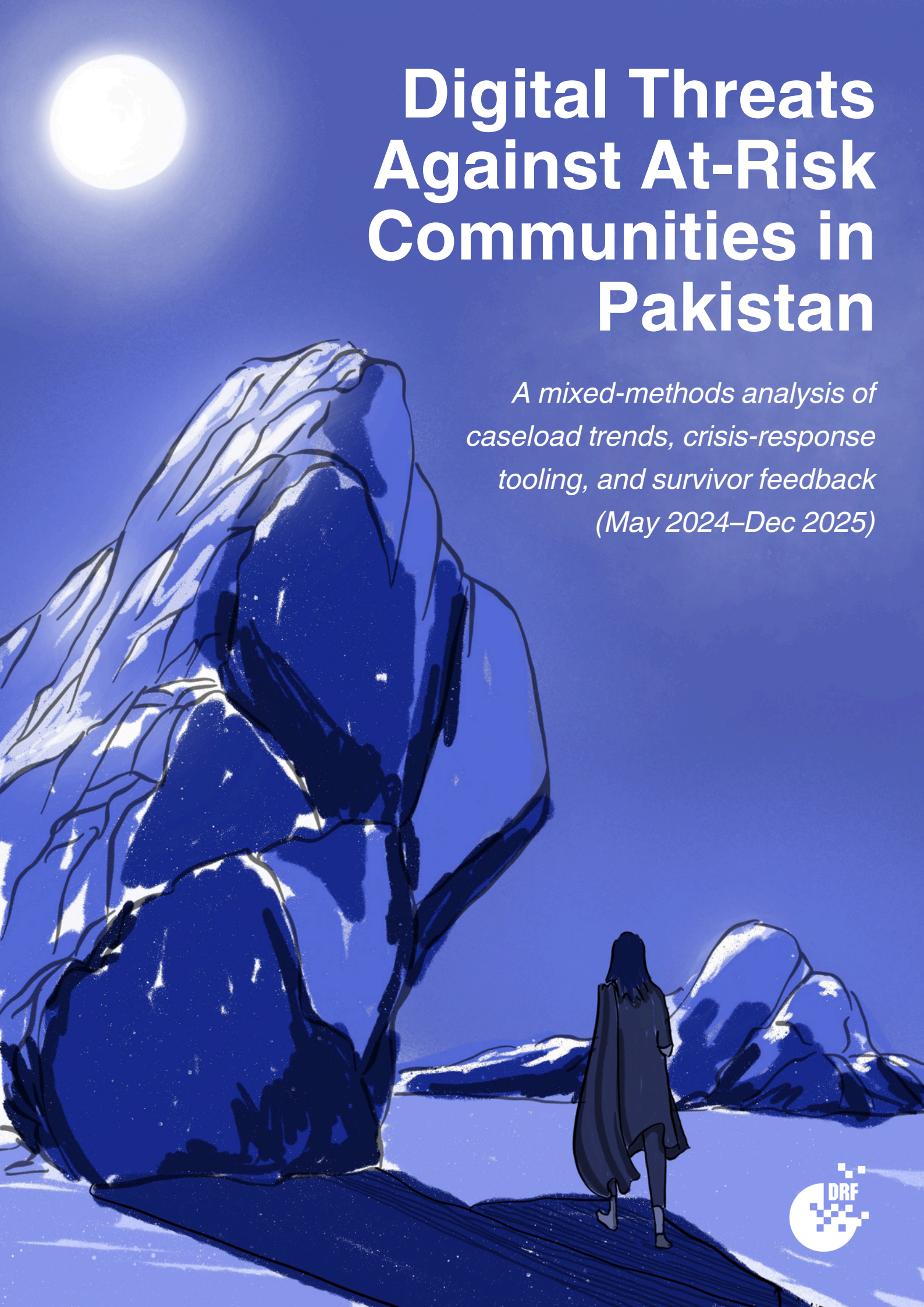
Digital Threats Against At-Risk Communities in Pakistan

A mixed-methods analysis of caseload trends, crisis-response tooling, and survivor feedback (May 2024–Dec 2025)



About Digital Rights Foundation

February 2026

Digital Rights Foundation (DRF) is a women-led, not-for-profit organization based in Pakistan, working since 2013 to advance digital rights, freedoms, and online safety for all. While our work centers on the lived experiences of women and gender minorities, we actively support and collaborate with religious minorities, human rights defenders, journalists, and civil society organizations across the region.

Contact information:
info@digitalrightsfoundation.pk
www.digitalrightsfoundation.pk

Author: Anam Baloch
Reviewer: Hyra Basit, Seerat Khan
Design and Layout: Ahsan Zahid and Talha Umar
Cover page: Bushra Saleem

Glossary

2FA — Two-factor authentication; an extra login step (e.g., app code/SMS) used to secure accounts.

CENO — Ceno Browser (“Censorship.no!”); a browser designed to help people access content during blocking/outages, including via peer-assisted delivery.

DRF — Digital Rights Foundation.

HTTPS — Hypertext Transfer Protocol Secure; encrypted web traffic between a user and a website.

IODA — Internet Outage Detection and Analysis; a system for monitoring large-scale Internet outages/disruptions.

KPK — Khyber Pakhtunkhwa (province of Pakistan).

n= — Sample size; the number of observations/respondents used in a specific statistic.

NCCIA — National Cyber Crime Investigation Agency (Pakistan).

NCII — Non-Consensual Intimate Images; sharing or threatening to share intimate images without consent.

NCUI — Non-Consensual Use of Images; broader non-consensual image misuse (used here to include edited/manipulated content, including deepfakes, where relevant).

OONI — Open Observatory of Network Interference; tools and research that measure internet censorship, filtering, and disruptions.

TOR — The Onion Router (Tor); privacy/circumvention network and browser used to reduce tracking and, in some cases, bypass restrictions.

VPN — Virtual Private Network; encrypts a user's connection and can help reduce surveillance and access restrictions depending on context.

Table of Contents

1. Executive Summary

This report documents the digital threats faced by at-risk communities in Pakistan and evaluates the usability and effectiveness of crisis response tools recommended through Digital Rights Foundation's (DRF) Digital Security Helpline (formerly known as Cyber Harassment Helpline). This aligns with this project's goal to capture real-time incidents and produce practical feedback on tools recommended to at-risk communities during their times of crisis. The analysis in this report triangulates cumulative helpline caseload trends, dedicated digital security support workstream (tool recommendation logs), structured surveys (n=76) from the standardized helpline feedback form (n=55 total; English 43 + Urdu 12) and the digital security help dedicated feedback form (n=21), and five qualitative interviews with high-risk personnel across journalism, law, minority rights activism, hate speech monitoring/student activism, and transgender community protection work.

During the period of data collection for this report i.e May 2024 - December 2025, DRF's helpline handled 5,041 new cases with 2,029 in 2024 and 3,012 in 2025. Across gender-segregated issue categories, a high volume of complaints included hacking, blackmailing or sextortion, threats, image-based abuse (NCII/NCUI including edited or deepfake imagery manipulation), and social engineering or financial fraud. The complaints dedicated to digital security help, which provides a subset of evidence on digital security tools, were logged as 97 cases (46 in 2024 and 51 in 2025). The dominant tool recommendation was that of MalwareBytes and limited instances of OONI, LastPass, and HTTPS, primarily to resolve account compromises and hacking-related incidents.

In May 2024 - December 2025, 64% (53) of helpline survey respondents received an initial response within minutes, 93% (54) received digital safety advice, and 92% (49) reported reduced risk after support. In the dedicated Digital Security form, 79% (19) rated tools/advice helpful, with MalwareBytes being the most frequently used recommended tool, according to the survey feedback.

Both the surveys and interviews showcase that survivors prioritize rapid, guided triage and recovery, and report improved perceived safety after support, and show uneven tool adoption driven less by awareness than by cost, usability, and platform responsiveness constraints.

2. Context

Digital Rights Foundation's Digital Security Helpline, formerly the Cyber Harassment Helpline, emerged from DRF's direct engagement with individuals facing online abuse and insecurity in Pakistan. First established in 2016, the helpline was shaped by the urgent need for practical, survivor-centered support after DRF's online safety trainings revealed how many women were experiencing harassment and seeking immediate guidance. Over time, the service expanded beyond responding only to technology-facilitated gender-based violence to address a broader range of digital threats affecting civil society actors, journalists, human rights defenders, and other at-risk communities. Its relaunch as the Digital Security Helpline reflects this wider mandate: providing specialized crisis support, tailored digital safety guidance, and informed tool recommendations to people navigating increasingly complex forms of online harm and surveillance.

At-risk communities, particularly women, religious minorities, and gender minorities, face digitally mediated harm that is based both on their identity and is amplified by platforms' algorithms and dynamics. This report aims to fill the gap in quantifiable evidence on digital threats in Pakistan by using helpline incidents as real-time signals and pairing them with tool efficacy feedback for crisis response.

Our collected data demonstrates that digital threats do not occur in silos and cannot be simplified as mere online abuse. Visibility of transgender identity and public service roles (akin to serving or activism) can lead to doxxing, sexualized abuse, and death threats, particularly when snippets of curated media clips go viral. Advocacy around religious minority rights attracts coordinated digital hate campaigns that increase offline and physical risk. Political speech and student activism face layered risks, such as becoming targets of propaganda and algorithmic suppression that shrinks the reach of human rights defenders and increases overall uncertainty. And women journalists and feminist lawyers face persistent sexualized harassment, self-censorship and content deletion to preserve professional credibility.

3. Methodology

This report combines DRF's helpline case data, surveys, and interviews to understand both threat patterns and tool efficacy and concurrent usability during any crisis. The data sources for this report in the May 2024 - December 2025 period are as follows:

- Cumulative helpline caseload in this time period: new cases 5,041. This includes all types of complaints received at the helpline.
- Dedicated Digital Security Help tool logs through helpline complaints: 97, including complaint types and tool recommendations.
- Surveys analyzed: total 76;

    Digital Security feedback form: 21 and Standard helpline feedback (English + Urdu): 55; English 43 and Urdu 12. The former was circulated amongst helpline complainants particularly for this report and the latter is the feedback form sent by the helpline to all complainants and its optional to fill.

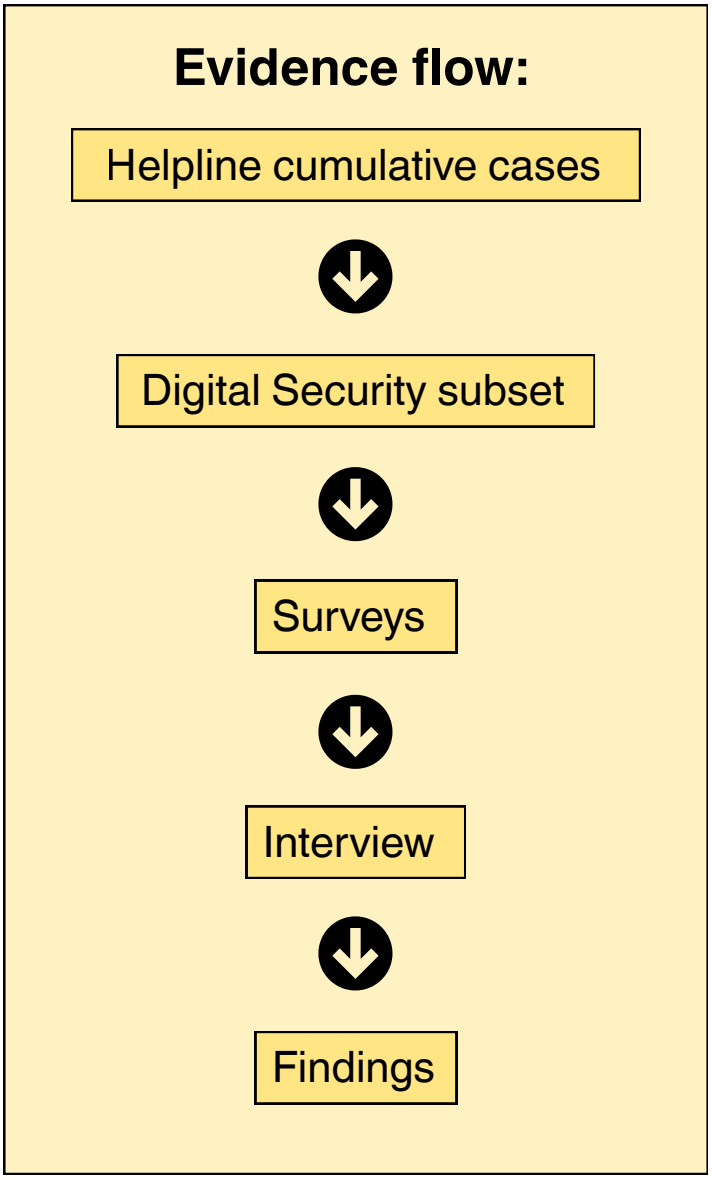
- Interviews: 5; journalist, lawyer, student activist/hate speech monitor, minority rights founder, transgender victim support officer. They were selected as individuals from at-risk communities/vulnerable occupations and have had contact with DRF's helpline before.

All the interviewees chose to keep their identities public and their profiles are as follows as their identities and work inform the risks they face:

Interviewee	Primary role/profession	At-risk community lens/constituency	Digital footprint (typical platforms)
AB	Victim Support Officer; transgender community focal person	Transgender community; marginalized communities policy advocacy	Primarily Facebook; harassment also linked to TikTok content reposted by influencers/news
JK	Final-year medical student; minority rights activist	Hindu community / religious minorities; broader minority coalition organizing	Active on X (Twitter), Facebook, Instagram; WhatsApp used for mobilization/forwarding
AR	Activist; hate speech monitor; researcher/writer	Student activism; rights-based advocacy	Facebook and X (Twitter) central; notes WhatsApp used for rumor circulation/private doxxing
MO	Human rights lawyer; activist educator	Women's rights / feminist organizing; work includes awareness on NCII	Heavy use of Instagram and X (Twitter); impersonation harm via TikTok
LZ	Woman journalist and activist	Women's rights activism; journalism targeting by political troll groups	Uses X (Twitter), TikTok, Facebook, Instagram, YouTube multiple times daily

Evidence Snapshot Box:
5,041 total cases; 97 dedicated digital security cases; 76 survey responses; 5 interviews.

There is a mix of quantitative and qualitative data to parse through. The quantitative data was analyzed using descriptive statistics such as counts and proportions. The gender-segregated categories are reported as complaint type entries rather than unique cases because a single complaint can involve multiple overlapping harms, for example, blackmail can be followed by hacking and threats and vice versa. The qualitative data from interviews and open-ended surveys, on the other hand, were thematically coded to identify recurring patterns across key dimensions, including threat type, platform attack, coordination dynamics, response pathways, tool use, barriers to safety, and resulting psychological, professional, and civic impacts.



4. Detailed Analysis of Digital Threats

4.1 Trends and Platforms

Across the evidence collected through helpline (logs and surveys) and interviews, threats seem to concentrate on mainstream platforms i.e Facebook, Instagram, WhatsApp, and TikTok, with platform-specific attack patterns. In minority rights organizing, an interviewee relayed that Facebook functions as the central site for harassment following any virality and has an abundance of comment-driven abuse, while WhatsApp plays a key role in forwarding mis/disinformation and mobilization of adversarial groups. Interviews conducted also showcased that X enabled coordinated trolling and targeted harassment more than other platforms and particularly around political flashpoints, and the human rights lawyer interviewee personally experienced that TikTok was at the forefront in accelerating impersonation and continued and exponential virality, increasing reputational harm and safety risk for women and at-risk communities, especially those in public-facing professions.

The survey responses reinforce what access to safety looks like in practice, as most standard helpline respondents accessed support through direct, low-friction channels (toll-free helpline number, WhatsApp, and email). This indicates that crisis response depends on fast and familiar contacts rather than complex reporting frameworks and pathways.

4.2 Case Types (Major Complaints)

Helpline's gender segregated complaint categories give rise to two intersecting profiles where women and trans women had elevated reports of account compromise, blackmail/sextortion, threats, and image-based abuse (including edited/deepfake manipulation), alongside social engineering and financial fraud, which reflects a coercion and reputation harm model. Men, on the other hand, have higher volumes of financial fraud and hacking, with social engineering and information requests also being prominent in their profile. This distinction highlights the gendered and sexualized nature of the threats women face in ways that differ from men. Even when the abuse takes the form of threats or account compromise, it is often framed through deeply gendered language and social expectations. For women, this can involve invoking notions of "honor," weaponizing their public visibility or participation, and sexualizing their image or social interactions in order to defame, shame, and control them. And the gendered threat profile can be visualized by the side-by-side top categories data of complaints below:

Rank	Women + Trans Women	Count	Rank	Men	Count
1	Hacking	531	**1**	Financial Fraud	732
2	Image-Based Abuse (NCUI + Edited/Deepfake)	514	**2**	Hacking	453
3	Blackmail	500	**3**	Social Engineering	295
4	Threat	491	**4**	Information Request	264
5	Financial Fraud	402	**5**	NCUI	97
6	Social Engineering	337	**6**	Defamation	91
7	Sextortion	282	**7**	Threat	89
8	NCII	270	**8**	Account Disabled	70

The interviews conducted clarified the mechanisms behind these categories, where coordinated harassment often becomes a problem of amplification as harmful content continues to gain reach even after reporting to the platforms' internal complaint systems. This compounds the amount of damage being done before the content is taken down (and content removal isn't always the resolution to complaints). For women journalists in particular, threats and exposure escalate to offline fear and long-term behaviour change; when threats are extended beyond their person and to their homes/families, many of them reduce or halt posting on sensitive topics and deliberately limit their public presence.

4.3 Impact - Psychological, Social, and Professional

The impact layer is consistent across the interviews as all of them reported a psychological toll in the form of chronic fear, exhaustion, hypervigilance, and anxiety as a result of combating online attacks and coordinated hate campaigns. There is also professional harm faced by the interviewees as attacks on their person or mis/disinformation about them can lead to reputational risk, reduced reach, and withdrawal from engagement to preserve credibility and safety; the latter was more common in speaking with the women and trans women interviewees. Their civic participation also faces constraints as online pressure from hate groups, as in the case of minority rights activism, shapes offline permissions and policing and the ability to organize safely is, therefore, directly linked to digital attacks.

Survey feedback responses also support this impact layer assessment practically. Across both survey streams, standard helpline feedback and dedicated digital security helpline feedback, respondents frequently describe support as reducing immediate risk and providing guidance that makes the situation feel more manageable during a crisis.

The impact layer can be visualized in the form of the impact ladder below:

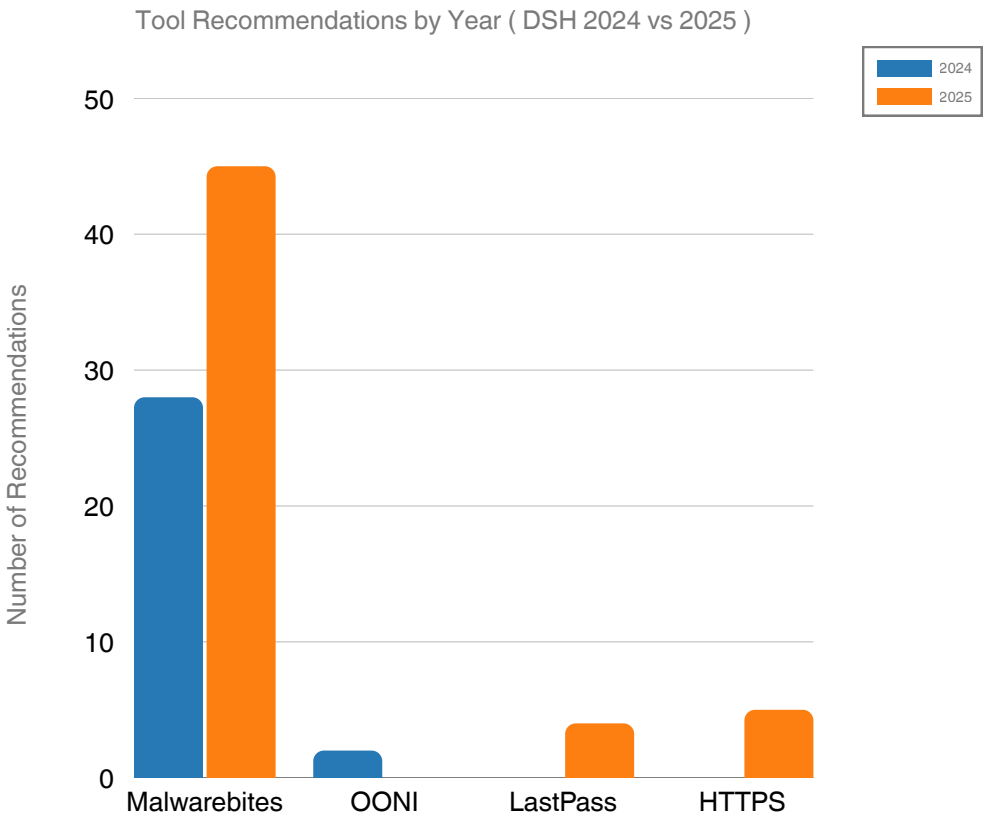
5. Tools and Efficacy

5.1 Distribution of Tool Recommendations Across Digital Security Crisis-Response Cases

The dedicated digital security help offered through the helpline showcases what crisis response tool usage looks like in practice.

In 2024, the helpline received a total of 46 cases requiring only/exclusively digital security assistance. The major complaints were hacking, blackmail, online stalking, and phishing attacks. 45 out of 46 times, MalwareBytes was recommended as a tool to these complainants and OONI was recommended one time for WhatsApp media download issues and broader application malfunction affecting a journalist in KP.

In 2025, the Helpline received 51 such cases. The major complaints here were also hacking, blackmail, online stalking, phishing attacks, and social engineering complaints. The tools recommendation in 2025 was also dominated by MalwareBytes, which was recommended 30 times, followed by LastPass, recommended 2 times and HTTPS Everywhere, recommended 2 times as well; the last two were both responses to cases of hacking.



Across the entire digital security helpline logs, tool recommendations overwhelmingly align with the most time-sensitive harm category, account compromise, suggesting that efficacy should be primarily assessed through containment and recovery speed, not only long-term tool usage for prevention and long-term behaviour change. In this setting, the Helpline favors tools that are fast to deploy, simple for non-technical users, and compatible with constrained devices and unstable connectivity. Tools that require higher technical prowess, sustained behaviour change, or complex installations are less likely to be recommended in an acute incident, especially when the complainants are already stressed and need immediate resolution and stabilization.

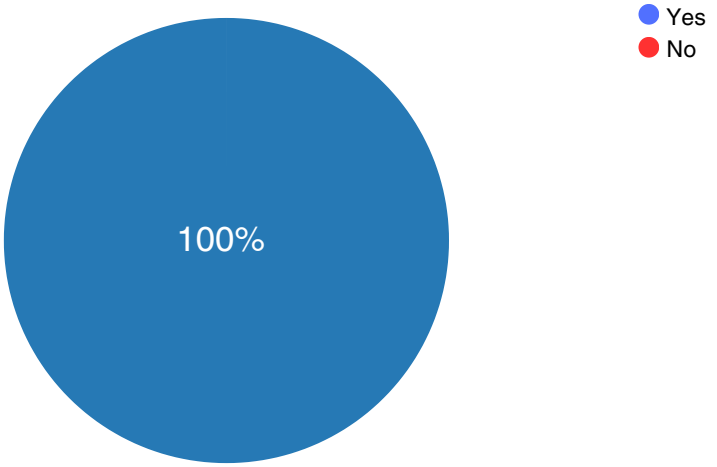
A second driver to explain the lack of variety in tool recommendations is that most protections are process-based rather than tool based i.e password resets, account recovery, 2FA activation, device and digital hygiene, evidence preservation, and platform escalation. And these processes are more decisive than installing additional applications. The interviews conducted also suggest adoption constraints that make some tools impractical even if they are effective in theory. The adoption constraints range from cost barriers, device storage limitations, and network-effect problems, where secure alternatives struggle because communities continue to engage on mainstream platforms. In practice, this means the helpline tends to recommend a small number of low-friction tools and prioritizes step-by-step recovery actions over introducing a large tool stack that complainants may not be able to maintain.

Lastly, platform accountability limitations influence the scope of tool recommendations: extra tools cannot replace prompt platform action when harm is caused by viral amplification, impersonation, and poor reporting results. Because of this, helpline recommendations for digital security complaints concentrate on things that the helpline can directly manage, such as account/device recovery and safety procedures, while escalation pathways deal with things that call for outside intervention, such as takedowns, impersonation enforcement, and coordinated harassment mitigation.

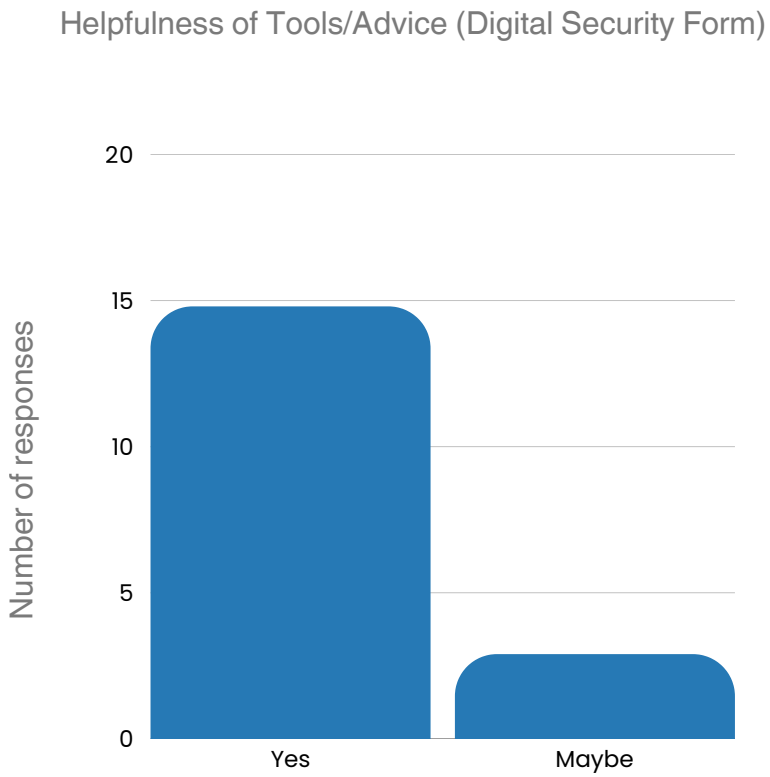
5.2 Survey evidence on adoption and perceived usefulness

In the Digital Security Helpline dedicated feedback survey (n=21), respondents overwhelmingly reported receiving digital safety advice and frequently described it as helpful. And among those who answered impact questions (90.5%), all of them reported that the advice or the tool recommended reduced the risk they were facing.

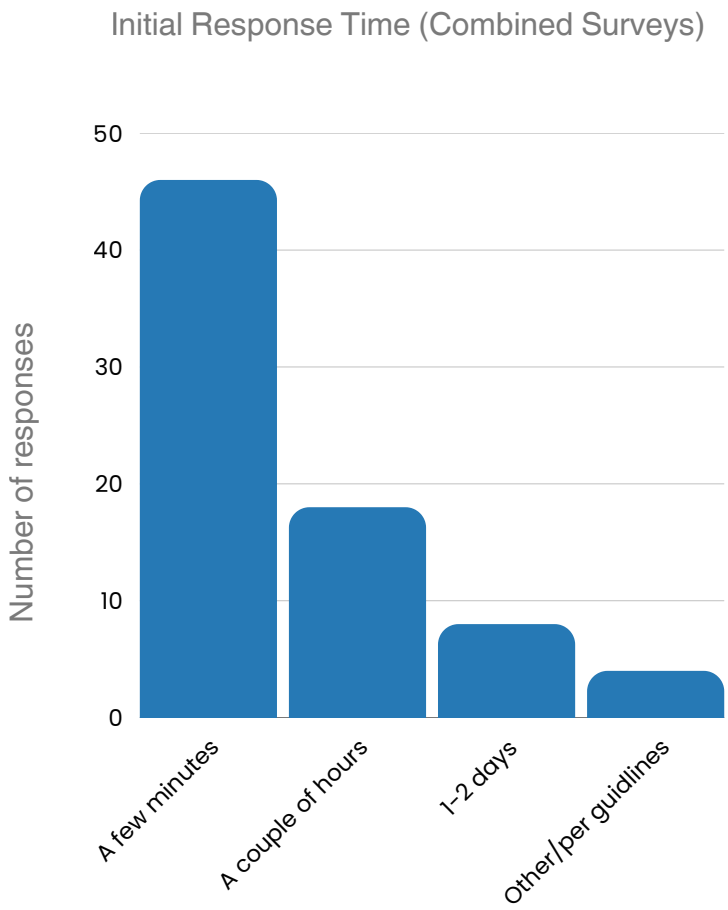
Did the digital help you received reduce the risk you were facing?
19 responses



Tool-use selections indicate that adoption concentrates around compromise and safety tools (notably MalwareBytes) alongside secure communications and access tools.



In the standard helpline feedback surveys, the dominant success signals are speed of response, receipt of actionable guidance, and perceived risk reduction. The intake channels recorded emphasize the importance of low-friction entry points such as the toll-free helpline number, WhatsApp and email during a crisis and for filing complaints.



In that manner, survey findings align with the anecdotal evidence from the interviews, where victims and survivors value immediate, guided triage and recovery steps the most. Additionally, they also support the notion that sustained adoption of secure tools is constrained by cost, convenience, device limitations, and coordination needs.

5.3 Tool Efficacy Table

In the Digital Security Helpline dedicated feedback survey (n=21), respondents overwhelmingly reported receiving digital safety advice and frequently described it as helpful. And among those who answered impact questions (90.5%), all of them reported that the advice or the tool recommended reduced the risk they were facing.

Tool category	Recommended tools	Evidence-linked efficacy summary
Security / recovery	MalwareBytes, 2FA, password hygiene	Dominant in digital security helpline logs for compromise recovery; interviews and surveys support triage-first workflows: secure accounts, scan devices, reset credentials, enable 2FA.
Browsers / access	TOR, CENO, Lantern	Used when access is constrained; circumvention becomes situational infrastructure during politically sensitive periods and network stress.
Communication	Signal, Briar, New Node	Effective for sensitive coordination, but adoption is constrained by network effects (WhatsApp remains default) and practical constraints (storage/usability).
Monitoring	OONI, IODA	Used selectively to verify disruption/outage conditions, especially during political unrest or platform malfunction.

6. Barriers to Digital Safety

6.1 Cost barriers to safety

Paid tools are often the difference between working reliably and failing at the worst moment. In practice, this creates a two-tiered security reality where users with stable income, payment accessibility, and technical know-how can maintain high-quality VPNs, password managers, and premium protections, while at-risk users (especially those with low tech literacy) must rely on free (and often unsafe) or unstable alternatives. Outside of major urban centers and cities, cost barriers are compounded by connectivity constraints; meaning even when people know what to use and how to use it, they cannot sustain it. The result is a predictable protection gap where the communities facing the highest harassment risk have the least access to reliable protective infrastructure.

6.2 Usability vs security trade-offs and network effects

Secure tools adoption is not just based on trust, it is also about the workflow. Complainants choose tools that preserve coordination, speed, and familiarity, especially during times of crisis. Secure tools lose their ground when they require additional steps, introduce any sort of friction, such as onboarding contacts from other platforms, storage and backups, or disrupt everyday communication norms.

This perceived ‘network effect’ is structural in nature because if the community remains on, for example, Meta platforms i.e WhatsApp, Instagram and Facebook, the move to safer alternatives can isolate the individual user socially and operationally. The practical outcome would be partial security where high-risk conversations move to secure channels and daily coordination remains on mainstream platforms, leaving persistent exposure points.

6.3 Platform reporting fatigue and weak responsiveness

Through data collected, this pattern emerged that reporting to the platforms is often experienced as an exhaustion pathway rather than protection. Complainants must repeatedly collect screenshots, fill forms, appeal decisions, and still face inconsistent outcomes. This leads to reporting fatigue where people just stop reporting, not because harm ends, but the cost-benefit imbalance is not worth it.

Meanwhile, platform review timelines rarely ever match the velocity of harm, as impersonation and harassment that goes viral can spread faster and before the take-down decision is even processed. In these scenarios, survivors turn to helplines such as the DRF's helpline and trusted partner networks because they offer a human response, guided triage, and containment steps, even if platforms remain slow.

6.4 Language and context moderation failures

Harassment in Pakistan relies on coded language, local slang, religious and political insinuation, and context-specific targeted hate campaigns. And when moderated systems, either human or automated, are not equipped to interpret such harassment, the hate and abuse are more likely to be dismissed as non-violating, even when it is clearly threatening or inciting offline harm to local communities. As such, it is more of a context problem, rather than a language problem, because understanding harassment requires interpreting identity markers, local political triggers, and community norms. This results in uneven protection as the same content that would trigger take-downs or an action in one context or region may be ignored in another, systematically disadvantaged marginalized communities.

6.5 Crisis conditions multiply friction

People default to harm-minimizing behaviour when they are under threat. They reduce their posting, delay speaking up on sensitive topics, delete content, and limit their online visibility. This can improve their immediate sense of safety but carries heavy long-term costs, i.e., self-censorship, professional withdrawal, and reduced civic participation. This, sadly, also creates an evidence paradox where the most severe environments of harassment and hate campaigns produce the least public documentation because victims and survivors are pushed offline or into private, secluded channels. That makes external measurement harder and reduces accountability pressure, reinforcing a cycle where high-risk moments coincide with lowest visibility and the greatest need for rapid and reliable response.

These barriers showcase that digital safety issues are structural, not individual. The risk doesn't go away because survivors "don't know what to do," but because affordable tools, usable secure workflows, and responsive platform enforcement are not always available. This means that platforms, government, and civil society need to work together to put proper safeguards in place for survivors and victims of online harassment and abuse.

7. Recommendations

7.1 Social Media Platforms and Intermediaries

- Emergency reporting mechanisms in place through trusted partners or otherwise for impersonation, account compromise, and NCII/NCUI, with time-bound timelines especially for at-risk groups.
- Safety and reporting tools should be more accessible through regional language support and audio support for differently-abled individuals.
- Anti-amplification safeguards should be put in place to reduce viral spread while credible reports are under review, preventing irreversible reputational harm.
- Investing in multilingual and low-bandwidth resources with visual guides for step-by-step recovery of accounts and content will be most effective for the users.
- Rapid response models based on trust to strengthen informal, relationship-based support pathways and accompaniment as formal mechanisms are often slow or ineffective.

7.2 State and Law Enforcement

- There needs to be emphasis on survivor-centric cybercrime response with standard intake, clear timelines, and safer reporting pathways.
- The law enforcement and investigation agencies, especially NCCIA, need to develop the capacity of their staff, through intense training, to have a survivor-centric response to complaints and handle cybercrime in a gender sensitive manner.
- They also need training on digital evidence preservation without retraumatizing the survivors/victims.
- There needs to be a non-retaliation guarantee in place so the trust deficit can be bridged, and it protects complainants from data leaks and surveillance misuse. And reporting does not become an exhaustion mechanism that deters justice seeking and civic participation.

7.3 International Community

- The international tech community needs to provide emergency access to paid security tools, especially trusted VPNs and account protection services, for frontline defenders.

- Support for long term funding for sustained reporting helplines and networks will result in better insights for advocacy and lobbying for better protections against online harassment and abuse. It should also allow a small fund to be kept aside for emergencies, for trauma and burnout support for survivors and victims.

8. Conclusion

For communities at risk navigating high-stakes digital harm, the Digital Security Helpline serves as an essential, reliable intermediary. While the dedicated digital security subset (97 cases) demonstrates crisis-response practice, the cumulative caseload (5,041 new cases) shows scale: account compromise recovery predominates, with monitoring tools used selectively during disruption events. According to respondents, platform design and a lack of institutional responsiveness exacerbate the worst harms, leading to self-censorship, harm to one's career, and effects on offline safety. Aligning tools and response systems with real-world circumstances, such as cost, device limitations, language/context moderation, and quick, survivor-centered escalation pathways, will be necessary for long-term safety improvements.

DigitalRightsFoundation
"KNOW YOUR RIGHTS"

@DigitalRightsFoundation

@digitalrightsfoundation

@digitalrightsfoundation

@digitalrightsfoundation

Digital Rights Foundation

@digitalrightspk.bsky.social

@DigitalRightsPK

@DigitalRightsPK

digitalrightsfoundation.pk