Technological Change and the UN Framework of Analysis for Atrocity Crimes

Federica D'Alessandra[1] and Ross James Gildea[2]

University of Oxford

EXECUTIVE SUMMARY

How might developments in new and emerging technologies (NETs) shape our understanding of the risk factors associated with atrocity crimes? This paper conducts an exploratory analysis of the potential need to revise and update the UN Framework of Analysis for Atrocity Crimes (FAAC). The FAAC provides "an integrated analysis and risk assessment tool for atrocity crimes", helping to identify areas where atrocity threats may be most pressing.[3] However, despite the value of the FAAC, particularly as a tool guiding UN analysis and forecasting in the context of its approach to mass atrocity prevention and response (MAPR), rapid changes in the tech landscape since its 2014 publication, both in relation to the commission and prevention of atrocity crimes, suggests there may be good reason to revisit the framework. Such revisions would not be unprecedented. The current FAAC replaced an earlier, more limited version published in 2009, taking account of "recent developments and new research into the processes that lead to those crimes."[4] To assess the need for future revision, we conduct a systematic evaluation of the FAAC, selecting six risk factors as representative cases. We initially focus on three common risk factors, which pertain to each atrocity crime. To further generalize our findings, we also examine three specific risk factors which relate to each category of crime – genocide, crimes against humanity, and war crimes – individually. In addition, we carry out an initial probe of the potential for NETs to be leveraged for detection efforts. As the FAAC is intended to aid monitoring and early warning to anticipate atrocity threats, we reflect on whether a future edition may benefit from incorporating information of how indicators of risk factors could be measured and assessed.

It is our contention that NETs, as well as a growing body of research on their development and impact, may be shifting our understanding of MAPR once again.

Our analysis indicates that while the current iteration of the FAAC continues to be of eminent value, with many of the risk factors and indicators reflecting enduring issues related to atrocity crimes, the rapid development of NETs may have shifted the MAPR landscape sufficiently to warrant some revision. In particular, NETs have markedly expanded the tools available to states and groups to commission atrocities, altering the dynamics and processes which may lead to such crimes. One pertinent example is NETs with surveillance applications, which can enhance the ability of perpetrators to track and monitor targets, increasing their capacity to commit atrocity crimes. Tech developments can also make significant contributions to both upstream and downstream elements of the MAPR toolkit. These findings are timely, given the 2021 report of the UN Secretary-General entitled *Advancing atrocity prevention: work of the Office on Genocide Prevention and the Responsibility to Protect* re-emphasized the responsibilities of not just member states in tackling

---

[1] Deputy Director, Oxford Institute for Ethics, Law, and Armed Conflict (ELAC); Executive Director, Oxford Programme on International Peace and Security (IPS); Blavatnik School of Government, University of Oxford.

[2] Postdoctoral Research Fellow, Oxford Programme on International Peace and Security (IPS), Oxford Institute for Ethics, Law, and Armed Conflict (ELAC), Blavatnik School of Government, University of Oxford.

[3] FAAC: 5. See: https://www.un.org/en/genocideprevention/documents/about-us/Doc.3_Framework%20of%20Analysis%20for%20Atrocity%20Crimes_EN.pdf

[4] Ibid.

atrocities, but technology and social media companies.[5] Relatedly, our paper identifies considerable scope for the FAAC to be revised to include a detection component, including details on how NETs and other tools can be applied in assessing indicators.

INTRODUCTION

The proliferation of NETs in the past decade appears to have far-reaching implications for MAPR. Technology is not only shaping the dynamics of interstate relations but has elevated the role of non-state actors in global politics.[6] On one hand, technological innovation is generating new challenges, providing would-be perpetrators with new means to incite and organize mass identity-based violence. This includes the growing use of AI-powered surveillance technology, information and communication technologies (ICTs) – such as social media and other digital platforms – and automated weapons systems. In Myanmar, for example, authorities have used "surveillance drones, phone cracking, and computer hacking software, to track citizens and steal data," contributing to systematic attacks on civilians.[7] Digital platforms have also been used by the Myanmar military to incite and instigate mass identity-based violence.[8] Conversely, on-going contextual changes also broaden the set of tools available to policymakers, enhancing our ability to prevent and respond to atrocity threats.[9] Geospatial intelligence (GEOINT) technology for instance, such as satellite imagery, has expanded the available tools for investigation and documentation of crimes, being used to track the actions of perpetrators in South Sudan, the Central African Republic, the Democratic Republic of the Congo, Niger and Myanmar, among other cases.[10] Given these tech-driven changes to the MAPR landscape, it is imperative to revisit policy frameworks such as the FAAC to ensure they are fit for purpose.

The FAAC was created by the United Nations Office on Genocide Prevention and the Responsibility to Protect to inform evaluations of global atrocity risks. Published in 2014, the framework builds on a prior iteration that was developed in 2009 by the United Nations Office of the Special Adviser on the Prevention of Genocide, which in turned emerged from foundational work on genocide prevention by United Nations Secretary-General Kofi Annan in 2004.[11] Its purpose is to serve as a "guide for assessing the risk of genocide, crimes against humanity and war crimes".[12] It is regularly (although not yet systematically)[13] used by various UN entities and agencies – as well as researchers and practitioners in the field of atrocity prevention – as part of the early warning and analysis tools currently available to carry out basic forecasting and risk assessments of both current atrocity threats and of how situations may deteriorate over time.[14] It is comprised of 14 risk factors – subdivided into eight common and six specific risk factors – and associated indicators. To evaluate the FAAC, we carry out a systematic review of the document, selecting six risk factors as representative cases. We

---

[5] See here: https://www.un.org/en/genocideprevention/key-documents.shtml

[6] Nye, Joseph S., Jr. 2014. The Information Revolution and Power, *Current History* 113(759): 19–22; see also: Drezner, Daniel W. 2019. Technological change and international relations. *International Relations*, 33(2): 286-303.

[7] https://slate.com/technology/2022/03/dual-use-surveillance-technology-export-controls.html

[8] https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html

[9] D'Alessandra, Federica, and Sutherland, Kirsty. 2021. The Promise and Challenges of New Actors and New Technologies in International Justice, *Journal of International Criminal Justice*, 19(1): 9–34.

[10] Ibid.

[11] Dieng, Adama, and Jennifer Welsh. 2016. "Assessing the Risk of Atrocity Crimes." *Genocide Studies and Prevention: An International Journal* 9:3, 4.

[12] FAAC: iii.

[13] For example, the recent reports of the UN Human Rights Council appointed Commission of Inquiry on Burundi and Fact-Finding Mission on Myanmar have both utilized the FAAC to identify risk factors and potential triggers for war crimes, crimes against humanity, ethnic cleansing or genocide. See: A/HRC/48/68; A/HRC/39/64. Yet, more could be done to systematize its use, particularly across UN field missions. See, for example: D'Souza. 2020. A Re-oriented Approach to Atrocity Prevention in UN Peace Operations, Oxford Programme on International Peace and Security Working Paper Series, https://www.elac.ox.ac.uk/a-reoriented-approach-to-atrocity-prevention-in-un-peace-operations/

[14] For further details on its applications, see: Dieng Welsh, 2016.

focus on three common risk factors, pertaining equally to each crime. To further generalize the implications of our findings, we also examine three specific risk factors which relate to each category of crime – genocide, crimes against humanity, and war crimes – individually (see Table 1). To enhance representativeness, risk factors were chosen to investigate broad and varying aspects of atrocity threats cited in the FAAC. For instance, Risk Factor 4 (Motives and incentives) pertains to the decision-making process of perpetrators, while Risk Factor 5 (Capacity to commit atrocity crimes) speaks to the material capabilities that underpin atrocity crimes.

Table 1. Risk Factors Selected for Analysis

Common Risk Factors	
Risk Factor 4	Motives and incentives
Risk Factor 5	Capacity to commit atrocity crimes
Risk Factor 7	Enabling circumstances or preparatory action
Specific Risk Factors	
Risk Factor 10	Signs of an intent to destroy in whole or in part a protected group (Genocide)
Risk Factor 12	Signs of a plan or policy to attack any civilian population (Crimes against humanity)
Risk Factor 13	Serious threats to those protected under international humanitarian law (War crimes)

In addition, we carry out an initial probe of the potential for NETs to be leveraged for detection efforts. As the FAAC is intended to aid monitoring and early warning to anticipate atrocity threats, we reflect on whether a future, revised edition may benefit from a systematic treatment of how indicators could be measured and assessed. To this end, in each section we briefly comment on the potential utility of applying a tech lens to a detection component of the FAAC.

Our analysis indicates that while the current iteration of the FAAC continues to be of indisputable value, with many of the risk factors and indicators reflecting enduring issues related to atrocity crimes, the rapid development of NETs may have shifted the MAPR landscape sufficiently to warrant some revision. In particular, NETs have markedly expanded the tools available to states and groups to commission atrocities, altering the dynamics and processes which may lead to such crimes. Tech developments can also make significant contributions to both upstream and downstream elements of the MAPR toolkit, ranging from early warning systems for prospective crimes to mitigation and response given the onset of atrocities. These findings are timely, given the 2021 report of the UN Secretary-General entitled *Advancing atrocity prevention: work of the Office on Genocide Prevention and the Responsibility to Protect* re-emphasized the responsibilities of not just member states in tackling atrocities, but technology and social media companies.[15] Relatedly, our paper identifies considerable scope for the FAAC to be revised to include a detection component, including details on how NETs and other tools can be applied in assessing indicators. Notably, these conclusions apply across both the common and specific risk factors, as our analysis below shows.

Common Risk Factors

Motives and incentives (Risk Factor 4)

This risk factor refers to the "reasons, aims or drivers that justify the use of violence against protected groups, populations or individuals, including by actors outside of State borders."[16] We may therefore

[15] https://www.un.org/en/genocideprevention/key-documents.shtml
[16] FAAC: 13.

ask, in what ways might developments in technology affect the motives and incentives leading to atrocity crimes? While it is likely that the core logics of atrocity crimes remain unaffected by NETs – the foundational role of factors such as ideology, as well as perceptions of grievances and interests, remain central to these crimes – NETs may exacerbate or create new forms of inequalities that re-shape actor motivations.

To illustrate this point consider, for example, one indicator of Risk Factor 4, which pertains to unequal distribution and control of economic resources that connects to differences in identity groups. Incentives to retain or exploit the status quo as a dominant group, or to reverse existing grievances, may heighten the probability of mass atrocities. Economic deprivation has been directly linked to participation in armed violence.[17] Previous work has connected participation by Hutu in the 1994 Rwandan genocide, for instance, to socio-economic status.[18] As technology plays an increasingly important role in society, the "digital divide" – that is varying access to key digital technologies like the internet – and differentiated patterns of use, could dramatically impact existing inequalities. Prior studies have suggested a nexus between technology, poverty, and widening of wealth gaps.[19] Modern digital technology has also been shown to condition the ability of groups to benefit from expanded education, job opportunities, healthcare and other critical social outcomes.[20] Additionally, such technologies can be deployed to organize and mobilize sectors of society, thus helping to consolidate political power and potentially exclude and disenfranchise weaker groups. Scholars of genocide have long posited a relationship between technological and bureaucratic architectures and the mentality of intent they can produce in those who wield them, as with studies of perpetrators during the Holocaust.[21] As such, while the core motives and incentives outlined in the FAAC remain very much applicable to today, particular associated indicators, such as those related to political and economic cleavages, may benefit from incorporating a technological dimension.

Extending from the above, and regarding detection, NETs could prove especially useful in identifying the drivers of certain political and economic sources of inter-group animosity and – ultimately – mass violence. For instance, tracking differentiated patterns of access to, and use of, key technologies with empirical links to salient socio-economic and political outcomes (access to public services, jobs, political office, and so on) could allow us to anticipate growing inequalities and thus improve our predictive capacities. Of course, scholars should remain cautious in prediction so as to avoid conflating deeply contextual challenges with correlations in aggregate data – for ethical reasons, in diffusing their findings researchers should remain cognizant of the potential impact of identifying false positives or in exacerbating inter-group tensions.

Capacity to Commit Atrocity Crimes (Risk Factor 5)

This risk factor pertains to "the ability of relevant actors to commit atrocity crimes."[22] As the FAAC suggests, atrocities are typically difficult to commit, requiring a certain level of personnel and resources. At present, the FAAC largely – and appropriately – focuses on material capacities such as personnel, arms, training, and financing. Insightfully, it highlights important yet less tangible factors such as cultures of obedience and group conformity. While these conceptualizations are valuable in their current guise, adopting insights related to technology may enhance our understanding of both.

---

[17] Justino Patricia. 2009. Poverty and Violent Conflict: A Micro-Level Perspective on the Causes and Duration of Warfare. *Journal of Peace Research*. 46(3): 315-333.

[18] Friedman, Willa. 2011. Local economic conditions and participation in the Rwandan genocide. In APSA 2011 Annual Meeting Paper.

[19] Mirza, M. Usman, Andries Richter, Egbert H. van Nes, and Marten Scheffer. 2019. "Technology driven inequality leads to poverty and resource depletion." *Ecological Economics* 160: 215-226.

[20] DiMaggio, Paul, Eszter Hargittai, Coral Celeste, and Steven Shafer. 2004 Digital inequality: From unequal access to differentiated use. *Social inequality*, 2004: 355-400.

[21] Akio Kimura. 2003. Genocide and the modern mind: intention and structure, *Journal of Genocide Research*, 5:3, 405-420.

[22] FAAC: 14.

Indicator 5.3 – which relates to perpetrators' ability to recruit and mobilize supporters – provides an instructive entry point for the utility of applying a tech lens. Previous studies have found that the diffusion of NETs, such as information and communication technology, can lead to increases in organized violence. A reputed mechanism driving this connection is that technology can radically shift collective action problems by lowering barriers of communication and organization. NETs may therefore improve perpetrators' capacity to recruit, mobilize and co-ordinate supporters.[23]

The relationship between NETs and the capacity to commit atrocity crimes is also apparent with surveillance technology. By surveillance, we refer to the collection and use of personal data to observe, influence, or manage those from whom data has been collected.[24] In China, a significant tech apparatus has been developed to gather information and track the domestic population being directly implicated in the repression of Uyghurs and other minorities.[25] Similar challenges have been identified in Myanmar, where access to surveillance technology has facilitated the targeting of victims by the military.[26] NETs with surveillance applications can enhance the ability of perpetrators to track and monitor targets, increasing their capacity for atrocity crimes. To date, this dimension of state capacity is largely absent from the FAAC, with Risk Factor 5 containing no reference to capacities to track and target victims.

NETs are also directly relevant to cultures of obedience and group conformity. Use of social media platforms, for instance, can "limit the exposure [of users] to diverse perspectives and favor the formation of groups of like-minded users framing and reinforcing a shared narrative."[27] Digital platforms may not only lead to polarized exchanges and intergroup distrust, but – especially in societies with identity-based divisions – serve as a tool to incite hatred and violence, as was the case with the targeting of Rohingya in Myanmar.[28] Incorporating a tech lens to the FAAC may therefore enhance our understanding of the dynamics producing cultures of obedience and group conformity, and how this may be conducive to committing of atrocities.

In terms of detection, identifying trends of investment in tech infrastructure, particular those with surveillance applications, could prove to be a useful signifier of future threats. Using tech to monitor the use of digital platforms, in particular the use of dehumanizing speech – a potential predictor of atrocities – or incitement by government officials and non-state actors, may also prove fruitful in pre-empting the orchestration of atrocity crimes.

Enabling circumstances or preparatory action (Risk Factor 7)

Risk Factor 7 relates to "events or measures, whether gradual or sudden, which provide an environment conducive to the commission of atrocity crimes, or which suggest a trajectory towards their perpetration."[29] The FAAC provides a substantial list of 14 indicators for this risk factor, including measures related to domestic laws, procurement of arms and other capacities for state violence, growing rights violations, and the targeting of marginalized groups.

Although the FAAC currently provides a comprehensive account of the environmental conditions that may enable atrocities, applying a tech lens may helpfully augment and add further specificity to this risk factor. Consider, for instance, Indicator 7.3, which refers to "strengthening of the security

[23] For example, see: Pierskalla, Jan H., and Florian M. Hollenbach. "Technology and collective action: The effect of cell phone coverage on political violence in Africa." *American Political Science Review* 107, no. 2 (2013): 207-224.
[24] Lyon, David. 2001. *Surveillance Society: Monitoring Everyday Life*. (Buckingham: Open University Press), 2.
[25] https://www.hrw.org/news/2021/11/24/mass-surveillance-fuels-oppression-uyghurs-and-palestinians
[26] https://www.lighthousereports.nl/investigation/eu-spy-tech-serves-myanmar-junta/
[27] Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. "The echo chamber effect on social media." Proceedings of the National Academy of Sciences 118: 9.
[28] https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html
[29] FAAC: 17.

apparatus" in states.[30] The global proliferation of artificial intelligence (AI) surveillance technology illustrates the benefits and limitations of this perspective. AI tech represents an increasingly important part of states' security infrastructure, and it would be erroneous to conflate its increasing and lawful use with atrocity threats. While a processional focus on the *strengthening* of security apparatuses is helpful, we might pay closer attention to the presence or absence of institutional safeguards before and after tech systems are in place. Experts suggests that algorithmic biases in AI-driven tech can compound and amplify societal inequalities, including identify-based marginalization.[31] Without adequate checks and balances, this tech also provides would-be perpetrators with powerful and direct means of targeting protected groups and individuals. Given the rapid spread of AI-driven surveillance, this is likely to be a growing challenge that requires additional focus on the norms and institutional frameworks surrounding the development and application of NETs.

A related challenge with AI technology, with direct relevance to "preparatory action" under this risk factor, is the emergence of "deepfakes".[32] The FAAC cites indicators such as increases in the politicization of past events, use of propaganda and inflammatory language.[33] Deepfakes add another dimension to these risks, allowing creators to generate sophisticated audio and visuals of individuals – such as political leaders – and events that appear realistic. The inauthenticity of this content can be very difficult for the public to detect. Deepfakes could include extremely dangerous scenarios such as fabricated declarations of war by leaders or incidents of identity-based violence.[34] Deepfakes therefore have serious potential to cause societal unrest and rationalize mass violence, helping perpetrators to quickly mobilize support and resources for the commission of atrocities. This feeds into a broader challenge of tech and misinformation, both within states and from third parties.[35] And while misinformation is hardly new, in combination with digital platforms and messaging applications, deepfakes and related kinds of misinformation can now be generated and spread with unprecedented speed. In the coming years, this form of NET may therefore prove to be a crucial form of preparatory action employed by perpetrators, one which policymakers must carefully examine and respond to.

Regarding detection, technology can be readily leveraged to evaluate indicators associated with Risk Factor 7. For instance, tech can be used to examine very different kinds of enabling factors, from mapping the import of arms and ammunitions, to the development of AI surveillance, and tracking trends in inflammatory rhetoric among political leaders and online. With respect to deepfakes, researchers are creating software tools which can be applied to images and video and detect manipulations.[36] On account of the rising sophisticated of deepfakes, it is essential that these tools be developed and leveraged to combat misinformation. The ongoing campaign of disinformation by the Russian government, to generate consent among its citizens for its actions in Ukraine, is one salient case where disinformation may be helping to facilitate serious crimes.[37] Tracking trends in the use of political deepfakes could serve as an important aid in identifying future atrocity risks.

---

[30] Ibid.

[31] Panch, Trishan, Heather Mattie, and Rifat Atun. 2019. "Artificial intelligence and algorithmic bias: implications for health systems." *Journal of global health* 9:2.

[32] This refers to "content, generated by an artificial intelligence, that is authentic in the eyes of a human being." See: Mirsky, Yisroel, and Wenke Lee. 2021. "The creation and detection of deepfakes: A survey." ACM Computing Surveys (CSUR) 54:2. 1-41. See also: https://lab.witness.org/backgrounder-deepfakes-in-2021/; https://www.aspi.org.au/report/weaponised-deep-fakes.

[33] FAAC: 17.

[34] Chesney, Bobby, and Danielle Citron. 2019. "Deep fakes: A looming challenge for privacy, democracy, and national security." Calif. L. Rev. 107. 1753.

[35] https://www.brookings.edu/wp-content/uploads/2020/06/The-role-of-technology-in-online-misinformation.pdf

[36] https://www.bbc.co.uk/news/technology-53984114

[37] https://odi.org/en/insights/a-war-on-many-fronts-disinformation-around-the-russia-ukraine-war/. In another example, hate speech and disinformation circulating online were found to have played a role in the deadly communal unrest (and anti-Muslim violence) that shook Sri Lanka in 2018: https://www.aljazeera.com/news/2020/5/13/sri-lanka-facebook-apologises-for-role-in-2018-anti-muslim-riots. Also see additional examples analyzed in GAAMAC, Preventing Hate Speech, Incitement, and Discrimination: Lessons on Promoting Tolerance and Respect for Diversity in the Asia Pacific. 2021.

Specific Risk Factors

Signs of an intent to destroy in whole or in part a protected group (Risk Factor 10)

This risk factor describes an intent to "to destroy all or part of a protected group based on its national, ethnical, racial or religious identity, or the perception of this identity."[38] In contrast to the general factors above, Risk Factor 10 specifically relates to the crime of genocide. Again, while the FAAC does an excellent job in mapping potential indicators of relevant intent, there may be areas where it could be fruitfully revised. Indicator 10.1 is an instructive example. This indicator cites "official documents, political manifestos, media records, or any other documentation" from which genocidal intent might be revealed or inferred.[39] However, this focus on official documentation and media records may ignore certain realities of political action and information diffusion in the digital age. In the recent Myanmar case, political leaders and military officials used online platforms to wage a "systematic campaign" to incite hatred and justify violence against the Rohingya minority.[40] Much of this campaign was propagated through unofficial accounts and channels. In countries such as Myanmar, digital platforms can function much like the internet as a whole and is an important source of information.[41] As such, focusing on traditional political sources, such as party documentation or official statements, may obscure how technology is being used by perpetrators to orchestrate mass atrocities such as genocide.

An additional challenge pertains to the platforms themselves through which perpetrators are demonstrating their intent to dehumanize and attack vulnerable identity groups. Facebook's failures in Myanmar are perhaps the most notorious example illustrating this. Investigations have revealed how the platform's failures around content moderation for users supporting armed violence against the Rohingya had a significant effect in bolstering the military's campaign. This not only facilitated the Tadmadaw's propaganda strategy, but also allowed them to drum up public support (and recruitment) for their activities. Even worse, Facebook's own algorithm was found to have promoted pages and posts being used by the military to incite violence, amplifying their messaging.[42] It would therefore be desirable to update the FAAC to reflect the new mediums – both official and unofficial – through which perpetrators may exhibit their intent to engage in identity-based violence. Moreover, the framework could be leveraged to inform (and even draw from) digital platforms's own indicators to assess and forecast crisis scenario. This would, of course, require partnerships between the private sector, the UN, and experts in the civil society, but such an exercise could lead to the gradual development of industry-wide standards for the identification (and removal) of content indicating an environment conducive to genocidal (or other atrocity) violence.

Thankfully, these issues are being increasingly recognized by the international community (and even the private sector). At a meeting of the UN Security Council in May 2022, for example, the UN Under-Secretary-General for Political and Peacebuilding Affairs, Rosemary DiCarlo, noted estimates that "malicious use of digital technologies for political or military ends has nearly quadrupled since 2015", urging states to take action in this area.[43] Similarly, learning from the failures in the context of Myanmar, industry actors (such as Twitter and Facebook itself) have themselves begun to revise their terms, policies, and practices to better and more promptly be able to detect and remove online content that could lead to serious identify-based offline harm.[44]

---

[38] FAAC: 19.

[39] Ibid.

[40] https://www.nytimes.com/2018/10/15/technology/myanmar-facebook-genocide.html

[41] Ibid.

[42] https://time.com/6075539/facebook-myanmar-military/

[43] See: https://media.un.org/en/asset/k1r/k1ryrk0n4o. Equally, pursuant to UN General Assembly resolution 75/240, two substantive sessions have already convened for the Open-Ended Working Group (OEWG) on security of and in the use of information and communications technologies.

Relatedly, to illustrate the applicability of a tech lens in respect of detection, we may refer to Indicator 10.7, which concerns "expressions of public euphoria at having control over a protected group and its existence."[45] As noted, for political leaders, military officials, and the public, digital platforms are increasingly the avenue through which socio-political communication is conducted. Monitoring this kind of speech produces technical challenges, and researchers have developed a variety of tools to effectively track and analyze this kind of content.[46] NETs may therefore prove to be a critical resource in measuring key indicators related to crimes such as genocide (among others).

Signs of a plan or policy to attack any civilian population (Risk Factor 12)

This risk factor concerns "facts or evidence suggestive of a State or organizational policy, even if not explicitly stipulated or formally adopted, to commit serious acts of violence directed against any civilian population."[47] As with previous risk factors, the FAAC provides an extensive list of indicators, several of which remain applicable without revision to take account of tech developments. For instance, Indicator 12.9, which refers to "widespread or systematic violence against civilian populations or protected groups," and Indicator 12.10, which alludes to "involvement of State institutions or high-level political or military authorities in violent acts," continue to be pertinent independent of NETs and do not require adjustment.[48]

This said, there are areas which may from the application of a tech lens. For example, Indicator 12.2, which relates to "adoption of discriminatory security procedures against different groups of the civilian population", could be updated given recent developments in surveillance technology.[49] With technologies such as facial recognition – now used widely in airport screening and law enforcement, but also in areas of social policy such as health and housing decisions – a major challenge is inherent bias against minority groups.[50] As such, tech which draws on AI systems and machine learning algorithms can have a significant impact on governance and the social fabric of societies, potentially widening pre-existing cleavages. Worryingly, such technology commonly fails in areas – such as gender and race[51] – with direct relevance to identity-based violence. This example alone highlights the potential for two tech-based updates to Indicator 12.2 (i) Processes of discrimination may arise in areas beyond security policy, such as health and housing, which may directly feed into wider societal inequalities and lead to the securitization of vulnerable minorities, and (ii) the adoption of discriminatory procedures may in fact be unintentional, hard to detect, yet nonetheless consequential in fostering an environment that may enable atrocity crimes. Moreover, this kind of updating may apply to other indicators. For instance, Indicator 12.5, which is termed as "preparation and use of significant public or private resources, whether military or other kinds", is currently somewhat vague in its relation to crimes against humanity. Added specificity on this indicator could incorporate the direction of NET in the commission of atrocities. For instance, reference to growing investment in, and operation of, AI-powered surveillance infrastructure, or access to phone cracking or computer hacking software through spyware such as Pegasus or similar – tech with demonstratable negative applications – would be a useful addition.[52] Related to this, Indicator 12.8 concerns the facilitations or

[44] On Twitter, see for example, https://www.nytimes.com/2022/05/19/business/twitter-content-moderation.amp.html; on Facebook, see https://www.thenationalnews.com/world/europe/facebook-announces-changes-to-content-moderation-1.1173392%3FoutputType%3Damp%26d%3D233; however, reports show that a lot more needs to be done to curb the spread of online hate speech on such platforms. See, for example https://www.google.com/amp/s/mg.co.za/africa/2021-11-22-facebook-fails-to-curb-the-spread-of-hate-speech-in-ethiopia/%3Famp

[45] FAAC: 19.

[46] See, for example: Laaksonen, Salla-Maaria, Jesse Haapoja, Teemu Kinnunen, Matti Nelimarkka, and Reeta Pöyhtäri. 2020. "The datafication of hate: expectations and challenges in automated hate speech monitoring." *Frontiers in big Data*: 3.

[47] FAAC: 21.

[48] Ibid.

[49] Ibid.

[50] See: https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/

[51] https://www.nature.com/articles/d41586-020-03186-4

[52] https://www.theguardian.com/news/2021/jul/18/what-is-pegasus-spyware-and-how-does-it-hack-phones

inciting of violence against civilians. Again, looking to norms and regulations around NETs, in particular information and communication technologies, could add important detail to this risk factor.

When it comes to detection and early warning signs, the issue of bias in AI-driven systems is an insightful case. Researchers are currently developing frameworks to identify the causes of biases in the development and implementation phases of tech such as facial recognition, as well as best practices to limit their negative consequences.[53] Incorporating these insights could help to establish novel indicators and inform revisions of the FAAC regarding detection, therein aiding our ability to pre-empt tech-driven moves toward crimes against humanity (among other crimes).

Serious threats to those protected under international humanitarian law (Risk Factor 13)

Risk Factor 13 pertains to "conflict-related conduct that seriously threatens the life and physical integrity of those protected under international humanitarian law."[54] Under this risk factor – which relates solely to conduct during war – the FAAC includes a broad list of 18 indicators. Many of these indicators speak to fundamental aspects of war crimes, such as evidence of orders for violent attacks that would leave no survivors, or that would otherwise contravene rights under International Humanitarian Law.[55] These indicators seem of enduring relevance, independent of the evolution of NETs.

How might a tech lens be applied to this risk factor? One area that a revised FAAC could engage with is the spread of lethal autonomous weapons systems (LAWS, also known as 'killer robots) in warfare.[56] Although LAWS aren't per se unlawful under humanitarian law, analysts have expressed doubt that this form of weaponry could ever "replicate human judgment and comply with the legal requirement to distinguish civilian from military targets."[57] LAWS also give rise to serious challenges for criminal accountability, given that they can "select and attack targets without direct human input."[58] At present, under Indicator 13.10, the FAAC cites use of "weapons, projectiles, materials or substances which are by their nature indiscriminate or cause superfluous injury or unnecessary suffering to people" as a sign of prospective war crimes.[59] The use of LAWS in war may add another layer to this risk factor, creating ambiguity as to when – and for who – criminal accountability will be forthcoming. Experts therefore speak of an "accountability gap", pointing to "hurdles to holding anyone responsible for the actions of this type of weapon."[60] For these reasons, there has been a growing campaign to ban them.[61] These challenges are far from resolved and new ones might still emerge as weapons technology continues to evolve. There thus appears to be some scope to explore within the context of the FAAC the distinct issues presented by LAWS (among other NETs), and therein better capture current challenges to anticipating and addressing war crimes.

Beyond this, technology could also (and perhaps more promisingly) provide a valuable detection function across a number of indicators in Risk Factor 13. For instance, software for examining information and communication tech can be effectively leveraged to examine patterns of radicalization and extremism among opposing parties in a conflict (Indicator 13.3), promotion of particular ethnicities and religions as signs of allegiance (Indicator 13.4), or dehumanization of groups within the population (Indicator 13.5). Evolving technology in the atrocity space, such as the use of

[53] https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

[54] FAAC: 22.

[55] Ibid.

[56] https://www.vox.com/2019/6/21/18691459/killer-robots-lethal-autonomous-weapons-ai-war

[57] https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots

[58] Walsh, James Igoe. 2015. "Political accountability and autonomous weapons." *Research & Politics* 2:4, 1.

[59] FAAC: 22.

[60] https://www.hrw.org/report/2015/04/09/mind-gap/lack-accountability-killer-robots

[61] https://foreignpolicy.com/2022/05/11/killer-robots-lethal-autonomous-weapons-systems-ukraine-libya-regulation/; https://www.nytimes.com/2021/12/17/world/robot-drone-ban.html

video and satellite imagery, could prove valuable – as it has in cases like Myanmar[62] – in tracking the movement of troops and logistical support (Indicator 13.9) and mobilization of weapons (Indicator 13.10). In conclusion, these examples suggest that emphasizing detection of early warning signs and indicators, as well as applying a tech lens to aid detection efforts, could complement and enhance the existing FAAC.

Conclusion

In this paper we carry out an exploratory analysis of the potential need to revise and update the UN Framework of Analysis for Atrocity Crimes in light of rapid developments in NETS since the document's publication in 2014. Our findings can be categorized under two rubrics. First, regarding the risk factors and associated indicators contained in the FAAC, it is clear that the framework remains an extremely pertinent tool for assessing risks of atrocity crimes such as genocide, crimes against humanity, and war crimes. Indeed, we find that the FAAC continues to do an admirable job in identifying core characteristics of atrocity threats. To take an illustrative example, Risk Factor 4, which pertains to "motives and incentives", astutely describes the role of ideology, as well as perceptions of grievances and interests, which seem foundational to atrocity threats, independent of the development of NETs. Likewise, indicators highlighted in Risk Factors 12 and 13, which pertain to evidence of orders or direct attacks on civilian populations, are also of enduring relevance.

While recognizing the continued value of the FAAC, we also identify several areas which may be revised and enhanced through application of a tech lens. For example, under Risk Factor 10 the FAAC cites "official documents, political manifestos, media records, or any other documentation" as key sources to identify intent to commit genocide. However, in a digital age – where perpetrators increasingly use online and informal channels to orchestrate mass violence, as we have seen in Myanmar and elsewhere – a focus on traditional documentation of this kind appears outmoded. In addition, and more generally, our paper suggests the advent of a "digital divide", AI-driven surveillance such as facial recognition technology, use of deepfakes, and autonomous weapons systems, among other tech-related phenomena, may be exacting profound changes on MAPR. NETs may not only exacerbate pre-existing social cleavages between groups, but – as with AWS – create new layers of complexity in assessing risks of atrocities and holding individuals accountable for crimes. It seems apparent that NETs, based on growing body of research on their nature and impact, are altering the dynamics of MAPR, a process likely to continue in coming years. As we approach a decade since the FAAC was published, we conclude that it may be time to incorporate insights on NETs on a revised version.

Second, we also carried out a preliminary analysis of the potential to augment the FAAC to include a detection component. It is our contention that, given the FAAC is intended to aid monitoring and early warning to anticipate atrocity threats, the lack of information on detection – such as methods of evaluating indicators – may be an oversight to be addressed in a revised edition. This is the case both in general and with specific allusion to NETs. For instance, our paper identifies the use of deepfake technology as a preparatory action which may be increasingly used by perpetrators to mobilize support and orchestrate violence in coming years. In an updated FAAC, it seems logical not only to include information on deepfakes as an indicator, but also to provide information on tools by which audio-visual manipulations may be identified. In a similar vein, on account of their ability to re-shape the dynamics leading to atrocity crimes, offering guidance on how to assess the use of AI-driven surveillance, AWS, or checks and balances in information and communication technologies, would be helpful. These steps would no doubt enhance the utility of the FAAC for practitioners involved in MAPR and should be considered in a future iteration of the document.

In light of this, we recommend:

[62] https://www.amnesty.org/en/latest/news/2017/09/myanmar-video-and-satellite-evidence-shows-new-fires-still-torching-rohingya-villages/

1. That common risk factors 4, 5, and 7 of the FAAC be revised to better reflect the impact and potential effects of technology tools (and their distribution) on perpetrator's motives, incentives, capacity, preparatory actions and enabling circumstances that might lead to the commission of atrocity crimes;
2. That specific risk factors 10, 12, and 13 of the FAAC be revised to better reflect the impact and potential effects of existing and emerging technology tools (including systems of surveillance and specific weaponry) on the risk of commission of, respectively, genocide, crimes against humanity, and war crimes;
3. That a broader study be carried out, building on this scoping exercise, to systematically review how each and every risk factor and indicator under the FAAC might be impacted by the application of a technology lens, and – similarly to our analysis above - whether revisions might necessary to any additional risk factors or related indicators, whether common or specific to any sub-set of atrocity crimes;
4. That the overall FAAC be expanded to incorporate a technology-driven detection component, to improve existing early warning systems and forecasting;
5. Furthermore, we recommend that public-private sector partnerships be fostered to develop industry-wide standards for the prompt identification and removal of online content that could lead to offline harm in the form of identity-based, mass violence;
6. Similarly, that internationally accepted norms be developed, informed by public-private sector partnerships, on the ethical development, procurement, sale and use of technological tools (such as AI-powered surveillance infrastructures, or certain weaponry systems such as LAWS) that, if misused, could lead or contribute to a heightened risk of commission of atrocity crimes.