Project COME-ON

Analysing and combating online gender-based hate speech from an intersectional approach

Collection of working papers

Grupo de investigación en Género, Identidad y Diversidad
Universidad de Barcelona (GENI)

Proyecto COME-ON: Analizar y combatir el discurso de odio en Internet y la discriminación de género desde una perspectiva interseccional

2024

Executive summary

The COME-ON project has employed an interdisciplinary methodology to examine online hate speech based on gender from various perspectives. This included analysing the legal framework, jurisprudence, and public policies, conducting in-depth interviews with legal experts, professionals, and victimized individuals, digital ethnography based on case studies on Twitter, and a survey of civil society organizations. This report collects five working papers elaborated during the research.

Findings shed light on how digital violence is experienced, perceived and contested in contemporary Spain. In this regard, various patterns characterising individualized hate attacks towards feminist women (LGBT+, racialised, etc.) were identified, often initiated by profiles with many followers on social media. These attacks are perpetrated by multiple individuals simultaneously and tend to be continuous over time, demonstrating the magnitude and persistence of the violence exerted. It is important to highlight that women are not only attacked for expressing their opinions on social media but also for doing so in any public medium that transcends a safe circle of like-minded people. Victims report a variety of impacts affecting their emotional and physical well-being, as well as their sense of security and behaviour in digital spaces. Real and bodily harm caused by digital violence is highlighted, with long-term consequences such as constant surveillance, behavioural changes, and self-censorship in public debates. Nevertheless, disconnecting from social media is not a viable option, especially for professional profiles. At the same time, the legal framework to address this violence is fragmented, and victims face multiple barriers to reporting it to the police or other institutions.

Professionals from both public and non-governmental sectors suffer from a significant lack of training in on safeguarding individuals from online attacks and responding to such incidents. Many rely on informal knowledge from younger colleagues. The binary view of online vs. offline violence downplays the severity of online violence, limiting support. Addressing this issue requires a multi-faceted response involving legal measures, education, digital protection, collaboration with social media platforms, and tailored support for victims, ensuring professionals are well-equipped to provide both structural solutions and empathetic care. A comprehensive response is required, including law enforcement and legal measures, educational and awareness initiatives, improvements in digital protection, and diverse support for victims. It is essential for professionals to be adequately trained to address this form of violence empathetically and effectively.

1. Introduction

Digital platforms offer a myriad of opportunities for personal, social, and cultural expression. However, the global, immediate, and participatory nature of these interactions has also created a fertile ground for exclusionary, intolerant, and extremist discourses. Online hate speech and violence can target individuals anytime and anywhere via mobile internet access (European Women's Lobby, 2017). Despite extensive academic and media attention, discussions about online hate speech often overlook the gender dimension, framing the issue as gender-neutral (Jane, 2014). This lack of gender focus distorts the understanding of the phenomenon, neglecting that most gender-based hate speech targets women, girls (UN Broadband Commission, 2015), and LGBT people (Citron, 2009). In 2017, UN Special Rapporteurs on Violence against Women and Freedom of Expression highlighted the prevalence of gender-based online abuse, including threats of rape, violence, and gross invasions of privacy. Such abuse not only harasses but also aims to silence and intimidate women, particularly those who are public figures or activists (Jane, 2017). Online hate speech has significant offline repercussions, leading to social, psychological, professional, financial, and political consequences for the victims (Citron, 2014; Jane, 2017).

Research on online hate speech often neglects the specificities of gender-based hate speech, either overlooking it or focusing on racist and xenophobic hate speech, with gender considered an intersecting element at best. There is also a lack of studies on the intersectional experiences of women and LGBT individuals online (Khosravinik and Esposito, 2018). Despite the EU's recognition of intersectionality in combating hate speech (ECRI, 2015), there is insufficient research to support effective legal and policy measures. Further, the EU Framework Decision on combating racism and xenophobia defines hate speech inclusively, but practical application of these laws in relation to gender-based hate speech is underexplored, especially in Spain.

To address this gap, this interdisciplinary research of the COME-ON project has explored "gender-based hate speech" in Spain, focusing on online hate speech against individuals based on gender, in intersection with other factors such as gender identity, sexual orientation, or racialisation. Based on this research, an important aim has been to develop tools and strategies to combat online hate speech and enhance understanding of this critical issue.

The core of this report consists of the collection of five working papers elaborated by the research and work teams during the course of the project, including a review of the state of the art; a review and analysis of the legal and policy framework in relation to gender-based online hate speech in Spain; the analysis of the experiences, impacts and responses of women targeted by gender-based online hate speech; the analysis of professionals' perceptions of online violence and their preparedness to provide support,

and the analysis of a survey on civil society perceptions on online hate speech and digital violence.

Objectives of the project and its research

Given the lack of scientific evidence, particularly from a qualitative approach, this project has aimed to analyse gender-based online hate speech from an interdisciplinary and intersectional perspective. The goal has been to gain a deeper understanding of this phenomenon, generate new knowledge specific to Spain, and propose social and legal measures. The main objective of this project has been to examine how gender interacts with other categories of inequality in online hate speech, to understand the underlying mechanisms of gender-based hate speech, its impact on the diversity of public discourse, and the experiences of the victims. Specific objectives include:

- Legal framework analysis: Map and analyse the legal instruments related to gender-based online hate speech and collect jurisprudence in Spain. Assess whether legal frameworks adequately recognize the intersectionality of individuals' experiences. Investigate whether more general laws on offensive speech and/or online harassment are being used to prosecute gender-based hate speech.
- Intersectional analysis: Investigate how gender intersects with other categories such as sexual orientation, gender identity, and racialisation in online hate speech.
- Victims' experiences: Explore the lived experiences of targets of online gender-based hate speech, focusing on identity constructions, the consequences for their lives, their responses, and how intersecting grounds of discrimination influence these experiences.
- Legislation effectiveness: Evaluate whether current hate speech legislation is effective in addressing gender-based online hate speech and in facilitating victim restoration, based on legal and social research findings.
- Policy influence: Influence public policies and responses regarding gender-based online hate speech through evidence-based recommendations.

A note on the methodology

The research has employed a mixed-methods approach to investigate the specific characteristics of gender-based online hate speech, considering intersecting identities and vulnerabilities. This has included the following elements:

- Desk research: State-of-the-art review and review of jurisprudence.
- Exploratory survey with civil society organizations: This exploratory survey aimed to analyse civil society perspectives and perceptions of digital violence and online hate speech, including perceptions of the legal framework and support offered by

the organizations. The survey collected responses from 44 NGOs active in the fields of feminism and gender-based violence, LGBT+ rights, anti-racism or other related fields.

- Digital ethnography: This included 26 case studies on Twitter. These cases included persons who auto-define as women and feminists, and who actively and openly debate feminist and and/or other political issues such as anti-racism or trans rights on social media, and who in the past few years have been targeted by gendered online violence. Information has only been collected from open public profiles and all cases have been anonymized through designated codes. No literal quotes have been used from these case studies in the reports, to avoid possible identification through a search engine. For the full sample, please see Annex I.

- Interviews with professionals and legal experts: This includes 20 interviews with legal experts and professionals, representing a diverse range of professionals. Among the interviewees were representatives from LGBT+ organizations and feminist associations, lawyers and legal experts specialised on hate crimes and gender-based violence, a prosecutor, public institutions at regional and local levels, police officers, local public services addressing gender-based violence, a sociologist affiliated with a feminist association, and a psychologist specialising on gender-based violence. For the full sample, please see Annex I.

- In-depth interviews with targets of online violence. This included interviews with 10 women, with a focus on feminist, LGBT+ and antiracist activists who are active on social media and who been targeted by online violence. For the full sample, please see Annex I.

2. Previous debates on online hostility: From flaming to hate speech

This section aims to examine the ongoing discussions surrounding gender-based online hate speech across various academic disciplines, including law, criminology, sociology, media, and communication studies. Before delving into these debates, however, it is crucial to outline broader discussions regarding online hostility and hate speech to clarify the diverse conceptualisations within this field of study. Precisely one of the primary challenges in conducting a literature review on online gender-based hate speech is navigating the multitude of concepts used to address this phenomenon.

Flaming, trolling and other concepts

In recent decades, scholars across various fields have explored online hostility through terms like 'cyber-bullying', 'cyber-stalking', 'cyber-violence', 'trolling', and 'flaming' (Jane, 2015). Initially, early digital researchers coined the term 'flaming' to describe aggressive and profanity-laden interactions in online communities, stemming from popular discourse (O'Sullivan and Flanagin, 2003). However, digital anthropologist Patricia G. Lange (2006) has offered significant insights into this concept. She notes that scholars often confine 'flaming' to online behavior, inadvertently reinforcing the perceived distinction between online and offline hostility. This approach may overlook potential similarities between the two realms.

Although the concept of flaming still sees some usage, terms like "trolling" or "hating" have gained prominence, particularly within social media contexts (McCosker, 2014:5). While trolling may encompass similar behaviors as flaming, it often involves intentional provocation through the posting of inflammatory or off-topic content in online discussions to evoke reactions. However, as highlighted by digital ethnographer Whitney Phillips (2015), the definition of trolling has become increasingly ambiguous, serving as a broad, catch-all term on the internet. Phillips suggests reserving the term specifically for certain subcultural trolling communities, such as those found on web forums like 4chan, rather than applying it broadly to the phenomenon of online hostility.

Here, it's essential to introduce the concept of 'e-bile,' as coined by Jane (2015: 66), aiming to encompass various denunciatory speech acts under one umbrella term, highlighting their shared characteristics and signaling the need for a comprehensive investigation. This underscores the importance of terminology that acknowledges the complexity and fluidity of the analyzed acts and the evolving nature of the practices and technologies involved (Jane 2017b: 46). However, as noted by KhosraviNik and Esposito (2018: 51), there's a suggestion to move away from terms like flaming and trolling towards alternatives such as cyberhate and "hate speech" (Citron and Norton 2011) or "e-bile" (). This shift aims to foster a more thoughtful approach to the violent and exclusionary aspects of online hostility, rather than trivializing the issue by focusing solely on its playful

and communal dimensions. This is crucial because, on a broader analytical level, the fixation on methodological framing, coupled with excitement over new communication affordances for both users and researchers, has often diverted academic discourse from the significant social, ethical, and political dimensions of this phenomenon (Khosravinik and Esposito 2018: 51).

As emphasized by KhosraviNik and Esposito (2018), concepts such as flaming, trolling, or e-bile fail to adequately address a crucial aspect: the grounds of oppression. Hate speech, as defined in EU law, involves the public incitement to violence or hatred based on specific characteristics, such as race, color, religion, descent, and national or ethnic origin. Although the EU Framework Decision on combating racism and xenophobia focuses solely on racist and xenophobic speech, it's widely agreed that the list should be interpreted as open-ended, with most Member States extending their national laws to include other grounds like sexual orientation, gender identity, and disability. The European Commission against Racism and Intolerance (ECRI) explicitly mentions sex, gender, and gender identity in its General Policy Recommendation No. 15 on combating hate speech, also addressing intersectionality. They acknowledge the significant problem of hate speech targeting women due to their sex, gender, and/or gender identity, particularly when combined with other characteristics (ECRI, 2016). What is lacking in concepts like flaming, trolling, and e-bile is the specific basis for the expressed hatred, whether it's racialization, religion, age, disability, gender identity, sexual orientation, or, in the context of this research, gender. While online hate speech often overlaps with flaming, trolling, and e-bile by targeting similar vulnerabilities or oppressed groups, the latter concepts can encompass any form of online person-to-person hostility, irrespective of structural factors. However, for discourse to be classified as hate speech, the presence of structural oppressive factors or vulnerabilities in the expressed hatred is crucial. Highlighting these elements is essential for incorporating significant structural and interactional factors into the analysis.

Online hate speech

Hostility and hate speech represent intricate social, cultural, and psychological phenomena (KhosraviNik and Esposito, 2018). According to Waldron (2012), hate speech communicates two simultaneous messages: one aimed at the targeted group, seeking to dehumanize and degrade them by attacking their fundamental human dignity. The other message is directed towards individuals with similar viewpoints, reinforcing a sense of group identity in opposition to and threatened by "the other," thereby uniting like-minded individuals. In this sense, hate speech both divides and unites simultaneously (Gagliardone et al., 2015). In our contemporary era, these discourses are amplified by the newfound capabilities of the participatory web (KhosraviNik and Esposito, 2018). Around 2006, the transition from the static web to the interactive and user-driven sphere known as web 2.0 gained momentum (Jane, 2015: 81). While social media has empowered ordinary users to engage in the creation and dissemination of online content, the

substantial increase in user-generated texts during this period marked a significant juncture in the circulation of online hate speech (Ibid.). The global, instantaneous, and participatory nature of social media communication has turned the online sphere into a fertile ground for the expression and propagation of various exclusionary, intolerant, and extremist discourses, practices, and beliefs (Kopytowska and Chilton, 2018). As scholars like Banks (2010: 234) have observed, the Internet can be characterized as "the new frontier for spreading hate".

Regarding this matter, a prevalent scholarly argument, particularly from the realms of Social Psychology, Criminology, and Media and Communication Studies, asserts that social media features serve as amplifiers, both in terms of quantity and intensity of interactions (KhosraviNik and Esposito, 2018). Certain digital characteristics and functionalities have been singled out for their role in perpetuating digital hate discourses. Firstly, anonymity, or the perception of anonymity, is widely recognized as a factor contributing to the escalation of online hostility. It is believed to empower individuals to disregard social norms and conventions as they do not perceive the threat of repercussions or accountability for their actions (Wallace, 2016). Online anonymity is often associated with the notion that interactions online are inherently different from face-to-face interactions, viewed as less significant, more liberating, and differently valued (KhosraviNik and Esposito, 2018). This perception is partly attributed to physical separation, which is also considered a factor exacerbating aggressive behavior online (Lewandowska-Tomaszczyk, 2017). In this context, as argued by Weisband and Atwater (1999), the absence of a face-to-face context and the consequent failure to acknowledge the humanity of the recipient may contribute to increased aggression online, as well as the absence of non-verbal cues (Epley and Kruger, 2005).

As previously mentioned, the bulk of research on online hate speech originates from legal studies, which either engage in debates concerning freedom of expression online (see, for example, Waldron, 2012; Heinze, 2016) or propose regulatory solutions for online hate speech (see, for example, Citron and Norton, 2011). Additionally, contributions come from criminology (e.g., Lewis, 2016), Communication and Media studies (e.g., Jane, 2014, 2015, 2017; Megarry, 2014; Khosravinik and Esposito, 2018), Information Technologies (e.g., Burnap and Williams, 2016), and sociology, which examines the mechanisms surrounding racist online hate speech (e.g., Daniels, 2008, 2009) or conducts quantitative studies on online hate speech (e.g., Oksanen, 2014).

In the specific context of Spain, legal studies dominate the field of online hate speech research. Serra (2018a, 2018b) has authored reports on Spain's legal landscape concerning hate speech, while Gomez (2016) analyzed the incorporation of gender as a category following the 2015 reform of the penal code in relation to hate crimes and hate speech. However, qualitative and ethnographic research on online hate speech remains scarce overall, with social anthropology largely absent from the discourse. There are a few notable exceptions, such as the work of Jubany and Roiha (2018), who explored

young people's experiences of and attitudes towards racist and xenophobic online hate speech through a multi-sited qualitative study across five European countries, including Spain. They concluded that online hate speech is increasingly common and normalized among youth. Additionally, Pohjonen and Udupa (2017) conducted a comparative study of online practice related to hate speech in Ethiopia and India from an anthropological perspective. Through a critical anthropological lens, they argue that unpacking concepts like hate speech and tracing the implications of adopting hate speech as a regulatory value in different national or regional contexts is essential.

Gender-based online hate speech

Conceptualising gender-based online hate speech

When delving into contemporary studies on gender-based online hate speech, one of the primary issues that arises is the absence of consensus regarding how to conceptualize this phenomenon. As noted by KhosraviNik and Esposito (2018), despite misogynistic violence representing a significant social problem, both institutional and scholarly research have often overlooked online misogyny as a form of gender-based hate speech. Scholars like Megarry (2014) contend that social media is rooted in the tangible realities of women's everyday encounters with sexism. Hence, it's crucial to examine online abuse against women as an extension of offline gender dynamics (Lewis, 2016). Gendered insults are often dismissed as mere protests or robust critiques, rather than recognized as instances of sex discrimination or gender-based hatred (Weston-Scheuber, 2012).

The omission of gender from most institutional definitions of hate speech underscores the institutional reluctance to acknowledge gender as a triggering social factor for hate (Titley, 2012; KhosraviNik and Esposito, 2018). As KhosraviNik and Esposito (2018) highlight, several factors potentially contribute to this: on one hand, the widespread belief that gender equality has largely been achieved may lead to hesitancy in discussing sexist speech as hate speech; on the other hand, while hate speech targets "members of vulnerable minorities" with the aim of completely eradicating them from a community, the same argument doesn't apply to women (Lillian, 2007). However, as argued by Khosravinik and Esposito, these assumptions fail to acknowledge the various ways in which women are targeted by sexist and sexually violent speech within patriarchal frameworks, which constitute one of the "most complex and pervasive systems of oppression" (Lazar, 2007: 143). Hate speech targets women precisely because of their gender, indicating that this type of speech can and should be categorized as hate speech. Another concept that could be employed in this context is online misogyny (Jane, 2016; 2017).In terms of definitions from a legal point of view, in the case of Spain, Gomez (2016) has analysed the new Article 510 CP, particularly in relation to the introduction of "razones de género", discussing whether this new article "podría resultar aplicable a hipotéticos supuestos de incitación al odio machista propios del llamado discurso racista".

The technological advancements intersecting discourse and society significantly influence the prevalence, characteristics, and mechanisms of new digital discursive practices involving symbolic violence, discrimination, and hate. However, any form of gendered violence online mirrors and amplifies the existing gendered power dynamics present prior to the advent of digital communication technologies (KhosraviNik and Esposito, 2018). As Lewis (2016) points out, delineating online abuse as distinct from real-world contexts poses a challenge. Approaching it from a technologically deterministic standpoint suggests that aggressive communications primarily stem from the technology utilized and its capabilities. Lewis further notes that "if online abuse is to be seen as an extension of real-world behavior, then it might follow that those who are not misogynistic offline are unlikely to become so online simply because they are in an uninhibited environment, as a technologically determinist position might imply" (2016: 1465).

Gendered dimensions of digital interactions and online hostility

To contextualize and grasp the historical background of current research on the gendered dynamics of digital interactions and online hostility, this section briefly surveys some early studies in this domain.

Initially, research in computer-mediated communication operated under the assumption that Internet communication was anonymous (Herring, 2001). This anonymity, coupled with the absence of visual and auditory cues, led to the belief that gender would be irrelevant to online interactions (Balka, 1993; Graddol and Swann, 1989). However, as the presence of women increased online in the early 1990s, studies exploring gender dynamics in online communication began emerging (Herring, 2001). These studies challenged claims of gender-neutral equality on the web. For instance, Selfe and Meyer (1991) found that men and participants of higher status offline dominated online interactions, even under conditions of anonymity. Additionally, scholars in the early 1990s observed more aggressive behavior by men in online discussions, often directed explicitly at female participants (Herring, 1993).

Susan Herring, from the field of Information Science, has extensively examined gendered dimensions in online interactions (2004, 2010; 2011), feminist responses to 'trolling' (2002), and gender differences in online posting behavior (Herring, 1996a; 1996b). She was among the first to recognize the gendered nature of much online abuse. For example, in a 1993 study, Herring found that women were more likely than men to respond aversively to aggression in online interaction, often by withdrawing from discussions. Other early studies revealed that women initiated fewer topics of discussion and received fewer public responses than men online (Herring, 1993; Herring, Johnson & DiBenedetto, 1995).

In contemporary studies, the analysis of gendered power dynamics on social media focuses on how the absence of traditional gatekeepers allows historically marginalized

groups, such as women, to potentially exert greater social and political influence (Megarry, 2014). However, despite women outnumbering men on many social media platforms, this doesn't necessarily translate to greater influence in online public spheres (Megarry, 2014). For instance, Aussenhofer and Maireder (2013) found significant underrepresentation of women in the Austrian political twittersphere, with Twitter users addressing men more frequently than women, irrespective of gender.

Moreover, considering the idealised anonymity and pursuit of equity in the early days of the internet, it's noteworthy that contemporary social media platforms encourage users to utilize their real names, photos, and define their gender. This model of 'radical transparency' (Kirkpatrick, 2011) promotes a single online identity. Consequently, the rise of social media has blurred the boundaries between our online and offline lives, making our physical bodies increasingly central to our online personas (Megarry, 2014).

Gender-based online hate speech: Experiences, consequences, contestations

In recent years, numerous international organizations have conducted studies on gender-based online violence, with hate speech emerging as a prominent dimension. Data from these studies indicate that women and girls are the primary targets of online violence (European Women's Lobby, 2017). For instance, Amnesty International (2018) revealed that 23% of women aged 18 to 55 surveyed across nine countries reported experiencing online abuse or harassment at least once, with 58% of these incidents involving racism, sexism, homophobia, or transphobia. These findings underscore how digital platforms serve as arenas for gender-based hate, often more openly than in offline environments (European Women's Lobby, 2017; UN Broadband Commission, 2015). However, studies from non-governmental organizations (NGOs) tend to approach the issue of gender-based violence online broadly, overlooking the specificity of hate speech. While hate speech is undoubtedly a crucial aspect of gender-based violence online, it requires focused attention and analytical frameworks tailored to this particular phenomenon. Studies conducted by international organizations often provide quantitative data to illustrate the prevalence of the issue and mobilize political action but lack in-depth analysis and theoretical frameworks for a broader understanding.

Regarding gender-based online hate speech, scholars have increasingly shown interest over the past decade, predominantly within the disciplines of law or criminology (Citron, 2009; Lewis et al., 2016), media and communication studies (Jane, 2014a, 2014b, 2015, 2017; Megarry, 2014), and psychology (Sills et al., 2016). Reviewing these studies, there is consensus on the targeting of women with a public presence, including journalists, politicians (Jane, 2017), academics (Beard, 2013), gamers (Burgess et al., 2017; Jane, 2016), and women engaged in feminist online discourse (Lewis, 2016). The content often contains sexually explicit and misogynistic elements, along with homophobia (Jane, 2014b), using derogatory stereotypes of femininity or insults targeting women's physical appearance. The range of insults spans from name-calling to threats of sexualized violence, often accompanied by pornographic images and language aimed at insulting

and degrading women (Megarry, 2014). As Megarry (2014) concludes based on a study of the Twitter hashtag #MenCallMeThings, gender-based hate speech perpetuates the notion that the internet is a male-dominated space where women's presence is tolerated solely based on their sexual appeal to men.

Lewis et al. (2016), drawing from a mixed-methods study involving feminists targeted by online abuse in the UK, emphasize the subjective nature of the experience of abuse. Their findings reveal that the majority of women who encountered online abuse deemed it 'significant,' with only a small fraction claiming to be 'not bothered.' Some responses indicated a tendency to minimize or normalize online abuse, a phenomenon also observed by Jubany and Roiha (2018) in the attitudes of young people towards online hate speech. Lewis (2016) notes that this normalization occurred even in response to death and rape threats, suggesting that women may perceive a 'simple' abusive message as relatively mild compared to more severe threats of physical and sexual violence. However, while normalization may serve as a coping strategy, it raises concerns about the long-term harm of considering such threats as 'normal.' The cumulative effect of routine abusive encounters can be profoundly impactful. Similarly, in a study by Sills et al. (2016) interviewing young people about their exposure and responses to 'rape culture' on social media, participants expressed a sense of 'living within a matrix of sexism,' where sexism, misogyny, and elements of rape culture are normalized aspects of everyday life.

These findings underscore the online and offline repercussions of gender-based online hate speech, with online abuse having enduring detrimental effects on the well-being of targeted women. Victims often experience feelings of anxiety, sadness, vulnerability, and terror, impacting them socially, psychologically, professionally, financially, and politically (Jane, 2017a). Consequently, women may alter their online behaviour, including the places they visit, the debates they engage in, and even the personas they adopt, in attempts to avoid harassment (Jane, 2017b). The victimization of women on social media serves to exclude certain voices from the internet, potentially jeopardising women's career prospects and economic livelihoods. Even when women opt to go offline, the effects of harassment persist, affecting them psychologically and professionally. Moreover, the use of pseudonyms may hinder women's professional networking opportunities and diminish the legitimacy of their online contributions (Citron, 2009b).

In the GamerGate controversy of 2014, examined by both Burgess et al. (2017) and Jane (2016), feminist gamer Anita Sarkeesian faced severe harassment and abuse after speaking out against misogyny in video games. She received a barrage of threats including death, rape, and bomb threats, as well as a threat to disrupt a speaking event with a mass shooting, ultimately leading her to cancel the engagement. As Burgess (2017) concludes, this case exemplifies how advocating for equal respect for women in male-dominated spaces can provoke conflict and aggression from certain community

members. While feminist gamers endure alarmingly high levels of misogyny, the GamerGate incident also reflects the broader backlash feminists face in online environments. Megarry (2014) underscores that achieving equity online necessitates not only occupying a space but also influencing it and speaking without fear of harassment. The ability to speak without fear is what is truly at stake for targets of online hate speech.

However, as Sassen noted in 2002, while digital spaces may replicate masculine cultures and power dynamics, they also afford women opportunities for new forms of resistance. Jane (2017b) provides evidence of increasing visibility and assertiveness among women pushing back against online misogyny, manifesting in heightened feminist rhetoric and greater individual and collective feminist activism. This activism often includes "feminist digilante responses" to gendered cyberhate, commonly known as "naming and shaming." These responses range from drawing attention to abuse without identifying perpetrators to exposing their online identities and publicly outing them to family, friends, or employers (Jane 2017b). Additionally, as Lewis (2016) points out, while targets of gender-based hate speech may view it as an attempt to discredit their online presence, it can paradoxically strengthen their resolve for political engagement. Thus, while online abuse can silence some women at certain times, it can also galvanize participation in civic life (Lewis 2016), complicating the narrative of online abuse as solely silencing women.

Intersectional perspectives

Intersectionality has received limited scholarly attention in the realm of hate crimes and hate speech, both legally and in social research. Burnap and Williams (2016) note that only recently has hate crime scholarship begun to explore how various identities intersect and are perceived by both victims and perpetrators. However, there are some quantitative studies shedding light on this issue. Citron (2014), a legal scholar, underscores that non-white women in the US encounter cyber harassment at a higher rate than any other group, with 53% reporting online harassment compared to 45% among white women. She concludes that being a woman elevates one's risk of cyber harassment, and this risk may be even higher for lesbian, transgender, or bisexual women and women of color. Outside academia, Amnesty International conducted a quantitative crowd-sourced study revealing that racialized women, particularly black women, are disproportionately targeted by online abuse on Twitter. Analysis of tweets directed at women journalists and politicians in the US and the UK in 2017 showed that black women were 84% more likely to receive abusive or problematic tweets than white women.

The intersectional perspective is most notably present in studies on Islamophobic hate crimes, where attention to the gendered aspect has increased in recent years. Islamophobia often manifests in physical aggression against veiled Muslim women due to their visible Muslim identities. Attacks on veiled Muslim women are considered particularly harmful as they strike at the core of the victim's identity. While it's unclear whether Muslim women are more targeted than Muslim men in terms of online hate

speech, some studies suggest that the majority of reported victims of online Islamophobia are women.

Although intersectionality is briefly acknowledged in Lewis's study (2016), where some respondents experienced online abuse intersecting with other forms of oppression, it is not further analyzed. Jane (2017b) clarifies that her focus on gendered cyberhate does not negate or diminish online hate speech related to race, class, sexual orientation, or gender identity. Burnap and Williams (2016) recognize the utility of 'othering' language in classifying cyber hate based on religious beliefs and have attempted to extend this model to develop a machine classifier for online hate speech based on intersectionality.

Zooming in on Spain

The vast majority of scholarly research on this topic, as identified through literature reviews, originates from Anglophone countries such as Australia, the UK, and the United States. In the Spanish context, however, there has been a recent emergence of legal studies explicitly focusing on issues like 'gender hate speech' (Souto Galván, 2015), 'sexist hate speech' (Álvarez, 2019), or 'anti-gender hate speech' (Igareda, 2022), particularly in the context of hate crimes and hate speech legislation following the 2015 reform of the penal code (Gómez, 2016). In the realm of media and communication studies, researchers such as Arcila-Calderón et al. (2021) and Arce-García and Menéndez-Menéndez (2022) have conducted big data analyses to explore gender-related hate speech on Twitter. Additionally, Piñeiro-Otero and Martínez-Rolán (2021) have conducted a quantitative study examining misogyny in Twitter conversations about Spanish women with high profiles, while Willem, Platero, and Tortajada (2023) have analyzed anti-trans discourse on social media. Furthermore, recent scholarship has also focused on hate speech against LGBTQI+ individuals on social media platforms in Spain, including works by Rivera-Martín (2022) and Martínez-Valerio (2021).

However, for research extending beyond the legal domain or big data analyses, especially in the Spanish context, it's essential to delve into the broader subject of gender-based online violence. It is important to note that the literature on this topic goes beyond academic studies, encompassing grey literature generated by various institutions and organizations. Grey literature often addresses all forms of online gender-based violence, potentially overlooking the distinct nature of hate speech. Consequently, it can be challenging to exclusively focus on online gender-based hate speech in literature reviews, as aspects of this phenomenon are often studied alongside other forms of gendered online violence. Moreover, studies published by international organizations tend to prioritize quantitative data to highlight the prevalence of the issue and stimulate political action, often lacking in-depth theoretical frameworks for comprehensive analysis.

Prevalence and targets of online gender-based violence

Piñeiro-Otero and Martínez-Roldán (2021) shed light on the toxic environment faced by women in communication and politics on Twitter. Their study revealed that more than one in ten tweets directed at high-profile women contained insults or derogatory terms. Complementing these findings, a policy brief from the Spanish National Observatory of Technology and Society (Ontsi) reported on a 2019 survey by the Government Delegation against Gender Violence. This survey, which included approximately 10,000 women, addressed violence against women, including online sexual harassment. Results indicated that 7.4% of women aged 16 or older had received inappropriate or offensive insinuations via social media. Among those who experienced sexual harassment, 18.4% encountered it on social media. Furthermore, over 15% reported repeated harassment from the same individual, and nearly a quarter endured offensive or embarrassing comments online. The brief highlighted that younger women, particularly those aged 16 to 25, were more vulnerable to digital harassment, with over 25% experiencing inappropriate insinuations on social media.

The association Calala, in collaboration with Hybridas and Komons, produced a report in 2020 on online gender-based violence against activists in Spain. This report, which focuses on violence against activists, aligns with the scope of the present research. It found that activists with visible online information, such as their name, photo, or work, were disproportionately affected, particularly those representing non-normative identities. Twitter was cited as the primary platform for such violence, mentioned by 72.7% of respondents.

Igareda et al. (2019) provided additional insights through a survey conducted by the association DonesTech. Despite a relatively small sample size of 262 respondents, the study highlighted the prevalence of online gender-based violence in Spain, a topic that has been under-researched. Of the respondents who reported experiencing online violence (98.9%), 70.0% had been targeted by more than one type of online violence, with insults or discriminatory expressions based on gender being the most prevalent (54.6%). Participants with public relevance were more likely to receive such insults, followed by digital threats, harassment, and defamation. The study also revealed variations in the frequency and duration of online violence, with some individuals experiencing it daily over extended periods.

These findings underscore the pervasive nature of online gender-based violence, particularly against women with public profiles. The data highlights the urgent need for comprehensive strategies to address and mitigate such violence, incorporating legal measures, digital protection, and enhanced support systems for victims.

Perpetrators and the manosphere

The Calala report (2020) brings attention to the presence of misogynistic communities on social media platforms in Spain, such as Facebook pages and groups characterized by a paramilitary culture. Igareda (2022: 101) observes that anti-gender activists are

particularly active online, leveraging the opportunities provided by new technologies. However, it's essential to note that not all perpetrators originate from organized movements. According to Calala's study on online gender-based violence against activists in Spain (2020), the majority of victims (76%) perceived the attackers as acting individually rather than collectively, with many attackers remaining anonymous (63.6%). Nonetheless, in around one-fifth of cases, the attackers were identified as belonging to specific communities, including political parties, abolitionist or transexcluding feminist sectors, groups advocating for shared custody, those denying gender-based violence, and organized far-right groups and parties. Similarly, Morena-Balaguer et al. (2021) delineate five types of perpetrator accounts involved in online violence against feminist activists in Spain, encompassing anonymous radical far-right profiles, bots, harassers, transphobic feminists, and transexcluding anonymous women.

'Triggers', content, discourses

Sobieraj (2018) proposes categorizing aggressors' strategies into three overlapping tactics: intimidation, shaming, and discrediting, arguing that their common purpose is to silence or diminish the impact of women in digital spaces. Delving into these strategies, "intimidation manifests in threats of physical violence, such as death and rape threats, as well as insinuations that the attacker knows the victim's whereabouts or vague but menacing messages suggesting that the target should remain silent or risk their family's safety" (2018: 5). In the Spanish context, various discursive strategies have been identified in cyber violence against activists and feminists: insults, sarcasm, imposition, intent to harm, sexual objectification, criminalization and defamation, general threats, sexual threats, as well as judgments and insults based on gender binarism, gender role assignments, sexualization and objectification of women, use of sexual terror, threats of sexual assault and rape, discrediting of opinions, and evaluations of physical appearance (Crosas and Medina-Bravo, 2018; Calala, 2020).

Regarding the "triggers" or topics that provoke violence against activists and feminists, previous research points to several subjects discussed by women that incite attacks, including general feminist messages, opinions on geopolitical tensions or politics, addressing or denouncing sexist violence, remarks about racism in Spain, criticisms of individuals, entities, or parties perceived as anti-women's rights (Calala, 2020), as well as tweets related to transfeminism (Morena-Balaguer et al., 2021). The latter is identified as a significant trigger for large-scale and coordinated attacks, originating from both the far-right and transphobic profiles.

Experiences and life-impacts

In Spain, Calala (2020) reveals a significant impact on both physical and mental health due to online violence, with frustration and helplessness being the most common emotions experienced by targets. More than half of the respondents reported engaging in self-censorship or reducing their online presence. According to the Igareda et al. (2019) survey, 76.1% of those targeted by online gender-based violence reported suffering

psychological harm, while over half experienced negative effects on their public participation, including self-censorship, withdrawal from social media, and abstention from public events. Additionally, respondents reported physical health issues such as headaches or nausea (26.3%), and economic repercussions such as income loss and difficulty finding work due to damage to their reputation or legal expenses (10.8%). International studies also highlight the enduring economic impact of online harassment, along with its obstruction of women's career advancement opportunities (Citron, 2009; Megarry, 2014).

Moreover, attacks on women politicians have been found to undermine the democratic process, as affected women may hesitate to participate in elections or disengage from online political discourse (Igareda, 2022). Those who continue to engage in political debate online often suffer psychological consequences and waste time evaluating whether the threats they receive pose real threats to their safety and well-being (Ibid.). Igareda underscores that online harassment functions as a punitive mechanism for female politicians who challenge the traditional male-dominated political space.

Responses and contestations
The Spanish survey studies discussed earlier shed light on various responses by targets of online violence. Morena-Balaguer et al. (2021) suggest that targets exhibit diverse reactions, ranging from ignoring the attack to publicly condemning it. Consistent with Jane's findings, some targets also employ strategies of "naming and shaming" in response to online violence.

3. The legal and policy framework in relation to gender-based online hate speech in Spain

Introduction: Hate crimes vs gender-based violence

In March 2021, the OSCE Office for Democratic Institutions and Human Rights (ODIHR) launched a factsheet on Gender-Based Hate Crimes, clearly showing the aim to encourage OSCE states to prosecute gender-based crimes as hate crimes. According to ODIHR, "gender-based hate crimes are criminal offences motivated by bias against a person's gender" and,

> One of the motivating factors behind this type of crime is the perpetrator's perceptions of gender norms. The victims of such crimes are often targeted due to their perceived deviation from gender norms, including on the basis of their sexual orientation and gender identity. Gender-based hate crimes may also target people or property due to their association, professional affiliation with or activism on gender issues, such as women's rights groups and civil society organizations working with victims of violence (ODIHR, 2021).

According to ODIHR, while gender-based hate crime can affect anyone regardless of gender, it disproportionately impacts women and girls. Victims may be targeted based solely on their gender or due to multiple identity traits, such as gender and religion. ODIHR also notes that certain cases of gender-based violence can qualify as hate crimes if the perpetrator demonstrates a gender bias during the incident.

At the collective level online, women as a group experience various form of technology-facilitated violence, including ridiculing gender theories, insulting or parodying women as a social group, and trivializing or denying gender violence (Igareda et al., 2019). Similar to hate crimes, online verbal abuse not only harms individual women but also sends a broader message to all women, implying that they must be cautious, refrain from raising their voices, and that their freedom is restricted. The result is that women, in all their diversity, are underrepresented, and their contributions are made invisible in the new democratic agora that is the Internet.

The legal framework applicable to gender-based online hate speech, or online gender-based violence in general, can be described as fragmented. Many aspects of this type of violence lack clear legal definitions:

> There is a diversity of nomenclature to name the same phenomenon, so to speak. And of course, no denomination has as an official or legal definition. Well, we are interested in the legal definition and of course you cannot find any legal instrument that defines misogynist cyber violence. Sometimes we find that they talk about cyberbullying, other times they talk about digital violence, but what happens is that all these denominations, what worried us is that the gender dimension was lost [EX-15].

The challenge of acknowledging the gender-based motivation behind online violence stems from two main issues. Firstly, the legal framework on gender violence in Spain primarily addresses intimate partner violence, leaving other forms of gender-based violence inadequately addressed. Secondly, although gender is listed as a protected characteristic in Spain's Penal Code, gender-based hate crimes are often not officially recognized as such. In the following sections, we will delve deeper into these frameworks and the gaps they present.

Hate crimes and hate speech in the Spanish criminal code

Although the Spanish Criminal Code does not specifically refer to "hate crimes", several articles take into consideration bias-motivated offences. In this sense, article 22.4a regulates the aggravating circumstances of crime, in which gender, sexual orientation and sexual identity are considered as particular circumstances, among others, which negate the principle of equality. That is, in order for a crime to be prosecuted as a hate crime it has to on the one hand, be categorised as a crime in the criminal code with an aggravating circumstance or correspond to a crime categorised within article 510 CP on incitement to hatred (or, as more commonly denominated, hate speech).

Article 510.1a of the Spanish criminal code refers to publicly encouraging, promoting or inciting, directly or indirectly, hatred, hostility, discrimination or violence against a group, part of it or against a specific person because of their membership of that group, for racist, anti-Semitic or other ideological reasons, religion or beliefs, family situation, the belonging of its members to an ethnic group, 'race' or nation, their national origin, their sex, orientation or sexual identity, for reasons of gender, illness or disability. Article 510.2a further refers to injuring the dignity of people through actions that involve humiliation, contempt or discredit of any of the groups referred to in the previous section, or of a part of these groups, or of any person determined by reason of belonging to these groups for racist, anti-Semitic or other reasons referring to ideology, religion or beliefs, family situation, the membership of its members to an ethnic group, 'race' or nation, their national origin, their sex, orientation or sexual identity, for reasons of gender, illness or disability. 510.2b refers to producing, elaborating, possessing in order to distribute, provide third parties with access, distribute, disseminate or sell writings or any other kind of material or supports that, due to their content, are suitable to damage the dignity of persons because it represents a serious humiliation, contempt or discredit of any of the mentioned groups, of a part of them, or of any person determined by reason of their belonging to these groups.

Legal experts agree upon the fine balance between what can be defined as hate speech and what cannot in a legal sense of the concept. Serra (2018) in this regard refers to a European consensus on the categorisation of hate speech on three levels:

> 1. Non-punishable "hate speech", which is harmful or disturbing speech in terms of coexistence and tolerance, but that does not deserve any type of sanction;

2. "Hate speech" of medium intensity, which deserves a civil or administrative sanction

3. Serious "hate speech", which deserves a criminal sanction.

Indeed, there's agreement that criminal sanctions may not be warranted for all instances of hate speech, opening the possibility for civil or administrative penalties instead. However, as we'll explore, options for civil or administrative sanctions are limited, particularly depending on the protected characteristics targeted by the discourses.

A review of jurisprudence and interviews with legal experts reveal that criminal cases prosecuted under article 510 are few and far between. This can be attributed to several factors. Firstly, widespread evidence indicates that underreporting remains a significant challenge for hate crimes and hate speech. Additionally, civil society organizations struggle to clearly delineate what constitutes hate speech from a criminal standpoint. As one NGO representative stated, "We believe that we need to polish and see what we are talking about when we talk about hate incidents on social media networks, because this also leads to a much deeper debate that has to do with freedom of expression..." (EX-04). The online environment introduces an additional layer of uncertainty regarding what should be reported to the police, as well as presenting logistical challenges in monitoring online activity. As another representative noted, "What happens online? It's true that we have a more, so to speak, passive attitude. That is, it has to come to us, we don't have the resources to go and look for incidents online" (EX-05). Moreover, online incidents are perceived to have been exacerbated by the Covid-19 pandemic, as much of social and work life - and consequently hate crimes - have shifted online: "So it's true that in 2020, with the lockdown, obviously the Internet and social media is an item that has skyrocketed. It's obvious" (EX-05).

Organizations and other stakeholders are generally aware that prosecuting these cases can be challenging, particularly due to the online dimension, which presents additional obstacles. A prosecutor highlights the difficulties in determining the authorship behind false or anonymous profiles:

> In principle, the difficulty we have with profiles that are false or anonymous is to know who is behind that profile. So most of the social media platforms are legally located in the US, when you have to request ownership of the accounts. And of course they don't facilitate it for freedom of expression in their country by the first amendment, okay? Or they simply tell you... They no longer tell you that, because they have agreed with the EU to withdraw content and all that. So if you want to know who the owner is, you have to do it with a rogatory commission through the US. They already know that the US judicial authorities are going to tell us that this is the first amendment and therefore they cannot give the information (EX-09).

A further reason for few cases successfully prosecuted under article 510 is the blurry interpretation of this article:

> There has already been a tremendous legal discussion about whether what counts is the motivation, in fact as 510 is articulated, it is a kind of strange mix, because on the one hand there is an international trend that defines hate crimes by discriminatory motivation, there is another sector that leaves in second place the motivation that in addition to being quite difficult to prove many times and puts on the table the effects on socially recognized groups as vulnerable (EX-06)

Despite the General Attorney of the State issuing a circular in 2019 aimed at clarifying the application of this article, some of the legal experts interviewed argue that it only added to the confusion and undermined the original purpose of hate crime laws, which is to protect vulnerable groups:

> The 2019 prosecution circular no longer speaks of motivation, it speaks of intolerance (...) I consider that it's the formula that the prosecution finds to finish exterminating the political foundation of hate crimes. Then it's no longer about categories or historical oppressions or sociodemographic justifications. It's about censoring intolerance and all radical critical discourses can be considered as such, and also here if we depoliticise it for historical or geographical reasons, hate speech of women against men could occur because it's an intolerance towards a category or a lifestyle or identity based on masculinity. And this worries me because it's directly at odds with the international warnings of the ECRI and company, keeping an eye on how the fight against hate speech is being applied. Here in Spain the aggravating factor of ideological discrimination against antifascists who have attacked Nazis is already being applied... (EX-06).

As previously mentioned, sex and gender are included as protected characteristics both under article 22 and article 510, however, whether gender-based hate speech can and should be prosecuted under article 510 is subject to different interpretations. This is further discussed in the next section.

Gender-based online hate speech and the criminal framework on hate crimes and hate speech

The application of article 510

In our investigation into cases of hate speech based on gender, through consultations with experts and an examination of jurisprudence, it is evident that Article 510 has seen limited application overall, particularly in instances of inciting or promoting gender-based hate. There exists a divergence of perspectives regarding the appropriateness of applying this article to cases of gender-based online hate speech. Notably, scholars like Gómez (2016) contend that Article 510 may not feasibly extend to sexist discourse, suggesting that for prosecution, there should be a demonstrable alteration in the collective security or honor of women as a group. Gómez argues that as such alterations are unlikely, prosecuting such discourse as hate speech may not be warranted, arguing that "it seems reasonable to demand, nonetheless, at least, that it is a minimally adequate behaviour to alter in some way the collective security or the honor of women as a collective. This, as is easy to see, will not happen easily".

Several experts we interviewed share similar viewpoints:

> Normally the classical or dominant interpretation of hate speech is when hate speech is apologistic of the holocaust, anti-Semitic, racist speech, but it would be very difficult for them to come to accept that anti-gender hate speech is really hate speech (...) If we try to apply hate speech against women or against certain women, it would be impossible to prove, because you cannot prove that this hate speech is inciting violence against all women in Spain and that it implies an imminent threat of persecution and annihilation of women, right? (EX-15).

Moreover, within the political context, discussions regarding freedom of speech regarding varying perspectives on gender identity, roles, and relations are often perceived as unrestricted to a certain extent. They are viewed as integral to democratic discourse among political representatives or parties:

> It's practically impossible that speech that is clearly sexist or denying of gender violence as we have in Spain with Vox, could be considered hate speech. Because you could never say that Vox is instigating people to murder women, starting tomorrow, nor could you really ever prove that their speech poses a threat to 51% of the Spanish population that are women. That's why it would never be considered hate speech no matter how much the content disgusts us. That is, being sexist is not a crime. That is also sometimes interesting, differentiating between hate speech and discriminatory or sexist or homophobic content, or xenophobic, being xenophobic or being sexist is not a crime in itself and that sometimes is difficult to understand (EX-15).

Other experts, however, argue that art 510 can be applied to gender-based hate speech: "Can you apply it? It can be applied. It's been contemplated like that. Another thing is that statistically I don't know of any case in which it has been applied" (EX-06). The scarce application is perceived, on the one hand, as linked to a lack of effort or decision in the feminist movement in including sexist or gender-based discourses in the broader category of hate speech:

> So there is a debate about whether it is more strategic for women to embrace this broader, more diffuse category. For example, when the Code of Conduct of the large internet intermediary platforms was approved in 2016, gender discourse has remained there half volatilized. Or, the other option to claim it as political violence that affects half the planet, and turn to tools for women's rights and the fight against sexist violence [EX-06].

Further, the underutilization of Article 510 in cases of gender-based online hate speech is also perceived as stemming from underreporting. Few victims report instances of hate speech to law enforcement, and when they do, it is not categorised under Article 510 or with the aggravating factor of gender. Moreover, the limited application of hate speech laws to address gender-based hate speech is also tied to the perception of violence against women as primarily occurring within the private sphere, particularly within (heteronormative) romantic relationships. As a result, the structural aspects of gender-

based violence in contexts beyond intimate partner relationships remain largely unseen, and such violence is not recognized as inherently gender-based or political:

> The system has a tendency for everything that is violence against women to pass through a kind of private filter or conflict, let's say personal. Therefore, it is always easier for a certain discourse against a feminist journalist to be read as an attack on honor instead of political violence. So between the under-reporting of women through 510, the internal dynamics of the judicial system and feminisms that have not made it a battle to include misogynistic hate speech in this category, there has been a reality of general underuse of this article (EX-06).

Another limitation to the application of art 510 for gender-based hate speech is, as argued by a lawyer, that "although it could be applied to legal persons, it is focusing on individual offenders when in the Spanish state those who exercise great misogynistic hatred are religious confessions, right-wing parties and various groups. That should also be taken into account" [EX-06]. Igareda et al. (2019) acknowledge the challenge of applying criminal law to collective online crimes, such as those involving multiple perpetrators, as discussed in their research on misogynistic cyberviolence. They also emphasise how the unique features of digital violence raise concerns about the effectiveness of current legal frameworks.

The failure to recognise gender-motivated crimes (beyond anti-LGBT hate crimes) as hate crimes in Spain is evident when consulting with a specialised police unit focused on hate crimes. They seldom receive reports concerning gender-based online hate speech and could only recall one instance of online hate speech targeting women—a hate website reported by a women's rights organization.

The aggravating factor of gender of article 22.4

With the recent focus from OSCE-ODIHR on gender-based crimes as hate crimes, it's worthwhile to briefly examine the aggravating factor of gender outlined in Article 22.4 of the Spanish criminal code. This article has been subject to widely differing interpretations regarding its scope, largely influenced by the legal definitions of gender violence within the Spanish criminal framework. Scholars such as Marin de Espinosa (2018) have scrutinized this aggravating circumstance, contending that, due to the flawed concept of gender violence in the Organic Law 1/2004, the factor of gender is limited to cases where the victim is or has been in a romantic relationship with the perpetrator. According to this perspective, online threats or harassment could only be prosecuted with this aggravating factor if the perpetrator is the victim's current or former partner, thereby situating such harassment within a heteronormative framework.

However, other experts argue that Article 22.4 can be applied and is indeed being applied beyond cases of intimate partner violence, albeit infrequently:

> The aggravating factor of gender of 22.4 of the penal code is hardly being applied, but since 2017-2018 the supreme court has begun to make sentences that are quite good, incorporating and validating the gender perspective. There are a few

sentences, like one that I got last week, that are beginning to apply the aggravating factor of gender in sexual violence, outside the framework of the couple (EX-06).

Jurisprudence linked to article 510

Upon examining jurisprudence, academic literature, and various reports, only two instances of convictions for online hate speech under Article 510 specifically based on gender have been identified between the introduction of gender as an aggravating circumstance in Spanish criminal law in 2015 and the conclusion of the current research in 2023. One case pertains exclusively to gender, whereas the other case involves publications that incite hatred on multiple grounds, including religion, nationality, sexual orientation, and gender. In the subsequent sections, we present an overview of these cases.

Cases prosecuted and sentenced under art.510 with a gender motivation

Case 2: STS 2085/2022 Supreme Court

The first case, STS 2085/2022, revolves around publications by Antonio, also known as Argimiro, who utilised three different profiles on Facebook to post numerous comments targeting various groups. Initially handled by the Provincial Court of Madrid, the case involved categorising the posts based on the groups they targeted, organizing them according to the specific grounds or protected characteristics under attack. The targeted groups included: other branches of Islam apart from the one the defendant had converted to (religion); Spanish individuals, particularly those from Andalusia (nationality/origin); LGBT+ individuals (sexual orientation); individuals affiliated with the political "right-wing" (ideology); and finally, women as a collective (gender). The defendant posted multiple comments relating to each of these groups. Regarding the group most pertinent to this research, namely women as a collective, the sentence outlines nine distinct posts published on seven different dates. Below are excerpts from four of these posts:

— July 20, 2017: The rebel woman / You hit her / The rebel woman/ you slap her a couple of times
— July 20, 2017: Whores and shitty drunks / They are the Spanish whores / They eat multiple dicks / They're not even valid as slaves
— August 13, 2017: (...) If a man obeys to feminist indications / It will be the man who suffers / Physical violence / Economic violence / Family violence / I'm tired of that feminine egocentrism that makes the harm that women suffer / more important than the harm that men suffer / I think I haven't murdered an ex-partner who kidnapped my daughter /Because of the enormous quantities of joints that I smoke / But if instead of smoking joints / I drink whiskey / I would have run over her or beat her up or shot her or anything to /end her drunk dirty cocaine addict existence.
— August 30, 2017: Whether it's a man or a woman / If they attack me, I defend myself / Violence against women is good

- January 1, 2019: You feminists are whores who eat pig / and you are what you eat / You try to destroy the biology of man forcing him to be cold in situations where you don't wear proper clothes that cover your erogenous zones / In conversations you don't put any limits / But the man has to stay cold / Is that what you want whores? To be as slutty as you can without suffering the consequences? Is that what you want to be, whores?

Antonio was convicted as the primary perpetrator of a crime against fundamental rights under Article 510.1a and 3 of the Penal Code and was sentenced to two years and six months in prison, along with a fine of nine months and one day, with a daily fee of two euros. The sentence was appealed and later upheld by the Supreme Court. Regarding the expressions of gender-based hate speech, in conjunction with hate speech directed at other groups, the Supreme Court emphasized that this behaviour transcends the mere expression of hurtful or offensive opinions via social media. Instead, some passages—either individually or in conjunction with others—indicate an encouragement to engage in action, violence, and armed conflict. The Court further noted that Antonio not only encourages others but positions himself as the primary aggressor against those he despises due to their ideology, gender, sexual orientation, or nationality. Consequently, the Court asserted that it cannot condone such hate speech, which encourages online users to participate in violence and advocates for violence against women as a model of coexistence.

Case 2: STS 396/2018 Supreme Court

The case STS 396/2018 reached the Supreme Court following an appeal by the representative of the accused party. The case can be summarized as follows: On January 26, 2017, the Fourth Section of the National Court issued a conviction against the defendant, Miguel, for incitement to hatred (Article 510 of the Criminal Code), stemming from his publication of a series of messages with illicit content on Twitter. Among the established facts that led to this verdict, Miguel had maintained ownership of two accounts on the social media platform for several years, collectively amassing around 2,000 followers. Through these accounts, he disseminated posts containing derogatory remarks aimed at women, while also endorsing and encouraging acts of violence against them. Specifically, he posted the following messages through his first account:

- December 17, 2015: "53 murdered by sexist gender violence so far this year, it seems few to me with all the whores that are loose"
- December 30, 2015: "And 2015 will end with 56 murdered women, it is not a good mark, but we did what we could, let's see if we double this figure in 2016, thank you"

As a consequence of these postings, the Police Unit Grupo de Redes II received multiple reports from concerned citizens regarding the comments. Additionally, on January 7, 2016, Twitter suspended the defendant's first account, prompting him to utilize his second account to continue publishing posts with similar content:

- January 14, 2016: "Beatriz was a feminist and she jumped into the river because women get wet for equality."
- January 14, 2016: "I like to fuck against the counter and the stove, because I put the woman in her place twice"
- January 16, 2016: Sharing the image of a woman, regarding whom it is unknown if she has been a victim of abuse or gender violence, the accused published the comment "I have already battered her, you're up next"

The Supreme Court found Miguel guilty as the primary perpetrator of a crime of incitement to hatred under Article 510, subsections one and three, and sentenced him to 2 years and 6 months in prison along with a fine of nine months, with a daily rate of €40. The Court clarified that Article 510 CP penalizes individuals who promote discrimination, hatred, or violence against specific groups or associations based on the criteria outlined in the provision. It underscored that the essence of such criminal acts lies in the use of epithets, qualifiers, or expressions conveying a general message of hatred. The Court further emphasized that the offense of incitement to hatred does not require a specific intent but relies on the presence of a fundamental malice discernible from the content of the expressions used. Intent in these crimes is established by confirming the voluntary nature of the act and ensuring that it is not a response to an uncontrollable situation or a momentary, even emotional, reaction to an external circumstance. In this context, the factual account pertains to the publication of the messages on various dates, indicating voluntary conduct rather than a reaction to an external stimulus. Conversely, the content of the verdict underscores the aggressive nature of the expressions and the manifestation of hatred when referring to situations in which the defendant seeks to position women, whom he addresses in aggressive terms within a gender-based context (Article 510 1a).

Cases prosecuted under art.510 and not sentenced

Article 510 has also been tried for incitement to hatred based on gender in a few other cases, but without sentence. One case involved threats against a woman on Facebook, including, amongst others, the following expressions: "If you do that to my daughters, I won't break your phone. I'll break your head, you disgusting bitch," "Next time her head is broken... if she plays the victim... let it be for good reason... ". Here, the court interpreted that this would not amount to incitement to hatred under art. 510, as

> these are expressions addressed to a specific person and as a result of a previous incident carried out by the same, without it being possible to estimate that it integrated a discourse whose purpose was only to propagate violence or promoting hatred against a certain group, even though the complainant may belong to that group and this without prejudice to the fact that, as we have already indicated, the expressions discharged through the social network FACEBOOK could constitute a minor crime of threats and/or harassment.

The ambiguous language of Article 510 and the widespread lack of understanding regarding what constitutes hate speech from a legal standpoint may contribute to the underreporting of online hate speech. Additionally, the limited dissemination of

convictions under the legal framework exacerbates this lack of awareness and subsequent lack of action (Lizardo, 2023). Consequently, analysing jurisprudence aids in fostering a clearer comprehension of the boundary between freedom of speech and hate speech within the legal context.

Both the tried and prosecuted cases and those tried without sentence highlight several key points. Firstly, they underscore the necessity for hate speech targeting women as a group to include generic hatred towards them to be legally defined as hate speech. However, as indicated in the court's statement in the case above, this hatred towards women as a collective can be integrated into an attack targeting a specific individual. Secondly, the cases emphasize that there is no requirement for a specific intention to cause physical harm to the collective for the speech to constitute a criminal act. The aggressive nature of the expressions against the group alone is sufficient to limit freedom of expression and sentence such expressions as incitement to hatred. Thirdly, the cases demonstrate that the intent of these crimes is established by the voluntary nature of the act and verification that it is not an isolated, uncontrolled situation or a momentary emotional reaction to an uncontrollable circumstance. Instead, the sentenced cases involve multiple publications of generic hatred against the same collective on different occasions, and in the cases presented here, even across different social media accounts and platforms.

In the next section, we will delve further into the gender violence laws in Spain and their limitations to better understand the various interpretations of the potential application of the aggravating factor of gender.

The legal framework on gender-based violence in Spain

The Istanbul Convention defines violence against women as a "violation of human rights and a form of discrimination against women" encompassing all acts of gender-based violation leading to, or likely to result in, physical, sexual, psychological, or economic harm or suffering to women, including threats of such acts, coercion, or arbitrary deprivation of liberty, whether occurring in public or private life. According to the convention, "gender-based violence against women" refers to violence directed against a woman because of her gender or that disproportionately affects women, stating that acts against a woman aimed at discriminating, dominating, or subjugating her can occur in any context.

Despite entering into force in 2014 and being signed and ratified by the Spanish state, the Istanbul Convention's principles have not been fully integrated into Spanish law. As noted by Marin de Espinosa (2018), the concept of gender violence has been inadequately incorporated into the Spanish legal framework. The Organic Law 1/2004, of December 28, on Comprehensive Protection Measures against Gender Violence, only considers gender violence in cases where the perpetrator is or has been in a romantic relationship with the victim. Thus, the ratification of the convention did not entail an adaptation of Spanish

laws to encompass the concept of gender-based violence comprehensively. As Espinosa (2018: 13) points out, this results in an incomplete law that views gender violence as a problem confined to the private sphere, failing to recognize gender-based violence occurring in the public sphere. Consequently, Spanish law does not acknowledge other forms of violence such as forced marriages, genital mutilation, sexual violence and abuse, or digital violence (outside of intimate relationships) as gender violence. As articulated by a legal scholar:

> The EU and Council of Europe instruments talk about gender-based violence but Spain has translated this into gender violence, which is not the same as gender-based violence, so it seems to be a translation trick so to speak, but of course it has very important legal consequences. So of course, we find laws that don't really talk about gender-based violence, because they only talk about violence within the couple [EX-15].

The Spanish definition of gender violence's incompleteness severely limits the recognition of numerous crimes as gender-based. While acknowledging that intimate partner violence is a significant issue in Spain and warrants attention, experts interviewed argue that this focus should not come at the expense of other types of crime motivated by gender. As one expert expressed: "I think it leaves out many types of violence, it ends up being explosive, it ends up reinforcing binarisms, heteronormativity, some of the things that I think are very interesting to be questioned but not to invalidate everything else" [EX-17].

Gender-based cyber violence

Igareda et al. (2019) conducted research on the legal articles under which gender-based cyber-violence can be reported and prosecuted. Their study encompasses various forms of online violence against women, extending beyond verbal abuse. Igareda et al. (2019) point out that different manifestations of gender-based cyber-violence may fit into different legal classifications of violence, yet they are not always recognized as sexist or gender-based violence.

The researchers list the legal articles applicable to gender-based cyber-violence, categorising them into articles falling under the framework of gender violence (i.e., instances involving a relationship between the victim and the perpetrator) and those falling outside this framework and thus not categorized as gender violence.

Within the context of intimate partner violence, the first category includes Articles 147 and 1489 - Injuries; Article 153 - Physical or psychological abuse; and Article 171 - Threats.

In the second category, outside of a relationship and thus not within the Spanish criminal code's definition of gender violence, the following articles can be included: Art. 171.7 - Minor threats; Art. 172.3 - Minor coercion; Art. 173 – Crime against moral integrity; Art. 178 - Sexual assault; Art. 184 CP - Sexual harassment; Art. 172 ter - Stalking; and Art. 197.7 CP - Dissemination without consent of intimate content.

As emphasized by Serra (2018), the new criminal provisions introduced in the Spanish criminal code under the influence of the Istanbul Convention, aimed at safeguarding women from gender violence, have been formulated in a gender-neutral manner within the Spanish context. Serra provides the example of stalking, noting that due to the limitation of state legislation on gender violence to the context of (ex)-partners, online violence against women outside of this framework is not recognized as gender violence. Consequently, while the sentence for stalking is aggravated if committed within the context of a partner or ex-partner, in other cases, the sentence remains the same regardless of whether the perpetrator is male or female. According to Serra, this failure to acknowledge the gender discrimination component in online violence occurring outside of intimate partnerships is significant.

Therefore, rather than introducing new laws or tightening the criminal code, there is a need for a reformulation of existing provisions in the criminal code, aligned with an updated definition of gender-based violence in line with the principles of the Istanbul Convention. As noted by Igareda et al. (2019: 84):

> When the behaviours constitutive of sexist cyber-violence are reported, it will most likely not be under a specialised court of gender violence and, therefore, the sensitivity or knowledge of the judge about this phenomenon is very arbitrary. Many times, this lack of knowledge about sexist cyber-violence ends up re-victimising women, blaming them for what has happened (author's translation).

Moreover, victims frequently minimise the violence and fail to preserve evidence. Additionally, when cases reach the court, legal professionals often underestimate the significance of the evidence provided or lack the specialised technical skills necessary to effectively utilize all available evidence (Igareda et al., 2019). This indicates a lack of recognition of the gender-based motivation behind these online crimes, resulting in a deficiency of essential specialisation, both in addressing victims' needs and in handling technical aspects. This deficiency in specialisation and awareness regarding the specific nature of these crimes has ramifications throughout the entire criminal justice process, including reporting, prosecution, court procedures, and outcomes, significantly impacting the victim's experience.

However, there is at least one potential positive aspect of gender-based online crimes not being classified as gender violence, namely that mediation or restorative justice measures may be employed in these cases: "Because these forms of violence are not recognized as forms of gender-based violence, this is possible, because gender violence laws and many autonomous laws of gender-based violence clearly state that mediation in cases of gender violence is not possible in crimes where there is not equality between the parties" (Igareda et al., 2019: 87 - author's translation). In Igareda's study, lawyers emphasize their attempts to utilize direct mediation to avoid lengthy court proceedings. However, this approach is limited to cases where the perpetrator can be identified and contacted, thereby excluding cases involving collective attacks or anonymous accounts.

Other legal tools

Civil and administrative laws

In Catalonia, the administrative law addressing gender-based violence underwent recent changes with the enactment of *Llei 17/2020, del 22 de desembre, de modificació de la Llei 5/2008, del dret de les dones a erradicar la violència masclista*. Prior to this reform, the Catalan context already had a broader definition of gender-based violence in line with the Istanbul Convention. However, this law reform introduced significant changes by incorporating the concept of digital violence.

According to the law, digital violence is defined as "acts of sexist violence and misogyny online committed, instigated, amplified, or aggravated, in part or in whole, using information and communication technologies, social network platforms, websites or forums, e-mail and instant messaging systems, and other similar media that affect the dignity and rights of women". These acts result in psychological and sometimes physical harm, reinforce stereotypes, damage dignity and reputation, violate women's privacy and freedom of action, cause economic losses, and hinder political participation and freedom of expression.

The law delineates various areas where gender-based violence manifests, with the following being particularly relevant to the scope of the present research:

Firstly, reprisals against individual and collective discourses and expressions of women demanding respect for their rights, as well as expressions and public discourses that directly or indirectly encourage, promote, or incite hostility, discrimination, or violence towards women.

Secondly, violence in the digital sphere, encompassing sexist violence occurring in digital communication networks, which are seen as new arenas of interaction, participation, and governance through information and communication technologies. This includes practices such as cyberbullying, surveillance, defamation, insults, threats, unauthorized access to social media accounts, invasion of privacy, manipulation of private data, impersonation, non-consensual disclosure of personal information or intimate content, damage to women's equipment or channels of expression, discourse inciting discrimination against women, sexual blackmail through digital channels, and the publication of personal information with the intention of others assaulting, locating, or harassing a woman.

Thirdly, violence in the sphere of political and public life of women, encompassing sexist violence occurring in public and political spaces such as political institutions, public administrations, political parties, the media, or social networks. When this form of sexist violence occurs in political institutions or public administrations and is tolerated and not sanctioned, it becomes a form of institutional violence.

While some of the experts interviewed view this law as a significant step forward in recognizing digital violence against women, others perceive it as merely a symbolic change, as

> the competences in criminal matters only belong to the state. Then there are autonomic laws like the Catalan that go beyond the definitions of the violence stipulated by the state law. It's very well from a political point of view, symbolic, but in legal effects it has no repercussions because it does not modify the penal code or modify the main laws [EX-15].

However, beyond its criminal implications, this law could potentially provide victims of online violence with access to support mechanisms designed for victims of gender-based violence. Regrettably, these support systems are already strained due to the high number of victims of intimate partner violence and would require additional resources to adequately address victims of other forms of violence.

Experts also point to how this law could contribute to compiling statistics:

> Perhaps this would allow that all these cases of digital violence that reach the Mossos d'Esquadra, will be identified as gender-based violence and will be counted in a specific way, because if not now, when you ask what's the prevalence of digital violence in Spain? We have no idea (…) And what's not quantified doesn't exist and you can never justify it as an issue of attention from a legal or political perspective [EX-15].

Returning to Serra's categorisation of hate speech, where hate speech of 'medium intensity' would be subject to civil or administrative sanctions, the reformed law could in theory open up for a path for administrative sanctions for gender-based online hate speech. However, some experts interviewed view this path as unlikely to have any real consequences, as the institution responsible for upholding this administrative law has no powers in investigating on the Internet and social media. Nevertheless, it could be a path for collecting statistics on digital violence which might not be a criminal matter, and also a way for victims to receive support if new services and resources are provided in accordance with this law.

Catalonia also has, since 2014, an administrative law to uphold the rights of LGBT people, *Llei 11/2014, del 10 d'octubre, per a garantir els drets de lesbianes, gais, bisexuals, transgèneres i intersexuals i per a erradicar l'homofòbia, la bifòbia i la transfòbia*. Through *the Àrea per a la Igualtat de tracte i no-discriminació de persones LGBTI of the Direcció General d'Igualtat*, the Generalitat de Catalunya collects civil infractions, including those taking place online. However, the balance is often complex - and even more so online - between behaviour that could constitute a criminal offense and incidents that should be treated as administrative infractions. Representatives of LGBT organizations interviewed admit the difficulty in categorising online incidents and recognise that "at the level of legal regulation it's much more complicated to manage this, when things happen on social media, because there is so much, so much. And we set the criteria: if you report to us, we record it" [EX-05]. Another issue that LGBT organizations struggle with is that a high

percentage of reports are presented by white gay men, whilst other groups within the collective, such as lesbian and bisexual women, as well as trans persons, report incidents to a much lesser extent:

> I think it has to do with the logic, it has to do with the logic that the white gay man can report. And in fact, he reports, or has the chance to report it. Maybe women, and I'm talking hypothetically because I haven't come across so many cases, maybe lesbian women have other itineraries that don't go through reporting many times when it comes to repairing their dignity [EX-05].

Nevertheless, regardless of whether it's a crime or an administrative infraction, LGBT associations encourage those that seek their advice to report:

> And we always say, go and report it, for legal reasons, if you don't report, at the legal level [the crimes are not recorded]. In fact, the vast majority of incidents don't have a legal itinerary. They don't go through a legal or criminal route, although it's the most common of our cases, this year I think we have received 30 criminal cases [EX-05]

Similarly, LGBT organizations in Catalonia express that very few infractions reported administratively lead to administrative sanctions. According to a prosecutor interviewed this is due to a blurring between administrative infractions and criminal offenses in the Catalan LGBT law: "It's an ineffective law, because all the behaviours it describes are criminal, they fall within the criminal jurisdiction. Of course, they complain that the law is not enforced, well they don't have administrative infractions to apply, because everything described in it is a crime" [EX-09].

However, a well-defined administrative law could facilitate addressing some infractions taking place online:

> On the Internet there are behaviours or there is dissemination of expressions that are not criminal and that don't have to be criminalised because not everything is criminal law, but it would be good if that type of insults were classified or this type of comments that are sometimes done when a piece of news is published in a digital newspaper or on social media, that this could be pursued at an administrative level [EX-09].

Detection and intervention protocol for victims of digital gender violence in Andalucia

Since 2015, the Andalusian government has had a detection and intervention protocol for victims of digital gender violence, which is defined as "Gender violence that is carried out taking advantage of ICTs" (Instituto Andaluz de la Mujer, 2015). Positively, the protocol does not only refer to gender violence as defined by Spanish law, i.e. violence within the (ex)-couple but lists a range of crimes, including discovery and disclosure of secrets (data theft of data, photos/videos, accounts, profiles); injuries, slander through ICT; denigrating treatment through ICT; dissemination of degrading or harmful images through ICT; identity theft; fraudulent use of cards; damages (equipment, companies, professional activity); grooming and crimes related to corruption and prostitution of

minors; induction to the abandonment of the home to minors through the Internet; induction to suicide through the Internet; threats and computer coercion; sextortion; child pornography with use of ICT. Finally, it also mentions that "At the collective level, there is also a crime that is increasing through ICT: The apology of discrimination and gender violence" (Ibid.). This implies a broader concept of gender violence, which could include what the present project refers to as online gender-based hate speech. As highlighted by Vergès et al (2017), the protocol seeks to facilitate the detection and registration of technology-facilitated gender violence as well as to incorporate this view in the actions of professionals who provide direct attention to victims.

Right to be forgotten

Among mechanisms of a civil nature, where there is not a criminal case, the protection of personal data can be highlighted. This can be, for example, the dissemination of images or distribution of other personal data, e.g. in cases of doxing, such as private addresses. In these cases, the right to be forgotten can be exercised, in accordance with European Regulation 2016/679 of the Parliament and of the Council of 27 April 2016, known as the General Regulation of Data Protection. The exercise of the right to be forgotten can be used as an instrument to empower women who suffer from cyber-violence to the extent that there is no need to report to police or judicial authorities. Rather, it is a tool where you can request directly the deletion and/or non-indexing of the content (Igareda et al, 2019). Nevertheless, we have not found any information on to what extent this is used in cases of cyberviolence against women, or to what extent the petitions of deletion of content are successful.

Protocol to combat illegal hate speech online

The protocol, coordinated by the Observatorio Español de Racismo y Xenofobia of the Ministry of Inclusion, Social Security and Migrations (OBERAXE), is inspired on the one hand, by the Code of Conduct, signed by the European Commission and several social media platforms, in 2016, and, on the other hand, by the Recommendation of the EU 2018/334 of the European Commission, of March 1, 2018, on measures to effectively combat illegal content online. The document agrees to combat illegal hate speech online; establish a national focal point - the Computer Crime Unit of the State Attorney General's Office - as the interlocutor of the public administration with internet companies; accredit and train trusted flaggers (or trusted users); preferably process communications from duly accredited trusted flaggers; establish homogeneous notification circuits for hate speech; and finally, implement the protocol and its monitoring, within the framework of the Interinstitutional Agreement to combat racism, xenophobia, LGBTI-phobia and other forms of intolerance. Whilst gender-based online hate speech is not explicitly included in the protocol, this type of hate speech and its intersections with other factors, can be interpreted as included in "other forms of intolerance".

Conclusion

The discussion on addressing gender-based online hate speech often centers around legislative frameworks, highlighting the need for comprehensive legal definitions and consistent application across jurisdictions. Effective solutions may not solely lie in creating new laws but in ensuring the constructive use of existing tools with a gender perspective. The current legal fragmentation poses practical challenges but serves to protect various rights across different legal domains.

The protection of legal goods ("bienes jurídicos") such as privacy, dignity, and security may require re-evaluation in light of evolving social values and the impact of digital spaces on these rights. Incorporating interdisciplinary insights from anthropology, sociology, and political science can enhance the legal response to digital violence.

Empowering women with digital self-protection skills and incorporating gender perspectives among judicial and law enforcement professionals are critical steps. Addressing the root causes of gender-based violence, including structural inequalities and the perpetuation of male supremacy, is essential for long-term solutions. Recognising institutional responsibilities and creating specific laws to address online and institutional violence are necessary advancements.

A holistic approach, involving both legal reform and societal change, is imperative to mitigate gender-based online hate speech effectively.

4. Being targeted by online violence as a feminist-, anti-racist or LGBT+ activist

Introduction

The participants in the fieldwork share several common characteristics: they all identify as women, engage publicly in debates on feminism and gender equality, LGBTQ+ rights, and anti-racism, and have experienced online hate speech or other forms of attack. These factors significantly influence the types of online violence they discuss in their narratives. Their roles as feminist and political voices in online debates were the primary reason for their inclusion in the study. The incidents that led to their participation in the research or their inclusion in the digital ethnography were directly linked to their activism in feminism, LGBTQ+ rights, or anti-racism. Therefore, the interviews and the digital ethnography focused on attacks, harassment, and abuse stemming from these activist roles. This delineation is crucial to understand when examining the experiences presented in this paper.

Lived experiences of online violence

When discussing online violence targeting these women based on their activist or professional profiles (as opposed to targeting them solely based on their gender and/or other intersecting factors), the incidents of violence were often linked to specific triggers. These included a media appearance, a tweet or other social media post by the targeted individual, or a publication by an anti-feminist or transphobic public figure that incited attacks from their followers. Most interviewees identified a clear "trigger" that precipitated the attacks.

Women who faced attacks due to their feminist positions are aware of the specific topics that typically provoke backlash in the public discourse of contemporary Spain. These topics include feminism and gender equality, trans rights, and any criticism of the so-called "manosphere." For instance, T-3 shared how a humorous meme she tweeted about Jordan Peterson led to a wave of attacks from his Spanish-speaking followers: "What I didn't take into account was that the fans of Jordan Peterson, who is an openly fascist person, well they said some barbarities in some messages, and in quotes on Twitter, telling me any kind of thing really" (T-3).

However, even seemingly innocuous topics such as astrology can trigger attacks, particularly when discussed by women who identify as feminists. T-8 noted, "Astrology is a sensitive topic because it stands out a lot and I think it triggers certain discourses and certain violence." Similarly, any discourse that challenges traditionally male-coded interests, like football, can provoke backlash. T-6 shared, "The problem is that I messed with football. Or I didn't mess with football, I said 'many people let football condition and determine their lives, the schedules, the moods.'" This suggests that the specific topics

triggering misogynistic attacks are interchangeable. As Jane (2017:11) points out, "a woman could receive a near-identical spray of rape threats for commenting about basketball games, bike riding, comic book covers, soft pretzel recipes, DIY fishtail braid videos on YouTube, and on and on ad infinitum."

T-6 linked this phenomenon to her public feminist stance, adding that "there are people who are quite angry with me". She perceives that individuals with strong opposing views seize any opportunity to insult or ridicule her.

Feminism itself is a significant trigger for online violence. T-5 highlighted how the mere announcement of a feminist talk on a national TV show's social media channel incited hate messages: "The fact that I talked about feminism was the trigger for this kind of hate... When it was announced that I would give a talk on feminism, there were already hate messages in relation to that announcement. Even before the talk...". It wasn't necessarily the content of the talk that enraged her attackers, but the mere fact that the concept of feminism was brought into the mainstream. Online abuse against feminists often aims to challenge both their views and their fundamental right to participate in public debates and spaces (Lewis et al., 2019). The interviews and digital ethnography reveal that whether a tweet, publication, or media appearance leads to attacks often depends on its reach. As T-5 noted, it's when a message "goes into the mainstream," and T-3 added, when it "reaches the incel timeline," that attacks are more likely to occur. Once someone has reached a wider audience, they become more susceptible to subsequent attacks. Massive harassment often begins when individuals discuss feminism—or simply appear as feminists—on mainstream platforms.

In the case of T-5, the harassment continued for an extended period. Even after the initial attacks subsided, men kept sending her private messages on Instagram. This persistent behaviour from numerous men suggests that such attacks are often more organized than they might initially appear. Years later, T-5 remains vigilant, wary that any media appearance might trigger new attacks.

Similarly, T-1, an LGBT+ activist, has come to expect negative reactions to any media appearance:

> You already assume that... For example, the other day when I did the interview [on TV] I knew that they would put it up [on the channel's social media] and that there would be bad messages, so you already expect it, and it's like, let's see where they get you today. And that's quite crappy now that I think about it (T-1).

This experience echoes that of DE-8, a racialised communicator, who notes that every time she appears on mainstream television, she receives hate messages. As a visibly Muslim woman, her mere presence on mainstream platforms triggers online attacks, which she perceives as coming from men with far-right political views. DE-15 and DE-17 also report being targeted by hate speech and racist discourses simply for being racialized women in mainstream media. As a Roma feminist, DE-17 has faced repeated gendered and Roma-phobic abuse across various platforms. In a tweet, she reported

receiving over 100 abusive messages in response to a TV appearance and publicly asked the TV channel to moderate the comments, which they eventually did.

Trans women face similar challenges. Their presence on social media often triggers attacks, especially in the current polarised debate over trans rights in Spain and beyond (e.g., Willem et al., 2022). The digital ethnography highlights this clearly. Both DE-6 and DE-9 have been targeted by transphobic discourses simply for sharing photos that do not conform to normative gender constructions.

Gender-based attacks often exhibit a wave-like pattern, particularly when perpetrated by many individuals. The experience of T-6 serves as an illustration of this phenomenon. She recounts three distinct waves of attacks directed at her over a short period. In two of these instances, prominent individuals with opposing views initiated the attacks by negatively mentioning her on their social media platforms, prompting their followers to engage in targeted harassment against her. Understanding the pivotal role of these attacks' origins from influential profiles is crucial. T-6 notes that as a public feminist, she has always faced occasional hate. However, the massive extent of these three waves had a significant impact on her. These waves had a clear starting point, grew to a peak, and then gradually decreased. As seen in the case of T-5, the decline can be prolonged, with abuse continuing for years after the initial peak.

In the experience of T-6, the starting point was a mention on social media by a prominent figure. The attacks then escalated as the followers of the original perpetrator joined in. In one instance, the attacks lasted 6-7 hours, while in another, they continued for several days before tapering off. The 6–7-hour attack only ceased because T-6 made her Instagram account private, where the main attacks occurred. This illustrates that a seemingly isolated comment by a high-profile individual can quickly escalate into a widespread attack, beyond the control of the original poster. The initial instigator can thus disassociate from the attack, while the target continues to suffer prolonged harassment from the instigator's followers.

Discourses and content of the attacks

The striking uniformity of gendered cyber-hate rhetoric across different contexts reflects what Jane describes as "a diachronic perspective on gendered e-bile," noting its persistence and rhetorical consistency over many years (2014: 560). She provides explicit examples to illustrate several recurring characteristics:

> They target a woman who is, for one reason or another, visible in the public sphere; their authors are anonymous or otherwise difficult to identify; their sexually explicit rhetoric includes homophobic and misogynist epithets; they prescribe coerced sex acts as all-purpose correctives; they pass scathing, appearance-related judgments and they rely on ad hominem invective (Ibid.).

These characteristics remain prevalent in the online hate directed at feminists and other women activists who engage in public debate, almost ten years after Jane's 2014 article

and in a different social and geographical context. The main exception compared to the situation in 2014 is the anonymity of the authors. In contemporary online spaces, anonymity is no longer a necessary condition for perpetrating gendered online attacks.

In our research, the discourses and content of the attacks do not primarily contest the feminist discourses expressed by the targets. In fact, the content of the targets' feminist discourses was rarely debated by the attackers – further than expressing offense at feminist opinions at large:

> My feeling from the comments I received was like they wanted to call me unpresentable, understanding feminism as a great setback, they felt particularly offended by the speech I had made, in some cases. Others directly had no greater intention than to humiliate and denigrate me (T-5).

The perception among those targeted is not that the attackers are interested in debating the topics under discussion. Instead, the aim is to ridicule, humiliate, and denigrate women who express feminist opinions. These attacks are not grounded in arguments or ideological debates but are characterized by insults and discrediting remarks (cf. Morena-Balaguer et al., 2021).

The discourses and content of the messages targeting the interviewees can be placed along a continuum, with death threats at one extreme and seemingly less serious insults at the other (cf. Lewis et al., 2017). Examples of the latter include repeatedly telling a comedian she is not funny or telling a communicator she is bad at communicating. Although some experiences may not appear directly threatening, this does not mean they are necessarily less impactful or harmful. Online violent expressions may not reach the threshold of being criminal, but they can still be experienced as violence. The repeated nature of these messages, especially from different individuals, can lead a person to question themselves both personally and professionally. As Harris and Vitis (2020) highlight, "some behaviours may be viewed as typical and normalised, others problematised or criminalised, but all have the potential to impact women's wellbeing, sense of safety, rights, and freedoms."

Situated between these extremes of the continuum are sexualized messages targeting or discussing the women's physical appearance, rating them as sexual objects, or posing sexualised threats, in line with much of the international literature referenced in Chapter 2 of this report. (e.g., Jane, 2014, 2017; Megarry, 2017). These sexualised discourses are often coupled with direct threats of rape or even murder and may move quickly from a comment on a woman's appearance to rape threats or -fantasies.

In addition to sexualised content, the discourses often include references to mental health or intelligence, suggesting that these women are either crazy or unintelligent. For instance, T-9 recalls being referred to on Twitter as "mentally destitute" by a man with a public profile, while DE-19 shares screenshots of a man calling her "rude, uneducated, unstable, and 'loca del coño'." Furthermore, nearly all the targets interviewed, along with many in the digital ethnography, have been labelled with the catch-all insult "feminazi."

Although intended as a pejorative, feminists have humorously embraced and to some extent reclaimed this term in recent years.

> Suddenly a comment 'feminazi, communist'. The feminazi thing has already become fashionable before, and it is still holding up. Out of nowhere a comment that 'you are a feminazi'. I don't feel insulted... More proud. In my profile it says that I'm a feminist. And of course, when they get into my profile and read that, well, they attack me with feminazi and so on (T-2).

Further examples of this humorous reappropriation can be found in media and culture, such as the video series "Feminismo para torpes" ("Feminism for dummies") by the national newspaper El País, with its first video titled "Am I a feminazi?"; or the book "Puta gorda, feminazi" by the comedian and feminist Penny Jay. This process can be compared to the reappropriation of other labels that were initially meant to be insulting or pejorative, like "queer" (e.g., Butler, 1997; Eichhorn, 2001).

Adding an intersectional lens

Both the interviews and the digital ethnography reveal that sexual orientation, gender identity, racialization, and religion exacerbate the abuse, as LGBT-phobic, racist, transphobic, and Islamophobic discourses often coexist with misogynist hate speech in attacks against individuals. Previously, examples of abuse against racialised women, trans women, and Muslim women have been drawn from the digital ethnography, indicating that their mere presence in mainstream channels or reaching mainstream audiences on social media often triggers attacks.

The interviews facilitated deeper reflections on the intersection of misogyny with racism and/or LGBT-phobia in the abuse. Narratives of racialised activists illustrate how misogyny and racism intertwine in attacks: "It started with questioning whether I could give that speech, and ended in some brutal, really brutal misogynist and racist attacks" (T-4). This aligns with experiences depicted in Gray's (2012) study, where online violence against Black women gamers turned racist and sexist upon unveiling their racialised identities. Similarly, as Jeong (2018) demonstrated, racialised women who changed their Twitter photos to that of a white man experienced a decrease in received hate speech, which shifted to a more benign tone.

Regarding the intersection of misogyny with LGBT-phobia, T-2 recounts how insults escalated when men in the gaming community discovered she was not only a woman but also a lesbian: "Once on Discord, I was playing on the server and they got super aggressive, well with insults of all kinds against my sexual orientation." This incident prompted her to report the individuals behind it to both Discord and the police.

The experience of T-1 offers another perspective on intersectionality. As an LGBT+ activist, her public engagements are consistently centered around LGBT+ issues, sometimes alongside a cis male activist colleague. T-1 can directly compare the reactions she receives with those directed towards her colleague: "I'm thinking about an

interview that we did for [a TV programme] and the comments were malicious, and I find curious the difference between how they attack him and how they attack me. Most of all, I think it's very much related to gender, what perspective they judge us from" (T-1). She clearly attributes the specificity of the hate messages she receives to the intersection of LGBT-phobia and misogyny, noting that her male colleague receives fewer and less aggressive messages, despite both advocating for LGBT+ rights.

Lastly, we must also consider intersectionality concerning the specific roles and positions of the women included in this research—women with public voices and profiles—and the intricate power dynamics at play. The misogyny evident in certain online abuse may not be universal to all women but rather targeted at feminist women, women who are outspoken, or those who challenge patriarchal and heteronormative structures (e.g., Lewis et al., 2019). The targeting of specific women suggests that this particular form of hate is selectively directed towards those who deviate from expected norms, rather than being aimed at all women. This specificity is also recognized by some of the women interviewed, who explicitly note the distinct nature of the violence targeting feminists. T-6 observes, "I think it's a very, very specific kind of hate that targets women who have a public feminist discourse". The perceived challenge to power relations posed by LGBT+, racialised, or Muslim women may be seen as even more pronounced when activism is added to their profiles. These women not only exist outside white heteropatriarchal norms but also assert their voices and advocate for rights for themselves and others.

Digital self-protection strategies

Throughout the research, spanning digital observations and in-depth interviews, a recurring theme emerges among women targeted by online gendered hate: the adoption of digital self-protection tactics. Those subjected to online hate speech seamlessly integrate general precautions into their digital social interactions. As a response to experiencing online hate speech and violence, these individuals modify their online behaviour to some extent, avoiding certain topics or reducing their activity on social media platforms. Additionally, they implement measures aligned with the capabilities of digital spaces to safeguard themselves. For instance, they refrain from sharing personal or geolocation information that could reveal their physical whereabouts to followers. These self-protection strategies are viewed as essential and are normalized to the extent that they are implemented and discussed without much contemplation. Thus, these strategies reflect the acknowledgment and normalization of online violence. A prevalent protective measure involves utilizing multiple accounts on the same platform. Typically, one account is entirely public, accessible for anyone to follow, while the other remains private, accessible only to trusted friends or associates. There is minimal overlap between these accounts to maintain the secrecy of the private profile from most followers of the public one. Additionally, various strategies on different social media

platforms, such as blocking, muting, filtering, or reporting, are employed as part of these protection mechanisms, which will be further explored in subsequent sections.

Blocking, muting, filtering, reporting

In addition to the broad digital protection methods discussed earlier, targets often employ platform-specific measures, which are nonetheless quite similar across different platforms. These responses vary based on factors such as the target's public visibility, their social media following, their experience in handling online attacks, and the severity of the attack. Among the most common of these strategies are blocking, silencing, muting users, and filtering out certain words. These tactics enable targets to prevent abusive accounts from further engaging with them. However, these actions are typically invisible to users not directly involved, except for the individual who initiates the block and the person who is blocked.

Blocking abusive or explicitly anti-feminist profiles on social media platforms is a prevalent strategy among the targets examined in this study. These measures are typically implemented in response to direct attacks, sporadic insulting comments, or when a target perceives that a profile is attempting to disrupt the conversation—a behaviour often termed as "sealioning". Blocking thus becomes an integral part of social interactions on platforms like Twitter. For instance, T-6 describes how she began blocking users who were attacking her: "I started blocking, one after another, all the persons that were insulting me. Maybe they were 500, I don't know, they seemed to be so many, so I spent like 6 or 7 hours during which the messages didn't stop, until I decided to cut it off." By "cutting it off," she means that she made her profile private and restricted non-followers from sending her messages. This action of privatising the profile is often the only effective strategy to halt collective and coordinated attacks (e.g., Morena-Balaguer et al, 2021).

Blocking is often accompanied by reporting to the platforms, as discussed further below. Some platforms, like Instagram, explicitly link these two options so that if you choose to block a user, the interface directly asks if you also want to report them. However, some targets believe that blocking may give haters or attackers a sense of accomplishment. Moreover, blocking a user who produces or spreads offensive content through comments does not entirely halt the abuse; it merely makes it invisible to the target or prevents that specific profile from sending hateful messages to the target (Smith, 2019). Essentially, it temporarily mitigates the impact on the individual target. Nevertheless, a user can easily create a new account or target other people with the old account even if it has been blocked by several other targets.

Some individuals prefer silencing, or as it is formally termed on Twitter/X, "muting," as it allows users to hide an account's tweets from their timeline without unfollowing or blocking that account. Additionally, some targets have incorporated into their protection strategies the practice of not quoting tweets or engaging in discussions with anyone who

attacks them. Instead, they opt to block the user directly to prevent the further spread of hate messages or the risk of instigating an attack by their own followers on that person. This decision to avoid arguments or quoting hateful content is seen as a way to avoid perpetuating a cycle of counter-attacks.

Another digital protection strategy involves filtering out specific words. This allows attackers to send messages and comments containing these words, but the targeted individual does not have to see them. Implementing filters or blocking users can also be automated, as evidenced by the experience of DE-13, who notes, "I would like to add that in my personal case this Twitter violence has reduced considerably since I use automatic blocking tools. I say this because even if you lose a lot of followers, visibility and there are accidental blocks, it is worth it. It's a matter of survival" (DE-13). While these strategies may not address hate speech or misogyny at a collective level, they do protect the individual target and contribute to a slightly more hospitable digital environment.

Reporting to the platforms

This section delves into the utilization of social media platforms' reporting functions, with a particular focus on Twitter, as the digital ethnography primarily centres around interactions on this platform. Overall, reporting to the platforms is a strategy seamlessly integrated into the digital protection measures of the targets. However, their accounts frequently express disillusionment with how the platforms manage their reports and consequently moderate content:

> There have been so many times that we have reported people with openly racist discourses, like saying that all immigrants or all Muslims must be killed, and nothing happens. You don't trust it anymore (...) At the beginning of using Twitter, I reported more, but now what I do more is blocking. And that is a bit to save yourself, because [when you report] you get the report that it doesn't break the rules and then you see people who haven't broken the rules but got reported and they get the account closed... (T-9).

This perception remains consistent regardless of the platform, with comparable experiences recounted across various platforms. These include conventional social media platforms like Instagram, Facebook, and Twitter, as well as video platforms such as YouTube. Additionally, platforms primarily associated with the gaming community, like Discord or Twitch, also feature similar narratives:

> It's like Discord, you report, and nothing happens. On Discord, I reported loads of times. They said they'd look at it and it's been a year. So I haven't seen any punishment, when I've reported I haven't seen any punishment, so it's normal that they do it with this impunity and they keep doing it because there's no punishment (...) So, they put the code of conduct and that shit just to look good, and because someone asks them to because it's a social network, but they don't do anything. I already told you that they do nothing. It's been the same all my life, reports to the platforms, and they don't cancel accounts or do anything. They tell you that they

are going to do it so you will shut up. Then you check and the server is still there, the account is still there (T-2).

The women interviewed strongly believe that platform operators are aware of the need for a code of conduct but are failing to enforce it effectively to safeguard those in need of protection. This failure has led to a widespread lack of trust among targets of online hate. Typically, after submitting a report to a platform, the most common response is a notification stating that the reported user or publication does not violate community norms. This leaves victims of online hate with a sense of impunity and a perception that decisions regarding norm infringement are highly arbitrary.

For any platform that relies on user-generated content and moderation measures, targets of gendered hate perceive these measures as ineffective, opaque, and arbitrary. This arbitrariness is evident when feminist accounts are reported by anti-feminists or far-right profiles and subsequently shut down. Consequently, there is a belief that the reporting system is not designed to protect them but can just as easily be used against them. This interpretation reflects broader societal structures that perpetuate inequality. It exacerbates distrust in platforms and fuels the perception that platforms lack genuine intent to protect vulnerable groups.

Despite feelings of impunity and arbitrariness, most targets continue to report instances of abuse and urge their friends and followers to do the same. However, not everyone opts to utilise the existing platform reporting mechanisms. Among the interviewed targets, one woman explains that she refrains from reporting, even to the platforms themselves. She attributes this decision to the normalisation of hate messages within the broader LGBT+ community to which she belongs. Like most response strategies, reporting and implementing self-protection measures unfairly burden targets, a burden exacerbated by platform operators' limited interest in establishing transparent measures to protect particularly vulnerable groups (Jane, 2017). Misogyny often falls into a gap where reported publications are not categorised as hate speech, likely due to the general exclusion of misogyny from the hate speech umbrella.

Furthermore, the specific intersectional form of misogyny, which overlaps with abuse on other grounds such as racism or LGBT-phobia, is not adequately addressed by existing reporting mechanisms. Platforms typically allow only one category to be chosen for reporting hate speech, posing a dilemma for targets facing multifaceted abuse. This limitation fails to accurately reflect the complexity of the abuse, or the protection needs of the targeted individual. The inability to select more than one hate speech category, coupled with platforms' lack of response to gendered hate, can convey to targets that they are undeserving of protection.

While digital protection strategies may not tackle the root causes of the issue or directly deter individual perpetrators of hate speech, effective mechanisms for blocking profiles and reporting content can at least help mitigate the impact on targets.

Ultimately, blocking, silencing, or reporting users may prove insufficient to halt the onslaught of abuse, especially during a large-scale attack. Moreover, these actions can be extremely time-consuming, requiring targets to block attackers individually. When faced with overwhelming abuse, targets may resort to making their profiles private as a last resort. This action is commonly referred to as "poner el candado" ("putting on the padlock"), where non-followers attempting to access the profile will see a padlock icon instead. On platforms like Twitter, making an account private restricts post visibility to only those who were following before the account was locked. This strategy is often employed temporarily until attacks subside. Turning an account private serves both as a protective measure and a subtle form of protest, as the change is visible to other users. Although users may not know the specifics of what occurred, the sudden shift to a private profile signals that something significant has transpired with the user (Morena-Balaguer et al, 2021).

Reporting to the police

Ample evidence suggests that both hate crimes and gender-based violence are significantly underreported to the justice system. For instance, the Survey Report on Hate Crimes by the Spanish Ministry of the Interior (2023) indicates that 89.24% of hate crime victims did not report their experiences to the police. When analysed by aggravating circumstances, 87% of victims targeted based on sexual orientation or gender identity and 90% of those targeted based on sex or gender did not report the crimes. In the digital realm, violent expressions are highly normalised, and reporting to the police is even less common, with official statistics being particularly scarce.

In this section, we will explore the experiences of those who have reported to the police, as well as delve into the reasons women provide for not seeking support from the formal justice system when victimised by online violence.

Experiences of reporting to the police

To begin with, it is important to note the evident scarcity of experiences from women who have reported online violence to the police, as revealed by the data collected during this research. Among the women interviewed, only two had attempted to engage the criminal justice system by reporting online gendered hate to the authorities. This finding aligns with prior qualitative research (see, e.g., Smith, 2019; Eckert, 2017; Mantilla, 2015). Smith (2019: 290) concludes that "where such attempts had been made, the experience had been disheartening and had discouraged them from adopting this form of response in the future". Additionally, in the digital ethnography, reporting to the police was noted in two cases. While additional women might have reported without publicly disclosing this on their social media profiles, police reporting is frequently suggested by Twitter users in comments as advice on how to proceed, particularly in severe cases.

Regarding the experiences of women who have reported to the police, T-2, a feminist gamer, recounted attacks on Discord targeting her as a lesbian woman:

> Once on Discord on the server, while playing, they became super aggressive, with all kinds of insults towards my sexual orientation. I reported them to Discord and went to the police, but even the kid himself laughed at me and said, 'report, report, hahaha.' And it's true because nothing happens. I went with image captures, but they didn't give me many answers. In other words, the police really don't do anything about these things (T-2).

T-4, an anti-racist activist, had been targeted on two occasions by a mix of racist and misogynist threats. Initially, she did not perceive a real threat to her physical safety as the attackers seemed far from her hometown. However, during new attacks, she realized that the perpetrator lived in the same autonomous community and might know how to find her, as she had been quite open about her location. This fear prompted her to report to the police. In both cases, the outcome was not what the women expected. Although the police filed reports, they did not follow up, and as far as the women know, no investigation was opened.

Similarly, DE-18, a feminist journalist and activist, experienced a combination of online and offline harassment. In addition to direct verbal abuse on social media, her phone number was published on Twitter, she was added to various WhatsApp groups, and an advert was placed in her name on an adult contact page, resulting in her receiving sexual photos from unknown men. Offline, someone painted insults outside her workplace. Despite consulting with a lawyer specializing in gender, her reports have not led to any action. The police failed to consolidate the various attacks against her into a single report, even though the attacks were clearly linked and targeted her as a journalist and feminist. This case illustrates the omission of a gender perspective in crimes not committed by a woman's partner or ex-partner, as per the Spanish legal framework. Additionally, most of the attacks were perpetrated by multiple men, adding complexity to reporting and addressing these crimes within the criminal justice system.

In summary, while only two women among those interviewed had reported online gendered hate to the police, their experiences reflect a broader trend noted in previous research: disheartening experiences with the justice system that may discourage future reporting. Looking at all the experiences that led these women to report, several common factors emerged. First, there was an intersection between various motivations behind the attacks, such as gender combined with LGBT-phobia or racism. Second, the perceived gravity of the harassment made the women fear that the online attacks posed a direct physical threat, linking their reporting to the need to protect their physical safety. Reporting is thus often associated with feeling a threat to one's physical integrity or identity.

A significant issue in these cases is the lack of information or updates from the police, leaving the victimized women without answers or follow-up, and a sense of impunity despite reporting. This lack of response reinforces the perception that such crimes are not taken seriously by the police and may contribute to a lack of trust in law enforcement and judicial authorities. For women reporting misogynist hate crimes or hate incidents, it is crucial to feel heard, receive reassurance from the police, and sense that their

concerns are treated seriously (cf. Mullaney and Trickett, 2018). Even when women understand the police may have limited ability to locate and act against the perpetrators, feeling heard, reassured, and taken seriously by law enforcement remains vital.

In the next section, we will discuss the narratives of women who chose not to report to the police.

Reasons linked to non-reporting

The reporting of online violence remains infrequent, even when the abuse includes direct threats. This observation aligns with previous research and official statistics. To delve deeper into this issue, part of the interviews aimed to understand participants' reasons for not reporting such incidents.

The case of the youngest interviewee (T-3, 22 years old) exemplifies several reasons for not reporting. T-3 had encountered various forms of verbal abuse, including direct threats. When asked if she had reported any of this abuse to the police, she admitted that it had not even crossed her mind. She elaborated: "I mean, I'm not that well known, and my threats haven't reached a level that would intimidate me to the point of changing my routine. In other words, these are anonymous profiles that don't know who I am, and that's it. Nothing unusual has happened to me in that sense" (T-3).

Similarly, T-6 explained that she never felt any of the attacks warranted police involvement:

> "Luckily, they haven't threatened to kill me, control me, or anything like that, which has happened to some colleagues (...) It's more about attacking my self-esteem and personal integrity rather than my physical integrity, so fortunately, I haven't had to report it. You can't report this—I wish you could report someone for sending their followers after you, knowing their reach and responsibility... There must be consequences for inciting hatred. I wish it could be reported, although I doubt I would, given what we know about complaints. I wish I lived in a system where this was possible because clearly, there is damage" (T-6).

Both stories illustrate how the severity of abuse and threats is often downplayed, despite the profound impact of verbal attacks. T-3 does not believe the threats are genuine since the attackers seem unaware of her identity outside social media, thus she does not feel the need for physical protection. On the other hand, T-6 mentions not having received any direct threats to her physical safety. While deeply affected by widespread verbal abuse, she feels there is no concrete evidence (such as a death threat) to report to the police. There is a prevalent belief among those targeted that only certain types of online violence are reportable to the police—specifically, violence perpetrated by a partner or ex-partner (within the context of gender violence) or direct, explicit, and severe threats by strangers. Moreover, the interviewed women feel that these threats must pose an imminent risk to their physical safety to be taken seriously.

T-3 also cited a significant reason for not reporting—distrust in the police:

And besides, I don't know if I would report it to the police because I have the feeling they would laugh in my face if I did… I mean, I have screenshots of someone saying 'I hope they rape you' and 'If I see you on the street, I'll do something to you.' I could report it, but I think they'd laugh at me. Like testimonies on social media, where women report and are ridiculed (T-3).

Previous research has extensively documented this mistrust as a barrier to reporting hate crimes (see, e.g., Perry, 2001; Awan and Zempi, 2015; Langarita et al., 2018; Paterson et al., 2018; Domínguez Ruiz et al., 2023). Collectives targeted by LGBT-phobia or racism, for instance, may have encountered discrimination from law enforcement themselves, which is also true for misogyny.

T-3's belief that the police would not take her seriously is partly based on social media accounts of other women who reported digital violence and received dismissive responses. While the participants in this research who did report were not met with overt indifference, the lack of subsequent action and feedback made them perceive law enforcement's commitment to addressing online violence as minimal.

In some narratives, the decision not to report is also tied to the normalisation of online violence. If an individual is subjected to hate messages daily, it would take something extraordinary for them to approach the police or use other reporting channels. This normalisation may prevent women from realising that such incidents could be addressed by the justice system. This link between normalisation, repeated victimisation, and underreporting has been highlighted in discussions about hate crime reporting, particularly among LGBT+ individuals (see, e.g., Chakraborti and Hardy, 2015; Paterson et al., 2018).

Lastly, T-5's explanation reveals another reason for not reporting, related to managing the online environment. T-5 acknowledged that some messages could have been reported, but these were deleted by friends who filtered her messages to protect her:

Many of the messages that might have constituted a crime—like threats of 'we are going to kill you'—were deleted by my friends. In the midst of a massive influx of messages, they needed to filter them, so I never managed to report anything (…) At that moment, I didn't see reporting as a useful option. What I needed most was to disconnect from technology, be at home, and let my loved ones take care of me while others filtered out the atrocities. I didn't see reporting as an effective solution to the crisis I was experiencing (T-5).

An immediate response strategy, which felt necessary at the time, inadvertently hindered another potential strategy. If she had wanted to report later, the evidence was already deleted. Nonetheless, at that moment, she did not view reporting as a helpful course of action. Reflecting years later, she now recognises some messages as crimes that could have been managed by the justice system, but she prioritised other needs at the time and did not see institutional support as beneficial.

Conclusion

Women anticipate online hate after public appearances, with media coverage triggering waves of abuse. Hate intensifies when addressing feminist issues, aiming to discredit and degrade feminists. Attacks often include explicit rape or death threats. Perpetrators may also share victims' content without consent, escalating the violation across various platforms. Attacks persist over time, with continuous streams of hate messages contributing to ongoing victimization. These attacks often originate from profiles with significant social media followings, exacerbating the impact. Perpetrators may also organise attacks, suggesting premeditated efforts, sometimes associated with political affiliations.

Tangible, embodied harm is inflicted, blurring boundaries between digital and non-digital realms. The frequency and magnitude of abuse exacerbate its impact. The collective impact is significant, creating a chilling effect on feminists and women with a public profile. At the same time, the challenge of "going offline" is notable: professional obligations hinder disconnection, and disconnection involves a risk of exclusion from public discourse.

To protect themselves, women targeted by online attacks put different digital protection measures into place, including blocking and silencing users, as well as reporting users to the platforms. Reporting to the police is less common, aligning with the existing underreporting for hate crimes and hate speech in general. Very few women report, and those who do so do not get the outcomes they might have wished for. As Powell and Henry (2017: 273) note, "'justice' gains meaning through the opportunity to recount one's victimisation in a public hearing, have that account validated by peers, and have action taken to hold the perpetrator accountable, correct the harm, and/or prevent its recurrence". Those who reported might have hoped for such justice. However, whether they report or not, the prevailing sentiment among victims is that reporting to the police leads nowhere, and law enforcement does not take gendered online violence seriously. This leaves victims feeling impunity prevails, with self-protection and peer support as their only options.

5. Professionals' perceptions of online violence and their preparedness to provide support

Introduction

In the realm of anthropological research, the examination of digital spaces necessitates an understanding of their integration within the broader societal, cultural, and subjective structures of lived experience. This perspective illuminates the ways in which behaviours in digital contexts are intertwined with everyday practices and embodied knowledge. Such an approach challenges the notion that digital technology inherently generates new forms of violence. Instead, it highlights that the structural roots of these behaviours are consistent with those underpinning gendered violence in other settings (Dodge and Spencer, 2017). Consequently, it is crucial for institutions and organizations dedicated to combating gender-based violence and hate crimes to comprehend both the mechanisms and impacts of gendered violence in online spaces and the underlying structural causes of this violence.

Despite the importance of this understanding, there is a notable scarcity of studies exploring the perceptions of practitioners and professionals who interact with victims of gender-based online violence or online hate speech. A significant exception is the work by Powell and Henry (2016), which investigates police and service sector perspectives on managing technology-facilitated sexual violence. Their research aims to elucidate the nature of the harms involved and the challenges associated with both legal and non-legal responses. Specifically, there is a lack of research examining the attitudes of Spanish practitioners towards gendered online violence. This working paper aims to address this gap by analysing the perceptions and categorizations of verbal gender-based violence in digital spaces within contemporary Spain. While gender is recognized as a protected characteristic under Spanish hate speech and hate crime laws, and gender-based violence is considered a "classic" form of hate crime (Perry, 2001), in practice, gender-based hate speech is seldom addressed within this legal framework. This is particularly true for verbal violence in digital spaces, which is often perceived as less serious due to the online environment being seen as a mitigating factor, an unmonitorable space, and a barrier to legal prosecution.

The analysis presented in this paper is based on ethnographic interviews with professionals who support victims of gender-based violence and anti-LGBT violence from various perspectives, as well as professionals from the justice system and law enforcement agencies. The focus is on these professionals' perceptions of the online context, their categorizations of gendered online violence, and their preparedness to respond to such violence. Through these interviews, the paper explores the nuanced understandings and categorizations of verbal gender-based violence in digital spaces by Spanish practitioners. It reveals the challenges they face in addressing such violence and

underscores the need for a more robust framework to tackle gender-based hate speech online. By filling this research gap, the paper aims to contribute to a deeper understanding of the complexities of gendered violence in digital spaces and inform the development of more effective strategies and policies to combat it.

Perceptions and categorisations of gendered online violence

As a starting point, many professionals interviewed for this study predominantly conceptualize online sexist violence within the context of heteronormative couple relationships. Specifically, they view it as a form of gender violence perpetrated by a man against a woman who is or has been his partner. This narrow understanding aligns with Spain's current legal framework, which defines gender-based violence primarily as intimate-partner violence. Consequently, a wide range of services from both public institutions and NGOs are designed to support victims of gender violence as legally defined. During fieldwork conducted, contacts with various public institutions across Spain consistently redirected the researchers to services dedicated to victims of gender violence. These services, although increasingly aware of technology-facilitated violence, are constrained by limited resources and thus cannot heavily focus on digital violence. Moreover, they typically address only intimate partner or ex-partner violence. Despite these limitations, interviews with professionals from these services provided valuable insights into their perceptions of various types of online violence and its impacts. These insights highlight the broader challenges in addressing the full spectrum of gendered online violence within the current legal and resource constraints.

Public services that support victims of gender violence generally recognize the dimensions of this violence in digital spaces. Both public institutions and civil society organizations acknowledge the importance and pervasiveness of these spaces, especially in the lives of young people. As one representative of a public service for women noted, "This is a kind of thing where you are constantly connected to your mobile... It's a violence that doesn't stop, 24 hours a day, every day of the year."

While most services focus on intimate partner violence, many service representatives also understand that digital violence extends beyond intimate relationships. This awareness points to the need for a broader approach to address all forms of gendered online violence effectively.

> Here we also differentiate between online violence received from someone known, and violence received by someone unknown or anonymous. We find the most cases of someone known, but we know that there is violence through persecution on social media, through platforms such as Wallapop[1], Instagram, Facebook, that someone starts to follow you and coerce you. We don't have it quantified, unless a woman comes and identifies very clearly that she is suffering from online violence. But I think that the aspect that we could find the most is the

---

[1] A mobile application to sell second-hand items.

digital element as another instrument used by a known aggressor or the abuser (Representative of a public service for women).

Representatives of services also recognize the possible impact of online violence on the individual victim:

> The level of violence is very extreme, it's very high and it has a big impact. We have that idea that physical violence is the one that generates the most impact, and obviously it does, but there are other, slightly more invisible forms of violence that can play a very powerful role (Representative of a public service for victims of sexist violence).

Despite recognizing the potential impact of online violence, professionals often refer to a continuum and hierarchy of violence. Within this hierarchy, physical violence is seen as "real violence" and is perceived as having the highest impact. This aligns with Harris and Vitis' (2020) observation that institutions and professionals from whom women seek assistance—such as platforms, tech companies, and justice agents—often view online violence as separate from or "less serious" than offline violence (see also Citron, 2009; Harris, 2020). Powell and Henry (2016) have similarly highlighted how law enforcement responses tend to minimize the harms of gendered violence when it occurs online or through email and text-based communications. As Powell and Henry emphasize, this minimization reflects a broader tendency within institutional responses to de-prioritize digital forms of gendered violence:

> the minimisation of digital forms of abuse, violence and harassment perhaps stems from a fixation on corporeal or bodily forms of harm, as opposed to harms that are structural, social, emotional or psychological. Conventional conceptualisations of criminality require an overhaul in both theory and practice in order to address relations of power and exploitation that are reproduced in digital contexts, and which can potentially serve to destroy another person's life (2016: 11).

Gender-based violence, whether physical or psychological, is the primary focus of the services designed to address it, often utilizing limited resources. However, these services typically lack specific training or formal protocols for handling online violence. In some cases, they do not even categorize incidents occurring in digital spaces within their internal systems.

One representative from a public information service for women explained:

> We would mark the item as sexist violence and look for psychological, physical, and economic support, but we don't have a category for quantifying violence that occurs through mobile phones or online. We know it exists, and in cases where psychologists or social workers are working with the women, WhatsApp messages and pressures like 'where are you?' are mentioned in interviews. But we haven't quantified it. We don't know how many cases we address each year.

This lack of categorization not only leads to a dearth of statistics but also hinders the preparedness of professionals to handle online violence cases effectively. When digital

space violence is not formally recognized in public services, it remains invisible, preventing the development of tools, protocols, or specific training necessary to equip practitioners with the knowledge and skills needed to support victimized women adequately.

Several obstacles hinder the provision of support to victims of online gender violence and other types of gendered online violence, beyond intimate partner violence. Firstly, the vast majority of services are tailored to victims of gender violence as defined by the Spanish legal framework, which is predominantly heteronormative. A representative from a local institution working on feminism and LGBT+ rights noted:

> I started working in the department three years ago and found an obviously very heteronormative approach to violence services, because it has a historical logic and a legal logic... I remember strong discussions with teammates, saying 'but let's see, not all women have the same sexual orientation', and that has forced us to analyse what gender violence we are leaving out or not attending to.

Some local public services, particularly in Catalonia, have begun to expand beyond strictly heteronormative definitions of gender violence. They are adopting a broader, more intersectional approach that includes considerations of gender and sexual orientation. This shift aligns with the autonomous law on gender violence in Catalonia, which better conforms to the Istanbul Convention compared to the broader Spanish framework.

Secondly, the online dimension itself poses a significant challenge to supporting victims. Interviews with professionals reveal that online violence is perceived as a major hurdle. Many professionals express doubts about victimization in digital spaces, largely due to the perceived immensity and complexity of the internet. NGOs, which often take a more proactive stance on various forms of violence, find it challenging to monitor violent expressions online due to resource constraints. A representative of an LGBT+ association explained: "What happens with the Internet? It is true that we have a more, so to speak, passive attitude. That is, it has to come to us, we don't have the resources to go looking for incidents online".

This passive stance may be linked to a perceived hierarchy of violence, where physical violence is seen as "real" violence. Despite this, professionals and civil society representatives also express uncertainty about the boundaries of freedom of speech in digital spaces. The threshold for what constitutes hate speech online is perceived as much higher than for face-to-face incidents, as the latter are seen as carrying a greater immediate risk of physical violence. However, recent years have shown that misogynistic discourses in online forums, particularly those linked to the manosphere and incel forums, can lead to physical violence (see e.g., Barcellona, 2022). Therefore, even when focusing solely on physical violence, it is crucial to consider also violent online expressions to fully understand the ramifications of heteropatriarchy and the interconnectedness of online and offline environments.

The perception of the online context as less important and too complex to control or respond to not only limits professionals' understanding of what constitutes violence but also discourages victims of gendered online violence from coming forward and reporting their experiences. This can lead to unrealistic advice from professionals meant to provide support, as we will discuss in the next section.

Professionals' level of preparedness to respond to online violence

Among the professionals and practitioners interviewed, there is a persistent perception of the online and offline worlds as fixed, binary categories rather than as a continuum of highly intertwined spaces. Deuze et al. (2012) emphasize that life is now lived in media rather than merely with media, yet a comprehensive understanding of this integration is still lacking. This gap in understanding is evident not only among professionals responding to incidents of violence but also within society at large. This contributes to a perception of the online world as a distinct space that those targeted by online violence can simply exit.

Powell and Henry (2016) argue that more attention needs to be given to the inherent and continuous links between online and offline interactions and communications. An individual's social world spans both spheres, making it crucial to understand how experiences in one sphere affect the other. When digital and physical spaces are perceived as separate, with greater importance placed on physical interactions, the significance of online experiences is downplayed. This perception often leads to unrealistic advice from professionals on how to avoid online attacks. For instance, a representative of a public care service for victims of trafficking noted their approach: "The guidelines we give them are to eliminate all their digital footprints—no Instagram, no Facebook, change the phone card, remove the phone's location settings. We try to suppress absolutely any possibility of locating them through any device." Such advice, while well-intentioned, overlooks the deep integration of digital and physical lives and the impracticality of completely disconnecting from the online world. This perspective fails to acknowledge the pervasive nature of digital spaces and the challenges in navigating safety within them.

In cases of digital gender-based violence, professionals often feel that the only advice they can offer is for victims to go offline. This can be particularly difficult for young individuals whose lives are deeply integrated with social media and for women who rely on their online presence to generate income. Although the professionals interviewed are aware of the challenges of disconnecting, their lack of in-depth knowledge and specific training about online protection mechanisms leaves them with limited options. Their primary objective is to protect victims from being tracked, leading them to prioritize physical safety over addressing online violence. This approach reflects a concern primarily with physical threats, with physical violence being seen as "real" violence, while online violence is not perceived with the same level of seriousness. As a result, the

responsibility for protection falls on the victims, who are often advised to disengage from social media entirely:

> We tell them that if they open another profile on social media again, because many are young women who use Instagram or TikTok, and now it is very difficult not to be present on social media... The guidelines are always not to take any photographs with elements that can locate or identify them. They need to be very prudent to avoid new re-victimizations in the future (Representative of a public service for victims of gender-based violence).

Powell and Henry (2016) have noted that advising victims to disconnect can have adverse impacts, such as extending the perpetrator's control and denying the victim digital citizenship (see also Henry and Powell, 2014). Viewing self-protection as the only viable mechanism to avoid attacks—essentially making the profile private or leaving social media—can silence certain voices, particularly activists involved in digital debates. Those most at risk of being silenced are individuals experiencing multiple forms of oppression, such as racialised or LGBT+ women.

Perceptions of responses by the police and the justice system

Civil society organizations focused on feminism and LGBT+ rights often express doubts regarding the online dimension, particularly concerning the limits of freedom of expression in digital spaces. There exists a complex balance, especially online, between behaviour that may constitute a crime and incidents that could be treated as administrative infractions. A representative of a feminist association emphasized this complexity: "We think we have to polish and see what we are talking about when we talk about online hate incidents because this also leads to a much deeper debate that has to do with freedom of expression". This ambiguity, coupled with the perceived difficulty in monitoring digital spaces, leads to reluctance in reporting incidents to the police. Additionally, there is a general perception among organizations and institutions that cases of digital violence, particularly online hate speech, are challenging to prosecute due to territorial limitations, that is, when the account that spread hate speech is located in another country.

Some professionals find their uncertainties about the online context reinforced through consultations with the police. A representative of a public anti-discrimination service highlighted that "the police themselves have recognised many times that tracking and reaching the origin of many of these discourses or situations of virtual discrimination is difficult, it's not easy". This perception of online crimes as inherently challenging to prosecute may lead professionals to hesitate in encouraging victims to report incidents, fearing a futile process.

Furthermore, professionals express doubts and insecurities about collecting evidence if a victim wishes to report digital violence to the police:

If you are going to report, from what perspective do the police advise on what to do? Because, of course, it's a crime, but between having evidence, that your profile on social media is being questioned, if you have published a photo like that and so on... Then we don't know very well how someone should have the mobile phone programmed so that it really becomes forensic evidence, how is that done, do women know how to export those conversations, are screenshots useful or not? (Representative of a public support service for women).

This uncertainty extends to the treatment victims may receive from the police and adds an additional layer of complexity regarding what can and should be reported and what information constitutes evidence in digital violence cases.

Interviews with legal experts shed light on the challenges of integrating hate speech or gender-based digital violence into police work:

You either have police forces specialised in gender violence, who primarily focus on violence or abuse by partners or ex-partners—a purely physical issue—or you have teams specialized in hate speech or cybercrime, who primarily deal with cybercrimes such as big tech scams. They don't usually consider attacks on public figures or gamers through social media as significant. To them, these cases are minor or irrelevant (Legal expert).

This statement highlights two key issues complicating the classification of online violence or hate crimes within police work: the online dimension and the lack of recognition of gender crimes as hate crimes or gender-based violence beyond intimate partner violence and anti-LGBT hate crimes. A conversation with a police unit specializing in hate crimes further illustrates this challenge. According to representatives of this unit, they can recall only one case of online hate speech against women, which involved a hate website targeting women, reported by a women's rights institution: "We have encountered cases, especially related to gender, where individuals have created websites with the intent of denigrating women or promoting male supremacy over women. In these cases, it was evident that a crime had been committed" (Police specialized in hate crimes).

In essence, the application of hate speech frameworks based on gender is limited to cases of hatred against the entire group of women, rather than individual attacks on women grounded in gender hate. Furthermore, according to the Spanish criminal code, the police classify as "gender-based violence" only events occurring within heteronormative relationships.

While civil society organizations have expressed doubts about the boundaries of freedom of expression, it's also challenging for the police to delineate these boundaries. There's a general perception that the circumstances must be rather serious to warrant an investigation:

I'll give you an example: some insults directed at an individual from the LGBTI community or a woman of recognized prestige, or insults targeting a woman for her gender, might be considered more of an administrative matter... For it to be

considered a criminal offense, it has to go beyond mere insults. In the case of LGBTI individuals, some insults might be subject to administrative sanctions based on civil laws, but when it comes to insults against a woman for being a woman, I'm not sure... (Police specialised on hate crimes).

As indicated, attacks against LGBT+ individuals may be subject to administrative sanctions in certain autonomous communities in Spain, like Catalonia. Therefore, the potential for reporting an online attack perceived as identity-based may depend on various factors. However, the online context presents additional obstacles for administrative processes, particularly related to territoriality and the jurisdiction of administrations, especially in identifying offenders behind anonymous accounts.

Moreover, professionals within the justice system also find online cases highly challenging. A prosecutor highlighted the difficulties in identifying the individuals behind anonymous profiles:

> The main challenge with false or anonymous profiles is identifying the person behind them. Most social media platforms are legally based in the United States... If we want to identify the owner, we have to go through a legal process with the United States. However, US judicial authorities often cite the first amendment and refuse to provide the information.

This difficulty in obtaining cooperation from Internet service providers has been previously noted by Powell and Henry (2016), who highlighted how it hampers obtaining evidence for legal action, thereby creating barriers to achieving justice for victims. Additionally, police have noted an increasing demand for forensic services to analyse electronic evidence and hardware, underscoring lawmakers' incomplete understanding of cybercrime. In the Spanish context, Igareda et al. (2019) found that legal operators often lack understanding of the evidence collected in digital spaces or lack technical expertise to utilize it effectively. Furthermore, cases involving "sexist cyberviolence" are often not brought before specialised courts dealing with gender violence, leading to arbitrary judgments. This ignorance can further victimize women, as they may be blamed for what happened. Moreover, victims themselves often downplay the violence or fail to report or retain evidence.

It is crucial to recognize that policing and criminalizing behaviours may not be the sole solution to addressing all forms of online violence. Scholars such as Coburn et al. (2015), Hasinoff (2015), and Shariff and DeMartini (2015), as summarised by Dodge and Spencer (2017), have raised concerns that criminal law approaches offer limited potential to change the cultural beliefs that underpin sexist, transphobic, racist, and homophobic harassment more broadly. This concern is also echoed by Powell and Henry (2016) regarding technology-facilitated sexual violence. However, as noted by Mullaney and Trickett (2018), the primary motivation for women to report misogynistic hate crimes to the police is not necessarily conviction but rather the acknowledgment that their concerns are taken seriously: "Officers need to ensure that they continue to give

empathetic and sympathetic responses to female victims and take them seriously" (Ibid: 55).

In the upcoming section, we will delve deeper into the training gap and explore potential pathways to better equip professionals with knowledge about digital spaces and enhance their ability to support victims of gendered violence in online environments.

Conclusion: Training needs and possible ways forward

Upon analysing interviews with professionals, it becomes evident that both public and non-governmental sectors lack training on safeguarding individuals vulnerable to online attacks and responding to online incidents or crimes. A representative of a public information service for women emphasizes the necessity for extensive training across various sectors involved in assisting women affected by such violence: "A lot of training work will have to be done, both to the services that provide attention to women victims of this violence, such as the legal field, the police field, the lawyers on duty". Professionals are acutely aware of this training deficit, often attributing the lack of knowledge to a generational gap. For instance, one representative from a public care service for victims of gender violence acknowledges the digital divide: "I, for example, am older, and the whole digital issue is not my forte. There really is a significant gap in the age that I am, of some women that we attend to who are perhaps 18 or 22 years old..."

To bridge this gap, proper training is essential. Currently, in the absence of protocols or formal training, many professionals rely on informal knowledge gleaned from younger team members who are perceived to be more familiar with social media platforms. Even within services supporting victims of trafficking, which heavily relies on digital spaces, professionals admit to lacking formal training on digital protection mechanisms. However, training is not only needed to address the online context but also to broaden perspectives on other forms of violence. As a representative from a local institution focusing on LGBT+ rights highlights: "One of the issues that came up in a meeting between several services was the need for training for professionals so that they have a broader view of violence..."

In Catalonia, recent modifications to regional laws against gender violence, such as Law 17/2020, aim to include digital violence within the scope of gender-based violence recognised by the law. While some view this as a significant step forward, others see it as merely symbolic, given that criminal jurisdiction rests with the state. Nevertheless, such laws could aid in collecting statistics on gendered online violence, offering victims access to support mechanisms available for victims of gender violence. However, existing support services are already strained and would require additional resources to accommodate victims of online violence.

To conclude, our research suggests that the perception of the online-offline continuum as a binary division between the "real" and the "virtual" contributes to downplaying the severity of online violence, despite its significant impact. This perception normalises

gendered online violence while relegating it to a lower position in the hierarchy of violence compared to physical violence. Consequently, support for victims of gendered online violence is often insufficient.

Professionals in Spain are largely unprepared to support victims of gender-based online violence due to a lack of specific training and protocols. At the same time, the rapid pace of technological change calls for a constant updating of knowledge to effectively address evolving digital spaces. This challenge is compounded by resource constraints and a general undervaluing of events occurring in digital spaces unless they pose an immediate physical threat.

Framing gendered online violence within a hate crime framework helps elucidate its structural roots and collective silencing effects. Viewing it on a continuum of violence reveals its interconnectedness with other forms of structural violence, even if not explicitly defined as criminal offenses. To combat this violence effectively, a multi-level networked response is necessary, encompassing policing, legal measures, education, digital protection, and collaboration with social media platforms. Additionally, tailored support, including legal and psychological assistance, must be provided to individuals targeted by online violence. Constructing this networked response requires equipping professionals with the knowledge to address the violence structurally while providing empathetic support and practical advice to victims.

6. Civil society perceptions of online hate speech and digital violence

Introduction

As part of the COME-ON research, an exploratory survey was elaborated with the aim to create a better understanding of the perceptions of non-governmental organizations and associations in Spain with regard to digital violence and online hate speech. In the survey, the following definitions were provided in relation to the aforementioned concepts:

"Digital violence" was defined as any act of sexist violence and misogyny in digital spaces. "Online hate speech" was defined as the promotion or incitement, directly or indirectly, of hatred, hostility, discrimination or violence against specific groups, in digital spaces.

The purpose of this survey was multifaceted. Firstly, it sought to gauge the awareness and understanding of digital violence and online hate speech among these organizations. Secondly, the survey intended to identify existing measures and strategies employed by these organizations to combat digital violence and hate speech. Lastly, the survey aimed to gather insights into the support needs and resource gaps that these organizations face in addressing these issues effectively.

The findings from this survey are expected to contribute significantly to the broader discourse on digital safety and human rights in the digital age. By highlighting the challenges and needs of non-governmental organizations and associations, the research aims to inform policy recommendations and the development of targeted interventions. Ultimately, the goal is to foster a safer and more inclusive digital environment, where the rights and dignity of all individuals are upheld.

Methodology and sample

The survey questions were divided in three sections:

1. Sociodemographic information, including age, gender, and autonomous community;
2. Information about the organization, including field of work, function in the organization and of the person responding to the survey, years active in the organization, whether the organization provides support to victims and if so to victims of what types of crimes, and what type of support the organization provides
3. Perceptions about digital violence and online hate speech. This third and main part of the survey included a combination of Likert scale questions on perceptions (e.g. In your opinion, to what extent is digital violence a problem that needs more attention in Spain") with response options ranging from "Not at all" to "Completely"; multi-choice questions; and open-ended questions. The questions

will be further explored and discussed in the presentation and analysis of the results, below.

The survey was set up on SurveyMonkey and was open for three months, from mid-June to mid-September, 2022.

To define the universe from which to draw the sample, we mapped civil society organizations across Spain, with a focus on feminist- and LBGT+ organization. This was done through searching for publicly available lists of national, regional and local organizations and associations. 307 relevant organizations from across Spain were invited by e-mail to participate in the survey. Whilst the original list of organizations was larger; we only count those organizations that had a valid e-mail address when the invitation was sent.

44 organizations responded to the survey. Among these, 19 responded that they were active in the fields of feminism and gender-based violence, 13 were active in the field of LGBT+ rights, 5 were active in the field of anti-racism and 5 in other fields, including migration, human rights, housing and social exclusion, and the rights of disabled persons. Despite inviting organizations from all autonomous communities, the responding organizations represented 13 different autonomous communities in Spain, including Andalucía (2), Aragón (1), Canarias (1), Cantabria (2), Castilla y León (2), Castilla-La Mancha (3), Cataluña (14), Comunidad de Madrid (8), Comunidad Foral de Navarra (5), Comunidad Valenciana (2), Extremadura (1), Islas Baleares (1) and Región de Murcia (2). The weight of the sample in Catalonia and the Community of Madrid reflects the mapping, where a larger number of organizations and associations were found in these communities, with a concentration around the main cities of Barcelona and Madrid.

35 of the representatives who responded to the survey were women, whilst 8 were men, and one respondent identified as "demigender". 41,9% of the respondents were in the age range of 36-45 years, 25,6% were between 26 and 35 years old, 18,6% were 46-55 years of age, 9,3% were in the age range of 56-65 years, and 4,6% were between 18 and 25 years old.

Presentation and analysis of the responses

Perceptions regarding digital violence and online hate speech

When asked to what extent digital violence is an issue that needs more attention in Spain, 77,3% of respondents (34) reply with the highest extent, and 22,7% (10) with the next to highest extent. Responding to a multi-choice question on what type of attention would be needed, the most frequent responses were:

- Improved legal framework (93,2% -41);
- Improved support to victims (81,8% -36)
- Improved response by the police (79,5% -35)

- More training for public institutions (79,5% -35)
- Improved response by social media platforms (72,7% -32)
- More training for NGOs (72,7% -32)
- More academic attention (47,7% -21)
- More media attention (40,9% -18)

Going into the perceptions of hate speech, these questions were divided into bias-motives, including gender, sexual orientation, sexual identity and other grounds including religion, racialization or ideology. When asked to what extent online hate speech is an issue that needs more attention in Spain, the responses for three of the grounds, gender, sexual orientation and other grounds were very similar (Figures 1-3 below), with 5 defined as "to the highest extent" and 1 as "not at all", with most responses divided between "to a very high extent" and "to the highest extent":

Figure 1: To what extent is gender-based hate speech an issue that needs more attention in Spain?
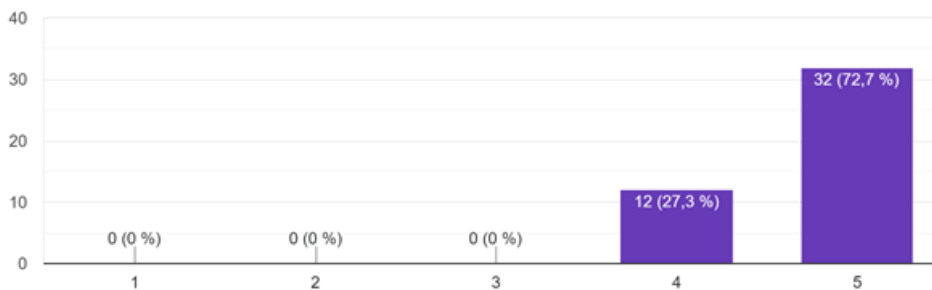
Figure 2: To what extent is hate speech based on sexual orientation an issue that needs more attention in Spain?
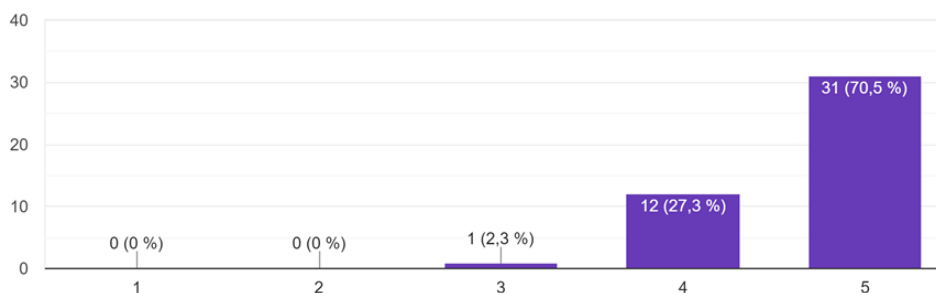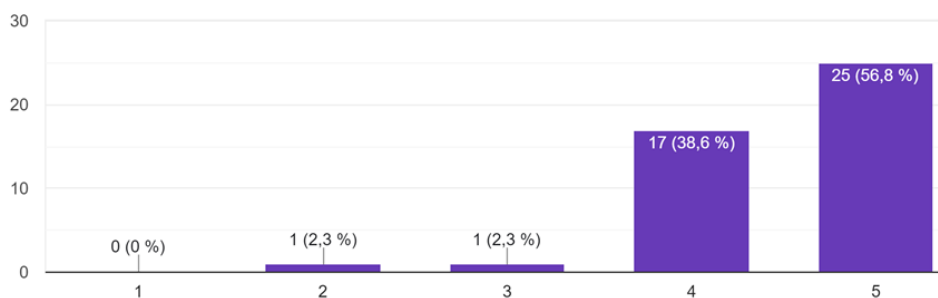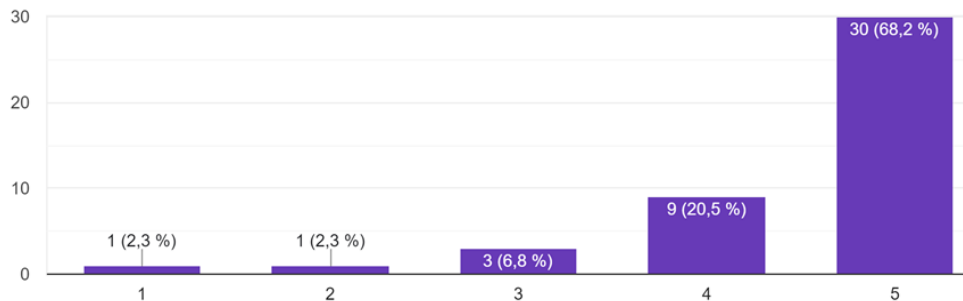

Figure 3: To what extent is hate speech based on other grounds, such as religion, racialization or ideology an issue that needs more attention in Spain?

However, for gender identity, the responses were slightly more varied, as observed in Figure 4, below.

Figure 4: To what extent is hate speech based on gender identity an issue that needs more attention in Spain?

Whilst the majority, 68,2% fully believe that hate speech on the grounds of gender identity deserves more attention, there are also a few deviating responses, which may reflect the highly polarised stances in relation to this topic even among different groups of feminists.

Several participants responded to an open-ended question on whether they had any reflections on the different grounds of online hate speech. In reflecting on the roots and consequences of online hate speech, survey respondents offered a multifaceted analysis. The proliferation of fake news, particularly from anonymous far-right accounts, was identified as a significant driver of intolerance, particularly towards marginalised communities like the LGTBI+ collective. Anonymity on social media platforms was highlighted as fostering impunity, with suggestions for implementing stricter verification processes for user accounts.

Furthermore, participants noted the correlation between disinformation and hate speech, pointing out that false information often fuels hateful rhetoric, while those consuming disinformation are more susceptible to engaging in hate speech themselves. Transphobia emerged as a particularly concerning form of online hate, with respondents calling for a clearer legal framework and increased public awareness to address the issue. Participants also advocated for a more nuanced understanding of the boundaries between hate speech and freedom of expression, stressing the importance of educating both users and policymakers on these distinctions. Critiques were directed towards mainstream media for their inadequate coverage of gender and racism, with calls for more feminist and anti-racist perspectives in media representation.

In summary, respondents underscored the need for greater education, awareness, and technological accountability to combat online hate speech effectively. They highlighted the role of societal norms, media representations, and political rhetoric in shaping online discourse and called for collective action to address these underlying factors.

Further, regarding the multi-choice question on what type of attention would be needed, the most frequent responses were:

- Improved legal framework (84,1% -37 responses);
- More training for public institutions (79,5% -35)
- Improved support to victims (72,7% -32)
- Improved response by the police (70,5% -31)
- Improved response by social media platforms (68,2% -30)
- More training for NGOs (65,9% -29)
- More academic attention (47,7% -21)
- More media attention (45,5% -20)

Activities and/or attention by the organizations in relation to digital violence and hate speech, and to victims of bias-motivated crimes and discrimination

A large part of the respondents (39,5%, 17 organizations) provide support to victims of gender-based violence; however, most of these organizations combine this with support to victims of other types of crimes, including anti-LGBT+ hate crimes, and racist or xenophobic hate crimes. This may also speak of an intersectional perspective, or that they may perceive some of the crimes or cases of discrimination as based on an intersection of factors. Only 5 organizations report that they provide support to victims of digital violence; however, always coupled with at least one other category (gender-based violence and/or anti-LGBT+ hate crimes). 18,6% of the organizations respond that they do not provide any kind of support to victims of crime or discrimination

Of those organizations that do provide some kind of support to victims of crime or discrimination, responding to a multi-choice question, the most common types of support are:

- General advice and counselling (81,8% - 27),
- Emotional and/or psychological support (72,7% -24),
- Social support (63,6% - 21),
- Accompaniment to report to the police (45,5% - 15),
- Legal advice (39,4% -13),
- Reception of reports (27,3% - 9)
- Support and advice in relation to digital spaces (27,3% -9),
- Economic support (21,2% - 7),
- Shelter (15,2% - 5).

As can be observed, some of the organizations (9) report that they do offer some kind of support and advice in relation to digital spaces. Regarding the question to what extent the organizations have addressed digital violence in their work, Figure 5 showcases the responses (1= not at all; 5= completely).

Figure 5: To what extent have you addressed digital violence in the work of your organization?
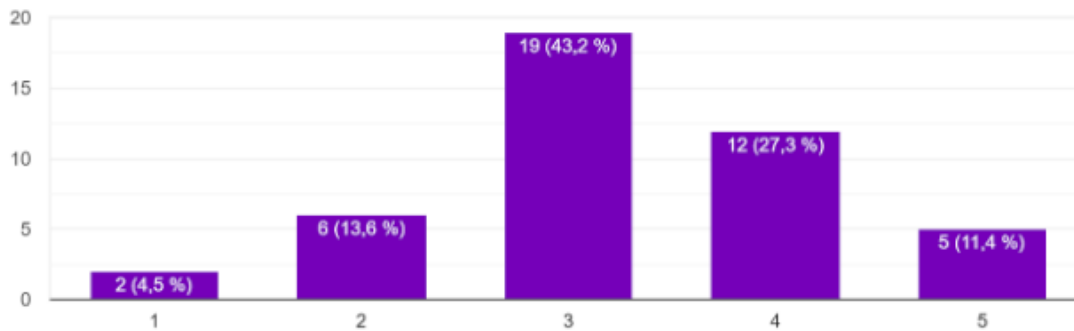
The same question, but in relation to online hate speech gives the following responses (1= not at all; 5= completely).

Figure 6: To what extent have you addressed online hate speech in the work of your organization?
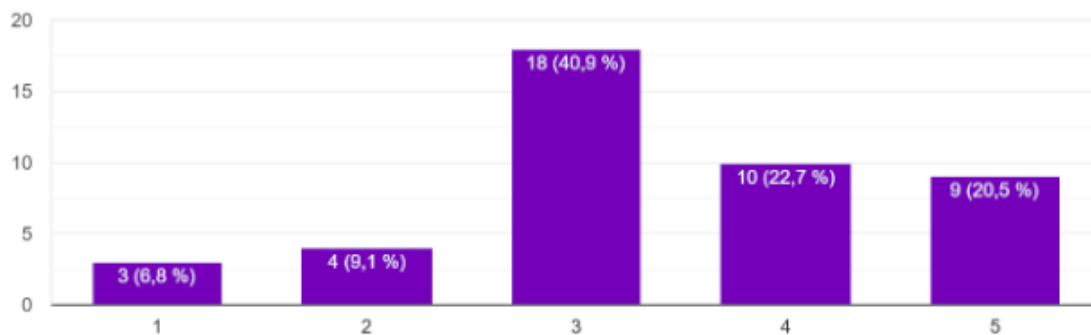
As we can observe, for both digital violence and online hate speech most organization position themselves as addressing these issues in their work to some extent, on the middle of the scale, and there is not much difference between the two issues in this regard. Looking at the organizations that report that they address violence in digital spaces (either digital violence or online hate speech) to a high extent or completely (17 organizations for digital violence and 19 for online hate speech), there is no clear pattern of the type of organization, that is, whether their main focus is on gender-based violence, LGTB+ issues or antiracism. However, almost all organizations that report that they address one issue completely also report that they address the other issue completely. This may indicate that these organizations perceive themselves as having a strong focus on (violence in) digital spaces in general. However, it may also indicate that the difference between digital violence and online hate speech is not fully clear either in our definitions or in their perception in general. Likewise, those that report that they do not address or address to a low extent one of the issues, report the same or a similar level for the other issue. Neither here is there a relation between the response and the type of organization (that is, it doesn't matter whether the main focus of their work is feminism, gender-based violence, LGBT+ rights or anti-racism).

Figure 7: To what extent do you feel confident to provide support or advice to victims of violence in digital spaces? (1= not at all; 5= completely):
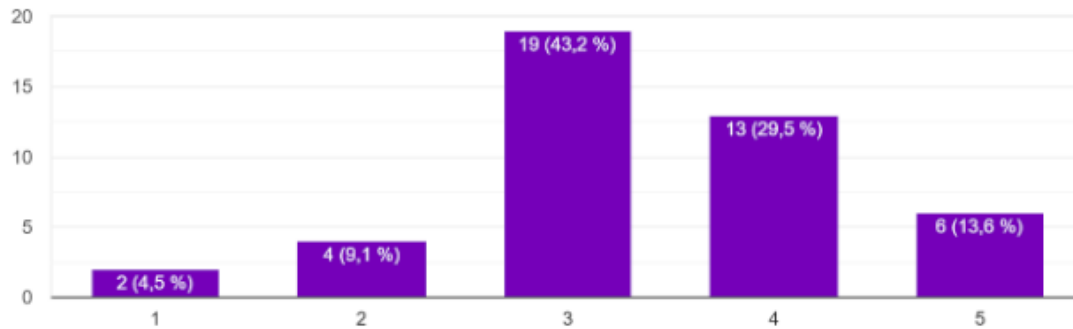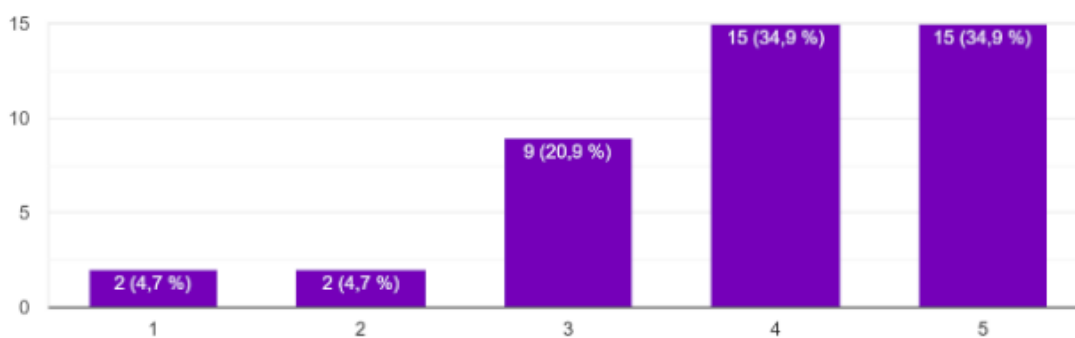
Figure 8: In your opinion, to what extent does your organization apply an intersectional perspective in its work? (1= not at all; 5= completely):

As we can observe in Figure 8, above, most organizations report that they apply an intersectional perspective to some extent. In this regard, the open-ended responses to the survey question about applying an intersectional perspective revealed a range of approaches and focal points. Intersectionality was seen as a crucial tool for understanding and addressing the specific vulnerabilities of different groups of women. It involved recognising and responding to the diverse needs and challenges faced by individuals based on their various identities and circumstances. Attention was paid to multiple factors, including gender, racialisation, class, sexuality, age, disability, nationality, and socioeconomic status, among others. Various organizations focused on specific intersections such as immigration, asylum, migrant experiences, transfeminism, sexual and functional diversity within the LGTBI collective, and issues faced by LBT+ women, particularly regarding harassment and employment assistance.

Several organizations report that they integrate intersectional perspectives structurally across all levels of the organizations, informing their activities, interventions, and networking efforts, recognising the interconnectedness of factors such as gender, ethnicity, nationality, disability, and sexual orientation. An emphasis was placed on offering support services that address multiple intersections, including mental health, housing, migration, substance use, and legal assistance. Despite these efforts, respondents acknowledged the ongoing need to further integrate intersectionality into

their work, make diversity visible, and continue internal training to advance their intersectional approach.

Perceptions related to the legal framework

In terms of perceptions of the legal framework applicable to digital violence and online hate speech, responses are on the lower end of the spectrum. That is, respondents perceive that the legal framework protects victims of this type of violence to some extent, but not completely, as can be observed in Figure 9 and Figure 10, below.

Figure 9: In your opinion, to what extent does the current legal framework in Spain protect victims of digital violence? (1= not at all; 5= completely):
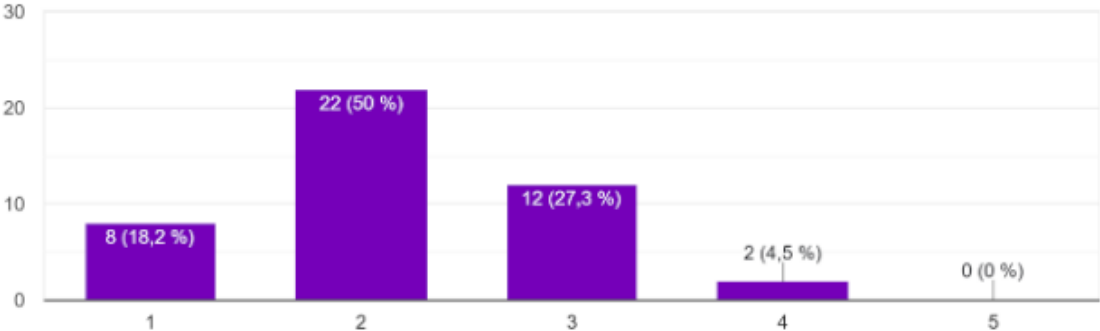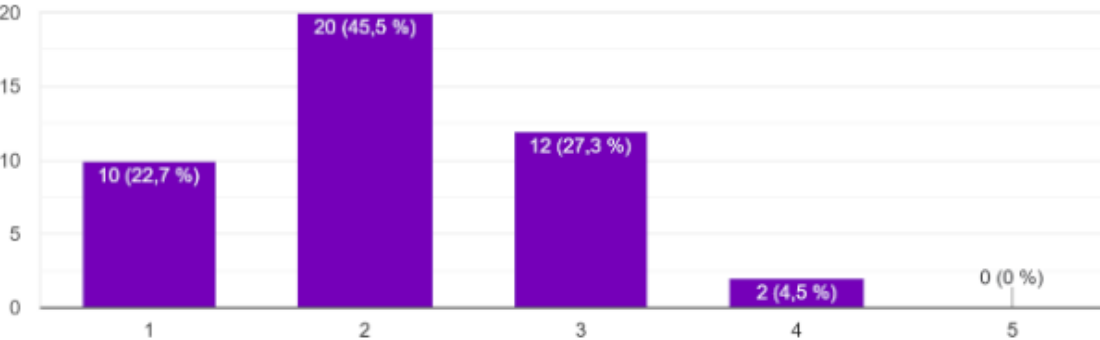

Figure 10: In your opinion, to what extent does the current legal framework in Spain protect victims of online hate speech? (1= not at all; 5= completely):

Further analysing the open-ended comments in relation to the legal framework, the legal framework in Spain for protecting victims of digital violence and online hate speech is widely perceived as insufficient. Additionally, there is a belief that the current legal framework lacks a feminist perspective. There is a perception of a significant underreporting of incidents, and a general lack of awareness or institutional responsibility. Judges are criticised for biased and inconsistent application of the law. Sanctions for aggressors are perceived as poorly understood by the public, leading to a perception of impunity. There is also a notable absence of information and socialization

campaigns. Some civil society respondents were unaware of the legal protections in place, suggesting a need for more information for victims on the legal options that exist, as well as more awareness campaigns targeting civil society organizations.

To conclude, the survey asked for open-ended final comments on digital violence and online hate speech regarding any issues not previously mentioned. Final comments emphasize the necessity for increased prevention efforts and highlight that the issue is often completely ignored. Disinformation is perceived as posing a significant threat to minority groups by fostering hate speech. Efforts to combat hate speech are perceived as needing to be peaceful and constructive, to further conflict. There is a general belief that awareness and institutional responsibility are lacking. Education at all levels is essential, along with early training to counteract violent attitudes. Finally, providing more information to victims about their options is also viewed as crucial, coupled with more general awareness campaigns.

Conclusion

Survey participants draw attention to a correlation between disinformation and the spread of hate speech. Contributing factors such as fake news from anonymous far-right accounts and the general anonymity on social media are identified as significant drivers of intolerance and impunity in digital spaces. Transphobia is highlighted as a particularly concerning form of online hate.

The survey results highlight significant concerns about digital violence and online hate speech in Spain. A vast majority (77.3%) of respondents believe digital violence needs the highest level of attention, with 22.7% assigning it the next highest level. When asked about necessary actions, the majority advocate for improving the legal framework; better support to victims; and for enhanced police response and training for public institutions. Additionally, respondents stress the need for better responses from social media platforms and more training for NGOs, while less than half perceive a need for more academic and media attention. Regarding online hate speech, concerns are similarly high, especially for issues related to gender, sexual orientation, and other grounds such as religion and racialisation.

The current legal framework is widely perceived as inadequate, lacking a feminist perspective and proper enforcement by judges. There are high levels of underreporting and a lack of public awareness about legal protections. Respondents recommend increased education and awareness campaigns to inform victims and civil society about legal options. There is also a call for unbiased and consistent application of laws by the judiciary.

Organizations surveyed provide various types of support to victims of digital violence and hate speech, offering general advice and counselling, emotional and psychological support, and social support. However, only 27.3% provide specific support related to

digital issues. Many organizations adopt an intersectional approach, addressing multiple forms of violence and discrimination.

Overall, the survey underscores an urgent need for comprehensive measures to address digital violence and online hate speech, emphasizing improved legal frameworks, support systems, education, and institutional accountability. The role of societal norms, media representation, and political rhetoric in shaping online discourse is also underscored, necessitating collective action and a multifaceted approach to these issues.

7. Final reflections

This research reveals that women often anticipate online hate following public appearances, with media coverage triggering waves of abuse. Hate speech intensifies when feminist issues are addressed, aiming to discredit and degrade feminists. The attacks frequently include explicit rape or death threats, and perpetrators often share victims' content without consent, spreading the violation across various platforms. These attacks persist over time, with continuous streams of hate messages contributing to ongoing victimization. Often, attacks originate from profiles with significant social media followings, amplifying their impact. Organized attacks suggest premeditated efforts, sometimes linked to political affiliations. The harm inflicted is tangible and embodied, blurring the boundaries between digital and non-digital realms. The frequency and magnitude of abuse exacerbate its impact, creating a significant chilling effect on feminists and women with a public profile. The challenge of "going offline" is notable, as professional obligations hinder disconnection, and going offline may risk exclusion from public discourse.

Reporting to the police is uncommon, with roots in the very normalisation of online violence, as well as a lack of trust in law enforcement, reinforced by previous experiences by peers. At the same time, reporting violence to the social media platforms is perceived as futile. This creates a sense of complete impunity among those who are targeted by online violence. The available strategies of protection consist of digital self-protection measures, as well as support from networks of peers.

Training for professionals is crucial, equipping them, including law enforcement, with the skills and tools to address violence in digital spaces, emphasizing empathy and realistic advice. Framing gender-based online violence within a hate crime framework can elucidate its structural roots and the collective silencing effects of such violence. This perspective underscores how various incidents, even those not defined as crimes by the penal code, can be experienced as significant acts of violence. A multi-level networked response is essential, moving beyond policing and criminal code to include education, enhanced digital protection, and increased collaboration with social media platforms. Comprehensive support for targets is necessary, providing legal, psychological, and tailored support to individuals targeted by online violence. Building this networked response requires equipping professionals with the knowledge to analyse and address violence from a structural perspective, recognising the specificities of the violence that targets women who express their opinions in public spaces. Understanding both the collective impacts and the connections between various attacks on an individual is essential. Professionals must be trained to offer empathetic support and practical advice to those directly affected, ensuring that the impact and severity of online violence are not underestimated simply because it occurs in digital spaces.

References

Álvarez Rodríguez, I. (2019). El discurso de odio sexista (en Construcción), *Revista Jurídica de Castilla y León*, 48, 43.

Amnesty International (2018). *Toxic Twitter*. Available at https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1-1/

Arce-García, S. and Menéndez-Menéndez, M.I. (2023). Inflaming public debate: a methodology to determine origin and characteristics of hate speech about sexual and gender diversity on Twitter, *Profesional de la información*, v. 32, n. 1, e320106. https://doi.org/10.3145/epi.2023.ene.06

Arcila-Calderón C., Amores J.J., Sánchez-Holgado P., Blanco-Herrero D. (2021). Using Shallow and Deep Learning to Automatically Detect Hate Motivated by Gender and Sexual Orientation on Twitter in Spanish, *Multimodal Technologies and Interaction*; 5(10):63. https://doi.org/10.3390/mti5100063

Ausserhofer, J., & Maireder, A. (2013). NATIONAL POLITICS ON TWITTER: Structures and topics of a networked public sphere, *Information, Communication & Society*, *16*(3), 291–314. https://doi.org/10.1080/1369118X.2012.756050

Awan, I. and Zempi, I. (2015). *We Fear for our Lives: Offline and Online Experiences of Anti-Muslim Hostility*. Tell Mama.

Balka, E. (1993). Women's access to on-line discussions about feminism, *Electronic Journal of Communication*, 3(1).

Barcellona, M. (2022). Incel violence as a new terrorism threat: A brief investigation between Alt-Right and Manosphere dimensions, *Sortuz: Oñati Journal of Emergent Socio-Legal Studies*, 11(2), pp. 170–186.

Beard, M. (2013). Internet Fury: Or Having Your Anatomy Dissected Online, *The Times Literary Supplement: A Don's Life*, available online at http://timesonline.typepad.com/dons_life/2013/01/internet-fury.html.

Burgess, M., Byars, F., Sadeghi-Azar, L. and K.E. Dill-Shackleford (2017). Online Misogyny Targeting Feminist Activism: Anita Sarkeesian and Gamergate. In Sturmey, P. (Ed.) *The Wiley Handbook of Violence and Aggression*. Hoboken, NJ: Wiley-Blackwell

Burnap, P. and Williams, M. (2016). Us and them: identifying cyber hate on Twitter across multiple protected characteristics, *EPJ Data Science* 2016 5:11:

Butler, J. (1997). *Excitable Speech. A Politics of the Performative*. New York: Routledge.

Calala Fondo de Mujeres (Coord.) (2020). *Las violencias machistas en línea hacia activistas. Datos para entender el fenómeno.*

Chakraborti, N. and Hardy, S.J. (2015). *LGB&T Hate Crime Reporting. Identifying Barriers and Solutions*. Equality and Human Rights Commission.

Citron, D.K. (2009a). Law's Expressive Value in Combating Cyber Gender Harassment, *108 Michigan Law Review*, 373-415

Citron, D.K. (2009b). Cyber civil rights, *Boston University Law Review*, 89. 61–125.

Citron, D.K. (2014). *Hate crimes in cyberspace*. Cambridge, MA: Harvard University Press.

Citron, D.K. and Norton, H.L. (2011). Intermediaries and hate speech: Fostering digital citizenship for our information age, *Boston University Law Review*, 91(4).

Coburn P., Connolly, D. and Roesch, R. (2015). Cyberbullying: Is federal criminal legislation the solution?, *Canadian Journal of Criminology and Criminal Justice* 57(4): 566–579

Crosas Remón, I. and Medina-Bravo, P. (2018). Ciberviolencia en la red. Nuevas formas de retórica disciplinaria en contra del feminismo*, Papers 2019*, 104/1

Daniels, J. (2008). Race, Civil Rights, and Hate Speech in the Digital Era. In Everett, A., The, J. D. and MacArthur, C.T. (Eds.) *Learning Race and Ethnicity: Youth and Digital Media*. Foundation Series on Digital Media and Learning. Cambridge, MA: The MIT Press

Deuze, M., Blank, P. and L. Speers (2012). A Life Lived in Media, *Digital Humanities Quarterly*, 6(1).

Dodge, A. and Spencer, D.C. (2017). Online Sexual Violence, Child Pornography or Something Else Entirely? Police Responses to Non-Consensual Intimate Image Sharing among Youth, *Social & Legal Studies* 1–22

Domínguez Ruiz, I.E., Roiha, M. and Jubany, O. (2023). Restorative solutions for anti-LGBT victimisation experiences: potential pathways for victims' wellbeing and key challenges and needs, *Culture, Health & Sexuality*, 25:12, 1626-1639, DOI: 10.1080/13691058.2022.2105401

Eckert, S. (2018). Fighting for recognition: Online abuse of women bloggers in Germany, Switzerland, the United Kingdom, and the United States, *New Media & Society* 20(1), DOI:10.1177/1461444816688457

European Commission against Racism and Intolerance (ECRI) (2016). *General Policy Recommendation Nº 15*. Brussels: ECRI.

Eichhorn, K. (2001). Re-in/citing linguistic injuries: speech acts, cyberhate, and the spatial and temporal character of networked environments, *Computers and Composition,* 18 (2001) 293–304.

Epley, N., & Kruger, J. (2005). When what you type isn't what they read: The perseverance of stereotypes and expectancies over e-mail. *Journal of Experimental Social Psychology, 41*(4), 414–422. https://doi.org/10.1016/j.jesp.2004.08.005

European Women's Lobby (2017). #HerNetHerRights. Mapping the state of online violence against women and girls in Europe.

Gagliardone, I., Gal, D., Alves, T. and Martinez, G. (2015). *Countering Online Hate Speech*. Programme in Comparative Media Law and Policy, University of Oxford.

Gómez, V. (2016). Incitación al odio y género. Algunas reflexiones sobre el nuevo art. 510 CP y su aplicabiliad al discurso sexista, *Revista Electrónica de Ciencia Penal y Criminología*, Vol. 18, 2016

Graddol, D. and Swann, J. (1989). *Gender Voices*. Oxford: Basil Blackwell.

Gray, K.L. (2012). Intersecting Oppressions and Online Communities: Examining the Experiences of Women of Color in Xbox Live, *Information, Communication & Society*, 15(3), 411

Harris, B. and Vitis, L. (2020). Digital intrusions: technology, spatiality and violence against women, *Journal of Gender-Based Violence*, vol 4, no 3, 325–341, DOI: 10.1332/239868020X15986402363663

Hasinoff, A. (2015). *Sexting Panic: Rethinking Criminalization, Privacy, and Consent*. Illinois: University of Illinois Press.

Heinze, E. (2006). Viewpoint Absolutism and Hate Speech, *The Modern Law Review*, 69(4), 543–582.

Henry, N. and Powell, A. (2014). The dark side of the virtual world. In Henry, N. and Powell, A. (Eds.) *Preventing Sexual Violence: Interdisciplinary Approaches to Overcoming a Rape Culture*. Basingstoke, UK and New York, US: Palgrave Macmillan.

Herring, S. (1993). Gender and democracy in computer mediated communication, *Electronic Journal of Communication*, 3(2), 476–489.

Herring, S. (1996). Bringing familiar baggage to the new frontier: Gender differences in computer-mediated communication. In Selzer, J. (Ed.) *Conversations* (pp. 1069-1082). Boston: Allyn & Bacon.

Herring, S. (2001). Gender and power in online communication, *Center for Social Informatics Working Papers*, no. WP-01-05.

Herring, S. (2002). Computer-mediated communication on the internet, *Ann. Rev. Info. Sci. Tech*., 36: 109-168. https://doi.org/10.1002/aris.1440360104

Herring, S. (2004). Computer-mediated communication and woman's place". In Bucholtz, M. (Ed.) *Robin Tolmach Lakoff Language and Woman's Place: Text and Commentaries*. NY: Oxford University Press.

Herring, S. (2010). Who's got the floor in computer-mediated conversation? Edelsky's gender patterns revisited, *Language@Internet*, 7, article 8.

Igareda, N., Pascale, A., Cruells, M. and Paz, O. (2019). *Les Ciberviolències masclistes*. Barcelona: Institut Català de les Dones.

Igareda, N. (2022). El discurso de odio anti-género en las redes sociales como violencia contra las mujeres y como discurso de odio, *Derechos y libertades*, Nº 47, Época II, junio 2022, pp. 97-122

Instituto Andaluz de la Mujer (2015). *Protocolo para la detección e intervención a víctimas de ciberviolencias*. Available online at https://www.juntadeandalucia.es/iam/catalogo/doc/iam/2015/143528391.pdf

Jane, E. (2014a). You're a Ugly, Whorish, Slut, *Feminist Media Studies*, 14(4)

Jane, E. (2014b). Back to the kitchen, cunt": Speaking the unspeakable about online misogyny, *Continuum: Journal of Media and Cultural Studies*, 28: 558–570, DOI:10.1080/10304312.2014.924479

Jane, E. (2015). Flaming? What Flaming? The Pitfalls and Potentials of Researching Online Hostility, *Ethics of Information Technology*, 17: 65–87.

Jane, E. (2016). Online misogyny and feminist digilantism, *Continuum: Journal of Media and Cultural Studies*, DOI: 10.1080/10304312.2016.1166560

Jane, E. (2017). Feminist flight and fight responses to gendered cyberhate. In Segrave, M. and Vitis, L. (Eds.) *Gender, Technology and Violence*. London: Routledge.

Jeong, S. (2018). *The Internet of Garbage*. Forbes Media.

Jubany, O. and Roiha, M. (2018). *Las palabras son armas*. Barcelona: Edicions UB.

KhosraviNik, M. and Esposito, E. (2018) Online hate, digital discourse and critique: Exploring digitally-mediated discursive practices of gender-based hostility, *Lodz Papers in Pragmatics*, 14.1, 45–68 Special issue on Narrating hostility, challenging hostile narratives: https://doi.org/10.1515/lpp-2018-0003

Kirkpatrick, D. (2011) *The Facebook Effect: The Inside Story of the Company That Is Connecting the World*. New York: Simon & Schuster.

Langarita, J.A., Sadurní, N., Dorado Caballero, A., and Albertín Carbó, P. (2018). From Zero to Hero: A Transnational Perspective on Countering Anti-LGBT Hate Crime in Europe. In Godzidz, P. and Viaggiani, G. (Eds.) *Running through Hurdles: Obstacles in the Access to Justice for Victims of Anti-LGBTI Hate Crimes*. Lambda Warsaw Association.

Lange, P.G (2006). What is your claim to flame?, *First Monday*, Vol 11, nº 9.

Lewandowska-Tomaszczyk, B. (2017). Incivility and confrontation in online conflict discourses, *Lodz Papers in Pragmatics*, vol. 13, no. 2, 2017, pp. 347-367. https://doi.org/10.1515/lpp-2017-0017

Lewis, R., Rowe, M. and Wiper, C. (2016). Online abuse of feminists as an emerging form of violence against women and girls, *British Journal of Criminology*, 57(6), 1462-1481.

Mantilla, K. (2015). *Gendertrolling. How misogyny went viral*. Santa Barbara: Praeger

Marin de Espinosa, E. (2018). La agravante genérica de discriminación por razones de género (art. 22.4 CP), *Revista electrónica de ciencia penal y criminología*, Nº. 20, 2018, p. 27.

Martínez-Valerio, L. (2021). Mensajes de odio hacia la comunidad LGTBIQ+: análisis de los perfiles de Instagram de la prensa española durante la "Semana del Orgullo", *Revista Latina de Comunicación Social*, 80, 363-388. https://doi.org/10.4185/RLCS-2022-1749

McCosker, A. (2014). Trolling as provocation: YouTube's agonistic publics, *Convergence*, 20(2), 201-217. https://doi.org/10.1177/1354856513501413

Megarry, J. (2014). Online incivility or sexual harassment? Conceptualising women's experiences in the digital age, *Women's Studies International Forum*, Vol 47, issue PA.

Ministerio del Interior (2023). *Informe sobre la evolución de los delitos de odio en España*. Available at https://www.interior.gob.es/opencms/export/sites/default/.galleries/galeria-de-prensa/documentos-y-multimedia/balances-e-informes/2022/Informe_Evolucion_delitos_odio_2022.pdf

Morena-Balaguer, D., García-Romeral, G. and Binimellis-Adell, M. (2021). *Diagnóstico sobre las violencias de genero contra activistes feministes en el ámbito digital*. Calala Fondo de Mujeres and Servei de Publicacions de UVic

Mullaney, L. and Trickett, L. (2018). *Misogyny Hate Crime Evaluation Report*. Available at https://www.nottingham.ac.uk/lipp/documents/misogyny-evaluation-report.pdf

Observatorio Nacional de Tecnología y Sociedad (Ontsi) (2022). *Violencia digital de género: una realidad invisible*. Available at https://www.ontsi.es/es/publicaciones/violencia-digital-de-genero-una-realidad-invisible-2022

ODIHR (2021). *Gender-Based Hate Crime*. OSCE-ODIHR Hate Crime Reporting. Available at https://www.osce.org/odihr/480847

Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M and Räsänen, P. (2014). Exposure to online hate among young social media users. In Warehime, M. (Ed.) *Soul of Society: A focus on the lives of children and youth*. Bingley: Emerald Group

O'Sullivan, P.B. and Flanagin, A.J. (2003). Reconceptualizing 'Flaming' and Other Problematic Messages, *New Media & Society*, vol 5, nº 1, pp. 69–94. http://dx.doi.org/10.1177/1461444803005001908

Paterson, J., Walters, M.A., Brown, R. and Fearn, H. (2018). *The Sussex Hate Crime Project. Final Report*. Leverhulme Trust.

Perry, B. (2001). *In the name of hate: Understanding hate crimes*. London: Routledge.

Phillips, W. (2015). *This Is Why We Can't Have Nice Things: Mapping the Relationship between Online Trolling and Mainstream Culture*. MIT Press.

Piñeiro-Otero, T. and Martínez-Rolán, X. (2021). Say it to my face: analysing hate speech against women on Twitter, *Profesional de la información*, v. 30, n. 5, e300502. https://doi.org/10.3145/epi.2021.sep.02

Pohjonen, M. and Udupa, S. (2017) Extreme Speech Online: An Anthropological Critique of Hate Speech Debates, *International Journal of Communication,* 11(2017), 1173–1191

Powell, A., & Henry, N. (2016). Policing technology-facilitated sexual violence against adult victims: police and service sector perspectives, *Policing and Society*, 28(3), 291–307. doi:10.1080/10439463.2016.1154964

Powell, A. and Henry, N. (2017). *Sexual violence in a digital age*. London: Palgrave MacMillan.

Rivera-Martín B. (2022). Hate speech towards LGBT people in Spain: Media and social audience, *Prisma Social*, 39, 213-233.

Sassen, S. (2002). Towards a Sociology of Information Technology, *Current Sociology*, 50(3).

Selfe, C. and Meyer, P. (1991). Testing Claims for On-Line Conferences, *Written Communication*, 8(2), 163-192. http://doi.org/10.1177/0741088391008002002

Serra, L. (2018a). *Las violencias de género en línea*. Pikara Magazine.

Serra, L. (2018b) *Discurso de incitación al odio: Análisis desde los derechos humanos y pautas interpretativas*. Barcelona: IDHC - Institut de Drets Humans de Catalunya

Shariff, S. and DeMartini, A. (2015). Defining the legal lines: egirls and intimate images. In Bailey J and Steeves V (Eds.) *eGirls, eCitizens*. Ottawa: University of Ottawa Press, pp. 282–307

Sills, S., Pickens, C., Beach, K., Jones, L., Calder-Dawe, O., Benton-Greig, P. and and N. Gavey (2016). Rape culture and social media: young critics and a feminist counterpublic, *Feminist Media Studies*, 16:6, 935-951, DOI: 10.1080/14680777.2015.1137962

Smith, J. (2019). 'When I saw women being attacked… It made me want to stand up and fight': Reporting, responding to, and resisting online misogyny. In Lumsden, K. and

Harmer, E. (Eds.) *Online othering: Exploring digital violence and discrimination on the web*. Palgrave MacmillanSobieraj (2018)

Souto Galván, B. (2015). Discurso del odio: género y libertad religiosa, *Revista General de Derecho Penal*, 23, 14.

Titley, G. (2012). *Hate speech in Europe: considerations for the proposed campaign Young People Combating Hate Speech in Cyberspace*. Council of Europe.

UN Broadband Commission for Digital Development Working Group on Broadband and Gender (2015). *Cyberviolence against women and girls. A world-wide wake-up call*. The Broadband Commission for Digital Development.

Vergès, N. (2017). *Redes sociales en perspectiva de género: Guía para conocer y contrarrestar las violencias de género online*. Instituto Andaluz de Administración Pública.

Waldron, J. (2012). *The Harm in Hate Speech*. Cambridge MA: Harvard University Press.

Weisband, S. and Atwater, L. (1999). Evaluating self and others in electronic and face-to-face groups, *Journal of Applied Psychology*, 84(4), 632.

Weston-Scheuber, K. (2012). Gender and the prohibition of hate speech. *QUT Law & Justice Journal*, *12*(2), 132-150.

Willem, C., Platero, R.L., and Tortajada, I. (2022). Trans-exclusionary Discourses on Social Media in Spain. In Krijnen, T., Nixon, P.G., Ravenscroft, M.D., Scarcelli, C.M. (Eds.) *Identities and Intimacies on Social Media*. London: Routledge

Annex

Table 1: Interviews with women targeted by online violence

Code	Profile	Age range
T-1	LGBT+ activist	25-29
T-2	Feminist and LGBT+ activist; Gamer	30-34
T-3	Feminist involved in youth politics	20-24
T-4	Feminist and antiracist activist	35-40
T-5	Journalist focused on issues of gender and feminism	30-34
T-6	Feminist journalist	30-34
T-7	Feminist activist	40-44
T-8	Feminist mainly active on social media	30-34
T-9	Antiracist and feminist activist	25-29
T-10	Feminist mainly active on social media	25-29

Table 2: Interviews with experts and professionals

Code	Profile/expertise
EX-1	Representative of an LGBT+ organization
EX-2	Representative of a feminist association
EX-3	Representative of a feminist association
EX-4	Representative of an LGBT+ organization
EX-5	Representative of an LGBT+ organization
EX-6	Lawyer specialised on hate crimes and gender-based violence
EX-7	Representative of a regional public institution in the field of gender equality and LGBT+ rights
EX-8	Representative of a regional public institution in the field of gender equality and LGBT+ rights
EX-9	Prosecutor
EX-10	Police officer
EX-11	Police officer specialised on victim support
EX-12	Representative of a public institution (local) in the field of gender equality and LGTB+ rights
EX-13	Representative of a public institution (local) in the field of anti-discrimination
EX-14	Representative of a public service (local) in the field of gender-based violence
EX-15	Law scholar specialised on gender-based and anti-gender violence
EX-16	Representative of a public service (local) in the field of gender-based violence
EX-17	Sociologist and representative of a feminist association

EX-18	Representative for an information service on gender-based hate crimes (NGO)
EX-19	Psychologist specialised on gender-based violence
EX-20	Researcher in the field of trans rights

Table 3: Cases from the digital ethnography

Code	Profile
DE-1	Feminist journalist
DE-2	Feminist journalist
DE-3	Antiracist activist
DE-4	Politician, councillor in a local government
DE-5	Non-binary person, local politician
DE-6	Trans woman, activist, academic
DE-7	Journalist
DE-8	Antiracist activist, feminist and communicator
DE-9	Trans woman and activist
DE-10	Trans woman and activist
DE-11	Trans feminist, activist
DE-12	Trans woman and comedian
DE-13	Feminist gamer
DE-14	Social media feminist
DE-15	Racialised journalist
DE-16	Trans woman, local politician
DE-17	Roma feminist
DE-18	Feminist journalist and activist
DE-19	Feminist journalist
DE-20	Social media feminist
DE-21	Local politician, feminist
DE-22	Anti-racist activist and feminist
DE-23	Journalist
DE-24	Online communicator and gamer
DE-25	Science communicator
DE-26	Sociologist specialised on gender-based violence