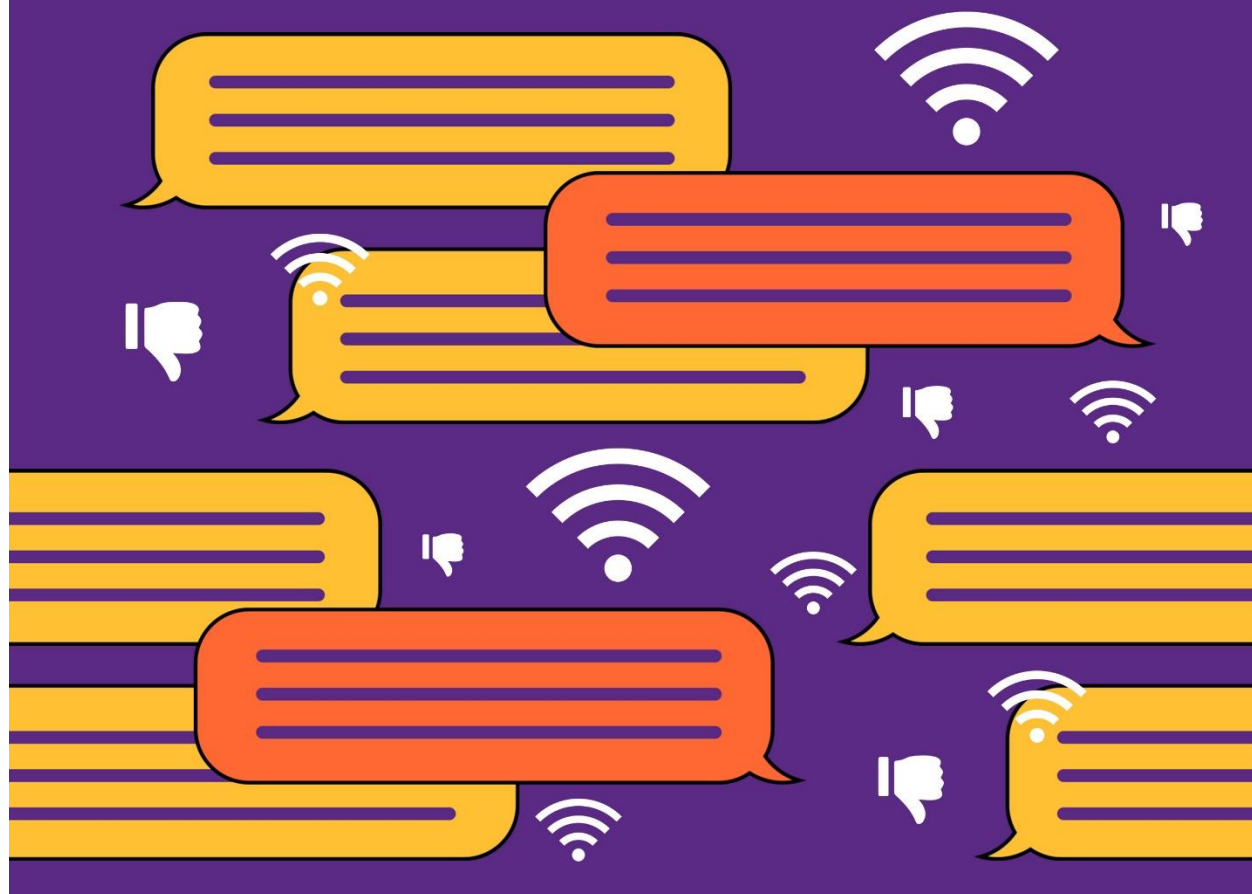# Digital Rights Foundation
"KNOW YOUR RIGHTS"

POLICY BRIEF

PLATFORM ACCOUNTABILITY ON ONLINE HARMFUL CONTENT AFFECTING MARGINALIZED GROUPS

About DRF

DRF's Cyber Harassment Helpline is Pakistan's first dedicated helpline dealing with complaints of cyber harassment and online violence. The Helpline is a gender-sensitive confidential service with qualified support officers who assist callers in complaints regarding digital security, legal assistance and psycho-social well-being. The Helpline additionally also carries out escalations to social media platforms to protect marginalized groups in the country particularly women, minors, minorities and vulnerable professions affected by online harassment. The Helpline started its operations in 2016 with operations running from Monday to Friday, 9 am to 5 pm on the toll free number 0800-39393. The helpline is also additionally available over email (helpdesk@digitalrightsfoundation.pk) and social media platforms like X (formerly Twitter), Facebook and Instagram to the general public. The Helpline has received over 14,000 cases with a wide range of these cases being from women in its six years of operations.

Introduction

Pakistan had a total of 87.35 million internet users as of 2023 with 71.10 million social media users in January 2023 [1]. Despite these staggering numbers of internet users in the country, according to the Freedom of the Net Report 2023 by Freedom House Pakistan ranks as 'not free' with internet freedoms still limited and curbed for its citizens.[2] Pakistan's internet connectivity and freedoms have been limited and skewed due to obstacles around access to the internet, limitations on online content and violations against users' privacy and data protection.

Since the passing of the Prevention of Electronic Crimes Act (PECA) in 2016 and the Removal Blocking of Unlawful Online Content Rules (RBUOC) in 2021, Pakistan's regulatory bodies hold immense authority over user data and content removal in the country. In 2022 alone, according to Meta's transparency report, the platform received 1,144 total requests by governments for user data.[3] The platform also restricted access to 6,428 items reported by the Pakistan Telecommunication Authority (PTA) for allegedly violating local laws, including 3,674 items for blasphemy and anti-religious sentiment, 2,010 items related to obscenity, 589 items related to sectarian enmity and hate speech,

---

[1]https://datareportal.com/reports/digital-2023-pakistan#:~:text=Internet%20use%20in%20Pakistan%20in%202023&text=Pakistan's%20internet%20penetration%20rate%20stood,percent)%20between%202022%20and%202023.

[2] https://freedomhouse.org/country/pakistan/freedom-net/2023

[3] https://transparency.fb.com/reports/government-data-requests/country/PK/

119 items for fraud and regulated goods violations, 22 items for terrorism content, and 14 items related to anti-judiciary content.[4] X (then Twitter)'s July - December 2021 transparency report shares that the platform received 17 legal government information requests for the platform.[5] The platform's change in governance structure did not allow for country-wide transparency reports in 2022 and 2023.[6] However, these numbers might have doubled with the government finding more means to regulate platforms.

The increased use of social media and the over regulation of online spaces has led to varied experiences for marginalized groups. The rise in their exposure to online harmful content, including but not limited to misinformation, disinformation, hate speech and online gender-based violence against marginalized groups, particularly women and transgender communities, is the reason why many vulnerable communities self-censor online and have limited representation in the space.

This policy brief sheds light on the online experiences of religious minorities, human rights defenders (HRDs), transgender activists and journalists in Pakistan. The policy brief aims to capture the complaints received by these groups on DRF's Cyber Harassment Helpline in the year 2023 from January to August. It also provides recommendations for platforms, governments, civil society and media to make the internet an accessible and safe space for all.

Big Tech and marginalized communities in Pakistan

Pakistan's religious minorities, human rights defenders, transgender activists and journalists are repeatedly intimidated, harassed and bullied online and for the purposes of this Policy Brief are classified as vulnerable groups in the country due to the nature of their work or gender or religious identity.

According to Reporters Without Borders's (RSF) World Press Freedom Index, Pakistan ranks 150 out of 180 countries on press freedom.[7] Additionally, the United Nations (UN) Human Rights Council has expressed concerns over reported incidents of reprisals against human rights defenders and civil society activists by state and non-state actors for cooperation with the UN.[8] The Pew Forum on Religion & Public Life / Global Restrictions on Religion Report highlights the problems that individuals from minority

---

[4] https://transparency.fb.com/reports/content-restrictions/country/PK/

[5] https://transparency.twitter.com/en/reports/countries/pk.html

[6] https://blog.twitter.com/en_us/topics/company/2023/an-update-on-twitter-transparency-reporting

[7] https://rsf.org/en/country/pakistan

[8] https://voicepk.net/2023/10/pakistan-under-scrutiny-for-reprisals-against-human-rights-defenders/

groups face in Pakistan. According to the report, individuals continue to face harassment or intimidation motivated by religious hatred or bias in the country.[9]

Human rights defenders in Pakistan are attacked and abused online with sophisticated digital attacks orchestrated against them in the country. According to a four-month investigation report by Amnesty International, HRDs experience a range of attacks online with their social media accounts getting hacked, and devices infected with spyware.[10] With the rise and spread of disinformation, digital HRDs are also subjected to attacks in the form of campaigns on platforms with some cases perpetuating offline violence.

Similarly, journalists are also subjected to disinformation campaigns and digital attacks on social media platforms. The Cyber Harassment Helpline has a dedicated helpline associate to deal with the complaints and escalations of journalists that are received at the Helpline. Additionally, in July 2022, DRF collaborated with Pakistan's National Commission of Human Rights (NCHR) to establish a Complaint Cell for Journalist Protection, particularly dedicated to resolving human rights issues that arise from the targeting of women journalists to ensure the freedom of the press. Journalists are very susceptible to attacks online as witnessed over the years and the Cyber Harassment Helpline policy brief in 2022 alone showed that the Helpline received a total of 75 complaints from journalists and media practitioners with 34 female, 40 male and 1 trans journalist.[11]

Religious minorities and gender minorities are also subjected to hate and abuse online, particularly by certain groups and these vulnerable groups have little or no institutional support for their protection in online and offline spaces. In a research by DRF on the online experiences of religious minorities,when the communities were asked whether they feel comfortable expressing their opinions online, 38.9% of respondents answered in the affirmative, while 34.5% said no. Additionally, 26.5% said they were unsure about whether they felt comfortable or not. Individuals were also asked about why they felt unsafe online with many sharing the lack of institutional support, abuse and harassment they had faced at the hands of strangers online and being "on the receiving end of a lynch mob" to name a few.[12]

---

[9] https://assets.pewresearch.org/wp-content/uploads/sites/11/2009/12/restrictions-resultsbycountry.pdf
[10]https://www.amnesty.org/en/latest/news/2018/05/pakistan-campaign-of-hacking-spyware-and-surveillance-targets-human-rights-defenders/
[11]https://digitalrightsfoundation.pk/wp-content/uploads/2023/07/Policy-Brief-Cyber-Harassment-Helpline-2022.pdf
[12] https://digitalrightsfoundation.pk/wp-content/uploads/2021/05/Religious-Minorities.pdf

Pakistan's transgender community has also been witnessing an unprecedented level of hate and orchestrated campaigns on social media platforms leading to offline violence. Since August 2022 the transgender community has been subjected to online attacks and hate speech when an activist and policy specialist was removed from a panel of speakers at a conference that was scheduled to take place at a school in Lahore.[13] Following that, the community has been attacked as a whole with a hate campaign that went so widespread that it led to sections being struck down in the Transgender Persons (Protection of Rights) Act 2018, and putting into question the identity of transgender individuals across the country.[14] In the wake of these attacks, the Cyber Harassment Helpline made efforts to escalate content to platforms, and hold meetings with them in order to push them to recognize the dangerous ways in which these platforms are being used to drive policy-level change in the country that discriminates and attacks the human rights of the transgender community.

Moreover, according to Minority Rights Group International's study on 'Hateful Places: Punjab, the epicenter of hate speech in Pakistan – A geospatial analysis of Twitter conversations', the major volume of hate speech in Pakistan occurs in three urban cities which are Karachi, Lahore and Rawalpindi.[15] Considering these are urban centers with high internet penetration and greater social media engagement, the situation is quite alarming. The study highlights how the major urban centers in the country are the epicenter of hate speech on social media platforms with little to no accountability serves as a chilling reminder of how urgent the need is to deal with this menace.

Journalists, particularly women, also face a range of attacks for reporting on key political issues online. In August 2020, 50 women journalists wrote a statement about the well-defined and coordinated campaign of harassment they had been facing on social media platforms which ranged from including abusive language to threats of violence.[16] The nature of these attacks have amplified to this day for journalists in the country with 23 Pakistani journalists targeted under the country's draconian cybercrime act in 2021[17]

---

[13]https://tribune.com.pk/story/2417727/worrying-rise-in-digital-hate-speech-against-transgender-community-report

[14]https://www.amnesty.org/en/latest/news/2023/05/pakistan-revocation-of-rights-of-transgender-and-gender-diverse-people-must-be-stopped/

[15]https://minorityrights.org/wp-content/uploads/2022/10/Geospatial_hate-speech-prefinal_finalMRG_Oct9.pdf

[16] https://docs.google.com/document/d/1DD8BQ53noKO6zHy-gysGnFjeKT4ride4uYtQsNNRYoc/edit

[17]https://www.aljazeera.com/news/2021/11/2/pakistan-journalists-targeted-cyber-crime-law-press-freedom

alone and more to follow in the coming years which establishes a pattern of legal action to silence dissent of the media in the country.

While platforms like X, Facebook and Instagram are used for self-expression and raising voices of the marginalized, they are also being used to attack and silence these groups. Platforms' community guidelines and transparency reports fail at times to capture the nuanced and complex issues of marginalized groups in the country and despite the many efforts of the cyber harassment Helpline in escalating content, there has been a rise in hate speech and disinformation on platforms leading to further complications for these particular vulnerable groups. Furthermore, social media platforms cater to a wide range of audience across the globe, however, their guidelines and content moderation policies do not necessarily align with the countries located in the Global South and the diverse experiences of the communities living in this region. Escalation of harmful content against these targeted groups needs to be handled in a time-sensitive manner and removed accordingly to avoid incitement to violence. However, there has been little transparency and accountability of platforms regarding policies and escalation mechanisms adopted for vulnerable communities in this region.

There has also been a recent shift in regulating big tech platforms across the globe. Approaches have been introduced by multi stakeholder bodies like UNESCO with the introduction of Guidelines for regulating digital platforms[18] for member states. Additionally there's been a recent shift in introducing functioning mechanisms for big tech platforms across the globe as well with the introduction of EU's Digital Services Act (DSA)[19], Artificial Intelligence (AI) Act[20] and UK's Online Safety Bill[21]. While these Acts ask for more user privacy from big tech platforms they also ask for big tech companies to work alongside governments which can be seen as fairly different from the situation in the Global South.

---

[18]https://www.unesco.org/sites/default/files/medias/fichiers/2023/04/draft2_guidelines_for_regulating_digital_platforms_en.pdf

[19] https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

[20]https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=Parliament's%20priority%20is%20to%20make,automation%2C%20to%20prevent%20harmful%20outcomes.

[21]https://www.gov.uk/government/news/britain-makes-internet-safer-as-online-safety-bill-finished-and-ready-to-become-law

Helpline Data and Analysis

The DRF Cyber Harassment Helpline became operational in December of 2016, in response to the advent of the country's main cybercrime law Prevention of Electronic Crimes Act (PECA) which was passed earlier in the same year.

The 3 main services offered by the Helpline include:

- Legal assistance: callers at the Helpline are offered assistance in terms of being advised on their legal rights, as they pertain to the particular issue being faced. DRF's Legal Department also extends support in terms of accompanying complainants to the local FIA Cybercrime division in order to assist them in pursuing a legal remedy.
- Digital security assistance: DRF has dedicated digital security trainers (DSTs) who can offer assistance in terms of device and data safety and how to best protect oneself in online spaces.
- Psycho-social support: Helpline Associates are trained in providing basic counseling to callers in distress. In certain situations, callers may be referred to an external vetted organization for mental health support.

How does the helpline process the complaints?

The Cyber Harassment Helpline receives complaints through the phone line, independent email address, mobile number and social media accounts on platforms such as Facebook, Instagram and X which are managed by the Helpline associates. DRF also has a number of networks with journalists and marginalized groups from where individuals reach out in case they need support.

The Helpline also works directly with Law Enforcement Agencies (LEAs) such as the Federal Investigation Agency (FIA) Cybercrime Wing. The FIA is the designated LEA named in Pakistan's cybercrime law, the Prevention of Electronics Crimes Act 2016 (PECA), and is thus the relevant authority for filing complaints under the law, especially in instances where technology-facilitated online harassment has been experienced by a citizen.

DRF, primarily because of the work undertaken by the Helpline, is a Trust and Safety partner with Meta. This means that we can offer support in terms of contextualizing, flagging and escalating cases that come due to concerns attached to the use of either Facebook or Instagram, the two most prominent platforms that fall within the Meta company head. This process is applied in cases where reporting directly to the social media company by the user has not resolved the issue. The process of contextualizing primarily involves applying a local lens on the specific instance to highlight why that particular harm to that particular person (based on their gender, ethnicity, occupation or religion) can amount to risk of them accruing grievous harm.

DRF is also an escalation partner with TikTok and a part of the YouTube Trusted Flagger program through which civil society organizations can effectively communicate violations of the platform's community standards.

The data that will be discussed in this Policy Brief is emanating from escalations made by the DRF Helpline from January to August 2023.

Data Breakdown:

The data set comprises 126 escalations made, where the youngest complainant is 13 and the oldest is 68 years of age.

### Gender Disaggregated Data (GDD):

This Policy Brief will employ a gender-disaggregated form of data breakdown to amplify the gendered nature of online harms and the visibility of vulnerable communities, women and gender minorities as sufferers of these crimes.

The Helpline collects non-personally identifiable information for research and follow-up purposes. This data set is reflective not only of the far smaller number of transgender individuals in Pakistan in comparison to the cisgender population but also demonstrates how heightened the risk and impact are for them due to the aggressive nature of the threats they face.

The primary portion of complainants are women and girls (65.87% or 83 out of a total of 126), then transgender women (17.46% or 22 out of a total of 126) where the escalations were for posts targeting members of the transgender community. Issues faced and escalated by men were 16.67% or 21 out of 126 complainants. The gender information pertaining to 13 of the escalations made was unavailable.

The numbers we see in 2023 (January to August) at the Helpline generally reflect that there is a gendered aspect towards online harms that individuals face. Our six-year review also reflects this where the Helpline had received 14,376 cases since December 2016 where 59% of those assisted since operations began have been of women.[22]
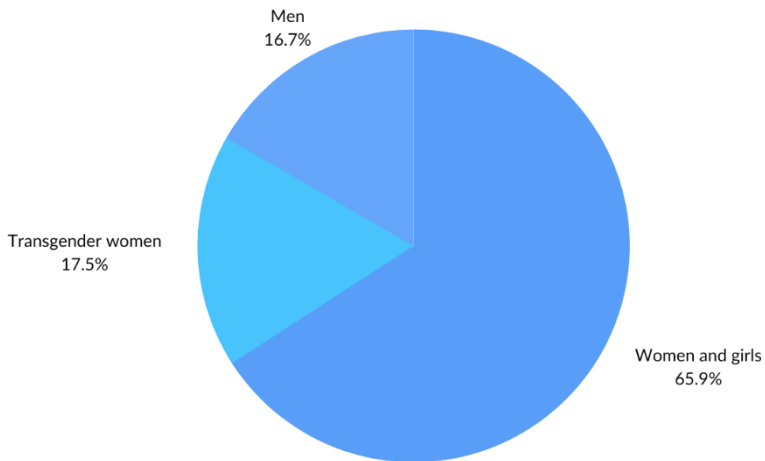

Gender breakdown of complaints received on helpline

---

[22] (2022) Cyber Harassment Helpline Annual Report 2022 - Digital Rights Foundation. Available at: https://digitalrightsfoundation.pk/wp-content/uploads/2023/05/Cyber-Harassment-Helpline-Annual-Report-2022-1.pdf (Accessed: 27 October 2023).

Breakdown by platform:

In terms of platforms, following global trends, Facebook commanded 33.3% (or 42 out of 126) of the source of harm distribution, with Instagram at 30.1% (or 38 out of 126) , X at 16.6% (or 21 out of 126) and TikTok at 7.9% (or 10 out of 126). The remaining percentage was divided amongst YouTube at 5.5%, multiple platform harm at 5.5% and harm emanating from websites at 0.79%.
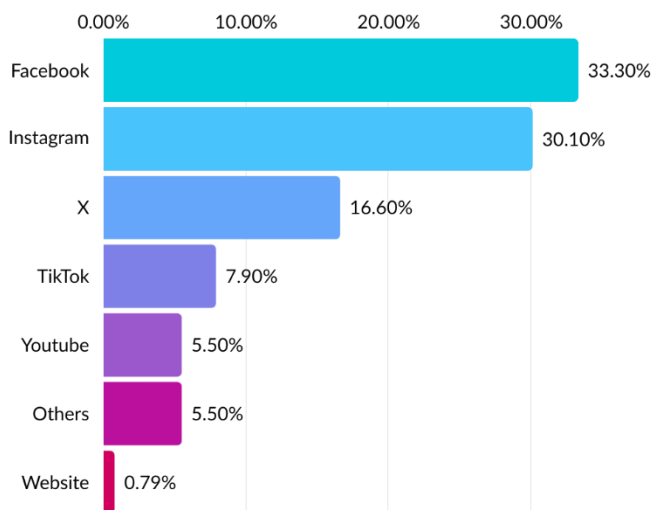

Breakdown of complaints

The multiple platform harm category signifies callers who are facing one or several forms of online harm through more than one platform.

Geographical data:

Following the general trend of richer data emanating from the most populous of Pakistan's provinces (and amplified perhaps by the location of the Helpline set up in the provincial capital, Lahore) Punjab was where most of our callers reached out from at 26.98% (34). Following that, the next highest number of complaints originated from Sindh at 16.6% (21) whereas Balochistan was at 8.7% (11) and KPK commanded 6.34% (8) of the geographical distribution. Complaints from Islamabad ranged to12.69% (16).

Complaints outside of Pakistan were 6.34% (8). It is pertinent to mention here that when people reach out to the helpline through DRF's social media platforms and helpdesk we do not always get the chance to ask them about their location in a non-intrusive manner and hence we do not have the information of all the complainants who reached out to us and these complaints were 22.2% (28).
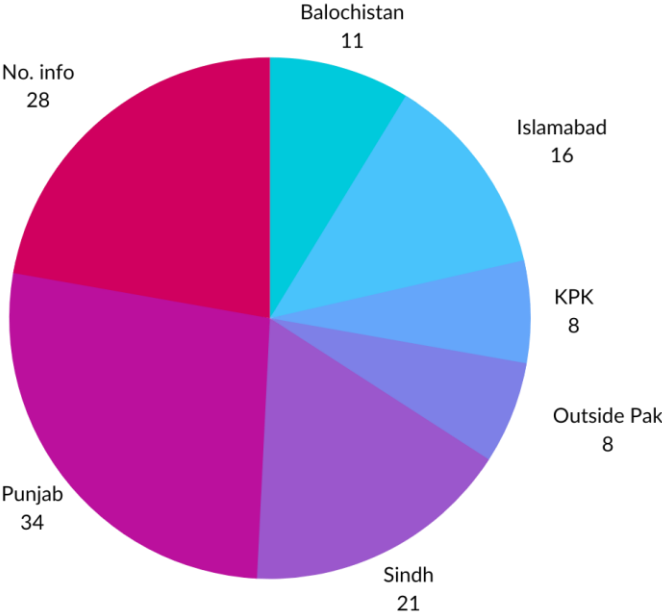
Geographical data

Vulnerable Identities:

For the purpose of this brief, and within the Helpline, human rights defenders (HRDs), transgender activists, and journalists have been categorized under vulnerable occupations and identities which would constitute those individuals or groups whose religious, sexual, and gender identities belong to the minority or rights-based minority sections of Pakistan.

Hate speech as defined under the Prevention of Electronic Crimes Act 2016 only includes information that 'advances or is likely to advance interfaith, sectarian, or racial hatred'.[23] However, the Helpline, and most social media platforms also consider language and information that attacks people on the basis of their gender or sexual identity as falling under hate speech.

From the 16 instances of hate speech recorded at the Helpline that were then escalated to social media platforms, 12 targeted activists, and 13 targeted a transgender individual or the community. There is a significant overlap between the two groups mentioned (activists and transgender identity), and it is pertinent to mention that these complaints came at a time of an intense online hate campaign targeting the transgender community, which eventually led to a withdrawal of the legal protections guaranteed to them by the state[24].

Technology-facilitated gender-based violence:

Technology-facilitated gender-based violence consists of harmful acts directed at an individual based on their gender. Harms filed under TFGBV as defined under the Cyber Harassment Helpline consist of online stalking, bullying, doxxing, fake profiles, non-consensual use of intimate images (NCII), non-consensual use of information (NCUI), hacking attempts, threats of physical violence, non-consensual use of porn (NCP), abusive messages, blackmailing and defamation. Out of the escalations made by the Helpline that concerned these specific types of harms, women formed 82.8% of the complainants, whereas transwomen constituted 9.1% of the complainants. The number of men whose cases regarding these specific claims were escalated also stands at 8.1%; however, it should be noted that the Helpline prioritizes escalating cases related to children, women, and other gender minorities, which may help to explain the low percentage of men in this data.

---

[23]
https://www.google.com/url?q=https://www.nr3c.gov.pk/peca16.pdf&sa=D&source=docs&ust=169935946 7246603&usg=AOvVaw1zXT5MM5kDFizT_IlDX1IO

[24] https://www.google.com/url?q=https://www.nation.com.pk/19-May-2023/shariat-court-strikes-down-three-sections-of-transgender-act-2018&sa=D&source=docs&ust=1699359467242488&usg=AOvVaw1_DHuZawemXwaDBLkIshKc

These harms account for approximately 1197 or 68.24% of the total online harms experienced by the callers at the Helpline in the year 2023, regardless of whether their case was escalated directly to social media channels or whether an alternate solution was provided. Among this set of complainants, 900 or 75% of them were women and girls and 19 or 1.58% of the complainants were transgender individuals.

'Pakistan ranks 145th out of 146 countries on the World Economic Forum's 2022 Global Gender Gap Index, and 161st out of 191 countries on the Human Development Report's 2022 Gender Inequality Index.' states a 2022 UNDP report[25] that looks at Gender Equality in Pakistan, Climate and Politico-Economic Stressors.

The impact of a deeply patriarchal society is evident not only from multiple global indicators but also from the nature of the cases the Cyber Harassment Helpline receives as per the 2022 DRF Annual Helpline Report. [26]

For the year 2022, Helpline recorded 58.6% of cases coming from those who identified themselves as women and 1% of the cases came from members of the transgender community.

An excerpt from the Annual Helpline Report 2022[27] reads:

'The gender distribution noted at the Helpline, however, cannot be considered to be exactly representative of the way online harassment manifests in the country. It has to be considered that many incidents of harassment, whether online or otherwise, go largely unreported. Furthermore, the Helpline is just one resource available in the country where complaints can be recorded, and it is possible that people reach out to other avenues instead.'

---

[25]https://www.undp.org/sites/g/files/zskgke326/files/2023-01/genderequality_in_pakistan-climate_politico-economic_stressors.pdf

[26](2022) Cyber Harassment Helpline Annual Report 2022 - Digital Rights Foundation. Available at: https://digitalrightsfoundation.pk/wp-content/uploads/2023/05/Cyber-Harassment-Helpline-Annual-Report-2022-1.pdf (Accessed: 27 October 2023).

[27]Ibid.

When looking at trends impacting vulnerable and intersectional identities, we saw 4 out of 5 cases where the caller was threatened with an indirect blasphemy allegation, coming from those who identified as (3) cisgender women and (2) transgender women.

Given the particularly grim state of Pakistan's growing religious intolerance, blasphemy allegations are often used to overpower and silence dissenting voices by employing Section 295 (c) of the Pakistan Penal Code (PPC) which states that the use of derogatory remarks when referring to Prophet Muhammad (pbuh) is punishable by death. [28] .

This law has been used, or rather misused, especially when it comes to the marginalized communities in the country, as a nuclear option. As we see from the numbers above, the record available with the Helpline is in line with these general trends.

Services Provided & Status Overview:

Service provided:

The primary strand of service provided in terms of numbers was Digital Help which was the form of help required by 60.6% of the callers at the Helpline. Solely legal assistance accounted for 11% of the services and legal and digital combined stood at 5.5%. The helpline provides additional services which are not reflected in the numbers here since it's not relevant to this policy brief.

These numbers indicate the prevalence of digital harms that exist and all communities, especially vulnerable ones, face in Pakistan.

---

[28] Pakistan Penal Code ,1860. Available at: https://pakistancode.gov.pk/english/UY2FqaJw1-apaUY2Fqa-apaUY2Npa5lo-sg-jjjjjjjjjjjjjj (Accessed: 27 October 2023).

In terms of processing these cases forward, the classification is simple: those cases that have been escalated, those that have then been rejected or resolved, those under processing and those that have been referred further.

On the Helpline out of the 1754 cases received from January to August 2023, 101 complaints were escalated to social media companies.

Of the 126 cases that were escalated 47 complaints were received from vulnerable occupations. The vulnerable occupations identified by the Helpline team are as follows: 63.8% activists, 2.13% civil society, 4.26% government figure, 17.02% journalists, 4.26% lawyers, 2.13% media practitioners, 4.26% politicians and 2.13% others.
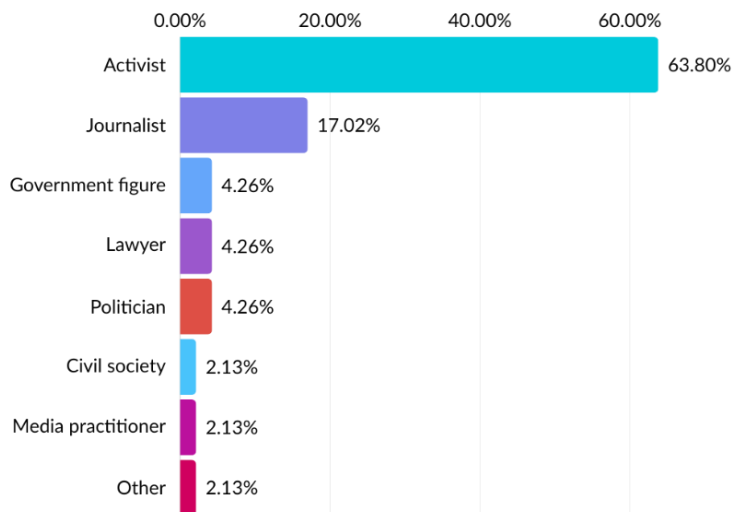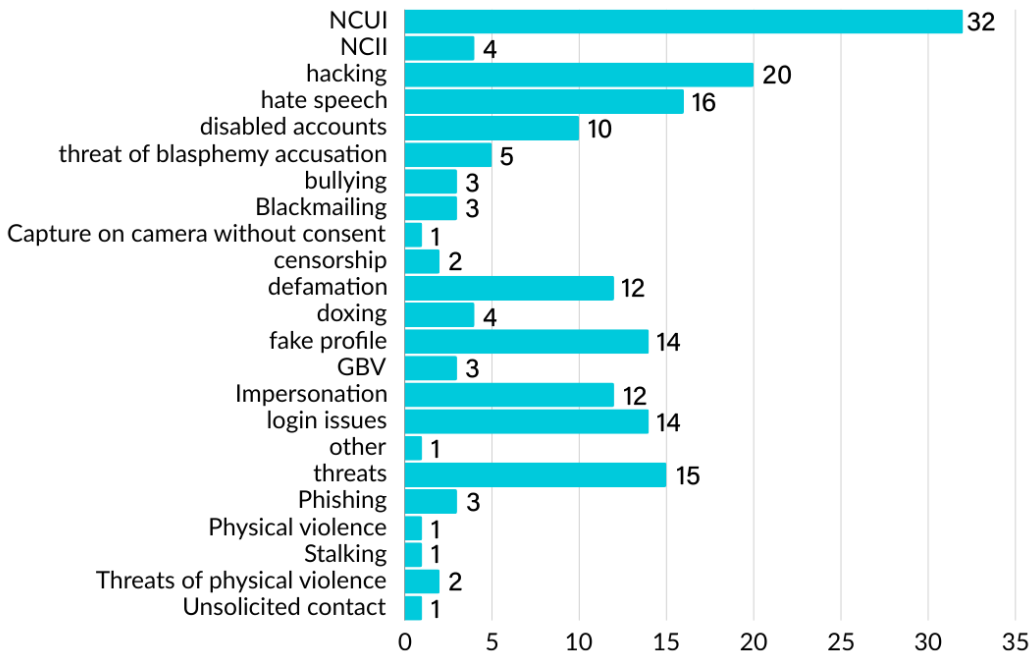

Status of case

A broader scale demarcation that the Helpline employs is to ask whether the caller requires help regarding cyber harassment or another issue pertaining to the digital realm. 87.3% or 110 out of 126 callers' data set required advice and/or services regarding cyber harassment.

When looking at the types of online harassment faced by callers, numerous harms emerge. The Helpline categorizes this data into 34 types including, but not limited to defamation, phishing, spam, doxing, disabled accounts, GBV, NCII, hate speech and fake profiles.

The largest share here can be attributed to non-consensual use of information (NCUI) which is defined as the violative spread of someone's personal data such as photos, videos, phone number, etc., and has been done so without their consent. NCUI (and NCII (non-consensual intimate image abuse) account for 36 complaints here or 28.5% of the total complaints escalated in the defined 8-month period. This is followed by concerns related to hacking which are the subject of 20 or 15.87% of the calls received by the Helpline.

In slightly smaller numbers we see hate speech, which accounts for 11.1% and disabled accounts which factor in at 7.9% of the overall categories of complaints received. The strongest trends we see in terms of the nature of complaints received at the Cyber Harassment Helpline are those that are similar across the board: gender and vulnerability play a decisively impactful role in determining who can be at a higher risk for being accosted in online spaces.

Type of harm	Count
NCUI	32
NCII	4
hacking	20
hate speech	16
disabled accounts	10
threat of blasphemy accusation	5
bullying	3
Blackmailing	3
Capture on camera without consent	1
censorship	2
defamation	12
doxing	4
fake profile	14
GBV	3
Impersonation	12
login issues	14
other	1
threats	15
Phishing	3
Physical violence	1
Stalking	1
Threats of physical violence	2
Unsolicited contact	1

Types of harms

* The Cyber Harassment Helpline recognizes that each case brought forward by a complainant can be multifaceted, and so gives space to record three categories of complaint per case. The numbers above, therefore, add up to a total higher than the total number of complaints because many callers who reach out to the Helpline have complained of more than one type of harm they've experienced.

The responses recorded from major social media platforms in terms of the escalation made to each stand at the following rate:

Of the 126 cases received at the Helpline, 42 pertained to online harms experienced on Facebook. Out of these, 10 (23.8%) stood resolved and 2 (4.7%) were rejected. In some cases, a response from the escalation team is not received which makes it more difficult to cross-check whether a case has been resolved, and this stands for Instagram and X as well. This, combined with the Helpline's policy of not following up with complainants without their express permission also makes it difficult to track escalation results.

For Instagram, the Helpline escalated 38 cases between January and August 2023, where 5 (13.1%) received a rejection from the platform, and in 10 (26.3%) of the instances the issue was resolved by Instagram.

We received 10 calls where the issue was emanating from the video recording and sharing platform TikTok. Here, in 9 (81.8%) instances the complaint was resolved by TikTok, whereas in 1 case, a response was received and forwarded to the complainant, but the Helpline was not informed further of any further developments.

For X, we received a total of 20 escalation requests, whereas in 8 (40.0%) of the cases that were received at the Cyber Harassment Helpline, the platform resolved the complaint. The remaining 12 (60%) cases did not receive a response.

YouTube was the source of an online safety issue for 7 of our Helpline callers, where six of the cases were escalated to the platform, of which two (28.5%) were resolved. Four cases were at the escalation stage and one was resolved without the need arising for an escalation request to be sent to the platform. The remaining 9 escalation requests were from other platforms.

Others
9

Youtube
7

TikTok
10

Facebook
42

X
20

Instagram
38
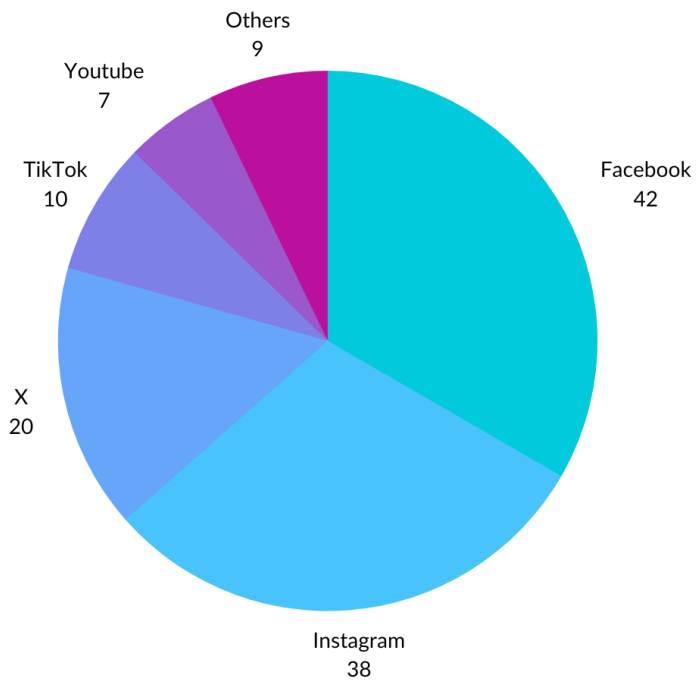
Response from social media companies

Recommendations

Big Tech Platforms

- Content moderation policies of social media platforms should be consistent with the obligations of corporations to respect and promote human rights, as set out in the UN Guiding Principles on Business and Human Rights[29] and other established international human rights standards.
- Platforms should take immediate, timely and decisive action to remove any content (textual, visual, live streaming etc) that is inciteful towards communities and groups which are already marginalized.
- Platforms must invest in training its human content moderators on understanding the cultural and linguistic complications tied to online harmful content and threats targeted at vulnerable groups in Pakistan. Much of the hateful content posted online is in languages other than English, and therefore, takes longer to report and remove. To cater to this, platforms must ensure that they are sufficiently staffed and trained to meet the urgency of certain content removal requests and that online harmful content in all languages are being treated with the same urgency
- Platforms must revise community guidelines to account for the media content that is produced on important subjects particularly keeping in mind the regional context. While regulation of content is important and in some cases necessary, especially when it is weaponized for hate speech and disinformation, over-regulation or regulation without discerning such content from genuine reporting leads to censorship and silencing of critical voices in journalism and media and restricts access to accurate information for the general public.
- A transparent responsive appeals mechanism for content decisions should be implemented by all social media platforms which needs to be adequately resourced and accessible to ensure context-appropriate and timely redressal. Given the non-transparent and arbitrary nature of content moderation decisions, individuals who are impacted rarely have redressal mechanisms to appeal decisions made against them and little control over how their content is regulated.
- Platforms should make it a practice to release bi-annual transparency reports regarding content removal across the globe. These reports are essential in developing policy frameworks for platforms and should be available in regional languages so that more people are aware of content removal requests and policies adopted by tech platforms and there's more accountability of these platforms.

---

[29] *UN Guiding Principles on Business and Human Rights* (no date) Business & Human Rights Resource Centre. Available at: https://www.business-humanrights.org/en/big-issues/un-guiding-principles-on-business-human-rights/#:~:text=The%20UN%20Guiding%20Principles%20on%20Business%20and%20Human%20Rights%20are,abuses%20committed%20in%20business%20operations. (Accessed: 26 October 2023).

- Big tech platforms when deploying the use of artificial intelligence (AI) must be mindful of the implications it can have for marginalized groups in countries residing in South Asia. Human rights impact assessment mechanisms should be adopted by big tech platforms to ensure that emerging technologies are not amplifying hateful content against marginalized groups
- Change in privacy terms and content regulation policies by big tech platforms must be transparent and available in regional languages in Pakistan so that more audiences are able to understand data privacy mechanisms adopted by tech platforms for countries in the Global South

Government

- The regulatory models developed by the government should focus specifically on content that is expressly illegal and harmful which is clearly defined with priority categories. Use of vague terms must be avoided for regulation. Any restrictions to the right to freedom of expression must be clearly prescribed by law, pursue a legitimate aim, be necessary in a democratic society, and be proportionate to the aim pursued so that the misuse of the law to silence dissent can be avoided.
- The government must enact human rights-compliant legislation on digital privacy and protection after meaningful consultations with civil society and the general public. The right to dignity and privacy, as guaranteed under Article 14 of the Constitution of Pakistan, must be protected for every citizen. DRF's comments to the Ministry of Information Technology and Telecommunications (MOITT) on the Personal Data Protection Bill 2021 should be taken into account when drafting the law.
- The government should include Internet education and safety courses in school curriculums. Topics such as consent, social media ethics, safety practices, and what is illegal online must be covered. This will empower the younger generation to be more confident and aware while exploring the internet.
- Government and Law Enforcement Agencies should work with civil society and media literacy organizations to create awareness regarding harassment, online harassment and rights around free speech online. Frequent sensitization training of law enforcement led by civil society on the importance of journalist welfare and safety need to be facilitated and supported by the government.
- Government needs to allocate more resources to the Federal Investigation Agency's cyber crime wing to increase its capacity to process cases quickly and in a sensitized manner. There must also be rules and protocols in place to safeguard vulnerable groups from being harassed by the agencies' personnel.
- The government must collaborate with other countries in the region and international organizations to address the similar nature of online harms that they face and how cross-border cooperation can be useful.

- They must also conduct regular reviews of the laws and regulations to be able to adapt and stay current with the evolving nature of online threats and platform technologies.

Civil society and the media

- Civil society and the media should develop methodologies to monitor platforms for harmful content that targets vulnerable groups, document evidence and conduct advocacy and awareness campaigns.
- They must put pressure on the platforms and demand transparency from them regarding their content moderation policies, practices, and the outcomes of their moderation efforts.
- They can support digital literacy and online safety programs aimed at vulnerable groups to empower them with the skills to protect themselves online.
- Civil society groups can collaborate with the media to run awareness campaigns on how online harmful content impacts those at the receiving end.
- Civil society groups can collaborate with big tech companies and governments to ensure that laws, policies and content moderation guidelines are human rights compliant and ensure the protection of the rights and privacy of the users.