JustPeace Labs' Submission to the Special Rapporteur Report on Freedom of Expression in Times of Armed Conflict and other Disturbances

July 11, 2022

Introduction

JustPeace Labs welcomes the opportunity to provide a submission to the Special Rapporteur on the issue of Freedom of Expression in Times of Armed Conflict and other Disturbances. We hope that our submission can help inform the Special Rapporteur's scoping report for submission to the 77th session of the UN General Assembly in October 2022.

We have been working with multiple stakeholders over the past several years to build awareness of the challenges posed by technology in conflict-affected situations, provide practical recommendations for policy and practice, and build a community of practice around these issues. We invite you to review previous publications and research on these issues in addition to the submission below, including:

- Comparing Guidance for Tech Companies in Fragile and Conflict-Affected Situations
- Technology in Fragile Contexts: Engagement, Partnerships, and Positive Action
- Technology in Conflict: Conflict Sensitivity for the Tech Industry
- Peacebuilding, Extremism, and Social Media, Part 1: A Problem
- Peacebuilding, Extremism and Social Media, Part 2: Social Media Account Suspensions
- Peacebuilding, Extremism and Social Media, Part 3: Algorithms

Disinformation, misinformation, and propaganda: Creating Digital Risks, Conflict, and Social Cohesion

Media—whether print, radio, television, or other communications systems—has long been used to cause harm and incite people to violence.[1] For example, the Radio Télévision Libre des Mille Collines spread hate speech before and during the 1994 Rwandan genocide, leading to convictions for the incitement of genocide before the International Criminal Tribunal for Rwanda.[2] However, the emergence of social media platforms and other digital technologies pose new and dire threats to countries around the world. Digital technologies allow false, deceptive, and dangerous speech to spread, target, and influence people at a speed, precision, and scale never before experienced.

Compared with legacy media, digital technology is faster, globally accessible, more affordable, simpler to use, searchable, mostly unmonitored or edited, and offers opportunities for both public and private conversations. Digital technologies enable vast new ways to track a user's location and data. Social media platforms operate largely on a social confidence method of information verification; people endorse

information on social media by sharing it with their friends. The rapid growth of new technologies is also unique. New forms of artificial intelligence and machine learning, for example, change social media algorithms that feed unique digital content to each separate user.[3]

Weaponizable digital technologies cause "digital harms" to individuals, communities, and states, including through:

- Cyberbullying and hate speech that dehumanizes individuals or groups (groups using slurs against ethnic or religious minority groups);
- Dangerous speech that threatens individuals or groups with real-world physical violence or harm (gangs or militias calling for violence against an individual or group);
- False or distorted information that leads to health risks;
- False or distorted information that leads to physical attacks on individuals or communities;
- False or distorted information that aims to undermine public trust in institutions or democratic elections; or
- Privacy violations that share personal information in ways that may reveal the location of individuals or communities under threat or enable cognitive and emotional manipulation through cognitive warfare.[4]

These harms are not limited to just areas affected by armed conflict as defined by international humanitarian law (IHL). They are also prominent in other communities that experience a lack of human security or social cohesion, and sometimes act as a precursor to more widespread forms of violence or the emergence of an outright armed conflict.

As documented in *Social Media Impacts on Conflict and Democracy: The Techtonic Shift*,[5] Indian social media users spread rumors accusing two men of kidnapping local children, leading to them being killed by a mob. In Brazil, false rumors about a political candidate reached millions of people all over the country on WhatsApp. In Zimbabwe, the government searched social media posts to enforce its ban on critiquing the government. In Northern Ireland, groups of youth sent messages to each other to organize fights along the peace lines that had divided their city. In Colombia, people posted messages spreading false information about the peace process. In Venezuela, the government created an ID system that linked food distribution to social media accounts, suggesting that people who "tweeted" a positive thing about the government might get access to food. In Myanmar and Venezuela, the governments set up troll armies to harness the power of social media in ways that would undermine democracy and human rights.[6]

There is devastating evidence of how social media was used to coordinate and direct hate-based violence in the United States[1] and promote a terrorist attack in New Zealand.[2] Social media has also been used to further large-scale human rights abuses, armed conflict, and mass killings in places like Myanmar[3], India[4], Sri Lanka[5], and elsewhere.[6] Governments are weaponizing internet access in conflict-affected and restive areas.[7]

Technology Companies and Digital Harms in Fragile and Conflict-Affected Settings

There are many ways that tech companies are inadvertently contributing to conflict dynamics through product design and release decisions. They can directly facilitate harm, incentivize harm, fail to conduct human rights due diligence, or fail to act to mitigate when they knew or should have known about potential harms. Sometimes technology products are used by third parties to intentionally foment conflict and abuse. Content moderation on social media platforms can also exacerbate a conflict. So can following government orders to shutdown internet services or collect and process sensitive data. Some business models reinforce structural inequalities and enflame conflict drivers. Sometimes just releasing a product or service in a conflict-affected market can have adverse impacts on the conflict.

Perhaps most importantly, many large tech companies operate on a profit model that rewards the amplification of outrage and disinformation. Social media offers users free access in exchange for their attention and data. Tech giants extract private information from users and then sell this information to advertisers, who pay tech companies to target their ads to specific users. Some companies design their products to keep users hooked—or even addicted—to these technology platforms. User attention is at the center of the profit model. In the "attention economy," tech companies require user attention to extract more private information to sell to political or business advertisers, and to show their ads to more people.[7] False, distorted, hateful, and violent content keeps user's attention. The economic model of many tech platforms correlates profits with user outrage in what some refer to the technology "race to the bottom of the brainstem."[8] The very core of many tech companies' business models can contribute to conflict.

Defining Jurisdictions Impacted by Digital Risks

Digital risks are impacting all countries, but some areas are more at risk than others due to preexisting factors. We focus on the use of digital technologies that amplify the spread of harmful information in "at-risk countries" or "fragile and conflict-affected situations" (FCS).[9]

Different companies measure what they consider "at-risk" in different ways based on unique risk tiering criteria. A common challenge companies face is defining risk categories and translating that to existing internal methods for allocating resources and measuring impact. Some focus on immediate threats of violence and physical harm, and others focus on other typs of digital harm stemming from hate speech or misinformation. Another common challenge is being able to indentify risk and engage corporate human rights policies before a situation intensifies or a disturbance evolves into outright conflict.

At-risk countries are by definition complex and dynamic by nature. They involve multiple, interconnected actors, drivers, and motivations; and many are based on long-standing, historical grievances. The absence of overt violence does not necessarily mean there is peace—situations are impacted by invisible social, political, and economic tensions. Situations of social unrest and cycles of violence can emerge with little warning and spark more intense and widespread conflict. Some conflict and human rights issues will be more prevalent in some contexts or developmental phases of a product than others.

JustPeace Labs welcomes the recently published "Heightened Human Rights Due Diligence for Business in Conflict-Affected Contexts: A Guide." Nevertheless, additional

guideance is needed, specific to the technology industry. In particular, companies need guidance on defining what a "conflict-affected" area is, and what events should trigger enhanced due diligence. Defining conflicts can be tricky and is debated even among conflict experts and scholars. To help, some organizations provide industry-specific guidance and lists. For example, the OECD Due Diligence Guidance includes supplements that provide a list of "red flag" situations related to mineral extraction that trigger the need for enhanced due diligence; the World Bank publishes a list of Fragile and Conflict Affected Situations (FCS).[14] The UN Working Group on the issue of human rights and transnational corporations and other business enterprises indentifies other circumstances which should trigger enhanced or heightened human rights due dilgience.[17]

However, these lists do not always pertain to the technology industry, the unique types of digital harm social media poses, and the lack of human security and social cohesion that may precede "fragility" or violent conflict. There are reports of digital risks in nearly every country on the planet. Even in countries that are mostly peaceful, there are communities and cities within those countries that may face unique digital risks to social cohesion. Using an indicator such as the number of deaths or preexisting human rights abuses or conflicts may not accurately measure the level of digital risks. While the number of deaths may be relatively low in general, a social media campaign to spread disinformation about electoral integrity could, for example, trigger public protests that could not only be deadly but could put a country's democratic institutions at risk. The costs of digital disinformation and hate speech on public trust in democratic institutions and social cohesion may be putting most or all societies at risk of public violence.

Insufficient Legislation and Legal Regulation

Existing regulation (both formal and informal or voluntary) may fall short in addressing the digital challenges related to fragility, conflict, and social cohesion. Tech companies are generally expected to self-regulate, whether by adopting codes of ethics, human rights due diligence processes, or similar. But these efforts have proven to be ineffective in many regards.

Having clear, enforceable, and rights-based rules would be an ideal approach for mitigating the risks of technology in society, especially as they relate to conflict. However, regulatory efforts to date have largely been reactive, slow, and focused on specific technologies (such as artificial intelligence) or issues (such as freedom of expression). Many jurisdictions are only starting to pass regulations that specifically address the risks posed by social media and other emerging technologies. At the time of writing, there is no international regulation specifically addressing the risks of technology in general or the specific risks of technology related to conflict. Such a multilateral effort remains well outside the realm of political feasibility at the time of writing.

Most applicable legislative and/or regulatory frameworks exist only at the domestic or regional level, although some states and cities are leaders in this space. As such, existing regulations are jurisdictionally narrow—and therefore limited in their ability to address a global problem. There is of course always the possibility that domestic laws which require certain compliance in one jurisdiction or with respect to that jurisdiction's users will lead to wider extraterritorial reach; for example, it was once thought that the implementation of the GDPR might lead to companies applying increased privacy protections for users across jurisdictions. In practice, however,

companies are seeking to limit the GPDR rules by moving user agreements to less restrictive jurisdictions.[19]

Mandatory human rights due diligence (mHRDD) legislation, where it exists, refers almost exclusively to conventional supply chains.[20] For example, the UK Modern Slavery Act and the California Transparency in Supply Chains Act apply exclusively to the very specific issue of forced labor in traditional goods and services supply chains. They do not present much opportunity to address the challenges posed by online platforms in areas affected by conflict. The French Duty of Vigilance Act, on the other hand, is much more broadly applicable across sectors and therefore could present a potential opportunity to advance the respect of human rights by online platforms, although as a practical matter this has yet to be tested.[21] Passed in 2017, it makes French multinational companies civilly liable for human rights violations committed by its subsidiaries, suppliers, and subcontractors, regardless of their jurisdiction. While the first of its kind and an ambitious first step into regulating human rights due diligence, a group of civil society organizations found that in its first two years, the law was ineffective and poorly implemented.[22]

Legal and regulatory efforts that specifically address conflict-affected areas, while well-meaning, can also risk unintentional consequences. For example, the US Dodd-Frank legislation requiring certain companies to disclose their use of conflict-minerals reportedly had negative impacts on the local communities it was intended to protect. [23] Some companies considered that it imposed too significant a compliance burden and weighty risk of legal or financial liability and opted to simply withdraw from those jurisdictions, proving detrimental to those local communities already suffering from conflict.[24] Legislation like this can also open local markets to other, less scrupulous companies—or local militia groups, as reportedly happened in the DRC.[25] While such legislation can have very positive impacts, it also risks exacerbating some conflict dynamics.

Recommendations

JustPeace Labs, together with Business for Social Responsibility, is currently engaged in further research to understand and make recommendtions for how technology companies can enhance existing human rights due dilgienc practices to best address the challenges of protecting freedom of expression and privacy in the context of conflict and other disturbances. We will be publishing those recommendations in September 2022. Our initial findings, however, demonstrate that particular attention is needed to understanding how to define "conflict" and what contistutes a "disturbance," and to help companies understand the risks and impacts related with their products and services in those areas. This includes the need for guidance on how to assess the impact on a conflict situation of any mitigations or actions taken to avoid negative human rights impacts—often, the steps taken to protect human rights can in turn exacerbate conflict. This requires more research, analysis, and guidance. Another critical need is to develop equitable and horizontal pathways for sustained engagement with rights holders and civil society, *before* fragile situations turn into conflict. Finally, particular attention needs to be paid to the providing guidance to companies on how to assess the delicate balance between protecting rights, such as freedom of expression and privacy, in light of larger conflict issues, especially in situations where local laws might contravene or work against international human rights law in these considerations.

We believe that by working with industry to define when enhanced due diligence would apply, understand the impacts of social media on conflict and social cohension with advanced research and data, and improving policies at government level to define and regulate what companies need to do in these situations, we can better address these challenges.

Citations

[1] Theo Dolan, Preventing Media Incitement to Violence in Iraq, USIP Peace Brief, April 7, 2010.

[2] The Prosecutor v. Ferdinand Nahimana, Jean-Bosco Barayagwiza, Hassan Ngeze, Case No. ICTR-99-52-T, Appeals Judgement, 28 November 2007.

[3] Schirch; 7-9.

[4] Lisa Schirch, editor. (2021). Social Media Impacts on Conflict and Democracy: The Techtonic Shift. Sydney: Routledge Press.

[5] Ibid.

[6] See chapters in Social Media Impacts on Conflict and Democracy (2021) by Spandana Singh (India); Diego Casaes and Yasodara Cordova (Brazil); Tendai Marima (Zimbabwe); Brendan McCourt (Northern Ireland); Diana Dajer (Colombia); Iria Puyosa (Venezuela); and Victoire Rio (Myanmar).

[7] Shoshana Zuboff. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. New York: Public Affairs.

[8] Tristan Harris. (2019). "Technology is Downgrading Humanity: Let's Reverse That Trend Now." Center for Humane Technology. Medium. July 17.

[9] There are several other ways that technology can increase risks to people in FCS, including exploitation of personal data, digital surveillance, irresponsible use of emerging technologies, and others. Together, these harms are often referred to as "digital risks." ICRC, Digital Harms (2021), 5.

[10] Casey Newton. (2021). "The Tier List: How Facebook Decides Which Countries Need Protection." The Verge. October 25. *https://www.theverge.com/22743753/facebook-tier-list-countries-leaked-documents-content-moderation*

[11] Lee Haleand Eyder Peralta. (2021). "Social media misinformation stokes a worsening civil war in Ethiopia." National Public Radio. October 15. *https://www.npr.org/2021/10/15/1046106922/social-media-misinformation-stokes-a-worsening-civil-war-in-ethiopia*

[12] International Alert. (2018). Human Rights Due Diligence in Conflict-Affected Settings, 15, available at https://www.international-alert.org/sites/default/files/Economy_HumanRightsDueDiligenceGuidance_EN_2018.pdf

[13] UN General Assembly, UN A/75/212 (2020). Issue of human rights and transnational corporations and other business enterprises, paras 50-51, available at *https://undocs.org/en/A/75/212* (hereinafter UN A/75/212).

[14] See World Bank. (2021). "Classification of Fragile and Conflict-Affected Situations." https://www.worldbank.org/en/topic/fragilityconflictviolence/brief/harmonized-list-of-fragile-situations

[15] UN A/75/212, para 17.

[16] UN A/75/212, para 18.

[17] UN A/75/212, paras 19-21.

[18] UN A/75/212, paras 58 – 71.

[19] Reuters, Facebook will move UK users to US terms, avoiding EU privacy laws, 15 December 2020, *https://www.theguardian.com/technology/2020/dec/15/facebook-move-uk-users-california-eu-privacy-laws*; Reuters, Exclusive: Google users in UK to lose EU data protection-sources, 19 February 2020, *https://www.reuters.com/article/us-google-privacy-eu-exclusive-idUSKBN20D2M3*.

[20] See, e.g., UK Modern Slavery Act; US FARS regs on the subject (Section 2(2)(A) of Executive Order 13627). Robert McCorquodale, Lise Smit, Stuart Neely, and Robin Brooks. (2017). Human Rights Due Diligence in Law and Practice: Good Practices and Challenges for Business Enterprises, Business and Human Rights Journal, 2, 195-224, CUP.

[21] Elsa Savourey and Stéphane Brabant. "The French Law on the Duty of Vigilance: Theoretical and Practical Challenges Since Its Adoption." Business and Human Rights Journal 6, no. 1 (2021): 141–52. doi:10.1017/bhj.2020.30.

[22] Amnesty International. (2019). Devoir de Vigilance: Les Entreprises Peuvent Mieux Faire, available at *https://www.amnesty.fr/responsabilite-des-entreprises/actualites/les-entreprises-dans-le-viseur-des-ong*; Sherpa. (2021). Creating a Public Authority to Enforce the Duty of Vigilance Law: A Step Backward? Available at *https://www.business-humanrights.org/en/latest-news/sherpa-publishes-critical-analysis-on-potential-creation-of-public-authority-to-enforce-french-duty-of-vigilance-law/*

[23] See, e.g., House Hearing, 113 Congress (2013), The Unintended Consequences of Dodd-Frank's Conflict Minerals Provision, available at *https://www.govinfo.gov/content/pkg/CHRG-113hhrg81758/html/CHRG-113hhrg81758.htm*

[24] Whereas this risk tends not to be accounted for in hastily passed legislation that can be clunky and overly broad, the report of the UN Working Group on the application of the UNGPs in FCS refer expressly to and provide guidance for companies' responsible exit in situations of last resort.

[25] See, e.g., *House* Hearing, 113 Congress.