Research Report

The Feminist Helpline Research Initiative

Holding Platforms Accountable

Table of

Contents

Table of

Contents

Introduction and Context

Technology-facilitated gender-based violence (TFGBV) has emerged as an escalating global crisis, intensifying as digital platforms become increasingly intertwined with everyday life. While expanded internet access has enhanced opportunities for political participation, economic activity, and social connection, these same online environments have simultaneously become hostile spaces for women and marginalized communities. TFGBV encompasses a wide range of harmful practices, including harassment, cyberstalking, non-consensual sharing of intimate images (NCII), disinformation campaigns, impersonation, and coordinated abuse that systematically target individuals based on gender, sexuality, religion, race, caste, and political identity. Far from being isolated virtual incidents, these harms represent the digital extension of offline patriarchal structures, where existing inequalities are amplified and weaponized through visibility, scale, and anonymity.

Global evidence highlights the severity and pervasiveness of this crisis. Amnesty International reports that 85 percent of women internet users have witnessed online violence, while 38 percent have experienced direct abuse themselves[1]. United Nations agencies further estimate that up to 95 percent of aggressive behavior and harassment online is directed at women, revealing a stark gendered imbalance in digital safety.[2] Online abuse disproportionately affects women journalists, human rights defenders, LGBTQ+ persons, minorities, and politically active women, who face systematic efforts to silence participation through intimidation, reputational attacks, and sexualized violence. These patterns reflect the continuation of patriarchal social control through digital mechanisms designed not only to marginalize voices but also to deter public engagement altogether.

Despite major social media companies' assurances of robust safety mechanisms and community guidelines, there exists a deepening crisis of trust between platforms and users. Survivors across the globe routinely report that harmful content remains online for extended periods, reporting processes are opaque and retraumatizing, and enforcement is inconsistent or culturally blind. Research has shown that the majority of women feel unprotected in digital spaces, with only eight percent reporting confidence in platform safety systems. As a result, disengagement from digital platforms is becoming increasingly common. A 2025 global survey revealed that 66 percent of women took breaks from social media due to harassment or safety concerns, while nearly half abandoned at least one platform entirely.[3] These trends signal a profound loss of faith not only in corporate accountability but also in the possibility of participating safely in digital public life.

---

1 https://www.amnesty.org/en/what-we-do/technology/online-violence/
2 https://blogs.worldbank.org/en/developmenttalk/protecting-women-and-girls-cyber-harassment-global-assessment
3 https://uplevyl.medium.com/online-harassment-is-silencing-women-what-are-we-doing-about-it-028abd248deb

In the absence of effective platform governance and accessible state remedies, feminist helplines and grassroots digital safety organizations have emerged as critical support structures for survivors of TFGBV. Operating at the intersection of care provision, digital security assistance, rights advocacy, and documentation, these helplines fill a void left by inadequate institutional responses. They offer survivors counseling, safety planning, reporting guidance, and community solidarity while simultaneously generating evidence about evolving abuse trends and systemic failures. Across the Global South and beyond, feminist helplines have become the frontline infrastructure responding to digital harm, building survivor-centered solutions grounded in empathy, privacy, and local context. This research recognizes these helplines not only as essential support systems but as knowledge leaders whose experiences provide vital insights into reimagining platform accountability and building safer digital futures.

Feminist Helplines as Digital Safety Infrastructure

Feminist helplines have emerged as a critical form of digital safety infrastructure globally, developed and operated by women's rights and digital rights organizations to respond to the growing crisis of technology-facilitated gender-based violence (TFGBV). These helplines offer essential, gender-sensitive support for women, LGBTQIA+ individuals, and other marginalized users facing online harassment, threats, and digital insecurity.[4] Often launched at local or regional levels and frequently operating with limited resources, feminist helplines fill the void left by underperforming institutional mechanisms, creating trusted, survivor-centered alternatives where mainstream platforms and legal systems have failed.

Holistic and Trauma-Informed Support Models

Feminist helplines are distinct due to their holistic and trauma-informed model of support. Rather than providing one-dimensional or purely technical services, these helplines integrate emotional care, digital literacy, legal navigation, and platform escalation into their responses. Survivors are met with empathy and active listening; support workers create a non-judgmental space where callers are first acknowledged and emotionally validated before being guided through technical solutions. The Vita-Activa helpline in Mexico, for instance, provides psychological first aid to women journalists experiencing online attacks, recognizing that trauma often precedes the need for digital intervention. This practice of placing care before response underlines the feminist ethos that informs these services.[5]

[4] https://www.digitaldefenders.org/feministhelplines/
[5] https://latamjournalismreview.org/articles/helplines-assist-women-journalists-in-latin-america-who-are-being-attacked-online/

Digital Security and Empowerment Through Accessible Guidance

Digital security guidance is another central feature of feminist helpline operations. Whether assisting a user in reclaiming hacked accounts, removing malicious content, or enhancing digital privacy, helpline teams prioritize accessibility and empowerment in their approach. Based in Brazil, MariaLab's "Maria da'Ajuda" initiative provides context-specific advice ranging from social media lockdown strategies to mitigation responses for organizational breaches. What unifies these efforts is their insistence on user agency, where recommendations are tailored, jargon-free, and aimed at building long-term digital autonomy.[6]

Legal Navigation and Platform Escalation Support

Legal and reporting support forms a crucial component of these helplines' work. Survivors navigating complex systems of law enforcement or opaque content reporting tools often face barriers in understanding their rights or articulating abuse within institutional frameworks. Helplines bridge these gaps by offering guidance on legal options, assisting with documentation, and, in many cases, escalating content removal through direct liaison with tech platforms. DRF's Digital Security Helpline in Pakistan works closely with both the National Cyber Crimes Investigation Agency (NCCIA) and platforms like Google, Facebook and TikTok to push for takedowns and report resolution.[7] This dual engagement, between state institutions and global tech companies, represents a strategic use of multiple avenues to maximize protection and accountability.

Embedded Legal Support and Pro Bono Aid

Legal aid is often embedded in helpline infrastructure, either through in-house counsel or partnerships with legal clinics and NGOs. DRF's model includes legal professionals who offer pro bono advice and representation in some cases, resulting in real-life consequences for perpetrators and legal victories for survivors. The integration of legal referrals with mental health and digital safety services highlights the multidisciplinary nature of feminist helplines, which recognize that TFGBV demands layered, not linear, responses.

[6] https://capiremov.org/en/experience/a-digital-security-help-line-by-feminists-from-brazil/
[7] https://digitalrightsfoundation.pk/digital-security-helpline-annual-report-2024/

Survivor-Centered Values and Privacy-First Design

Feminist digital helplines distinguish themselves through values-driven practices rooted in survivor agency, privacy, and contextual awareness. The support is not only confidential but also consent-based, personal data is stored minimally, access is tightly controlled, and survivors retain full autonomy over how their cases are handled. Many helplines have adopted encrypted communication tools like Signal to protect users' identities, particularly in regions where fear of reprisal or social stigma deters survivors from seeking help.[8] WhatsApp is also used by helplines, but for its greater accessibility with the general population rather than enhanced security. The emphasis on privacy, both as principle and practice, fosters the trust that is central to these helplines' credibility and reach.

Community Embeddedness and Trust-Building

Community embeddedness further strengthens this trust. Helplines are often staffed by individuals who speak local languages and understand regional gender dynamics, allowing them to respond with cultural nuance and relevance. Whether it is addressing honor-based online blackmail in a particular region or state-sponsored trolling in authoritarian regimes, feminist helplines craft responses informed by the intersectional realities of those they serve. In Ecuador, Navegando Libres exemplifies this approach by tailoring its services to women, children, and LGBTQ+ survivors whilst adapting to multiple threats that fall under sexual digital violence such as AI-generated abuse.[9] Over time, helplines cultivate deep-rooted relationships within their communities, becoming reliable and accessible resources, particularly for those who might otherwise remain invisible to formal systems.

Emergency Response and Real-Time Crisis Management

Beyond individual casework, feminist helplines serve as critical nodes in the broader ecosystem of digital rights advocacy. They operate as emergency responders, offering real-time interventions during digital attacks. They also function as observatories, collecting, analyzing, and publishing data on TFGBV that is often unavailable elsewhere. DRF's helpline, for instance, has recorded over 20,000 cases between 2016 and 2024, documenting patterns such as increased financial fraud and gendered misinformation.[10] These insights form the foundation for advocacy efforts, including recommendations for legislative reform and pressure on tech companies to improve platform design and content moderation.

---

[8] https://www.digitaldefenders.org/the-feminist-helpline-for-survivors-of-tech-facilitated-gender-based-violence-in-ecuador/
[9] https://www.digitaldefenders.org/the-feminist-helpline-for-survivors-of-tech-facilitated-gender-based-violence-in-ecuador/
[10] https://digitalrightsfoundation.pk/digital-security-helpline-annual-report-2024/

Data Collection as a Tool for Advocacy and Accountability

Collectively, feminist helplines are amassing a wealth of data on the patterns and prevalence of technology-facilitated GBV, data that was rarely documented before. Reports produced by helplines like DRF, Acoso Online (Chile), Luchadoras (Mexico), and Internet Bolivia SOS have mapped TFGBV trends regionally and globally.[11] In 2024, the Digital Defenders Partnership launched the Feminist Helpline Index, a collaborative repository of helplines across 13 countries, reflecting a commitment to shared knowledge and strategic learning.[12] These data resources are not just informational, they are foundational to evidence-based advocacy, providing civil society with the tools to challenge corporate negligence and push for responsive, survivor-informed design in digital governance.

Policy Advocacy and Structural Reform

Feminist helplines don't stop at helping survivors endure harm; they actively campaign for systemic change. Their frontline experience lends credibility and urgency to their advocacy. Many helpline organizations engage lawmakers, tech companies, and international forums to address the root causes of online GBV. Digital Rights Foundation (DRF) , for instance, has used its annual reports to call for reforms to Pakistan's cybercrime legislation, the Prevention of Electronic Crimes Act, 2016, gender-sensitization in law enforcement, and investment in digital literacy. These calls echo across borders as feminist helplines frame digital safety as a human rights imperative.

Feminist helplines represent a transformative model of care-driven, community-rooted, and survivor-led digital governance. They challenge reactive and punitive state or corporate responses with proactive, holistic systems grounded in trust and dignity. In a digital landscape increasingly marked by hostility and inequality, these helplines stand as resilient infrastructures of safety, solidarity, and justice. As TFGBV continues to evolve in scale and form, the work of feminist helplines signals the need and the possibility of building more equitable and humane digital futures.

[11] https://www.digitaldefenders.org/the-feminist-helpline-for-survivors-of-tech-facilitated-gender-based-violence-in-ecuador/
[12] https://capiremov.org/en/experience/a-digital-security-help-line-by-feminists-from-brazil

Literature Review

Violence enacted in online spaces is often reflective of existing structures of discrimination and social violence that are reproduced by intersecting regimes of patriarchy, coloniality, and marginalization.[13] Patterns of exclusion are often perpetuated through various forms of online violence that fall under harassment, gender-based violence and intimidation, and other measures threatening the privacy of already precarious communities. While online spaces also provide avenues for feminist activism and communities, their sustainability tends to be extremely fragile due to the weak privacy protections and user safety mechanisms of social media applications. Excessive collection of data at an unprecedented and increasingly centralized scale has proven to further exacerbate privacy concerns for vulnerable communities[14], as demonstrated by the rapid upsurge in stalking[15], online harassment, and identity theft that threatens existing vulnerable communities. Within this context, Feminist Helplines have emerged from grassroots civil society organizations as one alternative infrastructure emphasizing feminist solidarity against profit-centric platform biases perpetuating exclusion and violence.

TFGBV and Platform Negligence

The growing integration of technology into daily life has intensified concerns about the ways digital spaces enable and amplify gender-based violence, particularly in the form of TFGBV. This issue presents a grave threat to women and gender minorities as online spaces mirror the patriarchal system prevalent offline, which privileges certain identities while marginalising others. TFGBV encompasses a range of harmful behaviours, including cyberstalking, online harassment, non-consensual sharing of intimate images, and online threats.[16] Between 2021[17] and 2024[18], data from the Digital Security Helpline at DRF consistently reflected the gendered nature of rights violations in Pakistan. Across the four years, women constituted the majority of complaints, making them the most prominently affected group in cases of TFGBV, averaging around 60 percent of reported cases every year. Meanwhile, gender minorities, including transgender and non-binary individuals, represented around 1.1 percent of reported cases each year. Underreporting from transgender individuals suggests potential barriers to seeking help, including fear of discrimination, lack of trust in reporting mechanisms, and systemic legal challenges.

These patterns are also consistent with global research. A 2022 report by UN Women indicates that the prevalence of violence against women online ranges from 16 percent

[13] https://www.amnesty.org/en/what-we-do/technology/online-violence/
[14] https://epic.org/issues/consumer-privacy/social-media-privacy/
[15] https://www.forbes.com/sites/simonchandler/2019/10/11/social-media-proves-itself-to-be-the-perfect-tool-for-stalkers/
[16] https://www.womankind.org.uk/wp-content/uploads/2024/11/TFGBV-Policy-Brief.pdf
[17] https://digitalrightsfoundation.pk/wp-content/uploads/2022/05/helpline-annual-report-2021-1.pdf
[18] https://digitalrightsfoundation.pk/wp-content/uploads/2025/05/Digital-Security-Helpline-Annual-Report-2024.pdf

to 58 percent.[19] Similarly, a global survey conducted by the Center for International Governance Innovation (CIGI) found that of roughly 18,000 respondents across 18 countries, nearly 60 percent had experienced some form of online harm, and 25 percent said they were targeted because of their gender identity.[20] The report also states that the most pervasive and severe experiences of online harm occur among transgender and gender- diverse people. These figures, both locally and globally, are a testament that technology is not neutral. What should be a space for freedom of expression has become a medium through which existing social hierarchies are reproduced and amplified.

Despite the rising prevalence of violence against women and gender minorities online, one of the most significant challenges in curbing the issue is inaction by social media companies.[21] Reports of abuse and harassment experienced by the users are often dismissed or ignored. Community guideline standards have proven quite ineffective in protecting fundamental rights, leaving users vulnerable and without access to formal recourse. The Digital Security Helpline's Annual Report 2024[22], documented one such experience of a survivor of Image Based Abuse. The complainant reported a profile containing images and personal identifying information to Meta multiple times, but the platform deemed them non-violative of community guidelines, ignoring cultural contexts where such content could put individuals at serious risk. Attempts to seek help from local cybercrime authorities were frustrating and exhausting, requiring repeated visits and extensive documentation, while the perpetrator continued posting content rapidly. The survivor and their family, by extension, felt helpless and overwhelmed by the risk of exposure and potential social consequences. The accounts were taken down through escalation requests made using the Helpline's Trusted Partner Channel.

Increasingly, across the region perpetrators' tactics have evolved to bypass platform policies by exploiting regulatory gaps. Cases of such strategic evolution are evident in the 2025 annual report of Rati Foundation's "Meri Trustline" initiative[23]. The report remarks on the phenomena of coordinated "networked" abuse, which relies on memes, edited content, and voiceovers that tap into viral social media trends across platforms. In most of their documented cases, female content creators on social media discovered AI-generated versions of their original reels. These involved highly explicit and sexualized frames inserted at regular intervals of their original content, and had been posted by multiple accounts. Splicing methods in most cases allowed these creators to bypass automated detection, and while escalation channels did often lead to content takedown, this has become increasingly difficult with the mass proliferation of AI technologies.

---

[19] https://knowledge.unwomen.org/sites/default/files/2022-10/Accelerating-efforts-to-tackle-online-and-technology-facilitated-violence-against-women-and-girls-en_0.pdf

[20] https://www.cigionline.org/articles/tech-facilitated-gender-based-violence-is-an-international-human-rights-concern-finds-new-research/

[21] https://ijlsss.com/the-impact-of-social-media-in-combating-gender-based-violence/#:~:text=Social%20media%20also%20functions%20as,healthcare%2C%20and%20violence%20against%20women.

[22] https://digitalrightsfoundation.pk/wp-content/uploads/2025/05/Digital-Security-Helpline-Annual-Report-2024.pdf

[23] https://ratifoundation.org/wp-content/uploads/2025/05/Meri-Trustline-Year-2-Report-Final-Version_compressed.pdf

With both state systems and tech platforms often set up in ways that make it extremely difficult for survivors to receive assistance, Helplines across the globe play a vital role by stepping in and advocating for vulnerable individuals whose reports are dismissed due to the lack of understanding of cultural nuances. The specialized and contextually-specific nature of the Helplines infrastructure also enables them to identify new strategies of digital violence not accounted for by centralized private platforms.

As is evident from these statistics and anecdotes, private platforms clearly prioritize user-engagement over user safety.[24] As the intersection between digital spaces and technological monopolies becomes almost indistinguishable, the former's governance landscape is subsumed under the logic of profit maximisation.[25] Rights-based principles are often compromised in the process in favor of sensationalist or vitriolic sentiment that overwhelms digital landscapes. Moreover, attempts to expand community safety are quite limited, as is evident in the case of content moderation. For example, Facebook's content moderation laws do not account for unique expressions of TFGBV in localized contexts and settings.[26] Existing centers for complaints and user safety are not well-resourced or trained enough to cater to such forms of violence, forcing women to either exclude themselves from online spaces or limit their digital footprint through anonymity.

Rights-Based Internet Governance and Feminist Helplines

Rooted in principles of empathy and survivor-centric care, Feminist Helplines across the world have assisted in mitigating the damages caused to marginalized gender communities due to the exclusionary logics of mainstream social media platforms.[27] This infrastructure is guided by careful, and contextually-specific understandings of how gender identities structure access and safety barriers for individuals in digital spaces. According to a report from the Digital Defenders Partnership, these institutions perform a variety of functions that range from resource-creation, and legal and psychological assistance, to acting as a liaison between victims and state institutions.[28] Most Feminist Helplines act as 3rd party partners and trusted flaggers for social media platforms. This means they are able to report content to platforms on a priority notice through a contextually-specific analysis of how specific forms of online content cause harm to vulnerable communities[29]. Trusted Flaggers or Partners are usually recognized in an official sense by platforms like YouTube and Meta, allowing Helplines ease of access in escalating cases ignored by traditional platform moderation channels. The Digital Security Helpline, along with the 14 feminist helplines that participated in this research,

[24] https://ishr.org/how-tech-companies-foster-cyber-abuse-through-negligence/
[25] https://www.apc.org/sites/default/files/the_hidden_codes_that_shape_our_expression.pdf
[26] https://ishr.org/how-tech-companies-foster-cyber-abuse-through-negligence/
[27] https://www.digitaldefenders.org/feministhelplines/
[28] https://www.digitaldefenders.org/wp-content/uploads/2022/10/VMD_final-EN.pdf
[29] https://www.inhope.org/EN/articles/what-is-a-trusted-flagger

are all examples of Feminist Infrastructures performing these multiple roles. In the Global South, where legal channels intertwined with platform negligence leave citizens vulnerable to arbitrary forms of violence, such Helplines provide an alternative mode of recourse, psychological assistance, and gender disaggregated research output for policy.

Helplines often emerge within localized, often a national-level grassroots context, allowing for channels of online community safety that women and other gender minorities have secure and regular access to. The organizational structures and pertinent obstacles faced by these feminist helplines across national boundaries constitute the focus of our research. While recognizing the implicit critique their very emergence carries for mainstream social media companies, we also evaluate the role these alternative networks of feminist solidarity and infrastructure can play in developing guiding principles for new, genuinely democratic digital spaces.[30] According to the Feminist Principles of the Internet Project, new modalities of platform governance must be built around expanding democratic space across a wide range of vulnerable demographics. This necessitates an intersectional perspective on digital governance that accounts for structural discriminations and socio-economic injustices within platform governance frameworks.[31] New organizations such as Mastodon have already begun experimenting with technologies that enable alternative paradigms for digital spaces. In contrast to mainstream platforms that involve highly centralized and top-down policymaking, decentralized internet platforms are guided by principles of microdemocratic decision making, federalized community-run platform units, multiple context-specific frameworks for content moderation, and stringent safeguards of user safety.[32] Still, new applications such as Bluesky occasionally run into issues of scalability, and have yet to fully challenge the monopolistic control of traditional social media platforms.

We are witnessing radical transformations in digital technologies and increasing discontent towards the direction of centralized governance taken by traditional social media platforms. Against this background, our research hopes to derive insights gained from organizations acting as frontline responders to TFGBV and digital rights violations. Our interventions aim at providing concrete and actionable frameworks for governance rooted in a rights-based approach that prioritizes intersectionality and social welfare, inclusive of vulnerable gender identities and marginalized communities.

[30] https://www.apc.org/en/kefir-0
[31] https://www.cigionline.org/articles/building-feminist-internet/?utm_campaign=Ann%20Cathrin%27s%20Digital%20Digest&utm_medium=email&utm_source=Revue%20newsletter
[32] https://carnegieendowment.org/research/2025/03/fediverse-social-media-internet-defederation?lang=en

Methodology

The Feminist Helpline Research Initiative adopted a mixed-methods research design to capture both the systemic trends shaping TFGBV and the frontline realities of feminist helplines engaging with digital platforms. DRF, in partnership with the Social Web Foundation (SWF), conducted the research, which was implemented in 2024-2025 across 14 countries in Asia, Sub-Saharan Africa, Latin America, and, the Middle East and North Africa. The study combined a desk-based literature review, a quantitative cross-regional survey conducted through Google Forms with feminist helplines, and qualitative in-depth interviews to generate both comparative and contextual insights into survivor support pathways, platform accountability mechanisms, and emerging models of alternative platform governance.

The research begins with a comprehensive literature review, including academic publications, civil society research, reports by international organizations, and platform transparency disclosures on TFGBV, online content moderation, survivor reporting mechanisms, and digital safety governance models. The review established the global prevalence of gendered online abuse, documented the erosion of trust in major platforms' safety measures, and identified significant gaps in evidence addressing community-led digital safety interventions. This foundational research informed the development of survey instruments and interview protocols, enabling the project to build upon existing work while addressing critical knowledge gaps related to frontline practitioner experiences.

Research Objectives

The primary objectives of this research were:

01 **To document the lived experiences of feminist helplines engaging directly with digital platforms and institutional channels to support survivors of TFGBV.**

02 **To analyze the effectiveness of existing reporting and content moderation systems, including survivor accessibility, responsiveness, and enforcement outcomes.**

03 **To analyze the effectiveness of existing reporting and content moderation systems, including survivor accessibility, responsiveness, and enforcement outcomes.**

04 **To identify gaps in platform accountability and response mechanisms, particularly across diverse regional and political contexts.**

05 **To examine emerging threat vectors, including the role of AI-enabled abuse, deepfakes, impersonation, and automated disinformation campaigns targeting women and marginalized communities.**

06 **To assess community-driven alternatives to mainstream platform governance, including feminist, decentralized, or cooperative platform models oriented toward safety by design.**

07 **To co-create evidence-based, advocacy-focused recommendations that center on survivor needs and inform more transparent, accountable, and rights-respecting digital governance frameworks.**

These objectives guided the development of the research and shaped the findings across all phases of the project.

Quantitative Research

The quantitative component consisted of an online survey administered to 14 feminist helplines operating across 14 countries representing diverse geographic regions, including Asia, Sub-Saharan Africa, Latin America, and the Middle East and North Africa. Participating organizations were selected based on their direct engagement in providing support to survivors of TFGBV and their experience interacting with social media platforms' reporting and escalation systems. The participants were sent out surveys via Google Forms with about 10 days time to complete.

01 **Common forms of abuse reported by survivors include harassment, cyberstalking, non-consensual intimate image abuse, impersonation, financial extortion, hate campaigns, and AI-generated content.**

02 **Trends in case volume and frequency, including the evolution of complaints over time.**

03 **Survivor interactions with platform reporting tools, such as ease of use, accessibility, and procedural barriers.**

04 **Reported outcomes of takedown requests, assessing success and failure rates, speed of responses, and transparency in decision-making.**

05 **Barriers to survivor justice include platform inaction, unclear reporting requirements, institutional hesitancy, and fear of backlash.**

06 **Helpline perceptions of platform accountability, including trust in enforcement mechanisms and responsiveness to escalated complaints.**

This list of helplines that participated in the survey were the following:

	Helpline Name	Organization	Region
1	Meri Trustline	Rati Foundation	India, South Asia
2	Tech4Peace	Tech4Peace	Middle East and North Africa (MENA) region
3	Anonymous contribution	Anonymous contribution	Indonesia
4	Anonymous contribution	Anonymous contribution	West Asia and North Africa (WANA) region
5	Navegando Libres	Taller de Comunicación Mujer	Ecuador, South America
6	Prathya	Hashtag Generation	Sri Lanka
7	NEON (No to Online Violence)	BaBe	Croatia

8	Sexual Image Abuse Reporting Center, Ministry of Health and Welfare	Taipei Computer Association (TCA)	Taiwan
9	BLAST Helpline Number	Bangladesh Legal Aid and Service Trust (BLAST)	Bangladesh
10	Online Ambulance (by Stop Online Harm)	Stop Online Harm	Southeast Asia (Myanmar, Thailand, Laos, Cambodia)
11	Anonymous contribution	Anonymous contribution	Sub-Saharan Africa
12	Fembloc	Fembloc	Spain / Catalonia
13	Centro S.O.S. Digital.	Internet Bolivia	Bolivia
14	KURAM (It means 'Keep Me Safe')	Kuram	Nigeria

Data has been aggregated across regions and the analysis is based on recurring patterns, regional differences, and cross-platform trends. There was special attention given to documenting inconsistencies between platform safety claims and frontline experiences of moderation outcomes.

Ethical safeguards were integral to the research design. Participation was voluntary, organizations could opt for anonymity, and no personally identifiable survivor data was solicited or included. All participants participated in an online introductory call to familiarize themselves with the topic and the survey they'd be taking. All responses were stored securely using privacy-first protocols, with restricted access to core research personnel only.

Qualitative Research

To deepen the understanding of the trends identified through the survey, the research incorporated semi-structured interviews with representatives from five feminist helplines spanning multiple regions. Interviews explored the practical and emotional realities of frontline digital safety work and were focused on:

01 **Survivor pathways through reporting processes, capturing procedural complexity and emotional impact.**

02 **Direct experiences escalating harmful content to platforms, including communications with trust-and-safety teams and challenges obtaining removals.**

03 **Documented harms caused by slow, inconsistent, or opaque moderation, particularly the mental health consequences for survivors when abuse remains online.**

04 **Context-specific challenges, such as language barriers, political constraints, limited platform localization, or risks of offline retaliation.**

05 **Examples of advocacy successes and failures, including cases that resulted in policy changes or legal remedies as well as instances where survivors were denied justice.**

06 **Exploration of alternative governance visions, addressing feminist and community-led platform models that prioritize care, consent, and accountability over profit-driven engagement algorithms.**

Interviews were recorded with consent and transcribed thematically to identify cross-cutting issues related to platform governance, survivor protection gaps and promising intervention strategies.

Case Study Integration

DRF's own Digital Security Helpline served as a longitudinal case study for this research. Between 2016 to 2025, the helpline has recorded more than 22,716 complaints, including 3171 complaints in 2024 alone. Analysis of this long-term dataset provided insight into shifting online abuse patterns, platform escalation outcomes and emerging forms of harm, particularly with the growth of financial extortion schemes, impersonation and AI-enabled image-based abuse. This research report offers practical operational lessons the helplines have adopted when states and platforms do not facilitate and contextual grounding for interpreting cross-regional survey and interview findings.

Research Limitations

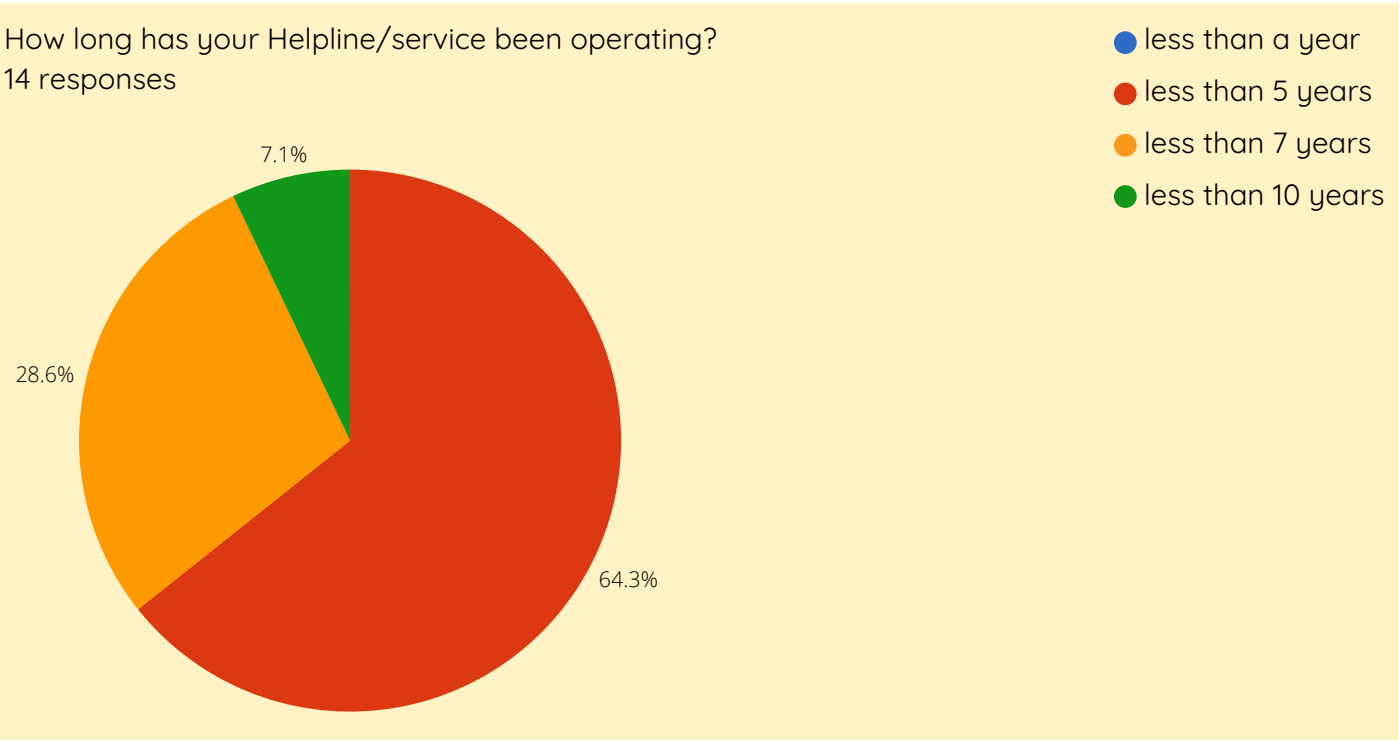
Several limitations shaped the scope of this research. While the participating helplines provided valuable geographic diversity, the sample represents only a portion of the feminist digital safety initiatives operating worldwide, and therefore, the findings may not fully capture the breadth of global experiences. Data collection was further constrained by time limitations and the high operational workload faced by helplines toward the end of the year, which affected the depth and consistency of participation across organizations. Additionally, while helplines shared frontline data and insights, not all available datasets could be integrated into the study due to variations in documentation formats and incomplete trend reporting. Efforts to incorporate perspectives directly from technology platforms on their moderation practices were limited by both time constraints and restricted access to corporate data and interview participants. Finally, the rapid evolution of artificial intelligence technologies means that patterns related to generative abuse are likely to shift faster than research cycles can track.

Findings: Trends in TFGBV

Demographics

For this report, 14 feminist helplines completed a quantitative survey designed to help the research team understand their operational structures, their experiences responding to TFGBV, and the challenges they encounter when escalating harmful content to social media platforms. Before delving into these experiences, DRF sought to establish a baseline understanding of how long each helpline had been operating and which platforms they primarily engaged with.

Survey findings show that 7.1% of participating helplines had been operating for less than 10 years, while 28.6% had been active for under 7 years. Notably, the majority 64.3% were relatively new initiatives, having established and run their helplines within the past five years. This highlights both the emerging nature of feminist digital safety infrastructures globally and the growing demand for community-centered responses to online gender-based violence which has grown over the years.



Moreover, all 14 feminist helplines served a diverse range of communities. Thirteen of the fourteen helplines supported women, while nine provided services to children, and ten worked with activists and human rights defenders. In addition, nine helplines supported journalists, and eight assisted minority groups such as religious and ethnic communities. Approximately ten helplines also catered to gender minorities, including trans and non-binary individuals.

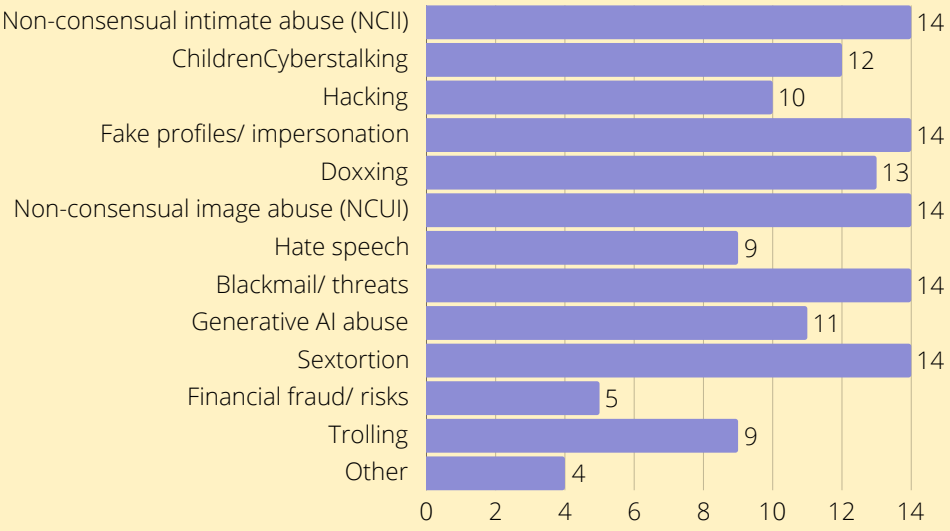
These numbers highlight both the breadth of harms faced by marginalized groups and the critical role feminist helplines play in addressing technology-facilitated violence across multiple vulnerable communities.

Who does your Helpline/services cater to?
14 responses

In terms of the types of online harms addressed, all 14 helplines supported survivors facing non-consensual intimate image abuse, fake profiles or impersonation, blackmail or threats, and sextortion. Twelve helplines assisted users experiencing cyberstalking, ten responded to cases involving hacking, and thirteen dealt with incidents of doxing. Additionally, eleven helplines reported handling generative AI–enabled abuse, while nine each supported cases related to hate speech and trolling. Five helplines also assisted individuals facing online financial fraud or related risks.

These patterns highlight the broad spectrum of technology-facilitated abuses handled by feminist helplines and the evolving nature of digital harms impacting their communities.

What problem areas does your Helpline assist with? If technology-facilitated gender based violence, then please select the types below
14 responses



From the survey, out of the 14 participating helplines, 9 reported being Meta Trusted Partners, 8 held partnerships with TikTok, 5 with Google, and 2 with X. Three helplines indicated that they were not trusted partners with any platform. Additionally, several helplines noted partnerships with other platforms including Reddit, Snapchat, Pornhub, Telegram, Line, and Microsoft.

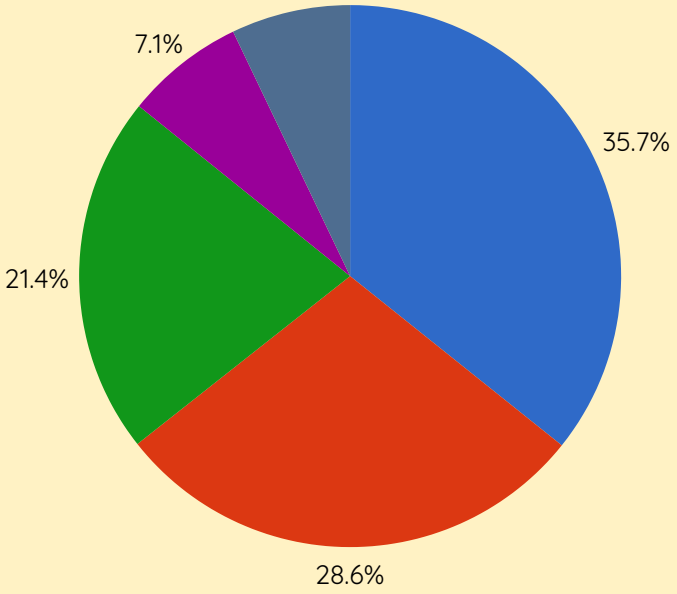
These partnerships reflect the varied levels of platform engagement across helplines, as well as the reliance on trusted channel relationships to escalate and resolve cases of online harm.

Is your Helpline/organization a Trusted Partner with any social media or tech companies?
14 responses

The survey also asked helplines which social media platforms they most frequently escalated cases to. The responses indicated that 35.7% of escalations were directed to Facebook, 28.6% to Instagram, 21.4% to TikTok, 7.1% to Google/YouTube, and 7.1% to Telegram.

These figures highlight where the largest volume of online harms is occurring and which platforms helplines most frequently rely on for urgent intervention.

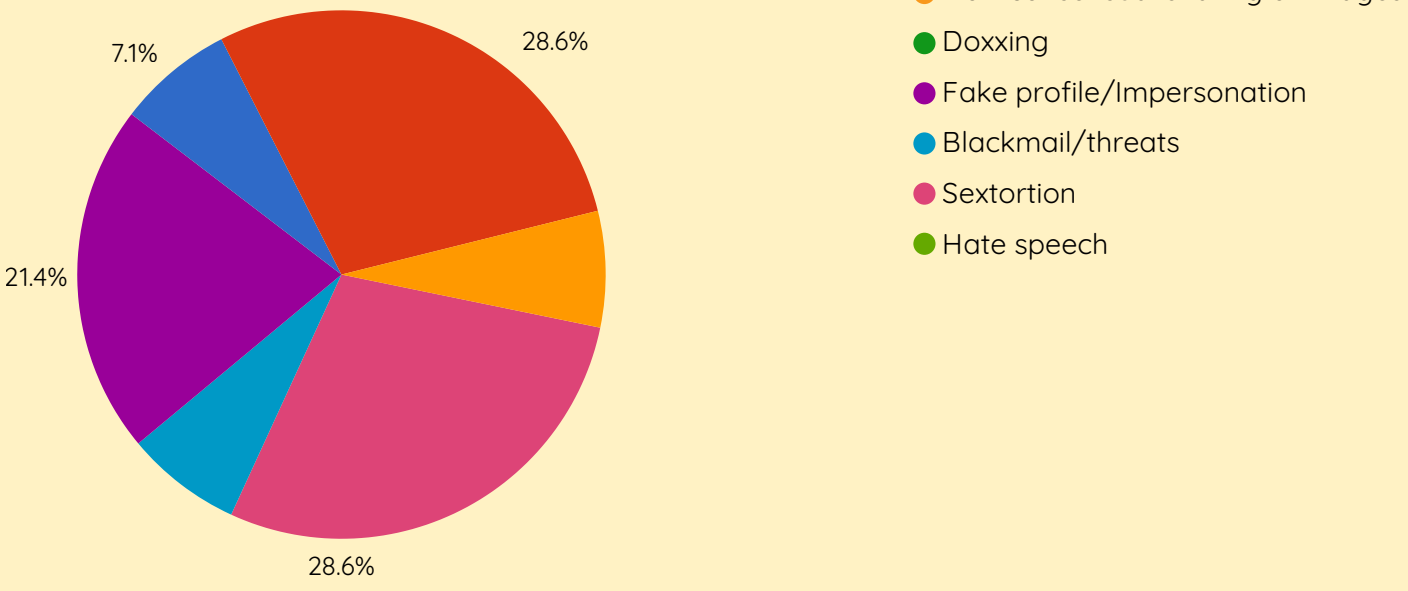
Which social media/tech company do you make the most escalations to?
14 responses



Regarding the types of complaints most frequently escalated by helplines, 28.6% involved sextortion—a form of online blackmail in which perpetrators threaten to share intimate images or information unless victims comply with demands (often monetary or sexual). Another 28.6% of escalations related to non-consensual intimate image abuse, while 21.4% concerned fake profiles or impersonation. Additionally, 7.1% of escalations were linked to hacking, 7.1% to blackmail or threats, and 7.1% to the non-consensual sharing of images.

These patterns reflect the severity and prevalence of image-based abuse and coercive online harms faced by users across the regions where these helplines operate.

Which type of complaint is most often escalated?
14 responses



Increasing Abuse Patterns

As emerging tech has completely transformed the online landscape, we see abuse patterns increasing and transforming in tandem. Among these patterns are image-based abuse, non-consensual AI-generated content, and the growth of deepfake sexual content targeting journalists, activists, and public figures.

Image-based abuse is a dominant category of TFGBV that emerged from our findings interviewing all helplines. This category takes many forms, depending on region and context.

Our respondent from Tech4Peace, an Iraqi helpline, for example, spoke about how in their region online users sometimes impersonate other women because it is easy to obtain IDs of women in Iraq as the personal data of Iraqis is not very secure. Impersonation cases were also flagged by RATI, who run the Meri Trustline, which operates in India. According to our respondent, helplines cannot escalate cases of impersonation as Trusted Flaggers, since these are not considered TFGBV within all platform policies; however, in many cases of impersonation in India, a man will impersonate a woman who is not allowed to have an account, and send a friend request to one of her family members, which then becomes TFGBV. Therefore, impersonation can be and is often weaponised as image-based abuse.

A recent trend in India flagged about a month ago by Meri Trustline involves videos featuring voyeuristic footage of women doing menial tasks such as walking on the street, working in fields, or household chores like shelling peas. The women are fully clothed, but the camera angles are invasive and often captioned with incest-themed keywords. These are not edited from public videos; they are secretly recorded, making the person behind the camera the primary perpetrator. According to our respondent, the trend felt “like a horror movie: you’re asleep and suddenly there’s a video of you sleeping.” Despite repeated reporting of this pattern to Meta over several weeks through established escalation channels, responses were largely limited to action taken on individual, escalated accounts. Broader measures to address the pattern at a trend or network level were not observed during this period, allowing similar content to continue circulating.

Another example of how seemingly harmless image sharing can become dangerous is seen in India, where queer partners sometimes refuse to delete intimate public photos after a breakup. Even if the content is not explicit, it can still have a toxic impact on the victim, especially when one partner reaches a “marriageable age” and fears discrimination from potential male suitors. The ex-partner’s refusal to remove the content, often for sentimental reasons, can leave the victim vulnerable.

Experiences from Indonesia highlighted an ongoing job scam that emerged in 2024, as reported by our respondent from SAFEnet Indonesia. The victims are women who are trying to find employment, but the perpetrators, posing as employers, ask them for "sexy photos", which will supposedly enable them to acquire that particular job. However, the job does not exist, and the perpetrators are using this scam to obtain blackmail material to extract money from the victims. Most victims are between 18 and 25, with women making up the majority. However, videocall sex-cam scams, where a random caller plays an explicit video, screenshots the victim's reaction, and then uses it for sextortion, are increasingly targeting men as well. These scams have been active on WhatsApp and Telegram in Indonesia since 2023 with multiple attempts from the helpline to escalate them to these platforms.

Our interview with the Cyberclinic Helpline by the Ghana Internet Safety Foundation revealed that there are also many cases of sextortion in Ghana. This is a type of image-based abuse in which victims do not know how to protect themselves. A partner may ask for explicit photos or videos during the relationship and later use them to demand money or to pressure the victim to return after a breakup, which was very similar to the experiences of the other helplines interviewed for this research.

However, the sharing of intimate images is also often how marginalised communities, such as the transgender community in Ecuador, find work. It then becomes an additional responsibility for helplines like Navegando Libres in Ecuador to help "give [the trans community] back tech" by teaching communities how to use tech as a tool for safe communication. This involves a judgement-free understanding of how these communities work with their intimate images, and finding out the safest ways for them to share these.

The escalation of non-consensual AI-generated content is a similarly growing observable pattern, as evidenced by our findings from the experiences of all the helplines participating in this research.

In 2024, Meri Trustline noticed a large volume of AI nudify content. While India has a very coercive takedown mechanism, the problem here is that the perpetrator changes or revolves every year or so, and according to our respondent, "Recidivism is a problem that is never addressed". Secondly, the harmful content is dumped on porn websites as opposed to other social media platforms, so when one searches the victim's name on the web, the results show porn, even if the videos are of someone else. Another trend noticed was the use of the voices of popular singers to post AI-generated explicit songs, which often leads users to links where someone has doxxed their ex-girlfriend. This constitutes TFGBV, and there have been cases where Meri Trustline has been able to prove that violation. Traumatic incidents of gender-based violence are often also turned

into monetised AI content, as perpetrators intuit an uptick of 'hits' for such incidents. For example, following the R.G. Kar Medical College rape and murder case in India[33], monetised AI videos started circulating with captions like "My last day at R.G. Kar", or videos captioned "Imagine she was happy" with the victim's face being photoshopped onto a dancing actress.

Similarly, morphing cases are on the rise in Indonesia following the rise in deepfakes, as evidenced by our respondent from SAFEnet Indonesia. This involves the use of AI to generate non-consensual explicit or compromising images to threaten and exploit victims.

In Iraq, nonconsensual AI generated content often assumes a political shape, where AI-generated nude or near-nude images and videos of politicians, especially hijabi politicians, are circulated during crucial times such as near elections. In cases of AI videos, Meta usually sends these to third party fact checker programs, and will not take the video down if it is not flagged by them. However, in a particular case of similar GenAI videos of a hijabi politician circulating, Tech4Peace's reporting yielded faster results than fact checkers', and the video was taken down, because they were able to prove that harm was imminent. According to our respondent, political and other influential families in Iraq are targeted by targeting the women in their family in the first instance, or targeting public figures who are women, because the way the societal norms are set up mean this is often the most 'effective' strategy to threaten and silence dissent or opposing viewpoints.

With revolving perpetrators, new strategies to avoid being caught, and novel abuse patterns, helplines are often thrown for a loop or have to play endless catch up with the newest TFGBV trends which are now becoming even more sophisticated due to the use of evolving AI tools.

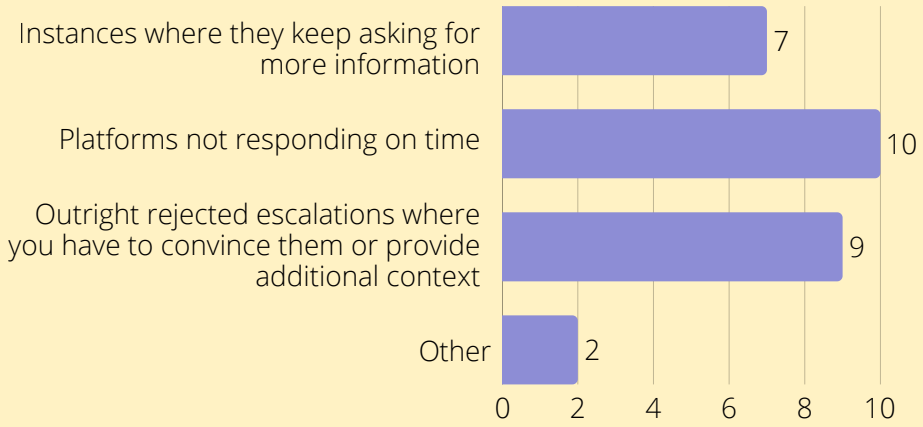
[33] https://www.bbc.com/news/articles/cwy7dyq4ezyo

Platform Failures

Findings revealed that in the face of disturbing patterns of violence and abuse, platforms fail to provide adequate and timely responses to reports and escalation requests. This failure includes inconsistent enforcement, language and cultural blind spots, and the reporting process itself being retraumatising for victims.

Most of the interviewees expressed dissatisfaction with some platform response rates. When asked what difficulties helplines face when communicating with platforms, 76.9% of our survey respondents' major difficulty was platforms not responding on time.

What difficulties do you face when communicating with platforms?
13 responses



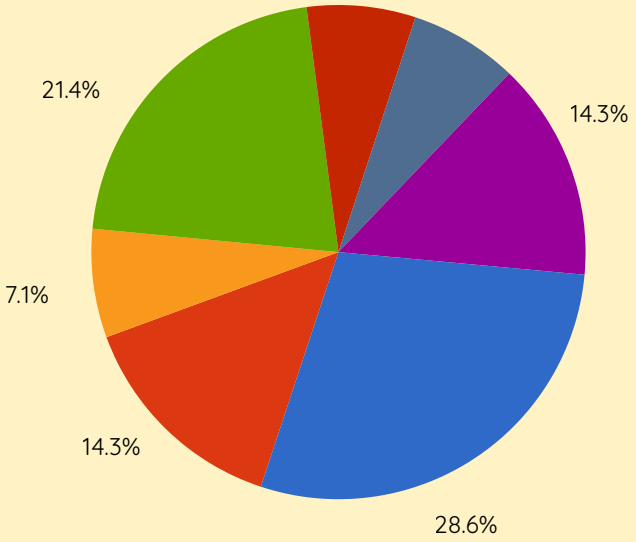
Meta had the weakest escalation response among Trusted Partners: Tech4Peace reported waits of up to a week, Navegando Libres from Taller de Comunicación Mujer reported over a week to a month despite the claim that "Facebook and Instagram and Whatsapp (Meta) has seen the most violence", and SAFEnet Indonesia noted response times of more than 10 days. In comparison, TikTok performed well across helplines. Tech4Peace reported less than an hour for TikTok takedowns, and SAFEnet Indonesia reported that TikTok's appeal process was better than Meta's, as one can appeal more than 2-3 times, which is Meta's appeal limit.

Meta's recent restructuring of their content moderation team, and the hiring of contractual employees were cited as possible reasons for the increasing delays in the past 1-2 years.[34] According to Tech4Peace, the new team often denies claims of restrictions on content even with video evidence, pointing to how they might not have the required access to check for restrictions in the first place.

[34] https://www.texastribune.org/2025/01/07/texas-meta-content-moderators-fact-checking/

Another reason cited by the helplines for the variation in response times is the type of content, with fully nude or child abuse content taken down the fastest, within one to forty-eight hours. According to the results of the survey, NCII complaints receive the greatest attention and resolution rate, followed by hacking cases for marginalized groups.

Which type of complaint receives the greatest attention and resolution rate by social media/tech companies?
14 responses

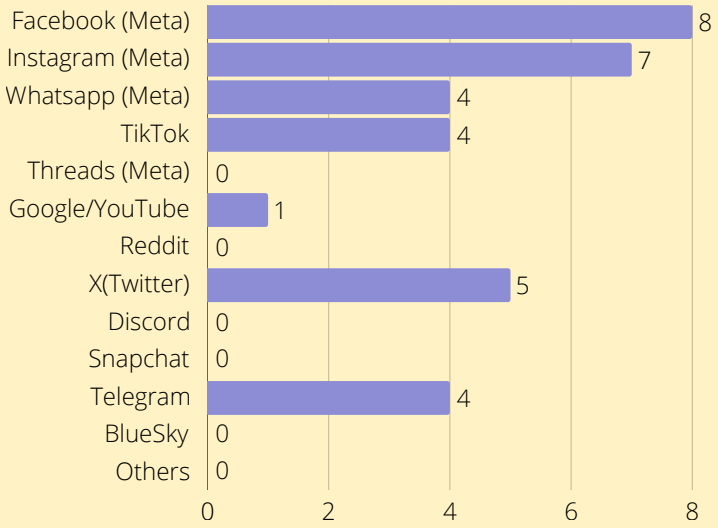


- Non-consensual intimate image abuse
- Non-consensual intimate image abuse for minors
- Hacking of vulnerable occupations (human rights defenders, activists, journalists, political activists)
- Hacking for marginalised groups (LGBTQI+ community, religious minorities, ethnic minorities)
- Hate speech Campaign
- Disinformation Campaign
- Incident/event based disinformation (eg during the elections, political unrest and protests)
- Fake profile sexualising minors
- Dangerous Organisations and individuals targeting marginalised groups and vulnerable occupations
- Dangerous Organizations and Individuals related content
- Fake profiles sexualising adults
- Fake profiles impersonating minors
- Other

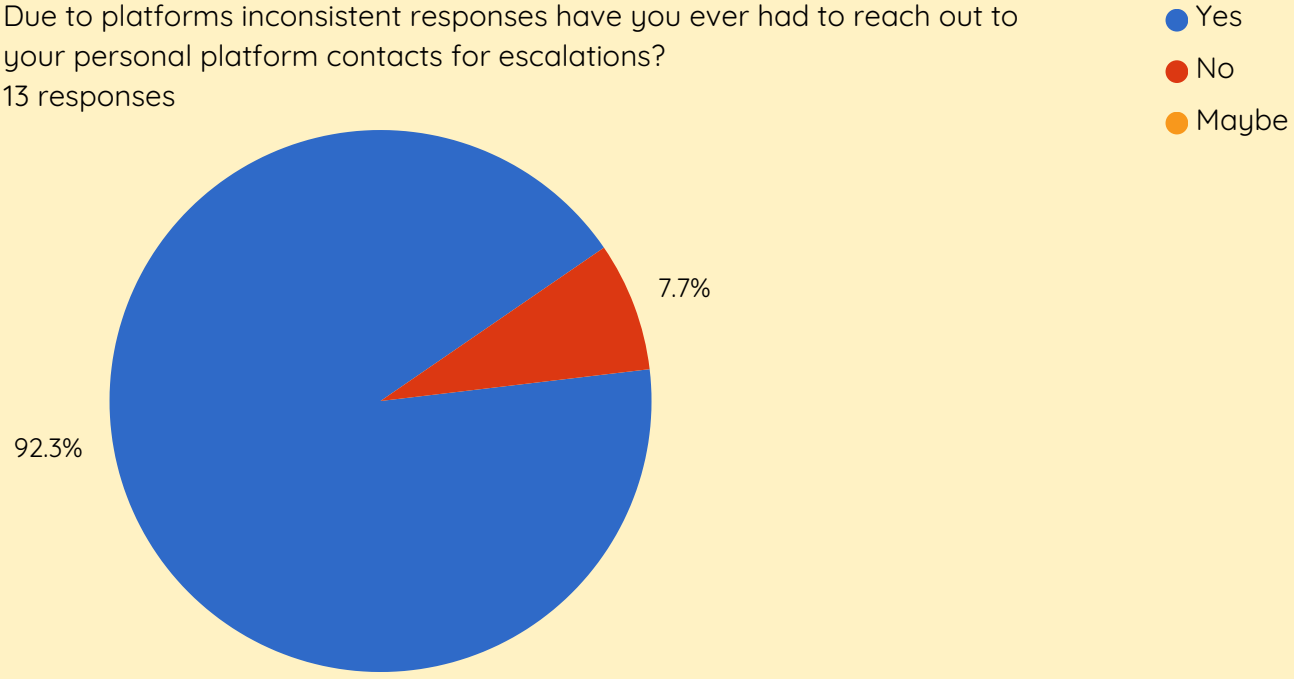
However, in our interviews, respondents claimed that images of victims in bras or bikinis, or even hijabs, are often not treated as NCII and therefore face slower removal or even rejection.

Inconsistent enforcement by platforms was also reported as an issue across helplines. According to the results of our survey, Meta platforms top the list in enforcement errors. Facebook sees the most enforcement errors, with Instagram and WhatsApp coming second and third. In fact, 92.3% of our survey respondents had to reach out to personal platform contacts for escalations due to inconsistent responses.

In which platforms do you see the most enforcement errors?
13 responses



Due to platforms inconsistent responses have you ever had to reach out to your personal platform contacts for escalations?
13 responses

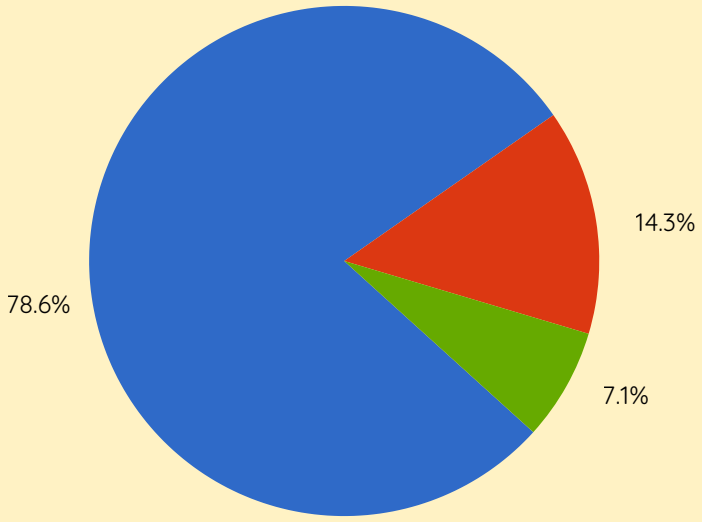


According to our interviewee for Tech4Peace, reports from collective organisations are taken more seriously than those from local ones. One example they shared is that Meta accepts requests for restoring accounts if a collective organisation like Access now files the report, but when a local organization such as Tech4Peace reports the same, Meta claims that they can't take action. Another major issue highlighted was how Meta is much more responsive to government takedown requests than requests by Trusted Partners, this being the case even when the content in question does not violate community guidelines. An example of inconsistent enforcement pertaining to government requests was seen in the case of Ithar Saeed, an Iraqi activist who was targeted for allegations of being a US spy. The Iraqi government filed a removal request with Meta for fake accounts impersonating and targeting her. Although Trusted Partners had already escalated the case, their efforts were limited to only the content being removed, and the account was only taken down after the government intervened. According to our respondent, Meta then recorded the removal as a government takedown in its transparency report.[35] Helplines sometimes have to use workarounds to overcome the problem of inconsistent enforcement. According to our respondent from SAFEnet Indonesia, platforms sometimes reject their reports, so they have to appeal cases again from a different angle to increase chances of being accepted.

Inaccurate rejections of reports and escalations often also occur as the result of language and cultural blind spots. SAFEnet Indonesia reports that they often have to appeal reports which have been rejected because platforms lack local context. They have to make platforms understand that even if the content does not violate guidelines, such as cases where leaked pictures are not necessarily nudes, "it is considered not polite or as an intimate image in our country". Similarly, since the local language in Indonesia is Sundanese, content moderation is carried out in that language, overlooking harmful content in other local languages. SAFEnet reported holding meetings with platforms to bring about policy action that is more inclusive of local languages and culture. A desire to close the gap between local context and broad-based policy enforcement was expressed by survey respondents, as 78.6% of the helplines believed platform policies should vary by region. However, 38.5% of respondents answered "No" to the question of whether platforms engage effectively with them on platform policies and implementation (e.g. holding consultations), with another 38.5% answering "Maybe", so confidence in the adaptation of platform policies to regional cultural contexts seems unlikely in the near future.

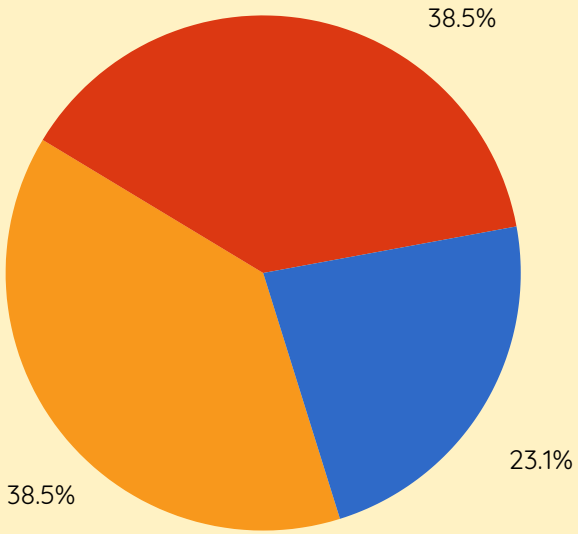
[35] https://transparency.meta.com/reports/content-restrictions/case-studies/

In your opinion, is global application of community guidlines and platform policies effective or should they vary by region
14 responses



- Platform Policies are effective
- Platform Policies are ineffective
- Platform Policies should vary by region
- Other

Do platforms engage with you effectively on platform policies and implementation (e.g. hold consultations)?
13 responses



- Yes
- No
- Maybe

Reporting a traumatic case to platforms can itself be a retraumatising process for victims. The Ecuadorian organization Taller de Comunicación Mujer (TCM) “Navegando Libres” reported escalating the case of a trans person with Meta, whose Facebook account had been wrongfully taken down. Meta accidentally released the wrong account, i.e. an older account of the victim’s, before they had transitioned, “so it was an outing by Facebook, basically”.

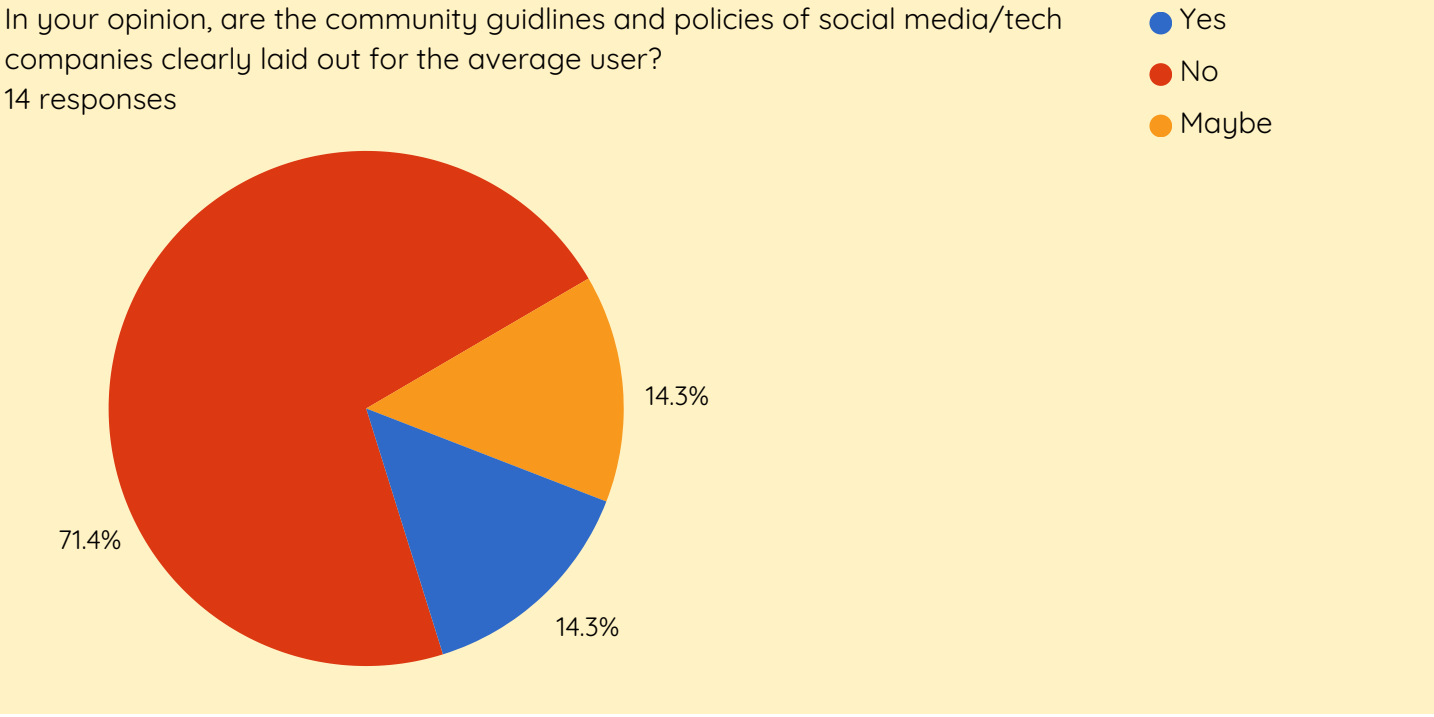
The numerous ways in which platforms have failed victims and Trusted Partners/Flaggers further exacerbates digital harms perpetuated against women and gender minorities, by not only taking up victims’ time and emotional labour, but also causing inefficiencies and frustration to helplines catering to escalation requests pertaining to TFGBV cases.

Feminist Helplines’ Community-Led Responses

Feminist Helplines across the world deploy contextually-specific strategies when escalating digital abuses to private social media platforms. Context variation is a result of differences in cultural sensitivities and digital literacy, as well as rather subtle factors like the size of a country, its market potential, and cross-border trends.

Our findings in this section outline the following pertinent themes that shape the ways in which Feminist Helplines organize, strategize, and innovate their support systems against the backdrop of a multitude of internal and external challenges. First, we find that Feminist Helplines operate entirely on principles of survivor anonymity and empathy, and these principles are not always defined in line with platform regulations. Second, Helplines situated within the Global South invariably participate in community and capacity-building exercises, targeting improvements in digital literacy and safety. Third, Helplines specifically from countries in Latin America and West Africa are actively involved in cross-border collaborative efforts as a way to pool their resources, engage in common advocacy programs, and also to improve their leverage vis-à-vis private platforms.

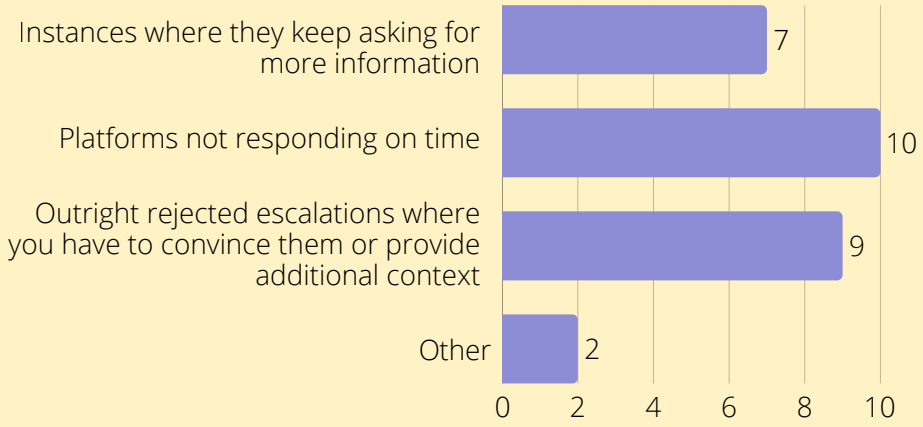
While anonymity and empathy constitute the rights-based underpinnings of all Helplines, such concepts are mediated differently depending on the cultural and political contexts. As platform rules are shaped primarily by individuals and departments located in the Global North, these specific nuances tend to be ignored in community guidelines. This discrepancy was indicated in our statistics, where 86% of participating Helplines have said that community guidelines tend to be quite inaccessible to users.



In the MENA region, our interviewee who works with Tech4Peace (T4P) spoke about the sensitive nature of documenting and advertising cases related to LGBT and Queer communities. Considering the conservative cultural landscape of countries like Iraq, where our interviewee was situated, such cases were invariably kept anonymous and not documented in the organization's publicly available research. T4P also relies on processing ID cards of complainants and confirming their identity using short videos, and this need arises out of the widespread phenomena of impersonation in Iraq which has led to the Helpline being misused by perpetrators.

At times, the principle of anonymity also clashes with platform policies, as evidenced by an associate from SAFENet Indonesia. They remarked on incidents where cases of sextortion and AI deepfakes were not being taken seriously by platforms in the absence of complete image evidence. Repeated appeals would have to be made to ensure takedowns, but the provision of evidence had to be balanced with cultural sensibilities and victim privacy. SAFEnet's story reflects wider communication issues our Helplines face when escalating requests to platforms. 7 Helplines expressed difficulties that stem from platforms requesting additional information that complainants were reluctant to share due to individual or cultural sensibilities. Therefore, while Helplines actively try to balance their principled rights-oriented commitments with effective case resolution, this proves exceptionally difficult when confronted with rigid platform requirements.

What difficulties do you face when communicating with platforms?
13 responses



Capacity-building is another key function performed by Feminist Helplines and their affiliated organizations. Our respondents operate in countries where rapid digitization has occurred only very recently. The expansion of digital spaces has also been quite uneven, with our T4P interviewee remarking on difficulties faced by women in accessing social media applications like Instagram. In addition to advocacy and educational campaigns, the T4P helpline has also made it much easier for complainants to register their complaints through direct messages on Instagram. In Ghana, a representative from Cyberclinic remarked on how the Helpline, in addition to escalating digital abuses, provides survivor-centric care to its victims. This includes cyber trauma counseling sessions with victims, training sessions with law enforcement agencies and educational institutions, and advocacy programs around TFGBV, financial fraud, and the ethical use of digital technologies. These exercises are often conducted in a cross-border space, with Cyberclinic regularly conducting sessions on digital rights with fraternal organizations in Nigeria and other African countries.

Finally, our findings indicate unique patterns of cross-border collaboration that extend beyond mutual capacity-building and conferences. Our conversations with a Helpline Associate based in the Ecuadorian helpline Navagando Libres were particularly illuminating. They identified the importance of a country's market size in determining its access to platform escalation channels. Ecuador, being a sparsely populated country in Latin America, has faced a myriad of challenges in escalating takedown requests or establishing contacts with platform associates. The issue of delayed communication by platforms was identified as pertinent by 10 of the 14 Helplines included in this study (as indicated in the bar chart above). In Ecuador, however, communication was – for the most part – non-existent. Meta, in particular, has not been responsive in the slightest even while digital abuses ran rampant against Ecuadorian accounts on Facebook. Navegando Libre's capacity to resolve cases was heavily curtailed as a result of platform negligence, compelling them to seek assistance from ally organizations located in Brazil. For many years, cross-border collaboration between organizations in

Brazil and Ecuador ensured that sensitive cases were being amplified to relevant platform authorities, evidencing the ways in which Feminist Helplines have adapted their strategy to overcome external obstacles.

Alternative Platform Governance

Participating helplines did not have many insights to share about their experiences with alternative platforms. No helpline reported receiving any major complaints or cases that concern decentralized applications such as Mastodon or Bluesky. As a matter of fact, some remarked on there being very little awareness of such platforms in the wider community within which they operate. This is understandable, considering that these applications are still quite new and their user base is very small compared to established platforms. For example, there are around 40 million Bluesky users, out of which only 3.5 million use the application daily. Over half of these users are located in the Global North, with the United States, U.K, and Germany constituting over 60% of Bluesky's user base[36]. These statistics pale in comparison to Blueksky's peer competitor X, which boasts over 500 million active users[37] dispersed across the globe.

It is important to note here that almost none of our interviewees reported having positive experiences in escalating complaints to X, especially after 2020. SafeNET Indonesia and Navegando Libres Ecuador stated that post-2020, escalation channels with the platform were hollowed out, severely limiting the ability of helplines to provide relief to abuse victims on this platform. While de-centralized platforms like Bluesky afford greater space for moderation, X has moved in the opposite direction by prioritizing efficiency and economic returns. Still, the effects of their contrasting moderation policies on attracting users has not been sufficient.

With this discrepancy in mind, our findings suggest that decentralized applications do not yet operate on the same global scale as popular social media platforms. Therefore, helpline engagement with or training in the alternative technical systems offered by these technologies is quite negligible. Considering the open-source nature of content moderation in BlueSky and the "de-federation" options provided by Mastodon servers, it may be possible that self-moderation within these platforms precludes the need for external complaint channel services offered by feminist helplines. However, this hypothesis cannot currently be tested considering the low user-base for these platforms in the wider Global South.

[36] https://backlinko.com/bluesky-statistics
[37] https://worldpopulationreview.com/country-rankings/twitter-users-by-country

Access to traditional and alternative social media platforms must be understood within the broader context of the former's seemingly impenetrable market power and scale. According to Navegando Libres Ecuador, the growing monopolization of mainstream media platforms coincides with major political shifts towards far-right and fascist forms of political discourse and government policies. Their parent companies operate on a global scale in partnerships with increasingly authoritarian governments. These trends indirectly underscore the prohibitively high barriers to entry for alternative social media platforms. How they can operate at the same scale as established companies without compromising their decentralized structures of governance remains subject to debate. For the purposes of this research report, however, this does inform a significant barrier to their feasibility in comparison to mainstream digital spaces.

Toward Feminist Platform Accountability

Having established the experiences of feminist helplines across the Global South and exploring how platforms have failed to address growing contextual concerns of first responders to online harms, it is important to go into detail about what should platform accountability look like as per the knowledge of these helplines who talk not only from a unique geographical point of view but also from a feminist perspective.

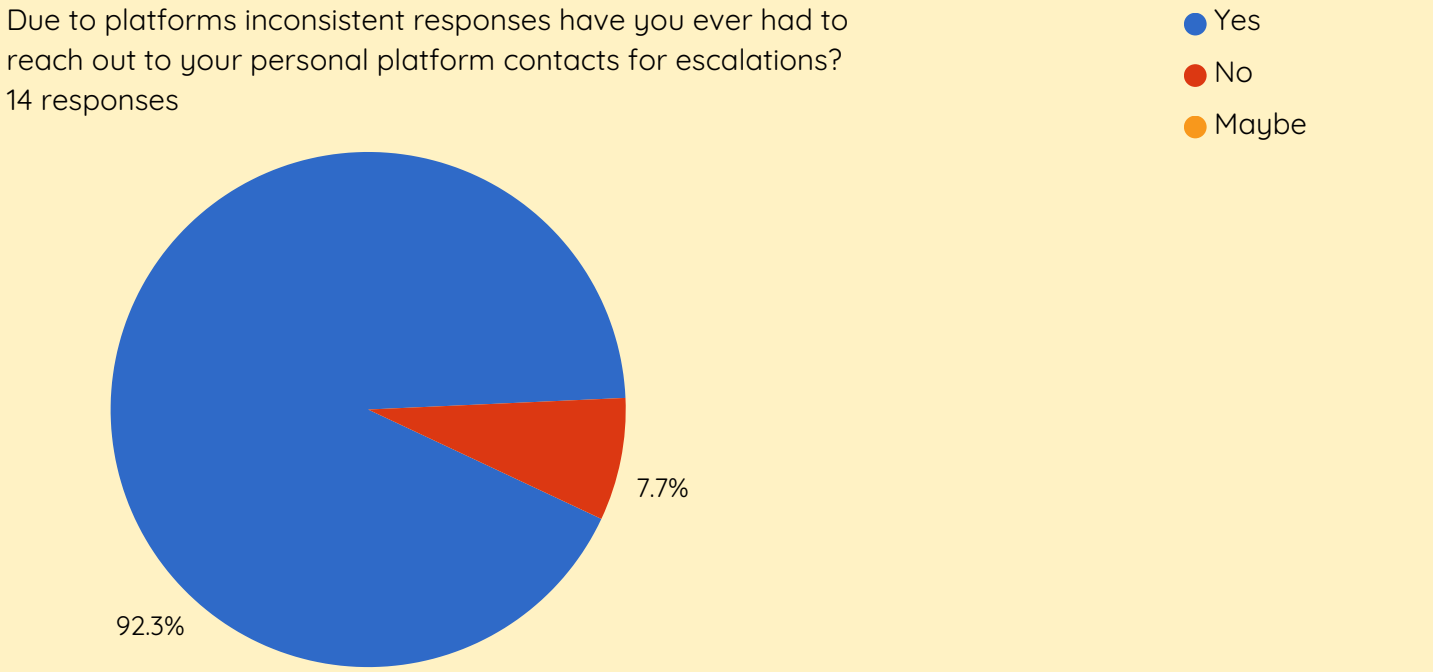
When asked about what accountability measures would make their work more effective and easier, all helplines noted the growing necessity for transparency; transparency in escalation methods, in prioritization mechanisms and what categorizations they use when deciding what is or isn't harmful content. When talking to Tech4Peace, operating in the greater MENA and North African region, helpline associates were surprised and confused about why requests for the removal of dangerous content originating from Iraq were left completely unheard when reported by them, but removed when asked by more "higher-up" institutions, such as the Ithar Saeed case previously mentioned.

While information about where the requests for content takedown came from are reflected in Meta's transparency reports[38] why they choose to listen to state institutions over identical case appeals made by Trusted Partners is unclear. According to our participating helplines, transparency in what procedures come into play when platforms decide which requests to approve and which to ignore – even when requests are exactly the same – will help helplines better understand how cases should be framed and passed on to platforms to ensure requests are heard and resolved without the need for third-party interference.

This is further corroborated by our survey results, where 92% of participating feminist helplines have said that due to inconsistent platform responses, they have had to reach out to platform personnel in a private capacity to get immediate help. Having to resort

[38] https://transparency.meta.com/reports/content-restrictions/case-studies/

to indirect connections and personal contacts via email, or more commonly WhatsApp, as a helpline associate from Cyberclinic in Ghana has said, not only becomes time consuming for these helplines who have to deal with multiple cases on a single day, but also takes away any official channels of liability, accountability and necessary remedy.



A turn towards feminist platform accountability requires social media platforms and Big Tech to openly share their guiding principles, as well as being open to changing these principles on the basis of recommendations given by experienced associates in the field. One such change in the principles that feminist helplines are asking for is the emerging need to concretely address emerging threats to content online. During our participant interviews, most feminist helplines, safe for Ghana, are finding it especially challenging to filter through incoming video content and identify it as AI-generated. Considering the sheer volume of information and content sent to these helplines alongside the fact that AI technology is becoming staggeringly advanced, feminist helplines can only be scratching the surface of the harmful AI content available to users on platforms today. Without further help from platforms, either in terms of verification or watermarking, feminist helplines will continue to find themselves stretched thin and much of AI-generated content, that either acts as TFGBV or as disinformation, will be left unchecked.

Advocacy Agenda

Regulation, in isolation, is not sufficient to make a big enough impact in the world of social media where content, trends, technology and policies are constantly changing and adapting into newer, better ways of interacting with the digital world. According to our participating feminist helplines, establishing frameworks and building relationships as trusted partners with platforms needs to go hand in hand with advocating for better policies alongside raising awareness around the localized, unique issues of the Global South.

Thus, equally important to feminist helplines operating in the digital sphere today is the need for platforms to incorporate and consider cultural context when addressing concerns from the Global South. When talking about the challenges faced when interacting with platforms such as X, Meta and TikTok, all feminist helplines have said that their requests to remove content, suspend accounts or restrict online material, are largely denied because content moderators at these platforms do not think they violate any platform guidelines. Tech4Peace, Rati, SAFEnet, Navegando Libres, and CyberClinic all expressed concerns on how content that is culturally considered to be harmful in their respective regions, including the Middle East, South America, South Asia and Southeast Asia, and reported to platforms for removal is more often than not rejected because it is not nude enough, not intimate enough, or simply not dangerous. Context setting, according to SAFEnet, can be the deciding factor between whether a woman feels safe enough to be on the internet.

Platforms can only be feminist in practice and accountable for their actions if they take active steps to incorporate cultural sensitivities prominent in every region where they operate – something that feminist helplines are constantly advocating for through their work and research. Through holding active dialogues with Trusted Partners (i.e. feminist helplines) and incorporating their unique perspectives on what does or doesn't make content from the region high-risk to the integrity of their users, platforms can create a much more personalized user experience, one that centers on their safety and societal norms.

Similarly, regulation through platforms alone has never been sufficient in and of itself – it has always needed to be paired with collaborations, coalitions and conversations between CSOs themselves as well as other institutions working in the digital arena. As most of our participating feminist helplines have been operating for over 5 years, and with such vast experience in an ever changing environment, they have come to realize the importance of strength in numbers and collaboration. As mentioned previously, a recurring theme we encountered was the collaboration between helplines in the region when trying to flag and escalate content on a larger scale. SAFEnet from Indonesia had also mentioned working alongside helplines in Malaysia, Thailand and the Philippines

specifically when reporting hate speech in the region (TFGBV was mostly handled locally given country-specific offenses that rarely crossed borders).

In addition to this alliance between helplines regionally (such as Ecuador and Brazil; Indonesia and Malaysia), we also noted an interesting reliance and cooperational relationship between helplines and other first-responder institutions. In India for example, helpline associates from RATI Foundation mentioned a government initiative referred to as the Grievance Appellate Committee[39] (GAC) - an online dispute resolution mechanism that appoints grievance officers for every platform operating in India with upwards of 500K subscribers and/or users. Rati found that reported cases that were rejected at first and then appealed through the GAC were more likely to go in the favour of the victim than if the helpline themselves were to submit an appeal - an avenue that the helpline finds beneficial. But, while favourable, the process does have its shortfalls with little to no transparency, lack of language inclusivity (forms can only be found in English) and a long turnover period (2-3 weeks).

International collaborations also hold significance among feminist helplines when taking into account the unfortunate reality that voices heard from the West are given more importance than those coming from the Global South. This is true in particular for Iraqi helpline Tech4Peace, who have said that they often work with Access Now, an international non-profit that works on digital rights, to take down or, more commonly, unban content.

The idiosyncratic nature of feminist helplines around the world, as mapped out and analysed in this report, highlights the resilience found in the work done by feminist helplines, the impact they have not only on the lives of their clients but also on how Big Tech frames its regulations and more importantly, how in spite of this, platforms are still failing their users and leaving the onus for getting remedy and recourse with feminist helplines themselves. Given these lived experiences, our findings, and the efforts taken by feminist helplines on a global stage to fight the rise of TFGBV amongst other digital harms, there is still much work to do by platforms and institutions alike.

[39] https://gac.gov.in/

Recommendations

Based on the findings from the survey and interviews, DRF developed a set of recommendations grounded in the insights shared by participating helplines and the patterns identified throughout the research.

Platforms:

- Several feminist helplines pointed out the absence of proper escalation channels with several major platforms. Meta had the highest partnership rate, with 9 helplines out of 14 stating that they were included in the Trusted Partner program. Reddit, Snapchat, and Telegram only had official escalation channels with one helpline each, while Discord was not officially connected with any helpline. With the rising significance and potential harmful effects of disinformation, hate speech, NCII, and child sexual abuse material (CSAM), escalation channels and trust and safety partnerships with civil society and feminist helplines are increasingly essential. Tech and social media platforms need to invest resources in establishing healthy channels in order to counter harmful online content and behavior, and diversify their partnerships over regions and target population groups.

- Improve response time and response quality, which includes context and updates on how the case is progressing. Escalation training can also be provided so helplines know exactly how to create a case to cater to platforms' needs.

- Evaluate type and number of cases from each country and region, and develop strong relationships based on that spread. Invest in supporting those helplines so they can act as strong mediators between the general population and the platforms.

- Establish and eventually expand Trusted Partner programs, and hold regular consultations with them over policy problems and regional concerns.

- Dis- and misinformation need to be countered with fact-checking programs that provide value and become almost essential in the age of unreliable information, particularly when disinformation powered by AI is rising in an increasingly polarized and charged online environment.

- AI generated content and synthetic media labels need to be highlighted, and in a way that is easily understandable for the common person.

- Most platforms tend to focus primarily on posts and captions, whereas the comments sections also generate significant hate speech, abuse, and overall harm. Posts, images, and videos in themselves might not be violating, but when the comments underneath encourage harm, they should be actioned by platforms, particularly when there are calls to violence and abuse.

- New and emerging social media and tech platforms should take into account the findings and experiences of feminist digital security helplines, and incorporate policies that best serve users, particularly women, minors, and other vulnerable and marginalized groups. Policies should also be implemented in a way that takes into account regional sensitivities and findings from this research.

State:

- Use the wealth of data and information that is collected through the helplines, and invest in them so they are able to research and collect more data, which will, in turn, inform state level policies to both counter TFGBV and demand accountability from platforms.

- Collaborate with civil society, and support their work because the onus of fighting TFGBV in all its aspects should not fall solely or primarily on civil society. The work is draining and requires a lot of emotional labor, and these civil society actors and helplines should receive the necessary support.

- Lead targeted digital literacy and awareness campaigns for survivors and the general public, with a focus on digital rights, online safety, and reporting cyber incidents

- Enhance and bolster law enforcement that deals with cyber crimes. Establish special wings for TFGBV against women and children within those law enforcement authorities, which provide holistic support.

- Cyber crime laws should center around the protection of women and children and evolving threats posed by AI.

- Data protection laws and cyber crime investigation procedures should prioritize the confidentiality and protection of women's sensitive data associated with complaints and legal proceedings..

- Introduce mandatory gender-sensitisation and survivor-centred training programmes for law enforcement officers, judges, and prosecutors to prevent secondary victimisation.

Civil Society:

- Maintain a proper database of their interactions with social media and tech platforms, particularly how many consultations they attend, how many cases they escalate, the response time for each case, and the time in which cases are resolved.

- Even when something is considered to be 'out of scope' for escalations with platforms, helplines should still consider escalating cases (after careful review of all other options), so that trends and patterns can be recognized by platforms.

- Develop and maintain regional and global collaborations with other helplines, civil society, and other stakeholders, to enhance knowledge sharing, build stronger communities, and exchange advocacy best practices.

- At a time when helplines and digital rights organizations are losing funding and have been pushed further down the priority list, it is important to highlight and emphasize the importance of the work that helplines have been doing, and the essential services they provide in their communities.

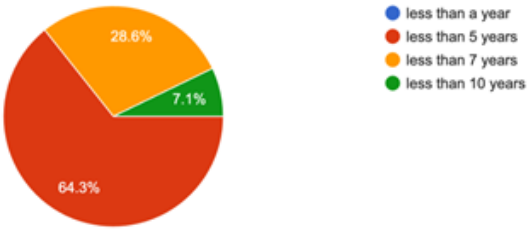
Appendix A

Would you like to be anonymous for this research report?

14 responses

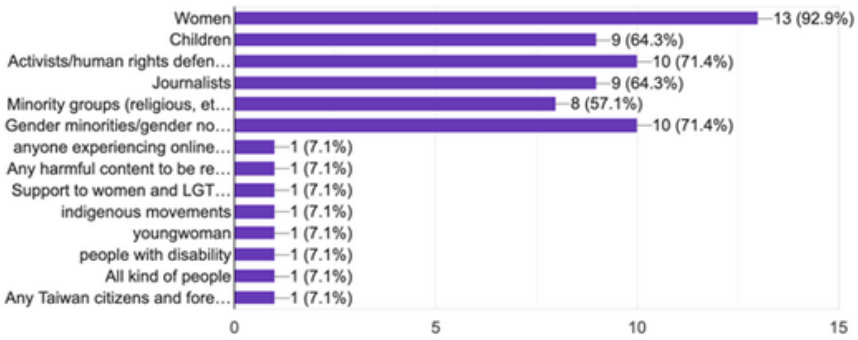
How long has your Helpline/service been operating?

14 responses



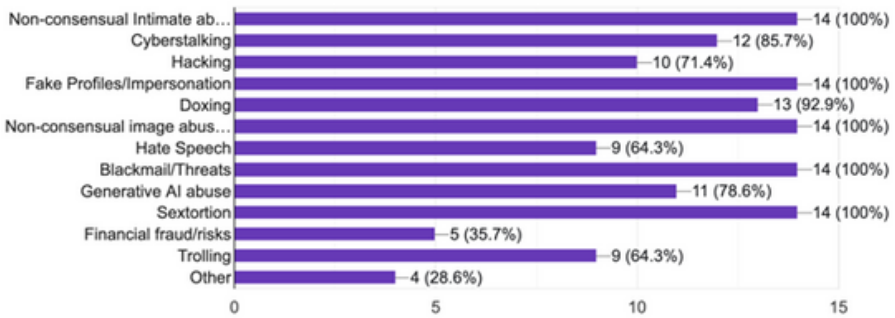
Who does your Helpline/services cater to?

14 responses



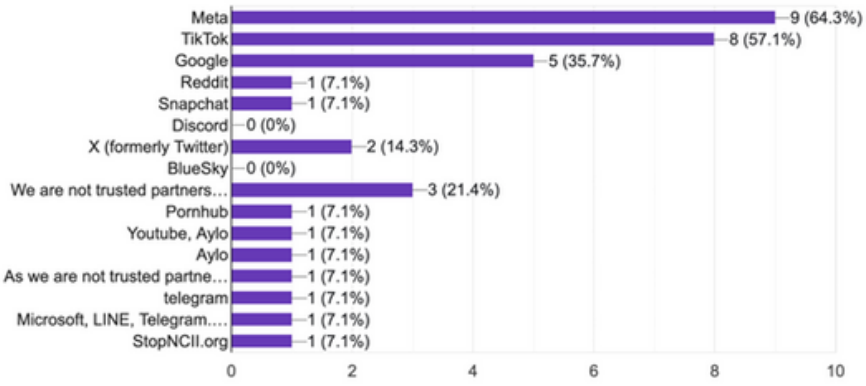
What problem areas does your Helpline assist with? If technology-facilitated gender based violence, then please select the types below

14 responses



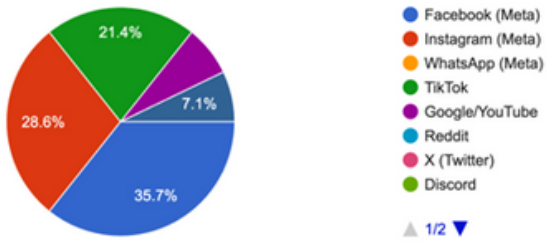
Is your Helpline/organization a Trusted Partner with any social media or tech companies?

14 responses



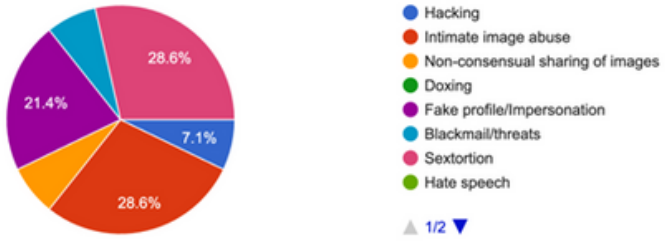
Which social media/tech company do you make the most escalations to?

14 responses

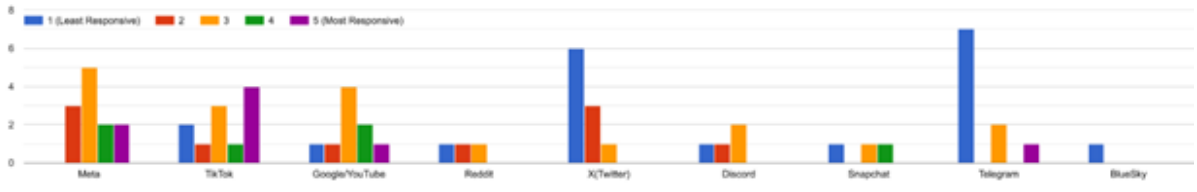


What type of complaint is most often escalated?

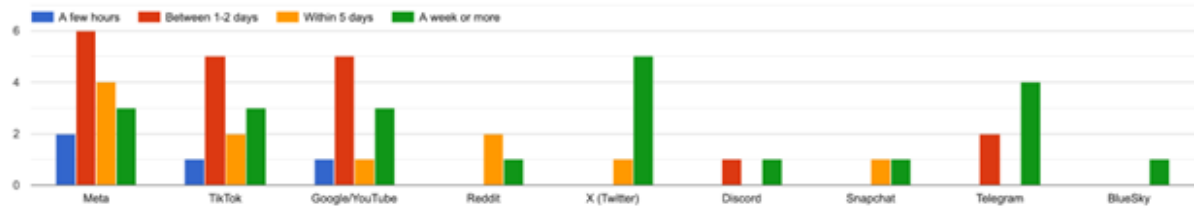
14 responses



Which social media/tech company has the best response rate? (Response rate refers to the time duration in which the company responds to an escalation, with 1 being least responsive and 5 being the most responsive)



For the question above, how long do they generally take to respond (either positively or negative)

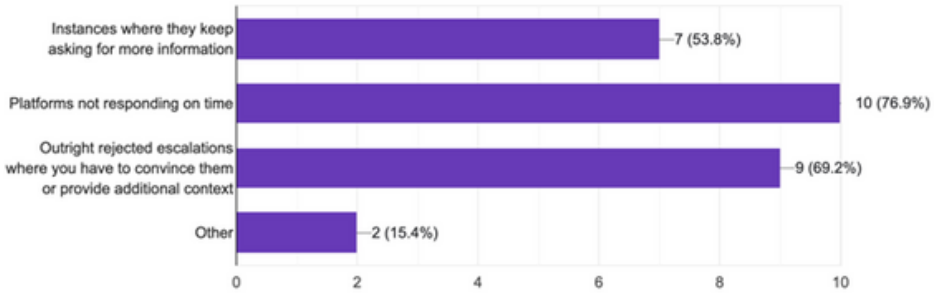


Which type of complaint receives the greatest attention and resolution rate by social media/tech companies?

14 responses

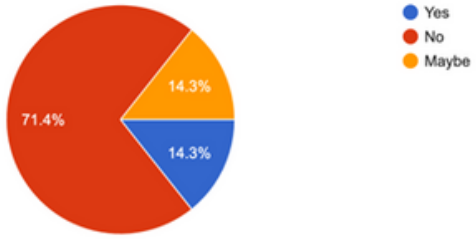
What difficulties do you face when communicating with platforms?

13 responses



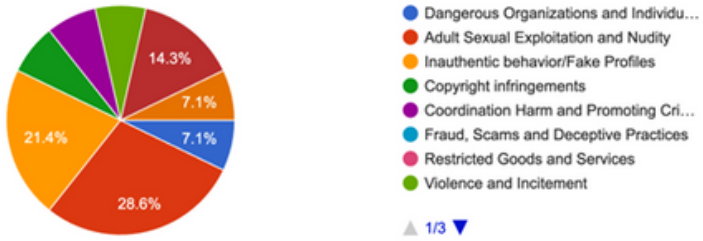
In your opinion, are the community guidelines and policies of social media/tech companies clearly laid out for the average user?

14 responses



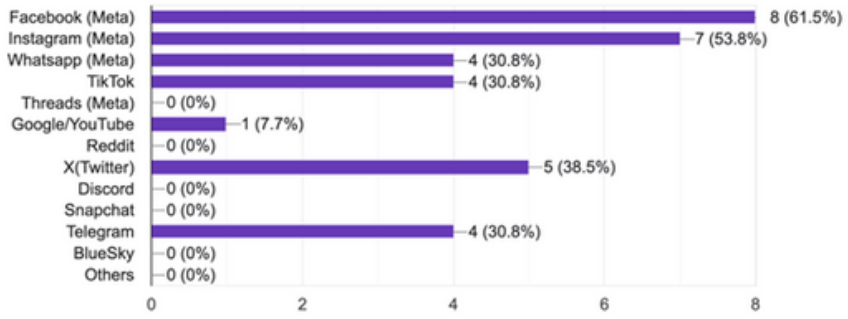
Under which community guidelines have you observed the most enforcement errors?

14 responses



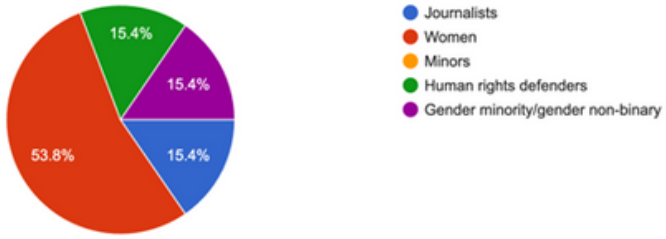
In which platforms do you see the most enforcement errors?

13 responses



Which vulnerable group do the enforcement errors most affect?

13 responses

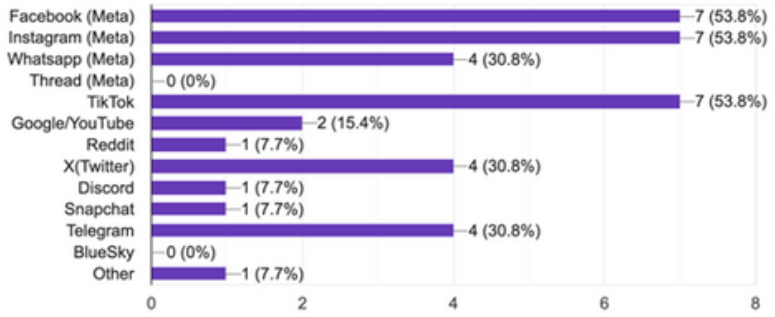


Have you ever come across an instance where NCII (non-consensual intimate image) has been allowed on a platform, or was not detected automatically?

13 responses

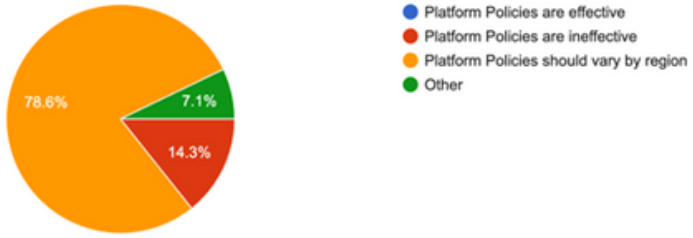
Please share which platform(s) where this has occured, if you are comfortable naming them:

13 responses



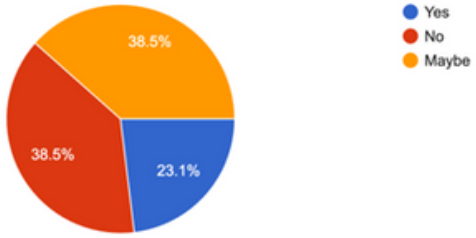
In your opinion, is global application of community guidelines and platform policies effective or should they vary by region?

14 responses



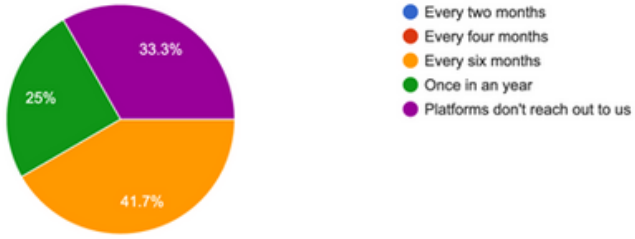
Do platforms engage with you effectively on platform policies and implementation (e.g. hold consultations)?

13 responses



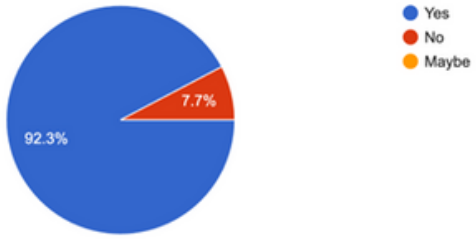
If they do engage, how regularly does that happen?

12 responses



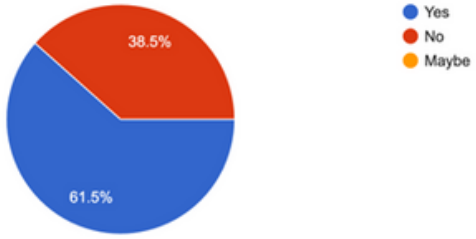
Due to platforms inconsistent responses have you ever had to reach out to your personal platform contacts for escalations?

13 responses



Have you ever reported local campaigns or trends to platforms?

13 responses



DigitalRightsFoundation

"KNOW YOUR RIGHTS"

@DigitalRightsFoundation

@digitalrightsfoundation

@digitalrightsfoundation

@digitalrightsfoundation

Digital Rights Foundation

@digitalrightspk.bsky.social

@DigitalRightsPK

@DigitalRightsPK