



Photo Credits, clockwise from top left:  
E. Millstein, Myanmar 2017 / B. Hoffman, Ukraine 2015 / M. Samper,  
Guatemala 2012 / Mercy Corps, Kenya 2013

# **SOCIAL MEDIA AND CONFLICT: UNDERSTANDING RISKS AND RESILIENCE**

## **An Applied Framework for Analysis**

JULY 2021



## Acknowledgements

This paper, and the related Research Findings paper, were authored by Keith Proctor, currently Technical Advisor and Team Lead, Evaluation and Research, at the International Republican Institute. The third publication in the series, the Policy Brief, was produced by Will Ferroggiaro at Mercy Corps. Adrienne Brooks, Lisa Inks, Adam Lichtenheld, and Will Ferroggiaro provided technical direction and guidance to the entire project, and all three publications benefited from the inputs and insights of other members of the Peace and Conflict, Technology for Development, Research and Learning, and U.S. and European Policy and Advocacy teams as well as Country Teams in Ethiopia, Iraq, Myanmar and Nigeria.

We would especially like to thank: Melaku Yirga, Country Director, and Aschalew Tekle, Peacebuilding Advisor (Mercy Corps Ethiopia); Danjuma Saidu Dawop, Head of Conflict and Social Cohesion, Mercy Corps Iraq; Mathieu Rouquette, Country Director, and Jonathan Bartolozzi, Director of Programs (Mercy Corps Myanmar); Ndubisi Anyanwu, Country Director, Phebe Banu, Program Management Officer, and David Gatere, Chief of Party for Community Initiatives to Promote Peace (Mercy Corps Nigeria); and from our global team members, Emma Whitaker, Peace and Conflict Advisor; Richmond Blake, Senior Director of Policy and Advocacy (U.S.); Ommera Ahmed, Senior Policy and Advocacy Advisor (UK); and Sasha Davis (Director, Corporate Partnerships). Many thanks to Heather Cummings for graphic design.

The author and Mercy Corps would also like to thank all of the individuals in Ethiopia, Iraq, Myanmar, and Nigeria who participated in interviews and questionnaires to inform these reports.

### Citation:

Proctor, K. 2021. *Social Media and Conflict: Understanding Risks and Resilience; An Applied Framework for Analysis*. Washington, D.C.: Mercy Corps.

By Keith Proctor<sup>6</sup>

The weaponization of social media is transforming conflict in the world's fragile states. From disinformation campaigns to electoral manipulation and online recruitment by violent extremist organizations, digital threats exacerbate conflict drivers, open new avenues for spoilers, and increase polarization, complicating efforts to prevent or counter violent conflict.

Peacebuilding initiatives therefore need to incorporate a digital lens. Addressing this challenge requires a sound risk assessment framework to empower local actors, international partners, and social media companies to anticipate and prevent harms arising from the interplay between digital ecosystems and off-line conflict dynamics.

Building on case studies examining the emergence of digital threats in Ethiopia, Iraq, Myanmar, and Nigeria, Mercy Corps' Peace & Conflict and Technology for Development technical support teams developed this report to inform approaches and methods for addressing the weaponization of social media, as well as advancing online and off-line social cohesion and peace. This research focused on qualitative community-level assessments to better understand how the interplay of online and off-line dynamics creates opportunities for social media narratives to gain traction and contribute to conflict.



Photo Credit: J. Zoeherman / Mercy Corps / South Sudan 2014

With extensive reference to the case studies and to secondary research, this framework employs six categories of analysis to understand pathways to violence between the online and off-line space by way of different risk factors:

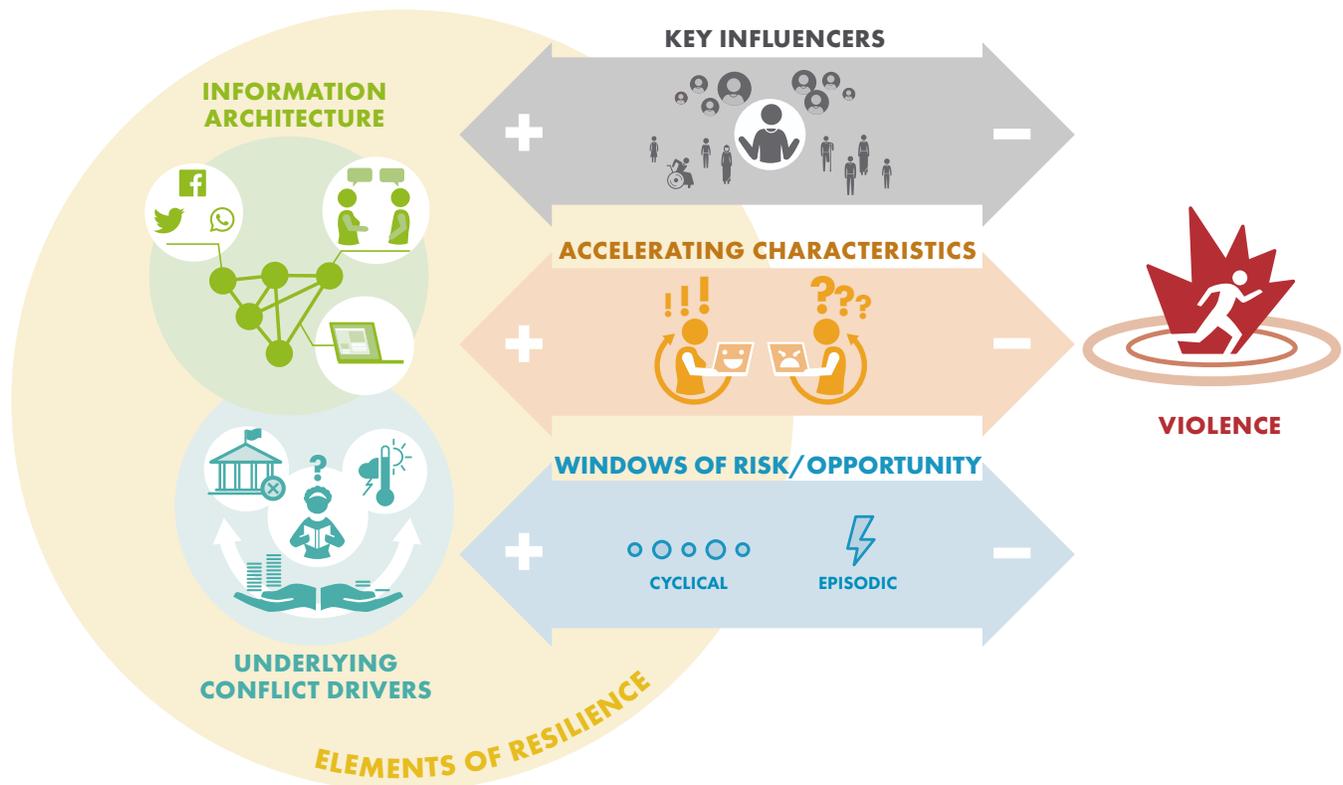
1. The **Information Architecture** describes how information flows between on- and off-line spaces. Relevant factors include internet access, dominant platforms, the regulatory environment, and user characteristics.
2. **Key Influencers** who can shape perceptions and mobilize on- and off-line constituencies.
3. **Underlying Conflict Drivers** that are susceptible to social media platform manipulation, such as intercommunal conflicts and community-state tensions.
4. **Windows of Risk** during which vulnerabilities to online harms are most pronounced. Examples include elections, religious festivals, and public health crises.
5. **Accelerating Characteristics** describe the mechanisms by which social media appears to transform conflict dynamics, for example by heightening perceptions of threat, normalizing hate speech, or facilitating mobilization.
6. The **Sources of Resilience** that appear to mitigate digital threats, such as online and off-line civil society actors who counter dis- and misinformation.

<sup>6</sup> Keith Proctor is currently Technical Advisor and Team Lead, Evaluation and Research, at the International Republican Institute.

For key findings from the case study countries that emerged through the process of refining this analytical framework, please refer to the Research Findings report from this research. In addition, a Research Summary and Policy Brief casts the framework and case studies in a policy light and concludes with a set of recommendations.

**Methodology.** This framework is based on a review of secondary literature; Mercy Corps research on social media and conflict; case study research, including key informant interviews (KIIs) and questionnaires, by the author in Ethiopia, Iraq, Myanmar, and Nigeria in 2020; and KIIs conducted by the author in Nigeria’s Middle Belt in 2019 that explored the intersection of social media and off-line conflict trends.

## A FRAMEWORK FOR UNDERSTANDING RISKS AND RESILIENCE OF SOCIAL MEDIA AND CONFLICT



### 1. Information Architecture

#### How information flows between online and off-line spaces

This research assessed social media platforms<sup>7</sup> not as a series of closed or discrete spaces, but as part of a larger information ecosystem, one in which the (highly localized) interplay of online and off-line dynamics create new realities for users and non-users alike. Social media does not operate in a vacuum. Identity and context frame how social media narratives are received, and which stories gain traction and which do not. The character of the information ecosystem is key to assessing how and where social media weaponization is likely to succeed.

<sup>7</sup> Platforms assessed included Facebook, Twitter, Instagram, YouTube, WhatsApp (and other messaging platforms), and any other social media relevant to the case study country context.

In addition to outlining key findings and considerations, this section concludes with a list of relevant factors for future assessments.

- › **Social media threats are not restricted to social media users.** As connectivity grows, opportunities for social media weaponization increase. The falling price of SIM cards and the expansion of telecommunications coverage in Myanmar rapidly magnified the risks of hate speech and disinformation. But even when internet penetration is low or confined to urban centers, the impact of social media can reach far beyond the user base. This research documented cases in which online narratives appeared to “spill over,” reaching populations with limited or no internet connection. According to respondents across multiple country contexts, social media is shaping perceptions and driving discourse in off-line spaces, from church pulpits to newspapers, TV, and radio, reaching those who are illiterate, do not have smartphones, or lack social media accounts. False or inflammatory narratives that become popular online travel by word of mouth or text. In public spaces—football fields, restaurants, the market—non-social media users hear stories second-hand or gather around a peer’s device to watch a video or to listen to someone read a story off of their phone. In many cases, social media is driving news, with print outlets, radio, and television stations recycling online stories. In Nigeria’s Middle Belt, rural participants described religious leaders echoing online conspiracy theories while preaching to their congregants. Among traditional media actors, such as radio stations or newspapers, the desire to stay relevant in the face of competition from bloggers and e-journalists leads some editors to publish stories from the internet without verification or fact-checking. “Lots of people don’t use social media here,” said a youth in a rural community. Still, he said social media shapes the information landscape. “The radio stations,” he said, “just pick up and repeat the stories on Facebook.”
- › **Off-line social networks inform trust in social media narratives.** Across the case studies, social media was generally either the primary or secondary source of news for social media users. However, research participants were not naive to the problems of online mis- and disinformation. They expressed skepticism about many of the stories circulating on social media platforms. Most emphasized the importance of checking a story’s sources. In general, however, validating a story’s source simply meant considering who was forwarding it. Social relationships provide a shortcut for assessing the plausibility of social media news stories: an online story was more likely to be believed if it had been implicitly validated by a friend, family member, or respected peer.<sup>8</sup> “If people trust the messenger, they trust the message,” said an Ethiopian participant. “People never actually bother to check the veracity of the story.”
- › **Social media access and vulnerability vary across class, age, gender, and geography.** Internet penetration varied across the case study countries, from 17.8 percent in Ethiopia to 75 percent in Iraq. Social media use generally lags behind these numbers.<sup>9</sup> The most prevalent and savvy social media users are younger (generally age 40 and under). Most users are concentrated in urban centers, although this is beginning to change as networks expand into rural areas. Internet penetration and network robustness influence platform popularity. For instance, while YouTube is a prominent channel for posting inflammatory videos, it may have limited reach in poorer communities where data costs are relatively high.

User characteristics, such as age, gender, and geography, shape opportunities and vulnerabilities to platform manipulation. For example, in Nigeria’s Middle Belt, a focus group of Salafi Muslim women emphasized that social media was a “lifeline” providing social connection for wives confined to the home by conservative husbands, but this dependence also increased susceptibility to online disinformation and misinformation.

<sup>8</sup> In general, and across the country contexts, stories traveling by word-of-mouth were the most highly trusted sources of news.

<sup>9</sup> The exception is Myanmar, where the data-cheap Facebook app is, for many, the most cost-effective way to access the internet. The vast majority of users there continue to access the internet via social media, and analysis from January 2020 suggests the number of social media users approximately matches the internet user population at 22 million people. See: <https://datareportal.com/reports/digital-2020-myanmar>

- › **Some risks are platform-specific.** For example, in Nigeria the popularity of WhatsApp has confounded efforts to regulate or counter political manipulation and digital hate speech. Closed WhatsApp discussion groups, in which members are vetted by moderators, reduce opportunities for external monitoring and help ensure the ‘echo chamber’ effect that contributes to polarization. Social media influencers who sow online disinformation on behalf of political leaders preferred to use WhatsApp given the potential for creating multiple online identities: given that WhatsApp accounts are tied to mobile phone numbers, an influencer can have as many proxy accounts as he has SIM cards.

Violent online content flourishes where it is unpoliced. In Ethiopia, e-activists working to flag dangerous online content for removal said they found YouTube far slower to act upon complaints than other platforms. Crackdowns also drive content to more accommodating (if less popular) channels; VKontakte (VK), Russia’s most popular social media platform, has been a haven for Myanmar influencers banned from Facebook for inciting violence.

- › **Government policy, liberalization, and crackdowns.** Government policies open and close spaces for social media activities. In 2018, the Ethiopian government liberalized media policies, lifting restrictions on press freedoms, releasing jailed journalists and bloggers, and restoring access to more than 200 news sites that had been blocked for years. The reforms opened up opportunities for Ethiopian TV channels from abroad, many of which also have a strong social media presence, to broadcast more freely within the country. This has expanded the reach and influence of the Ethiopian Diaspora to more directly influence political discourse inside the country.

However, Ethiopia—like other countries—has retained the ability to block social media usage, and during periods of instability has shut down internet access entirely. In Myanmar, telecommunications law authorizes the government to block websites during an “emergency,” a power it has used to shut down websites maintained by critics or activists; since the February 1 military takeover, the military has ordered social media blocks, and throttling and shutdowns of fixed-line and mobile internet. Journalists may also be jailed under a counter-terrorism law. And the government has utilized selective shutdowns of 3G and 4G services to reduce access to real-time news and increase information dependence on official announcements arriving by SMS.

Across the case studies, there has been an observable uptick in state-led efforts to police online content. The Iraqi parliament has introduced a cybercrime law under which “harming the reputation of the country online” would carry a life sentence. Similarly, in Nigeria the “Protection from Internet Falsehood and Manipulation Bill,” introduced in 2019, would empower authorities to imprison and/or fine individuals who transmit online statements that are deemed false, likely to influence an election, and/or are “prejudicial to the security of Nigeria.”<sup>10</sup> These efforts are popular in some quarters, due to a growing recognition of the dangers of digital hate speech and disinformation. However, anti-hate speech laws may be used to curtail speech, muzzle journalists, and silence online dissent.

- › **Social media narratives fill the gaps left by eroded trust in traditional sources of information.** Across country contexts, relative measures of trust appeared to be important. Online conspiracy theories flourish where traditional sources of information—governments or news organizations—lack credibility. While many participants viewed online stories with some skepticism, they distrusted other media and political leaders more. In Ethiopia, long-standing suspicions of state-dominated media have fostered an appetite for alternative, non-official sources of information. When competing narratives cannot be reconciled, these dynamics can be self-reproducing. The deficit of faith in official sources reduces elite capacities to counter false narratives. Across contexts, this appeared to drive potentially dangerous cynicism. In Iraq, online rumors and disinformation

---

10 <https://africlaw.com/2019/12/05/a-socio-legal-analysis-of-nigerias-protection-from-internet-falsehoods-manipulations-and-other-related-matters-bill/>

about COVID-19 had potentially crippling effects on legitimate public health messaging. A June 2020 assessment in Anbar and Ninewa found that 43 percent of respondents indicated they had been inundated by conflicting information about the virus and were unable to distinguish between rumor and fact.<sup>11</sup>

### Relevant factors for information architecture assessments.

- › *Levels of digital connectivity*, including internet and smartphone use, broadband subscriptions, costs, infrastructure coverage, and connection speeds.
- › *How social media use varies by class, age, gender, and geography.*
- › *Relative trust and influence of information sources in the community.* Relevant indicators might include attitudes toward traditional media and/or official news sources, reliance on social media as a primary or secondary news source, etc.
- › *Relative popularity of social media platforms* (Facebook, WhatsApp, Twitter, etc.) and how the design features of popular platforms shape digital threats.
- › *Marketplace and infrastructure shifts that may expand coverage* (e.g., the falling price of SIM cards, telecommunications expansions).
- › *Social media impacts on off-line information channels*, including ways in which social media narratives move online and off-line, and how they reach communities with limited digital connectivity.

## 2. Key Influencers

Key influencers are individuals, organizations, and institutions with the capacity to shape perceptions and mobilize key constituencies. Motives, methods and tactics may vary widely. Those with an ethnic or sectarian following may be more likely to deploy digital hate speech as a method to drive division. Social media's low barriers to entry can empower online militants who may sideline customary leaders or traditional mechanisms for dispute resolution. Authoritarian political elites may engage in political manipulation to influence the media, silence dissent, or undermine reformers. External actors, keen to meddle in internal affairs or shift balances of power between groups, may utilize online information campaigns to promote economic or political interests. Yet, given the potential for anonymity online, it may be difficult to discern who is pushing a narrative, or why. "On social media, ethnic conflicts can be created by ghosts," said an Ethiopian participant. "It's like these people do not even exist."

Where possible, these case studies attempted to map key social media influencers, both violent and non-violent. Among the latter, peace-oriented influencers, such as digital activists or online religious leaders, may be sources of resilience to digital threats, but they can also be lightning rods that attract attacks. For example, Iraqi youth activists have used online platforms to mobilize peaceful anti-government protests, but these have spawned a violent backlash from the movement's enemies, including Iranian-backed militias and prominent religious leaders associated with the government.

Key influencers vary, though some archetypes are common. This research identified categories of online influencers with the ability to mobilize key constituencies either to promote social cohesion or to sow division. Examples common across these contexts include:

- › **Online diaspora communities**, many of whom are well-resourced, digitally-savvy, and shielded from retaliation by living abroad. Social media has long provided opportunities for Ethiopia's global Diaspora—

---

11 Mercy Corps (2020). Connecting or Dividing Communities? The Impact of COVID-19 in Iraq. August.

numbering over two million, according to some estimates—to help shape domestic politics. Operating from the safety of the U.S. or Europe, this Diaspora, many of whom fled the country amidst political turmoil, wield substantial influence as a source of remittances and as vocal political activists utilizing social media and satellite TV to shape events.

- › **State actors**, including military or security sectors, who use social media to spread propaganda, identify and prosecute critics, or, via service denials and internet blackouts, use their authority and capability to reduce communications and the visibility of ongoing conflicts. In Iraq, the state security sector and its allies, including Iranian-backed militias, appear to have used social media to identify and target bloggers and digital activists who were part of the 20 October Revolution. Given the low cost and minimal infrastructure needed, even poorer countries may have security apparatus that deploy ‘digital armies’ against internal and external critics.
- › **Non-state armed actors**, including violent extremist organizations and insurgent groups. Nigeria’s Boko Haram, and its offshoot the Islamic State West Africa Province (ISWAP), have used various social media platforms—including Facebook, WhatsApp, SnapChat, Instagram, and YouTube—to attack the state and Western aid organizations, spread disinformation about COVID-19, and publicize acts of violence. In Iraq, ethnic militias operating as Popular Mobilization Units (PMUs) trawl online platforms to identify and target enemies or supposed ISIS sympathizers.
- › **Politicians and political parties** who use social media as a tool for politics and campaigning. In the worst cases, they use social media to attack rivals, intimidate opposition constituencies, and flood the internet with propaganda. In Nigeria, politicians are almost universally cited as the Middle Belt’s most divisive figures. Prominent politicians are routinely central figures in ongoing social media conflicts, either as catalysts or targets, or both. For example, the governor of Kaduna state, Nasir El-Rufai, appears to make extensive use of both traditional and social media to shape opinion, according to respondents, but this has also made him a magnet for online attacks, which in turn galvanize his supporters, and thus, negative feedback loops ensue.
- › **Religious leaders** and their followers who are active on social media and use the platform variously to advance interfaith understanding or to sow ethnic and sectarian divisions. Myanmar’s Buddha Dhamma Parahita Foundation (BDPF) is an ultra-nationalist Buddhist organization and the successor of the Ma Ba Tha.<sup>12</sup> BDPF leaders, such as Ashin Wirathu, utilize social media platforms to drive ethno-sectarian hate speech, much of it targeting the country’s Muslim Rohingya minority.
- › **E-activists**, primarily youth, who use social media platforms to mobilize protests and advocate for political change. For example, Nigerian youth activists protesting police profiling and unjust arrests perpetrated by the government’s Special Anti-Robbery Squad (SARS) have used the #EndSARS hashtag to mobilize massive protests. In central and southern Iraq, digital youth activists have made extensive use of social media to organize anti-government protests in 2019 and (to a lesser extent) in 2020. Government supporters, including Iranian-backed militias, have responded with violent crackdowns and a number of prominent activists have disappeared.

An assessment of key influencers can shed light on the nature and scope of social media risks in a given context. Where possible, assessments should identify specific individuals and organizations. An assessment of key influencers should encompass:

- › *Interests and incentives*, which may be ideological, financial, political, etc., and how the actor fits into existing *patterns of risk and grievance*.

---

12 The Ma Ba Tha was associated with driving some of the worst anti-Rohingya hate speech, online and off-line, in the lead-up to the 2017 crises. Disbanded in the aftermath of the crisis, the group was reborn as the BDPF.

- › *Operational capacities*, which include organizational attributes, links, and social networks that facilitate mobilization and outreach (including an influencer’s social media following).
- › *Relevant communication channels*, including online and off-line platforms.
- › *Transnational links* with organizations and actors outside the state that offer resources or support, or enhance the actor’s status.
- › *Methods and tactics*, particularly the types of social media operations the actor employs (e.g., information operations, political manipulation, digital hate speech, radicalization and recruitment).
- › *Competition* from other key actors for the same base of support. Where competition is intense, acts and rhetoric may escalate, promoting more radical or hard-line views (e.g., toward out-groups or the state), a process described as “outbidding.”<sup>13</sup>



Photo Credit: M. Madukwe / Mercy Corps / Nigeria 2019

### 3. Underlying Conflict Drivers

#### The social, economic, and political issues susceptible to platform manipulation

Social media risks are rooted in context. The local environment encompasses the social, economic, cultural, and historical issues that govern intercommunal and state-society relations. While these relations can be amplified, or even transformed, in social media, the socioeconomic and political conditions provide the raw material for digitally-mediated conflicts. Social media manipulation is more likely to succeed where it capitalizes on existing divisions, frustrations, and fears. Where social cohesion—defined as “a sense of shared purpose and trust among members of a given group or locality and the willingness of those group members to engage and cooperate with each other to survive and prosper”—is weak, societies may be vulnerable to social media weaponization.<sup>14</sup>

While the purpose of this research was not to reiterate conflict drivers, the case studies highlighted a number of sub-themes relevant to the study of social media and conflict.

- › **Ethnic and sectarian tensions appear particularly vulnerable to the weaponization of social media.** Across the case studies, online platforms routinely fueled communal tensions centered on ethnic and sectarian identities. In Ethiopia, inter-ethnic tensions provide ready fodder for online conflicts that have spilled over into

13 Mia Bloom describes this phenomenon in the context of violent extremist groups. See Bloom, M. (2005). *Dying to kill: The allure of suicide terror*. New York: Columbia University Press.

14 Working with the World Bank, Mercy Corps elaborated a conceptual framework for measuring social cohesion. Relevant components include: (i) *trust*; (ii) *shared purpose*, including belonging, common identity, and attitudes towards out-groups; (iii) a capacity for *collective action*, due to shared beliefs, norms, and commitments; and (iv) *civic engagement*, which measures individual and group commitments to cooperate to improve local conditions and shape the community’s future. See Kim, J., Sheely, R., & Schmidt, C. (2020). *Social Capital and Social Cohesion Measurement Toolkit for Community-Driven Development Operations*. Mercy Corps & The World Bank, February.

off-line violence. In Myanmar, years of deliberate disinformation campaigns by Buddhist nationalists helped build resentment against the Muslim Rohingyas in Rakhine state, setting the stage for a brutal campaign of ethnic cleansing in 2017. Anti-Muslim speech continues to circulate today. In Iraq, ethnic and sectarian tensions play out routinely in online spaces, with conflict actors utilizing disinformation to attack minorities, or to gain the sympathy and support of the oppressed. For example, the ISIS media network has circulated online videos of Shia or Kurdish militias perpetrating alleged human rights abuses among Sunni communities as part of an outreach campaign to potential supporters.

- › **Multiple identities may overlay an existing conflict—which can offer multiple entry points for use of social media as a weapon.** For example, what begins as resource competition may exacerbate other intergroup differences, such as tribal or religious divides. According to participants, social media appears to be a space in which conflict-prone identities are nurtured and activated. In Nigeria’s Middle Belt, amid the decades-long conflict between farmers and herders, discourse around resource competition has been politicized along ethnic and sectarian lines, eroding social cohesion. Participants highlighted declines in trust and positive intergroup interactions, as well as the increasing politicization of religious identity. Online rumors of sectarian killings, perceived insults to one’s faith, highly publicized acts of conversion from one religion to another, or prohibited interfaith marriages can all become viral online events that quickly spiral. Meanwhile, local ethnic militias and community vigilante groups increasingly appear to be better-resourced and better-armed. Such hair-trigger environments are ripe for platform manipulation.

- › **Where governing systems are dysfunctional, oppressive, or weak, they foster grievances and trust deficits that social media may exacerbate.** In the midst of an emergency, such as the Ebola crisis or the current COVID-19 pandemic, citizens may be persuaded by online conspiracy theories to discount guidance from state and health authorities. Videos and stories shared on social media may highlight repressive actions by state security forces—providing



Photo Credit: D. Evans / Mercy Corps / Guatemala 2019

a welcome source of transparency, but also one that may be manipulated. In Northeastern Nigeria, ongoing insecurity has sapped faith in the security sector and has fed online theories that elites have a vested interest in prolonging the conflict, as it enables them to siphon funds from the massive annual budget of international humanitarian aid. Boko Haram has attempted to capitalize on frustrations by claiming that efforts to mitigate COVID-19’s spread—e.g., the suspension of pilgrimages to Mecca, modifications to Ramadan observances—are really part of a war on Islam. In Iraq, frustrations with government corruption, unemployment, and poor services have fueled outrage on social media, including condemnations of the government, and off-line protests. In October 2019, social media activists in central and southern Iraq ignited a mass protest movement against government corruption, a lack of jobs, and meddling by Iran and the U.S. in Iraqi affairs. The “20 October Revolution” lasted months, paralyzing cities in central and southern Iraq, and helping prompt the resignation of Prime Minister Adel Abd al Mahdi.

- › **Frustration with security forces is a prominent flashpoint on social media.** Oppressive security actors, police profiling, government-sanctioned violence, and unjust arrests animated social media narratives across the country contexts. In some cases, they have erupted into off-line demonstrations (which can draw further

retaliatory action by the state). In Nigeria, the research highlighted a number of online movements targeting the security sector. In urban areas, frustration with the police profiling of marginalized youth, particularly by the government's Special Anti-Robbery Squad (SARS), have led to popular condemnations on Twitter—via the hashtag #EndSARS—and widespread off-line protests. Similarly, in the Northeast simmering frustrations with the security forces fighting Boko Haram are reaching a boiling point on social media, which have provided a fora for highlighting human rights abuses perpetrated by the military. On Twitter, the hashtag #SecureNorth has provided a focus for online discourse demanding an end to rampant insecurity.

Relevant factors for assessing conflict drivers and their relationship to social media:

- › *Intergroup perceptions that are shaped by social media.* Online platforms can amplify perceptions of threat and vulnerability, in turn influencing off-line interactions. Upticks in digital hate speech targeting an ethnic or sectarian minority may reduce intergroup interactions, undermining social cohesion, and reducing peacebuilding opportunities.
- › *Online misinformation and disinformation may inflame perceptions of unfairness or inequality.* Examples include unequal access to natural resources, land, employment, or government assistance vis-a-vis other identity groups.
- › *Attitudes toward violence, both generally and toward outgroups, which may be fostered and licensed in online 'echo chambers.'* Measures include respondents' professed acceptance of the use of violence against the other group under a range of situations, such as to protect their family or avenge past aggressions.
- › *Prevalence of online hate speech and conflict triggers,* including routinely deployed false narratives, pejorative terms, and insults.
- › *Social media impacts on conflict dynamics,* with an emphasis on shifting intragroup and intergroup dynamics. Examples include perceptions of increased polarization, online mobilization leading to violence, and hate speech driving the isolation of vulnerable groups.

## 4. Windows of Risk

### When vulnerabilities to online harms are most pronounced

Windows of risk describe events or periods in time during which an uptick in online and off-line tensions make violence more likely, providing malign actors with increased opportunities for escalation. While windows of risk are apparent in any conflict environment, they are often amplified by social media. For example, a high-profile act of violence, such as an assassination or a terrorist attack, may spawn a wave of disinformation, hate speech, scapegoating, and/or calls for reprisal. Similarly, an economic or climate change-driven shock may imperil livelihoods and increase intergroup competition, catalyzing spikes in digital hate speech that further aggravate tensions.

Windows of risk highlighted in the case studies include:

- › **Elections and political campaigns** are often characterized by spikes in partisanship and contests between groups. Across the country contexts, politicians and their supporters employed social media to smear opponents and orchestrate intergroup conflict to drive (or depress) turnout. During elections in Nigeria's Middle Belt, politicians employ digitally-savvy youth—"e-rats"—to drive disinformation targeting political opponents. According to participants, social media is becoming so central to political campaigns that some describe a social media "arms race" among candidates who hire the services of "e-rats" to attack and defend themselves in social media spaces.

- › **Cultural celebrations and religious festivals** can also be opportunities for magnifying or exploiting differences between ethnic or sectarian identity groups, sparking online and off-line confrontations. According to Ethiopian participants, celebrations associated with particular ethnic or religious groups may prompt social media attacks, but they also present opportunities for disinformation.<sup>15</sup> Examples include Irreecha, the Oromo festival of thanksgiving in October; Timkat, the Orthodox celebration of Epiphany in January; or Chambalaalla, the new year festival for the Sidamas. In Myanmar, Ramadan is typically characterized by an uptick in anti-Muslim hate speech and disinformation.<sup>16</sup> In Nigeria, public celebrations of Ashura by the Shia minority are typically flashpoints, especially in urban areas.
- › **Viral acts of violence**, including terrorist attacks or assassinations. For example, the killing of renowned Ethiopian musician and activist Haacaaluu Hundeessaa in 2020 catalyzed widespread conflict principally fueled through social media and fomented by political leaders, activists, and the global Diaspora.<sup>17</sup>
- › **Periods of resource-gathering**, particularly where land use or access is a flash point. In Nigeria’s Middle Belt, planting and harvest seasons routinely see an uptick in intergroup hostilities between farmers and herders that also play out in the digital space. Following the wet summer and the onset of the dry season, herders move south, towards greener pasture land and water supplies, often moving across farm land, which can spark confrontations. In addition, some participants suggested that harvest season is a time of increased criminality—mature crops, in addition to providing cover, may tempt thieves. The rise in criminality provides increased opportunities for identity-based scapegoating. As during the planting season, spikes in conflict may be exaggerated, or even manufactured, in online spaces, spawning intercommunal retaliation.
- › **Politically-salient anniversaries**, including the deaths of prominent leaders or past acts of violence. In Jos, Nigeria, the 7 September anniversary of the 2001 crisis—which saw bloody clashes between Muslims and Christians—has perennially been a day of intercommunal rancor in online and off-line spaces. In Ethiopia, the Tigrayan holiday of Yekatit 11 celebrates the onset of armed struggle in 1974 that eventually overthrew Ethiopia’s communist military regime in 1991. According to participants, the anniversary celebrations emphasize Tigrayan identity and are associated with an uptick in digital hate speech across ethnic groups.
- › **Public health crises**, such as COVID-19, appear particularly well-suited to social media weaponization. Across the case study countries, disinformation about COVID-19 exacerbated intergroup tensions and community-state relations. For example, in Nigeria, the perceived inadequacy of government assistance has exacerbated community-state frustrations and fed conspiracy theories that COVID-19 is a hoax perpetrated by government officials and shadowy outsiders. In Myanmar, fears associated with the pandemic appeared



Photo Credit: M. Roquette / Mercy Corps / Egypt 2021

15 The infamous events of the 2016 Irreecha celebration, and subsequent online disinformation and misinformation, provide a prime example. For more, see: <https://www.pri.org/stories/2018-02-22/social-media-america-are-stoking-ethiopia-ethnic-violence>

16 Government actors have played to the popular will by suppressing Islamic religious activities, e.g., canceling public prayers or shutting mosques.

17 <https://www.theguardian.com/global-development/2020/aug/03/how-a-musicians-death-unleashed-violence-and-death-in-ethiopia>

to drive social media attacks on ethnic and sectarian minority groups, including Muslims and Christians. “Social media use has increased a lot during the pandemic, because everyone is staying home,” said a Myanmar participant. “And we have witnessed an explosion in online propaganda, hate speech, and rumors.”

- › **International disputes** may also supply an accelerant to online conflicts. For example, the international dispute between Egypt and Ethiopia over the Grand Ethiopian Renaissance Dam (GERD) is fueling jingoistic rhetoric online. In Myanmar, the 2017 ethnic cleansing of Muslim Rohingya is now a genocide case before the International Court of Justice (ICJ). Some analysts fear future rulings have the potential to further stoke Buddhist nationalism and anti-Rohingya sentiment.<sup>18</sup>



Photo Credit: E. Millstein / Mercy Corps / Nigeria 2018

Steps for assessing windows of risk and their impact on digital threats include:

- › *Identify likely cyclical windows of risk* that can intensify existing intergroup and/or community-state tensions. While previous incidents of tension or conflict associated with these cyclical events may serve as a guide, they may be insufficient as events, conflict drivers, relevant actors, and the character of the digital ecosystem may all have evolved.
- › *Sporadic windows of risk* may be identified through conflict analysis exercises, based on existing sources of risk, to assess the types of events that produce or exacerbate existing tendencies for harm.
- › Identify the *key influencers* (local and external) who have the motives, resources, and followers to capitalize on identified windows of risk.
- › *Relevant sources of resilience* (e.g., respected leaders, peace-builders) who may counter or prevent violent conflict during periods of elevated tensions (for more, see “Sources of Resilience” below).

## 5. Accelerating Characteristics

### The mechanisms by which social media appears to transform conflict dynamics

Social media reduces the costs of violent incitement. Low barriers to entry, combined with relative anonymity and geographic flexibility, empower influencers to reach new constituencies, push disinformation, or incite conflicts. It also allows disparate individuals to feel a part of something they perceive as unifying, successful, or bigger than

<sup>18</sup> <https://www.crisisgroup.org/asia/south-east-asia/myanmar/myanmar-international-court-justice>

themselves—whether or not those are true. Connections fostered through social media may reduce the isolation of would-be extremists, enabling them to connect with like-minded others, or practice digital hate speech in a manner with little personal risk.

Critically, social media does not simply provide an additional communication channel; rather, social media spaces—by virtue of their mobilizing, value-setting, and perception-shaping powers—shape how conflicts manifest. Social media appears to increase the emotional salience of conflict, expand audiences, and speed up processes of provocation and reaction. Events that might not otherwise turn violent, might spiral uncontrollably in digital space. In short, social media is transforming how, when, and whether conflicts manifest.

This research aimed, in part, to document different pathways by which social media influences behaviors, erodes social cohesion, and influences perceptions (of self and other) in a manner that heightens risk of conflict. The following factors, while not exhaustive, may inform a subsequent typology of social media and conflict.

- › **Online ‘echo chambers’ intensify perceptions of threat.** Social media naturally creates insulated information environments that promote consensus-building and polarization. Particularly in environments of intergroup conflict or community-state tensions, social media spaces appeared to intensify perceptions of threat, raise emotional stakes, and lend urgency to calls for action. By appearing to confirm existing fears and biases, social media may catalyze shifts in intragroup and intergroup dynamics that make conflict more likely. In Myanmar, fears associated with COVID-19 have fed online scapegoating narratives targeting ethnic and religious minorities. Christian communities, some of which have allegedly held religious gatherings in spite of lock-down orders, are being blamed on social media for spreading the virus. Similarly, a subsequent outbreak that appeared to originate in Rakhine state prompted accusations online that Muslim Rohingya there are intentionally spreading the virus.
- › **The speed of disinformation can manufacture new realities.** The impulse to share alarming or inciteful stories is often automatic. Verification and correction simply take too much effort and may not be timely. Conflicts between Christians and Muslims in Nigeria’s Middle Belt often have a “wag the dog” dynamic, with fake stories online about intergroup attacks catalyzing rapid off-line retaliation. The fake stories may eventually be debunked, but not before catalyzing real acts of violence and sowing new sources of hostility. “Even if you can eventually correct a story,” said a civil society leader, “by the time you do, it’s too late. The damage has been done.”
- › **The breadth of communication facilitates mass mobilization.** The diffusion of peer-to-peer communication has reduced the costs of information-sharing and collective action, providing a channel for identity-based groups to coordinate across geographic space. An inciting event, which might otherwise have fizzled out or remained highly local, can now spread virally online. As noted above, online conspiracy theories surrounding the 2020 assassination of Haacaaluu Hundeessaa in Ethiopia, a prominent Oromo singer, triggered bitter intercommunal conflicts. Oromo and Amhara influencers and their followers attacked one another online, there were calls for revenge attacks, and Diaspora influencers piled on, hyping up the emotional salience and identifying culprits. Waves of violence engulfed the country; mobs assembled via social media clashed with government security forces. Following a three-day rampage, at least 178 people had been killed, hundreds more injured, and over 9,000 people arrested.<sup>19</sup>
- › **Online hate speech can heighten the perceived vulnerability of marginalized groups, undermining social cohesion.** Social media can have a chilling effect on real-world interactions across intercommunal divides. Respondents described how hate speech reduced off-line interactions, for example by dissuading

---

19 <https://theconversation.com/ethiopias-political-crisis-plays-out-in-the-regions-why-its-a-federal-problem-144893>

vulnerable groups from visiting neighborhoods they perceived as increasingly unsafe. When casual interactions between communities become rare, it is easier for one side to demonize the other. Where online hate speech exacerbates perceptions of vulnerability, it may accelerate ethnic or sectarian sorting like that evident in Nigeria’s Middle Belt. For instance, in Kaduna, Christians increasingly live in the southern part of the city, Muslims in the north. Similarly in Jos, the Congo-Russia neighborhood is sharply

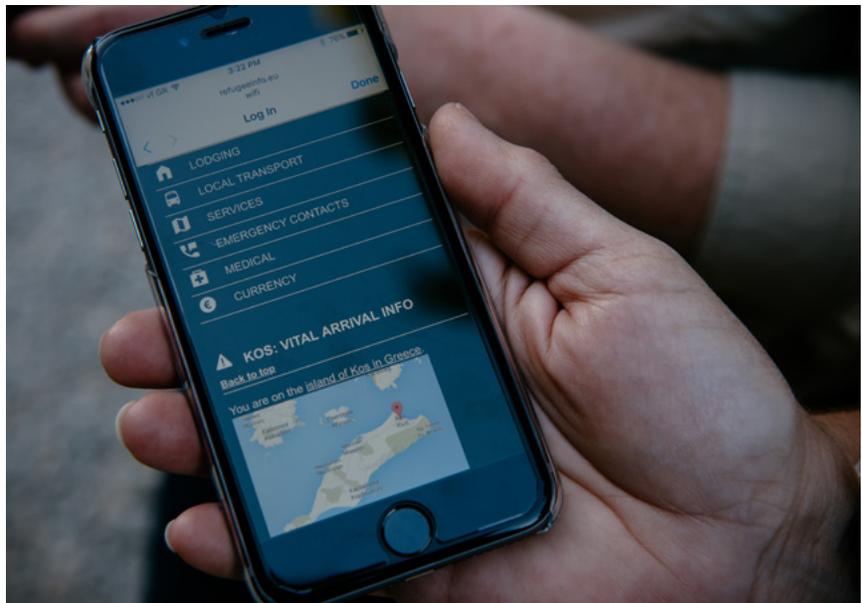


Photo Credit: C. Robbins / Mercy Corps / Greece 2015

demarcated by Christian and Muslim neighborhoods that are “no go” zones for members of the other group. According to respondents in these communities, the rise in sectarian hate speech online has helped drive the perceptions of threat and vulnerability that are leading local minorities to emigrate to safer neighborhoods. As communities become more homogeneous, isolation from other groups deepens, leading to further deterioration in social cohesion.

- › **Online platforms incentivize not only connection but performance.** Influencers competing for clicks or constituencies might amp up online rhetoric attacking marginalized groups, or a live-streamed act of vandalism can win accolades and imitators. Social media communities may give license to violent behavior, and online platforms can provide a venue for militants to practice hate speech in a personally low-risk manner. In Nigeria, a Sunni man described responding to an online call to violence against the minority Shia. He joined a group to burn down the Shia leader’s house, afterward proudly posting photos documenting the crime. When asked if he was afraid of being prosecuted, he said no. “Everyone is aware that the Shia are against our country,” he said. “Everyone talks about it online....they are criminals.”
- › **Social media may accelerate tit-for-tat.** Social media provocations tend to escalate into dynamics of one-upmanship between antagonistic groups. For example, amid the ongoing dispute between Egypt and Ethiopia over the Grand Ethiopian Renaissance Dam (GERD), a massive hydroelectric project that potentially threatens Egypt’s downstream water access, Egyptian and Ethiopian social media influencers have engaged in spiraling rounds of provocation. According to participants in this research, popular opinion inside Ethiopia is becoming more militant online and off-line as people rally around nationalist themes. Videos shared on Instagram—in which Ethiopians threaten to cut off Egypt’s water, and Egyptians threaten retaliation—rack up tens of thousands of views. Ethiopian respondents reported that the ratcheting up of online tensions is feeding dangerous nationalist discourses. On Twitter, prominent Ethiopia-based influencers have advocated for construction of the dam whether or not a diplomatic solution is reached, while prominent Egyptians have tweeted threats of war.<sup>20</sup>

20 <https://www.arabnews.com/node/1702516/media>

## 6. Sources of Resilience

### Factors that appear to mitigate digital threats

Social media does not solely divide. Social media can promote peace-building by forging connections between communities, improving awareness about conflict prevention, and empowering grassroots actors. Platforms offer spaces for marginalized or publicly maligned groups to “correct the story” and offer their own version of events. The diffusion of mobile devices, and the existence of social media platforms, expand opportunities for individuals and groups to hold political elites accountable, document discrimination and violence, and share information with the outside world. In Nigeria, a group of female activists emphasized that social media has opened vital new spaces for public engagement. “Social media,” said one, “provides a way for us to demonstrate our value.”

Resilience in international development refers to the ability of people and systems to advance and protect community well-being in the face of complex shocks and stresses, such as violent conflict, an economic downturn, or a climate-based disaster—events that may be exacerbated or weaponized in social media.

Resilience factors may be diverse in type, encompassing people, institutions, and norms. And many resilience factors that are common in peacebuilding generally may reduce opportunities for, and the impact of, platform manipulation. For example, online hate speech may be less likely to drive a wedge between communities where robust off-line intergroup peacebuilding mechanisms exist. Examples of conventional resilience factors include:



Photo Credit: Mercy Corps / Gaza 2013

- › *The presence of respected, peaceful local leaders, such as religious leaders and social media influencers, who can counter social media weaponization in online and off-line spaces.*
- › *Non-violent dispute resolution mechanisms that effectively facilitate resolution of intergroup tensions.*
- › *Inclusive governance and policy making, including community-informed processes for defining service and infrastructure priorities.*
- › *Interfaith organizations that build bridges across sectarian or ethnic divides.*
- › *Presence of community groups, which provide a channel for civic engagement and opportunities for intergroup collaboration.*
- › *A culture of positive intergroup interactions, as evidenced by trade, social interactions, intermarriage, etc.*

In addition, the case studies highlighted sources of resilience that are more specific to social media.

Examples include:

- › **Off-line actors who correct online rumors**, including civil society organizations working to counter digital threats, particularly disinformation and hate speech. In Nigeria, interfaith groups and some traditional media—such as Unity FM Radio in Jos—work to mitigate digital threats by fact-checking misinformation and disinformation circulating in social media.
- › **Online fact-checkers** working to counter false narratives in real time, flag inflammatory content for removal by social media companies, and/or provide non-partisan online spaces for intercommunal engagement. Examples include Ethiopia’s Ethio Check or Myanmar’s “Real or not?” online service.
- › **Digital reformers** who are pushing social media companies to address hate speech and disinformation. In Myanmar, a Facebook reform campaign is being spear-headed by civil society organizations to push the technology company to reform its practices and commit greater resources to combating disinformation and hate speech.
- › **Online discussion boards and marketplaces** that promote positive intergroup interactions. In northern Iraq, female participants described how online platforms facilitate local economic activity—by connecting buyers and sellers of used goods—and provide spaces where people can seek advice, share tips, and engage in the kinds of peaceful transactions that are critical to social cohesion and yet are difficult to promote in off-line spaces.
- › **Respected, non-partisan traditional media**, such as radio and newspapers, that adhere to professional journalistic standards, carefully vet stories, and counter disinformation and misinformation. For example, Myanmar participants highlighted Myanmar Now and The Myanmar Times for their integrity and relative lack of bias.
- › **Institutions that promote journalistic integrity**. For example, the Nigerian Press Council (NPC) organizes seminars to train journalists to identify fake stories and authenticate sources before publication.
- › **Online fora that promote intergroup dialogue**. Mediated online fora can create opportunities for different ethnic and sectarian groups to connect. For example, in Nigeria, the Kaduna Youth Forum was founded in 2012 to create a non-partisan Facebook page for young people from different backgrounds to communicate. The Forum now has 92,000 members.
- › **Digitally-savvy youth**. Across the country contexts, participants highlighted their large, growing youth populations, as vulnerable to conflict risks but also as providing potential opportunities. The rising generation tends to be more technically literate and skeptical of disinformation campaigns, and may represent an important bulwark against future efforts to weaponize social media.
- › **Even-handed internet and media regulations**. While a number of central governments have moved toward implementing hate speech laws that would likely be counterproductive, many participants believed there is space for more limited internet and media regulation that could reduce the volume of disinformation. However, they emphasized that such efforts should be transparent, moderate, and the result of proactive dialogue with civil society and grassroots actors.

## CONTACT

ADRIENNE BROOKS

Digital Communities Advisor | Technology for Development

[abrooks@mercycorps.org](mailto:abrooks@mercycorps.org)

WILL FERROGGIARO

Senior Specialist | Peace and Technology

[wferroggiaro@mercycorps.org](mailto:wferroggiaro@mercycorps.org)

ADAM LICHTENHELD, PHD

Senior Researcher | Peace and Conflict

[alichtenheld@mercycorps.org](mailto:alichtenheld@mercycorps.org)

## About Mercy Corps

Mercy Corps is a leading global organization powered by the belief that a better world is possible. In disaster, in hardship, in more than 40 countries around the world, we partner to put bold solutions into action — helping people triumph over adversity and build stronger communities from within. Now, and for the future.



45 SW Ankeny Street  
Portland, Oregon 97204  
888.842.0842

[mercycorps.org](http://mercycorps.org)