UN Office on Genocide Prevention and the Responsibility to Protect and the UN Office of the High Commissioner for Human Rights

Call for Input

"Impact of technological advances on prevention of genocide efforts and on the risks of the perpetration of genocide"

Federica D'Alessandra, Dr. Ross Gildea

12th March 2023

Introduction

Federica D'Alessandra is Deputy Director of the Oxford Institute for Ethics, Law and Armed Conflict (ELAC) and Executive Director of its Programme on International Peace and Security (IPS), Blavatnik School of Government. Dr. Ross Gildea is Visiting Scholar at ELAC, Blavatnik School of Government, and Fulbright Scholar at the Arnold A. Saltzman Institute of War and Peace Studies, Columbia University. The authors have a professional interest in the improvement of public policy related to atrocity and genocide prevention, and have been engaged in research on the relationship between developments in technology and the risks and opportunities this poses for prevention activities. This submission draws from research and analysis carried out by the authors and other colleagues at the University of Oxford precisely to this effect.[1]

Overview

This submission highlights the role that technology can play with respect to the prevention, commission, and accountability for mass atrocities. It is divided into three parts. The first part addresses the need to carry out a systematic assessment of the impact of technology on states' responsibilities under each of the three Pillars of R2P, precisely through the incorporation of a tech lens in our understanding of standards of conduct arising for states under each pillar. The second part of the submission provides an overview of some of the practical ways in which technology can be instrumental to bolster early warnings and accountability efforts. And the third part is concerned with taking a closer look at some of the ways in which technologies affect key

[1] See: Federica D'Alessandra and Ross James Gildea. 'Technological Change and the UN Framework of Analysis for Atrocity Crimes.' Policy Brief, The Stimson Center (2022). Available here: https://www.stimson.org/2022/technological-change-and-the-un-framework-of-analysis-for-atrocity-crimes/. Also by the same authors, 'Technological Change and the Practical Tools of Mass Atrocity Prevention' *Global Responsibility to Protect* (forthcoming 2023). Also see Federica D'Alessandra and Kirsty Sutherland. 'The Promise and Challenges of New Actors and New Technologies in International Justice.' *Journal of International Criminal Justice*, Volume 19, Issue 1, March 2021, Pages 9–34, https://doi.org/10.1093/jicj/mqab034; and Federica D'Alessandra, Stephen Rapp, Kirsty Sutherland, and Sareta Ashraph. 'Anchoring Accountability for Mass Atrocities: the Permanent Support Needed to Fulfil UN Investigative Mandates' Oxford Institute for Ethics, Law and Armed Conflict, May 2022. Available here: https://www.bsg.ox.ac.uk/sites/default/files/2022-05/Anchoring%20Accountability%20for%20Mass%20Atrocities%20Report.pdf

political and societal dynamics underpinning the commission of atrocities, and how these need to be better captured in key instruments for the operationalization of R2P, such as the UN Framework of Analysis for Atrocity Crimes given rapid developments in digital and cyber technology.

The Framework is a key tool in the atrocity prevention toolkit, providing "an integrated analysis and risk assessment tool for atrocity crimes."[2] However, given major changes to the technological landscape since the Framework's publication in 2014, we contend that enacting revisions through a tech lens would ensure it is more attuned to contemporary atrocity dynamics. Our comments can be categorized under two main rubrics. In our assessment, while the Framework continues to do an admirable job in capturing core risk factors of genocide and other atrocity crimes, there is important scope for limited revisions of associated indicators to better reflect the tech-shaped dynamics which now precede and enable atrocity crimes. Given that the Framework is used not only as a monitoring and early warning tool, but is also aimed to "promote action" to address atrocity threats, we also propose the inclusion of a detection component in a future iteration of the document. By detection, we refer to advice on operationalization of key concepts and approaches to measurement. The submission concludes by proposing some additional steps, outside of the UN Framework of Analysis, to mitigate atrocity challenges posed by new technologies.

Written Submission

I. Technology and R2P: a disciplinary gap

A systematic assessment of the relationship between technology and R2P is currently lacking and urgently needed, as technological developments may require us to re-assess our understanding of standards of conduct arising for states under each pillar of the norm, of the dynamics that underpin atrocity scenarios, and of the many tools now available for the operationalization of the norm and its underpinning legal principles.[3]

The scope and nature of states' preventive responsibilities under R2P remains the topic of heated debate. What we do know about the status of the international legal framework on preventive duties however is that: "(1) preventive duties arise for States at the instant they learn or should have learned of the serious risk that atrocities would be committed; (2) their obligation is one of conduct, requiring States to use the means reasonably available to them to try to avert mass atrocities; and perhaps most importantly, (3) States are held to a due diligence standard, with the extent of each State's responsibility evaluated against its capacity" to act meaningfully to avert the situation from precipitating.[4]

---

[2] "Framework of Analysis for Atrocity Crimes: A Tool for Prevention" (United Nations, 2014), 5, https://www.un.org/en/genocideprevention/documents/about-us/Doc.3_Framework%20of%20Analysis%20for%20Atrocity%20Crimes_EN.pdf.
[3] Submission authors, 'Technological Change and the Practical Tools of Mass Atrocity Prevention' *Global Responsibility to Protect* (forthcoming 2023).
[4] Federica D'Alessandra and Shannon Raj Singh. Operationalizing Obligations to Prevent Mass Atrocities: Proposing Atrocity Impact Assessments as Due Diligence Best Practice, *Journal of Human Rights Practice*, (2022).

Such capacity-driven assessments can only be contingent on knowledge of the specific dynamics of the atrocity situation in question, which are being seriously altered by technology, both in terms of expanding potential risks and the means of preventing them. Incorporating tech insights into both our normative commitments and practical assessment tools surrounding RtoP thus appears necessary, for it will not only generate a better analytical understanding of the impact of certain technological developments on the *practical* realities of mass atrocities but will also have wider positive implications for clarifying states' material capacities and, thus, their responsibilities (including their duties under international law) arising under each Pillar of R2P.[5] The twentieth anniversary of the RtoP norm may constitute the perfect opportunity for the Joint Office to partner with scholars in the academy and with legal, policy, and tech experts to carry out a systematic assessment of the impact of technology on standards of conduct arising for states under each pillar of the norm.

II. Technology to bolster prevention and accountability

There is immense scope to better and more systematically incorporate new technologies and tech-derived methods to bolster monitoring, early warnings and accountability efforts.[6] At a minimum, certain low-cost documentary technologies — including satellite imagery, radio, radar and other forms of remote sensing capacity, such as commercial unmanned aerial vehicles — should be more systematically used alongside open-source information and other methodologies (including social network and big data analysis) for monitoring and fact-finding purposes, for they can assist with tracking the movement of individuals or groups (including militias and refugees), or even the 'mood' of specific groups[7] — sometimes being able to predict with amazing precision the outbreak and location of identity-based protests or other atrocity risk factors.[8]

The same low-cost technologies can also be instrumental for fact-finding, documentation, and investigative purposes, in combination with more traditional methods. In fact, in this space, the advantages of new technologies, digital documentary and investigative methods can be significant. Certain documentary tools - like the Eyewitness to Atrocity app, for example - can be instrumental in safely collecting, verifying, and transmitting evidence of specific incidents.[9] Satellite imagery and other forms of remote sensing capacities can also be used to verify and corroborate the construction and use of military installations and other buildings, to track the movement of convoys, and document and geolocate specific incidents, as well as their

[5] Supra 3.
[6] Federica D'Alessandra and Kirsty Sutherland. 'The Promise and Challenges of New Actors and New Technologies in International Justice.' *Journal of International Criminal Justice*, Volume 19, Issue 1, March (2021), Pages 9–34, https://doi.org/10.1093/jicj/mqab034
[7] R. Rotberg, 'Deterring Mass Atrocity Crimes: The Cause of Our Era', in R. Rotberg (ed.) *Mass Atrocity Crimes: Preventing Future Outrages* (Brookings International Press, (2010) 1–24.
[8] C. Mahony, E. Albrecht and M. Sensoy, *The Relationship Between Influential Actors' Language and Violence: A Kenyan Case Study Using Artificial Intelligence*, Commission on State Fragility, Growth and Development (2019), available online at https://www.theigc.org/wp-content/uploads/2019/02/Language-and-violence-in-Kenya_Final.pdf.
[9] https://www.eyewitness.global/

aftermath.[10] For example, satellite has been instrumental in both tracking and showing the trail of destruction left by militias and military forces in Sudan, South Sudan, the Central African Republic, the Democratic Republic of the Congo, Niger and Myanmar, and to demonstrate the rapid development of detention camps in the Chinese Autonomous Region of Xinjiang, among others.[11] In addition, because digital technologies can theoretically be used anywhere, they permit meaningful monitoring, documentation and investigation of possible atrocities remotely even for so-called 'black hole' environments, where information is deliberately hidden by local authorities or otherwise scarce. Indeed, corroborated by witness affidavits, it has also unveiled the operation and reporting structures of special political prisoners' camps (*kwanli'so*), north of Pyongyang -where 80,000–130,000 people are currently being detained- and helped establish the exact location of camps 14, 15, 16, 18, 22 and 25, the very existence of which the North Korean government has vehemently denied.[12]

While measures can be taken to 'fool' satellites, and simple 'birds eye' views may not adequately show destruction or erosion of building complexes, the prevalence, independence and consistency of satellite oversight renders it more difficult to sustain denials of wrongdoing. Open-source investigations and tech-derived evidence can indeed be instrumental to help overturn state narratives, allow greater oversight and transparency, and even significantly contribute to shaping judicial processes.[13] Such information cannot, of course, entirely substitute more traditional forms of evidence and information. However, it can act as a 'force multiplier'. For this reason, countless grounds exist to expand its potential application and use, including but not limited to judicial and non-judicial accountability, including criminal and civil proceedings,[14] other transitional justice strategies, truth and reconciliation efforts, memorialization and restorative justice processes.[15]

III. Technology and the UN Framework of Analysis

The incorporation of a tech-aware lens can also be vital to ameliorate key analytical and policy tools aimed at the operationalization of RtoP. In this submission, we focus on potential revisions to the UN Framework of Analysis for Atrocity Crimes (UNFAAC). The revisions we propose to the UNFAAC may be categorized under two rubrics. First, while we believe the existing risk factors capture core atrocity dynamics, we recommend several updates to relevant indicators to incorporate tech insights. Second, we suggest the potential inclusion of a detection component,

---

[10] Federica D'Alessandra and Kirsty Sutherland. 'The Promise and Challenges of New Actors and New Technologies in International Justice.' *Journal of International Criminal Justice*, Volume 19, Issue 1, March 2021, Pages 9–34, https://doi.org/10.1093/jicj/mqab034

[11] Ibid.

[12] International Bar Association, *Report: Inquiry on Crimes Against Humanity in North Korean Political Prisons*, December 2017, available online at https://www.ibanet.org/IBA-War-Crimes-Committee–-Inquiry-on-Crimes-Against-Humanity-in.aspx (visited 28 November 2020) at 21–26.

[13] Ibid. section 2.

[14] Federica D'Alessandra, Stephen Rapp, Kirsty Sutherland, and Sareta Ashraph. 'Anchoring Accountability for Mass Atrocities: the Permanent Support Needed to Fulfil UN Investigative Mandates' Oxford Institute for Ethics, Law and Armed Conflict, May 2022. Available here: https://www.bsg.ox.ac.uk/sites/default/files/2022-05/Anchoring%20Accountability%20for%20Mass%20Atrocities%20Report.pdf

[15] Supra 16.

whereby users are provided with guidance on the operationalization of key concepts and approaches to measurement.

1. Revising Risk Factors and Indicators

As it would be impractical to attempt an exhaustive list of updates for the UNFAAC, in this section we seek to illustrate, using three instructive examples, how and why insights about the digital tech and cyber environment might be integrated into the framework.

- One of the current limitations of the UNFAAC lies in its conceptualization of perpetrators' capacity to commit atrocity crimes (most directly in Risk Factor 5). Capacity is viewed primarily in terms of the possession and deployment of items such as arms, training, personnel, support, and financing.[16] In today's digital world, this emphasis appears outmoded. AI-powered surveillance, in allowing the state to constantly track, monitor, and target individuals and groups in the population, has radically altered relations between the state and citizenry, notably atrocity dynamics. Beyond capacity, escalation in the surveillance of vulnerable groups may, for instance, be a useful indicator of other risk factors, such as "Enabling circumstances or preparatory actions" (Risk Factor 7), "Triggering factors" (Risk Factor 8), and "Intergroup tensions or patterns of discrimination against protected groups (Risk Factor 9).
- Another possible weakness of the UNFAAC is that it does not recognize growing threats from ICTs, such as social media platforms, in promulgating disinformation and incitement. Insights on the diffusion of hateful rhetoric and incitement online could be fruitfully incorporated into the framework as indicators for several risk factors, including as a sign of "Intergroup tensions or patterns of discrimination against protected groups" (Risk Factor 9); "Signs of an intent to destroy in whole or in part a protected group" (Risk Factor 10); "Signs of a plan or policy to attack any civilian population" (Risk Factor 12); and "Serious threats to those protected under international humanitarian law" (Risk Factor 13).
- While possession of many new technologies does not pose inherent atrocity threats, raising issues of governance and regulation, certain tech poses acute challenges. For example, "Deepfakes" - audio-visual media that can portray and mimic real people and events - may be particularly dangerous in the context of the rapid spread of media across social platforms.[17] Similarly, spyware such as Pegasus, an invasive tool which can crack encrypted communications and dubbed "the world's most powerful cyberweapon", poses major risks to civil and political rights and could enable atrocities.[18] It is vital that an updated UNFAAC takes stock of the development and use of these dangerous technologies.

[16] United Nations. 'Framework of Analysis for Atrocity Crimes: A Tool for Prevention' (2014): 14.
[17] Nina I. Brown, 'Deepfakes and the Weaponization of Disinformation,' *Va. JL & Tech.* 23 (2020).
[18] Ronen Bergman and Mark Mazzetti, "The Battle for the World's Most Powerful Cyberweapon", *The New York Times*, (2022).

As these examples illustrate, the current version of the UNFAAC does not account for significant changes in the tech landscape since its publication in 2014, with major implications for atrocity dynamics. Incorporating changes to the Framework from a “tech-aware” perspective would likely improve its capacity to evaluate atrocity threats, including risks of genocide. To expand this analysis and to ensure the Framework remains fit for purpose, we propose a systematic review of the UNFAAC using a tech lens.

2. Detection Component

- As the UNFAAC is intended not only to serve as an early warning tool, but also to “promote action” to address atrocity threats, it may prove useful to include a “detection component” in a future iteration of the document. By this, we refer to guidance on the operationalization of key concepts and approaches to measurement. With the increasingly advanced technical knowledge required to evaluate the role of digital technology in atrocity risks, a detection component would not only improve the utility of the Framework but also help to widen its user-base.

3. Additional Steps

Although an updated UNFAAC would prove immensely valuable, to preclude the rise of atrocity threats additional policy steps must be taken to regulate the development and use of potentially dangerous technologies. Some preliminary measures in this direction might include:

1. The development of internationally accepted norms, informed by public-private sector partnerships, on the ethical development, procurement, sale, and use of technological tools such as AI-powered surveillance infrastructures or dangerous technologies such as Deepfakes which, if misused, could increase the risk of atrocity crimes.
2. The fostering of public-private sector partnerships to develop industry-wide standards on ICTS, such as for the prompt identification and removal of online content that could lead to offline harm in the form of identity-based mass violence.