IEAI Technical University of Munich
School of Social Science and Technology
Institute for Ethics in Artificial Intelligence

Research Brief – Q1 2024

AI For Human Security – Applications and Ethical Considerations

By Caitlin C. Corrigan

Traditional conceptions of security and the prevention of intrastate conflict have in many ways become less relevant to individual daily lives compared to threats from extreme weather, civil conflict, violent crime or rights abuses. Complex interconnections and porous borders mean that a threat to the security of those in a neighboring region is also a threat to those elsewhere. The prominence of these vulnerabilities calls for a rethinking of how we define a secure environment and has led to a discourse on human security. At the same time, humanitarian and government organizations are working with very scarce resources, multiple simultaneous crises and ever-increasing amounts of information to sift through. Given these conditions, AI-enabled tools, and the way they can handle and interpret vast amounts of diverse information quickly, have become of increasing interest to responsible parties focused on improving human security. How can AI help with improving human security around the world? Where might it exacerbate problems? And how can ethical frameworks help us to maximize positive impacts? These questions are explored in this research brief.

By the end of 2022, 108.4 million people worldwide had been displaced due to violence, instability and human rights violence (UNHCR, 2022). The physical and mental insecurity that accompanies displacement is unthinkable and is exacerbated by factors such as poverty, climate change and political tensions. Extreme weather is becoming increasingly common worldwide (IPCC, 2023), meaning finding resources to prepare for and react to these events has become a priority. Extreme wealth inequalities (Chancel et al., 2021) leave many without access to basic needs, a violation of personal human rights and is a source of instability everywhere.

At the same time, the world has become smaller in many ways, with information from all over the globe flowing into our homes, devices and public discussions at a rate previously unseen. The challenges humanity faces have also become increasingly intertwined, complex and dire. The combination of these facts means that daily threats to individuals worldwide are highly apparent to those beyond the borders of a given crisis. On the other hand, we are often confronted with too much information about too many crises to be able to react in an effective or logical way.

The prominence of these vulnerabilities calls for a rethinking of how we define a secure environment. Traditional conceptions of security (i.e. based around borders or military strength) and the prevention of intrastate conflict have in many ways become less relevant to individual daily lives compared to threats from extreme weather, civil conflict, violent crime or rights abuses. Complex interconnections and porous borders mean that a threat to the security of those in a neighboring region is also a threat to those elsewhere.

It becomes clear from these observations that traditional concepts of security are not necessarily the most, or at least not only, relevant concern for individuals. From this observation, a discourse on the concept of *human* security emerged. The diversity of threats we face daily, particularly those in developing or conflict-prone countries, is great and complex. Thus, we require innovative approaches for new ways to manage and promote human development and individual security.

At the same time, humanitarian and government organizations are working with very scarce resources in terms of time, manpower and finances. They are confronted with multiple simultaneous crises, such as violent conflicts or natural disasters. Moreover, given the way technology is being used for population tracking, communication and supply management, agencies are being confronted with an ever-increasing amount of information to sift through. Given these conditions, AI-enabled tools, and the way they can handle and interpret vast amounts of diverse information quickly, have become of increasing interest to those actors focused on improving human security (Beduschi, 2022; European Parliament, 2019).

AI-enabled tools, and the way they can handle and interpret vast amounts of diverse information quickly, have become of increasing interest to those actors focused on improving human security.

How can AI help with improving human security around the world? Where might it exacerbate problems? How can ethical frameworks allow us to maximize positive impacts? These questions are explored in this research brief. The following section outlines the concept of human security and justifies the focus of the brief. This is followed by an overview of where AI has the potential to play a role in improving human security with a focus on (1) the protection of civilians in war, (2) dealing with humanitarian disasters, (3) promoting human rights and (4) preventing and documenting mass atrocities. The ethical implications and potential negative impacts that need to be assessed are then discussed, and a way forward is suggested.

What is Human Security

Originating in the Cold War era, the concept of *human security* was brought to the forefront of discourse with the 1994 UN Human Development Report (UNDP, 1994), attributing “security with people rather than territories, with development rather than arms”. The concept continued to develop and be debated from this point on, but generally focused on a different approach to thinking about security. Instead of the traditional focus on protecting national borders and interstate power balances, human security focuses on threats to the individual.

With the center of action and analysis placed on increasing security at the individual level, the prominence of different security threats emerges beyond traditional interstate warfare to be combated with arms, to include threats such as violent crime, forced migration and internal displacement, human rights violations, crimes against humanity, extreme poverty, the spread of disease or natural disasters. Moreover, these threats, or protection from threats, do not end at any specific border. This is particularly relevant given the transnational nature of major threats today and the transformative impact of the subject of this brief: AI.

The concept of human security is often divided into two aspects (Hanlon and Christie, 2016):

- Freedom from want
- Freedom from fear

Derived from President Franklin D. Roosevelt's "Four Freedoms" Speech[1] and reinforced by the UN in 1945 (UNDP, 1994), this categorization reflects a perception of positive and negative definitions of peace as well (Galtung & Fischer, 2013). A negative definition of peace focuses on the absence of violence (direct, cultural and structural) or, from an individual perspective, much of what encompasses the human security concept of "freedom from fear". This narrower approach to human security focuses on protection from violence and fundamental rights abuses to individuals.

Broadening the focus, we can identify positive definitions of peace: a state that enables development through cooperation, equality, equity and dialog. This goes hand in hand with the human security concept of "freedom from want". Under the Universal Declaration of Human Rights, this is termed "the right to an adequate standard of living" (United Nations, 1948). It moves beyond physical protection as the benchmark of human security to include aspects closely related to our current sustainable development agenda, arguing that these are necessary prerequisites for anyone to be "secure". To be certain, the concepts of freedom from fear and want are intricately connected, as they both build off of and enable each other. Similarly, Amertya Sen argues for the need to see the interconnections between personal freedoms such as political, economic, social and protective, as we examine how they enable true "development" (Sen, 1999).

> ***Human security, in its broadest sense, embraces far more than the absence of violent conflict. It encompasses human rights, good governance, access to education and health care and ensuring that each individual has opportunities and choices to fulfill his or her potential***
>
> -Kofi Annan, 2000

While both aspects are key to creating real and comprehensive human security, this brief will focus on AI as it impacts *freedom from fear*. This is for two reasons. First, because of the wide range of topics that can be included under threats to *freedom from want*, it is simply beyond the capacity of this brief to effectively cover them all. Secondly, significant attention has rightly been paid to so-called "AI for Good" applications in healthcare, agriculture, education and environmental protection - the ethical considerations of which have been reviewed in several past IEAI Research Briefs.[1] Thus, the remainder of this brief will focus on the potential and challenges to the use of AI in relation to threats against human security that relate to the right to freedom from fear.

The Use of AI for Human Security as Freedom from Fear

In this section, potential applications for AI use in four main areas related to the narrow definition of human security as "freedom from fear" are briefly explored. Namely, reducing human implication in conflict and war, reducing and managing humanitarian crises, protecting fundamental human rights and identifying and preventing mass atrocity crimes. While the section is organized around these use areas (type of

Source Title Page Image: Maxim Hopman, Upsplash

[1]See IEAI Research Briefs on AI and mental health, public health and the green deal: https://www.ieai.sot.tum.de/publications-and-reports/research-briefs/

threat/crises), it may also be helpful to think about a categorization in terms of the purpose of tools used.

Below we will see examples of AI-enabled tools use for:

(1) *Identification*: examples include face recognition or automated detection of objects with aerial imagery.
(2) *Data analysis and pattern recognition*: examples include sifting through vast amounts of diverse data to alert organizations to patterns of hate speech on social media, signs of famine, natural disasters or political instability.
(3) *Information distribution*: examples include automated mapping of potential flood or rescue areas based on requests or chatbots that help refugees find information.

It is important to note that regardless of the type of tool or its application area, the use should be to support human decision-making, not completely automate it. Blind use or misuse of AI-enabled systems can lead to the exacerbation of already dire or complex dynamics related to human security, examples of which will be touched upon later on in this Brief.

- *Reducing human implications of war:*

In discussions about the ethical implications of AI and automated decision-making, there has been consistent talk about the role of technology in conflict. Much of these conversations have centered around the use of automated weapons used by militaries to take out or inflict damage on adversaries. While certainly the misuse or inaccurate use of these tools is a direct threat to military and individual security, the conversations often overlook a key aspect of modern conflict: civilians are the major casualties in most conflicts (Nohle & Robinson, 2017).[2] Thus, while considering the ethical use of AI in conflict, a more important focus might be on how it can be used to *protect* non-combatants and alleviate civilian suffering. Indeed, a clear ethical approach to AI use in military contexts or conflicts would be that it should only be used to avoid harm and not to inflict it.

While civilians are, on occasion, directly targeted in conflict (see section below on human rights and atrocity crimes), much of the casualties are due to error. AI can help with unintentional loss through improved information at improved speeds to better target weapon deployment and battlefield decision-making to protect non-combatants (Devitt et al., 2023). This can be achieved for instance, through ensuring buildings targeted for military reasons are correctly identified (i.e., hospitals, schools or civilian shelters vs. military objectives) and that people are correctly categorized as military targets or civilians. These AI applications have the potential to aid decision-making for defining no combat zones or protected areas in complex and changing environments.

The ability to process diverse and real-time information allows AI-enabled systems to improve the accuracy of these decisions. However, data bias is a major issue, as the distinction between civilian and non-civilian populations is hard to define and categorize, particularly in asymmetric wars. It is an issue that has been made clear in the supposed use of AI systems in the recent conflict in Gaza (Brumfiel, 2023).[3] These serious challenges will be discussed in more detail in the following sections.

AI applications have the potential to aid decision-making for defining no combat zones or protected areas in complex and changing environments.

[2] Current major conflicts in Ukraine and Gaza speak to the fact that traditional military tactics and geopolitical concerns are certainly still alive and relevant in conflict, but they also exemplify the pronounced public and international concern over human security aspects of conflict: the extreme humanitarian crisis in Gaza for instance, or civilian suffering and flow of refugees to western Europe from Ukraine.

[3] For instance, in the current bombardment of Gaza with the aid of an AI-system called the Gospel, some have argued that even though the system is claimed to have helped reduced civilian casualties, it also helps to justify the continued use of bombing without confirmed evidence about the accuracy in terms of preventing civilian casualties (Brumfiel, 2023).

- *Dealing with humanitarian crises:*

While humanitarian crises can be man-made or the result of natural disasters, a human security perspective does not put the focus on the origin of the crisis, but on the similar threats they create to individual security. In these states of crisis, which we are tragically increasingly familiar with, information is murky and scarce, and decisions about severely limited resources have life-or-death implications. Thus, the capability for AI to speed up, clarify and inform decision-making under these stressful conditions has immense potential for impact. There are three distinct areas in which this potential exists: preparedness, response and recovery (Beduschi, 2022; European Parliament, 2019).

The first area where AI-enabled tools have potential to be of increased use is in the **early warning and preparedness** phase of a crisis. There is potential for AI-enabled drones or automated processing of earth observation imagery to provide information that can predict or respond to things such as natural disasters much faster (European Parliament, 2019). "AI technologies have the potential to support humanitarian actors as they implement a paradigm shift from reactive to anticipatory approaches to humanitarian action in conflicts or crises (Beduschi, 2022)". This ability to anticipate crisis rather than react allows for faster movement of aid and more efficiency with resources, lowering human costs and suffering in crises.

For example, the UNHCR, the UN's refugee agency, developed Project Jetson to predict the displacement of people using machine learning and combine data science, statistics, qualitative research and design thinking (UNHCR, 2023). The International Federation of Red Cross and Red Crescent Societies (IFRC) uses ML-based techniques to "forecast extreme weather events, take early action, and thereby prevent human suffering". Through this, they are able to rapidly and automatically allocate resources to geographical areas when a certain threshold is met in the forecast, allowing humanitarian aid to be put into action as efficiently as possible (IFRC, 2019). In a last example, the World Food Program is using AI-based tools to map expected food insecurities, combining "key metrics from various data sources – such as food security information, weather, population size, conflict, hazards, nutrition information, and macro-economic data – to help assess, monitor and predict the magnitude and severity of hunger in near real-time" (WFP, 2023).

This ability to anticipate crisis rather than react allows for faster movement of aid and more efficiency with resources, lowering human costs and suffering in crises.

Response to and management of humanitarian crises is a second application area. In this phase, information dissemination is key. Applications such as semi-automated flood mapping, where "images of flood-prone areas are automatically downloaded and processed by machine learning algorithms to output disaster maps (…) allow for the implementation of live streaming mapping services triggered by direct partner requests or automatic activations." These tools allow response teams to request information at faster paces to respond most effectively on the ground (UNOSAT & UN Global Pulse, n.d.). In other examples, AI techniques are used to gather crowdsourced and micro-level data to create maps for response in natural disasters or refugee areas (HOTOSM, 2023).

Finally, AI tools have the potential to impact crisis **recovery**, where populations in protracted situations need long-term help in reestablishing familial links and basic services. For instance, chatbots can be used to gather information from those on the ground or answer tailored questions in real time to assess community needs. Often used until now in setting more related to sustainable development or "freedoms from want", the potential applications for assessing needs or concerns in conflict or refugee areas are immense (ICRC, 2017). For instance, some chatbots are being used to help refugees integrate into their new host countries (Culbertson et al., 2019). Finally, there are multiple implemented applications using face recognition technology to help people find family members, particularly children, displaced from their homes, such as in large refugee camps (Hirani, 2022), or to manage access to services in refugee areas (Culbertson et al., 2019).

- *Protecting human rights:*

In many ways, modern human rights bring together concepts from both freedoms from want *and* fear to codify these protections in law, thus, enabling human security. Human Rights can be, at a basic level, understood as "a basic set of guarantees that all individuals possess from birth" (Hanlon and Christie, 2016). While the list of these guarantees is debated and involves normative discussions, we can generally think about human rights protections (freedom and equality before the law) and human rights realizations (entitlements such as the right to health care, food or education) as we consider where AI impact to human security in this respect (see IEAI Research Brief - Kriebitz, 2023).[4&5]

If we think about entitlements, AI-enabled tools have immense potential to improve things such as the right to healthcare or education. As acknowledged at the start of the brief, the positive and negative impacts of AI on these issues have been outlined in many other IEAI research briefs, as well as the increasing literature on AI for Good. (Cowls, 2021; Vinuesa, 2020).

Some scholars argue that algorithms are actually less biased than judges themselves toward certain populations, questioning whether improvements in, but not the elimination of, bias is enough to call an AI system "ethical"

Considering protections or freedoms enabled by human rights, we could take the example of equality before the law and rights related to fair treatment. While examples of using AI to promote equal treatment are plentiful (in areas such as banking, hiring or social services), the use of AI in the justice system provides a particularly dramatic case in relation to human rights.

AI-enabled tools are already playing a role in the justice system, including predictive policing, bail, parole and sentencing recommendations (Bagaric et al., 2021). Their use has been heavily criticized with good reasons related to the overreliance of judges or police on systems that have displayed inherent bias (Angwin et al., 2016; Buolamwini & Friedman, 2024) or lack of inclusiveness in their design (Okidegbe, 2021). While the challenges of bias will be further elaborated upon in the next section, it is important to note already here that the use of these type of AI systems can create adverse impacts on procedural rights specifically. The case of predictive policing, for instance, is not only problematic from the perspective of equality before the law, but also from the standpoint of privacy and the presumption of innocence.

On the other hand, some scholars argue that algorithms are actually less biased than judges themselves toward certain populations, questioning whether improvements in, but not the elimination of, bias is enough to call an AI system "ethical" (Bagaric et al., 2021). Moreover, *IF* continuous human oversight, transparency and data bias detection mechanisms are better employed, there are potentials for AI to help with inefficiencies in the legal system that leave people waiting unreasonable periods for bail or parole, for instance, or by actually reducing opportunities for human bias in decision-making.

Thinking about other application areas, AI-enabled tools have been demonstrated to help decision-makers arrive at ethical decisions, for instance, in the context of healthcare (Meier et al., 2022). With efforts to reduce the known bias in data, similar tools could be key in the justice system, again given proper human-in-the-loop mechanisms for oversight, such as incorporating explainable AI techniques.

Moreover, tools that enable the *detection* or *prevention* of human rights violations have the potential for impact. This could include the development of automated detection systems for monitoring violations of political rights or self-

[4] The Universal Declaration on Human Rights gives a detailed overview of these rights (United Nations, 1948).
[5] It is important to note that AI also has the potential to threaten human rights without proper governance in place. Kriebitz (2023) categorizes these threats within (1) purpose, or the possibility that the goal of the AI system is to threaten human rights, (2) process, related to the opacity of AI systems and the rights of those affected by the process and outcomes, or (3) the idea that decisions made may violate the human rights of those impacted, particularly in relation to discrimination of certain groups.

determination, such as scraping social media data for evidence of restrictions on freedom of expression or monitoring voting violations to bring public attention to them in real time (and hopefully correct for them) (P, Simoes, & MacCarthaigh, 2023). Current research undertaken by the Alan Turing Institute[6], for instance, supports this view and points to the potential of due diligence of AI for human rights. Other relevant examples include the most dramatic category: the context of mass atrocity crimes. It is this, in particular, that we turn to next.

- *Identifying and preventing mass atrocity crimes:*

In international law, a clear distinction between international humanitarian law and the identification of atrocity crimes and human rights is matter of debate. In many cases, they cover similar matters (United Nations, 2011). Nonetheless, understanding where AI may interact with these crimes, in particular, is a worthwhile exercise. The Global Center for the Responsibility to Protect identifies four mass atrocity crimes: genocide, war crimes, crimes against humanity and ethnic cleansing (Center for Responsibility to Protect, 2018).

Regardless of the crime, AI generally has the potential to play a role in automating the detection of circumstances ripe for committing atrocities. For instance, through monitoring online behavior and social media posting, AI could quickly comb through vast amounts of information to detect worrisome trends, such as hate speech aimed at inciting violence, persecution or genocide of a certain population. Quick identification allows for earlier intervention, deletion or counter speech[7] of inciteful posts or enhanced situation monitoring (Adams, 2020).

AI generally has the potential to play a role in automating the detection of circumstances ripe for committing atrocities.

Moreover, while still facing technical hurdles; AI-assisted earth observation has the potential to improve the detection time of spotting and reacting to major population movements emblematic of forced displacement or people fleeing mass atrocities.[8] It could also be used for surveillance of well-known potential perpetrators or to sift through extensive documentation for patterns of abuse to document and prosecute crimes (Hao, 2020; Milard and Smith, 2021).

Clarifying Threats from AI to Human Security and Ethical Considerations

The concept of human security and the related threats it prioritizes is an ethical exercise at its core. Are there universal considerations that bring us together as humans and look beyond the geographical and political borders that make us "different" from each other? Does a threat to my neighbors' physical security or human rights represent a threat to my own? If these answers are yes, how do we change our values and/or priorities to improve security for all? Answering these questions and deciding how to prioritize threats and resources requires an evaluation of tradeoffs and understanding relevant ethical principles.

The last section discussed the potential for different AI applications to advance human security, largely covering principles related to promoting rights and doing "good". In this section, highlighted concerns or challenges with AI use for Human Security are organized around ethical themes often identified in AI ethics analyses.[9] While there are too many examples to touch on all potential challenges, through this process, we can begin to think about the tradeoffs between the priorities that emerge under a human security approach.

[6] https://www.turing.ac.uk/ai-human-rights

[7] The IEAI project *Personalized AI-Based Interventions Against Online Norm Violations: Behavioral Effects And Ethical Implications* and related research brief looks at effective mechanisms for counter speech in the context of hate speech (Cypris et al. 2022).

[8] This idea was conceived in the context of the mass atrocities in the Sudan in 2010, researching ways to provide real-time information on potential crimes being committed using satellite data (Enough Project, n.d.).

[9] Building off of well-known frameworks from the OECD (2023) and UNESCO (2021), for instance.

- *Safety, privacy and prevention of harm:*

In using any AI application, safety and prevention of harm must be considered. This can include an AI system functioning incorrectly or being misused, as well as intentional or unintentional impacts that adversely affect users and their environment. When thinking about the narrow definition of human security, as individual protection from harm and acute rights abuses, conflict settings are front and center in terms of where threats may occur and are a clear space where threats to safety, in particular, may result from misuse of AI-enabled tools.

In terms of ensuring safety in using AI at its most dramatic level, it is clear that AI should only be used in conflict settings to avoid harm, not inflict it.

The use of AI in conflict settings has long been contentious, as many argue that lethal automated weapon systems (LAWS) or AI-based decision-making systems that result in military-initiated killing should be a red line (Asaro, 2012; Wareham & Goose, 2016). In framing the discussion around the concept of human security, one has to think about the often more prevalent unintentional impacts of automated systems in conflict zones (civilian casualties), not only intentional and automated deployment of weapons for military targets. To this end, in terms of ensuring safety in using AI at its most dramatic level, it is clear that AI should only be used in conflict settings to avoid harm, not inflict it. The risk of misuse and mistakes (that are amplified by the possible biases that will be expanded below) is too high.

Moving beyond the clearer idea of "do no harm", Devitt et al. (2023) argue for a "minimally-just ethical machine or "MinAI" … (that) could deal only with what is ethically impermissible. That is, MinAI could make "life" decisions. This includes tasks like identifying protected areas or objects that militaries should avoid. If an AI-enabled system can more accurately or quickly identify these symbols or areas compared to a human, there is potential for reductions in threats to civilians.

Another issue related to harm prevention is privacy. There is a lack of standardization *within* the humanitarian community about data use and privacy. Additionally, many humanitarian organizations may be low on resources to devote to concerns such as data protection or AI impact assessments (Culbertson et al., 2019). Moreover, Beduschi (2022), for instance, discusses "surveillance humanitarianism". This implies an over-surveillance that many populations enduring humanitarian crises face, often without their consent (or ability to refuse the releasing of data). Face recognition may help unite families in camps or make service access more efficient. On the other hand, refugees or other people trapped in crisis are vulnerable in ways that make the idea of informed consent for submitting personal data tricky. If access to safety zones or emergency services is contingent on being exposed to data collection or AI-based observation, refusal to be monitored is not a realistic option. Children or traumatized populations also require different considerations for consent to make sure it is given in a truly informed way.

A related ethical challenge to privacy and safety is the misuse of AI-enabled tools. For instance, some agents might use AI as a means to deliberately violate human rights (Kriebitz & Lütge, 2020), as was mentioned earlier in use of technologies at the expense of minorities and vulnerable groups in crises. In this case, the misuse (i.e. to strategically suppress speech or voting) is and should be a major concern.

When speaking about atrocity crimes, the misuse or intentional use of AI-enabled technologies to actually contribute to or even enable mass atrocities also cannot be overlooked. Milard and Smith (2021) have looked at the potential for AI use specifically within the 10 stages of genocide (Fig. 1). The early stages outline the increasing intensity of creating an "us vs. them" mentality or solidifying a "different group" that can then be dehumanized in order to be easily persecuted. While AI has the potential to detect group dehumanizing behavior, it also can play a role in exacerbating, deepening or speeding up this process.

While AI has the potential to detect group dehumanizing behavior, it also can play a role in exacerbating, deepening or speeding up this process.

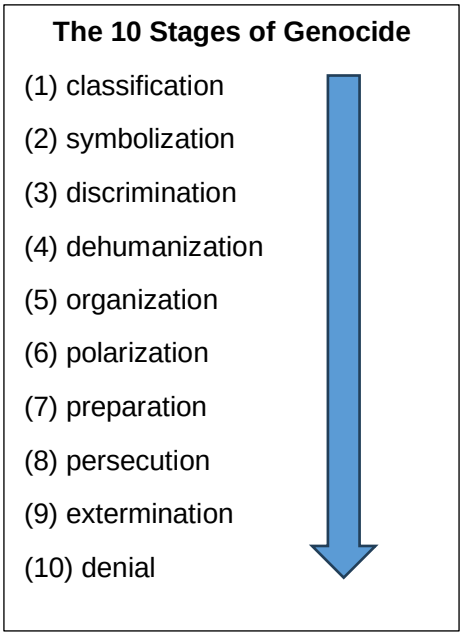

Figure 1: 10 Stages of Genocide (Stanton, 1996)

The spread of information through AI-enabled algorithms on social media may contribute to misinformation waves or create an environment where an individual user perceives opinions to be converging, normalizing discrimination and enhancing polarization. AI tools can also be used to create deep fakes that support a storyline of "otherness" or perceived threats from a categorized group. Moreover, automated social media bots or other applications could be similarly targeted to deliberately spread misinformation about minority groups with the aim of exploiting tensions in a region, something companies and governments have the responsibility to monitor and combat (Adams, 2020).

In terms of moving from polarization to the later stages of genocide aimed at preparing for and implementing violent persecution, AI-enabled surveillance is a particularly worrisome application. For instance, the Chinese tech company Huawei has drawn attention in recent years for indicating the use of their face recognition technology specifically for identifying people from the Uighur minority, a group targeted for persecution in China (D'Alessandra & Gildea, 2022; Kelion, 2021). AI-enabled application for smartphone monitoring have been used to track suspicious behavior of "othered" groups in China as well (Human Rights Watch, 2019).

Such application cases need therefore to be seen in a wider context of other policies or measures that target specific minorities, for instance the mass incarceration of ethnic minorities. The same applies also to atrocities committed in wars. Here, AI guided missile strikes could be coordinated in a way which inflicts more causalities for civilians or which leads to the deliberate displacement of individuals in the context of ethnic cleansing, whether intentional or due to data bias. AI-based technology that intentionally targets ethnic or minority groups for identification or attack has worrisome potential to speed up genocide or ethnic cleansing.

All of these examples are related to a general caution about using tools that work well in one scenario for another purpose. For instance, AI tools that gather information quickly to assist with crises (such as refugee movement or expected needs, damage to buildings or identification of safe zones) may work well for events such as natural disasters, but this does not mean we can automatically apply it to, for instance, combat zones.

In a politically generated crisis, we must be more careful about how information may be used, as there could be military or political incentives to use data to attack civilian groups, promote genocidal activities, delay relief to certain areas or deliberately alter information to create confusion. For example, an AI-enabled face recognition tool may help with organizing humanitarian assistance in a refugee camp. If, however, the refugees are related to political crisis, these tools also have the potential to be misused by governments or political groups to accelerate the identification of members of an opposition or a discriminated against an ethnic group. AI applications that monitor population movements may help humanitarian workers anticipate the amount and location of need, but they may also help armed groups track potential targeted groups. Thus,

We must be careful about how information may be used, as there could be military or political incentives to use data to attack civilian groups, promote genocidal activities, delay relief to certain areas or deliberately alter information to create confusion.

agreement about use, data governance and security will be key in these settings.

- *Fairness and accountability:*

Another challenge, even if safety, misuse and privacy considerations are made, is the AI systems' accuracy and fairness in promoting human security. As with many AI-enabled systems, if biased or inappropriate data is used to train a system, the results, suggestions or decisions made by the system will reflect this inaccuracy and biases. Several specific issues occur with respect to this.

One is the **definitional bias** problem. Meredith Whittaker alluded to an example of this is an interview in 2023. If we were, for instance, to build a tool to predict genocide, we first must ask questions about how to define genocide and what data reflects this definition. Approaching these questions without appropriate thought could change or even exacerbate the problem (Perrigo, 2023). Over-prediction results in loss of resources or unwarranted surveillance of a group, and under-prediction could have dire consequences and human suffering.

A second problem has to do with **historically biased data**. If one were, for instance, to use AI to predict effective aid distribution, this would necessarily be based heavily on historical data on who received aid or what aid distribution looked like in the past. However, this would not necessarily be fair or effective for the current situation. Other factors would have to be incorporated, drawing on contextual expertise or experiences of different stakeholders (Beduschi, 2022).

As also described in the section on human rights abuses above, historical bias against people from certain races or ethnic groups in judicial or policing systems, for instance, would result in training an AI system on biased data, leading to discriminatory results affecting certain groups that share characteristics amplified in the data. This can maintain or even exacerbate already unjust systems. Another example would be the use of face recognition for refugee access to services. These systems have been criticized for having a significant racial bias, again because of the data they were trained on (Buolamwini & Gebru, 2018). This could lead to misidentification for certain groups, sometimes limiting their access to services, or in more dramatic cases, such as with the case of the Uighur in China or the use by police in the US, false arrests or detainment.

This relates also to the third problem of **geographically biased data**. While there is tremendous potential to use AI-enabled systems in conflict and humanitarian crises to protect human rights or monitor and deter mass atrocities, many of these systems will be developed or designed outside of the region where they intend to be used.

While there is tremendous potential to use AI-enabled systems in conflict and humanitarian crises … many of these systems will be developed or designed outside of the region where they intend to be used.

A lack of relevant data or a lack of concern about collecting local data will result in AI-based tools working better in some locations than others. This is unjust when it means that victims of an earthquake or other disasters are able to rely on advanced tools in somewhere like California, but not in Afghanistan. It's also unjust that civilians in a conflict zone in Eastern Europe are able to be better protected than those in Somalia because they had more accurate AI systems assisting crisis response efforts. One example of this comes from AI-based identification of earth observation data. Schools or hospitals may look very different in African countries compared to where the AI tools are trained. If developers do not take this into account and emphasize combining training data with ground truthing, the systems will be less effective for identifying non-military objects in conflict zones where they could perhaps make the most effective impact for humanitarian relief.[10]

Moreover, given the complexity of the environments where threats to human security are occurring, as will

[10] For more on this topic, see the IEAI interview from 2022 with Manuel García-Herranz of UNICEF - https://www.ieai.sot.tum.de/wp-content/uploads/2022/05/Reflections-on-AI-Ethics_Manuel-Herranz-FINAL.pdf

be discussed more in the next section, accountability for ensuring fairness and misuse becomes a cloudy but important discussion point. While some larger organizations, such as UNICEF, may have data science teams building tools in-house, many organizations will be users of tools developed by outside technology companies. Responsibility for appropriate data gathering and training, as well as for protecting against misuse, is a conversation that stakeholders need to have in these scenarios.

- *Inclusion, oversight, and transparency:*

Empowerment is a cross-cutting issue between freedom from want and fear. AI and other technologies can enable feelings of empowerment. Still, without consideration, they can also leave people with feelings of worry or uncertainty about what is going on in their environment and if they have any control over how technology is used around them. This makes inclusion, transparency and oversight vital considerations for effective and ethical AI use.

As the brief has shown, human security threats are complex and often unstable. This means the stakeholders involved may be diverse in terms of language, age, sensitivities and technological awareness (Culbertson et al., 2019), and they may be changing throughout a given crisis. This creates a challenging environment in terms of inclusion and oversight of AI-enabled technologies. Moreover, children or traumatized populations may require different attention or considerations in the inclusion process.

The stakeholders involved may be diverse in terms of language, age, sensitivities and technological awareness and they may be changing throughout a given crisis.

To deal with the challenges of inclusion, as well as privacy and consent, and to increase knowledge for stakeholders about where and when it might be beneficial to employ AI, UN-based scientists have suggested expanding the *society-in-the-loop* algorithm concept. This concept of embedding the general will into an algorithmic social contract would allow for "both humanitarian responders and affected populations to understand and oversee algorithmic decision-making that affects them" (Oroz, 2017). Increasing inclusion, in this sense, not only makes for more ethical AI, but if stakeholders are brought in on the information sharing and development, there is also a higher likelihood that tools will be accepted, useable and beneficial once deployed. RAND analysts have also advocated for and introduced inclusion mechanisms through surveys (for instance with refugee populations) of what tools are actually working and where (Culbertson et al., 2019).

Final Thoughts - A Human Security Approach to AI

The challenges related to improving human security explored in this Brief remain immense and require a collective effort to promote development and peace worldwide. Whether it be the threat of conflict, natural disaster, enduring poverty or unjust treatment, defining the threat through the lens of individual insecurity and suffering can inform our response to it. AI has tremendous potential to help under-resourced organizations provide desperately needed services. Used correctly, it could support decision-making or reduce biases in protecting civilians and their human rights. There is, however, a sincere need for human-in-the-loop processes that guard against bias, harm and lack of inclusion in developing and deploying these tools. This means that AI is developed with mechanisms for human monitoring and input throughout the tool's lifecycle and that is aids rather than automates decision-making.

To be sure, there is a strong and continuous relationship between individual security and collective security. But a human security approach allows us to dissect this relationship and think beyond the traditional players and borders that have long defined the international security regime. The concept reminds us that regardless of nationality, the basic needs that define our day-to-day lives are the same and should be an intrinsic right for everyone.

This way of thinking could help advance an approach to AI governance as well. Rethinking how we define the "players", how we think about governance beyond the concept of borders and how we consider and prioritize AI risk as it affects individual security and the day-to-day flourishing of communities can create insights into how we approach governing the use of AI.

References

Adams, S. 2020. Hate Speech and Social Media: Preventing Atrocities and Protecting Human Rights Online. Center for the Responsibility to Protect. https://www.globalr2p.org/publications/hate-speech-and-social-media-preventing-atrocities-and-protecting-human-rights-online/ (accessed 21.12.23)

Angwin, J.L., S. Mattu &L. Kirchner. 2016. Machine Bias: There's software used across the country to predict future criminals. And it's bias against blacks. ProPublica, 16 May 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed 30.01.24)

Annan, K. 2000. Press Release SG/SM/7382, Secretary-General Salutes International Workshop on Human Security in Mongolia. Ulaanbaatar, May 8-10, 2000. https://press.un.org/en/2000/20000508.sgsm7382.doc.html (accessed 30.01.24)

Asaro, P. 2012. On banning autonomous lethal systems: human rights, automation and the dehumanizing of lethal decision-making. Int. Rev. Red Cross 94, 687–709.

Bagaric, M., J. Svilar, M. Bull, D. Hunter, & N. Stobbs. 2022. The Solution to the Pervasive Bias and Discrimination in the Criminal Justice System: Transparent and Fair Artificial Intelligence. American Criminal Law Review, Vol. 59, 1.

Beduschi, A. 2022. Harnessing the potential of artificial intelligence for humanitarian action: Opportunities and risks. International Review of the Red Cross, 104 (919). doi:10.1017/S1816383122000261

Brumfiel, G. 2023. Israel is using an AI system to find targets in Gaza. Experts say it's just the start. NPR. 14 Dec 2023. https://www.npr.org/2023/12/14/1218643254 (accessed 25.1.2024)

Buolamwini, J. & B. Friedman. 2024. How the Federal Government Can Rein In A.I. in Law Enforcement. New York Times, 2 Jan 2024. https://www.nytimes.com/2024/01/02/opinion/ai-police-regulation.html (accessed 3.1.2024)

Buolamwini, J. & T. Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Conference on Fairness, Accountability and Transparency. PMLR, pp. 77–91.

Center for the Responsibility to Protect. 2018. Background Briefing: Defining the Four Mass Atrocity Crimes.

Chancel, L., T. Piketty, E. Saez, G. Zucman et al. 2021. World Inequality Report 2022. World Inequality Lab. www.wir2022.wid.world (accessed 30.01.2024)

Cowls, J., A. Tsamados, M. Taddeo et al. 2021. A definition, benchmark and database of AI for social good initiatives. Nat Mach Intell 3, 111–115. doi:10.1038/s42256-021-00296-0

Culbertson, S., J. Dimarogonas, K. Costello, & S. Lanna. 2019. Crossing the Digital Divide: Applying Technology to the Global Refugee Crisis. RAND Corporation.

Cypris, N.F., S. Engelmann, J. Sasse, J. Grossklags & A. Baumert. 2022. Intervening Against Online Hate Speech: A Case for Automated Counterspeech. IEAI Reserch Brief, April 2022.

D'Alessandra, F. & R.J. Gildea. 2022. Technological Change and the UN Framework of Analysis for Atrocity Crimes. Stimson. https://www.stimson.org/2022/technological-change-and-the-un-framework-of-analysis-for-atrocity-crimes/ (accessed 21.12.23)

Devitt, S.K., J. Scholz, T. Schless, et al. 2023. Developing a trusted human-AI network for humanitarian benefit. Digi War. doi:10.1057/s42984-023-00063-y

Enough Project. N.d. Satellite Sentinel Project. https://enoughproject.org/about/past-campaigns/satellite-sentinel-project (accessed 21.12.23)

European Parliament. 2019. Technological Innovation for Humanitarian Aid and Assistance. European Parliamentary Research Service, Scientific Foresight Unit (STOA), PE 634.411 – May 2019.

Galtung, J. & D. Fischer. 2013. Johan Galtung. SpringerBriefs on Pioneers in Science and Practice, doi:10.1007/978-3-642-32481-9_17.

Hao, K. 2020. Human rights activists want to use AI to help prove war crimes in court. MIT Technology Review. 25 June 2020. https://www.technologyreview.com/2020/06/25/1004466/ai-could-help-human-rights-activists-prove-war-crimes/ (accessed 21.12.23)

Hanlon, R.J. & K. Christie. 2016. Freedom from Fear, Freedom from Want: An Introduction to Human Security. University of Toronto Press.

Hirani, S. 2022. SexyFace and Recognizing Refugees – Dr. Vivienne Ming's Journey of Using AI for Good. Harvard: Digital Innovation and Transformation. https://d3.harvard.edu/platform-digit/submission/sexyface-and-recognizing-refugees-dr-vivienne-mings-journey-of-using-ai-for-good/. (accessed 19.12.2023)

Human Rights Watch. 2019. China's Algorithms of Repression: Reverse Engineering a Xinjiang Police Mass Surveillance App. https://www.hrw.org/report/2019/05/01/chinas-algorithms-repression/reverse-engineering-xinjiang-police-mass (accessed 21.12.23)

International Federation of Red Cross and Red Crescent Societies (IFRC). 2019. Forecast-based Financing: A new era for the humanitarian system. https://www.forecast-based-financing.org/wp-content/uploads/2019/03/DRK_Broschuere_2019_new_era.pdf (accessed 19.12.2023)

International Federation of Red Cross and Red Crescent Societies (IFRC). 2017. Humanitarian Futures for Messaging Apps: Understanding the Opportunities and Risk for Humanitarian Action.

Intergovernmental Panel on Climate Change (IPCC). 2023. Summary for Policymakers. In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 1-34. doi: 10.59327/IPCC/AR6-9789291691647.001

Kelion, L. 2021. Huawei patent mentions use of Uighur-spotting tech. BB News. 13 Jan 2021. https://www.bbc.com/news/technology-55634388 (accessed 21.21.23)

Kriebitz, A. & C. Lütge. 2020. Artificial Intelligence and Human Rights: A Business Ethical Assessment. Business and Human Rights Journal, 5(1):84-104. doi:10.1017/bhj.2019.28

Kriebitz, A. 2023. Protecting and Realizing Human Rights in the Context of Artificial Intelligence: A Problem Statement. IEAI Research Brief - April 2023.

Meier, L.J., A. Hein, K. Diepold & A. Buyx. 2022. Algorithms for Ethical Decision-Making in the Clinic: A Proof of Concept. The American Journal of Bioethics, 22:7, 4-20, doi: 10.1080/15265161.2022.2040647

Milard, M. & S. Smith. 2021. How AI can either exacerbate or prevent genocides: Reflection based on the 10 Stages of Genocide. Budapest Center for Mass Atrocities Prevention. https://www.genocideprevention.eu/files/10_stages__AI.pdf (accessed 28.2.24)

Nohle, E., & I. Robinson. 2017. War in cities: The 'reverberating effects' of explosive weapons. ICRC Blog. Retrieved from https://blogs.icrc.org/law-and-policy/2017/03/02/war-in-cities-the-reverberating-effects-of-explosive-weapons/ (accessed 30.01.2024)

OECD. 2023. Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (accessed 2.1.2024)

Okidegbe, N. 2022. The Democratizing Potential Of Algorithms? 53 Connecticut Law Review 739.

Okidegbe, N. 2021. Discredited Data. Cornell Law Review, Vol. 107, Issue 7.

P, D., S. Simoes, & M. MacCarthaigh. 2023. AI and core electoral processes: Mapping the horizons. AI Magazine 44: 218–239. doi:10.1002/aaai.12105

Perrigo, B. 2023. Time 100 AI: Meredith Whittaker. Time Magazine, 7.9.2023. https://time.com/collection/time100-ai/6309018/meredith-whittaker/ (accessed 3.1.2024)

Sen, A. 1999. Development as Freedom. Oxford, U.K.: Oxford University Press.

Stanton, G. 1996. The Logic of the Ten Stages of Genocide. Genocide Watch. https://www.genocidewatch.com/tenstages (accessed 21.12.23)

UNESCO. 2021. Recommendation on the Ethics of Artificial Intelligence SHS/BIO/REC-AIETHICS/202. https://unesdoc.unesco.org/ark:/48223/pf0000380455 (accessed 2.1.2024)

United Nations. 1948. Universal Declaration of Human Rights. https://www.un.org/en/about-us/universal-declaration-of-human-rights (accessed 30.01.2024)

United Nations. 2011. Guiding Principles for Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework. https://unglobalcompact.org/library/2 (accessed 25.1.2024)

United Nations Development Program (UNDP). 1994. Human Development Report. https://hdr.undp.org/system/files/documents/hdr1994encompletenostatspdf.pdf

United Nations High Council on Refugees (UNHCR). 2023. Project Jetson. https://www.unhcr.org/innovation/project-jetson/ (accessed 19.12.2023)

United Nations High Council on Refugees (UNHCR). 2022. Global Trends Report 2022. https://unhcr.org/global-trends-report-2022 (accessed 21.12.2023)

UNOSAT & UN Global Pulse. N.d. Fusing AI into Satellite Image Analysis to Inform Rapid Response to Floods. https://unitar.org/about/news-stories/news/fusing-ai-satellite-image-analysis-inform-rapid-response-floods. (accessed 19.12.2023)

Vinuesa, R., H. Azizpour, I. Leite et al. 2020. The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications 11 (233). doi:10.1038/s41467-019-14108-y

Wareham, M & Goose, S. 2016. The growing international movement against killer robots. Harvard Int. Rev. 37, 28–34. https://www.hrw.org/news/2017/01/05/growing-international-movement-against-killer-robots (accessed 07.02. 2024)

World Food Program (WFP). 2023. Hunger Map. https://hungermap.wfp.org/ (accessed 19.12.2023)