



DigitalRightsFoundation
"KNOW YOUR RIGHTS"

Digital Battlegrounds: Gendered Disinformation, TFGBV, and Hate Speech in the Indo-Pak Escalations



ABOUT

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.

We believe in the power of a free and secure internet, one that enables expression without fear, protects personal privacy, and resists censorship. At DRF, digital rights are not abstract ideals; they are the foundation for dignity, equality, and justice in the digital age.

Through direct interventions, policy advocacy, and strategic partnerships, DRF delivers digital security trainings, provides emergency assistance, and builds resilience among those most at risk. We work on the frontlines to combat technology-facilitated gender-based violence, promote access to information, and strengthen protections for those under threat, especially in marginalized communities.

As surveillance intensifies and digital spaces become more hostile, DRF confronts both state and corporate actors through research, monitoring, and rights-based advocacy. We challenge repressive policies, expose violations, and propose practical, rights-affirming solutions.

Our goal is clear: to create digital environments where safety, agency, and freedom are not privileges but guarantees for everyone.

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

ACKNOWLEDGMENTS

This short investigative report is released in the aftermath of the India-Pakistan escalations between May 6th and 10th, following the Pahalgam attack. Drawing on real-time data collected during this existing period of heightened tension, the report examines the surge in technology-facilitated gender-based violence (TFGBV), gendered disinformation, and hate speech that accompanied the conflict. It highlights how misogynistic, sexist, and gendered narratives were weaponized online, causing real harm to women and gender minorities on both sides of the border.

The report comes at a tragic moment, as gender was systematically weaponized in digital warfare to fuel hostility and amplify online attacks. This investigation was made possible by the tireless efforts of DRF's research team, who monitored and analyzed digital trends while navigating the unfolding crisis, including firsthand experiences of drone infiltrations and attacks in their cities.

Ultimately, the report seeks to understand how gender becomes a casualty of conflict in the digital space and how online hate can escalate into real-world harm. The report concludes with key recommendations for social media platforms, urging them to take greater responsibility when TFGBV, gendered disinformation, and gendered hate speech are on the rise on their platforms. These recommendations are grounded in international human rights frameworks and best practices for platform accountability and transparency.

Authors: Sara Imran, Research Associate

Seerat Khan, Senior Research Associate

Maria Nazar, Research Associate

Editor: Nighat Dad, Executive Director

INTRODUCTION

On the midnight of May 7, 2025, India launched [Operation Sindoor](#) as a retaliatory response to the [Pahalgam attack](#), in which 26 tourists were killed in the Indian-administered region of Kashmir. The operation targeted [nine sites](#) described by Indian authorities as terrorist infrastructure across Pakistan's major cities and Pakistan-administered Kashmir. The airstrikes reportedly resulted in 31 fatalities and 57 injuries. In retaliation, Pakistan claimed to have shot down [five Indian fighter jets](#) and attacked Indian air bases in a counter-strike, [Operation Bunyan-um-Marsoos](#).

Throughout the operation, social media platforms became critical battlegrounds for public discourse and narrative control by the [state](#). Users from both sides of the border flooded platforms with unverified footage, satellite imagery, and commentary, contributing to an intense and fast-moving information war. Moreover, during this time, public narratives were also restricted via [censorship](#) within both countries, which was at an all-time high, with content from across the border being restricted at both ends to control changing narratives. Hashtags related to the conflict trended globally, further amplifying public sentiment. This digital confrontation increased the physical hostilities between the two countries. Widespread dissemination of misinformation, nationalist propaganda, and competing truth claims, with censorship and geo-blocking of content on the rise in the region, created an information war that the global audience is still trying to understand to date.

Both countries claimed to have carried out these escalations until May 10, 2025, when they agreed to a [ceasefire](#) brokered under significant international diplomatic pressure. However, the situation remains fragile, and security analysts warn that the risk of renewed escalation persists if underlying disputes remain unresolved.

Gendered Disinformation and Digital Misogyny

During this period, online spaces saw a troubling surge in TFGBV, gendered disinformation, and hate speech targeting women and gender minorities in both India and Pakistan. Social media platforms were overflowing with misogynistic content in multiple local languages, alongside AI-generated and doctored visuals.

One particular example involved the circulation of a photo claiming to show a "downed female Indian Rafale pilot" captured in Pakistan. The image, widely shared by pro-Pakistan accounts, was a mis-captioned photo from a 2023 Indian Air Force crash, as verified by [Reuters](#) and independent fact-checkers. Indian authorities, including the Press Information Bureau (PIB), debunked the claim, and the Pakistan military also denied holding any Indian pilot at that time. Despite these clarifications, the narrative triggered a wave of gendered and misogynistic slurs across platforms, with users on both sides weaponizing the false story to further exacerbate gendered hate speech and disinformation during this time of conflict.

Meanwhile, Indian social media accounts circulated stylized images of women as nationalistic symbols. A particularly viral example was an anime-style illustration of [Himanshi Narwal](#), a young widow post the Pahalgam attack, shown sitting beside her martyred husband. The image was used to raise patriotic sentiment, but when [Himanshi](#) later spoke publicly in favor of peace and urged against communal hate, she became the target of online abuse, including slut-shaming and graphic rape threats. This particular example highlights how even those portrayed as symbols of national sacrifice can become victims of abuse when they don't follow mainstream political narratives.

Additionally, AI-generated images were widely shared as supposed "[evidence](#)" of women on the

frontlines. One viral post depicted a group of Indian Army women soldiers allegedly dining near the Pakistan border in desert conditions. Fact-checkers confirmed the image was AI-generated, pointing to visual anomalies such as distorted limbs and missing shadows. Despite being fake, the image fueled nationalistic and misogynistic commentary online. Users drew false [comparisons](#) between the imagined sacrifices of “army women” and derogatory portrayals of women who wear makeup or dance, claiming the latter were undeserving of respect. The viral spread of this narrative, amplified by generative AI, grows to show how TFGBV, gendered disinformation and hate speech can converge into a dangerous tool for reinforcing patriarchal norms and intensifying online harassment.

Humor and Memes Fueling Gendered Violence in a Digital War

Across digital platforms, users from both India and Pakistan engaged in dehumanizing and hyper-sexualized rhetoric. Social media threads were filled with “[war trophy](#)” narratives, where women from the rival country were discussed as sexual conquests. On X (formerly Twitter), Facebook, and WhatsApp, users shared memes and posts joking about which celebrity or public figure would be “taken” after a hypothetical victory. These posts objectified women, framing them as symbolic property of the victorious nation.

Furthermore, during this time, attacks on public figures escalated significantly too, in both India and Pakistan, with attacks shaping themselves in a gendered manner. In [India](#), the National Commission for Women (NCW) strongly condemned the targeted online abuse directed at Foreign Secretary Vikram Misri and his family, particularly his daughter. In a statement, the NCW Chairperson described the circulation of her personal contact information as a “grossly irresponsible act” and a serious breach of privacy that could compromise her safety. The Commission emphasized that such personal attacks on the families of senior public officials are not only unacceptable but also morally indefensible.

Technology-Facilitated Gender-Based Violence in War and Platform Accountability under International Human Rights Law

The India-Pakistan conflict of May 2025 exposed how gendered violence becomes central, not collateral, in digital warfare. Women and gender minorities were deliberately targeted through fake visuals, hyper-sexualized [memes](#), and dehumanizing slurs, turning gender into a strategic tool of online hostility. Across platforms, this content wasn’t just visible, it was viral.

DRF investigations documented a surge in gendered attacks during the conflict and looked at the role of algorithms in amplifying and spreading such content. The findings show that content promoting hate, disinformation, and misogyny was not only tolerated by platform algorithms but was frequently amplified, with no content takedown measures in place.

Social media platforms failed to adequately moderate content in local languages or act on reports of abuse. Terms in Hindi, Urdu, and Punjabi, including deeply misogynistic slurs, largely remained unflagged. Platforms like Meta (Facebook, Instagram), X (Twitter), and YouTube have not adequately addressed the gender-based violence that unfolded during this period, and hyper-gendered graphic content still remains online.

This failure to act stands in stark contrast to the international human rights obligations platforms are under. Under the [United Nations Guiding Principles on Business and Human Rights \(UNGPs\)](#), social media companies are expected to respect human rights and prevent adverse impacts linked to their operations. This includes taking action against online abuse that constitutes a

violation of women's and minorities' rights to dignity, safety, and participation in public discourse.

Reports by [UN Women](#), the [UN Special Rapporteur on freedom of expression](#), and digital rights organizations such as [Access Now](#) and [Ranking Digital Rights](#) have consistently emphasized the responsibility of technology companies to act preemptively and remedially in high-risk conflict zones. In the case of India and Pakistan, this obligation becomes especially urgent given the deeply rooted patriarchal norms and systemic barriers that disproportionately impact women and gender minorities in the region. In this instance, the role of platforms in mitigating harm becomes even more critical than ever before.

DRF'S METHODOLOGY AND ANALYSIS

Methodology

DRF analysed 295 unique posts covering the scope of the Indo-Pak escalation by social media users across 5 social media platforms, X (formerly Twitter), Facebook, Instagram, TikTok, and YouTube. An alarming 25% of these posts (73 out of the total 295) were directly connected to gendered disinformation, TFGBV, and gendered hate speech that victimised Pakistani and Indian women and other gender minorities. The data collection period was from 7 to 20 May, which covers the period from the inception of Operation Sindoor, up to recent events following the [May 10 ceasefire](#).

Limitations

While every effort was made to ensure the accuracy and reliability of this report, a key limitation is that the majority of data collection took place during the escalation period, when active military attacks were being carried out on both sides of the border. However, TFGBV, along with gendered disinformation and hate speech, continued well beyond the end of escalations and the announcement of the ceasefire agreement.

Findings and Analysis

DRF organised the collected data on gendered disinformation, TFGBV, and gendered hate speech into five key content categories, identified through prominent data clusters around a few recurring topics.

These five categories include:

- 1) Explicit threats and jokes pertaining to rape or sexual harassment, exchanged between Indian and Pakistani users in the context of imagined city invasions or takeovers;
- 2) Narratives invoking sexual violence as a symbol of conquest, exemplified by references such as Operation "Suhaag Raat"/"Honeymoon";
- 3) "Maal-e-Ghanimat" (spoils of war) and "laundiyan" (enslaved women) tropes, frequently couched in dark humor and used to justify or normalise gendered violence, at times amplified by public figures;
- 4) The targeted harassment of specific individuals, including Indian Air Force pilot Shivangi Singh and YouTuber Jyoti Malhotra, even by users from their own countries; and
- 5) Transphobic jokes linking the 'effects' of nuclear radiation to gender nonconformity.

1. Rape threats and misogynistic humor in the context of city invasions

As the escalations heated up, a recurring theme in the war 'humor' that permeated both sides of the border was borderline rape threats and misogynistic jokes that made women's bodies the punchline.

A user [posted](#) a film scene depicting the rape of a woman, labelling the rapist as "Indian Drones" and the woman as "Pakistan". Posts like these, shared and amplified by multiple users, highlight how dominance, even in the context of war, is expressed through symbolic violence against women. Such suggestive memes trivialize sexual violence and normalize graphic threats, revealing how gendered hate speech and TFGBV are increasingly masked as humor or satire in digital spaces.



Sexist language and AI-generated gendered hate speech were used to [criticize](#) Pakistani military officials and to [accuse](#) Pakistan of colluding with China.



This latter trend in particular reflects how gendered tropes are weaponized even in ostensibly non-gendered political conflicts. By feminising or sexualising political figures and national institutions, online bad actors use women's bodies and identities as symbolic battlegrounds for nationalist and geopolitical narratives.



Posts referencing cross-border relationships also became targets of online criticism, with users branding Indian men involved with Pakistani women as "traitors" or disloyal to their nation. Similarly, images from the popular Indian Netflix series Heeramandi, set in a historic red-light district in Lahore, were weaponized online. Users shared pictures of women depicted as courtesans, using them to label Pakistani women calling for something as benign as de-escalation as "prostitutes," and men supporting peace as their "customers."

These deeply gendered and sexist attacks reflect a growing trend in which cross-border relationships are demonized, fueling hate speech and disinformation. Such narratives reduce calls for peace and dialogue to matters of moral character, reinforcing misogyny and hostility through nationalistic and patriarchal lenses.

2. Narratives Invoking Sexual Violence as a form of conquest: Operation "Suhaag Raat"/"Honeymoon"

One of the earliest trends of TFGBV was observed as a 'humorous' play on the name of the Indian military operation, Operation Sindoor. Sindoor refers to the vermilion that Indian women traditionally wear on their foreheads after marriage.

Immediately following the Indian attacks of 7 May, a trending hashtag of [#OperationSuhaagRaat](#) was observed, with numerous Pakistani social media users [calling](#) for Pakistan to retaliate with an operation named after "Suhaag Raat", which translates to consummation night.



The use of the term “Operation Suhaag Raat” (or “[Operation Honeymoon](#)”, in other framings) in [response](#) to an already gendered military operation name reflects how rape culture is embedded in digital discourse, where the language of sexual violence is co-opted to express nationalist aggression.



This framing, even if meant to be ‘humorous’, not only trivialises sexual harm but also reinforces TFGBV by framing the violation of women’s bodies as a symbol of victory. In South Asia, widowed women are already subject to deep-rooted stigma, often seen as inauspicious or dishonorable. Posts that celebrate ‘wiping off sindoor’ (a traditional symbol of marital status) as an act of conquest reflect a disturbing pride in turning women into widows, reinforcing harmful gender norms and violence against women. These narratives deepen the existing discrimination and marginalization faced by widowed and unmarried women in the region, weaponizing gendered trauma in the name of nationalist sentiment.

3. “Maal-e-Ghanimat” (spoils of war) and “Laundiyan” (slaves) narratives

This particular example of TFGBV was mainly spread by Pakistani social media users and may be seen to have gained license via Pakistani public figures.

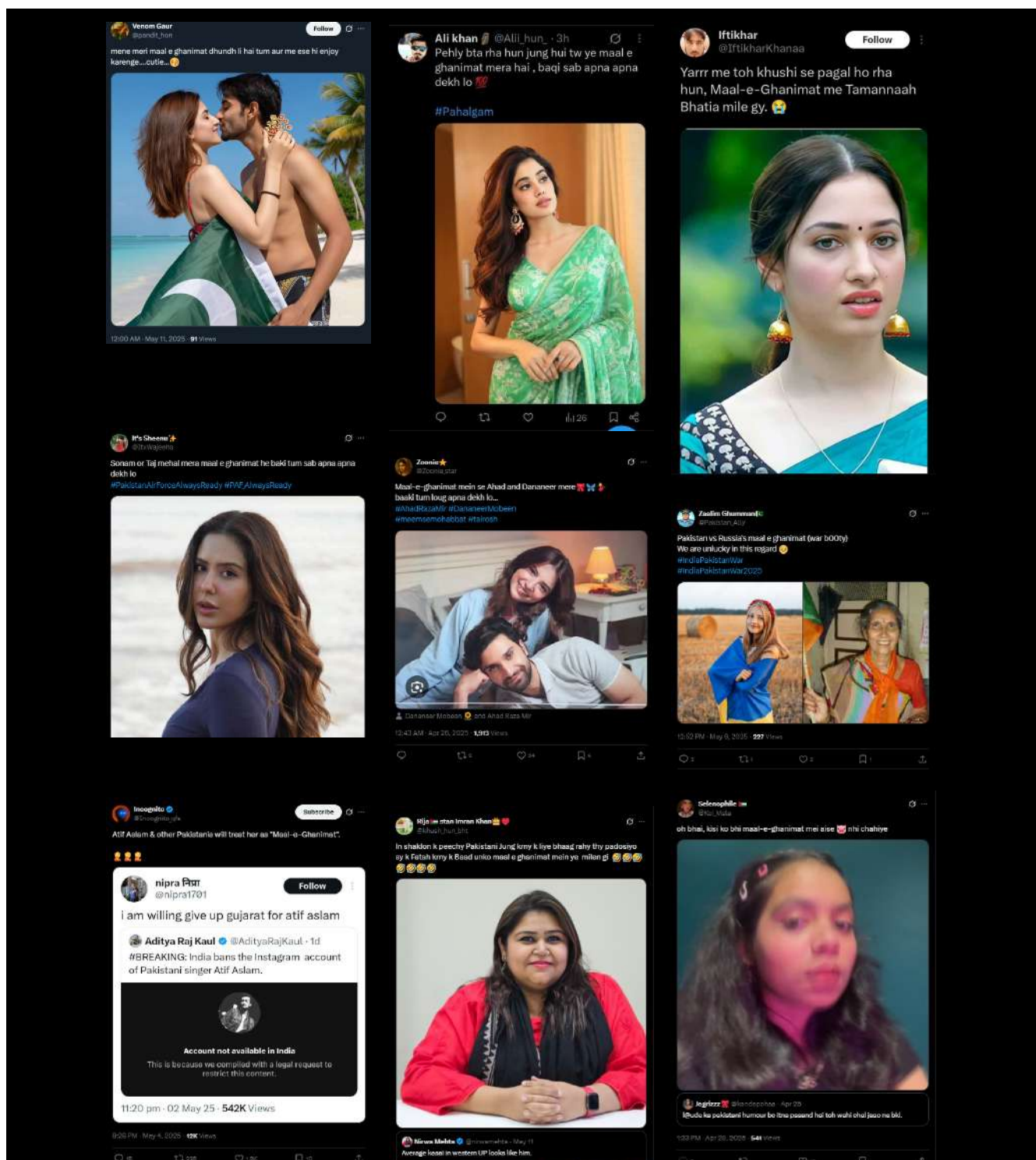
“Maal-e-Ghanimat,” an Urdu term meaning spoils of war, was frequently used in this discourse, while “laundiyan”, which [literally translates](#) to slave-girls or bondmaids, was used interchangeably with the term sex slaves. Users frequently used both terms in tandem to evoke imagery of ancient warfare, objectifying women as “booty” to be claimed following a hypothetical conquest of land.

The misinformation claim about an Indian woman pilot, Shivangi Singh, having been captured by the Pakistan Air Force, was also [portrayed](#) in this light, by terming Singh as an example of “Maal-e-Ghanimat” following her alleged capture.

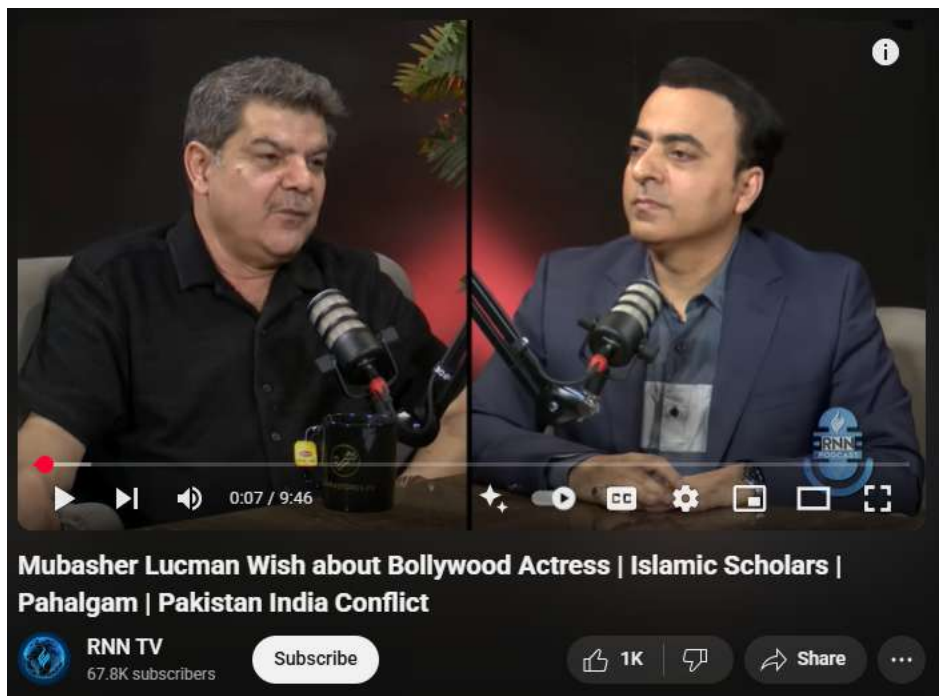


Another TFGBV trend promoting objectification of Indian women, and rape culture, was Pakistani social media users posting a number of popular Indian women celebrities, and ‘[picking](#)’ their “Maal-e-Ghanimat” based off their attraction to that celebrity. There were also some cases noted of Indian users doing the same with Pakistani celebrities, as well as Indian and Pakistani women social media users being targeted.

On both sides of the border, the trope of ‘Maal-e-Ghanimat’ kept resurfacing in disturbing social media commentary, with users comparing the perceived attractiveness of women from India and Pakistan. Many posts reflected a misogynistic mindset, expressing a preference for “more attractive” women as hypothetical opponents, reducing women to objects of conquest rather than individuals, further reinforcing the gendered nature of digital abuse during conflict.

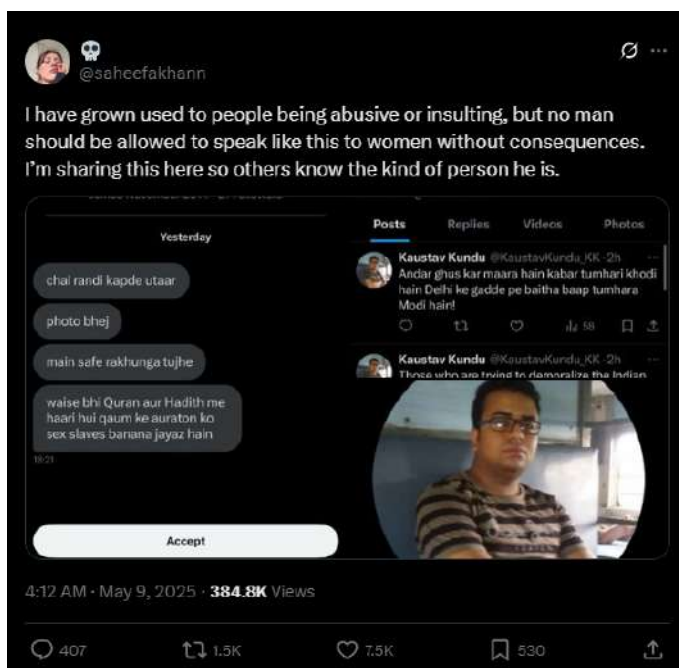


On 6 May, prior to the Indo-Pak escalations, the Pakistani media organisation Rasta News Network (RNN) posted a podcast episode on their YouTube channel featuring two prominent Pakistani journalists, Naeem Hanif and Mubasher Lucman. In the episode, Lucman [states](#) that if war breaks out with India, he will want to take Indian women actors as “Maal-e-Ghanimat” (spoils of war), referring to them as “Laundiyan”, or slaves, with an implied sexual connotation.

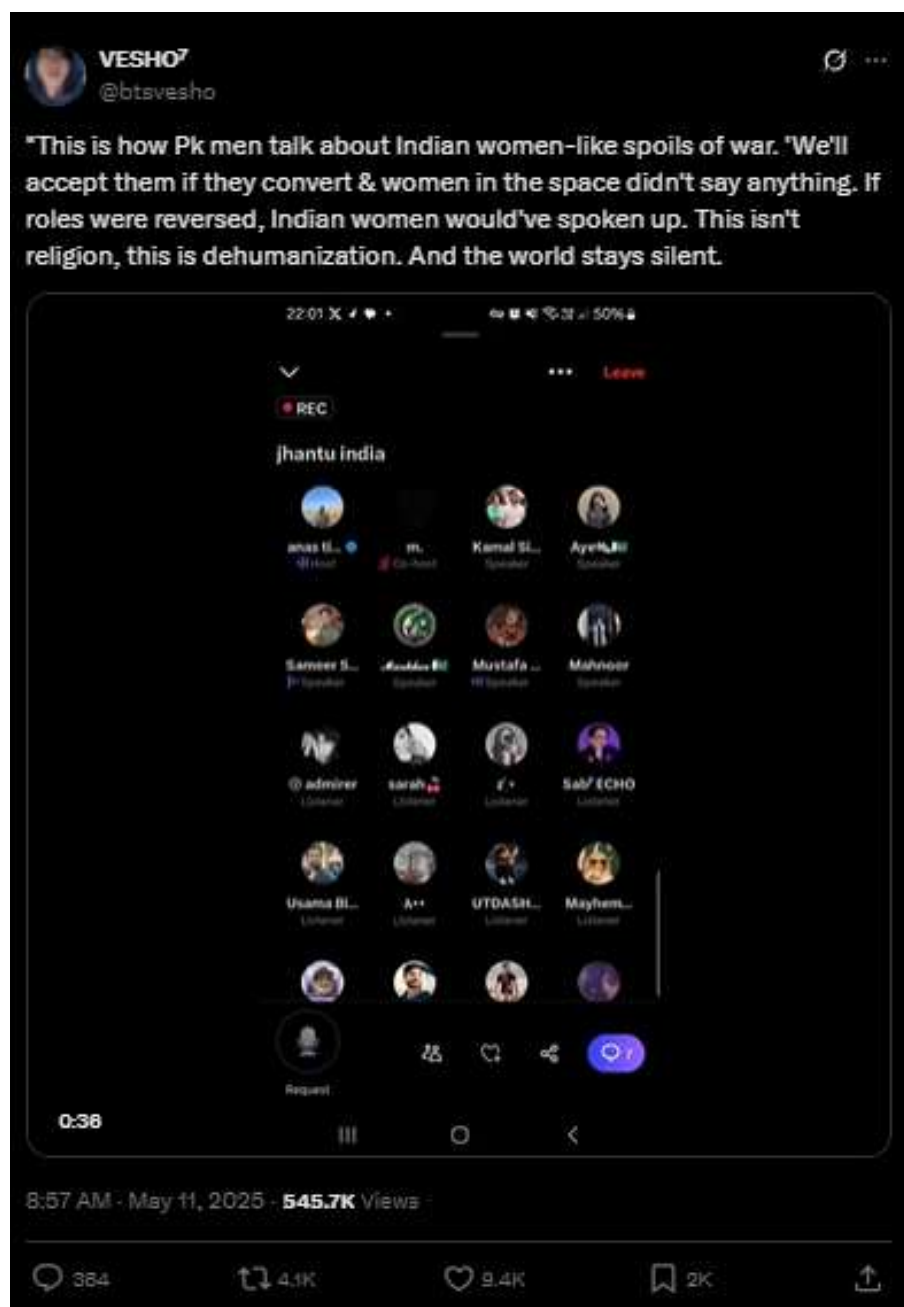


Such remarks, even when framed as ‘humor’ by public figures, as Lucman certainly implied, help normalise misogynistic and violent rhetoric, creating a permissive environment where similar language and threats are directed at real women with real-world consequences.

One such real-world consequence was suffered by a woman online during the escalations, when an Indian X user allegedly sent a Pakistani woman threatening messages online, according to screenshots [posted](#) by the Pakistani user. The alleged perpetrator referred to the concept of sex slaves to sexually harass the victim.



This same real-world consequence was multiplied manifold as [reports](#) came out of an X (formerly Twitter) space in which Pakistani X users allegedly discussed treating Indian women as spoils of war, and forcibly converting them to Islam.



The casual use of militarised and misogynistic language, especially by influential figures, not only reflects deeply embedded patriarchal and nationalist ideologies, but also actively enables a culture where TFGVB is trivialised and reproduced. Terms such as “Maal-e-Ghanimat” and “Laundiyan”, often couched in humor, serve to legitimize gendered violence and dehumanisation, with tangible impacts on women’s safety both online and offline.

4. Cross-border TFGBV case studies with a focus on how AI-driven systems reinforce gendered violence and platform-enabled misogyny

DRF identified two distinct cases of targeted TFGBV involving Indian women. Notably, in one case, the perpetrators were Pakistani users, while in the other, the abuse originated from Indian users.

Shivangi Singh

The first case of TFGBV came shortly after Pakistan launched its Operation Bunyan-um-Marsoos, in the wake of which Pakistani social media users spread a major misinformation claim that an Indian woman pilot, Shivangi Singh, had been taken into custody by the Pakistan Air Force.

These claims were soon accompanied by misogynistic humor, referring to the Indian pilot as Pakistan's "[bhabhi](#)" (sister-in-law), and terming her "[Maal-e-Ghanimat](#)" (spoil of war) while making crass sexist commentary about her appearance.



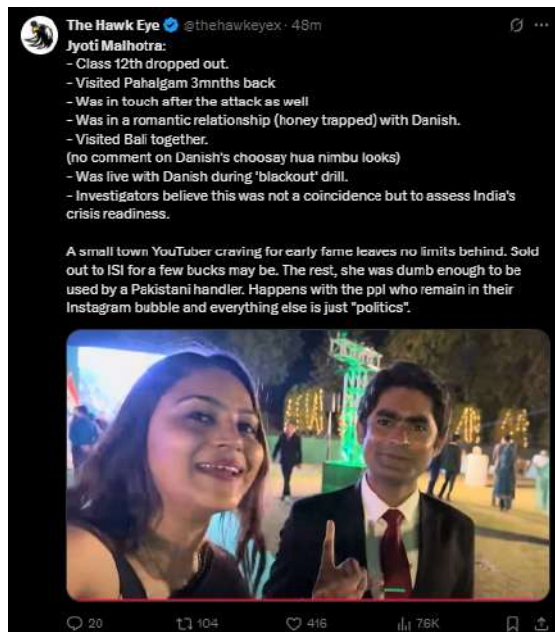
In this thread, DRF observed an AI-generated fabricated viral video, which depicted the Indian pilot crying in front of Pakistan's Chief of Army Staff General Asim Munir, supposedly after having been taken into custody. The video is meant to make 'light' of the situation by depicting the Indian pilot asking the general for Sting, a popular Pakistani beverage, amid her tears. Beyond its inaccuracy, the video reinforces harmful gender stereotypes by portraying the woman as overly emotional and fragile, even in a military role, in contrast to the trope of brave and strong male soldiers. This not only trivializes conflict but also weaponizes gender norms to undermine the presence and credibility of women in uniform.



Jyoti Malhotra

The second case study DRF analyzed was that of Jyoti Malhotra, an Indian YouTuber from Haryana who has recently been [arrested](#) in India for allegedly spying for Pakistan, after allegedly maintaining contact with a Pakistan High Commission official and making multiple trips to the country.

It was observed that many Indian users on X (formerly Twitter) took the opportunity to hurl hate speech and misogynistic allegations on Malhotra, which included [accusing](#) her of being in a [romantic relationship](#) with the Pakistan High Commission official, insulting the official's appearance, and terming her a "small town YouTuber craving for early fame leav[ing] no limits behind."



Some accounts went as far as to [claim](#) that Malhotra had had a child with the Pakistan High Commission official, and that the child was "living in Pakistan with [his] first wife". These fabricated claims intensified the online attacks on Jyoti, with users targeting her perceived promiscuity and questioning her "honor" for allegedly having a child with the so-called enemy, reflecting how nationalist narratives are often weaponized through deeply gendered and moralistic shaming.



It was also noticed that there was an attempt to objectify Malhotra, with certain users [posting](#) scantily clad images of Indian actors to speculate which actor would “play” Malhotra in a possible film made on current events.

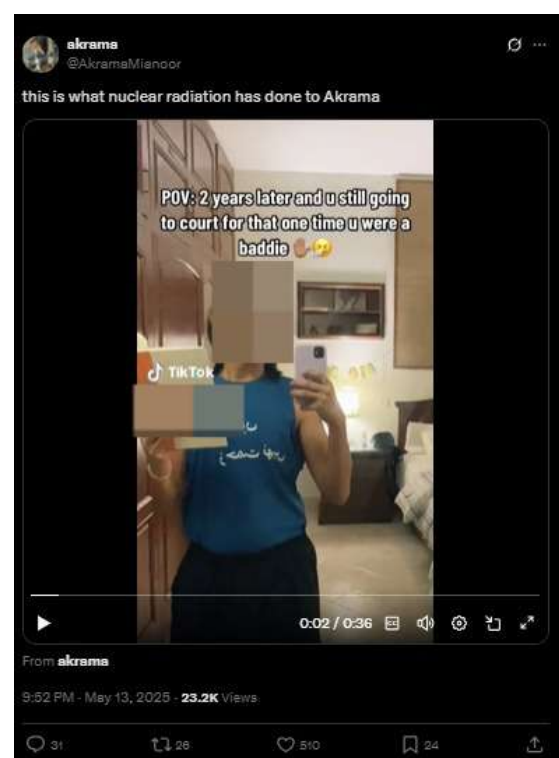


The cases of Shivangi Singh and Jyoti Malhotra illustrate how TFGBV can transcend national borders, with women often becoming the first targets of gendered online abuse, the causes of which are driven by geopolitical tensions and nationalist sentiments, which have ostensibly nothing to do with the targeted women.

5. Gender-based hate speech and dehumanisation stemming from nuclear radiation memes

After the ceasefire agreement, Indian social media users began to circulate a new piece of disinformation, claims of nuclear radiation spreading in Pakistan due to one of the Indian missile strikes having damaged a nuclear power plant. As had been seen often during the escalation, Pakistani social media users responded to this claim with humor, posting depictions of people acting in bizarre and comical ways, and claiming that these were the ‘effects’ of nuclear radiation.

However, even this ‘nuclear’ humor took a turn for the worse when the jokes devolved into gender-based hate speech, transphobia, and dehumanisation of gender minorities in Pakistan.



The posts utilised the nuclear radiation ‘effects’ punchline to target specific, well-known transgender Pakistanis.

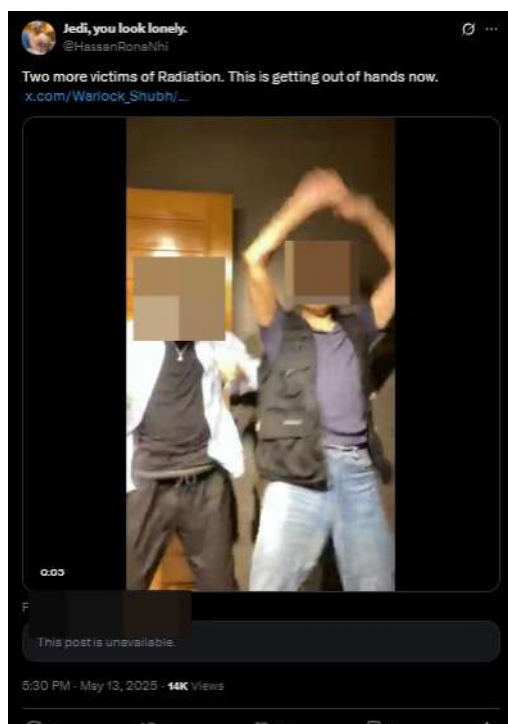
One post was a picture of two transgender activists, claiming that the nuclear radiation had “caused” them to “switch their gender”. This was shared by multiple users, reinforcing the narrative that transgender individuals are an abomination or unnatural. In the context of Pakistan, where transgender individuals are already persecuted and face systematic discrimination, content like this further fuels smear campaigns.



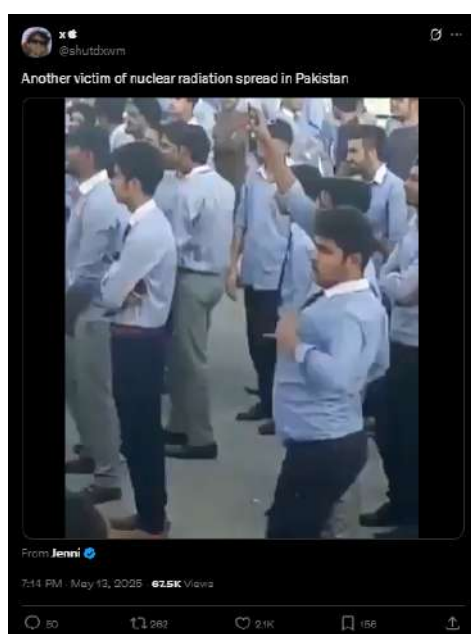
Another user posted a video of a transgender individual, using the nuclear radiation meme format to refer to them as a “mutated cockroach” who had turned into a human demanding “homosexual rights”. These posts were especially alarming, not only because they featured the faces of real transgender activists without consent, but also because they falsely implied that these individuals were advocating for ‘homosexual rights’ in Pakistan, where same-sex relationships are criminalized. Such content not only dehumanizes transgender people but also puts them at serious risk by misrepresenting their advocacy in a hostile legal and social environment.



The same user posted other videos of prominent Pakistani transgender individuals, calling them “victims of Radiation”.



The trend seemed to spread outward from solely targeting prominent Pakistani gender minorities to spreading hate speech about seemingly random Pakistanis acting outside of rigidly stereotypical gender roles. A post containing a video of a uniformed student dancing claimed that the student had become “another victim of nuclear radiation spread in Pakistan.”



This trend reveals how seemingly humorous meme formats can serve as vehicles for transphobia and gendered hate speech, reinforcing harmful stereotypes already dominant in Pakistani society, and dehumanising marginalised identities under the guise of satire. DRF, in its [case study](#) on gendered disinformation, has highlighted in the past that these narratives further intensify the smear campaigns against the community ongoing since 2018, and endanger the community in online as well as offline spaces.

THE GAP BETWEEN PLATFORM POLICIES AND THEIR IMPLEMENTATION

Despite platform policies that ostensibly aim to curb gendered violence, sexual abuse, hate speech, misinformation, and harm, DRF's analysis found a significant gap between the guidelines these companies publish and their actual implementation, particularly when it comes to content involving TFGBV, gendered disinformation, and gendered hate speech.

X [claims](#) to have a comprehensive set of policies that “promote and protect the public conversation,” citing posts violating their Terms of Service made during the “Israel-Hamas conflict” as an example of implementation. The enforcement of these policies includes:

- Restricting the reach of a post
- Removing the post
- Account suspension
- Ineligibility of such posts for monetisation
- Community notes

Despite these stated tools, out of all the posts analysed by DRF in this particular instance, barely any faced any of the above-mentioned repercussions. Some accounts posting the harmful content were suspended or deleted, but it remains in question what the reason for these suspensions/deletions was. This is notable given that X's rules also [state](#): “You may not attack other people on the basis of race, ethnicity, caste, sexual orientation, gender, gender identity, religious affiliation... disability, or serious disease,” and prohibit “malicious, unreciprocated targeting,” “denial of violent events,” and “unwanted sexual discussion of someone's body”. However, as noted in the section above, several posts attacking gender identities and women's bodies remain up, till date.

Meta's [Bullying and Harassment policy](#) seeks to “remove content that's meant to degrade or shame,” including content that targets people based on “experience of sexual assault, domestic abuse,” or involves “claims that individuals are lying about being a victim of a violent tragedy.” Yet much of the gendered disinformation and harassment, such as targeted attacks and doxxing of individual Pakistani and Indian women, went unremoved and unreduced in reach.

In YouTube's case, it was noticed that channels with subscribers in the millions posted content featuring derogatory remarks about women, such as the ones by Mubasher Lucman documented above. There was no visible implementation of the strike system, and the harmful content remains on the platform to date. This is despite YouTube's [stated](#) ban on content that uses “extreme insults to dehumanize,” expresses “a wish for death or serious injury,” or “threatens someone's physical safety.”

TikTok, similarly, [outlines](#) strict rules in its community guidelines against targeted abuse and hate speech. According to its policy:

“We do not allow content that contains slurs, hate symbols, or hate speech. This includes mocking victims of hate crimes or terrorism, denying the occurrence of such events, or portraying them positively.” TikTok also prohibits *“Content that praises, promotes, or denies any violent tragedy”* and *“Content that includes dehumanizing, slur, or hateful speech towards a specific group.”*

Despite these stated positions, DRF observed that many TikToks containing gendered hate speech, violence against women, or targeted disinformation against women remained accessible and continued to be widely circulated. None of the videos carried content warnings, and none

were labelled for being either disinformation or generative AI or taken down at the time of reporting.

Across platforms, a clear gap remains between the protections claimed by these companies and their actual implementation. While all five major platforms have publicly committed to removing or demoting harmful and misleading content, especially that which could cause real-world harm or target vulnerable individuals, such policies appear to exist more in theory than in practice when it comes to gendered disinformation and TFGBV in online spaces.

RECOMMENDATIONS

Guided by the UN Guiding Principles on Business and Human Rights (UNGPs), recommendations by UN Women, the UN Special Rapporteur on Freedom of Expression, and digital rights civil society groups, the following are actionable steps platforms must adopt to curb TFGBV, gendered disinformation, and hate speech during high-risk conflict scenarios:

1. Explicitly Restrict TFGBV and Gendered Hate in Policies

Social media platforms must explicitly recognize TFGBV as restricted content. Online abuse, including sexualized deepfakes, should be recognized as distinct from nude or semi-nude image categories with rape threats, non-consensual sharing of personal information (doxxing), and coordinated harassment campaigns must be restricted under platform community guidelines.

This aligns with the [UN Special Rapporteur's](#) recommendation to treat gendered disinformation and abuse as violations of international freedom of expression standards when they aim to silence or intimidate women.

Platforms must ensure prompt takedown of gendered hate speech and doctored images, along with meaningful enforcement actions (e.g., account suspensions, law enforcement referrals in the case of credible threats). During conflict periods, trust and safety teams must proactively monitor trending topics to swiftly remove hashtags or content that glorify sexual violence or incite harm against women.

Moreover, while platform policies often state that public figures are subject to a higher standard under bullying and harassment guidelines, to allow for public discourse and critical commentary, this should not be misused to justify the sexualization or degradation of women in the public eye, including actors, journalists, and politicians. The current implementation of these policies requires urgent revision to distinguish between legitimate critique and gendered abuse, ensuring that public visibility does not become a license for targeted misogyny.

2. Conduct Human Rights Due Diligence with a Gender Lens

In line with the UNGPs, companies must regularly conduct human rights and gender impact assessments to identify and mitigate systemic risks to women and gender-diverse users. These assessments must evaluate how platform algorithms, content promotion systems, and business practices might contribute to gendered harm, particularly in volatile contexts such as India and Pakistan in May 2025.

Preventive measures can include:

- Gender-responsive product design (e.g., privacy by default, content sharing limits)

- Context-aware adjustments to algorithmic virality, especially for unverified or inflammatory content
- Regular consultation with local gender and human rights experts to identify abuse patterns in regional languages such as Hindi, Urdu, and Punjabi.

As [DRF's 2024 Digital Security Helpline Report](#) shows, online harms are highly contextual and intersectional; hence, standardized globalized approaches often fall short of addressing region-specific abuse.

3. Strengthen Local-Language Moderation and Human-in-the-Loop Reviews

Content moderation must prioritize regional languages and culturally informed human insight. During the 2025 India-Pakistan conflict, a significant portion of gendered hate and disinformation circulated in Hindi, Urdu, Punjabi, and other regional dialects. These posts often relied on historical and cultural references that, while seemingly non-violent on the surface, were contextually harmful and violative. However, due to the limitations of English-centric AI moderation systems, such content frequently escaped detection, highlighting the urgent need for localized, context-aware moderation frameworks.

Platforms should:

- Hire and train more regionally knowledgeable moderators, including women and marginalized groups;
- Pair these moderators with AI systems in a “[human-in-the-loop](#)” model for more accurate review of coded or context-specific abuse;
- Improve machine learning models for South Asian languages, while retaining human oversight to avoid false positives and cultural misreadings.

Trusted civil society groups should be integrated into rapid escalation channels to flag viral abuse (e.g., coordinated rape threats or deepfake content) for immediate review and takedown.

4. Deploy Gender-Sensitive Safety Tools and Reporting Mechanisms

Platforms must invest in user-facing tools that center survivor safety, especially during digital violence surges. This includes:

- Easy-to-use, trauma-informed reporting mechanisms with specific categories for TFGBV, available in regional languages
- Features like “Safety Mode” that temporarily auto-block abusive accounts;
- Content filters to screen graphic or misogynistic material, particularly AI-generated images (deepfakes);
- Integration with support systems like DRF’s Digital Security Helpline, which offers users a pathway to assistance.

These safety features require timely action by moderators and provide victims with updates on outcomes, especially in cases of reported threats to life or sexual violence.

5. Coordinate with Civil Society for Real-Time Conflict Monitoring

Platforms should institutionalize partnerships with local civil society organizations, journalists, and digital rights groups, especially during political crises or conflict.

To do this, platforms must:

- Establishing a dedicated crisis advisory group for India and Pakistan, involving trusted partners and regional experts
- Opening secure, real-time reporting channels for trusted partners to escalate harmful gendered content is essential, particularly by improving escalation forms and reporting portals. These improvements should reflect previous feedback from trusted partners and account for the accessibility needs of users in the region, including language, digital literacy, and ease of navigation.
- Holding weekly or bi-weekly calls during active conflicts to share trends and policy feedback.

6. Monetization of GBV Content

Crucially, platforms must also address the monetization of GBV content. During the 2025 India-Pakistan conflict, DRF observed coordinated campaigns where inflammatory, misogynistic, or AI-fabricated content targeting women was not only amplified but also monetized via ad revenue, paid engagements, or creator payouts. This creates a dangerous incentive structure where bad actors profit from virality rooted in harm.

Platforms must de-incentivize the production and spread of gendered disinformation by:

- Demonetizing content that targets individuals based on gender or sexual identity;
- Suspending monetization privileges for repeat offenders;
- Removing algorithmic boosts for engagement-driven GBV content;
- Publicly reporting data on demonetized harmful content during crisis periods.

This will ensure that platforms align with the **UN Guiding Principles on Business and Human Rights** and uphold their duty to prevent adverse human rights impacts while ensuring that crisis response mechanisms are regionally grounded, survivor-centered, and structurally deterrent to abuse-for-profit models.

7. Ensure Transparency, Accountability, and Appeals

DRF urges platforms to publish detailed transparency reports that track and disclose data specifically related to TFGBV and gendered disinformation.

These reports should include:

- Number of takedowns involving gender-based abuse
- Language and regional breakdowns of violations
- Average response times and appeal outcomes.

Additionally, platforms must offer robust, human-reviewed appeal processes with cultural competence to prevent unjust removals or dismissals of legitimate reports.

Platforms should refer decisions to independent oversight bodies for high-stakes cases, particularly those involving journalists or public-interest speech. The Oversight Board is one example of this, when it [reversed Meta's decision](#) to leave up a Facebook post that mocked a victim of gender-based violence. The post featured a photo of a woman with visible injuries and a caption implying she deserved the abuse due to a typographical error. The Board determined that the post violated Meta's Bullying and Harassment policy and highlighted a gap in Meta's policies regarding content that normalizes gender-based violence. The Board recommended that Meta develop policies to address content that praises, justifies, celebrates, or mocks gender-based violence.

Civil society organizations must be able to request audits of platform enforcement, especially to assess bias, systemic silencing of women, or failures to act on credible threats.

8. Prioritize Protection of Women in Crisis Response Protocols

Women and marginalized users must receive heightened, proactive protection during emergencies. Crisis response protocols should include:

- Temporary enhanced security for verified accounts of women journalists, activists, and public figures
- Priority review and takedown of high-risk content, including sexualized or humiliating deepfakes
- Use of detection algorithms to monitor for AI-generated abuse

CONCLUSION

The gendered abuse and disinformation that emerged during the 2025 India-Pakistan conflict is a stark example of how platform inaction can exacerbate real-world harm. As outlined above, social media companies possess both the capacity and the responsibility to take decisive, rights-respecting measures before, during, and after periods of crisis.

DRF's analysis and recommendations are grounded in the lived experiences of women and marginalized communities in South Asia during the conflict and are aligned with the minimum global standards set by the UN bodies, civil society organizations, and survivors themselves. Addressing crisis-related technology-facilitated gender-based violence, hate speech, and gendered disinformation is not optional, it is essential for human rights and digital rights.

Platforms must respond with urgency, transparency, and accountability to ensure that digital spaces are not weaponized as tools of modern information warfare. Failing to act allows bad actors to exploit information flows, turning online abuse into another front in the conflict.



Digital**Rights**Foundation
"KNOW YOUR RIGHTS"



@digitalrightsfoundation



@DigitalRightsFoundation



@DigitalRightsPK



@digitalrightsfoundation



@digitalrightsfoundation



@DigitalRightsPK