



Poll on digital violence

Unsatisfied and helpless – how social media platforms are failing users

A new poll published by the NGO HateAid and the Landecker Digital Justice Movement illustrates how users frequently cannot rely on the reporting channels of social media platforms to remove harmful and illegal content. An overwhelming majority now demands effective internal complaint mechanisms to pursue their rights and feel safe again.

Violence on social media is pervasive, and its impacts are dreadful: from election interference, deaths caused by covid disinformation¹ to the silencing of voices of users through hate campaigns². To make matters worse, digital violence has become a mass phenomenon: Every second young adult in the European Union has already been affected by online violence themselves³. While the numbers of victims rise steadily, social media platforms continuously claim to spare no effort to protect users from online violence. A new poll commissioned by HateAid and the Landecker Digital Justice Movement now presents a clear grasp of the situation. The results are worrisome: users are unsatisfied with the quality of content moderation and find the content decisions of the platforms incomprehensible. Moreover, respondents demand low threshold options to complain about wrongful content decisions directly to the platforms.

This report, based on surveying of 10036 people, aged 18–70, across Germany, France and Sweden in February and March 2022, provides better understanding of user experiences with reporting violent content and presents analysis of their opinions, experiences, concerns, and a way forward for a human-centred approach to resolve problems arising in the online environment.

The European Union is right now working on the Digital Services Act, a new set of rules for online platforms. It is a once in a generation chance to listen to the needs of the citizens. And the poll shows that their message is clear.



Key Findings

1 A new golden standard

An overwhelming majority of 82% agree that all users should have a right to challenge content decisions of online platforms through an internal complaint mechanism.

82%

2 Unsatisfied and helpless

Reporting mechanisms of social media platforms do not deliver: 1 in 2 (49%) respondents, who has reported violent content, is unsatisfied with the notification system and the way platforms handle their notifications.

49%

3 Highly non-transparent

36% reported that they do not know what happened with their notice after submitting the report to the platform.

36%

4 Rational apathy

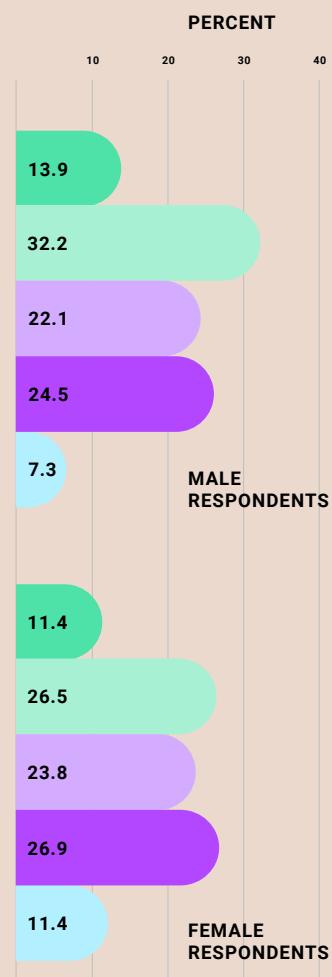
Only 3% of respondents have turned to courts to resolve issues concerning illegal content online, while 14% have considered it.

Reasons: High financial risk, lengthiness of legal proceedings and not enough information.

3%

To what extent were you satisfied with how the platform handled your notification?

- Fully satisfied
- Somewhat satisfied
- Somewhat unsatisfied
- Fully unsatisfied
- Hard to say



The status quo: Reporting systems do not deliver

For those targeted, online violence is a frightful experience, reflecting on psychological wellbeing, relationships, work and their daily lives⁴. Female journalists⁵, politicians⁶ and other public figures are especially affected, and easily become targets of organised online hate attacks with the objective to silence them and make them withdraw from public life. Official reporting mechanisms often are the only practical and affordable way to deal with the online hate. But what happens when the content moderation is not effective?

In fact, unresponsive platforms who fail to remove violence even fortify these harmful effects. Violent content that remains online is further used to shame and scare the victims. The poll shows that most reporting options with which users can report hate speech, threats and other violence on social media platforms do not deliver. Nearly 1 out of 2 respondents (49%; n= 1362) who have stated to have reported digital violence to a social media platform said that they are unsatisfied with how the platform handled their notification.



A recent study in France supports this claim: It shows that Facebook systematically failed to delete violent content during the time running up to the French election⁷.

Online violence targeted at women is often more extreme and highly sexualised. When looking at the gender divide, only 38% (n=543) of respondents identifying as women reported that they were satisfied with how platform handled their notification.

Recent findings by WWW Foundation suggested that reporting systems need to be improved to better support women, including giving more control and transparency about the report status, as well as allowing women to include context when reporting violence⁸.

Main reasons for respondents' unsatisfaction with reporting system:

- Content not being deleted – platform not taking an action (48%)
- Platform not providing a response (42%)
- Not understanding decision made by the platform (25%)

FURTHERMORE, 36% (N=1003) OF RESPONDENTS WHO REPORTED CONTENT, DID NOT KNOW WHAT DECISION WAS TAKEN BY THE PLATFORM AFTER THEIR NOTICE WAS SUBMITTED.

That is consistent with HateAid's findings from a study on Facebook content moderation in France, where the platform failed to inform the user of their decision in 39% of cases⁹.

Power to the users: Users want a say against wrongful platform decisions

Users expect quick answers from platforms and a fast reaction to delete the violent content. In the poll, users ask for an effective remedy: An overwhelming majority of 82% (n=8247) of respondents agree that all users should have a right to appeal content decisions of online platforms through an internal complaint procedure.

Everyone should have a right to appeal content decisions through internal complaint mechanisms to the platform in question.





Apart from the complaint mechanisms of the platforms, users have also the possibility to appeal content decisions in a court. But this seems to be a viable option only for a handful of users: Only 3% of respondents have turned to courts to resolve issues concerning illegal content online, while 14% have considered it. High financial risk, lengthiness of the proceedings and not enough information are the top 3 reasons why people have not turned to courts to have illegal content online removed. One respondent from France added that "a fear of violence" is another reason to not seek help in a court. These findings show that all users,

especially in cases where the platform has not acted upon the reported content, urgently need low threshold access routes to challenge platform decisions. Those are often a result of insufficiently resourced content moderation. In addition to that, users demand access to an internal complaint procedure with the platform directly, as well as the external out of court dispute settlement.

Reasons for not reporting: channels are too complicated

28% of all respondents have stated to have reported online violence through the reporting channels of social media platforms. The figure is much higher among specific groups:

- Young people are more likely to report abusive content online. 56% of surveyed 18–24-year-olds confirmed they have reported digital violence already.
- Users who already had experienced discrimination on a prior occasion also tended to report violent content more often. 48% of this group stated to have at least once reported online violence to the platform.

Have you ever reported violence on social media?



Online violence is on the rise: 41% of all respondents have witnessed it already

52% in Sweden

37% in Germany

36% in France

The main reason for not reporting online violence according to the poll results was that respondents had not been exposed to it (67%, n=4863).

MEANWHILE, 25% (N=1814) OF RESPONDENTS WHO HAVE NOT REPORTED ONLINE VIOLENCE, SAID THAT THEY DID NOT KNOW HOW TO REPORT IT OR THAT THE REPORTING IS TOO COMPLICATED.

Recommendations

1 User Redress

Give all users access to internal complaint mechanism and out of court dispute settlement, also in cases where the platform has rejected a notification or not reacted.

2 Reporting channels

Make reporting channels user-friendly and located close to the content in question.



Methodology

HateAid commissioned Respondi to survey a sample of 10 036 people, aged 18-70, in Germany (3721), Sweden (3005) and France (3310). The surveying was done in between February 22 and March 10, 2022, via Ad-hoc-Research. 49.81% of respondents identified as Male, 49.75% identified as Female, and 0.44 % as Diverse.

Notes

- 1** Due to the rounding, percentages in the text do not always add up to the percentages in the tables and graphs.
- 2** All respondents were asked if they have ever felt discriminated against based on race, gender, sexual orientation, ethnicity, disability, religion. We have used this answer to compare the results between respondents who have felt discrimination and all respondents.
- 3** Definition used for digital violence: Hate speech, insults, sending unsolicited nude photos, defamation, threats, extortion, cyberbullying, cyberstalking, defamation, threats of bodily harm or sexual violence, publication of personal data or pictures, other forms of online violence
- 4** Questions concern violence on social media platforms broadly – users were not asked to report their experiences concerning specific online platforms.

References

- ¹ I've been lied about and others get death threats. Covid has shown the power of misinformation; The Guardian; 1.1.2022; <https://www.theguardian.com/commentisfree/2022/jan/01/death-threats-covid-disinformation-public-health-expert-pandemic>; last visited : 22.03.2022
- ² Boundless hate on the internet – Dramatic situation across Europe; HateAid report; 3. 11. 2021; <https://hateaid.org/eu-survey/>; last visited: 22.03.2022
- ³ Boundless hate on the internet – Dramatic situation across Europe; HateAid report; 3. 11. 2021; <https://hateaid.org/eu-survey/>; last visited: 22.03.2022
- ⁴ The dynamics of Hate Speech and counter speech in the social media; Center for Internet and human rights; Dr. Katarzyna Bojarska; 2018; https://cihr.eu/wp-content/uploads/2018/10/The-dynamics-of-hate-speech-and-counter-speech-in-the-social-media_English-1.pdf; last visited : 22.03.2022
- ⁵ The Chilling: Global trends in online violence against women journalists; UNESCO; April 2021; <https://en.unesco.org/sites/default/files/the-chilling.pdf>, last visited : 22.03.2022
- ⁶ Hate Towards the Political Opponent: A Twitter Corpus Study of the 2020 US Elections on the Basis of Offensive Speech and Stance Detection; University of Stuttgart, Grimminger and Klinger; 02.03.2021; <https://arxiv.org/pdf/2103.01664.pdf>, last visited : 22.03.2022
- ⁷ Months before the French election, Facebook gives a free pass to far-right hate; HateAid report; March 2022; https://hateaid.org/wp-content/uploads/2022/03/HateAid_LICRA_2022_French_elections_Content_Moderation_report.pdf; last visited : 22.03.2022
- ⁸ Tech Policy Design Lab: Online Gender-Based Violence and Abuse; World Wide Web Foundation; June 2021; <https://techlab.webfoundation.org/ogby/overview>; last visited: 22.03.2022
- ⁹ Months before the French election, Facebook gives a free pass to far-right hate; HateAid report; March 2022; https://hateaid.org/wp-content/uploads/2022/03/HateAid_LICRA_2022_French_elections_Content_Moderation_report.pdf; last visited : 22.03.2022

About HateAid

The non-profit organization HateAid gGmbH was founded in 2018 and is headquartered in Berlin. HateAid offers support for victims of digital violence: legal cost funding, emotionally stabilising initial, safety, and communication counselling. The founding managing director is Anna-Lena von Hodenberg.

The study was commissioned in March 2022 by HateAid and funded by The Landecker Digital Justice Movement.

