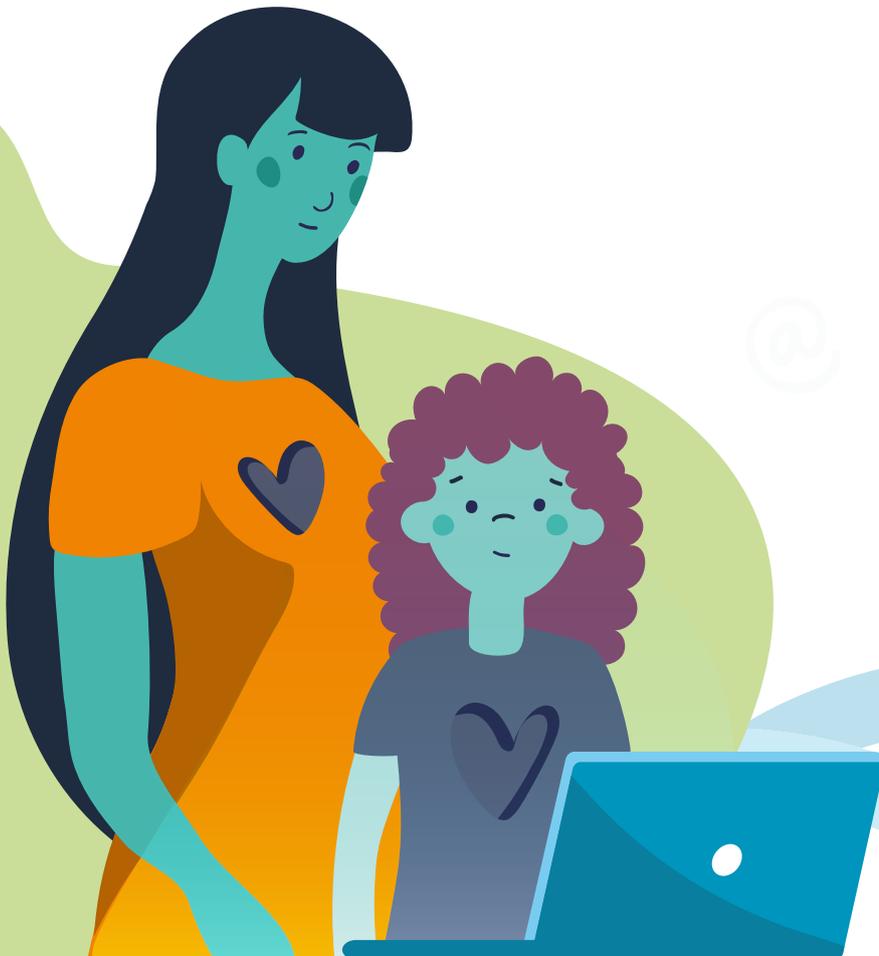




European Institute for  
Gender Equality

# Tackling cyber violence against women and girls: The role of digital platforms







European Institute for  
Gender Equality

# Tackling cyber violence against women and girls: The role of digital platforms





The European Institute for Gender Equality (EIGE) produces independent research and shares best practice to promote gender equality and eliminate discrimination based on gender. As the EU agency for gender equality, we help people achieve equal opportunities so everyone can thrive, independent of their gender and background.

We combine research, data and tools to help policymakers design measures that are inclusive, transformative and promote gender equality in all areas of life. We communicate our expertise and research effectively. We work closely with partners to raise awareness. We do this at the EU and national levels, and with EU candidate and potential candidate countries.

Cite this publication:

EIGE (2024), *Tackling cyber violence against women and girls: The role of digital platforms*, Publications Office of the European Union, Luxembourg.

Manuscript completed in October 2024

The provision of services for the preparation of this policy brief was contracted by EIGE to the Centre for Strategy and Evaluation Services.

Luxembourg: Publications Office of the European Union, 2024

© European Institute for Gender Equality, 2024

Reproduction is authorised provided the source is acknowledged.

For any use or reproduction of elements that is not owned by EIGE, permission may need to be sought directly from the respective rightsholders.

Cover and pages 6, 7, 10, 11, 12 and 15: © Salomi art / stock.adobe.com

Page 5: © Hannah Syvak and © Salomi art / stock.adobe.com

Page 8: © Salomi art / stock.adobe.com and © Mental Health

Page 9: © Bro Vector

Print ISBN 978-92-9486-263-1 doi:10.2839/4861873 MH-01-24-002-EN-C

PDF ISBN 978-92-9486-260-0 doi:10.2839/1955989 MH-01-24-002-EN-N

European Institute for Gender Equality

EIGE Gedimio pr. 16  
LT-01103 Vilnius  
LITHUANIA

Tel. +370 52157444

Email: [eige.sec@eige.europa.eu](mailto:eige.sec@eige.europa.eu)

Find us on



# Contents

The significant impact of gender-related cyber violence	5
Research process overview	7
Important policy and legislative advances to combat gender-based violence	8
What are digital platforms doing to address cyber violence?	8
Dealing with the wide variety of definitions	10
Digital platform reporting and moderation in practice	12
Summary of the key challenges when combating CVAWG	14
Recommendations for digital platforms	16



- Women and girls are more likely to be the targets of cyber violence on digital platforms and, as a result, they experience significant physical, sexual and psychological distress and/or financial difficulties.
- There is limited provision in digital platforms' standards and trust and safety policies for keeping users safe from gender-related cyber violence online, despite the high incidence of this phenomenon.
- The standards and trust and safety policies of the platforms make little reference to relevant human rights acts or important legislative advances in the fight against gender-based violence and cyber violence.
- Digital platforms dealing with the significant challenges of cyber violence acts and behaviours would benefit from greater collaboration across platforms. This would enable cross-platform reporting and the harmonisation of the wide variety of definitions of cyber violence.
- Digital platforms do not have cyber violence data that is disaggregated by sex available for incident reporting, response and follow-up practices. This renders the assessment of the true extent of cyber violence acts and behaviours against women and girls challenging. More transparency is needed in relation to moderation and follow-up practices.
- This lack of a gender-sensitive approach at the level of reporting, recording and responding to different forms of violence online renders the scale of the phenomenon of cyber violence against women and girls largely invisible and contributes to obscuring its dynamics.

## The significant impact of gender-related cyber violence

Digital platforms such as Snapchat, Instagram, X and Reddit are forums for public and private expression and social interaction, and they provide many other useful services. However, the online space is not always welcoming to everyone. In an environment of pseudo-anonymity and limited accountability, pervasive exclusionary and harmful narratives are often unleashed. Spreading virally, these practices are categorised as cyber violence and can have significant emotional and physical impacts on human lives.



## The significant impact of gender-related cyber violence

While cyber violence practices and behaviours can affect anyone, women and girls are more likely to be the targets of this form of violence <sup>(1)</sup> and, as a result, they experience significant physical, sexual and psychological distress and/or financial difficulties <sup>(2)</sup>. Incidents of cyber harassment, cyber stalking, coercive control, hate speech and the non-consensual sharing of intimate images tend to spill over into the physical environment, causing severe distress.



Cyber violence against women and girls (CVAWG) occurs as part of the broader continuum of gender-based violence (GBV), which is embedded in persisting unbalanced power structures. This form of violence against women and girls (VAWG) is linked to the societal reproduction of gender stereotypes. In addition, emerging technologies <sup>(3)</sup> are exacerbating the impact of cyber violence on victims by providing perpetrators with automated intelligence mechanisms they can use to intrude on their victims' personal and professional lives, monitoring them without consent and/or spreading disinformation in their online and offline communities.

The accurate assessment of the prevalence and incidence of CVAWG is essential to further inform EU Member State policies and legislation across the European Union, and to advise stakeholders on how to strengthen their collaborative efforts to combat this phenomenon. To address and assess CVAWG, a standardised, collaborative approach is required across legislative jurisdictions, digital platforms and digital service providers. This brief by the European Institute for Gender Equality (EIGE) aims to support digital and social media platforms by providing evidence-based actionable recommendations on how to strengthen their policies and practices to combat CVAWG.

<sup>(1)</sup> Council of Europe (2021), GREVIO General Recommendation No 1 on the digital dimension of violence against women, Strasbourg (<https://rm.coe.int/grevio-rec-no-on-digital-violence-against-women/1680a49147>). GREVIO is the Council of Europe's Group of Experts on Action against Violence against Women and Domestic Violence.

<sup>(2)</sup> EIGE (2022a), *Combating Cyber Violence against Women and Girls*, Publications Office of the European Union, Luxembourg (<https://eige.europa.eu/gender-based-violence/cyber-violence-against-women>); UN Women (2022), 'Accelerating efforts to tackle online and technology-facilitated violence against women and girls', New York (<https://www.unwomen.org/en/digital-library/publications/2022/10/accelerating-efforts-to-tackle-online-and-technology-facilitated-violence-against-women-and-girls>).

<sup>(3)</sup> UNESCO (2023), 'Your Opinion Doesn't Matter, Anyway' – Exposing technology-facilitated gender-based violence in an era of generative AI, Paris (<https://unesdoc.unesco.org/ark:/48223/pf0000387483>).

# Research process overview

Building on ongoing cyber violence research <sup>(4)</sup>, this study, conducted from April 2023 to March 2024, analysed selected online platforms' standards and trust and safety policies and how they play out in practice <sup>(5)</sup>. Box 1 illustrates the research process undertaken. Publicly available documentation on 17 platforms was reviewed <sup>(6)</sup> and analysed through a gendered lens. Then a series of anonymised <sup>(7)</sup> interviews took place with 10 <sup>(8)</sup> platform representatives about moderation practices. Based on the findings, recommendations for digital platforms were drawn up.



## Box 1: Multi-stage model of the research conducted by EIGE on digital platforms' online standards and trust and safety policies

### 1. Analysis of publicly available digital platform trust and safety standards and regulations

An in-depth qualitative analysis of trust and safety standards and platform community guidelines was undertaken using a sample of 17 prominent online platforms. These were examined with reference to the CVAWG context.

### 2. Identification of challenges and concerns

Commonalities and gaps in the standards and regulations were identified, as well as common challenges in terms of gender-sensitive policies and specific approaches to dealing with CVAWG. A series of interview questions was drawn up to ascertain what happens in practice.

### 3. Filling the gaps: anonymised interviews with platform representatives

Anonymous interviews were conducted with 10 platform representatives about how these standards and regulations are implemented, with a specific focus on content moderation, reporting processes and follow-up.

### 4. Recommendations for platforms to address CVAWG

Based on the evidence uncovered, several actionable recommendations have been provided to strengthen the fight against CVAWG and provide a more inclusive and safe digital space.

Source: Developed by EIGE (2024).

<sup>(4)</sup> EIGE (2022a), *Combating Cyber Violence against Women and Girls*, Publications Office of the European Union, Luxembourg (<https://eige.europa.eu/gender-based-violence/cyber-violence-against-women>); EIGE (2022b), 'Cyber violence against women and girls – Key terms and concepts', Vilnius ([https://eige.europa.eu/sites/default/files/cyber\\_violence\\_against\\_women\\_and\\_girls\\_key\\_terms\\_and\\_concepts.pdf](https://eige.europa.eu/sites/default/files/cyber_violence_against_women_and_girls_key_terms_and_concepts.pdf)).

<sup>(5)</sup> Digital platforms call their standards and policies by a wide variety of names, such as community standards; trust and safety policies; user terms; codes of conduct; and abuse or bullying and harassment policies. In this brief, EIGE uses 'online platforms' standards and trust and safety policies' as an overarching term to refer to these policies.

<sup>(6)</sup> This sample of 17 service providers was selected to be representative of different platform focuses and sizes. The platform standards and policies examined were from Facebook, Instagram, WhatsApp, X (Twitter), TikTok, Google/YouTube, Snapchat, Reddit, Hinge/Tinder/Match.com, eHarmony, Uber, Depop, Bumble and Aylo.

<sup>(7)</sup> Except where permission was provided to include details.

<sup>(8)</sup> A total of 60 representatives from 30 online platforms were approached for interview; anonymity was guaranteed. However, only 10 representatives of 7 digital media platforms and meta-platforms, covering 11 distinct online platforms and services, agreed. Getting platform representatives to engage with this research was challenging, due in part to their busy schedules and in part to the fact that the names of key management staff responsible for platform trust and safety and governance are not, in general, publicly available. In addition, when representatives did engage, they provided information related to moderation practices and responsiveness to reporting with a high degree of caution.

# Important policy and legislative advances to combat gender-based violence

The European Commission's [gender equality strategy 2020–2025](#) <sup>(9)</sup> and the EU's accession to the Istanbul Convention <sup>(10)</sup> represented important milestones in the fight against GBV. These instruments put increased focus on ending all forms of GBV. In addition, the Digital Services Act (DSA) <sup>(11)</sup>, which regulates online intermediaries and platforms, now places increased responsibility on service providers to create a safer digital space for users. Furthermore, [Directive \(EU\) 2024/1385](#) of the European Parliament and of the Council of the European Union of 14 May 2024 on combating violence against women and domestic violence (the VAW/DV directive) constitutes a significant legislative commitment to combating CVAWG, as it obliges Member States to act against specific forms of cyber violence crimes.



Although digital platforms in EU jurisdictions are obliged to comply with the DSA (see Box 2), the other aforementioned important legislative acts and directives are not, as yet, reflected in their standards and trust and safety policies.

## Box 2: Explanatory note on digital platform compliance with the Digital Services Act

Platforms were found to be generally working towards compliance with DSA standards and demands. Digital platforms either had already published or were preparing to publish transparency reports <sup>(12)</sup> providing the public with insights into moderation, safety and legal requests. However, not all platforms provided reports on:

- time taken to act on violative content <sup>(13)</sup>;
- the proportion of violative content out of total content <sup>(14)</sup>;
- the different sanctions imposed on users violating standards, based on types of incidents.



Regulation (EU) 2022/2065 on a single market for digital services (the DSA) was adopted in October 2022 and has been applicable since February 2024. It places a range of legal obligations on online platforms, including reporting on moderation practices.

<sup>(9)</sup> <https://ec.europa.eu/newsroom/just/items/682425/en>.

<sup>(10)</sup> <https://www.coe.int/en/web/istanbul-convention>.

<sup>(11)</sup> [Regulation \(EU\) 2022/2065](#) on a single market for digital services and amending Directive 2000/31/EC (Digital Services Act) (OJ L 277, 27.10.2022, p. 1).

<sup>(12)</sup> The DSA was adopted in October 2022 and became applicable on 17 February 2024. The DSA places a range of legal obligations known as 'transparency reporting obligations' on online platforms, including requirements related to content moderation, with additional obligations set up for very large online platforms. Note also that the transparency reports that platforms need to submit are gender blind, making it difficult to collect data on GBV, as highlighted by EIGE (<https://eige.europa.eu/newsroom/news/eige-stresses-need-gender-dimension-dsa-transparency-reporting>).

<sup>(13)</sup> Snapchat does provide this, whereas Facebook/Instagram do not.

<sup>(14)</sup> Snapchat and Facebook/Instagram do provide this, but X does not.

## What are digital platforms doing to address cyber violence?

Digital platforms have many initiatives in place to deal with abuses of standards, breaches of user agreements and incidents of technology-related violence, in addition to many other violations. These standards and trust and safety policies focus, in the main, on a general user audience.

Significant challenges are faced by the platforms in ensuring compliance with these standards and trust and safety policies. The platform representatives interviewed referred to the difficulty they have with user age verification and the verification of other personal characteristics given digital privacy rights. Another challenge faced relates to the immense volumes of data and content posted by users, which require constant monitoring efforts across a multitude of geographical and legal jurisdictions.

In spite of these challenges, all platforms do:

- have core policies and regulations on general content moderation and continuously monitor and review moderation processes <sup>(15)</sup>;
- provide generally clear information that is publicly available to users on reporting processes;
- have terms and conditions in place stating the agreement between the user and the platform and covering the platform's expectations of the user in terms of conduct, privacy, platform rights and payment <sup>(16)</sup>;
- provide general community guidelines or standards, which have been developed on a global basis in the main <sup>(17)</sup>.

<sup>(15)</sup> Google highlights differences in the permissibility of content of a sexual nature, particularly differences across Western and Middle Eastern populations.

<sup>(16)</sup> In some cases, these are supplemented by other documents, for example, Reddit's code of conduct for moderators and Snapchat's explanations of each type of prohibited content.

<sup>(17)</sup> Exceptions to this are WhatsApp, X and TikTok, which use specific terms of service for different geographies.



## What are digital platforms doing to address cyber violence?

In addition, all serious violations are reported by the platforms to the relevant state agencies. For example, criminally defined and/or illegal content is reported to law enforcement agencies (e.g. child sexual abuse material).



### Box 3: How platforms deal with the gender perspective

Although platforms have generic policies to combat cyber violence, they do not apply a specific gender perspective in policy development, nor do they explicitly refer to the cyber violence experiences of women and girls<sup>(18)</sup>. Where gender is highlighted as a protected category, it is often just within the context of hate speech online. Although some specific forms of abuse and harm to users that are listed in standards and trust and safety policies have an implicit gender aspect, it is not explicitly stated (examples include ‘sexual content’, ‘harassment and bullying’, ‘hate speech’ and ‘non-consensual nudity’). Neither do platform policies reference international legislation on VAWG in their approaches to trust and safety (e.g. GREVIO General Recommendation No 1)<sup>(19)</sup>.

## Dealing with the wide variety of definitions

Platforms’ definitions of cyber offences<sup>(20)</sup> vary widely as, to date, they have not typically consulted with each other to harmonise definitions at the initial stage of policy development<sup>(21)</sup>. The platform representatives interviewed described the lack of common definitions as a barrier to tackling cyber violence. Urging policymakers to harmonise definitions across global legal jurisdictions, they stated: ‘Harmonisation is key and will greatly simplify our operations.’

An example of how the definitions and details of different offences vary widely across platforms is provided in [Table 1](#) regarding non-consensual intimate image (NCII) or material abuse.

<sup>(18)</sup> <https://www.coe.int/en/web/istanbul-convention>.

<sup>(19)</sup> See note 1.

<sup>(20)</sup> Definitions vary in terms of both the behaviours and acts described and the level of detail provided.

<sup>(21)</sup> Roblox and Depop called for more collaboration across platforms, and the consensus from interviews is that a trust and safety community is emerging.

**Table 1:** Non-consensual intimate image abuse: samples of variations in definitions across platforms

Platform	Definitions related to NCII abuse
Reddit	<p>'Rule 3 prohibits sharing intimate or sexually explicit media of a person created or posted without their permission. Intimate media include a depiction of the person in a state of nudity or engaged in any act of sexual conduct, including depictions that may have been AI-generated or faked. Images or video of intimate parts of a person's body, even if the person is clothed or in public, if contextualized in a salacious manner (e.g. "creepshots" or "upskirt" imagery), are also prohibited. The Rule applies to leaked, stolen or privately shared images of an individual where the individual, or their representative, reports that they do not consent to the public sharing of the images. Additionally, images or video of another person posted for the specific purpose of faking explicit content or soliciting "lookalike" pornography (e.g. "deepfakes" or "bubble porn") is also against the Rule' <sup>(22)</sup>.</p> <p>The guidance on this point goes on to provide explicit examples of content that violates Rule 3, before providing links to where users can report non-consensual intimate media and where users can find more information and support. The support signposted includes StopNCII and the international resources page of the US-based Cyber Civil Rights Initiative, which provides links to organisations focused on image-based sexual abuse in 10 Member States (Belgium, Denmark, Germany, Ireland, Spain, France, Italy, the Netherlands, Austria, Finland) <sup>(23)</sup>.</p>
TikTok	<p>'Image-based sexual abuse is the creation, manufacture, or distribution of nude, partially nude, or sexually explicit content without the consent of the person in the content, for the purpose of sexualizing their body, or portraying them in a sexual manner' <sup>(24)</sup>.</p> <p>This definition is accompanied by a link through which users can report content; a link to sexual assault resources, including advice for victims of sexual assault and their friends and family; and links / contact details for specialist local (i.e. country-specific) organisations, including in 11 Member States (e.g. Violences Femmes Info and Solidarité Femmes in France, or NANE Egyesület in Hungary).</p>
Tinder	<p>'Don't post images or private messages from other people unless you've been given consent to do so' <sup>(25)</sup>.</p>
Hinge	<p>Prohibition of uploading or sharing content that 'includes the image or likeness of another person without that person's consent' <sup>(26)</sup>.</p>

Source: Developed by EIGE from publicly available policies.



<sup>(22)</sup> <https://support.reddithelp.com/hc/en-us/articles/360043513411-Never-Post-Intimate-or-Sexually-Explicit-Media-of-Someone-Without-Their-Consent>.

<sup>(23)</sup> <https://cybercivilrights.org/intl-victim-resources/#1652883217730-29d4a1a0-71b2>.

<sup>(24)</sup> <https://www.tiktok.com/community-guidelines/en/safety-civility/>.

<sup>(25)</sup> <https://policies.tinder.com/community-guidelines/intl/en-gb/>.

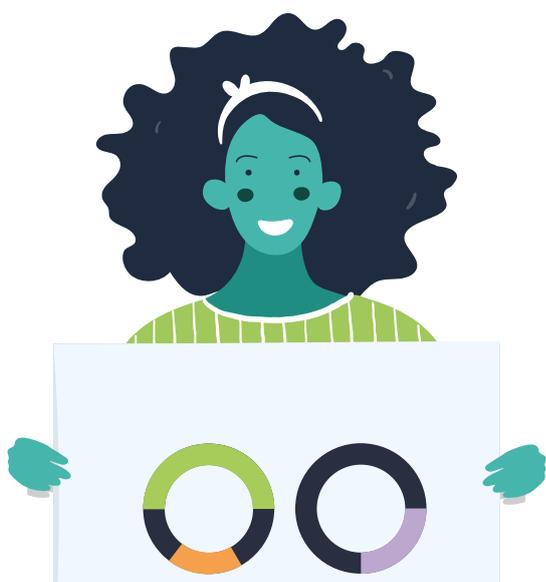
<sup>(26)</sup> <https://hinge.co/terms>.

## Digital platform reporting and moderation in practice

It was challenging <sup>(27)</sup> to determine exactly how the digital platform standards and trust and safety policies and regulations play out in practice. With regard to the internal platform workings of moderation, user reporting, follow-up processes and subsequent sanctioning, findings showed that:

- online platforms lack transparency in their internal mechanisms and processes for reporting, recording and managing complaints;
- the division of labour between machine learning (ML) technology and human moderators across the platforms is difficult to ascertain;
- the time limit for someone to make a complaint and the timeline for addressing reports are not always clear;
- reporting follow-through procedures are not always transparent;
- there are no specific provisions with regard to gender within the descriptions of penalties and sanctions that platforms impose on perpetrators;
- there is no indication that tailored considerations are used in relation to CVAWG cases;
- the exact percentages of user reports resulting in some action being taken, and the exact outcome of these reports, are largely unknown <sup>(28)</sup>;
- finally, with few exceptions, the breakdown of reports by sex/gender, language and geographical region is not publicly available.

Improving women's safety online implies that appropriate consequences will be enforced in cases of CVAWG. Cyber violence is under-reported by victims <sup>(29)</sup>, and those who do report it often find it difficult to navigate platforms' reporting systems to find and define the specific violations they have experienced. Many women who experienced cyber violence and reported it to the relevant platform said they were not satisfied with the outcome of their complaint <sup>(30)</sup>.



<sup>(27)</sup> See also note 8.

<sup>(28)</sup> With very few exceptions – notably Snapchat, which offers a report on 'violative view rate', which is the proportion of violative content out of the total content reported.

<sup>(29)</sup> Refer to note 2.

<sup>(30)</sup> Jurasz, O. (2024), *Online Violence against Women: A Four Nations study*, The Open University (<https://oro.open.ac.uk/96398/>).

Furthermore, the statistics related to GBV reports are not often made public by digital platforms. However, to comply with DSA reporting requirements, platforms are now required to provide this information, in addition to providing more public detail on the time taken to act on violative content <sup>(31)</sup>; the proportion of violative content out of total content <sup>(32)</sup>; and the different sanctions imposed based on the type of content. As yet, however, in the platforms' reporting mechanisms for the DSA, sex-disaggregated data is not available, so it is not currently possible to gather detailed CVAWG statistics from these reports to better inform future legislation, standards and policy development.

In terms of moderation processes, many digital platforms engage the services of external agencies to support them. [StopNCII](#) is one such provider. In addition, digital platforms do engage in the continuous refinement of their approaches to human-assisted and artificial intelligence (AI)-assisted content moderation. Reddit, a digital platform that hosts thousands of online communities, has a particularly innovative crowd-sourced approach to moderation, as illustrated in [Box 4](#).

#### Box 4: Spotlight on Reddit's moderation system regarding non-consensual intimate images

Reddit <sup>(33)</sup> is divided into thousands upon thousands of different subreddits created by users and primarily moderated by users according to rules appropriate to the particular subject of the subreddit. Reddit supports the moderators' ability to have those rules and provides them with tools to enforce them more efficiently. Generally, Reddit will not intervene with rules specific to subreddits unless there is behaviour violating Reddit's platform-wide [content policy](#) <sup>(34)</sup>. Every individual user on Reddit has a role to play in content moderation through democratic means, as content on Reddit is primarily sorted by the upvotes and downvotes of the users.

Most of the NCII abuse that Reddit takes action on is flagged by either automation or humans. When Reddit is taking that content down, most of the reports (around 75 %) come from Reddit's automation rather than user reports or moderator actions. Reddit recognises that it is very distressing for people to be victims of this type of behaviour; therefore, it has a help centre article that directs victims to resources for this situation and gives them instructions on how to engage with the StopNCII image-hashing database, if they would like to put their content therein. Reddit also has a partnership with Crisis Text Line, which people can use to get one-to-one help from trained crisis counsellors.

Source: Provided with permission from Reddit interview participant.

<sup>(31)</sup> Snapchat does provide this, whereas Facebook/Instagram do not.

<sup>(32)</sup> Snapchat and Facebook/Instagram do provide this, but X does not.

<sup>(33)</sup> Reddit is a platform that 'hosts a variety of online communities – thousands of subreddits address topics ranging from books and TV shows to science and politics ... Reddit was explicitly founded on the principles of free speech', according to Chandrasekharan et al. So, when violative content is flagged by a community member, Reddit employs quarantining as a moderation strategy, helping the platform fulfil its mission to be 'the best place online to have truly authentic conversations', according to Stephen. See Chandrasekharan, E., Jhaver, S., Bruckman, A. and Gilbert, E. (2022), 'Quarantined! Examining the effects of a community-wide moderation intervention on Reddit', *ACM Transactions on Computer–Human Interaction*, Vol. 29, No 4, pp. 1–26 (<https://arxiv.org/pdf/2009.11483>); Stephen, B. (2018), 'Reddit updates its quarantine policy with an appeals process', *The Verge*, 28 September (<https://www.theverge.com/2018/9/28/17914240/reddit-update-quarantine-policy-appeals-process>) (as quoted in Chandrasekharan et al.). This example was chosen as Reddit employs an innovative community moderation mechanism.

<sup>(34)</sup> <https://redditinc.com/policies/content-policy>.

It is important to support the human moderators who process and identify violative content, as they often experience psychological stress as a result of this work. The burden of viewing graphic and/or distressing content can cause severe strain. Digital platforms employ various technologies to mitigate this effect, and these technologies are intended to 'spare the reviewers from having to view upsetting content' (research interview participant). However, they protect moderators only to a certain extent, for the following reasons.

- Aspects of CVAWG and the violations that are interconnected with offline offences and offences on other platforms are frequently too subtle or contextual to be flagged by ML technologies. Therefore, human moderators are often tasked with the burden of identifying cases of CVAWG. In addition, views on what constitutes online hate may differ; it is unclear whether 'appropriate training on legal thresholds for identifying online hate' <sup>(35)</sup> is provided across the digital platforms.

- Human content moderators face significant challenges, as they are exposed to emotionally taxing content and are given often quite unclear guidelines. This raises concerns about ensuring that adequate support mechanisms and training from technology companies are provided. Little evidence was provided in the course of this study of such comprehensive support for moderators.

Finally, with regard to the provision of direct support or tailored guidance offered to complainants:

- it is unclear whether direct support or tailored guidance is offered to complainants as potential victims of cyber violence;
- while there are some links provided by platforms to suicide prevention helplines, very few relate specifically to GBV or reference specialised support services for victims of GBV.

## Summary of the key challenges when combating CVAWG

Standards and trust and safety policies could be improved, as 'integrating gender-based perspectives into policy development and content moderation would be appropriate and would promote inclusivity and enhance effectiveness of moderation efforts'. In addition, closer collaboration among platforms to improve policy and practice in trust and safety is beneficial. This has already been

prompted by DSA regulatory requirements: 'The emergence of a moderation community and closer cooperation among platforms regarding moderation [was] prompted by the DSA.' However, 'further cooperation is needed across platforms' to address the challenges they face online <sup>(36)</sup>.

<sup>(35)</sup> European Union Agency for Fundamental Rights (2023), *Online Content Moderation – Current challenges in detecting hate speech*, Publications Office of the European Union, Luxembourg ([https://fra.europa.eu/sites/default/files/fra\\_uploads/fra-2023-online-content-moderation\\_en.pdf](https://fra.europa.eu/sites/default/files/fra_uploads/fra-2023-online-content-moderation_en.pdf)).

<sup>(36)</sup> Quotes in this paragraph are all from (anonymised) digital platform representatives interviewed during the course of this research.

In summary, a significant set of challenges need to be overcome by digital platforms to effectively tackle CVAWG. These issues are summarised in [Table 2](#). Note, however, that EIGE acknowledges that policies and technologies in the digital and social media field evolve rapidly, and that research in this field needs to be continually updated. In the final section of this brief, a series of key policy recommendations for digital platforms is provided to help them in addressing the challenges they face.



**Table 2: Challenges to be tackled by digital platforms in the fight against CVAWG**

A lack of gender focus in standards and trust and safety policies	Data fragmentation due to variations in cyber violence definitions	Insufficient provision for moderation, reporting and response mechanisms related to GBV, VAWG and CVAWG	Insufficient cooperation, transparency and collaborative support across relevant stakeholders
<ul style="list-style-type: none"> <li>• Persistence of gender-neutral language and approaches in tackling different forms of violence on online platforms</li> <li>• Lack of reference to relevant national legislation or international women and/or human rights treaties on GBV, CVAWG and VAWG</li> <li>• Limited acknowledgement of gender-related and intersectional cyber violence vulnerabilities</li> </ul>	<ul style="list-style-type: none"> <li>• Multiplicity of diverse definitions of CVAWG across platforms, hindering the comparability of data and giving rise to the fragmentation of data (*)</li> <li>• Lack of homogeneity, hampering efforts to accurately measure CVAWG across jurisdictions and platforms and hampering the ability of users to report harmful behaviours</li> </ul>	<ul style="list-style-type: none"> <li>• Unclear guidelines, standards and support related specifically to GBV, CVAWG and VAWG</li> <li>• Absence of sex-disaggregated data collection at the level of incidents moderated and investigated</li> <li>• Unavailability of harmonised standards to provide swift responses to users' reports</li> <li>• Lack of recognition that CVAWG crosses platforms and physical boundaries</li> </ul>	<ul style="list-style-type: none"> <li>• Lack of clear and publicly available data on the reporting of GBV-related violations and behaviour and the outcomes of reporting</li> <li>• Insufficient willingness of platforms to engage in discussions with other relevant stakeholders about their safety standards and practices and their AI and ML developments</li> <li>• Adequate CVAWG training and support needed for human content moderators</li> </ul>

(\*) Data recorded in DSA transparency reports and data recorded by individual platforms.

NB: Evidence from EIGE's 2024 study of online platform standards and trust and safety policies and interviews with platform representatives.

Source: Developed by EIGE.

# Recommendations for digital platforms

## On addressing the challenge of inadequate gender focus in standards and trust and safety policies

Incorporate a gender-sensitive approach into the creation of policies and moderation procedures, fostering a more inclusive environment and improving content moderation.

Refer to significant legislative frameworks combating GBV and CVAWG, such as the Istanbul Convention and the VAW/DV directive, when developing and reviewing standards and trust and safety practices.

Implement policies and follow-up procedures attuned to distinct types of CVAWG and acknowledge intersectional cyber violence vulnerabilities.

## On tackling data fragmentation due to variations in cyber violence definitions

Aim to employ common, harmonised definitions of the different forms of CVAWG, referring to EIGE's publications and the CVAWG measurement framework developed in line with the VAW/DV directive.

## On improving provision for moderation, reporting and response mechanisms related to GBV, VAWG and CVAWG

Put in place user-friendly guidelines and mechanisms at the level of reporting and recording, tailored specifically to incidents of CVAWG <sup>(37)</sup>. Ensure guidelines are language and context specific.

Provide for trust and safety policies and support that adequately acknowledge CVAWG and use clear and uniform terminology across platforms.

Design reporting systems to permit the collection of sex-disaggregated and other relevant data.

Recognise that violence crosses boundaries. Facilitate the reporting of CVAWG that happens outside of the platform's immediate scope.

React promptly and respond swiftly to reports of harmful content – for example, by immediately suspending potential offenders' profiles and escalating the case to law enforcement where appropriate.

<sup>(37)</sup> Actions and penalties: the actions that platforms already take as a result of complaints depend on the nature of the platform and the severity of the offence; however, in most cases (including for Facebook/Instagram, TikTok, YouTube and Snapchat), these actions include removing the offending content and imposing some form of penalty on the account that shared the offending content (e.g. temporary suspension of the user's account). However, the approaches to warning, suspending or terminating accounts differ slightly across the platforms examined. For example, both TikTok and YouTube implement a warning/strike system. However, TikTok requires that strikes be counted by category of violation in relation to a particular policy (e.g. bullying and harassment) in order to reach a category-specific threshold, whereas YouTube's strikes are counted inclusive of all types of violative content. In addition, where required, content is reported to relevant law enforcement agencies by all platforms. Most platforms provide a clear means by which to appeal any penalties or sanctions. Note that there is no consideration of the gender dimension within the descriptions of penalties and sanctions researched, and no indication that tailored considerations are used in relation to CVAWG cases.

### On improved cooperation among relevant stakeholders

Collaborate further with researchers and contribute data, thus providing greater public transparency when it comes to the reporting and follow-up of GBV-related behaviour and violations. Assist in the mapping of the various forms of online violence.

Continue to expand cross-platform cooperation. Exchange best practices on moderating CVAWG. Work together to use AI responsibly and adopt gender-sensitive

moderation practices, and continue to create partnerships for tackling CVAWG.

Further train and support internal platform stakeholders, particularly human content moderators, as they play a crucial role in evaluating and addressing such cases. Acknowledge the burdens of this emotionally taxing job, continuing to invest in AI and ML to support moderation efforts and lessen the burden of viewing harmful content while ensuring the gender perspective is considered.

# Read more about EIGE's research in the field

EIGE's previous reports and publications on CVAWG are available at:

<https://eige.europa.eu/gender-based-violence/cyber-violence-against-women>

<https://eige.europa.eu/publications-resources/publications/combating-cyber-violence-against-women-and-girls>

[https://eige.europa.eu/sites/default/files/cyber\\_violence\\_against\\_women\\_and\\_girls\\_key\\_terms\\_and\\_concepts.pdf](https://eige.europa.eu/sites/default/files/cyber_violence_against_women_and_girls_key_terms_and_concepts.pdf)

<https://eige.europa.eu/gender-statistics/dgs>

[https://eige.europa.eu/gender-mainstreaming?language\\_content\\_entity=en](https://eige.europa.eu/gender-mainstreaming?language_content_entity=en)



# GETTING IN TOUCH WITH THE EU

## In person

All over the European Union there are hundreds of Europe Direct centres. You can find the address of the centre nearest you online ([european-union.europa.eu/contact-eu/meet-us\\_en](https://european-union.europa.eu/contact-eu/meet-us_en)).

## On the phone or in writing

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696,
- via the following form: [european-union.europa.eu/contact-eu/write-us\\_en](https://european-union.europa.eu/contact-eu/write-us_en).

# FINDING INFORMATION ABOUT THE EU

## Online

Information about the European Union in all the official languages of the EU is available on the Europa website ([european-union.europa.eu](https://european-union.europa.eu)).

## EU publications

You can view or order EU publications at [op.europa.eu/en/publications](https://op.europa.eu/en/publications). Multiple copies of free publications can be obtained by contacting Europe Direct or your local documentation centre ([european-union.europa.eu/contact-eu/meet-us\\_en](https://european-union.europa.eu/contact-eu/meet-us_en)).

## EU law and related documents

For access to legal information from the EU, including all EU law since 1951 in all the official language versions, go to EUR-Lex ([eur-lex.europa.eu](https://eur-lex.europa.eu)).

## EU open data

The portal [data.europa.eu](https://data.europa.eu) provides access to open datasets from the EU institutions, bodies and agencies. These can be downloaded and reused for free, for both commercial and non-commercial purposes. The portal also provides access to a wealth of datasets from European countries.



European Institute for  
Gender Equality



Publications Office  
of the European Union

[eige.europa.eu](http://eige.europa.eu)