

TOXIC FEEDS

Mapping the Digital Harms Ecosystem, Escalation and Accountability in Somalia

July 2026

BUILD UP A

PDRC

PUNTLAND
DEVELOPMENT &
RESEARCH CENTER



SOMALI
PUBLIC AGENDA

About this report

This analysis forms part of a wider Build Up programme on digital harms in conflict that connects work across several contexts. In Somalia, this digital harms analysis was conducted as part of a broader effort to scope and initiate dialogue on digital harms in the region, in the context of ongoing challenges to key governance and democratisation processes. A complementary social media analysis was conducted alongside this analysis, which can be found on the Build Up website.

For questions or to request support: team@howtobuildup.org

Research Team and Acknowledgements

Research team and authors: Mohamed Abdulkadir Mohamud, Research Consultant, Build Up; Nura Abdi, Communications Lead, PDRC; Muctar Hersi, Head of Programs, PDRC; Ibrahim Jibril, Researcher, Somali Public Agenda; Megan Grazier, Peacebuilding Strategy Lead, Build Up.

The authors would like to recognise the contributions and reviews of Mahad Wasuge (Executive Director, Somali Public Agenda), Allan Cheboi (Digital Innovations and Insights Lead, Build Up), and Meghan Brown (Inclusive Politics Lead, Build Up).

To cite

Mohamud, M.A., Abdi, N., Hersi, M., Jibril, I., Grazier, M. 'Toxic Feeds: Mapping the Digital Harms Ecosystem, Escalation and Accountability in Somalia.' Build Up / Somali Public Agenda / PDRC, July 2026.

Partners & funders

This material has been funded by UK International Development from the UK government; however, the views expressed do not necessarily reflect the UK government's official policies.



EXECUTIVE SUMMARY	4
INTRODUCTION	6
METHODOLOGY	10
THE INFORMATION ECOSYSTEM	13
Platform Landscape and Harm Concentration	13
Actors and the Paid Influence Economy	14
Audience Vulnerabilities	16
DIGITAL HARMS IN SOMALIA	18
I. Disinformation and Misinformation	18
II. Hate Speech and Identity-Based Incitement	21
III. Tech-Facilitated Gender-Based Violence (TFGBV)	23
IV. Intimidation, Doxxing and Censorship	27
V. Cyber Intrusion, Fraud, and Threats to Public Digital Infrastructure	29
ESCALATION DYNAMICS	31
Five Stage Escalation Model	31
Case Studies of Escalation	32
The Diaspora Feedback Loop	33
LEGAL AND INSTITUTIONAL FRAMEWORK	34
Legal Framework	34
Institutional Architecture	35
The Enforcement Gap and Selective Application	36
RESILIENCE & MITIGATION	37
Trusted Messenger Correction	37
Formal Verification and Reporting Mechanisms	38
Emerging Correction Practices	39
The Timing Problem	39
CONCLUSION	40
RECOMMENDATIONS	42
BIBLIOGRAPHY	45
ANNEXES	46

EXECUTIVE SUMMARY

Somalia's digital information environment has expanded faster than the institutions, norms, and laws needed to manage it. The country had 10.7 million internet users at the start of 2025, equivalent to 55.2 percent internet penetration, and 3.05 million social media user identities.¹ Public platform figures should be interpreted carefully because advertising reach, registered identities, shared devices and actual users are not the same measure. Connectivity has created real benefits: civic voices can reach national audiences, diaspora communities remain closely connected to local debates, mobile money underpins everyday transactions, and remote education and information-sharing are more accessible than in previous decades. Yet the same infrastructure is also used to manufacture narratives, weaponise clan identity, silence women, intimidate journalists and accelerate conflict dynamics.

This report presents findings from twelve KIs and a multi-stakeholder convening that took place in the first half of 2026. It maps five overlapping categories of digital harm – disinformation and misinformation; hate speech and identity-based incitement; technology-facilitated gender-based violence (TFGBV); intimidation, doxxing, and censorship of journalists and critics; and cyber intrusion, fraud, and threats to public digital infrastructure. The core argument is that digital harms in Somalia are not simply harmful content appearing online. They emerge from the interaction between platform speed, political incentives, clan identity, weak institutional trust and limited corrective infrastructure.

The report organises these harms into five overlapping categories: (1) Disinformation and Misinformation; (2) Hate Speech and Identity-Based Incitement; (3) Tech-Facilitated Gender-Based Violence (TFGBV); (4) Intimidation, Doxxing, and Censorship; and (5) Cyber Intrusion, Fraud, and Threats to Public Digital Infrastructure. Across all five categories, a consistent five-stage escalation pathway is identified: Trigger → Origination and Amplification → Identity Framing → Mobilisation → Offline Outcome.

Across all five harm categories, the research points to six conclusions about how digital harms operate, escalate, and evade response in Somalia:

1. Somalia's most serious digital harms operate through a public-private escalation system: content becomes more dangerous as it moves between public platforms and private channels.
2. Paid influence and coordinated amplification are now structural features of the information environment, especially around elections, political disputes and clan-linked controversies.
3. Clan identity is the main escalation frame: narratives become more dangerous when they portray a community as threatened, humiliated or under attack.
4. TFGBV is the most urgent protection gap. It is severe, underreported and structurally silencing for women in public life and women using digital spaces.

¹ DataReportal, Digital 2025: Somalia, January 2025.

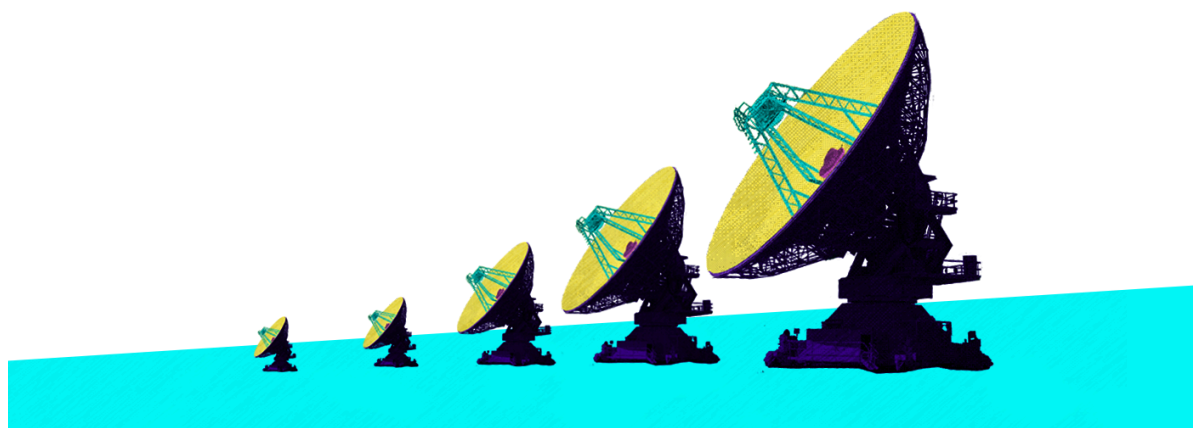
5. Diaspora amplification can intensify local disputes but can also be redirected toward constructive civic and economic engagement.
6. Platform takedown often arrives too late once content has moved into screenshots, voice notes, clipped videos and closed groups.
7. Somalia's legal framework is developing, but implementation, judicial capacity, victim support, public trust and platform engagement remain weak.

These conclusions are not abstract, but illustrative of documented patterns across the country. The escalation pathway documented in this report — from online trigger to offline mobilisation — has been active in three documented cases: Laascaanood, Galkayo, and the case of the female Bajaj driver detained in Mogadishu. The same pathway is structurally present in Somalia's recurring election disputes and ongoing federal-FMS tensions. As Somalia approaches its next electoral cycle, the risk of harmful digital content accelerating physical confrontation is not hypothetical — it is documented and predictable.

“Events are more likely to move offline when they involve clan identity, territorial disputes, armed groups, revenge narratives, or direct accusations against communities or leaders. If the issue is only a political debate, it may remain online. But if it touches clan honour, land, revenge, or violence, it can quickly create fear or confrontation offline.”

Academic respondent, Mogadishu

The report concludes that effective response requires earlier intervention in the escalation pathway. **Takedown, legal reform and fact-checking remain necessary, but they cannot carry the response alone.** Somalia needs stronger protection for women and public-facing groups, clearer accountability for paid influence, community-based verification systems, operational data protection and cyber incident response, and more meaningful engagement by major platforms with Somali language, law and context.



INTRODUCTION

Somalia's digital transformation has unfolded faster than its institutions can govern it. In a context shaped by federal tensions, clan identity, active armed conflict and weak accountability, platforms have become more than communication tools. They are spaces where political legitimacy is contested, grievances are amplified and social harm can move quickly offline. Commercial satellite connectivity is now part of this landscape. In April 2025, Starlink became available in Somalia after the government granted the company an operating licence, with officials expressing hope that the service would improve internet quality and expand access to remote areas². This kind of connectivity growth may bring clear benefits, especially for under-connected areas and institutions. But without parallel investment in digital literacy, platform accountability and trusted correction systems, it may also extend exposure to harmful narratives faster than communities can manage them.

This report is produced as part of Build Up's wider work on the impact of digital narratives and behaviours in fragile and conflict-affected settings. It follows the analytical approach used in related work and refined during an inception workshop in Nairobi in February 2026, while applying a methodology grounded in qualitative primary research rather than platform data analysis. The primary audience is practitioners and policymakers working on digital governance, conflict prevention, media freedom, and gender justice in Somalia. The secondary audience is international donors and platform companies whose decisions shape the environment in which Somali digital harms occur. A parallel social media analysis was conducted to assess where the dynamics identified in the desk review and KIs are also present in online data.

Somalia's political context in 2025–2026 makes this analysis urgent. Constitutional tensions between the Federal Government of Somalia (FGS) and Federal Member States (FMSs), particularly Puntland and Jubaland, have intensified the information environment, as rival administrations compete for public legitimacy online. Electoral processes at both federal and state level have generated documented spikes in harmful content. The legacy of the Laascaanood conflict and ongoing operations against armed extremist groups in Somalia continue to produce disturbing content and extremist propaganda in digital channels. At the same time, a new wave of AI-generated synthetic content is reaching audiences with limited ability to distinguish authentic from fabricated material, while platform moderation remains weak in Somali-language and local-context content. These dynamics all interact with a highly social, oral and identity-driven information environment.

As Somalia approaches its next electoral cycle, the question is not whether digital harm will shape the political environment; it is whether institutions, platforms and communities have the tools to interrupt the escalation pathway before it reaches offline harm.

² Reuters, "Musk says Starlink now available in Somalia," 13 April 2025.

Somalia's Digital Landscape

Connectivity and Infrastructure

Somalia's digital infrastructure is unusual for a country at its level of institutional development. Private telecommunications companies, mobile money providers and social media platforms have expanded faster than formal state regulatory capacity. There are an estimated 11.3 million active cellular mobile connections, 10.7 million internet users, 55.2 percent internet penetration and 3.05 million social media user identities at the start of 2025. At the same time, 51.2 percent of the population remained rural, and the median age was 15.6, which means that exposure to digital information increasingly overlaps with youth vulnerability and uneven access to education³.

10.7M Internet users (Jan 2025) DataReportal 2025	55.2% Internet penetration DataReportal 2025	3.05M Social media identities DataReportal 2025	70%+ Adults using mobile money GSMA / Hormuud 2024
--	---	--	---

Facebook remains one of the largest addressable social media audiences in Somalia, while TikTok's ad tools indicate a large adult reach. TikTok's user numbers are not reliably measurable through public data sources, but every respondent in this study placed it as the fastest-growing and most influential platform for political communication, particularly among youth.

Mobile money is also central to Somalia's digital life. According to a World Bank survey in 2017, 73 percent of Somali adults used mobile money services, including high reported use among urban, rural and nomadic populations⁴. This financial inclusion is a major achievement, but it also means that social engineering, SIM-swap fraud, false promotions and digital scams have broad social reach. Digital harm therefore includes not only political content but also economic and infrastructure-related risks.

Connectivity remains uneven. Urban centres such as Mogadishu, Garowe, Hargeisa, Kismayo and Bosasso generally have better electricity access, more reliable internet connectivity and more diverse platform use. Rural and nomadic communities often rely on voice calls, WhatsApp voice notes, forwarded clips and short videos rather than long written posts or formal news sources. This matters because content arriving through trusted social networks is often treated as credible, particularly where independent verification is not easily available.

Starlink's arrival in 2025 is particularly significant because it may bring high-speed internet to areas previously unreachable by mobile networks. In the near term it is more likely to benefit

³ DataReportal, Digital 2025: Somalia, January 2025
⁴ World Bank / Altai Consulting, Mobile Money in Somalia: Household Survey and Market Analysis, April 2017

institutions, businesses, NGOs, and diaspora-supported households than low-income rural users; over time, however, falling prices or shared access models could expand connectivity among previously less connected communities, including pastoralist populations⁵.



The Somali Political and Clan Context

Digital harms in Somalia cannot be understood without the structural context of clan identity and political fragmentation and active conflict in some parts of the country. Somalia's federal system includes five Federal Member States – Puntland, Jubaland, South West, Galmudug and Hirshabelle - alongside the Banadir Regional Administration and Somaliland, which remains outside federal control while continuing to seek international recognition. At the same time, clan identity continues to shape political life, social protection, conflict mobilisation and public legitimacy. These identities are not incidental to digital harms; they are one of their primary organising forces.

Since the collapse of the Siad Barre government in 1991, clan networks have served simultaneously as sources of social protection and drivers of communal tension. The digitalisation of Somali society has not dissolved this structure; it has accelerated and geographically expanded it. Clan grievances that might once have taken days or weeks to mobilise now travel across platforms in hours. Diaspora actors in Minneapolis, London, Nairobi, Toronto and elsewhere can insert content into cities, towns and villages across Somalia's information environment within minutes.

⁵ Reuters, "Musk says Starlink now available in Somalia," 13 April 2025

Somalia's governance structure adds a second layer of complexity. Each administration operates its own information ecosystem and deploys its own digital communication strategy. Constitutional disputes between the FGS and member states, particularly tensions involving Puntland and Jubbaland, have become major triggers for disinformation spikes and coordinated online campaigns in 2025 and 2026. These political disputes matter because they create information demand and narrative competition. When institutions compete over legitimacy, social media becomes the arena in which political claims are tested, reframed and mobilised.

Al-Shabaab and other armed extremist groups, including ISIS-linked actors, add a third dimension. These groups maintain digital communication capabilities, and while government-led content removal campaigns have pushed some activity increasingly toward private messaging channels, their narratives, often repackaged through religious framing and grievance politics, continue to circulate periodically in mainstream digital spaces.

The Media and Information Environment

Somalia has a plural and contested media environment. Private radio stations, television outlets, and online news portals operate across the country, alongside a growing ecosystem of individual content creators, TikTokers, YouTube channels, and WhatsApp group administrators who perform journalism-adjacent functions with varying levels of professional training, editorial accountability and legal recognition.

The line between journalism, political communication, and influence operations has become increasingly blurred. Influencers with hundreds of thousands of followers are hired by political actors to amplify narratives; media workers can face arrest or intimidation for covering protests or politically sensitive issues; and official government communication strategies may overlap with paid promotion through the same informal influencer networks that also participate in contested or harmful online discourse.

Press freedom evidence remains an important context. Reporting by press freedom organisations and international media has documented continuing risks to journalists in Somalia, including arrests and intimidation during politically sensitive periods; one recent report documented journalists being detained and beaten in Mogadishu in May 2026.⁶ This external context supports, but does not replace, the interview evidence on digital intimidation and selective enforcement.

⁶ Guardian, "Guardian reporter and colleagues detained and beaten by Somali police," 9 May 2026; Somali press freedom reporting

METHODOLOGY

This study uses qualitative primary research. The evidence base consists of twelve semi-structured Key Informant Interviews (KIIs) and one Multi-Stakeholder Convening conducted in Mogadishu in April 2026. The KIIs followed a semi-structured guide covering the information ecosystem and platforms; actors and incentive structures; harms and escalation dynamics; doxxing, abuse and public-facing groups; the regulatory and institutional environment; resilience, counter-narratives and verification practices; and closing reflections. Interviews were conducted in Somali and English.

Respondents were selected to represent a cross-section of institutional positions, professional backgrounds, geographic coverage, and gender perspectives. The respondent pool included government and regulatory perspectives, media and journalist perspectives, digital safety and civil society practitioners, a technical cybersecurity respondent, an academic/AI respondent, and a women's media representative. The Multi-Stakeholder Convening brought together civil society, media, academic, and government-linked participants. The sample was not designed to be statistically representative; it was designed to generate grounded, comparable and conflict-sensitive qualitative evidence.

Evidence was organised through a qualitative framework covering: A) information ecosystem and platform channels; B) actor types and incentive structures; C) harm typology; D) escalation dynamics; E) vulnerability and affected groups; F) regulatory and institutional environment; G) resilience, counter-narratives and verification practices; and H) cross-cutting analytical flags. Evidence was coded at the incident or observation level, not simply by interview. This allowed comparison across respondents and themes, and helped identify areas of convergence, divergence and uncertainty. Where relevant, corroboration status was tracked as multi-source, partial or single-source.

The coding process was used to structure qualitative evidence and identify patterns, not to quantify prevalence or produce statistically representative claims. The report therefore uses evidence-confidence language throughout. Claims appearing across several KIIs and the convening are treated as stronger findings; claims based on one or two examples are presented as illustrative or requiring further inquiry. Respondent accounts are treated as qualitative evidence to be interpreted, compared and triangulated, rather than as independently verified fact.

Throughout this report, evidence is attributed to its source type. Claims drawn from Key Informant Interviews are identified by respondent type, date, and location in footnotes (e.g. Interview with Journalist, 22/5/2026, Mogadishu). Contributions from the Multi-Stakeholder Convening are cited as: Digital Harms Convening, 27/5/2026, Mogadishu. Desk research sources are cited in-text. A bibliography is provided at the end of the report.

Sources and Confidentiality

Given the sensitivity of the findings, this report avoids naming respondents or attributing quotes to specific institutions except where the information is already public or institutionally necessary. Quotations are attributed only by broad respondent type and location to reduce political, professional, and personal risk. Given the small size of Somalia's digital policy, media, and civil society community, even broad respondent-type attribution carries some identification risk. All attributions in this report have been reviewed against that risk, and some descriptions have been further generalised where specific combinations of role, location, and topic would narrow the pool to a single identifiable person.

Respondents consented to participation with awareness of the research purpose and its public outputs. Evidence from the Klls and convening was systematically coded and organised through the Digital Harms coding framework developed for this project. External sources are cited in footnotes.



Key Analytical Concepts

This report introduces several analytical concepts developed from the research. They are defined in full in Annex C. Key terms used throughout include:

- Qarxis — The exposure or threatened exposure of intimate or identifying material to destroy reputation or coerce compliance; the Somali term for a digitally mediated form of gender-based harm that extends beyond 'doxxing'.
- La yiri / 'It was said' — The epistemic formula that has replaced source attribution in much of Somalia's digital information environment: content is accepted and shared because 'it was said', without traceable origin.
- The public-private escalation pathway — The mechanism by which harmful content moves from public platforms into private messaging channels, where it is recontextualised, amplified and acted upon. This movement is how platform moderation is circumvented and offline mobilisation is organised.

- Identity framing — The process by which content is recast in clan or group-threat terms as it escalates, transforming a political dispute or individual incident into a communal grievance.
- The paid influence economy — The network of political actors, media officers, influencers and amplifiers operating on commercial and patronage terms to shape the information environment.

Methodological Limitations

Gender representation: Women were underrepresented in the KII sample. TFGBV findings should therefore be read as pattern-level observations from practitioner and media perspectives rather than direct survivor testimony. Further women-focused and survivor-centred research remains a priority.

No direct rural or nomadic voices: All respondents were professionals — journalists, government-linked actors, NGO practitioners, academics and technical experts. No rural or nomadic community members were interviewed directly. Findings on rural vulnerability are therefore based on practitioner observations and should be interpreted with caution.

Evolving legal landscape: Somalia's digital governance framework is evolving. The Cybercrime Bill had been approved by Cabinet but was not yet enacted at the time of research, while the Data Protection Authority was established in law but not yet fully operational. Some findings may require updating as implementation progresses.

Methodological scope: This study does not include computational social media analysis, platform data or quantitative content monitoring, and platform companies did not participate. Findings therefore rely on qualitative primary research, convening evidence and desk review.

THE INFORMATION ECOSYSTEM

Platform Landscape and Harm Concentration

Somalia's digital public sphere is dominated by four platforms that serve distinct but overlapping functions.

Platform	Primary function in Somalia	Primary harm type
Facebook	Public news, political debate, community discussion	Harassment, defamation, coordinated political targeting
TikTok	Youth-facing political and social content	Clan-coded messaging, emotional content, fast political reaction
WhatsApp	Private communication infrastructure	Narrative coordination, harmful content 'second life', voice note disinformation
Telegram	Unmoderated distribution channel	Graphic content, intimate material, political propaganda

Facebook remains one of the central public spaces for news, political debate and community discussion. It functions as a visible public record of political statements, accusations, live broadcasts, news posts and community reactions. Harassment, defamation and coordinated political targeting are often easiest to observe on Facebook because content is public or semi-public. Respondents also noted that **Facebook moderation has improved in some areas, but Somali-language and context-specific moderation remains limited, especially when harm depends on clan-coded meanings rather than explicit slurs.**⁷

TikTok is described by respondents as the fastest-growing and most influential platform for youth-facing political and social content. Its short-video format, live-streaming features and recommendation system make it particularly effective for emotional content, satire, mockery, clan-coded messaging and fast political reaction. Respondents perceived **TikTok's recommendation system as deepening repeated exposure to politically charged or identity-coded content.** This is a perception based on respondent experience, not platform data, but it was widely reported.⁸

WhatsApp serves as the private communication infrastructure underlying much of Somalia's digital information ecosystem. It is where political actors coordinate narratives before releasing them publicly; where community members receive and forward breaking

⁷ Multiple KII respondents, April–May 2026, Mogadishu and Garowe; Digital Harms Convening, 27/5/2026, Mogadishu.

⁸ Interview with Media Practitioner, May 2026, Garowe; Interview with Regulatory Respondent, May 2026, Mogadishu; Digital Harms Convening, 27/5/2026, Mogadishu.

news; where harmful content gains its ‘second life’ after being shared from public platforms into trusted private groups; and where rural and nomadic communities, who rely on voice notes rather than text, are most exposed to unverified information. **WhatsApp’s end-to-end encryption makes it effectively unmonitorable from outside, and its group size limits create relatively insulated communities where false information circulates without external challenge.**⁹

Telegram occupies a distinctive and deeply troubling position in Somalia’s harm ecosystem. Unlike Facebook and TikTok, which have content moderation policies, however imperfectly enforced, **Telegram operates with minimal content restrictions, enabling the distribution of material that other platforms remove: graphic conflict footage, intimate content, political propaganda, and material that has migrated from mainstream platforms** through embedded links. ‘What makes Telegram different from all other platforms is that it has absolutely no content screening or restrictions’ (media association respondent). Respondents described Telegram channels operating on a commercial subscription basis, charging as little as one dollar per month to access intimate and exploitative content. During the Cal-Miskaad Hilaac military operation, Telegram became the primary distribution channel for graphic footage of casualties, with non-journalist content creators seeking audience growth distributing material that mainstream media outlets declined to publish. At the same time, Telegram serves important legitimate functions: educational institutions, media organisations, and civil society groups use it for file sharing and coordination.¹⁰

“The most dangerous moment is not when content first appears publicly, but when it is reinterpreted and intensified inside messaging groups.”

Regulatory respondent, Mogadishu

Actors and the Paid Influence Economy

Somalia’s online information environment is shaped by a layered set of actors with different motivations, capabilities, and accountability relationships.

Political actors and their media teams are the dominant force in the paid influence economy. Politicians retain media officers who create WhatsApp coordination groups to distribute talking points, draft narratives, and simultaneously contact networks of influencers before content is released publicly. A media trainer described the mechanics: ‘Politicians work through media officers or communication teams. These media officers create WhatsApp groups to coordinate narratives, prepare talking points, and simultaneously send messages to influencers, asking them to share or promote specific content.’ The media association respondent described a model in which a single well-connected influencer receives a

⁹ Interview with Regulatory Respondent (NCA), May 2026, Mogadishu.

¹⁰ Interview with Regulatory Respondent (NCA), May 2026, Mogadishu; Interview with Technical Cybersecurity Respondent, May 2026, Mogadishu.

lump-sum payment and then uses their network to amplify the message across platforms. Government institutions are also direct actors: respondents at the multi-stakeholder convening described government use of troll and bot networks to coordinate attacks against critics and spread misleading information.

“Politicians work through media officers who create WhatsApp groups to coordinate narratives and simultaneously send messages to influencers asking them to share specific content. Sometimes one well-connected influencer receives a lump-sum payment and then uses their network to amplify the message across TikTok, Facebook, and YouTube.”

Media association respondent

TikTok influencers have become among the most powerful narrative actors in Somalia, with the most prominent commanding audiences exceeding one or two million followers. They operate within what the media association respondent described as a ‘popularity circle’ – an informal prestige economy in which follower counts above approximately 100,000 unlock access to paid campaign networks and cross-regional amplification. The transnational scale of these networks was vividly illustrated by the media association respondent, who described an encounter in Djibouti at a conference in October 2024: ‘While we were having tea at a big mall, a group of popular Somali TikTokers walked in. All women and children suddenly rushed to meet them. I never thought a TikToker from Somalia would be so popular in Djibouti.’ The influencer economy creates what a digital safety practitioner described as ‘online conflict as business model’: staged disputes, manufactured controversies, and clan-coded content generate the engagement metrics that translate into advertising revenue, online gifts, and political sponsorship. The women’s media practitioner described a specific variant: the ‘extort-then-praise’ cycle, in which influencers direct sustained negative campaigns at politicians until financial incentives are offered, at which point coverage shifts to positive framing.

“There are cases where influencers appeared to be publicly attacking one another while privately collaborating to increase followers, views, and financial rewards. These manufactured disputes generate significant attention despite being coordinated behind the scenes.”

Independent journalist, Garowe

Diaspora actors constitute a distinct and particularly significant category. Somali diaspora communities – concentrated in North America, Western Europe, and the Gulf – have resources, connectivity, and strong clan identifications that make them disproportionately

active in the online information environment. ‘Problems rarely start from within Somalia. The harmful narrative often originates outside Somalia and finds its way into the country. They cook narratives and share them online’ (media association respondent). An independent journalist in Garowe identified a ‘new diaspora cohort’ – those who emigrated in the last five to ten years and remain more deeply embedded in clan dynamics than either long-established diaspora or local populations – as particularly active in amplifying divisive content. Diaspora actors provide financial support, fundraise for clan militias, and operate outside Somali jurisdictional reach, making them simultaneously among the most influential and least accountable actors in the ecosystem.¹¹

Anonymous and pseudonymous accounts function as the distribution infrastructure for both organic misinformation and coordinated disinformation. Respondents described networks of accounts using stolen photographs of familiar influencers or trusted community figures such as religious leaders, women representatives, youth icons to lend credibility to harmful narratives. ‘False information is often accepted and shared without verification, particularly when it originates from fake accounts with familiar influencer profile pictures’ (independent journalist, Garowe). These accounts enable narrative laundering: extremist content stripped of explicit affiliation markers is repackaged through accounts that appear to represent mainstream community voices.

Audience Vulnerabilities

Digital harm does not affect Somali audiences uniformly. Three dimensions of differential vulnerability are particularly important.

Urban-rural: Urban audiences have better access to multiple information sources, higher average digital literacy, and greater familiarity with the dynamics of content manipulation. Rural and nomadic communities receive information primarily through forwarded voice notes, short videos, and clips shared through trusted family or community networks. An independent journalist described receiving calls from remote village contacts who had accepted as true a falsely reported road accident death toll circulating on Facebook. In communities where independent verification is structurally unavailable, trust-by-kinship is the primary information filter, and this trust is specifically exploited by harmful actors. It is important to note that all evidence on rural community vulnerability in this report comes from practitioner-observers rather than from community members directly: the rural experience of digital harm is described here as practitioners understand it, not as communities have reported it themselves.

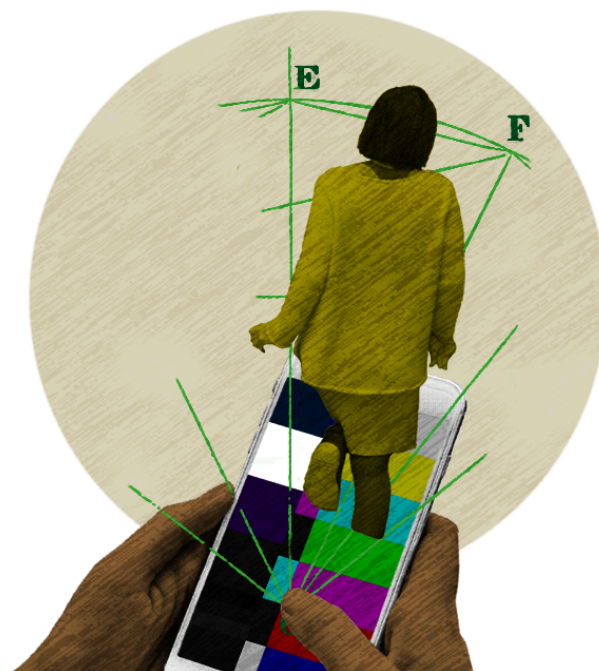
¹¹ The research did not include interviews with diaspora-based respondents. Findings on diaspora dynamics draw on accounts from Somalia-based interviewees and multi-stakeholder convening participants, supplemented by desk review. Dedicated diaspora-community research is recommended for a subsequent phase.

“I was in Hargeisa while my family was in Galkayo. My parents became very worried because they were following false information online. It took me a long time to convince them that some of the news was false and that the situation being shared on social media was exaggerated or fabricated.”

Academic respondent, Mogadishu

Gender: Women face a qualitatively different harm profile from men. While male journalists, politicians, and public figures face harassment and intimidation, the specific harms targeting women combine sexual exploitation, reputational destruction, family-level social consequences, and physical risk in ways that do not parallel the harm profile for men. The Somali Women Journalists Network’s 2022 survey found that more than 60 per cent of women in public-facing roles had received threats through social media. Many responded by reducing online engagement, self-censoring, or withdrawing from professional activities entirely.

Literacy and age difference: Young people, particularly those under 25, are the heaviest platform users and most exposed to TikTok’s algorithmically curated content, but also the least likely to have established verification practices. Older populations are more likely to believe misinformation because digital verification is less familiar; the media association respondent described elderly users who trusted whatever appeared on their feeds from accounts claiming institutional authority.



DIGITAL HARMS IN SOMALIA

The research identifies five primary categories of digital harm operational in Somalia. These categories overlap and interact: the same piece of content may constitute disinformation, hate speech, and TFGBV simultaneously, and escalation from one category to another is a common pattern. The taxonomy is therefore a practical organising tool, not a rigid classification system. The findings in this section are organised around these categories.

- I. Disinformation and Misinformation – False or misleading content produced with political, clan-based, or commercial intent (disinformation), and false content produced through errors or inadequate verification (misinformation). Includes AI-generated synthetic content and verification culture collapse.
- II. Hate Speech and Identity-Based Incitement – Content that attacks individuals or communities based on clan identity, ethnicity, religion, or gender, including content that encourages violence, retaliation, or mobilisation against targeted groups.
- III. Tech-Facilitated Gender-Based Violence (TFGBV) – Digitally mediated harms targeting women and girls, including sextortion, non-consensual intimate image sharing (Qarxis), coordinated harassment, reputational destruction, and content designed to silence women from public life.
- IV. Intimidation, Doxxing, and Censorship of Journalists and Critics – Targeted exposure of personal information, coordinated harassment campaigns, and legal or extrajudicial pressure against journalists, activists, political opponents, and critics.
- V. Cyber Intrusion, Fraud, and Threats to Public Digital Infrastructure – Hacking, SIM-swap fraud, social engineering scams, mobile money fraud, data breaches, and attacks on state digital systems.

I. Disinformation and Misinformation

Disinformation and misinformation are the most pervasive categories of digital harm in Somalia, present across all political contexts, all regions, and all platforms. They are also the category that interacts most directly with clan dynamics and political competition, making them structurally difficult to address.

The Political Economy of Disinformation

Respondents consistently described disinformation in Somalia as a political product rather than an information accident. During elections, political transitions, constitutional disputes, and security crises, political actors, including government institutions, deliberately manufacture false or misleading content to delegitimise opponents, shape public interpretation of events, and mobilise clan support. The multi-stakeholder convening in Mogadishu explicitly identified government actors as sometimes direct sources of disinformation, operating troll and bot networks and coordinating attacks against critics.¹²

¹² Multiple KII respondents, April–May 2026; Digital Harms Convening, 27/5/2026, Mogadishu.

The infrastructure for paid disinformation is well documented across respondents. Political narratives typically begin in WhatsApp coordination groups established by political media officers, from which they are simultaneously distributed to networks of influencers who release them publicly in a coordinated burst. Timing is itself the primary detection signal: fifteen to twenty accounts posting highly similar framing within the same hour is a reliable indicator of coordination. A government technology adviser acknowledged the political constraints on government intervention: ‘In many cases, people say highly inflammatory or harmful things online, but the government avoids intervening because of fears of political backlash or the possibility of triggering clan disputes’ – an unusually candid admission that non-enforcement can be a deliberate political choice, not merely a capacity limitation.

Independent election monitoring during the Banadir local elections found that harmful content followed a distinctive temporal pattern: disinformation was moderate to high in the weeks before polling, dropped slightly during voting itself, then rose sharply in the aftermath as it is amplified by both government-aligned and opposition communicators, with some activity linked to accounts outside Somalia. This pattern suggests that election monitoring and rapid-response capacity should focus particular attention on the pre-election and post-results windows, not only on election day itself.

The Collapse of Verification Culture

Beyond deliberate disinformation, respondents documented a broader epistemological shift that they described as the collapse of verification culture. Historically, Somali social norms required that information be accompanied by attribution – ‘who told you this?’ was a standard response to a new claim. That norm has eroded. As a government technology adviser observed: ‘In recent years Somalis have increasingly relied on statements such as “it was said” or “people have said” without questioning the source of information. Today, if you question the accuracy of information, people may instead attack you personally.’ This shift is distinct from a simple lack of verification skills: it represents a social norm change in which questioning information has become socially costly. The pattern is reinforced by platform dynamics that reward speed over accuracy and by content creators who add speculative framing to breaking news to be first.

AI-Generated Disinformation

AI-generated content represents the most rapidly escalating sub-category. Confirmed cases include an AI-manipulated audio recording of a prominent Puntland politician, originally delivering a standard speech, being repackaged to make it sound as though he was calling for the destruction of Mogadishu. A regulatory respondent observed that ‘many people believed it immediately, and some continue to believe it is genuine to this day.’ The Somali Journalists Syndicate issued a formal warning in late 2025 that even professional journalists could not reliably detect AI-generated audio. AI-generated satirical and mockery content, fabricated images of politicians in humiliating scenarios, manipulated videos of officials singing or dancing, has normalised as a daily feature of Somalia’s online political discourse,

with a government technology adviser estimating that approximately a third of political content on major platforms now involves AI assistance.¹³

A particularly concerning dynamic is what respondents described as ‘AI as plausible deniability’: the post-hoc attribution of genuine harmful content to AI fabrication in order to escape accountability. Once AI-generated content becomes normalised, any damaging recording, image, or video can be dismissed as ‘probably AI’ – creating a new category of epistemic harm that undermines accountability even for authentic evidence.¹⁴

A second and analytically distinct AI risk — less discussed but equally significant — is the use of AI as an accountability-evasion mechanism. Multiple respondents described a pattern in which real harmful content (an authentic recording, a genuine statement, an actual incident) is retrospectively claimed to be AI-generated in order to deny responsibility or create sufficient ambiguity to prevent consequences. This is structurally different from AI-generated disinformation: it does not require the content to be fake. It requires only that the claim of fabrication be plausible enough to muddy the evidentiary record. As AI becomes normalised in public discourse, this denial dynamic is likely to degrade the credibility of authentic evidence and create a ready-made defence for perpetrators of online harm. Somalia's judicial system — already lacking in digital forensics capacity — is not currently equipped to distinguish authentic from AI-generated content in a legally meaningful way.

“Harmful narratives move from private groups to public platforms through screenshots, forwarded messages, downloads, reposting, and screen recordings. A rumour may start in a Telegram or WhatsApp group, then be shared on Facebook pages, TikTok videos, YouTube discussions, or X posts.”

Media trainer, Puntland

The Content Lifecycle and the Hydra Problem

A consistent finding across all respondents is that platform takedown, even when it occurs, arrives too late to prevent the primary harm. A narrative typically originates in WhatsApp or Telegram groups, where it is shaped and emotionally amplified before being released to public platforms. Once on TikTok or Facebook, it spreads through reposts, screenshots, voice-note summaries, and clipped videos. By the time a platform removes the original post, it has already fragmented into hundreds of derivative forms; a dynamic described by an academic respondent as the ‘content hydra’ problem: ‘Removing the original post does little

¹³ Interview with Regulatory Respondent (NCA), May 2026, Mogadishu; Interview with Technical Cybersecurity Respondent (DPA, Hormuud), May 2026, Mogadishu.

¹⁴ Interview with Technical Cybersecurity Respondent (DPA, Hormuud), May 2026, Mogadishu. The respondent described the Abdirahman Faroole case in which an audio recording was AI-manipulated to appear as though he was calling for violence against Mogadishu.

to suppress the narrative because it has already reproduced in hundreds of derivative forms.’ The most consequential phase of the lifecycle often occurs in the private messaging channels where content is reinterpreted with stronger emotional and clan-based framing; what the regulatory respondent described as ‘the second life of content.’

“By the time harmful content is removed, many people have already downloaded it, taken screenshots, or forwarded it to other groups. The removal of the original post achieves very little — the content lives on in personal archives, private groups, and re-upload cycles.”

Academic respondent, Mogadishu

II. Hate Speech and Identity-Based Incitement

Hate speech in Somalia is inseparable from clan identity. Content that targets individuals or groups based on clan affiliation; portraying them as enemies, weak, criminal, or treacherous, is the most prevalent and the most consequential form of identity-based incitement. It functions as both a political weapon and a commercial product, generating the emotional engagement metrics that drive platform reach.

Clan Incitement as Political Weapon

Political actors across Somalia use clan-coded narratives to mobilise supporters, delegitimise opponents, and frame political disputes as existential conflicts between communities. The specific ‘weak vs strong clan’ narrative mechanic, portraying one community as victorious and another as humiliated or defeated, is particularly effective at provoking reaction from the targeted community and encouraging mobilisation. This content often originates in diaspora networks before entering Somalia's domestic information ecosystem, where it is amplified through TikTok lives and Facebook posts.¹⁵

Election periods and constitutional disputes generate the sharpest spikes in hate speech. The 2023-2024 Puntland democratisation process created a period of sustained public confusion that political actors exploited: without public understanding of how party systems worked, narratives framing the process through clan competition spread rapidly and unchecked. Multiple respondents described Puntland-specific triggers: the 2021 senators’ election, the 2023-24 democratisation tensions, and the Garowe conflict involving a clan militia attack on the government as sequential moments in which online hate speech escalated to documented offline violence.

¹⁵ Multiple KII respondents, April–May 2026; Digital Harms Convening, 27/5/2026, Mogadishu.

A specific case from Somaliland illustrates the diaspora mechanism clearly. In El Afweyn district in the Sanaag region, diaspora members were reported to have played a significant role in fuelling intercommunal tensions, providing financial support to armed actors, amplifying clan-based narratives online, and mobilising resources that made local peacebuilding and mediation efforts considerably more difficult. Large sums of money were raised through diaspora networks, and the sustained online amplification increased public pressure and emotional intensity in ways that reduced opportunities for calm dialogue and reconciliation (media trainer, Garowe). This Somaliland case complements the Laascaanood and Galkayo evidence in demonstrating that the diaspora amplification pattern is consistent across geographic and administrative contexts.¹⁶

“In El Afweyn, diaspora members played a role in fuelling tensions by providing financial support and amplifying clan narratives online. Large amounts of money were mobilised from diaspora communities, which contributed to prolonging disputes and making local peacebuilding and mediation efforts more difficult.”

Academic respondent, Mogadishu

The TikTok Algorithm and Chronic Exposure

Several respondents emphasised TikTok’s specific role in systematically deepening exposure to clan-coded and politically incendiary content. ‘Its algorithm selects the clips you watch, and once you watch them, it continues feeding you hate clips. Imagine a boy or a girl growing up with such a negative mindset? The problem is chronic’ (media association respondent). This algorithmically enforced exposure is particularly significant for young audiences who may encounter clan antagonism through TikTok before they encounter it through any other social context, and who lack the established cross-clan relationships that older generations used as verification anchors.



¹⁶ Interview with Media Practitioner, May 2026, Garowe.

Narrative Laundering and Extremist Content

Al-Shabaab and other extremist actors contribute to the hate speech ecosystem through a specific mechanism that respondents described as ‘narrative laundering’: explicit extremist content is stripped of its organisational markers and reintroduced into mainstream discourse through religious framing, community grievance narratives, and anonymous accounts using the photographs of respected religious leaders or community figures. ‘Some individuals intentionally reinterpret or redistribute these narratives in different forms in order to influence public opinion. In some cases, people with political motives may repackage extremist narratives using different language, labels, or messaging styles to make them more acceptable or appealing to wider audiences’ (government technology adviser).

One technically informed respondent described Al-Shabaab’s apparent shift from public platform distribution, which was increasingly disrupted by a government task force that removed approximately 4,500 pieces of extremist content over three years, toward one-to-one private channel communication through WhatsApp and Telegram. This is a finding that would require confirmation from security-sector or platform-level sources to be treated as confirmed, but if accurate, it would effectively evade the content removal architecture that has been developed to address it, requiring a fundamental rethinking of counter-extremism strategy.

III. Tech-Facilitated Gender-Based Violence (TFGBV)

TFGBV is simultaneously the most severe category of digital harm in Somalia and the most systematically underreported. It operates through multiple overlapping mechanisms and produces consequences that extend far beyond the digital sphere into family relations, physical safety, mental health, and women’s participation in public and economic life.

The Somali Context: Compounding Factors

TFGBV in Somalia operates within a specific cultural and social context that amplifies its severity beyond what comparable experiences in higher-trust institutional settings would produce. Three Somali-specific factors compound the harm and must be understood before the mechanisms can be properly assessed.

The honour and reputation system: In Somali culture, a woman’s social standing is deeply intertwined with family and community perceptions of her moral conduct. Content that compromises her reputation, whether true, fabricated, or manipulated, can cause irreversible social damage that extends beyond her own life to affect her family’s position, her children’s prospects, and her marriage or livelihood opportunities. The harm is not proportionate to the content’s accuracy; fabricated or manipulated material causes the same reputational injury as genuine material, sometimes more, because there is no corrective institutional mechanism.

Family-level consequences of shaming: When intimate or shaming content about a woman circulates online, the response of family members is often to withdraw her from public life, removing her from education, employment, or civic roles, in order to manage community

perception. The harm to the woman is thus compounded by the family's protective response. What looks from outside like a retreat is often a forced withdrawal managed by people who love the victim but prioritise community standing over her individual agency.

Structural anonymity barriers: Somalia's small social world, dense clan networks, and limited formal institutional trust mean that reporting or seeking help is structurally more difficult than in societies with functioning anonymous reporting systems. A woman who reports abuse risks her case becoming known throughout her community and clan network; a second layer of exposure that deters disclosure even when mechanisms nominally exist. In a context where everyone knows everyone, anonymity is not simply a preference but a structural impossibility for most women.

The Access vs Safety Gap

A digital safety practitioner articulated the central problem precisely: the challenge for women in Somalia's digital environment is not access to technology but safety within digital spaces. Women are able to reach platforms, but the hostile environment those platforms create, through coordinated harassment, sextortion, reputational attack, and non-consensual intimate content, means that meaningful participation is consistently unsafe. This framing – the 'access vs safety gap' – represents a fundamental reorientation of the digital inclusion agenda: connectivity targets have largely been met for urban women, but safety infrastructure has not been built alongside it.

“The primary challenge is not access to technology but safety within digital spaces. Women are able to access digital platforms, but the hostile online environment often discourages meaningful participation and limits their ability to engage freely in public discourse.”

Digital safety practitioner, Mogadishu

Mechanisms of TFGBV

Sextortion and Qarxis: The practice known in Somali as Qarxis – the exposure or threatened exposure of intimate or identifying material to destroy reputation or coerce compliance – is the most severe documented form of TFGBV. Perpetrators establish relationships with victims, sometimes romantic, sometimes professional, record intimate situations without consent, and use the recordings to threaten victims into sending money, providing further material, or complying with other demands. A technical cybersecurity respondent described the pattern: ‘Some perpetrators establish relationships with victims, sometimes involving sexual relationships, and then secretly record intimate situations in order to later manipulate, threaten, or shame the victim.’¹⁷

¹⁷ Interview with Digital Safety Practitioner, May 2026, Mogadishu; Interview with Regulatory Respondent (NCA), May 2026, Mogadishu.

Institutional GBDV – The Qardho Case: TFGBV is not confined to personal or romantic relationships. A case documented by a media practitioner in Puntland illustrates how harm can extend into institutional and professional settings: women receiving medical care at a hospital facility in Qardho were reportedly filmed without consent by individuals with access to the premises, and footage was subsequently used for coercive purposes. This case illustrates how digital harm can occur within spaces that should be safe, like health facilities, workplaces, public services, and how institutional failure to protect privacy creates harm vectors that conventional digital safety awareness does not address.

Coordinated harassment campaigns: Women journalists, politicians, and activists face coordinated harassment that combines personal abuse, sexual insults, threats of violence, manipulation of images, and sustained reputational attacks. These campaigns are often executed through networks of anonymous accounts and amplified by influencer networks, creating an experience of being attacked from multiple directions simultaneously. The Somali Women Journalists Network documented that more than 60 per cent of women in public-facing roles had received social media threats.

Monetised Sexual Violence on Telegram

Beyond individual harassment, a more organised form of TFGBV has been documented: the commercialisation of non-consensual intimate content through Telegram’s subscription-based channel structure. The women’s media practitioner described channels that ‘charge subscription fees, sometimes as little as one dollar, to access videos and images involving well-known individuals, private citizens, and members of marginalised communities, including LGBT individuals.’

These channels represent a categorically different harm level from individual harassment or opportunistic content sharing. They constitute commercial enterprises built on the distribution of sexual violence. The business model, small per-user fees generating income from large subscriber bases, creates persistent economic incentives for the production and distribution of harmful content. Unlike single acts of non-consensual image sharing, subscription channels are structured, recurring businesses that require active management and generate stable income. The subscription model means that once content is uploaded, it continues to circulate and generate revenue regardless of what the perpetrator does afterward; making effective intervention structurally more difficult.

The targeting of LGBT individuals is particularly significant: it illustrates that TFGBV in Somalia’s digital ecosystem extends beyond the silencing of women in public life to encompass exploitation of all groups that face social vulnerability. In a context where same-sex relationships are criminalised, the threat of exposure carries physical as well as reputational danger.

Consequences: Beyond Individual Wellbeing

The consequences of TFGBV extend systematically beyond individual wellbeing into Somalia's broader social, political, and economic fabric. Women who experience online harassment reduce their online presence, self-censor their views, avoid political discussions, disengage from civic activities, and in extreme cases withdraw from professional life entirely. These individual withdrawals aggregate into structural distortions with serious development implications.

“Women are able to access digital platforms, but accessing them safely is a different matter. The shame and the fear are bigger barriers than the technology itself.”

Digital safety practitioner, Mogadishu

In a political context where women's representation is already challenged, the systematic silencing of women who seek to participate in public discourse narrows the diversity of perspectives available to Somalia's democratic processes. Civic life, electoral politics, and community decision-making are impoverished when women are absent from online spaces where these discussions now occur. In the media sector, the targeting of women journalists and content creators reduces editorial diversity and creates self-censoring dynamics that affect coverage quality and public accountability. Economically, the 'access vs safety gap' means that women are effectively excluded from the digital commerce, remote work, and platform-based income opportunities that the same digital expansion has opened for men; a gender-differentiated economic consequence of digital harm that is rarely captured in development assessments or digital inclusion frameworks.

Under-Reporting and the Scale Problem

The scale of under-reporting means that the harms documented in this research almost certainly represent a small fraction of actual TFGBV affecting Somali women and girls. When cases are not reported – because of shame, fear of victim-blaming, family pressure, and distrust of authorities – they do not enter any monitoring system. Scale assessments based on reported cases therefore radically undercount the problem. This matters directly for resource allocation: if the true prevalence of sextortion, image-based abuse, and coordinated harassment is five or ten times higher than the number of cases that surface through formal or informal support channels, the current level of response infrastructure is severely inadequate. Practitioners and donors should treat reported case volumes as a floor, not an estimate, and fund accordingly.

Reporting is rare. Shame, fear of victim-blaming, distrust of authorities, and low expectations of meaningful response are the primary barriers. Even where women seek help, formal mechanisms are absent: Somalia lacks legal provisions specifically addressing gender-based digital violence, image-based abuse, or coordinated online harassment. Many cases are

handled through family or clan networks; channels that may not prioritise the victim's interests and that may themselves amplify shame or pressure victims into silence.

Available Support and the Service Gap

What survivors currently have access to, and what is structurally absent, shapes the harm experience significantly. Informal family and clan networks remain the primary recourse for most women, though these channels may not prioritise the victim's interests or guarantee confidentiality. Digital Shelter, a Mogadishu-based NGO, provides dedicated digital safety support for individuals experiencing online abuse, including case management, referral pathways, and emergency digital security assistance, and is the only specialised mechanism of its kind in Somalia. For women who know it exists and can physically access it, it offers a meaningful alternative to family-mediated resolution. But for most women outside Mogadishu, and for many within it, Digital Shelter's services remain unknown or inaccessible.

Legal aid for TFGBV cases, police support with the mandate and technical capacity to handle digital abuse, confidential reporting pathways, and any formal institutional support system that treats digital violence as a crime rather than a private social matter are all absent. The result is that victims choose between an informal system that may re-victimise them, a single NGO with limited geographic reach, and silence. Given the structural barriers to disclosure described above, silence is the overwhelmingly common choice.

IV. Intimidation, Doxxing and Censorship

Journalists, online commentators, civil society activists, and political opposition figures face a documented pattern of digital harm in Somalia that combines online harassment, legal weaponisation, content suppression, and in some cases physical intimidation. This harm category sits at the intersection of digital dynamics and the broader press freedom context.

Between April 2023 and March 2025, the Somali Journalists Syndicate recorded 206 separate incidents against journalists. NUSOJ documented that in 2023, approximately 87 per cent of attacks on journalists were attributed to police forces – a striking figure that identifies security services as the primary enforcement mechanism against the media, rather than a neutral guarantor of their safety. The pattern intensified: NUSOJ documented 41 violations of press freedom in the first half of 2025 alone, the highest mid-year figure in three years, alongside 69 attacks on individual journalists across the full year.¹⁸

Coordinated Harassment as an Intimidation Tool

Respondents with direct journalistic experience described coordinated harassment campaigns targeting individuals who publish content that challenges political actors, exposes corruption, or covers sensitive security issues. A former media association respondent

¹⁸ National Union of Somali Journalists and Somali Journalists Syndicate reporting on press freedom violations and journalist incidents 2023–2025; EUAA country focus data on journalist attacks. <https://nuso.org>; <https://cpj.org/africa/somalia/>

described personally receiving approximately 60 harassing messages within hours of publishing a report for a major outlet. ‘Many of the accounts involved used photographs of real people, despite not belonging to those individuals. These accounts function as critical amplification nodes because they rapidly distribute content across networks and create the appearance of widespread public opposition.’ This coordinated use of impersonation accounts, fake accounts using stolen photographs of real, identifiable individuals to create false impressions of community consensus against a target, is a distinctive and underappreciated mechanism in the Somali harassment ecosystem.

Doxxing and Personal Information Exposure

The targeted exposure of personal information such as home locations, phone numbers, family details, private communications, and intimate content, is used to intimidate individuals into silence and to mobilise community-level harassment against targets. A government technology adviser described a normalised pattern of governance documents like appointment letters, official communications, and even diplomatic cables, circulating on social media platforms before their intended recipients have seen them. While this specific form of leak involves institutional data rather than personal intimate content, it reflects the same underlying absence of data privacy norms and enforcement mechanisms that enables personal doxxing.



Weaponisation of Regulation

A consistent and disturbing pattern documented across multiple respondents is the deployment of legal and regulatory mechanisms against journalists and critics rather than against genuinely harmful actors. Respondents described arrests of journalists for content criticising political leaders, coverage of security incidents, or reporting on protests. NUSOJ documented that police forces were responsible for the majority of 2025 arrests in contexts involving security reporting, protests, and land eviction coverage. In March 2025, nineteen

journalists covering an Al-Shabaab attack were arrested, their equipment confiscated and footage deleted.

A regulatory respondent downplayed weaponisation as ‘not a significant pattern at the moment’ – a characterisation that contrasts sharply with the accounts of civil society respondents and journalists.

V. Cyber Intrusion, Fraud, and Threats to Public Digital Infrastructure

The digitalisation of Somalia’s public services, financial systems, and governance processes has expanded the attack surface for cyber intrusion and fraud. This category of harm has received less public attention than disinformation or TFGBV, but it has direct consequences for institutional trust, state legitimacy, and economic security.

Somalia’s digital ambitions are expanding: in February 2026, the government hosted the National Digital Public Infrastructure Conference in Mogadishu, with an explicit focus on connected digital systems, measurable service delivery, and improved public services. The more Somalia digitises, the more urgent robust data protection, cybersecurity architecture, and public trust become. The e-visa breach documented below illustrates what is at stake when infrastructure expands faster than security.

The E-Visa Data Breach

The most consequential documented case of public digital infrastructure failure is the November 2025 breach of Somalia’s e-visa platform, through which the personal data of an estimated 35,000 visa applicants, including passport scans and biometric information, was exposed. The government’s response was delayed by approximately 45 days: ‘Many people considered the issue normal and did not fully understand the seriousness of the leak. It took the government approximately one and a half months to publicly acknowledge the failure and respond to the issue’ (technical cybersecurity respondent).

A second area of data protection concern raised in interviews is the biometric voter registration system used in Somali elections. KII respondents noted that the collection of biometric data at registration creates a dataset that, if inadequately protected, could be misused for targeting, voter suppression, or identity fraud. The absence of an operational Data Protection Authority means there is currently no independent body verifying the security and data governance standards applied to voter data.¹⁹

Several embassies issued a security alert citing the breach. The third-party company managing the system continued to operate it after the breach was identified, raising fundamental questions about procurement standards and vendor accountability. This incident illustrates what happens when Somalia’s rapidly expanding digital public infrastructure

¹⁹ Interview with Regulatory Respondent, May 2026, Mogadishu; Digital Harms Convening, 27/5/2026, Mogadishu.

outpaces the regulatory oversight, data protection enforcement, and institutional awareness needed to secure it.

Mobile Money Fraud and Social Engineering

Somalia's dependence on mobile money – over 70 per cent of adults use mobile money as their primary financial tool, and EVC Plus facilitated \$914 million in aid transfers between 2019 and 2024 – makes fraud vectors in this space a matter of social welfare, not merely commercial risk. Social engineering scams in which individuals are manipulated into sharing authentication credentials or OTP codes, SIM-swap attacks enabling fraudulent account takeovers, and false promotional schemes targeting digitally inexperienced users are documented patterns. These harms disproportionately affect populations with lower digital literacy, and are largely invisible in formal reporting because victims lack mechanisms to report effectively and rarely trust that doing so will produce results.

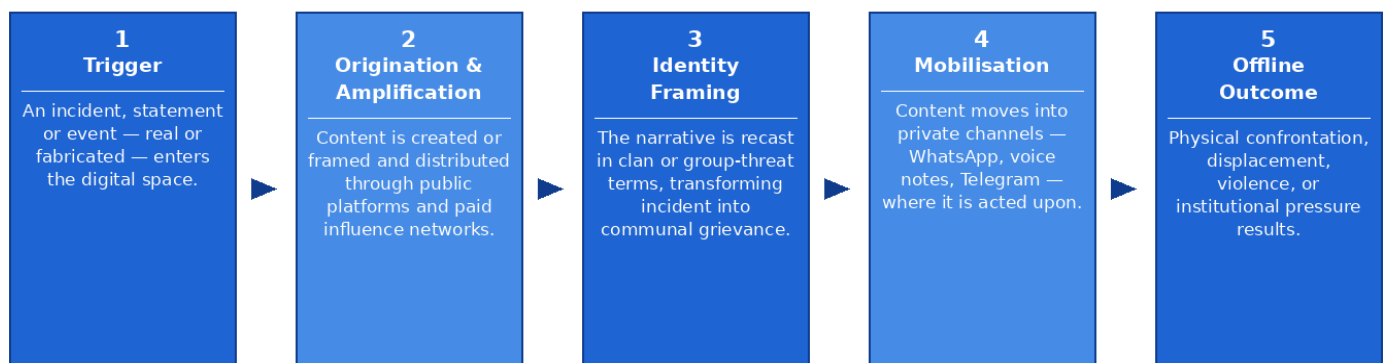
Judicial Incapacity for Digital Crime

A cybersecurity expert described multiple cases in which digital crimes that reached court were resolved through means other than the law; not because evidence was absent, but because judicial actors lacked the technical understanding to adjudicate them. In one case involving alleged telecommunications hacking, the accused argued that he had 'used the air' in the same way the company does, and the court, lacking technical expertise to counter this framing, referred the case to traditional elders for resolution. In another, a Kenyan national alleged to have attempted to infect a bank with malware via a USB device was acquitted because no adequate legal provision existed under which to charge him. These cases illustrate a fundamental judicial capacity gap: Somalia's courts cannot currently adjudicate digital crime cases even where evidence is available and prosecutorial will exists.

ESCALATION DYNAMICS

Five Stage Escalation Model

Across the five harm categories, a consistent escalation pathway emerges from the research: harmful digital content moves from online narrative to offline consequence through five recognisable stages. This model applies to all five harm categories and is illustrated through specific case studies in the following section. Understanding the model is essential for designing effective interventions: interrupting the escalation at Stage 2 or 3 is far more effective than responding at Stage 5, when harm has already materialised offline.



1. Stage 1 – Trigger: A political event, security incident, arrest, election result, or local conflict creates information demand. The trigger may be real, exaggerated, or entirely fabricated, but it activates public interest and creates space for harmful content.
2. Stage 2 – Origination and amplification: Content originating in private WhatsApp or Telegram groups is shaped, emotionally charged, and then released to public platforms. Influencer networks and diaspora actors amplify the narrative to large audiences. This is typically where paid coordination occurs.
3. Stage 3 – Identity framing: The narrative is reframed through clan identity – positioning one community as threatened, humiliated, or under attack, and another as responsible. This is the critical escalation step: it transforms information into mobilisation pressure and makes the narrative resistant to correction.
4. Stage 4 – Mobilisation: Emotional, identity-based content activates community response. Online calls to support a clan, fundraise for a conflict, attend a protest, or retaliate against perceived aggressors translate into real-world action. At this stage, local mediators begin to lose the initiative.
5. Stage 5 – Offline outcome: Physical confrontation, clan violence, displacement, protest, or harassment follows. In several documented cases, local mediation efforts were made more difficult or impossible by the sustained online amplification of the conflict.

The harms documented in the previous section can each escalate through this pathway. The following section provides detailed case studies of escalation in practice, examining

Laascaanood, the Bajaj driver case, and the longitudinal evidence from Kismayo and Lower Shabelle.

Case Studies of Escalation

Laascaanood (2023–2024)

The conflict in Laascaanood is the most extensively documented case of digital-to-offline escalation in the dataset, confirmed across multiple independent respondents. Diaspora communities amplified clan narratives about the conflict, framed as a liberation struggle or a territorial defence depending on clan affiliation, through TikTok, Facebook, YouTube, and WhatsApp. Online fundraising for clan militias was conducted through diaspora networks. Influencers published content designed to provoke emotional reactions and encourage mobilisation. Local mediation efforts were repeatedly complicated by the sustained online amplification of grievances, which arrived back in Laascaanood faster than mediators could work.²⁰

The women's media practitioner offered an analytically important nuance that the harm-focused account should not obscure: diaspora investment in Laascaanood appears to have subsequently transitioned, at least in part, from conflict support to economic and state-building investment. Infrastructure projects, business ventures, and remittance channels initially mobilised around the conflict have in some cases persisted as constructive engagement after hostilities subsided. This suggests that diaspora engagement is not irreversibly harmful: the financial and relational networks that sustain conflict amplification could, under the right conditions, be redirected. Practical channels for constructive diaspora engagement – formal remittance mechanisms linked to development priorities, diaspora investment platforms tied to infrastructure projects – may represent a meaningful policy opportunity for those working on conflict prevention and peacebuilding.

The Female Bajaj Driver Case

A technical cybersecurity respondent described in detail a case in which a female Bajaj (auto-rickshaw) driver became associated with anti-government protests and began spreading government-critical narratives online. Opposition figures reportedly encouraged and financially supported her to continue. She was detained, released after three days, travelled to Baidoa to publicly support an opposition figure against the federal government, and was detained again upon return to Mogadishu. Simultaneously, opposition actors circulated online narratives claiming the government intended to harm her, creating reputational pressure that complicated the government's calculation about her release. This case illustrates several dynamics simultaneously: the monetisation of political opposition through digital platforms; the use of social media to create political pressure around individual cases; and the real-time interaction between online narrative management and government decision-making. The

²⁰ Corroborated across: Interview with Media Practitioner, May 2026, Garowe; Interview with Academic Respondent, May 2026, Mogadishu; Interview with Community Respondent, May 2026, Garowe; Digital Harms Convening, 27/5/2026, Mogadishu.

respondent noted that online narratives claiming the government intended to assassinate the driver created a situation in which releasing her quickly became politically risky for the government: public accusation of planned harm meant any subsequent incident — even unrelated — risked being attributed to state action. This suggests the online narrative itself functioned as a form of protection and complication simultaneously, potentially extending her detention.²¹

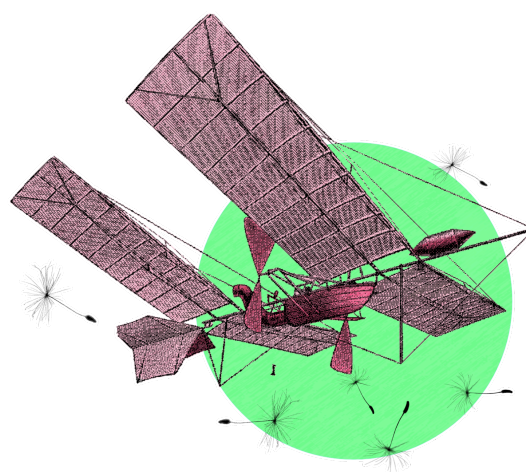
Kismayo (2017) and Lower Shabelle (2021)

Two earlier cases, cited with academic sourcing by a media practitioner in Garowe, provide longitudinal evidence that the online-to-offline escalation pattern predates the current generation of platforms. In Kismayo in 2017 and Lower Shabelle in 2021, digitally amplified clan narratives contributed to confrontation, displacement, and violence. These cases confirm that the escalation mechanism is structural rather than specific to current platforms or political actors; it has been a feature of Somalia's information environment since social media penetration began.

The Diaspora Feedback Loop

The diaspora's role in escalation dynamics is distinctive enough to merit separate analytical treatment. Diaspora actors are not simply amplifiers of local content: they often originate harmful narratives, fundraise for armed actors, provide financial incentives that sustain influencer campaigns, and operate from jurisdictions beyond Somali regulatory reach. A regulatory respondent described the mechanism precisely: 'Local events produce digital reactions, those reactions are amplified sometimes by diaspora voices, and the amplified version returns to the local context with more pressure, more emotion, and sometimes more danger.'

The independence of diaspora actors from local consequences is analytically significant. A community member in Garowe who publicly amplifies a clan grievance may face social pressure from neighbours, community leaders, or family members to moderate their position. A diaspora actor in the UK or US faces no equivalent local accountability and may in fact gain social status within their diaspora community through the same amplification. This structural asymmetry contributes to the diaspora's disproportionate role in escalation dynamics and explains why the feedback loop is so difficult to interrupt from within Somalia.



²¹ Interview with Technical Cybersecurity Respondent (DPA, Hormuud), May 2026, Mogadishu.

LEGAL AND INSTITUTIONAL FRAMEWORK

Somalia's legal and regulatory environment for digital governance is evolving but enforcement remains almost entirely absent. The gap between what is written in law and what happens in practice is the defining feature of this environment, and it is understood as such by every respondent in this study.

Instrument / Institution	Status	Key Gap
Cybercrime Bill	Approved by Cabinet; not yet enacted	No enforcement mechanism in place
Data Protection Authority	Established in law; not yet fully operational	No operational oversight of data processing
National Communications Authority	Operational	Limited jurisdiction over content; focus on licensing
Media law & journalist protections	Partial; inconsistently applied	Selective enforcement against critics of the state
Platform engagement	Minimal	No Somali-language content policy coordination

Legal Framework

The most significant recent legal developments are the Data Protection Act 2023 (signed into law in March 2023) and the Cybercrime Bill approved by Cabinet in August 2025. The Data Protection Act establishes a Data Protection Authority, sets out data subjects' rights, and requires controllers to notify the DPA of breaches within 72 hours. The Cybercrime Bill – which had not been enacted into law at the time of research – is intended to address cyberbullying, digital abuse, hacking, theft, and attacks on institutional digital systems. The National Cybersecurity Law (2025) establishes the NCA as the lead cybersecurity governance authority and SOMCIRT as the national incident response centre.

The regulatory landscape is in active development and findings in this section should be interpreted accordingly. The Cybercrime Bill had been approved by Cabinet but not yet enacted; the Data Protection Authority had been established in law but was not yet operationally active. These institutional configurations may change rapidly as laws are enacted and institutions are staffed, and some findings may need to be updated accordingly.

Additional relevant provisions include the Communications Act of 2017 (which establishes the NCA's mandate), the Anti-Terrorism Law of 2023, and initiated amendment of the Somali Penal Code, which introduces Chapter 10 provisions on online content. Puntland operates its own Puntland Media Law (last amended 2014–2016) and recently established the Puntland Media

Council. A regional comparator is instructive: Somaliland's Cybersecurity and Cybercrime Bill (2023) includes under Article 42 a specific definition of social media influencers, anyone with more than 5,000 followers, with associated obligations and penalties for spreading false information. While the provision's scope and proportionality remain contested, its existence demonstrates that a legislative framework addressing the influencer ecosystem specifically is both possible and precedented within the region. Somalia's federal-level frameworks currently contain no equivalent definition or obligation.

The most sensitive concern is selective enforcement. The evidence should be framed carefully since respondents across several categories perceived legal and security tools as more likely to be used against journalists, critics or politically inconvenient voices than against harmful content aligned with powerful interests. This perception itself matters because it reduces trust in regulation. If people believe regulation is political, they will be less likely to report harm or support institutional intervention.

Critical Legal Gaps Identified by Respondents

1. No specific legal provisions addressing gender-based digital violence, image-based abuse, or coordinated online harassment
2. No legal obligation for influencers to disclose sponsorship, payment, or political affiliation
3. The Somali Penal Code (1964) predates social media and is inadequate for digital crime
4. The Data Protection Authority has been established but is not yet operationally active
5. Courts lack technical capacity to adjudicate digital crime cases — multiple cases referred to traditional elders or resulting in acquittal due to absence of applicable legal provision
6. No cross-border legal framework for diaspora-originating harmful content
7. No legal classification of influencers, content creators, or journalism-adjacent practitioners

Institutional Architecture

Responsibility for digital harms is fragmented across multiple institutions: the NCA (communications regulation and cybersecurity), the Ministry of Communications and Technology (policy), the Ministry of Information (media governance), the Data Protection Authority (data and privacy), the Media Council (media accreditation), and security actors including NISA (counterterrorism and security-adjacent digital content). This fragmentation means that no single institution has a clear, operational mandate across all categories of digital harm, and coordination across institutions is weak.

The most significant institutional coordination achievement to date is the government task force, including the NCA, which removed approximately 4,500 pieces of extremist content over three years. This represents meaningful institutional capacity, but it is narrowly focused

on terrorism-related content and does not address the broader categories of disinformation, TFGBV, or political harassment that affect far larger numbers of Somalis.

The Enforcement Gap and Selective Application

The enforcement gap between legal framework and operational reality is acknowledged by all respondents, including government actors. The regulatory respondent identified attribution, case handling, victim support, and platform engagement as the primary implementation failures. A technical cybersecurity respondent noted that even the CID, which has received international support for digital investigation capacity, lacks proactive prevention capabilities.²²

“Individuals who publicly insult others online are often openly welcomed and celebrated. Perpetrators are rarely arrested or prosecuted. People engaged in insults or blackmail are rarely seen being held accountable.”

Government minister

A more concerning dimension is the selective application of legal tools. Multiple respondents described legal and regulatory mechanisms being used against journalists, critics, and opposition voices rather than against genuinely harmful actors. The pattern is sufficiently consistent across civil society and independent journalist respondents to be treated as a structural feature rather than exceptional abuse. The government technology adviser’s observation that government avoids intervening on harmful clan content due to political backlash or clan dispute risk sits uncomfortably alongside the documented pattern of enforcement against critics. The implication is that enforcement is selective precisely along the axis of political convenience.

²² Multiple KII respondents including government, regulatory, and civil society, April–May 2026; Digital Harms Convening, 27/5/2026, Mogadishu.

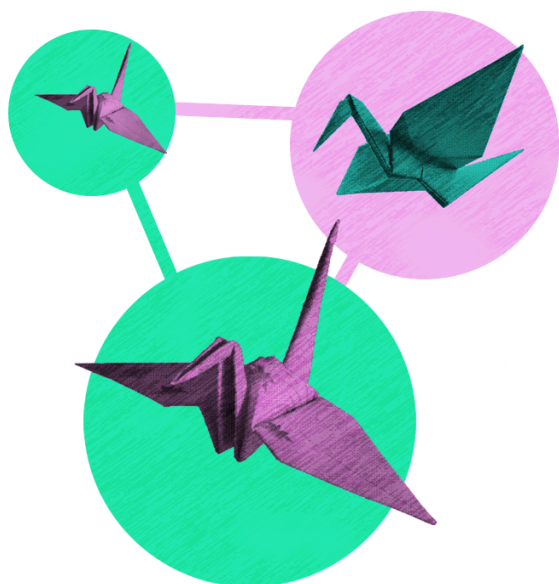
RESILIENCE & MITIGATION

Somalia's current resilience against digital harms is limited, fragile, and predominantly dependent on individual personalities rather than institutional systems. Several mechanisms operate at different levels, but none currently has the scale, speed, or reliability to address the harms documented in this study.

Trusted Messenger Correction

The most consistently identified effective counter-narrative mechanism is correction by trusted individuals, respected religious leaders, community elders, and locally known journalists, who intervene when harmful narratives begin to circulate. A government technology adviser described a concrete example: planned protests in Mogadishu were reportedly prevented because religious and traditional leaders monitored heated online debates and intervened directly to calm tensions before they escalated.

However, this mechanism has fundamental limitations. It depends on specific individuals who cannot be scaled, systematised, or deployed consistently. The correction credibility gap is structural: institutional communication – from government, from regulators, from formal media – lacks the credibility to counter narratives effectively because public trust in institutions is low. Trusted messengers are credible but cannot be everywhere at once. And once a narrative becomes tied to clan identity or moral outrage, 'even trusted messengers find correction significantly harder, particularly in private groups where messenger credibility and timing are decisive' (regulatory respondent).



An academic respondent proposed the Social Ecological Model as a more effective framework: rather than training individuals in digital literacy, strengthen community gatekeepers – elders, religious leaders, and community radio operators – so they can identify, verify, and counter harmful content within their existing authoritative roles.

This approach matches the trust structures that already operate in Somali communities and avoids creating parallel institutional systems that compete with established authority.

Formal Verification and Reporting Mechanisms

Two mechanisms — one formal, one CSO-driven — exist but operate at limited scale and public awareness. Somali Fact Check (somalifactcheck.org) is a formally structured fact-checking organisation providing verification on political, security, and misinformation claims. A second, community-driven mechanism is operated by a civil society and voluntary group without formal institutional backing. No other fact-checking sites were cited by interview respondents, suggesting a significant gap in the verification landscape. Both mechanisms face the fundamental timing constraint described in Section 10.4. exist but operate at limited scale and public awareness. Somali Fact Check (somalifactcheck.org) provides fact-checking on political, security, and misinformation claims, publishing its findings in Somali language to reach audiences beyond the English-literate elite; making it the only dedicated fact-checking organisation in Somalia’s information ecosystem. Its operational constraints reflect the wider challenges of counter-narrative work: limited staff, constrained resources, and response speeds that are inevitably slower than the platforms it monitors. During fast-moving events – an election, a security incident, a viral disinformation spike – Somali Fact Check can produce corrections within hours, but the harmful content it addresses has typically already completed its primary impact cycle before the correction reaches its audience.

“Most people place their trust in the individual who delivers the correction or actively follows up on the issue — not in any institution or formal system.”

Government minister

Cyber888, established by the NCA as a national reporting centre, allows citizens to report terrorism-related content, criminal activity, and online harm confidentially. The mechanism is still in early stages of operationalisation: the NCA respondent confirmed it has been established and is operational in principle, though public awareness remains limited. The NCA reported that the government task force, including Cyber888 infrastructure, has been involved in removing approximately 4,500 pieces of extremist content over the current administration. Whether Cyber888 has developed systematic public trust, independent from wider NCA institutional trust challenges, warrants further monitoring. Both mechanisms are valuable and merit sustained investment, but neither currently has the geographic reach, response speed, or institutional authority to address harms at the scale and pace at which they occur.

Emerging Correction Practices

There are tentative signs of emerging community-level resilience practices. The women’s media practitioner noted the emergence of trusted content creators who attempt to correct false information before it causes significant damage – ‘there are trusted content creators and influencers that attempt to correct before significant damage, and people are trying to not

jump to conclusions; this is a recent phenomenon.’ If this trend represents a genuine shift – influencers building reputations for accuracy rather than controversy – it may represent a market-based complement to regulatory and digital literacy interventions. However, it remains early-stage and fragile, dependent on the individual choices of specific influencers without structural incentives to sustain it.

The Timing Problem

The fundamental constraint on all resilience mechanisms is what respondents consistently described as the ‘timing weakness’: harmful content spreads in minutes, while institutional responses require days. The lifecycle of a harmful narrative in Somalia, from private group origination to viral public platform spread to private channel amplification, typically completes its primary impact cycle within two to four hours. No currently existing mechanism such as government takedown requests, trusted messenger intervention, fact-checking publication, or platform moderation can reliably operate within that window. Addressing this timing gap requires not just faster institutional response but structural interventions: pre-positioning trusted messengers with prepared counter-narratives for predictable trigger events, algorithmic changes to slow the spread of unverified content in sensitive political windows, and building community inoculation capacity before harmful content appears.



CONCLUSION

Somalia's digital information environment presents a paradox: the same rapid connectivity expansion that has brought mobile money, remote education, diaspora connection, and civic voice to millions of Somalis has also created the infrastructure for coordinated disinformation, clan mobilisation, systematic silencing of women, and organised harassment of journalists. These are not incidental side effects of a generally benign technology transition. They are structural features of an environment in which strong commercial and political incentives drive harmful content production, legal frameworks exist on paper but not in practice, platform companies operate without local accountability, and communities lack the literacy and institutional support to protect themselves. Six conclusions follow from the research findings.

First, the enforcement gap is not primarily a legal text problem. Somalia has enacted a Data Protection Act, approved a Cybercrime Bill, and established SOMCIRT. The gap between these frameworks and operational reality is not about the absence of law; it is about institutional capacity, judicial technical knowledge, fragmented mandates, political will, and the structural advantages that harmful actors, particularly diaspora-based ones, have over domestic enforcement systems. Passing more laws without addressing these constraints will not reduce harm.

Second, TFGBV is the most severe and most invisible harm. The silencing of women from Somalia's digital public sphere through sextortion, coordinated harassment, and reputation destruction is not a secondary concern. It is both a serious rights violation in itself and a structural distortion of public discourse, depriving political and civic processes of the perspectives of roughly half the population. The commercialisation of sexual violence through Telegram subscription channels represents an escalation beyond individual harm to organised exploitation. The absence of legal provisions specifically addressing TFGBV, combined with the cultural barriers to reporting documented in Section 7.3, means that the true scale of harm is almost certainly far larger than what appears in any monitoring data. The absence of legal provisions addressing TFGBV is the most urgent legal gap identified.

Third, the diaspora is neither simply a problem nor simply a resource. Diaspora actors are among the most significant originators and amplifiers of harmful content in Somalia's digital ecosystem, and their distance from local consequences insulates them from accountability. At the same time, the potential transition of diaspora investment from conflict support to state-building, evidenced in Laascaanood, suggests that diaspora engagement is not fixed in its harm direction. Policy that treats the diaspora purely as a threat will miss the opportunity to redirect significant financial and social capital toward constructive ends.

Fourth, platform companies are absent stakeholders with enormous responsibilities. No major platform – Meta, TikTok, YouTube, Telegram – has a local office in Somalia, operates meaningful Somali-language content moderation, or has established formal engagement with Somali regulatory institutions. Platform decisions about algorithm design, content moderation standards, monetisation of controversial content, and data-sharing with regulators have enormous consequences for Somali communities. Without formal engagement mechanisms,

these decisions are made without reference to Somali context, law, or community needs. Research published in 2025 on transformer-based language models applied to Somali-language social media content demonstrates that automated detection of fake news and toxic messages in Somali is technically feasible, with meaningful accuracy. That research exists but is not integrated into any major platform's moderation systems, which means that the technical means to address Somali-language harmful content at scale are available, but the commercial will and regulatory pressure to deploy them have not materialised.

One partial exception to the complete absence of content governance mechanisms is the Trusted Partner system operated by the African Policy Dialogue (APD). This mechanism provides a channel through which verified civil society actors can escalate content removal requests to Meta platforms. While the system is imperfect — scope is limited, response times vary, and not all harmful content categories are covered — it represents an existing lever that Somali civil society organisations and digital safety practitioners can utilise. The research recommendation for formal platform engagement should be understood as building on and expanding this foundation rather than starting from zero.

Fifth, resilience investment must be matched to Somalia's trust architecture. The most effective counter-narrative interventions documented in this research operate through existing community trust structures like respected elders, religious leaders, and community radio operators who hold authority within their communities precisely because they are not institutional actors. The Social Ecological Model, applied to this context, proposes investing in these community gatekeepers as trained, supported information intermediaries rather than relying on individual digital literacy programmes. This approach is Somalia-appropriate for three reasons: it works through trust structures that already exist and do not need to be created; it is scalable without requiring institutional infrastructure that takes years to build; and it can operate at a speed closer to that of harmful content itself; community gatekeepers can intervene within hours, while institutional responses typically take days or weeks. The practical implication is that investment should flow toward digital literacy and verification capacity for elders, religious leaders, and community radio operators, not only toward urban individual-level programmes that do not reach communities most vulnerable to manipulation.

Sixth, the diaspora offers a policy opportunity that has not been taken. This report's evidence confirms that diaspora actors are among the most significant originators and amplifiers of harmful content in Somalia's digital ecosystem. But the same research documented a case, the aftermath of the Laascaanood conflict, in which sustained diaspora investment that initially supported armed mobilisation appears to have transitioned toward economic development and state-building activity. The diaspora's resources, networks, and attention are not fixed in their harm direction. Developing channels for constructive diaspora engagement, platforms for investment in digital infrastructure, remittance mechanisms tied to development priorities, diaspora inclusion in peacebuilding dialogue processes, may enable the redirection of significant financial and social capital toward outcomes that support rather than undermine Somalia's stability. This is not an argument for ignoring diaspora harm dynamics; it is an argument for addressing them through engagement rather than restriction.

RECOMMENDATIONS

Recommendations are organised by stakeholder and prioritised within each section. Immediate actions address the most urgent gaps and can be initiated within existing frameworks or with minimal new resources. Medium-term actions require planning, coordination, or new legislative processes. Longer-term structural reforms address root-cause governance and platform failures that require sustained effort across multiple years.

For the Federal Government of Somalia and Federal Member State Authorities

The recommendations in this section apply primarily to the Federal Government of Somalia (FGS) and the National Communications Authority (NCA). Federal Member State (FMS) authorities, including Puntland, Jubaland, South West State, Hirshabelle, and Galmudug, have overlapping mandates on media regulation and public order, and are encouraged to adapt the immediate and medium-term recommendations to their own contexts. Somaliland authorities are similarly encouraged to engage with these recommendations as appropriate to their governance structure, given the distinct but parallel dynamics documented in this research.

- [Immediate] Prioritise TFGBV legislation. Develop and enact specific legal provisions addressing gender-based digital violence, image-based abuse, coordinated online harassment, and sextortion. This is the single most important legal gap identified in this study. Consult women's organisations, survivor-centred service providers, media safety groups, and digital rights practitioners in drafting. Ensure provisions include anonymous reporting mechanisms and victim confidentiality protections.
- [Immediate] Operationalise the Data Protection Authority. The DPA was established in 2023 but is not yet operationally active. Staffing, mandate clarification, and operational procedures, including a functioning reporting pathway for data breaches, should be prioritised, with lessons learned from the e-visa breach used to establish accountability standards for third-party platform vendors.
- [Medium-term] Introduce influencer transparency requirements. Require content creators and influencers above a defined follower threshold to disclose paid political sponsorships, commercial relationships, and institutional affiliations when publishing political or electoral content. Somaliland's Cybercrime Bill Article 42 (defining influencers with 5,000+ followers) provides a useful regional precedent. Align with emerging international practice on platform political advertising transparency.
- [Medium-term] Develop a pre-election digital harm early warning protocol. Drawing on the documented pattern of disinformation spikes before and after elections, including the Banadir local election evidence, develop standardised protocols for monitoring, rapid response, and trusted-messenger intervention in the weeks before and after major political events.
- [Medium-term] Build judicial digital literacy. Establish structured training programmes for prosecutors, judges, and police investigators on digital crime, digital evidence, and the technical dimensions of cybercrime cases. The judicial incapacity documented in Section 7.5, including cases resolved through traditional elders for lack of technical

understanding, underscores the urgency. Partner with international legal and cybersecurity institutions where domestic expertise is unavailable.

- [Longer-term] Clarify and coordinate institutional mandates. Resolve the fragmentation of digital harm responsibilities across the NCA, Ministry of Information, Ministry of Communications, Data Protection Authority, and security actors. Assign clear lead responsibility for each category of digital harm with defined inter-agency coordination mechanisms and accountability lines.

For Civil Society and Media Organisations

- [Immediate] Expand specialist digital safety support. Donor investment should strengthen dedicated, NGO-led digital safety support for people facing online abuse; including case management, referral pathways, emergency digital security assistance, and geographic coverage beyond Mogadishu. Digital Shelter's model is the only specialised mechanism of its kind in Somalia and deserves significant scaling.
- [Immediate] Document TFGBV cases systematically. Establish a confidential, survivor-centred case documentation system, ideally hosted by an existing women's organisation in partnership with Digital Shelter, to build the evidence base for legal reform and to monitor patterns over time. The current absence of reliable TFGBV data perpetuates the invisibility of the most severe harm category and enables resource misallocation.
- [Medium-term] Invest in community gatekeeper capacity. Design and pilot digital literacy programmes targeting community elders, religious leaders, and community radio operators, the trusted messengers who already operate as information gatekeepers, rather than focusing solely on individual user literacy. The Social Ecological Model proposed by academic and technical researchers provides a practical framework for programming design.
- [Medium-term] Rebuild verification culture. Develop campaigns and community education approaches that address not only the skills of information verification but the social norms around it; countering the 'it was said' dynamic by making questioning information socially rewarded rather than punished. This is a social norm change intervention, not only a literacy programme.

For Platform Companies

Note on engagement: Platform companies are the most structurally difficult stakeholder group to reach through research of this kind. None of the major platforms operating in Somalia - Meta, TikTok, Google/YouTube, Telegram - has a dedicated Somalia presence, a Somali-language moderation team, or a formal engagement channel with Somali civil society or regulators. The recommendations below are addressed to platforms directly, but the most realistic near-term pathway for implementation is through intermediary mechanisms: the APD Trusted Partner relationship with Meta, civil society escalation routes, and international advocacy by donors and digital rights organisations with existing platform relationships.

- [Immediate] Establish formal engagement mechanisms with Somali regulators. Meta, TikTok, Google/YouTube, and Telegram should establish formal points of contact with the NCA, the Data Protection Authority, and relevant Somali civil society organisations for reporting harmful content, discussing moderation policies, and receiving context-specific guidance on Somali cultural and political dynamics.
- [Medium-term] Invest in Somali-language content moderation. Meaningful content moderation requires human review capacity in Somali. Platforms should establish or contract Somali-language review capacity, including expertise in clan-coded language and local political context. Emerging technical tools for Somali-language harmful content detection (see Annex D) should be integrated into moderation pipelines.
- [Medium-term] Review monetisation policies for Somali-market political content. The ‘online conflict as business model’ dynamic documented in this study – in which inflammatory clan and political content generates greater revenue than accurate or constructive content – is in part a product of platform monetisation design. Platforms should review how engagement-based monetisation interacts with political and clan content in the Somali context.
- [Longer-term] Require influencer sponsorship disclosure. Platforms operating in Somalia should implement technical mechanisms to support and enforce sponsorship disclosure for political content, as part of their existing political advertising transparency frameworks.

For International Donors and Partners

- [Immediate] Fund dedicated TFGBV research and intervention. The most severe harm category is the least resourced and the least understood. Dedicated funding for survivor-centred case documentation, legal reform advocacy, victim support services, and prevention programming, including in Puntland and other states beyond Mogadishu, is urgently needed.
- [Medium-term] Support diaspora engagement programmes. Develop programming that enables diaspora actors to redirect their resources and networks toward constructive engagement with Somalia’s political and economic development. The transition of Laascaanood diaspora investment from conflict support toward economic development documented in this study suggests that redirection is possible when formal channels exist.
- [Medium-term] Integrate digital harms into conflict prevention programming. The escalation dynamics documented in this study are inseparable from Somalia’s broader conflict dynamics. Conflict prevention programmes that do not include a digital harms component are working with an incomplete picture of the drivers of violence.
- [Longer-term] Fund systematic digital harm monitoring. Establish sustained, longitudinal monitoring of Somalia’s digital information environment, including both quantitative platform data analysis and qualitative KII-based research, to track trends over time and evaluate the impact of interventions.

BIBLIOGRAPHY

DataReportal. 2025. "Digital 2025: Somalia." January 2025.

<https://datareportal.com/reports/digital-2025-somalia>.

EUAA (European Union Agency for Asylum). 2024. *Country Focus: Somalia — Digital and Information Environment*.

FSNAU / FHH Somalia. 2024. *Mobile Money Usage Data 2019–2024*.

Guardian. 2026. "Guardian Reporter and Colleagues Detained and Beaten by Somali Police." May 9, 2026.

Hassan, A. 2025. "Somalia Social Media Report 2025."

Mohamed, Muhidin A., et al. 2025. "Detection of Somali-Written Fake News and Toxic Messages on Social Media Using Transformer-Based Language Models." arXiv:2503.18117.

National Union of Somali Journalists (NUSOJ) / Somali Journalists Syndicate (SJS). 2025. *Press Freedom Reports 2023–2025*.

Reuters. 2025. "Musk Says Starlink Now Available in Somalia." April 13, 2025.

Sheikh, M. 2025. "Platform Use and Conflict Narratives in Somalia."

Udanyi, E., and M. Papadaki. 2022. "Telegram and Harmful Content: A Policy Analysis."

World Bank / Altai Consulting. 2017. *Mobile Money in Somalia: Household Survey and Market Analysis*. April 2017.

ANNEXES

ANNEX A: RESPONDENT OVERVIEW

Source Code	Scope and Key Topics	Text Attribution Label Used in Report
SRC-01	Regulatory/government perspective, national scope: regulatory landscape, platform engagement, extremist-content response, and AI harms.	regulatory respondent / Mogadishu
SRC-02	Media training/civil society perspective, Puntland and Somaliland: platform dynamics, diaspora influence, rural vulnerability, and TFGBV.	media trainer / Puntland or Garowe
SRC-03	Media association/civil society perspective, Puntland and national scope: influencer networks, paid amplification, business extortion, and diaspora narratives.	media association respondent
SRC-04	Regional government/media perspective, Puntland scope: judicial capacity gaps, extremist threats, and online-to-offline harm.	regional government respondent / Puntland
SRC-05	Media practitioner perspective, Puntland scope: Telegram harms, gendered harms, and electoral and democratisation triggers.	media practitioner, Puntland
SRC-06	Digital safety/civil society perspective, national scope: TFGBV, the access-vs-safety gap, and influencer-government relationships.	digital safety practitioner / Mogadishu
SRC-07	Academic/technical perspective, national scope: AI risk, content lifecycle, and community gatekeeper resilience.	academic respondent / Mogadishu
SRC-08	Independent journalist perspective, Puntland and national scope: Telegram exploitation, coordinated harassment, and impersonation accounts.	former media association respondent / independent journalist, Puntland
SRC-09	Cybersecurity/private-sector perspective, national scope: cybercrime, the e-visa breach, extremist adaptation, and judicial technical gaps.	technical cybersecurity respondent / Mogadishu
SRC-10	Government technology policy perspective, national scope: AI normalisation, legal reform, and political enforcement avoidance.	government technology adviser
SRC-11	Independent journalist perspective, Puntland scope: Telegram conflict content, influencer disclosure gaps, and diaspora generational dynamics.	independent journalist, Garowe
SRC-12	Women's media/journalism perspective, Puntland and national scope: women journalists' harms, Telegram exploitation, and AI concerns.	women's media practitioner
SRC-13	Multi-stakeholder convening, national scope: cross-sector validation of platform, legal, literacy, and governance findings.	multi-stakeholder convening / Mogadishu

Annex B: Key Analytical Concepts and Local Terms

Term	Language / Origin	Definition
Qarxis	Somali	Exposure or threatened exposure of intimate or identifying material to destroy reputation or coerce compliance. Extends beyond doxxing to include honour, family consequences, and sexual coercion.
La yiri / 'It was said'	Somali / English translation	The epistemic formula that has replaced source attribution in Somali information culture: accepting claims because 'it was said' without demanding to know who said it or on what basis. Verification culture collapse in practice.
Second life of content	Analytical	The process by which content removed from public platforms continues to circulate through screenshots, voice notes, and forwards in private messaging channels, often with stronger emotional and clan framing.
Content hydra	Analytical (SRC-07)	The property of viral harmful content whereby removing the original post cannot suppress the narrative because it has already been reproduced into hundreds of derivative forms across multiple platforms and private channels.
Cook narratives	Somali diaspora practice	The deliberate preparation and fabrication of harmful narratives by diaspora actors before distributing them into Somalia's digital ecosystem. Implies premeditation and external origin.
Popularity circle	SRC-03	The informal prestige economy structuring Somalia's TikTok influencer networks, where follower count thresholds determine access to paid campaign networks and cross-regional amplification.
Online conflict as business model	SRC-06	The commercial monetisation of staged disputes, manufactured controversies, and divisive clan/political content by influencers generating engagement metrics that translate into revenue.
Access vs safety gap	SRC-06	The distinction between digital access (broadly achieved for urban women) and digital safety (largely absent): the challenge is not getting women online but enabling safe participation.