SOCIAL MEDIA, CONFLICT, AND PEACEBUILDING: ISSUES AND CHALLENGES

A Discussion Paper

Toda Peace Institute

IREX

fhi360
THE SCIENCE OF IMPROVING LIVES

SEARCH FOR COMMON GROUND

PCFK

MERCY CORPS

Acknowledgements

Citation
Ferroggiaro, W. Mercy Corps. 2021. Social Media, Conflict, and Peacebuilding: Issues and Challenges; Discussion Paper. Washington, D.C.: Mercy Corps.

Purpose

Violent conflict is on the rise,[1] and increasingly, social media is playing an important role:[2] political actors have used social media to target critics with disinformation, while COVID-19 misinformation spread online has spurred hate speech against stigmatized groups. Current off-line and online approaches and tools to address these challenges seem insufficient to the task. Moreover, there are real dangers inherent in some current responses, such as exposure of human rights defenders to attack and restrictions on speech by governmental authorities.[3]

Peacebuilders are increasingly addressing social media harms alongside other conflict drivers, but it remains a relatively new area of peacebuilding. However, there are lessons to be drawn from recent experience. Given that sources of conflict, whether actors or drivers, may originate in the off-line space, both off-line and online approaches and tools are necessary to address social media's role in conflict.

This paper aims to be a practical statement about the 'weaponization' of social media and peacebuilding responses to it. While it discusses some examples and highlights some innovations, the paper does not catalogue social media's various impacts on conflict[4] or serve as a reference guide to digital tools.[5] Rather, it attempts to articulate an emerging consensus among peacebuilders about what we know about working in this area, and very practically, what approaches are effective or promising. The goals are to provide a common starting point for addressing social media drivers of conflict for practitioners, both local and global; suggest some priorities for organizations, donors, and governments; and serve as a foundation for other related discussions.

What we know about the problem: social media drivers of conflict

Defining social media harms

A foundational challenge for peacebuilders – as well as those from other sectors concerned with social media and conflict – is the lack of agreement around concepts and definitions.[6] This is important because definitions help frame an issue and affect the design and implementation of responses, the ability to form coalitions to address challenges, and more. 'Information disorder' is used to refer to specific types of

---

[1] Uppsala University's Uppsala Conflict Data Program tracks the various types of violence (e.g., state-based, non-state). See https://ucdp.uu.se/.

[2] That social media is having complex, multifaceted, and important impact on conflict is increasingly documented; Lisa Schirch provides a useful overview of these impacts at https://toda.org/assets/files/resources/policy-briefs/t-pb-73_lisa-schirch_san-diego-report-social-media-impacts-on-conflict-dynamics.pdf. Nonetheless, the extent of social media's influence on various conflicts is an area still requiring systematic research.

[3] According to Freedom House's *Freedom on the Net 2021*, "officials suspended internet access in at least 20 countries, and 21 states blocked access to social media platforms," while "authorities in at least 48 countries pursued new rules for tech companies on content, data, and competition over the past year." https://freedomhouse.org/report/freedom-net/2021/global-drive-control-big-tech.

[4] For example, in Mercy Corps' June 2021 report *A Clash of Contagions: The Impact of COVID-19 on Conflict in Nigeria, Colombia, and Afghanistan*, misinformation about the origins of the virus or about a local government's response is cited as undermining social cohesion. In addition, as a guide to the ways social media contributes to conflict, Mercy Corps released *Social Media and Conflict: Understanding Risks and Resilience: An Applied Framework for Analysis* in August 2021.

[5] Regarding disinformation, for example, USAID's *Disinformation Primer*, published in February 2021, provides its users (primarily field officers) with a section on 'What are some emerging solutions for disinformation?' that itemizes the great variety of tools being used in different sectors.

[6] The UN Special Rapporteur on Freedom of Opinion and Expression recently noted the definitional challenge concerning disinformation. See https://undocs.org/A/HRC/47/25, p. 3.

information 'pollution' including misinformation, disinformation, and malinformation. But it also is commonly used to characterize the broader information environment within which those disorder types exist.[7] A general consensus has emerged that:

- **Misinformation** is content that is shared by someone who doesn't realize it is false or misleading.
- **Disinformation** is content that is created with malign intent and is shared in order to cause harm.
- **Malinformation** is content that is genuine and is shared purposely to cause harm, such as leaked medical records.

These are not hard and fast categories; for example, a person might unwittingly share misleading content that was created for disinformation purposes.[8]

'Social media' is commonly acknowledged as "a form of electronic communication and networking sites that allows users to follow and share content (text, pictures, video, etc.) and ideas within an online community."[9] Peacebuilders and others also speak of social media 'threats' and 'harms,' sometimes using them interchangeably.

- **'Threat'** may suggest more of a sense of potentiality and is frequently used where the threat may lead to conflict.
- **'Harm'** is often used to suggest both immediate harm, such as that felt by a target of online gender-based hate speech or more indirect such as the harm resulting from polarization, radicalization, or loss of public trust.
- **'Weaponization'** is a specialized term but also has various iterations: social media might be mobilized to demonize or target a political opponent, such as 'red-tagging' by Duterte allies in the Philippines; hate speech on social media may generate actual physical attacks; and social media might be used to recruit for armed conflicts.[10]

Consequently, peacebuilders will utilize different strategies and programs tailored to the different origins and intents of weaponized social media efforts as well as the impact they have.

The role of social media as a conflict driver

The role of digital technologies in exacerbating existing threats and creating new threats is still under-appreciated amongst peacebuilders in general. Indeed, issues as various as public health and climate change show the effect of digital threats on underlying issues (e.g., natural resource competition spurred by online disinformation that leads to off-line conflict). Likewise, how communication technologies, including social media, are used varies enormously by context and actor. Technology tools also present opportunities, advantages, and benefits.

---

[7] See the 2017 Council of Europe report "Information Disorder: Toward an Interdisciplinary framework for research and policymaking," https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.
[8] These definitions draw on those developed by First Draft, which is focused on media and civil society responses to misinformation. See https://firstdraftnews.org/long-form-article/understanding-information-disorder/.
[9] This definition is from Patrick Hiller's Peace Science Digest article, "How Social Media is Changing Conflict," at https://peacesciencedigest.org/social-media-changing-conflict/, citing Zeitzoff, T. (2017). How Social Media Is Changing Conflict. *Journal of Conflict Resolution*, 61(9), 1970–1991.
[10] There are a variety of definitions and perspectives on this concept. Mercy Corps, in its 2019 landscape assessment, noted four types of weaponized social media: information operations, political manipulation, digital hate speech, and radicalization and recruitment. See https://www.mercycorps.org/research-resources/weaponization-social-media. For an engaging read on this phenomenon, see P.W. Singer and Emerson T. Brooking's 2019 book *Like War: The Weaponization of Social Media*, New York: First Mariner Books.

Nigeria, Boko Haram, and social media

Technology has many uses in peacebuilding, and peacebuilders have been early adopters of digital tools: In the Nigerian context, for example, peacebuilders are using social media interactions, analytics and online presence to identify peacebuilding influencers and opinion shapers with genuine and well-established links to Boko Haram. These people are then engaged for mediation purposes, and to test the receptiveness of conflict actors to peacebuilding initiatives. Social media is also used by peacebuilders in this context to understand alternative perspectives to the conflict because generally, the Nigerian military and by extension the Nigerian government, have the tendency to monopolize the narratives about the conflict with Boko Haram.[11]

The breadth of social media's reach and its adoption, the volume of content that can be shared or disseminated, the speed of transmission, and algorithmic amplification[12] – are all technological game-changers for conflict. However, these same characteristics also continue to be repurposed for constructive ends, as peacebuilders have demonstrated.

Evidence indicates that misinformation, disinformation, and hate speech readily jump platforms, while individual users also increasingly adopt multiple social media platforms and share content across platforms. Indeed, for disinformation agents this dynamic may be a primary element of their overall campaign. In addition, in contexts where journalists lack ethical and professional norms, such information may be posted across media modes without critical review. Even in contexts where access to the internet is limited, available technologies such as a mobile phone or DVD have been used to disseminate and promote content, such as anti-Muslim hate speech by supremacist Buddhist monks in villages in Myanmar.[13]

Some weaponized social media efforts may be specifically targeted on underlying problems with which peacebuilders are familiar but where the ultimate motivation of malign actors has been less of a focus, e.g., disinformation campaigns to exploit racial discrimination or tensions in order to more broadly undermine public trust as was done by Russian government-backed entities in the US during the 2016 election campaign.[14]

A lack of evidence about how to address social media drivers of conflict

There is a need for more evidence and assessment about what peacebuilding approaches and tools work to address social media harms and why. Indeed, there are fundamental questions as to whether interventions against such harms actually reduce violence or conflict events - beyond evidence that they can alter certain behaviors or change perceptions. These latter impacts may matter for reducing the risk of social media fueling or exacerbating conflict but there is insufficient evidence linking them to actual changes in violence-related attitudes or events. The need for increased research on this topic is increasingly urgent because of the ever-changing technology as well as social adaptations to it. There are several reasons for this lack of evidence, among them that the field remains comparatively young, it is not sufficiently defined for categorization as an academic discipline, and there is some lack of agreement over how to measure impact.

Another challenge for assessment, but also for building the capacity in the sector, is insufficient data-sharing and information exchange, including between global and local actors. However, the underlying obstacle here is a lack of consistent, quality data-sharing by social media and technology companies about the information on their platforms and the ways in which it is shared and amplified. Platforms do not readily share this information for a variety of reasons, including fear of regulatory action, public backlash around how they

---

[11] For a further discussion of this case, see Malefakis, Medinat. 2021. "Social Media Dynamics in Boko Haram's Terrorist Insurgence" in Schirch Lisa, *Social Media Impacts on Conflicts and Democracy*, New York: Routledge Publishers. This work features case studies from a variety of countries.

[12] See https://www.newamerica.org/oti/reports/report-series-content-shaping-modern-era/.

[13] In 2013, before telecoms and the internet were widespread in Myanmar, radical Buddhist monks like Wirathu were composing anti-Muslim speeches and sharing them via DVD and other means. https://www.theguardian.com/world/2013/apr/18/buddhist-monk-spreads-hatred-burma.

[14] See https://www.intelligence.senate.gov/sites/default/files/documents/Report_Volume2.pdf, especially pp. 38-40.

manage content on their services, and concerns around disclosing information that would shed light on the design and operation of their business. In addition, state-run information operations are often countered or mitigated by national authorities and given their national security implications, are not publicly disclosed. A final challenge is that research and learning has not been a donor priority, or where it is a priority of donors, the focus is on investigating specific responses rather than on the role information disorders play on conflict dynamics. However, this may be changing given the prominence of information disorders in the global public debate.

Organizations involved in digital peacebuilding seeking to address social media harms may focus on disparate aspects of peacebuilding and may not share the same analysis of a problem or approaches to addressing it. This is in part due to the aforementioned lack of agreement on definitions and concepts, the scale of the problem, and the lack of data and evidence about the problem and the efficacy of solutions. But it is also due to legitimately different 'lenses' on the issue given that it affects so many areas of life: freedom of expression advocates may view the challenge differently than a group concerned with the rights of those stigmatized by online hate speech. Perhaps more prosaically, they are only able to focus on that aspect of the problem which they understand and have a mandate to address. Consequently, it is difficult to come together to wield influence, especially in advocacy before government regulators and legislators or before technology companies.

Which responses have been effective, and which are promising?

Knowledge sharing; skill-strengthening; and tools for communities and individuals to analyze their context and to design their own responses to digital harms are more promising over the long term than directly countering specific instances of misinformation, hate speech, or radicalization. First, the former have generally been grounded in the lived realities and agency of conflict actors and affected communities, while countering has tended to be more concerned with a narrative or ideology (in terms of governmental authorities) or with regulations and sanctions (in terms of technology companies). Also, despite misinformation or hate speech being rampant across many online platforms, technology companies have not sufficiently invested in staff who can adequately monitor and moderate harmful language.[15] In addition, local initiatives, in part due to their proximity to underlying conflict issues, increasingly see the benefit of building resilience to digital harms among communities and society with their peacebuilding. Nonetheless, a challenge for both sets of responses is sustainability; local groups are often dependent on external donors, who have their own timeframes, funding requirements, and priorities.

Addressing the weaponization of social media requires holistic programming that spans both online and off-line spaces. While a threat may manifest online, it likely has roots or proximate causes off-line and thus, peacebuilding responses to online harms need to be focused there as well. For example, PeaceTech Lab developed a methodology and lexicons that not only identify hate speech occurring online in a country, but analyze the context – from past violent incidents to ongoing corruption and governance issues – within countries from Iraq to South Africa that when amplified online may incite discrimination, hostility, or violence.[16]

As indicated above, there is a need for increased research and assessment as to what types of programs work and in what contexts – yet there is evidence of impact in a number of areas. For the purposes of this

---

[15] See for example https://www.mei.edu/publications/flaws-content-moderation-system-middle-east-case-study and https://www.mei.edu/publications/content-moderation-trends-mena-region-censorship-discrimination-design-and-linguistic.
[16] Since 2016, PeaceTech Lab has produced more than one dozen country-focused lexicons of hate speech. See https://www.peacetechlab.org/hate-speech.

paper, these are grouped as follows: Monitoring, detection and assessment; Mitigation; Prevention; and Building Resilience.

Monitoring, detection, and assessment

Civil society has produced innovative efforts to monitor and detect social media threats and harms, the networks where they persist, and the actors who perpetrate them. These innovations include: fact-checking initiatives and organizations; information and threat mapping; rumor monitoring and management; identification and analysis of hate speech; and social network monitoring, analysis, and reporting. They have provided important insight concerning online dynamics as well as the context and shape of conflicts. Still, the efforts by civil society, media, and researchers have limitations in terms of their ability to design timely responses to the analytics they provide as well as the narrowness of their audience. However, these limitations are in part due to the tools available for this task and in part due to the companies, as they provide insufficient transparency around their policies and practices and the impact of these efforts.[17] Many of the tools that peacebuilders use to analyze this space have been built for commercial purposes and are limited in their ability to monitor phenomena like tensions or polarization; rather, they are primarily made to monitor sentiment around products. Moreover, while social media platforms have themselves designed and developed internal monitoring mechanisms, it is not possible to verify whether the companies' automated tools are actually able to detect harmful content equally across languages or to confirm beyond company statements whether or what action has been taken against harmful content that can incite violence.[18] Without such information, civil society is hindered in conducting analyses and in making responsive recommendations.

Mitigation

Social media harms can not only exacerbate conflict but can undermine efforts to prevent or mitigate conflict. Organizations ranging from NGOs to government ministries have developed crisis plans to respond to weaponized social media, yet many entities still lack mechanisms to mitigate against these risks. In addition, these organizations are developing their plans under the assumption that they have control over what happens on social media platforms, but in reality, what content gets shared, what gets amplified, and what does or does not get taken down is at the discretion of the platform in most cases. This inability to control or regulate content and design of social media has led to some government efforts to attempt to change this power balance with a resulting infringement upon free expression and privacy rights.[19]

Civil society has pioneered referral or warning and response mechanisms for mitigation. For example, Ushahidi's early crowd-sourced reporting platform in Kenya inspired subsequent generations of mobile reporting, including initiatives to involve enforcement responses by governmental authorities.[20] In large part through the pressure of civil society, companies have established trust and safety mechanisms for civil society input into the platforms. 'Trusted flagger' programs like those of YouTube, Twitter, and Facebook provide selected civil society groups priority review by the company of material they believe is abusive, inciting, or violent. However, groups must be able to access the companies, which is difficult in certain countries, and must be vetted; even if they are approved and able to contribute, company reviews of these flaggers' alerts may not be timely for action particularly for non-English language reporting.[21]

---

[17] New America's Open Technology Institute developed a platform Transparency Report Tracking Tool at https://www.newamerica.org/oti/reports/transparency-report-tracking-tool/.

[18] See https://www.mei.edu/publications/flaws-content-moderation-system-middle-east-case-study, and https://www.mei.edu/publications/content-moderation-trends-mena-region-censorship-discrimination-design-and-linguistic.

[19] The multi-stakeholder Global Network Initiative applied a human rights framework to its assessment and recommendations regarding government efforts on content regulation in this 2020 policy brief https://globalnetworkinitiative.org/wp-content/uploads/2020/10/GNI-Content-Regulation-HR-Policy-Brief.pdf.

[20] Among the efforts for Kenya's 2017 election, Mercy Corps and PeaceTech Lab supported local civil society early warning and response efforts in four counties. See https://www.peacetechlab.org/prevent-election-violence.

[21] See again https://www.mei.edu/publications/flaws-content-moderation-system-middle-east-case-study and https://www.mei.edu/publications/content-moderation-trends-mena-region-censorship-discrimination-design-and-linguistic.

In addition, significant focus and resources have been expended on addressing/countering polarized or extremist narratives online by government or government-sponsored entities, yet there is insufficient evidence of its effectiveness. Some key challenges are that counter-messaging often aims to undermine beliefs or narratives rather than present alternative information and that the messaging itself is insufficiently sophisticated for the complexity of the audience it targets.[22] Nonetheless, some efforts, such as the 'Redirect Method' pioneered by Moonshot CVE, have shown promise,[23] but in that case as others, there is insufficient data as to its effectiveness. Conversely, some alternative and positive narratives have shown effectiveness and influence in a variety of contexts.[24]

Prevention

In focusing on reducing the frequency of incidences and the scope of weaponization, there have been efforts, often controversial, to use law and regulation to set parameters for intermediaries as well as initiatives to establish norms and set guidelines concerning content or product design.[25] There are several challenges and dangers with the former. Efforts by national governments through legislation or decrees concerning the internet and social media have frequently had negative effects, especially in authoritarian and repressive contexts, where they have been used to threaten freedom of expression and privacy (e.g., platform bans or takedowns, internet throttling/shutdowns) and attack social activists and political opponents (e.g., Law Against Hate in Venezuela and Law of Cybercrimes in Nicaragua). While the UN Special Rapporteur on Freedom of Opinion and Expression and his regional counterparts have warned governments to provide a legal basis for any restrictions on the internet and intermediaries and should have explicit policies about content blocks or takedowns,[26] they lack the ability to enforce compliance.

In addition, civil society has played an essential role in creating guidelines and best practices for internet platforms on issues such as transparency of operations and policies.[27] However, this requires sustained effort, technical skill sets, and resources which are often beyond the scope of NGOs, peacebuilding initiatives, and broad-based coalitions. Other civil society advocacy, including hashtag campaigns on social media platforms, has resulted in improvements in community standards established by technology companies. Despite the intent of some political actors to use regulation for their political ends, there is nonetheless a growing lack of public trust with platforms and a declining belief in the self-regulation model in a variety of countries. As a result, advocates are increasingly turning to governments to pass reasonable legislation.[28]

Building resilience

Efforts in these areas are focused on building resistance to the weaponization of social media in society, particularly amongst and affecting vulnerable populations. Media and information literacy initiatives, such as

---

[22] The Institute for Strategic Dialogue produced a useful reference on this topic at https://www.isdglobal.org/isd-publications/the-impact-of-counter-narratives/.

[23] See https://www.rand.org/content/dam/rand/pubs/research_reports/RR2800/RR2813/RAND_RR2813.pdf.

[24] For a discussion of 'positive interventions,' see https://gifct.org/wp-content/uploads/2021/07/GIFCT-CAPI2-2021.pdf.

[25] See again https://globalnetworkinitiative.org/gni-content-regulation-hr-policy-brief/, as well as https://foundation.mozilla.org/en/campaigns/trained-for-deception-how-artificial-intelligence-fuels-online-disinformation/.

[26] The UN Special Rapporteur, joined by counterparts with the OSCE, OAS, and African Commission on Human and Peoples' Rights, issued the warning in March 2017. https://www.osce.org/files/f/documents/6/8/302796.pdf.

[27] For reference, see https://santaclaraprinciples.org/; https://rankingdigitalrights.org/index2020/; https://www.newamerica.org/oti/reports/transparency-reporting-toolkit-content-takedown-reporting/; and https://www.newamerica.org/oti/reports/report-series-content-shaping-modern-era/.

[28] At this writing, a key development to watch concerns the EU's Digital Services Act, which would provide some degree of regulation of social media platforms. See https://www.newamerica.org/oti/blog/the-eus-digital-services-act-makes-a-positive-step-towards-transparency-and-accountability-but-also-raises-some-serious-questions/.

IREX's 'Learn to Discern,'[29] have shown evidence of increasing the ability of consumers of both online and off-line media and information to identify hate speech and disinformation and assess the credibility of information as well as change their own information engagement behaviors to become more responsible and empathy-driven. Digital literacy efforts have provided increased understanding of the uses of information technology, including social media platforms, behavior and norms in online environments, and social issues affecting online environments. The challenge that these human-centered approaches face is ensuring accessible formats and incentives for people to build this resilience. This is especially acute in conflict or fragile environments where people may lack digital or formal literacy, funds for internet subscriptions, and reliable electricity. While it is possible to reach new generations through education systems, at scale efforts for decision-making adults, especially in conflict contexts, require innovative partnerships with trusted local leaders and institutions. Given this need for scale and the need for equitable access and administration, these efforts may well be a government function (although resources from partnerships with the private sector might enhance reach and adoption).

In addition, awareness campaigns, whether online (e.g., hashtag issue campaigns on particular platforms) or off-line (e.g., enlisting artists and musicians to spread awareness through community events) have shown impact in addressing weaponization of social media in a variety of contexts. The COVID-19 pandemic has provided evidence of non-traditional information sources such as social media influencers on TikTok enlisted to share official and factual information to counter misinformation and rumor.[30] In addition, there is some innovation occurring, particularly in the social behavior change communication area, to apply an 'inoculation theory' approach where audiences are exposed to small doses of disinformation (e.g., through 'edutainment' programs or interactive games like Harmony Square) that then help build their resilience to the phenomenon.[31] Finally, social media videos and memes that promote other media and information literacy tips have been shown to be effective in reducing engagement with disinformation about elections, especially among the most partisan news consumers.[32]

Traditional and non-traditional off-line programming, particularly those in the areas of social cohesion building and civic engagement, also provide opportunities to address the impact of social media on conflict, such as in addressing online and off-line stigmatization of war returnees.

Next questions

Conflict is increasingly multidimensional and interconnected; social media both reflects and contributes to this phenomenon with its own online dynamics and off-line impacts. Peacebuilders are also increasingly employing more holistic approaches for addressing conflict. The COVID-19 pandemic has exposed weaknesses and vulnerabilities of societies; perhaps it also suggests the need to work differently to build sustainable peace. Social media is immensely popular and has tremendous reach, particularly amongst younger generations; for example, Big Brother Nigeria generated nearly 700 million digital impressions over

---

[29] The approach is described at https://www.irex.org/project/learn-discern-l2d-media-literacy-training.
[30] Finnish national emergency authorities utilize social media influencers to disseminate information on COVID-19 and other emergency alerts. See https://www.politico.eu/article/finland-taps-influencers-as-critical-actors-amid-coronavirus-pandemic/.
[31] Harmony Square is explained as a tool that can help 'inoculate' against misinformation in this November 2020 *Misinformation Review* article https://misinforeview.hks.harvard.edu/article/breaking-harmony-square-a-game-that-inoculates-against-political-misinformation/. The game itself can be accessed at https://harmonysquare.game/books/default/.
[32] The RAND Corporation report, *Russian Propaganda Hits Its Mark: Experimentally Testing the Impact of Russian Propaganda and Counterinterventions*, utilized IREX's media literacy messages developed as part of the Learn to Discern (L2D) approach and revealed that Kremlin propaganda content is effective because it generates a strong emotional response with a stark partisan divide and that even brief exposure to media literacy can shift the behaviors of hard-to-reach groups.

two months. Given its inherent characteristics, what are the new or forthcoming opportunities where social media might contribute to conflict - or be leveraged to build peace?[33]

Peacebuilders have pioneered innovative programming in response to social media harms, yet they are often not at scale to address the problems. From off-line civic awareness campaigns to fact-checking initiatives or advocacy inputs on content moderation, these innovations have not proved sufficient to reigning in the threats or minimizing harms[34]- although they have empowered groups, catalyzed new types of coalitions, and helped shed light on the scale of the problems. How do peacebuilders grapple with this challenge of scale, including the business models that have produced it? Do conflict analysis guidelines, conflict transformation frameworks, or other frames from the broader peacebuilding sector offer insight?

If peacebuilders want to generate meaningful interventions for social media harms, they will also need to develop expertise in content and product design as well as public policy associated with both. For many organizations, these issues may seem beyond their scope and mandate, but given the scale of current challenges and foreseeable trends, peacebuilders will need access to this expertise. By contrast, civil liberties groups have moved from concerns about scope to developing this expertise, and they are better able to advocate. Where and how should peacebuilders focus this development of expertise? Where does collaboration with key actors in adjacent sectors, such as civil liberties advocates, come in?

Finally, there are broader questions about the power that social media platforms have in countries and globally. For example, Facebook effectively *is* the internet in some countries, i.e., people use it alone for commerce, news, or social connection, while in still others it serves as the primary news source for people. Consequently, the platforms touch every facet of life, yet the broader public – and the political leadership of countries – know little about how they operate. Given their size and influence, and the opacity with which they operate, how do peacebuilders and ordinary citizens hold technology companies to ethical standards and push governments to enact reasonable reforms? How can technology companies be held accountable for their roles in fueling online and off-line conflict?

Recommendations

Even as peacebuilders look to address new questions about their work on social media and conflict, this work is influenced by the contexts and environments in which it occurs. The recommendations below are offered as reflections to help guide this work.

Recommendations for peacebuilding practice:

1. <u>Analyze and address the off-line/online nexus</u>. Access to online spaces is increasing, diffuse, and inexpensive. The line between off-line and online is dynamic and permeable; traditional off-line conflicts such as cattle rustling are discussed online, amplified, and incite reactions on the ground--which then are played back online. Consequently, peacebuilding responses to online harms need to also reflect understanding of this dynamic, incorporating analysis of each sphere and tailoring responses to both online manifestations and off-line origins.

---

[33] Lisa Schirch identifies 25 ways technology can contribute to peacebuilding goals and outlines the evolution of technology's impact on peacebuilding in this concise paper
https://toda.org/policy-briefs-and-resources/policy-briefs/25-spheres-of-digital-peacebuilding-and-peacetech.html.
[34] For insight into the problem of scale, New America has examined the business models of the platforms as well as their own internal efforts to address the problems. See https://rankingdigitalrights.org/its-the-business-model/;
https://www.newamerica.org/oti/reports/how-internet-platforms-are-combating-disinformation-and-misinformation-age-COVID-19/.

2. <u>Employ broad-based, holistic approaches to the multifaceted challenges of social media</u>. The characteristics of social media that make it potent for conflict are beyond the scope of any one entity or sector to address as they have their own mandates and finite resources. Given the multifaceted nature of the problem, there is learning and practice to be shared across sectors. Thus, the nature and scope of social media and conflict necessitates increasing collaboration from design and implementation to advocacy and reform. This must begin with dialogue amongst diverse actors in order to align definitions, perspectives, and approaches.

3. <u>Prioritize longer-term and locally-driven resilience efforts</u>. Peacebuilding responses to social media harms are increasingly focused on building resilience within societies. These resilience tools are more effective when they are designed for the long-term and when they are locally-driven as local peacebuilders are best placed to understand context and potential for success. Donors should prioritize and adequately resource this focus. If peacebuilders are to tackle the off-line/online dynamics with holistic approaches, the people most affected by social media harms and threats need to be the central actors in the response.

Recommendations for peacebuilders' engagement with other actors:

1. <u>Ensure increased and consistent access to key data</u>. A key challenge to further understanding of social media dynamics and its impact on conflict is the need for data. Limited, ad hoc access to data generated by platforms is insufficient for peacebuilding responses and is not in the self-interest of the companies over the long-term. As a first step, peacebuilders should agree on common metrics for access to data, then identify what information they need. This would be made easier if social media companies enabled greater transparency about their organizations and operations. Existing or newly-established coalitions of peacebuilding organizations can provide guidance - and may be able to suggest interim or hybrid approaches such as vetted researchers. After that, social media platforms should then engage publicly and consistently on data access.

2. <u>Obligate states to preserve and bolster civic space</u>. Perhaps a first place to start in addressing social media impacts on conflict is the political and civic space in which they occur. Foremost, this means ensuring legal and political protections for civic activists, media representatives, researchers, and the public at large to be free from off-line threats and violence. This need knows no geographic or ideological boundaries, although it is most acute in states experiencing authoritarianism. While peacebuilders, both digital and those focused on other areas, should pressure governments for these protections, it is first incumbent on these governments that they live up to their existing legal and treaty commitments on human rights. To this end, UN and intergovernmental leaders recently issued a joint statement noting that "Political statements enjoy a high level of protection under international law, and that places a responsibility on politicians and public officials to uphold the rule of law, human rights, media freedom, intercultural understanding, and public trust in democratic systems of governance." States and their political authorities have an "obligation to create an enabling environment for freedom of expression and of the media, and the right to information."[35] It is possible that bolstering these protections, and promoting those efforts as they occur, will have positive online impacts as well, helping foster a safe and more productive online space.

---

[35] Organization for Security and Cooperation in Europe, "2021 Joint Declaration On Politicians And Public Officials And Freedom
Of Expression: Declaration by the United Nations Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression, and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information," 20 October 2021. See https://www.osce.org/files/f/documents/9/4/501697.pdf.

CONTACT

WILL FERROGGIARO
Senior Specialist | Peace and Technology
wferroggiaro@mercycorps.org

ADRIENNE BROOKS
Senior Advisor | Technology for Development
abrooks@mercycorps.org

LISA INKS
Director | Peace and Conflict
links@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.

MERCY CORPS

45 SW Ankeny Street
Portland, Oregon 97204
888.842.0842
mercycorps.org