic (−1) containing documents that did not align with a single topic. Table 3 provides an overview of the topics with publication count and topic labels. The detailed output of the BERTopic analyses is provided in: https://osf.io/sma4f/

The larger part of topics centered around detection models and data analysis (topics 0, 1, 4, 5, 7, 8, 9, 11 and 12) and grouped together publications that trained and evaluated models for hate speech detection, specifically focusing on models for text classification, including BERT and RoBERTa and CLIP for visual data. Indicative of textual modality, a handful of topics referred to comments (topics 4, 12, 13), while

Table 3. Overview of key topics identified by BERTopic.

Topic	Count	Topic label	Topic representations
−1	30	Datasets and language modeling on comments and tweets	dataset, model, data, comment, language, social, tweet, word, media, task
0	19	BERT models for detection in tweets	model, toxic, dataset, data, detection, target, tweet, base, performance, bert
1	15	Language models for homophobia and transphobia	language, model, tamil, english, homophobia, transphobia, dataset, task, comment, data
2	14	Transgender rights and TERFs	people, trans, transgender, social, gender, woman, group, user, terf, law
3	10	Minority groups in political posts	post, muslim, group, political, party, topic, study, community, account, platform
4	8	Sentiment analysis of homophobic comments	model, sentiment, dataset, task, homophobic, language, comment, mix, analysis, english
5	8	Task-based detection with transformers for Spanish tweets	task, model, label, tweet, result, spanish, transformer, fine, subtask, detection
6	7	Italian online communities on Twitter and Telegram	italian, tweet, chat, term, community, word, telegram, gay, refer, stereotype
7	6	Transphobia analysis with language models	language, model, condition, data, task, tulu, transphobia, subword, macro, script
8	6	Homophobic language detection with BERT and RoBERTa	model, homophobic, bert, dataset, offensive, language, roberta, english, data, content
9	5	Models for data in Tamil and English language	tamil, model, english, dataset, layer, comment, indicbert, language, score, word
10	5	Twitter analysis of minority group targets in Portuguese	portuguese, group, target, aggressive, comment, social, tweet, annotator, roma, analysis
11	5	Multimodal image, text, meme analysis via CLIP	image, multimodal, model, text, meme, clip, dataset, sentiment, accuracy, task
12	4	Detection of aggressive comments with annotated data	dataset, aggression, comment, language, annotation, cue, model, tag, different, meitei
13	3	Right-wing channel topics and comments	channel, right, comment, topic, wing, video, word, court, amendment, caption

others referred to tweets (0, 5, 6, 10), of which the latter also indicated a focus on platform-specific analyses, heavily centered around Twitter/X. Other topics centered on language-specific data in Spanish (topic 5), Italian (topic 6), Tamil and English (topic 9) and Portuguese (topic 10). Several topics focused on homophobia and/or transphobia (topics 1, 4, 7 and 8), indicating that studies paid attention to specific types of hate speech targeting sexual and gender minorities rather than treating it as an overall undifferentiated category. Ideology or politics appeared in a handful of topics (topics 2, 3 and 13), which points to a broader contextual consideration of the hate speech context, specifically considering TERFs (Trans Exclusionary Radical Feminist) (topic 2) and right-wing ideology (topic 13). The outlier topic (topic −1) included around 20% of the publications that did not fit into specific topics. Even so, the representations for the outlier topic indicated a focus on detection models and data analysis of comments and tweets in accordance with findings from the larger part of topics identified in the remaining publications in our full text database. Rather than constituting noise, this cluster reflected publications that discussed detection and data analysis with greater variety compared to the studies that were assigned specific topics. Overall, the topic modeling revealed that publications investigating social media platforms and hate speech targeted at sexual and gender minorities are heavily focused on detection models and data analysis for hate speech classification, often driven by specific datasets (e.g., for Tamil or Spanish language).

Discussion

The current social media environment provides opportunity structures for the spreading of hate speech against sexual and gender minorities. We do, however, lack a systematic understanding of the interplay between platform characteristics and the spreading of hate speech towards sexual and gender diversity. In this work, we conducted a scoping review focusing on $n = 145$ empirical studies to contribute to closing this gap in the literature. We presented a comprehensive overview of the relationship between social media platforms with their unique characteristics, such as modalities, affordances, and the spreading of hate speech towards sexual and gender minorities, as well as uncovered topics addressed by this body of literature. In the following, we discuss our findings, outline opportunities for future research, and discuss methodological limitations of our research design.

In answering RQ1, we established that mainstream platforms, and particularly Twitter/X, was by far the most researched platform among publications in our database. Sánchez-Sánchez et al. reported a similar finding in their bibliometric analyses of homophobia and transphobia on social media (2024). It is noteworthy in the interpretation of results that several datasets were reused across publications, which may skew the overall impressions of platform coverage. This is, for instance, the case for papers who analyzed the same publicly available dataset or for conference papers where participants analyzed the same dataset for a task and afterwards published individual papers of their analyses. Additionally, the fact that most empirical investigations rely on Twitter data may also be attributed to the platform's historical data availability. Twitter was long considered a 'researcher-friendly' platform due to its free and open API (Murtfeldt et al., 2024), which facilitated academic studies. However, following Elon Musk's acquisition of the platform (now called X), free API access was discontinued in 2023 (Murtfeldt et al., 2024). In the years ahead, it remains to be observed whether researchers shift their focus to alternative platforms that offer more accessible data for academic analyses. Besides larger mainstream platforms, including Twitter/X, YouTube, Facebook and Instagram, future studies could investigate hate speech targeting sexual and gender minorities on alternative platforms, which were largely understudied in our database to capture a more comprehensive picture of hate speech across diverse social media environments.

In answering RQ2, we found that textual content was the most researched content modality, leaving visual and audiovisual content comparatively under-researched. Analyzing visual or audiovisual content is more complex than analyzing solely textual content, because multiple elements (text, image and audio) require interpretation as individual elements and in combination with each other to account for multimodality (Hee et al., 2024). The complex nature of multimodal content makes it less suitable for automated methods and analysis, which aids in explaining why the current hate speech literature is predominantly centered on textual content, which requires fewer resources in data collection and analysis (Hee et al., 2024). In our findings, it is notable that $n = 53$ studies analyzed data from YouTube (video sharing platform), however only $n = 2$ studies analyzed audiovisual content (R. Kumar et al., 2024; Wang et al., 2024), while the remaining studies ignored the audiovisual content and focused solely on written comments to videos. However, since hate speech is not a text-only problem and

video-based platforms such as TikTok and YouTube are highly popular (Kemp, 2025), it is necessary to move beyond a textual focus and account for visual and audiovisual content to understand hate speech dynamics in current social media environments.

In answering RQ3, we established that the concept of affordances was rarely mentioned in publications in our database. The publications that did, focused on anonymity and visibility without necessarily labeling them as platform affordances or providing empirical analyses to demonstrate the relationship between affordances and the spreading of hate speech. Apart from anonymity and visibility, other affordances, including collectivity (the building and maintaining of online communities (Schulze et al., 2024)), have been linked to hate speech dynamics, however, without receiving attention by publications in our database. For instance, in a study of anti-Islam Facebook groups, the group atmosphere and specifically the way group members used humor in their discussions legitimized spread of hateful sentiment (Fangen, 2020). The low uptake of the affordance concept in our database may also be ascribed to the criticism the concept has received due to the ambiguity of the term and its inconsistent use in research, which has made it difficult for scholars to translate the concept into usable methodology (Ronzhyn et al., 2023). Even so, the concept offers a useful framework to study platform design in relation to hate speech dynamics, thus leaving potential for future studies to investigate platform affordances in greater depth, and particularly with empirical data. Alternative analytical lenses to the concept could add valuable nuances. Even though not covered by any publications in our database, the concept of gendered affordances, introduced by Schwartz and Neff (2019) and later expanded in other works (e.g., Díaz-Fernández & García-Mingo, 2024; Kettrey et al., 2024; Semenzin & Bainotti, 2020), may be relevant since sexual and gender minority identities inherently challenge heteronormative ideals and can become targets of hate speech for doing so. Research on non-consensual dissemination of intimate images on Telegram shows how affordances such as anonymity and community formation facilitated the spread of harmful content and performances of toxic masculinity (Semenzin & Bainotti, 2020). This exemplifies how a gendered lens can provide insights into how affordances contribute to the spread of harmful content and the reproduction of gendered inequalities, providing a meaningful starting point for future research.

In answering RQ4, we found that key topics discussed in the research literature in our database primarily focused on the creation of datasets and hate speech detection. However, the mere detection of hate speech does not necessarily solve its spread on platforms, as it is heavily dependent on subsequent moderation, which leaves room for reflection on the direction for future research. Future research could, besides detection, expand its focus to how platform design influences hate speech dynamics to provide insights into how we can create more tolerant digital environments. Important dimensions, therefore, remain unexplored with regard to how platform dynamics contribute to the spread of hate speech on social media platforms.

Findings from the descriptive analyses illustrated several further starting points for future research. For example, only $n = 4$ publications considered multiple or intersectional discrimination (Fortuna et al., 2019; Hara et al., 2024; Izzaddien et al., 2024; Sachdeva et al., 2022). Belonging to multiple vulnerable social groups can increase exposure to hate speech (Obermaier & Schmuck, 2022), which may result in cumulative effects and future research could explore this intersectionality and added burden on victims in

greater detail. Future studies could furthermore investigate hate speech in low-resource languages as publications in our database primarily focused on English, Spanish or Tamil.

Limitations

Our study had several methodological limitations which must be acknowledged. First, we focused solely on publications written in the English language, which may have excluded relevant publications written in other languages. Second, we acknowledge that the screening process may involve subjective interpretation, leading to inconsistent inclusion/exclusion, as it was conducted by the first author. Decisions of the title/abstract screening were, however, validated by the second author to ensure consistency in the application of selection criteria. Similarly, the labeling of topics was conducted manually by the first author, which may also involve subjective interpretation. Third, we acknowledge that the manual coding of descriptive data, necessary to answer RQs 1–3, is an approach inherently dependent on researcher judgment, which may have introduced bias. Nevertheless, the manual coding provided detailed data which was unlikely to be captured by automated methods alone. Lastly, we did not provide an overview of the theoretical and methodological underpinnings of the analyzed publications. Future research that systematizes and integrates the scattered theoretical lenses would thus be a valuable conceptual contribution to the field.

Conclusion

The current study used a scoping review to answer what is known about the relationship between social media platforms and the spreading of hate speech towards sexual and gender minorities. The review focused on empirical evidence and identified $n = 145$ scholarly publications which were used as the basis for the analysis. Our findings call for future studies to investigate alternative platforms, visual and audiovisual content, more in-depth research on platform affordances and the application of a research focus beyond the development of detection models to understand how platform dynamics contribute to the spreading of hate speech targeting sexual and gender minorities.

Note

1. The final pre-registration includes refinements to the formulation of research questions to improve methodological coherence.

Author contributions

CRediT: **Anne Clausen:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Lena Frischlich:** Conceptualization, Methodology, Project administration, Supervision, Validation, Writing – review & editing; **Peter Mayer:** Conceptualization, Data curation, Methodology, Project administration, Resources, Software, Supervision, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Anne Clausen is a PhD Fellow with the Digital Democracy Centre at the University of Southern Denmark. Her research focuses on investigating the spreading of intolerance on social media platforms [email: aclausen@sam.sdu.dk].

Lena Frischlich is an associate professor with the Digital Democracy Centre at the University of Southern Denmark. Her research focuses on the interplay between the digital realm, human cognition, and behavior. She is particularly interested in the staging and effects of manipulative online communication, such as propaganda, disinformation, or hate speech, and how to strengthen democracy in the digital realm [email: lefr@sam.sdu.dk].

Peter Mayer is an associate professor in the Artificial Intelligence, Cybersecurity and Programming Languages section at the Department of Mathematics and Computer Science at the University of Southern Denmark. He is also a KASTEL Security Research Labs Fellow at the Karlsruhe Institute of Technology. The goal of his research is end-user viable information security & privacy solutions [email: mayer@imada.sdu.dk].

Data availability statement

The supplemental material can be found in: https://osf.io/sma4f/

References

Adami, E. (2017). Introducing multimodality. In García O., Flores N., & Spotti M. (Eds.), *The Oxford handbook of language and society* (pp. 451–472). Oxford University Press.

Adekola, A. P. (2025). Digital wounds, lived realities: A synthesis of evidence on online hate and its impact on gender and sexual minorities in South Africa. *Journal of Ecohumanism*, *4*(4), 1238–1253. https://doi.org/10.62754/joe.v4i4.6857

ADL Center for Technology & Society. (2023). *Online hate and harassment: The American experience 2023*. https://www.adl.org/sites/default/files/pdfs/2023-12/Online-Hate-and-Harassmen-2023_0_0.pdf

ADL Center for Technology & Society. (2024). *Online hate and harassment: The American experience 2024*. https://www.adl.org/sites/default/files/documents/2024-06/online-hate-and-harassment-the-american-experience-v2024.pdf

Alvisi, L., Tardelli, S., & Tesconi, M. (2025). Mapping the Italian telegram ecosystem: Communities, toxicity, and hate speech. arXiv. https://doi.org/10.48550/arXiv.2504.19594

Amnesty International Danmark. (2025). *Had skader – 22 ÅRS Retspraksis PÅ Straffelovens § 266 B*. https://amnesty.dk/wp-content/uploads/2025/01/Amnesty_HadskaderRapport2024_Digital_FINAL.pdf

Analyse & Tal, & Os & Data. (2025). *Angreb & had i den offentlige debat på Facebook*. https://www.tryghed.dk/-/media/files/pdf/publikationer/2025/2025-angreb-og-had-i-den-offentlige-debat.pdf

Arce-García, S., & Menéndez-Menéndez, M.-I. (2022). Inflaming public debate: A methodology to determine origin and characteristics of hate speech about sexual and gender diversity on Twitter. *El Profesional de la Información*, *32*(1). https://doi.org/10.3145/epi.2023.ene.06

Badaoui, A. (2024). The affordance of visibility in the social media videos of LGBT nonprofits in Lebanon: (In)visibilities, representativity and socio-political engagement. *Athens Journal of Mass Media and Communications*, *10*(2), 131–148. https://doi.org/10.30958/ajmmc.10-2-4

Bayer, J. B., Triệu, P., & Ellison, N. B. (2020). Social media elements, ecologies, and effects. *Annual Review of Psychology*, *71*(1), 471–497. https://doi.org/10.1146/annurev-psych-010419-050944

Ben-David, A., & Matamoros Fernández, A. (2016). Hate speech and covert discrimination on social media: Monitoring the Facebook pages of extreme-right political parties in Spain. *In International Journal of Communication*, *10*, 1167–1193. https://ijoc.org/index.php/ijoc/article/view/3697/1585

Bimber, B., & Gil de Zúñiga, H. (2020). The unedited public sphere. *New Media & Society*, *22*(4), 700–715. https://doi.org/10.1177/1461444819893980

Brody, E., Greenhalgh, S. P., & Sajjad, M. (2024). Free speech or free to hate?: Anti-LGBTQ+ discourses in LGBTQ+-affirming spaces on gab social. *Journal of Homosexuality*, *71*(8), 2030–2055. https://doi.org/10.1080/00918369.2023.2218959

Carr, C. T., & Hayes, R. A. (2015). Social media: Defining, developing, and divining. *Atlantic Journal of Communication*, *23*(1), 46–65. https://doi.org/10.1080/15456870.2015.972282

Chakravarthi, B. R., Priyadharshini, R., Ponnusamy, R., Kumaresan, P. K., Sampath, K., Thenmozhi, D., Thangasamy, S., Nallathambi, R., & Mccrae, J. P. (2021). Dataset for identification of homophobia and transophobia in multilingual YouTube comments. *arXiv Preprint*. https://arxiv.org/pdf/2109.00227

covidence.org. (2025). https://www.covidence.org/

Díaz-Fernández, S., & García-Mingo, E. (2024). The bar of forocoches as a masculine online place: Affordances, masculinist digital practices and trolling. *New Media & Society*, *26*(9), 5336–5358. https://doi.org/10.1177/14614448221135631

Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication*, *22*(1), 35–52. https://doi.org/10.1111/jcc4.12180

Evans, W. (2024). Groomer discourse: A transgender-sensitive critical discourse analysis of queerphobic hate speech on twitter. *Honors Theses*, *966*. https://aquila.usm.edu/honors_theses/966

Fangen, K. (2020). Gendered images of us and them in anti-Islamic Facebook groups. *Politics, Religion & Ideology*, *21*(4), 451–468. https://doi.org/10.1080/21567689.2020.1851872

Fortuna, P., Rocha da Silva, J., Soler-Company, J., Wanner, L., & Nunes, S. (2019). A hierarchically-labeled Portuguese hate speech dataset. In S. T. Roberts, J. Tetreault, V. Prabhakaran, & Z. Waseem (Eds.), *Proceedings of the Third Workshop on Abusive Language Online* (pp. 94–104). Association for Computational Linguistics. https://doi.org/10.18653/v1/W19-3510

Frischlich, L. (2023). Hate and harm. In S. Paasch-Colberg, C. Strippel, & M. Emmer (Eds.), *Challenges and perspectives of hate speech analysis: An interdisciplinary anthology* (pp. 165–183). Digital Communication Research. https://doi.org/10.48541/dcr.v12.10.

Geschke, D., Klaßen, A., Quent, M., & Richter, C. (2019). *#HASS IM NETZ: DER SCHLEICHENDE ANGRIFF AUF UNSERE DEMOKRATIE – EINE BUNDESWEITE REPRÄSENTATIVE UNTERSUCHUNG*. https://www.idz-jena.de/fileadmin/user_upload/_Hass_im_Netz_-_Der_schleichende_Angriff.pdf

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure (No. arXiv:2203.05794). arXiv. https://doi.org/10.48550/arXiv.2203.05794

Hara, N. T., Siregar, M., Br. Perangin-angin, A., & Anshary, E. P. (2024). A cross-cultural examination of hate speech targeting transgender celebrities across Indonesian and international social media platform. *LingPoet: Journal of Linguistics and Literary Research*, *5*(3), 184–193.

Hee, M. S., Sharma, S., Cao, R., Nandi, P., Nakov, P., Chakraborty, T., & Lee, R. K.-W. (2024). Recent advances in online hate speech moderation: Multimodality and the role of large models. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Findings of the association for computational linguistics: EMNLP 2024* (pp. 4407–4419). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.254

Hickey, D., Fessler, D. M. T., Lerman, K., & Burghardt, K. (2025). X under Musk's leadership: Substantial hate and no reduction in inauthentic activity. *PLoS One*, *20*(2), e0313293. https://doi.org/10.1371/journal.pone.0313293

Izzaddien, Y. F., Sarno, R., Haryono, A. T., Septiyanto, A. F., Sunaryono, D., & Handoyo Winarso, R. A. (2024). Multilabel aspect-based sentiment analysis with diverse embedding techniques and

finetuned transformers for label detection. In *2024 Beyond Technology Summit on Informatics International Conference (BTS-I2C)* (pp. 95–100). https://doi.org/10.1109/BTS-I2C63534.2024.10941997

Kakavand, A. E. (2024). Far-right social media communication in the light of technology affordances: A systematic literature review. *Annals of the International Communication Association*, *48*(1), 37–56. https://doi.org/10.1080/23808985.2023.2280824

Kemp, S. (2025). *Digital 2025: Global overview report* (p. 372). DATAREPORTAL. https://datareportal.com/reports/digital-2025-global-overview-report

Kettrey, H. H., Nwajei, M., Quinn, S., Leslie, M., Paradise, E., & Wishon, D. (2024). Gendered affordances of digital technology in mitigating the perceived risk of dating app matches perpetrating sexual assault or "making stories" of assault. *Social Media + Society*, *10*(3), 20563051241269296. https://doi.org/10.1177/20563051241269296

Klutse, E. K., Nuamah-Amoabeng, S., Lyu, H., & Luo, J. (2023). Dismantling hate: Understanding hate speech trends against NBA athletes. In *14161 LNCS* (pp. 74–84). https://doi.org/10.1007/978-3-031-43129-6_8

Kumar, G., Singh, J. P., & Kumar, A. (2021). A deep multi-modal neural network for the identification of hate speech from social media. In D. Dennehy, A. Griva, N. Pouloudi, Y. K. Dwivedi, I. Pappas, & M. Mäntymäki (Eds.), *Responsible AI and analytics for an ethical and inclusive digitized society* (pp. 670–680). Springer International Publishing.

Kumar, R., Ratan, S., Singh, S., Nandi, E., Devi, L., Bhagat, A., Dawer, Y., Lahiri, B., & Bansal, A. (2024). A multilingual, multimodal dataset of aggression and bias: The ComMA dataset. *Language Resources and Evaluation*, *58*(2), 757–837. https://doi.org/10.1007/s10579-023-09696-7

Leets, L. (2002). Experiencing hate speech: Perceptions and responses to anti-Semitism and anti-gay speech. *Journal of Social Issues*, *58*(2), 341–361. https://doi.org/10.1111/1540-4560.00264

Luke, H. (2021). *Hate crime report: Supporting LGBT+ victims of hate crime*. https://galop.org.uk/wp-content/uploads/2021/06/Galop-Hate-Crime-Report-2021-1.pdf

Maarouf, A., Pröllochs, N., & Feuerriegel, S. (2024). The virality of hate speech on social media. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1–22. https://doi.org/10.1145/3641025

Menini, S., Aprosio, A. P., & Tonelli, S. (2020). A multimodal dataset of images and text to study abusive language. In *Proceedings of the Seventh Italian Conference on Computational Linguistics CLiC-It 2020*. https://doi.org/10.4000/books.aaccademia.8725

Mkhize, S., Nunlall, R., & Gopal, N. (2020). An examination of social media as a platform for cyber-violence against the LGBT+ population. *Agenda*, *34*(1), 23–33. https://doi.org/10.1080/10130950.2019.1704485

Murtfeldt, R., Alterman, N., Kahveci, I., & West, J. D. (2024). RIP Twitter API: A eulogy to its vast research contributions. *arXiv Preprint arXiv:2404.07340*.

Nguyen, T. T., Yue, X., Mane, H., Seelman, K., Mullaputi, P. S. P., Dennard, E., Alibilli, A. S., Merchant, J. S., Criss, S., Hswen, Y., & Nguyen, Q. C. (2025). Decoding digital discourse through multimodal text and image machine learning models to classify sentiment and detect hate speech in race- and lesbian, Gay, bisexual, transgender, queer, intersex, and asexual community–related posts on social media: Quantitative study. *Journal of Medical Internet Research*, *27*, e72822. https://doi.org/10.2196/72822

Obermaier, M., Schmid, U. K., & Rieger, D. (2023). Too civil to care? How online hate speech against different social groups affects bystander intervention. *European Journal of Criminology*, *20*(3), 817–833. https://doi.org/10.1177/14773708231156328

Obermaier, M., & Schmuck, D. (2022). Youths as targets: Factors of online hate speech victimization among adolescents and young adults. *Journal of Computer-Mediated Communication*, *27*(4), zmac012. https://doi.org/10.1093/jcmc/zmac012

Oehmer-Pedrazzi, F., & Pedrazzi, S. (2024). An image hurts more than 1000 words? *Communications-European Journal of Communication Research*, *49*(3), 421–443. https://doi.org/10.1515/commun-2023-0117

Oliveira, E., Oliveira, L., & Baldi, V. (2024). Analysis of cyberaggression in social networks involving students and university environments. In *Advances in design and digital communication*

IV. DIGICOM 2023. Springer series in design and innovation (Vol. 35, pp. 287–300). Springer Nature. https://doi.org/10.1007/978-3-031-47281-7_23

Paterson, J., Walters, M., Brown, R., & Fearn, H. (2018). *Sussex hate crime project: Final report*. University of Sussex. https://research.tees.ac.uk/ws/portalfiles/portal/4175549/621941.pdf

pypi.org. (2023). *Scholarly 1.7.11*. https://pypi.org/project/scholarly/

Rieger, D., Kümpel, A. S., Wich, M., Kiening, T., & Groh, G. (2021). Assessing the extent and types of hate speech in fringe communities: A case study of alt-right communities on 8chan, 4chan, and Reddit. *Social Media + Society*, *7*(4), 1–14. https://doi.org/10.1177/20563051211052906

Ronzhyn, A., Cardenal, A. S., & Rubio, A. B. (2023). Defining affordances in social media research: A literature review. *New Media & Society*, *25*(11), 3165–3188. https://doi.org/10.1177/14614448221135187

Rossini, P. (2020). Beyond incivility: Understanding patterns of uncivil and intolerant discourse in online political talk. *Communication Research*, *49*(3), 399–425. https://doi.org/10.1177/0093650220921314. (Original work published 2022).

Sachdeva, P., Barreto, R., Von Vacano, C., & Kennedy, C. (2022). Targeted identity group prediction in hate speech corpora. In K. Narang, A. Mostafazadeh Davani, L. Mathias, B. Vidgen, & Z. Talat (Eds.), *Proceedings of the Sixth Workshop on Online Abuse and Harms (WOAH)* (pp. 231–244). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.woah-1.22

Sánchez-Sánchez, A. M., Ruiz-Muñoz, D., & Sánchez-Sánchez, F. J. (2024). Mapping homophobia and transphobia on social media. *Sexuality Research and Social Policy*, *21*(1), 210–226. https://doi.org/10.1007/s13178-023-00879-z

Santos Fernández, F. J. (2024). Online homophobia: Hate speech and conspiracy theories towards LGBTQI+ people on Twitter in Spain. *Culture e Studi Del Sociale-CuSSoc*, *9*(1), 39–56.

Schmitz, M., Muric, G., & Burghardt, K. (2022). Quantifying how hateful communities radicalize online users. In *2022 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 139–146). IEEE.

Schulze, H., Greipl, S., Hohner, J., & Rieger, D. (2024). Social media and radicalization: An affordance approach for cross-platform comparison. *Medien & Kommunikationswissenschaft*, *72*(2), 187–212. https://doi.org/10.5771/1615-634X-2024-2-187

Schwartz, B., & Neff, G. (2019). The gendered affordances of craigslist "new-in-town girls wanted" ads. *New Media & Society*, *21*(11–12), 2404–2421. https://doi.org/10.1177/1461444819849897

Semenzin, S., & Bainotti, L. (2020). The use of telegram for non-consensual dissemination of intimate images: Gendered affordances and the construction of masculinities. *Social Media + Society*, *6*(4), 2056305120984453. https://doi.org/10.1177/2056305120984453

Shah, S. B., Shiwakoti, S., Chaudhary, M., & Wang, H. (2024). MemeCLIP: Leveraging CLIP representations for multimodal meme classification. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 17320–17332). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.959

Siegel, A. A. (2020). Online hate speech. In N. Persily & J. A. Tucker (Eds.), *Social media and democracy* (pp. 56–88). Cambridge University Press.

Sívori, H., & Zilli, B. (2022). COVID-19, homophobia and the bolsonarista vernacular: Hate speech on Brazilian social media. *APRIA Journal*, *4*(4), 35–49. https://doi.org/10.37198/APRIA.04.04.a3

Stefanita, O., & Buf, D. (2021). Hate speech in social media and its effects on the LGBT community: A review of the current research. *Romanian Journal of Communication and Public Relations*, *23*(1), 47–55. https://doi.org/10.21018/rjcpr.2021.1.322

Strippel, C., Ziegele, M., Schindler, M., Laugwitz, L., Domahidi, E., Bormann, M., Reiners, L., Langmann, K., Frischlich, L., Naab, T. K., Rieger, D., Puschmann, C., Schemer, C., & Ross, B. (2023). *Hate speech, incivility and related concepts of disrespectful language on the Internet: A scoping review*. ICA 2023, 73rd Annual Conference of the International Communication Association, Toronto, Canada. https://madoc.bib.uni-mannheim.de/65804/

Suler, J. (2004). The online disinhibition effect. *CyberPsychology & Behavior*, *7*(3), 321–326. https://doi.org/10.1089/1094931041291295

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

United Nations. (2019). *The UN strategy and plan of action on hate speech*. https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_EN.pdf

United Nations. (2023). *Countering "dark age of intolerance" starts by tackling hate speech*. https://news.un.org/en/story/2023/06/1137822

Unlu, A., Truong, S., Sawhney, N., Tammi, T., & Kotonen, T. (2025). From prejudice to marginalization: Tracing the forms of online hate speech targeting LGBTQ+ and Muslim communities. *New Media & Society*, *0*(0), 1–32. https://doi.org/10.1177/14614448241312900

Venegas Carrasco, C., Alarcón Hernández, P., & Maldonado Delgado, P. (2024). Attitudes towards LGBTIQ+ visibility: A qualitative analysis of Facebook comments in Chilean press. *Psykhe*, *33*(2), 1–17.

Vergani, M., & Navarro, C. (2023). Hate crime reporting: The relationship between types of barriers and perceived severity. *European Journal on Criminal Policy and Research*, *29*(1), 111–126. https://doi.org/10.1007/s10610-021-09488-1

Vergani, M., Perry, B., Freilich, J., Chermak, S., Scrivens, R., Link, R., Kleinsman, D., Betts, J., & Iqbal, M. (2024). Mapping the scientific knowledge and approaches to defining and measuring hate crime, hate speech, and hate incidents: A systematic review. *Campbell Systematic Reviews*, *20*(2), e1397. https://doi.org/10.1002/cl2.1397

Wang, H., Yang, T. R., Naseem, U., & Lee, R. K.-W. (2024). Multihateclip: A multilingual benchmark dataset for hateful video detection on YouTube and bilibili. In *Proceedings of the 32nd ACM International Conference on Multimedia* (pp. 7493–7502). https://doi.org/10.1145/3664647.3681521

Weise, J., Courtney, S., & Strunk, K. K. (2021). *"I didn't think I'd be supported": LGBTQ students' non-reporting of bias incidents at southeastern colleges and universities*. http://doi.org/10.31235/osf.io/cd9qh

Williams, M. L., & Tregidga, J. (2014). Hate crime victimization in Wales: Psychological and physical impacts across seven hate crime victim types. *The British Journal of Criminology*, *54*(5), 946–967. https://doi.org/10.1093/bjc/azu043

Yang, C., Zhu, F., Liu, G., Han, J., & Hu, S. (2022). Multimodal hate speech detection via cross-domain knowledge transfer. In *Proceedings of the 30th ACM International Conference on Multimedia* (pp. 4505–4514). https://doi.org/10.1145/3503161.3548255.

Zannettou, S., Caulfield, T., Blackburn, J., De Cristofaro, E., Sirivianos, M., Stringhini, G., & Suarez-Tangil, G. (2018). On the origins of memes by means of fringe web communities. In *ACM Internet Measurement Conference*. http://arxiv.org/abs/1805.12512

Zeng, J., & Schäfer, M. S. (2021). Conceptualizing 'dark platforms'. COVID-19-related conspiracy theories on 8kun and gab. *Digital Journalism*, *9*(9), 1321–1343. https://doi.org/10.1080/21670811.2021.1938165