DigitalRightsFoundation
"KNOW YOUR RIGHTS"

Combatting Flood Misinformation in Pakistan:

Generative AI and Platform Accountability in the Age of Climate Crisis

Acknowledgement

This investigative analysis comes at a time when Pakistan faces a climate catastrophe that has left millions homeless and claimed thousands of lives. DRF's findings highlight how disaster-related misinformation amplifies the hardships faced by affected communities, particularly marginalized groups. The report also highlights the growing role of generative AI in spreading false information, with bad actors exploiting crises to destabilize an already fragile environment.

In such moments, online platforms have a critical responsibility to curb the spread of misinformation, while internet users must remain vigilant in identifying false or AI-generated content online. Without this shared responsibility, flood-related misinformation risks amplifying chaos and deepening the suffering of those already struggling to survive.

Authors: Seerat Khan (Research Lead), Hassan Khan (Research Associate), Sara Imran (Research Associate) and Maria Nazar (Research Associate)
Edited by: Nighat Dad (Executive Director) and Seerat Khan (Research Lead)
Design and Layout by: Ahsan Zahid (Senior Graphic Designer) and Talha Umar (Graphic Designer)

Introduction

Over the last few years, climate change induced flash floods and cloudbursts have completely overwhelmed the fabric of Pakistani society. Despite these natural disasters happening every year, the Federal and Provincial governments have been inadequate in addressing the governance gaps in disaster management, forecasting, and warning systems, leaving citizens across the country in a state of anxiety and sheer helplessness. This year, monsoon floods have badly impacted all major provinces of the country, resulting in over 300 reported casualties in Khyber Pakhtunkhwa, and forcing an upward of 4 million people to flee their homes.

As citizens scatter for real-time weather updates, relief efforts, and the extent of damage caused in nearby communities, social media platforms act as substitutes for the perceived inadequacies of state entities responsible for disaster management. However, the pitfalls of using these platforms as a source of immediate information and guidance have become quite apparent. Especially during times of crisis and natural disasters, over the course of months, there has been ample evidence proving a surge in misinformation intended to garner views through the sensationalization of catastrophic tragedies and by spreading confusion.

With the unregulated proliferation of Generative AI content on social media platforms, misinformation has become even more elusive, as platforms have been found not acting with enough urgency to flag posts that deploy AI-generated footage in local languages. As we highlight in this analysis, misinformation is also often powered by employing local, culturally sensitive, and politically charged discourses within the broader crisis- allowing for the crisis to be used as a conduit for advancing polarized debates that cloud the truth under layers of propaganda and fabrications.

In this devastating and urgent time for relief, social media platforms also have a responsibility to recognize their role in regulating content during periods of crisis. The ease of access in content creation, coupled with profit-centric algorithmic patterns, incentivizes users to produce disorienting and oftentimes harrowing content designed to tap into audience sensitivities for views. Structural constraints on authentic news have now become quite pressing considering that citizens are relying on platforms for guidance during moments of unmitigated disasters. Even when such responsibilities are recognized- as in the case of TikTok, which introduced warning signals for Pakistani's searching for flood updates- DRF's observations indicate that the surge in misinformation has remained largely unchecked.

Background: Floods, Misinformation & Digital Vulnerability

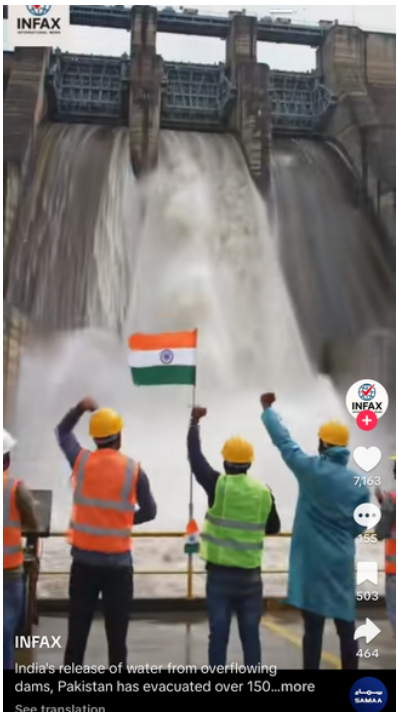
Pakistan finds itself at the crossroads of a climate disaster driven by glacial outbursts, unprecedented heavy monsoon rains, and floodings caused by major swellings in the country's largest rivers. Floods in 2022 caused immense human and economic losses, submerging almost 1/3rd of the country, destroying livestock and crops, while overwhelming hundreds of villages for months. According to World Bank estimates, economic losses and overall damage amounted to over $15 billion, affected 33 million people, displaced over 8 million, and resulted in almost 2000 documented casualties. The Human Impact Assessment further highlighted the compounding economic hardships imminent with the major disruptions caused to rural livelihoods, and rightly predicted a surge in both absolute and multidimensional poverty rates.

Now, in 2025, floods have ravaged the country yet again. Over the last month, multiple catastrophic flash floods have been reported in Khyber Pakhtunkhwa (KP) and Punjab, leading to over 700 official casualties, scores of inundated villages, and nearly half a million people displaced in Punjab alone. In KP's Buner district, monsoon rains led to disastrous landslides killing over 200 individuals. Punjab, which is at present witnessing the worst of these floods, has called a state of emergency in various districts owing to extremely high flood alerts. While the extent of damage in Punjab cannot yet be quantified, the magnitude of the disaster has heightened public anxieties across the country.

The immediacy of such a disaster usually prompts affected people to seek real-time communication updates. There is often little time for preparation, and for those looking to evacuate, updates from social media can play a decisive role in shaping potentially life-altering decisions. In our observations, however, flood-related misinformation has proliferated in Pakistan, blurring factual updates and complicating disaster management efforts. Such tasks become especially difficult with the onset of generative AI content, which has become increasingly difficult to properly identify. Lastly, fabricated narratives appended onto "flood updates" feed off of sensationalizing tragedies, crowding out essential updates and authentic reports. These often spill over into polarizing discussions as well, such as narratives singularly attributing the blame for floods to India's decision to release dam water downstream in Pakistani Punjab.

These concerns are further compounded when we factor in the low levels of digital literacy in Pakistan. According to recent research conducted on the interaction between digital literacy and truth discernment among Pakistan's social media users, authors find a highly positive statistical correlation between educational attainment, digital literacy, and the ability to identify false information. Considering that 40% of Pakistan's population is illiterate, it is critical that initiatives expanding internet and broadband access are integrated within broader literacy programs. While mobile subscriptions have seen a steady uptick, with better internet penetration and mobile access, this access has not been complemented by targeted literacy programs. In rural areas where literacy rates are generally quite low, users may not be able to discern between different forms of content, making them more susceptible to misinformation.

Examples of Flood-Related Misinformation Weaponizing AI in Climate Crises



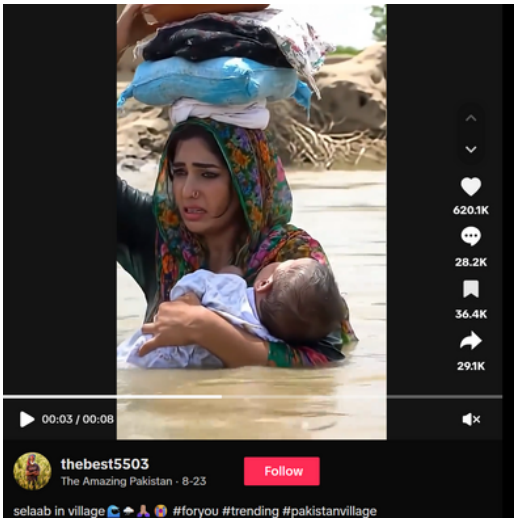
During DRF's initial investigation of how flood-related updates and information were being shared on social media platforms, we observed an alarming number of easily identifiable AI-generated clips that were deliberately weaponizing and sensationalizing tragedy to garner engagement. The following clips were posted during the Punjab Floods, and are purposefully fuelling anxieties across the province over Pakistan's potential vulnerability to hostile decisions made by India regarding water management. As shown in the two examples below, India's decision to release dam water to overflowing rivers running downstream into Pakistani Punjab is portrayed within a larger context of growing antagonism

between the two countries. These tensions are further dramatized using AI-generated footage, showing Indian workers maliciously releasing water downstream as a tactic of war. Attaching a singular geopolitical narrative onto a natural catastrophe not only fuels existing polarizations, but more importantly allows domestic gaps in governance, flood management and relief to slide unscathed from critique. When these narratives are corroborated through Generative AI content rooted in loose narratives that ignore the complex realities of water management, tailored content moderation strategies and effective enforcement become urgent platform responsibilities. Regarding the issue of enforcement, comparing these two clips reveals another dimension of platform inefficiencies. Both clips are evidently AI-generated, but only the latter has been aptly tagged as such.

Religious Framing of Flood Misinformation

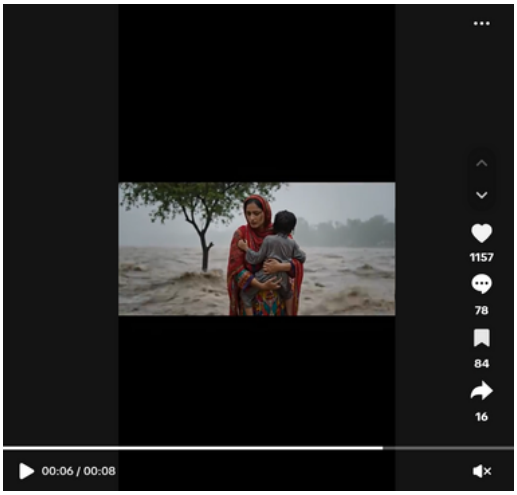
Another problematic narrative we observed over the course of our research revolved, once again, around the use of generative AI. This time, floodings in Punjab and North Pakistan were visualized as divine catastrophes. This was noted at times explicitly, with one clip showing an apocalyptic disaster unfolding upon Pakistan against the background of Quranic verses. In other instances, clearly fabricated videos showed individuals using religious symbols like the Quran or a red flag to successfully ward off floods right from their doorstep. Such content is clearly meant to disorient viewers, deliberately drawing upon their religious sensibilities for profit-driven engagement. Most of these clips have not been tagged as AI by TikTok, and content of this variety has only proliferated across the platform, crowding out critical real-time updates by sensationalizing tragedy.

AI Women in Crisis Content



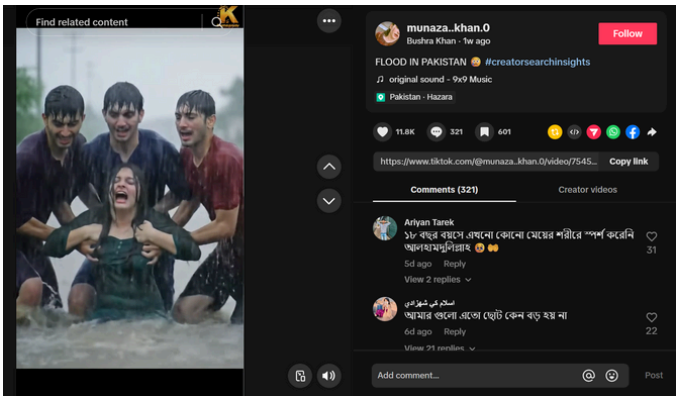
A distinct trend of misinformation noticed was the use of AI-generated women, both to elicit sympathy engagement, and to exploit a natural disaster as a vehicle to sexualise women's bodies. TikTok is filled with videos of AI-generated village scenes depicting a sole woman trying to navigate a flooded landscape, often with a baby in tow. These videos rack up views in the thousands, although only a fraction of them have the TikTok AI content label.

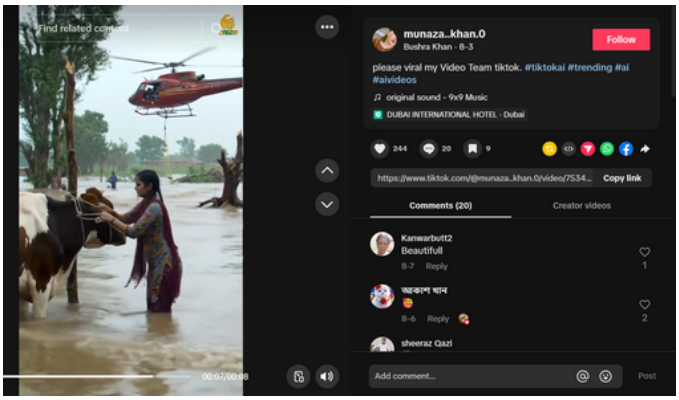
In cases where the creator has labelled the content as AI-generated, this label is not prominent, and instead appears as a transparent banner underneath the caption.

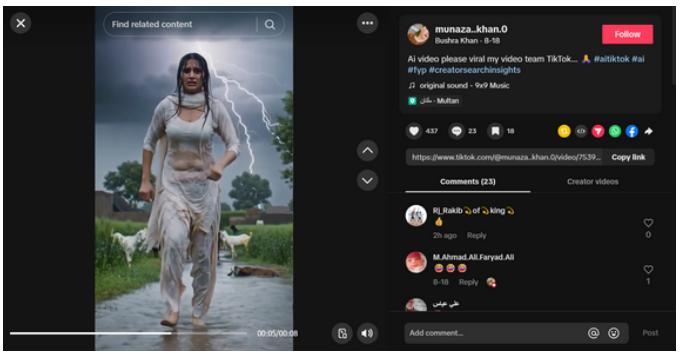


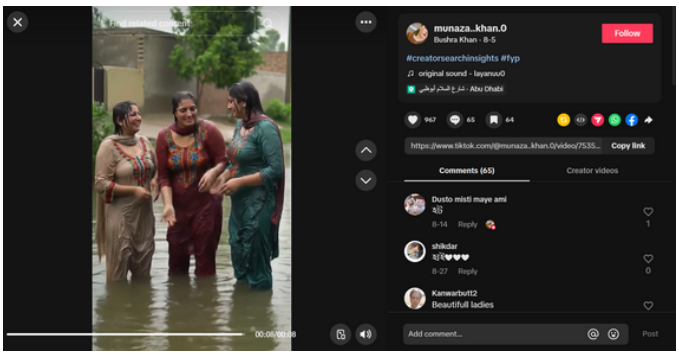
Following the same theme of depicting GenAI rural women caught in floods for engagement, many users on platforms like Instagram and TikTok started posting highly sexualised depictions of women, focusing on their bodies rather than the floods they were supposedly affected by. Across TikTok, Pakistani accounts which are engaged in creating AI village 'fetish' content that objectifies women's bodies (such as village men giving sleeping women 'chest massages'), saw the floods as another opportunity to garner engagement.

DRF uncovered several videos showing GenAI women caught in rural flooding, in drenched, figure-hugging clothing, with some videos even depicting men groping them under the guise of 'rescue'.



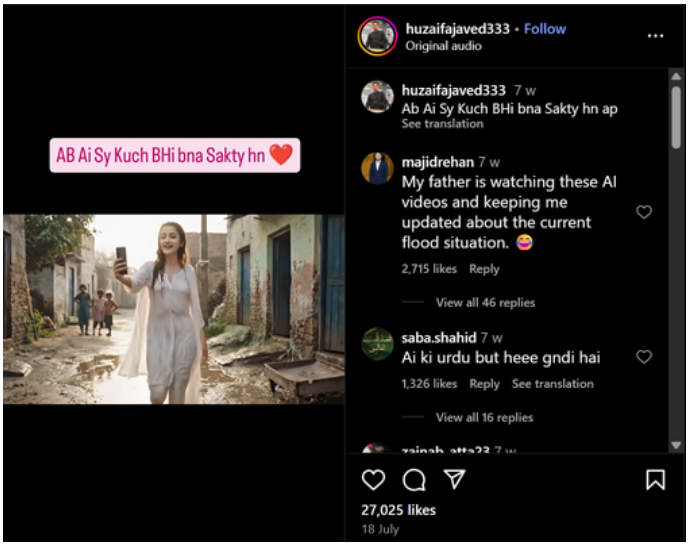


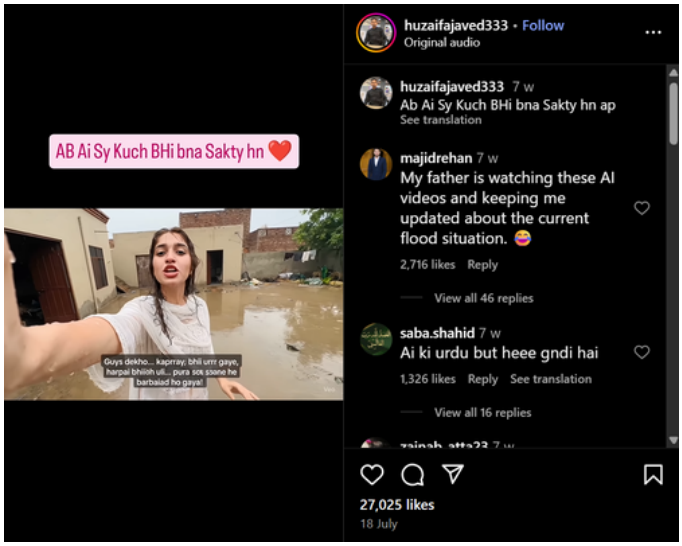




These videos rack up tens of thousands of views on TikTok, yet the AI label is notably missing on all of them.

On Instagram, this sexualisation of women took another turn, with village-vlog-style videos of an AI-generated villager showing her viewers the destruction caused by floods in her village. However, the focus of these videos is on the woman's appearance, rather than the damage caused by the floods. While not as explicit as the ones discussed above, these clips still indirectly sexualise the woman from her appearance and body language, to maximise engagement. Additionally, as evidenced from the comments, older Pakistani users of social media platforms tend to take this misinformation at face value, believing it to be a factual depiction of the situation on the ground in Pakistan. This is an alarming notion which reiterates the need for prominent AI labels on such content.

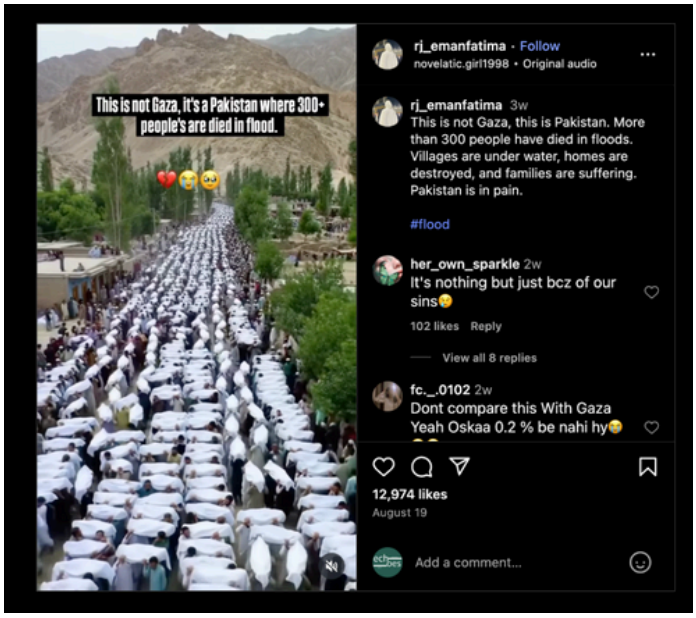




Exploiting Swat's Tragedy with AI-Generated Content

The Instagram reel shown above attempts to recreate the June 2025 tragedy in Swat, where 18 members of a single family drowned during flash floods. The clip, however, is entirely AI-generated, showing fabricated visuals and screaming voices of the family drowning while simultaneously speaking to a reporter and claiming their rescue calls went unanswered. This kind of synthetic content is both highly insensitive and dangerous since it distorts the memory of a real disaster, trivialises the suffering of survivors, and fuels misinformation for audiences already grappling with fear and grief.

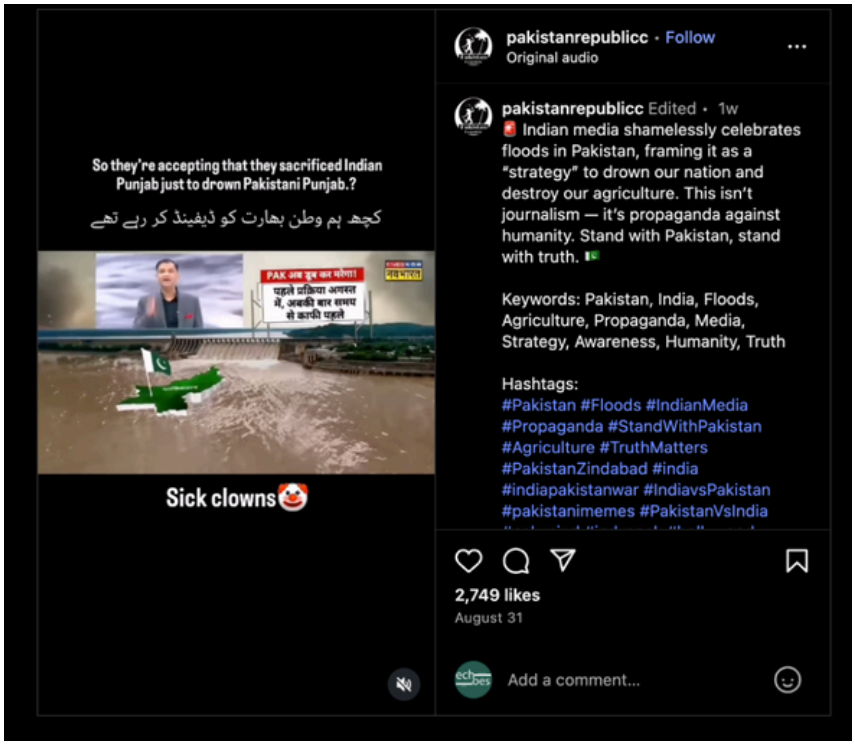
Misleading Narratives Linking Floods to Gaz



Another Instagram reel we documented declared, "*This is not Gaza, this is Pakistan. More than 300 people have died in floods. Villages are under water, homes are destroyed, and families are suffering. Pakistan is in pain.*" The video combines authentic footage of floods in Pakistan with out-of-context and likely AI-generated clips, including a mass funeral scene, children eating in relief camps, and other visuals that may not be related to the current floods or from Pakistan.

By comparing these images with the Gaza conflict, the reel blurs distinct tragedies with vastly different political and humanitarian contexts. This approach is harmful firstly because it disorients audiences who may struggle to distinguish factual flood coverage from fabricated visuals. Secondly because it instrumentalises suffering by framing one crisis through the lens of another, thereby fueling misinformation and polarised discourse. Instead of centering relief needs and accountability for Pakistan's floods, such content diverts attention and exploits grief for engagement.

Water Politics, Misinformation, and Floods

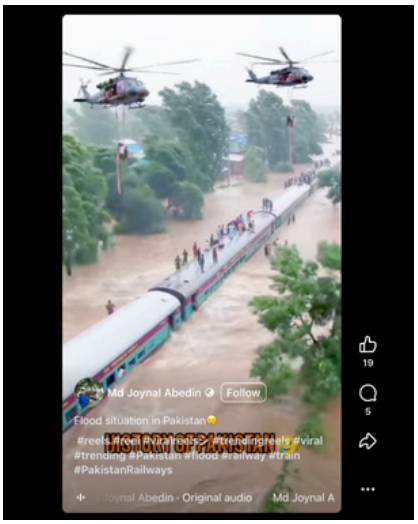


In addition to the devastating cloudbursts and glacial outbursts across Pakistan in August 2025, India released water into rivers flowing downstream into Pakistan, triggering further flooding that inundated cities and villages already reeling from disaster. While the scale of destruction continues to unfold, segments of Indian news media have actively spread misinformation, claiming that the water release was a deliberate act to drown Punjab in India itself as part of a supposed strategy to devastate Pakistan.

This framing is not only illogical, suggesting self-harm to justify hostility, but also deeply dangerous. It fuels nationalistic hate, erodes trust between communities, and distracts from urgent questions of governance, preparedness, and relief on both sides of the border. Even more troubling, Indian outlets have deployed AI-generated images and videos to dramatize and politicize these narratives, turning a climate catastrophe into a geopolitical weapon.

Water politics between India and Pakistan are a long-standing reality, but exploiting natural disasters to incite hatred represents a new low. In moments of shared suffering, the focus should be on humanitarian relief, not weaponizing misinformation.

Manufactured Rescue Narratives Through AI Content



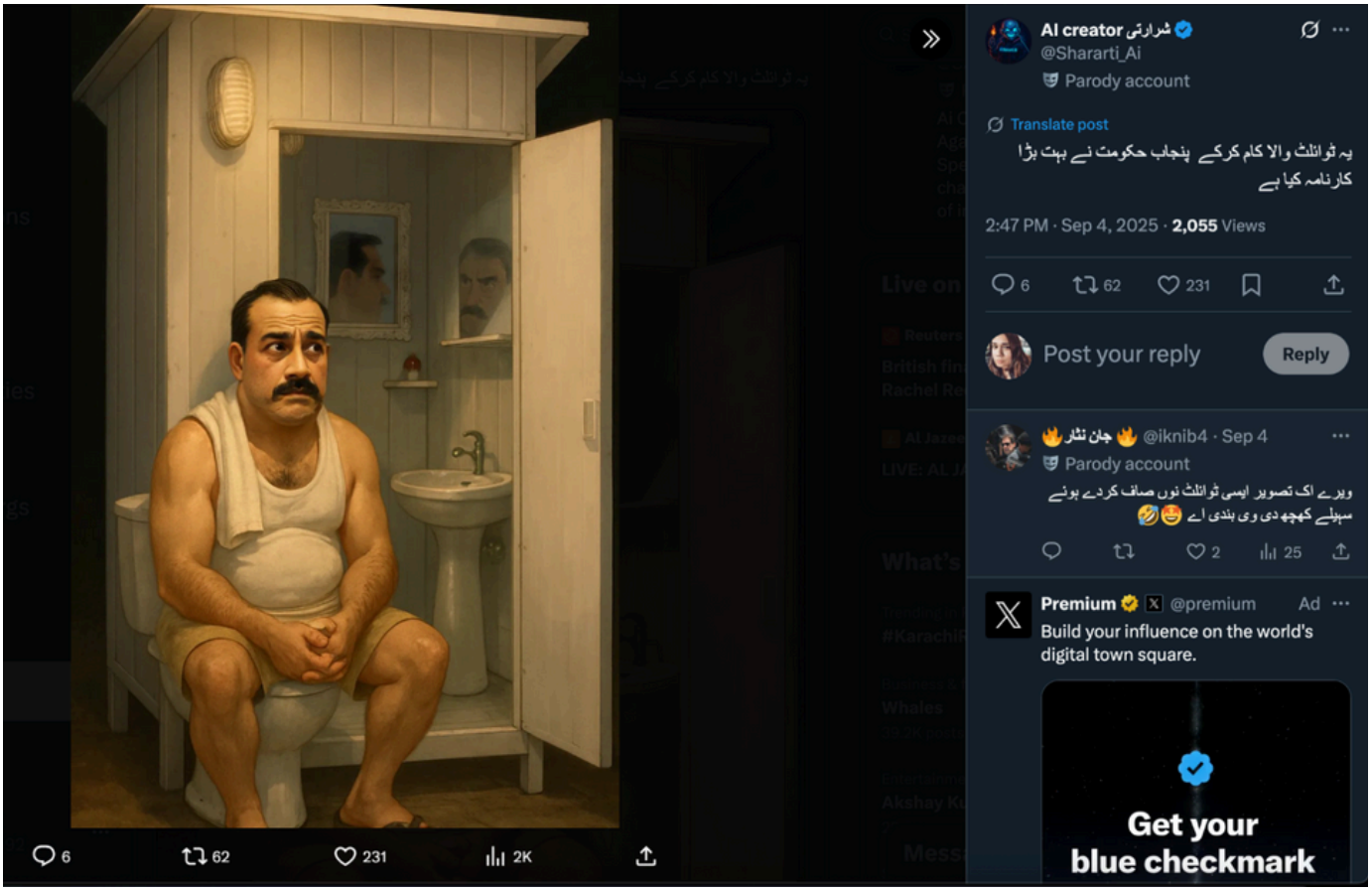
A widely shared reel on Facebook depicts a dramatic rescue operation, showing two helicopters airlifting stranded passengers from a train engulfed in floodwaters. While emotionally striking, the video is AI-generated and not authentic footage from Pakistan's floods. Its synthetic nature is evident in the exaggerated rendering of helicopters and the unrealistic movements of people, yet the absence of platform labeling allows viewers to mistake it for real documentation of events.

This type of content is deeply problematic because it creates a false impression of the scale and capacity of ongoing rescue efforts, potentially fostering both misplaced optimism among affected communities and unwarranted criticism of actual relief operations. By prioritizing virality over accuracy, AI-generated disaster visuals exploit public emotion, distort realities on the ground, and highlight the urgent need for stronger moderation and clear labeling of synthetic content on social media platforms.

Politicization of Flood Relief Efforts through AI Content

The devastating floods have also been appropriated for political commentary within Pakistan, with AI-generated content emerging as a tool to mock and criticize public figures. Notably, fabricated visuals circulated online depicted Punjab Chief Minister Maryam Nawaz and Chief of Army Staff Asim Munir "relaxing" amid flood-hit areas, an attempt to comment on their politics and their indifference during a humanitarian crisis.

This trend coincided with real controversies around flood relief communication. Maryam Nawaz faced backlash after posting an image of portable washrooms, presenting them as part of provincial flood relief in Chiniot. Social media users quickly identified the photo as a 2023 stock image from OLX, sparking allegations of misinformation and political self-promotion during a sensitive time. The opposition leveraged this misstep to demand greater transparency, while the government countered by sharing verified images and videos of actual portable facilities in relief camps.



The incident, however, opened the floodgates for further digitally manipulated attacks. Users began generating satirical AI content targeting Nawaz, blending misinformation with political mockery. Such examples highlight the dangerous overlap of politics, disaster communication, and generative AI, where fabricated visuals both exploit tragedy and deepen public mistrust in institutions.

Legal Precedents in Flood Misinformation Cases

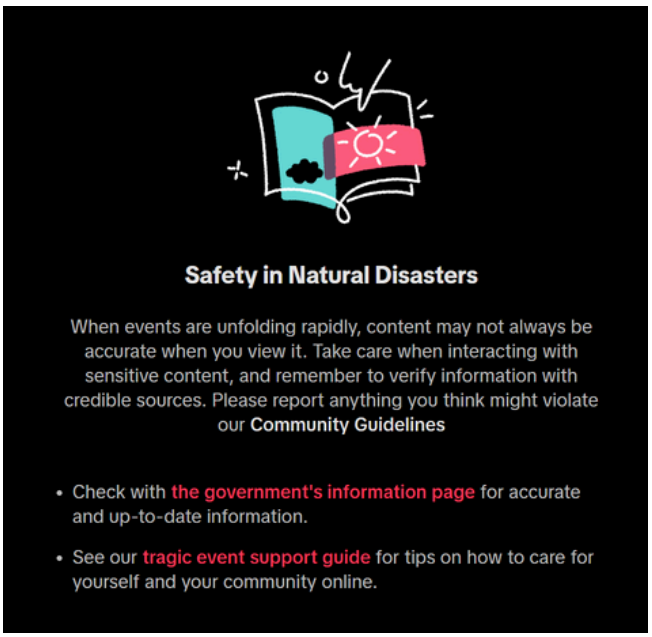
The spread of misinformation during the floods has already led to troubling legal precedents for online users. Recently, a social media influencer was booked under Section 26A of the Prevention of Electronic Crimes Act (PECA), 2016, for circulating false claims about "flooding in Jhelum villages due to a breach in the embankment" of the Mangla Dam. According to the FIR, the video alleged that several villages had been destroyed and that water had entered Jhelum city, sparking widespread panic.

Authorities stated that the video had caused fear and disorder, and under Section 26A, intentionally disseminating fake or misleading information that creates panic or unrest is a punishable offence, carrying penalties of up to three years imprisonment, a fine of up to two million rupees, or both. While the suspect has not yet been arrested, the case highlights how online misinformation can translate into severe offline consequences and how the state's reliance on punitive measures risks setting dangerous precedents for crisis communication.

Platform Spotlight: TikTok's Flood Misinformation Guide

In the midst of a torrent of misinformation in online spaces, only one platform provided cautionary content warnings and relevant links to flood-related information to the Pakistani public.

Users attempting to look up Pakistan floods on TikTok will see the banner below, which reminds them that not all content on the platform is accurate due to the rapid escalation of events, and to verify information from credible sources. It links to the National Disaster Management Authority (NDMA) website, where citizens can find up-to-date information, such as flood warnings, advisories, and GLOF (Glacial Lake Outburst Floods) alerts. It also links to TikTok's tragic event support guide, which contains tips to get support for a tragic event, as well as the procedure to report concerning TikTok content.



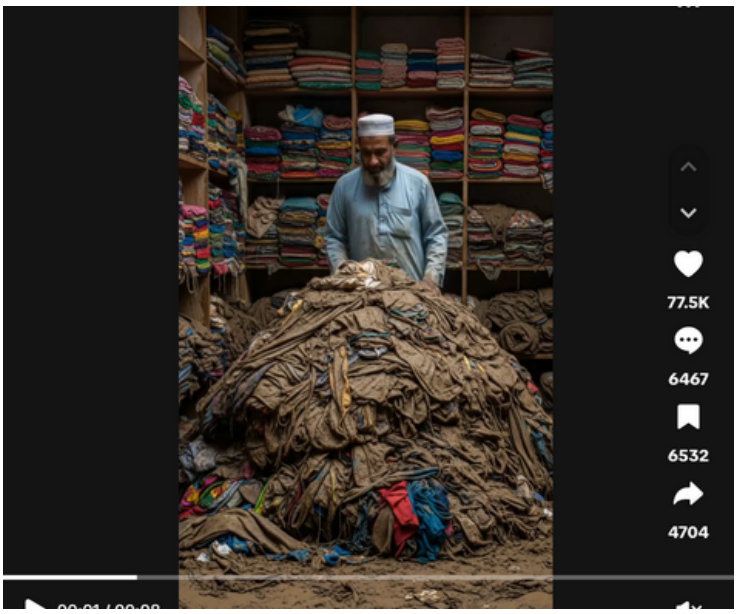
TikTok's response to this natural disaster shows a proactive step by a platform to curb misinformation, and demonstrates responsibility during climate emergencies, which are times when such action is needed.

Key Challenges and Gaps in Platform Responses

While TikTok's flood safety guide is a step in the right direction, this effort is limited in its scope, since it does not address the full extent of misinformation spreading on the platform. It also does not address regional languages or reflect local media landscapes, which means outreach is limited to the select few Pakistanis who can read and comprehend largely English-based information.

The guide does not equip TikTok users with the required critical tools to assess sources of mis/information independently, such as how to cross-check images and videos, and which trusted national and international news organisation sources to consult for verifiable information regarding the floods. Nor does it offer any tips for identifying GenAI content, which tends to garner the highest levels of engagement not only on TikTok, but across platforms.

Moreover, platform algorithms largely favour sensational content with no regard for their veracity, and TikTok is no stranger to this tendency. Proof of this can be found in the viral misinformation content littering the platform, much of which is AI-generated. The same is true for all other social media platforms, especially Instagram, Facebook, and X/Twitter.



Most of these platforms lack real-time crisis moderation strategies tailored to the Pakistani context, which means there is no tangible connection between the problems caused by these floods, and any concrete solutions, such as efforts by NGOs like Al-Khidmat Foundation and Mahwari Justice working on the ground with victims in flood-stricken areas.

The Case for Broader Platform Accountability

In an era of climate-driven disasters, social media platforms have a responsibility to act responsibly and uphold global standards when crises strike. According to human rights experts, online platforms should be guided by frameworks like the UN Guiding Principles on Business and Human Rights (UNGPs), which establish a global benchmark for respecting users' rights and mitigating harm. In practice, this means companies have a responsibility to curb harmful content, including disaster-related misinformation, as part of their duty to respect human rights in public discourse.

Misinformation during emergencies is not a trivial matter; it can "severely compromise humanitarian aid and disaster response by skewing perceptions, damaging credibility, and disrupting coordination," ultimately undermining the capacity of authorities to manage crises effectively. The climate catastrophe in Pakistan highlights why platform accountability during climate emergencies isn't just idealistic, it's necessary to protect lives and public order.

The current climate crisis in Pakistan has exposed major gaps that remain in how platforms currently handle crisis misinformation, especially in regions like South Asia. Researchers have noted that platforms' content rules are often unevenly enforced across different countries and languages, creating "unstable, unpredictable, and unsafe environments for users." In particular, moderation is much stronger in some languages than others, leaving many local communities vulnerable. Digital rights groups in the region have repeatedly called for stricter transparency and moderation standards in regional languages, since a lack of local-language content monitoring allows mis/disinformation to spread unchecked. Without robust policies for non-English content, misleading posts in Urdu, Pashto, Punjabi, and other local languages tend to repeatedly slip through, often with real-world consequences in disaster situations.

Accountability issues also lie in platforms' profit-driven algorithms. Emotionally charged or sensational posts tend to get more clicks, and algorithms have historically amplified such content for engagement. Unfortunately, this means during a flood or cyclone, dramatic misinformation can be algorithmically boosted, outcompeting genuine emergency updates. A 2024 youth survey by Mental Health America found that social media algorithms "often amplify content that is harmful… including misinformation and material that is polarizing, sensational and graphic." This incentive structure directly clashes with the public interest in a crisis, when accurate and verified information is needed. Platforms must therefore reevaluate their algorithms and curation practices in emergencies, for example, elevating official alerts over shocking clickbait, as part of a broader accountability approach.

Moreover, it is essential that platform accountability must be locally informed and inclusive. In February 2025, Digital Accountability Collective South Asia (DACSA) warned of the same and highlighted that the recent Big Tech policy shifts (like reducing fact-checking and safety safeguards) were "exacerbating misinformation, political instability, and communal violence" in the region. DACSA has urged social media companies to "engage meaningfully with civil society to develop policies that prioritize user well-being", rather than imposing one-size-fits-all solutions from afar.

This could mean partnering with local disaster management agencies and fact-checkers (as TikTok did by linking Pakistan's NDMA into its flood info guide), hiring more content moderators fluent in regional languages, and establishing clear crisis response protocols. Policymakers in the Global North have already begun pushing in this direction, for example, the EU's new Digital Services Act even introduced a crisis response mechanism for online platforms, but globally, much more needs to be done. Social media companies should develop tiered moderation strategies for crisis periods, ramping up real-time fact-checks and false content takedowns when a natural disaster or emergency unfolds. Digital rights groups have pointed out that media regulators and governments "must enforce stronger accountability measures" on both social media and news outlets, because combating misinformation is fundamental to protecting public safety and peace in times of disaster.

Empowering Users: Digital Awareness in Crisis

While platforms shoulder a major responsibility, users themselves are a crucial line of defense against misinformation, especially during chaotic climate emergencies. Digital awareness and media literacy can save lives by ensuring individuals make informed decisions rather than reacting to panic-inducing rumors. Experts stress a few key habits for staying safe and informed:

1) Verify first: go to official sources

- Start with trusted channels: Pakistan's NDMA, provincial disaster authorities, Met Dept, and reputable newsrooms.
- Official alerts cut through rumor cycles and give actionable guidance (routes, shelters, weather warnings).

2) Pause before you share

- If a post makes you feel angry, scared, or shocked, stop. That's often by design.
- Do a 30-second check:
    1. Has a credible outlet also reported it?
    2. Is the account identifiable and trustworthy?
    3. Does the claim cite a source you can open and read?
- If answers are "no" or "not sure," don't forward.

3) Spot the red flags of misinformation

- Sensational framing: ALL-CAPS, extreme language, dramatic thumbnails.
- Vague sourcing: "Experts say…" with no names, dates, or agencies.
- Missing details: Real alerts include who/what/where/when and contact info.
- Old or miscaptioned media: Check dates and locations; crisis photos often get recycled.
- Too-perfect visuals: Be wary of AI-edited or "impossible" footage (odd hands, inconsistent shadows, warped text).

4) Use fact-checks & quick verification tools

- Local fact-checkers: Soch Fact Check, Geo Fact Check, iVerify regularly debunk viral crisis claims.
- Reverse image search: See if that "new" flood photo is from another year/country.
- Platform tools: Look for Community Notes on X, fact-check labels on Facebook/Instagram.
- When in doubt: Treat surprising WhatsApp forwards as unverified until confirmed.

Lastly, there's a broader need for digital literacy education so that communities can identify misinformation before the next disaster hits. It's important to equip people with critical thinking skills and practical tools to identify misinformation and foster ethical digital engagement starting from schools and local community centers. Media literacy should be part of education curriculums and community programs, users will be better prepared to spot fake news about floods, fires, or other crises. When users know how to assess sources, cross-check news, and spot manipulative content, they are far less likely to fall victim to panic or propaganda during an emergency.

DigitalRightsFoundation
"KNOW YOUR RIGHTS"

@digitalrightsfoundation

@DigitalRightsFoundation

@DigitalRightsPK

@digitalrightsfoundation

@digitalrightsfoundation

@DigitalRightsPK