IRRC**_**

Harnessing the potential of artificial intelligence for humanitarian action: Opportunities and risks

Ana Beduschi
Ana Beduschi is an Associate Professor of Law at the University of Exeter and a Senior Research Fellow at the Geneva Academy of International Humanitarian Law and Human Rights.

Abstract

Data-driven artificial intelligence (AI) technologies are progressively transforming the humanitarian field, but these technologies bring about significant risks for the protection of vulnerable individuals and populations in situations of conflict and crisis. This article investigates the opportunities and risks of using AI in humanitarian action. It examines whether and under what circumstances AI can be safely deployed to support the work of humanitarian actors in the field. The article argues that AI has the potential to support humanitarian actors as they implement a paradigm shift from reactive to anticipatory approaches to humanitarian action. However, it recommends that the existing risks, including those relating to algorithmic bias and data privacy concerns, must be addressed as a priority if AI is to be put at the service of humanitarian action and not to be deployed at the expense of humanitarianism. In doing so, the article contributes to the current debates on whether it is possible to harness the potential of AI for responsible use in humanitarian action.

Keywords: artificial intelligence, do no harm, humanitarian action, humanitarian principles, human rights.

: : : : : : :

Introduction

The use of digital technologies in humanitarian action is not a new phenomenon. Humanitarian actors have been utilizing digital technologies to assist and protect populations affected by conflict and crisis for decades.[1] Yet, contemporary advances in computational power, coupled with the availability of vast amounts of data (including big data), have allowed for more widespread use of digital technologies in the humanitarian context.[2] The COVID-19 pandemic has further accelerated the trend of the use of digital technologies to help maintain humanitarian operations.[3]

Artificial intelligence (AI) is one such digital technology that is progressively transforming the humanitarian field. Although there is no internationally agreed definition, AI is broadly understood as "a collection of technologies that combine data, algorithms and computing power".[4] These technologies consist of

> software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal.[5]

This definition comprises two main elements: knowledge-based systems and machine learning systems. Knowledge-based systems are seen in computer

1   Patrick Meier, "New Information Technologies and Their Impact on the Humanitarian Sector", *International Review of the Red Cross*, Vol. 93, No. 884, 2011; Anja Kaspersen and Charlotte Lindsey-Curtet, "The Digital Transformation of the Humanitarian Sector", *Humanitarian Law and Policy Blog*, 5 December 2016, available at: https://blogs.icrc.org/law-and-policy/2016/12/05/digital-transformation-humanitarian-sector/ (all internet references were accessed in April 2022); Dzhennet-Mari Akhmatova and Malika-Sofi Akhmatova, "Promoting Digital Humanitarian Action in Protecting Human Rights: Hope or Hype", *International Journal of Humanitarian Action*, Vol. 5, 2020.
2   Ana Beduschi, "The Big Data of International Migration: Opportunities and Challenges for States under International Human Rights Law", *Georgetown Journal of International Law*, Vol. 49, No. 4, 2018; Michael Pizzi, Mila Romanoff and Tim Engelhardt, "AI for Humanitarian Action: Human Rights and Ethics", *International Review of the Red Cross*, Vol. 102, No. 913, 2021.
3   Saman Rejali and Yannick Heiniger, "The Role of Digital Technologies in Humanitarian Law, Policy and Action: Charting a Path Forward", *International Review of the Red Cross*, Vol. 102, No. 913, 2021; Jo Burton, "'Doing no Harm' in the Digital Age: What the Digitalization of Cash Means for Humanitarian Action", *International Review of the Red Cross*, Vol. 102, No. 913, 2021; John Bryant, Kerrie Holloway, Oliver Lough and Barnaby Willitts-King, *Bridging Humanitarian Digital Divides during Covid-19*, Overseas Development Institute, London, 2020; Theodora Gazi and Alexandros Gazis, "Humanitarian Aid in the Age of COVID-19: A Review of Big Data Crisis Analytics and the General Data Protection Regulation", *International Review of the Red Cross*, Vol. 102, No. 913, 2021.
4   European Commission, *White Paper on Artificial Intelligence: A European Approach to Excellence and Trust*, COM (2020) 65 final, 2020, p. 2.
5   European Union High Level Expert Group on Artificial Intelligence, *A Definition of AI: Main Capabilities and Scientific Disciplines*, Brussels, 2019, p. 6.

programs that use an existing knowledge base to solve problems usually requiring specialized human expertise.[6] Machine learning is "the systematic study of algorithms and systems that improve their knowledge or performance with experience".[7] Through machine learning, machines can be trained to make sense of data. For example, AI systems can be trained to perform tasks such as natural language processing, utilizing the computer's capacity to parse and interpret text and spoken words.[8] Deep learning, a subset of machine learning, is particularly used to perform tasks such as image, video, speech and audio processing.[9] The analysis in this article applies to both categories of systems.

AI systems often draw on large amounts of data, including information directly collected by humanitarian actors and other sources such as big data, to learn, find patterns, make inferences about such patterns, and predict future behaviour.[10] Big data, or "large volumes of high velocity, complex and variable data",[11] is also increasingly relevant in the humanitarian context. An important part of big data originates in user-generated content available on social media and online platforms, such as text, images, audio and video.[12] Social media platforms tend to provide specific channels for users to engage and communicate during conflicts or crises.[13] For example, Facebook has enabled safety checks whereby users can report their status as natural disasters or other conflicts or emergencies unfold.[14] AI systems can build on these different types of data to map the evolution of conflicts and crises.

In this regard, 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.[15] For example, in 2019, AI-supported

---

6   Martin Swain, "Knowledge-Based System", in Werner Dubitzky, Olaf Wolkenhauer, Kwang-Hyun Cho and Hiroki Yokota (eds), *Encyclopedia of Systems Biology*, Springer, New York, 2013.

7   Peter Flach, *Machine Learning: The Art and Science of Algorithms that Make Sense of Data*, Cambridge University Press, Cambridge, 2012, p. 3.

8   *Ibid.*; Jacob Eisenstein, *Introduction to Natural Language Processing*, MIT Press, Cambridge, MA, 2019.

9   Yann LeCun, Yoshua Bengio and Geoffrey Hinton, "Deep Learning", *Nature*, Vol. 521, 2015; Neil Savage, "How AI and neuroscience drive each other forwards", *Nature*, Vol. 571, No. 7553, 2019.

10   Jenna Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms", *Big Data & Society*, Vol. 3, No. 1, 2016; Sandra Wachter, Brent Mittelstadt and Chris Russell, "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR", *Harvard Journal of Law & Technology*, Vol. 31, No. 2, 2018.

11   Tech America Foundation, *Demystifying Big Data: A Practical Guide to Transforming the Business of Government*, Washington, DC, 2012.

12   Amir Gandomi and Murtaza Haider, "Beyond the Hype: Big Data Concepts, Methods, and Analytics", *International Journal of Information Management*, Vol. 35, No. 2, 2015.

13   Billy Haworth and Eleanor Bruce, "A Review of Volunteered Geographic Information for Disaster Management", *Geography Compass*, Vol. 9, No. 5, 2015; A. Beduschi, above note 2; Pankaj Sharma and Ashutosh Joshi, "Challenges of Using Big Data for Humanitarian Relief: Lessons from the Literature", *Journal of Humanitarian Logistics and Supply Chain Management*, Vol. 10, No. 4, 2020; T. Gazi and A. Gazis, above note 3.

14   Facebook, "Crisis Response", available at: www.facebook.com/about/safetycheck/.

15   Mark Lowcock, "Anticipation Saves Lives: How Data and Innovative Financing Can Help Improve the World's Response to Humanitarian Crises", speech delivered at the London School of Economics, 2019, available at: https://reliefweb.int/report/world/mark-lowcock-under-secretary-general-humanitarian-affairs-and-emergency-relief; United Nations Office for the Coordination of Humanitarian Affairs (OCHA), *From Digital Promise to Frontline Practice: New and Emerging Technologies in Humanitarian Action*,

disaster mapping helped humanitarians to provide a swift emergency response in Mozambique.[16] Data-driven AI systems can also build on predictive analytics techniques, which seek to identify patterns and relationships in data, to predict developments in the field.[17] For example, Project Jetson, an initiative of the Office of the United Nations High Commissioner for Refugees' (UNHCR), uses predictive analytics to forecast forced displacement of people.[18]

However, scholars and activists have increasingly voiced concerns about the risks posed by the deployment of AI in the humanitarian context. These concerns range from the dangers of "surveillance humanitarianism"[19] to the excesses of "techno-solutionism"[20] and the problems related to a potential rise in "techno-colonialism".[21] These are significant risks, as they may expose populations already affected by conflict or crises to additional harms and human rights violations.

Against this backdrop, this article investigates the opportunities and risks of using AI in humanitarian action. It draws on legal, policy-oriented and technology-facing academic and professional literature to assess whether and under what circumstances AI can be safely deployed to support the work of humanitarian actors in the field. Although the academic and professional literature points to the heightened interest in using AI for military action in armed conflicts, that area remains outside of the scope of this article.[22] This choice is justified by the

---

New York, 2021; Christopher Chen, "The Future is Now: Artificial Intelligence and Anticipatory Humanitarian Action", *Humanitarian Law and Policy Blog*, 19 August 2021, available at: https://blogs.icrc.org/law-and-policy/2021/08/19/artificial-intelligence-anticipatory-humanitarian/.

16 OCHA, above note 15, p. 7.

17 A. Gandomi and M. Haider, above note 12, p. 143.

18 See the Project Jetson website, available at: https://jetson.unhcr.org.

19 "Surveillance humanitarianism" is a term that refers to the increase in data collection practices by humanitarian organizations that, without the appropriate safeguards, may inadvertently amplify the vulnerability of individuals in need of humanitarian aid. See Mark Latonero, "Stop Surveillance Humanitarianism", *New York Times*, 11 July 2019. See also Keren Weitzberg, Margie Cheesman, Aaron Martin and Emrys Schoemaker, "Between Surveillance and Recognition: Rethinking Digital Identity in Aid", *Big Data & Society*, Vol. 8, No. 1, 2021.

20 "Techno-solutionism" is a term that refers to decision-makers' willingness to utilize digital technologies to solve complex societal problems which require more than solely technical solutions. See Mark Duffield, "The Resilience of the Ruins: Towards a Critique of Digital Humanitarianism", *Resilience*, Vol. 4, No. 3, 2016; Petra Molnar, *Technological Testing Grounds*, EDRi and Refugee Law Lab, Brussels, 2020, available at: https://edri.org/wp-content/uploads/2020/11/Technological-Testing-Grounds.pdf.

21 "Techno-colonialism" is a term that broadly refers to practices in digital innovation which can lead to reproducing the colonial relationships of dependency and inequality amongst different populations around the world. See Mirca Madianou, "Technocolonialism: Digital Innovation and Data Practices in the Humanitarian Response to Refugee Crises", *Social Media & Society*, Vol. 5, No. 3, 2019; Nick Couldry and Ulises A. Mejias, "Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject", *Television & New Media*, Vol. 20, No. 4, 2019.

22 Rain Liivoja, Kobi Leins and Tim McCormack, "Emerging Technologies of Warfare", in Rain Liivoja and Tim McCormack (eds), *Routledge Handbook of the Law of Armed Conflict*, Routledge, London, 2016; Ronald Alcala and Eric Talbot Jensen, *The Impact of Emerging Technologies on the Law of Armed Conflict*, Oxford University Press, Oxford, 2019; International Committee of the Red Cross (ICRC), *Artificial Intelligence and Machine Learning in Armed Conflict: A human-centred Approach*, Geneva, 2019; Hitoshi Nasu, "Artificial Intelligence and the Obligation to Respect and to Ensure Respect for IHL", in Eve Massingham and Annabel McConnachie (eds), *Ensuring Respect for International Humanitarian Law*, Routledge, London, 2020; Jai Galliott, Duncan MacIntosh and Jens David Ohlin,

growing uses of AI outside of military action, in support of humanitarian assistance in situations of conflict, disaster and crisis.

The analysis proceeds in three steps. Firstly, the article examines the opportunities brought about by AI to support humanitarian actors' work in the field. Secondly, it evaluates the existing risks posed by these technologies. Thirdly, the article proposes key recommendations for deploying AI in the humanitarian context, based on the humanitarian imperative of "do no harm". Finally, the article draws conclusions on whether it is possible to safely leverage the benefits of AI while minimizing the risks it poses for humanitarian action.

AI in support of a paradigm change: From reactive to anticipatory approaches to humanitarian action

As noted earlier, AI has the potential to support humanitarian actors as they implement a paradigm shift from reactive to anticipatory approaches to humanitarian action.[23] This shift entails acting as soon as a crisis may be foreseen and proactively mitigating the adverse impact on vulnerable people.[24] In this regard, AI technologies may further expand the toolkit of humanitarian missions in their three main dimensions: preparedness, response and recovery.

Preparedness is the continuous process that aims to understand the existing risks and propose actions to respond to those risks, thus supporting a more effective humanitarian response to crises and emergencies.[25] Response focuses on the delivery of assistance to those in need,[26] while recovery refers to programmes that go beyond the provision of immediate humanitarian relief.[27] As such, recovery is an important element, as contemporary humanitarian crises tend to be increasingly complex and protracted, transcending the boundaries between humanitarian aid and development cooperation.[28]

Lethal Autonomous Weapons: Re-examining the Law and Ethics of Robotic Warfare, Oxford University Press, Oxford, 2021.

23 M. Lowcock, above note 15; OCHA, above note 15.

24 M. Lowcock, above note 15.

25 Inter-Agency Standing Committee, *The Implementation of the Humanitarian Programme Cycle*, Geneva, 2015.

26 *Ibid.*

27 International Federation of Red Cross and Red Crescent Societies (IFRC), "Recovery", available at: www.ifrc.org/recovery.

28 Executive Board of the United Nations Development Programme and of the United Nations Population Fund, *Role of UNDP in Crisis and Post-Conflict Situations*, UN Doc. DP/2001/4, Geneva, 2000, para. 48; Lucy Earle, "Addressing Urban Crises: Bridging the Humanitarian–Development Divide", *International Review of the Red Cross* Vol. 98, No. 901, 2016; Atsushi Hanatani, Oscar A. Gómez and Chigumi Kawaguchi, *Crisis Management Beyond the Humanitarian-Development Nexus*, Routledge, London, 2018; Jon Harald Sande Lie, "The Humanitarian-Development Nexus: Humanitarian Principles, Practice, and Pragmatics", *Journal of International Humanitarian Action*, Vol. 5, 2020.

Preparedness

AI technologies can support humanitarian preparedness as AI systems can be used to analyze vast amounts of data, thus providing essential insights about potential risks to affected populations. These insights can inform humanitarians about such risks before a crisis or humanitarian disaster unfolds.[29] In this regard, predictive analytics, which builds on data-driven machine learning and statistical models, can be used to calculate and forecast impending natural disasters, displacement and refugee movements, famines, and global health emergencies.[30] To date, such systems have performed best for early warnings and short-term predictions.[31] Yet, their potential is significant, as AI systems performing predictive analytics can be instrumental for preparedness.

For example, the Forecast-based Financing programme deployed by the International Federation of Red Cross and Red Crescent Societies (IFRC) enables the swift allocation of humanitarian resources for early action implementation.[32] This programme uses a variety of data sources, such as meteorological data and market analysis, to determine when and where humanitarian resources should be allocated.[33]

Another example is UNHCR's Project Jetson, which uses predictive analytics to forecast forced displacement contributing to the escalation of violence and conflict in Somalia.[34] Project Jetson builds on various data sources, including climate data (such as river levels and rain patterns), market prices, remittance data, and data collected by the institution to train its machine learning algorithm.

In another context, the World Food Programme has developed a model that uses predictive analytics to forecast food insecurity in conflict zones, where traditional data collection is challenging.[35] This model provides a map depicting the prevalence of undernourishment in populations around the world.

But would deploying AI systems, particularly those using predictive analytics models, lead to better preparedness in humanitarian action? Any answer

---

29 Kevin Hernandez and Tony Roberts, *Predictive Analytics in Humanitarian Action: A Preliminary Mapping and Analysis*, Institute for Development Studies, London, 2020.

30 *Ibid.*; Petra Molnar, "Technology on the Margins: AI and Global Migration Management from a Human Rights Perspective", *Cambridge Journal of International Law*, Vol. 8, No. 2, 2019; Ana Beduschi, "International Migration Management in the Age of Artificial Intelligence", *Migration Studies*, Vol. 9, No. 3, 2020; Centre for Humanitarian Data, "OCHA-Bucky: A COVID-19 Model to Inform Humanitarian Operations", The Hague, 2021, available at: https://centre.humdata.org/ocha-bucky-a-covid-19-model-to-inform-humanitarian-operations/; T. Gazi and A. Gazis, above note 3.

31 K. Hernandez and T. Roberts, above note 29; Jessica Bither and Astrid Ziebarth, *AI, Digital Identities, Biometrics, Blockchain: A Primer on the Use of Technology in Migration Management*, Migration Strategy Group on International Cooperation and Development, Berlin, 2020.

32 IFRC, "Forecast-based Financing: A New Era for the Humanitarian System", 2021, available at: www.forecast-based-financing.org/wp-content/uploads/2019/03/DRK_Broschuere_2019_new_era.pdf.

33 Toke Jeppe Bengtsson, *Forecast-based Financing: Developing Triggers for Drought*, Lund University, Lund, 2018.

34 See the Project Jetson website, above note 18; UNHCR Innovation Service, "Is It Possible to Predict Forced Displacement?", *Medium*, 2019, available at: https://medium.com/unhcr-innovation-service/is-it-possible-to-predict-forced-displacement-58960afe0ba1.

35 See the World Food Programme HungerMap, available at: https://hungermap.wfp.org/.

to this question must be nuanced. On the one hand, in some contexts, AI systems may be beneficial to humanitarian action as they may contribute to a better understanding of the situation and better anticipation of responses. For instance, better preparedness can contribute to early allocation of resources, which may be crucial for the effectiveness of operations on the ground. On the other hand, the analysis of historical data should not be the only way to inform and frame future action. Models based on the analysis of past data may fail to consider variables such as changes in human behaviour and the environment, and may thus provide incomplete or erroneous predictions. For instance, during the COVID-19 pandemic, most AI models failed to provide efficient support to medical decision-making in tackling outbreaks of the disease.[36] That was partly due to the low quality of the data (historical data not relating to COVID-19) and the high risk of bias.[37] In addition, AI systems focusing on the analysis of past data might continue to reproduce errors and inaccuracies and perpetuate historical inequalities, biases and unfairness.[38] Accordingly, careful consideration of the specificities of the humanitarian context in which AI systems are to be deployed may help avoid unnecessary recourse to technologies and prevent exacerbated techno-solutionism.

Techno-solutionism, or faith in technologies to solve most societal problems, has proven to yield mixed results in the humanitarian field. For instance, studies have shown that focusing on big data analysis for anticipating Ebola outbreaks in West Africa was not always as effective as investing in adequate public health and social infrastructure.[39] Working closely with affected communities – for example, through participatory design[40] – could help to tailor anticipatory interventions to key community needs, thus better informing and preparing humanitarian action before a conflict or crisis unfolds. This can also apply to AI systems used in humanitarian response, as discussed in the following subsection.

Response

AI systems can be used in ways that may support humanitarian response during conflicts and crises. For instance, recent advances in deep learning, natural

---

36 Laure Wynants *et al.*, "Prediction Models for Diagnosis and Prognosis of Covid-19: Systematic Review and Critical Appraisal", *BMJ*, Vol. 369, 2020.

37 *Ibid.*, pp. 5–6.

38 See the discussion in the section below on "AI at the Expense of Humanitarianism: The Risks for Affected Populations". See also Rashida Richardson, Jason Schultz and Kate Crawford, "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice", *New York University Law Review*, Vol. 94, 2019, p. 224.

39 Dilon Wamsley and Benjamin Chin-Yee, "COVID-19, Digital Health Technology and the Politics of the Unprecedented", *Big Data & Society*, Vol. 8, No. 1, 2021, p. 3.

40 Participatory design is a process that includes a variety of stakeholders from the early stages of technology design. See Peter M. Asaro, "Transforming Society by Transforming Technology: The Science and Politics of Participatory Design", *Accounting, Management and Information Technologies*, Vol. 10, No. 4, 2000; Elizabeth Rosenzweig, "UX Thinking", in Elizabeth Rosenzweig (ed.), *Successful User Experience: Strategy and Roadmaps*, Elsevier, Amsterdam, 2015.

language processing and image processing allow for faster and more precise classification of social media messages during crisis and conflict situations. This can assist humanitarian actors in responding to emergencies.[41] In particular, these advanced AI technologies can help identify areas that would benefit from streamlined delivery of assistance to those in need.

For example, the Emergency Situation Awareness platform monitors content on Twitter in Australia and New Zealand to provide its users with information about the impact and scope of natural disasters such as earthquakes, bushfires and floods as they unfold.[42] Similarly, Artificial Intelligence for Disaster Response, an open platform that uses AI to filter and classify social media content, offers insights into the evolution of disasters.[43] Platforms such as these can triage and classify content, such as relevant images posted on social media showing damages to infrastructure and the extent of harm to affected populations, which can be useful for disaster response and management.[44]

Another example is the Rapid Mapping Service, a project jointly developed by the United Nations (UN) Institute for Training and Research, the UN Operational Satellite Applications Programme, and UN Global Pulse.[45] This project applies AI to satellite imagery in order to rapidly map flooded areas and assess damage caused by conflict or natural disasters such as earthquakes and landslides, thus informing the humanitarian response on the ground.

Could these examples indicate that AI can lead to more effective responses in the humanitarian context? Depending on their design and deployment, AI systems may support humanitarian responses to conflict and crisis. However, much is context-dependent.

Using AI technologies to map areas affected by disasters seems to yield satisfactory results. For instance, the Humanitarian OpenStreetMap project relies on AI systems capable of mapping areas affected by disasters.[46] This project uses crowdsourced social media data and satellite and drone imagery to provide reliable information about which areas are affected by disaster situations and need prioritization. However, such a project might not produce relevant results in the context of humanitarian responses in situations of armed conflict. For instance, disinformation campaigns may affect access to trustworthy data

---

41 Swati Padhee, Tanay Kumar Saha, Joel Tetreault and Alejandro Jaimes, "Clustering of Social Media Messages for Humanitarian Aid Response during Crisis", 2020, available at: https://arxiv.org/pdf/2007.11756.pdf; Firoj Alam, Ferda Ofli, Muhammad Imran, Tanvirul Alam and Umair Qazi, "Deep Learning Benchmarks and Datasets for Social Media Image Classification for Disaster Response", *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2020.

42 See the Emergency Situation Awareness website, available at: https://esa.csiro.au/aus/about-public.html.

43 See the Artificial Intelligence for Disaster Response website, available at: http://aidr.qcri.org/.

44 Declan Butler, "Crowdsourcing Goes Mainstream in Typhoon Response", *Nature*, 2013, available at: www.nature.com/articles/nature.2013.14186; Wenjuan Sun, Paolo Bocchini and Brian D. Davison, "Applications of Artificial Intelligence for Disaster Management", *Nature Hazards*, Vol. 103, No. 3, 2020.

45 Felicia Vacarelu and Joseph Aylett-Bullock, "Fusing AI into Satellite Image Analysis to Inform Rapid Response to Floods", United Nations Institute for Training and Research, 2021, available at: https://unitar.org/about/news-stories/news/fusing-ai-satellite-image-analysis-inform-rapid-response-floods.

46 See the Humanitarian OpenStreetMap website, available at: www.hotosm.org.

during armed conflicts.[47] More generally, problems with access to good-quality data, which can be scarce during armed conflict situations, might affect the design and development of AI systems in that context and thereby compromise the suitability of their mapping tools.

Accordingly, while AI technologies may present opportunities to support effective humanitarian relief responses, they should not be understood as a ready-made, "one-size-fits-all" solution for any context within the realm of humanitarian action.

Recovery

AI may be effectively used in the context of recovery, as the complexities of contemporary crises often lead to protracted conflict situations.[48] Information technology can be an additional asset for facilitating engagement between humanitarians and affected communities in such contexts.[49]

AI technologies may support humanitarian action in protracted situations. For example, the Trace the Face tool developed by the International Committee of the Red Cross (ICRC) was designed to help refugees and migrants find missing family members.[50] This tool uses facial recognition technologies to automate searching and matching, thus streamlining the process. Another example can be found in the AI-powered chatbots that may provide a way for affected community members to access humanitarian organizations and obtain relevant information. These chatbots are currently providing advisory services to migrants and refugees.[51] Similarly, humanitarian organizations may use messaging chatbots to connect with affected populations.[52]

---

47 ICRC, *Harmful Information. Misinformation, Disinformation and Hate Speech in Armed Conflict and Other Situations of Violence*, Geneva, 2021, available at: https://shop.icrc.org/harmful-information-misinformation-disinformation-and-hate-speech-in-armed-conflict-and-other-situations-of-violence-icrc-initial-findings-and-perspectives-on-adapting-protection-approaches-pdf-en.html.

48 Edwin Odhiambo Abuya, "From Here to Where? Refugees Living in Protracted Situations in Africa", in Alice Edwards and Carla Ferstman (eds), *Human Security and Non-Citizens: Law, Policy and International Affairs*, Cambridge University Press, Cambridge, 2010; ICRC, *Protracted Conflict and Humanitarian Action: Some Recent ICRC Experiences*, Geneva, 2016, pp. 9–11; Ellen Policinski and Jovana Kuzmanovic, "Editorial: Protracted Conflicts: The Enduring Legacy of Endless War", *International Review of the Red Cross*, Vol. 101, No. 912, 2019.

49 Mirca Madianou, Liezel Longboan and Jonathan Corpus Ong, "Finding a Voice through Humanitarian Technologies? Communication Technologies and Participation in Disaster Recovery", *International Journal of Communication*, Vol. 9, 2015; ICRC, above note 48, pp. 15, 37.

50 ICRC, "Rewards and Risks in Humanitarian AI: An Example", *Inspired: Innovation to Save Lives and Defend Dignity*, 2019, available at: https://blogs.icrc.org/inspired/2019/09/06/humanitarian-artificial-intelligence/.

51 Ana Beduschi and Marie McAuliffe, "AI, Migration and Mobility: Implications for Policy and Practice", in Marie McAuliffe and Anna Triandafyllidou (eds), *World Migration Report 2022*, International Organization for Migration, Geneva, 2021; Marie McAuliffe, Jenna Blower and Ana Beduschi, "Digitalization and Artificial Intelligence in Migration and Mobility: Transnational Implications of the COVID-19 Pandemic", *Societies*, Vol. 11, No. 4, 2021.

52 ICRC, The Engine Room and Block Party, *Humanitarian Futures for Messaging Apps*, Geneva, 2017, available at: www.icrc.org/en/publication/humanitarian-futures-messaging-apps; Joanna Misiura and Andrej Verity, *Chatbots in the Humanitarian Field: Concepts, Uses and Shortfalls*, Digital Humanitarian

However, it is vital to question whether it is possible to generalize from these examples that AI contributes to better recovery action. As noted earlier in the analysis of preparedness and response, the benefit of using AI depends very much on the specific context in which these technologies are deployed. This is also true for recovery action. Community engagement and people-centred approaches may support the identification of areas in which technologies may effectively support recovery efforts on the ground, or conversely, those in which AI systems would not add value to recovery efforts. This should inform decision-making concerning the use of AI systems in recovery programmes. Moreover, AI technologies may also pose considerable risks for affected populations, such as exacerbating disproportionate surveillance or perpetuating inequalities due to algorithmic biases. Such risks are analysed in the following section.

AI at the expense of humanitarianism: The risks for affected populations

While AI may lead to potentially valuable outcomes in the humanitarian sector, deploying these systems is not without risks. Three main areas are of particular relevance in the context of humanitarian action: data quality, algorithmic bias, and the respect and protection of data privacy.

Data quality

Concerns about the quality of the data used to train AI algorithms are not limited to the humanitarian field, but this issue can have significant consequences for humanitarian action. In general terms, poor data quality leads to equally poor outcomes.[53] Such is the case, for instance, in the context of predictive policing and risk assessment algorithms. These algorithms often draw from historical crime data, such as police arrest rates per postcode and criminal records, to predict future crime incidence and recidivism risk.[54] If the data used to train these algorithms is incomplete or contains errors, the outcomes of the algorithms (i.e., crime forecasts and recidivism risk scores) might be equally poor in quality. Studies have indeed found that historical crime data sets may be incomplete and may include errors, as racial bias is often present in police records in some

---

Network, Geneva, 2019, available at: www.digitalhumanitarians.com/chatbots-in-the-humanitarian-field-concepts-uses-and-shortfalls/.

53 Thomas Redman, "If Your Data Is Bad, Your Machine Learning Tools Are Useless", *Harvard Business Review*, 2 April 2018, available at: https://hbr.org/2018/04/if-your-data-is-bad-your-machine-learning-tools-are-useless; R. Richardson, J. Schultz and K. Crawford, above note 38.

54 Andrew Ferguson, *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*, New York University Press, New York, 2017; Sarah Brayne, *Predict and Surveil: Data, Discretion, and the Future of Policing*, Oxford University Press, Oxford, 2020; R. Richardson, J. Schultz and K. Crawford, above note 38.

jurisdictions such as the United States.[55] If such algorithms are used to support judicial decision-making, it can lead to unfairness and discrimination based on race.[56]

In the humanitarian context, poor data quality generates poor outcomes that may directly affect populations in an already vulnerable situation due to conflicts or crises. AI systems trained with inaccurate, incomplete or biased data will likely perpetuate and cascade these mistakes forward. For instance, a recent study found that ten of the most commonly used computer vision, natural language and audio data sets contain significant labelling errors (i.e., incorrect identification of images, text or audio).[57] As these data sets are often used to train AI algorithms, the errors will persist in the resulting AI systems.

Unfortunately, obtaining high-quality data for humanitarian operations can be difficult due to the manifold constraints on such operations.[58] For instance, humanitarians may have problems collecting data due to low internet connectivity in remote areas. Incomplete and overlapping datasets that contain information collected by different humanitarian actors may also be a problem – for example, inaccuracies can be carried forward if outdated information is maintained in the data sets.[59] Errors and inaccuracies can also occur when using big data and crowdsourced data.[60] Accordingly, it is crucial that teams working with these data sets control for errors as much as possible. However, data sets and AI systems may also suffer from algorithmic bias, a topic that relates to data quality but has larger societal implications and is thus discussed in the following subsection.

Algorithmic bias

Connected to the issue of data quality is the question of the presence of bias in the design and development of AI systems. Bias is considered here not only as a technological or statistical error, but also as the human viewpoints, prejudices

---

55 S. Brayne, above note 54, pp. 33–34, 105; A. Ferguson, above note 54, p. 23; C. Dominik Güss, Ma Teresa Tuason and Alicia Devine, "Problems with Police Reports as Data Sources: A Researchers' Perspective", *Frontiers in Psychology*, Vol. 11, 2020.

56 See, notably, Sonja B. Starr, "Evidence-Based Sentencing and the Scientific Rationalization of Discrimination", *Stanford Law Review*, Vol. 66, No. 4, 2014; Supreme Court of Wisconsin, *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), 2016, p. 769 (requiring a warning prior to the use of algorithm risk assessment in sentencing and establishing that risk scores may not be used "to determine whether an offender is incarcerated" or "to determine the severity of the sentence").

57 Examples can be explored in Curtis G. Northcutt, Anish Athalye and Jonas Mueller, "Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks", 35th Conference on Neural Information Processing Systems, 2021, available at: https://arxiv.org/pdf/2103.14749.pdf.

58 Christopher Kuner and Massimo Marelli, *Handbook on Data Protection in Humanitarian Action*, 2nd ed., ICRC, Geneva, 2020, p. 39; OCHA, above note 15, p. 10; ICRC, The Engine Room and Block Party, above note 52, p. 32.

59 Anne Singleton, *Migration and Asylum Data for Policy-Making in the European Union: The Problem with Numbers*, CEPS Papers in Liberty and Security in Europe No. 89, 2016, available at: www.ceps.eu/ceps-publications/migration-and-asylum-data-policy-making-european-union-problem-numbers/; European Union Agency for Fundamental Rights, *Data Quality and Artificial Intelligence: Mitigating Bias and Error to Protect Fundamental Rights*, Vienna, 2019.

60 B. Haworth and E. Bruce, above note 13; P. Sharma and A. Joshi, above note 13.

and stereotypes that are reflected in AI systems and can lead to unfair outcomes and discrimination.[61] AI systems can indeed reflect the biases of their human designers and developers.[62] Once such systems are deployed, this can in turn lead to unlawful discrimination.

International human rights law prohibits direct and indirect forms of discrimination based on race, colour, sex, gender, sexual orientation, language, religion, political or other opinion, national or social origin, property, birth or other status.[63] Direct discrimination takes place when an individual is treated less favourably on the basis of one or more of these grounds. Indirect discrimination exists even when measures are in appearance neutral, as such measures can in fact lead to the less favourable treatment of individuals based on one or more of the protected grounds.

Bias in AI systems may exacerbate inequalities and perpetuate direct and indirect forms of discrimination, notably on the grounds of gender and race.[64] For instance, structural and historical bias against minorities may be reflected in AI systems due to the pervasive nature of these biases.[65] Bias also commonly arises from gaps in the representation of diverse populations in data sets used for training AI algorithms.[66] For example, researchers have demonstrated that commercially available facial recognition algorithms were less accurate in recognizing women with darker skin types due in part to a lack of diversity in training data sets.[67] Similarly, researchers have shown that AI algorithms had

61 Kate Crawford, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, New York, 2021, pp. 133–135.

62 Batya Friedman and Helen Nissenbaum, "Bias in Computer Systems", *ACM Transactions on Information Systems*, Vol. 14, No. 3, 1996; James Zou and Londa Schiebinger, "AI Can Be Sexist and Racist — It's Time to Make It Fair", *Nature*, Vol. 559, 2018; Harini Suresh and John V. Guttag, "A Framework for Understanding Unintended Consequences of Machine Learning", 2020, available at: https://arxiv.org/pdf/1901.10002.pdf.

63 Universal Declaration of Human Rights (UDHR), 10 December 1948, Arts 2, 7; International Covenant on Civil and Political Rights (ICCPR), 16 December 1966, Art. 26; European Convention on Human Rights (ECHR), 4 November 1950, Art. 14; American Convention on Human Rights (ACHR), 22 November 1969, Art. 1; African Charter on Human and Peoples' Rights, 27 June 1981, Art. 2. See also Rachel Murray and Frans Viljoen, "Towards Non-Discrimination on the Basis of Sexual Orientation: The Normative Basis and Procedural Possibilities before the African Commission on Human and Peoples' Rights and the African Union", *Human Rights Quarterly*, Vol. 29, No. 1, 2007; Human Rights Council, *Report of the United Nations High Commissioner for Human Rights on Discriminatory Laws and Practices and Acts of Violence against Individuals Based on Their Sexual Orientation and Gender Identity*, UN Doc. A/HRC/19/41, 17 November 2011; Human Rights Council, *Protection against Violence and Discrimination Based on Sexual Orientation and Gender Identity*, UN Doc. A/HRC/RES/32/2, 15 July 2016.

64 Noel Sharkey, "The Impact of Gender and Race Bias in AI", *Humanitarian Law and Policy Blog*, 28 August 2018, available at: https://blogs.icrc.org/law-and-policy/2018/08/28/impact-gender-race-bias-ai/; UN Secretary-General's High-Level Panel on Digital Cooperation, *The Age of Digital Interdependence*, New York, 2019, available at: www.un.org/en/pdfs/DigitalCooperation-report-for%20web.pdf; UN General Assembly, *Report of the Special Rapporteur Tendayi Achiume on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*, UN Doc A/75/590, 10 November 2020.

65 H. Suresh and J. V. Guttag, above note 62.

66 J. Zou and L. Schiebinger, above note 62.

67 Joy Buolamwini and Timmit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification", *Proceedings of Machine Learning Research: Conference on Fairness, Accountability*

more difficulties identifying people with disabilities when such individuals were using assistive technologies such as wheelchairs.[68]

In this regard, biased AI systems may go undetected and continue supporting decisions that could lead to discriminatory outcomes.[69] That is partly due to the opacity with which certain machine learning and deep learning algorithms operate – the so-called "black box problem".[70] In addition, the complexity of AI systems based on deep learning techniques entails that their designers and developers are often unable to understand and sufficiently explain how the machines have reached certain decisions. This may in turn make it more challenging to identify biases in the algorithms.

The consequences of deploying biased AI systems can be significant in the humanitarian context. For example, in a scenario where facial recognition technologies are the sole means for identification and identity verification, inaccuracies in such systems may lead to the misidentification of individuals with darker skin types. If identification and identity verification by those means is a precondition for accessing humanitarian aid, misidentification may lead to individuals being denied assistance. This could happen if the system used for triage mistakenly indicates that an individual has already received the aid in question (such as emergency food supplies or medical care). Such a situation would have dramatic consequences for the affected individuals. If the AI systems' risks were known and not addressed, it could lead to unlawful discrimination based on race. This could also be contrary to the humanitarian principle of humanity, according to which human suffering must be addressed wherever it is found.[71]

Accordingly, safeguards must be put in place to ensure that AI systems used to support the work of humanitarians are not transformed into tools of exclusion of individuals or populations in need of assistance. For example, if online photographs of children in war tend to show children of colour with weapons (i.e., as child soldiers) disproportionately more often, while depicting children of white ethnic background as victims, then AI algorithms trained on such data sets may continue to perpetuate this distinction. This could in turn contribute to existing biases against children of colour in humanitarian action, compounding the suffering already inflicted by armed conflict. Awareness and control for this type of bias should therefore permeate the design and development of AI systems to be deployed in the humanitarian context. Another example relates to facial

and Transparency, Vol. 81, 2018. But see Stewart Baker, "The Flawed Claims about Bias in Facial Recognition", Lawfare, 2 February 2022, available at: www.lawfareblog.com/flawed-claims-about-bias-facial-recognition.

68   Meredith Whittaker et al., Disability, Bias and AI, AI Now Institute, New York University, 2019, pp. 9–10.
69   M. Pizzi, M. Romanoff and T. Engelhardt, above note 2.
70   The black box problem occurs when AI systems' operations are not be visible to users and third parties. See Frank Pasquale, The Black Box Society: The Secret Algorithms that Control Money and Information, Harvard University Press, Cambridge, MA, 2016.
71   UNGA Res. 46/182, 19 December 1991; Statutes of the International Red Cross and Red Crescent Movement, adopted by the 25th International Conference of the Red Cross, Geneva, 1986 (amended 1995, 2006), preamble.

recognition technologies – as long as these technologies remain inaccurate in recognizing people with darker skin types, they should not be used to assist decision-making essential to determining humanitarian aid delivery.

Data privacy

As is internationally agreed, "the same rights that people have offline must also be protected online".[72] This should include AI systems.

International human rights law instruments recognize the right to privacy.[73] In addition, specific legal regimes, such as the General Data Protection Regulation (GDPR), establish fundamental standards for protecting personal data. While the GDPR is a European Union (EU) law regime that does not bind all humanitarian actors across the globe, it remains relevant beyond the EU as it has inspired similar regulations worldwide.[74]

The principles set forth in the GDPR have also been taken into account by the *Handbook on Data Protection in Humanitarian Action,*[75] which is considered a leading resource that sets a minimum standard for processing personal data in the humanitarian context. These principles include lawfulness, fairness and transparency in the processing of personal data (Article 5 of the GDPR).

Having a lawful basis for the processing of personal data is a legal requirement (Article 6 of the GDPR). Consent is often used as a lawful basis for processing personal data in the humanitarian context. According to the legal standards, consent must be fully informed, specific, unambiguous and freely given (Article 4(11) of the GDPR). Yet, in the humanitarian context, consent may not be entirely unambiguous and freely given due to the inherent power imbalance between humanitarian organizations and beneficiaries of humanitarian assistance. A refusal to consent to collecting and processing personal data may, in practical terms, lead to the denial of humanitarian assistance.[76] However, it may be difficult for humanitarian actors to ensure that recipients of humanitarian assistance effectively understand the meaning of consent due to linguistic barriers and administrative and institutional complexities.

Fully informed, specific, unambiguous and freely given consent may also be challenging to achieve given that AI systems often use data to further refine and

---

72 UNGA Res. 68/167, 21 January 2014, para. 2; Human Rights Council, *The Promotion, Protection and Enjoyment of Human Rights on the Internet*, UN Doc A/HRC/20/L.13, 29 June 2012, para. 1.

73 UDHR, Art. 12; ICCPR, Art. 17; ECHR, Art. 8; ACHR, Art. 11.

74 According to Article 45 of the GDPR, the European Commission can issue an adequacy decision recognizing that a third country's domestic law offers an adequate level of data protection that is essentially equivalent to the GDPR. The consequence of such a decision is that data flows can continue without the need for further safeguards. To date, the European Commission has issued adequacy decisions regarding Andorra, Argentina, Canada (commercial organizations), the Faroe Islands, Guernsey, Israel, the Isle of Man, Japan, Jersey, New Zealand, the Republic of Korea, Switzerland, the United Kingdom and Uruguay. See European Commission, "Adequacy Decisions", available at: https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en.

75 C. Kuner and M. Marelli, above note 58, p. 23.

76 M. Madianou, above note 21, p. 9; M. Pizzi, M. Romanoff and T. Engelhardt, above note 2, p. 152.

develop other AI solutions. While individuals may agree to have their personal information processed for a specific purpose related to humanitarian action, they may not know about or agree to that data being later used to develop other AI systems.[77] Such concerns are further aggravated by the criticisms concerning "surveillance humanitarianism", whereby the growing collection of data and uses of technologies by humanitarians may inadvertently increase the vulnerability of those in need of assistance.[78]

These practices require even more scrutiny due to the increasingly common collaborations between technology companies and humanitarian organizations.[79] These companies play a central role in this area as they design and develop the AI systems that humanitarians later deploy in the field. Arguably, technology companies' interests and world view tend to be predominantly reflected in the design and development of AI systems, thus neglecting the needs and experiences of their users.[80] This is particularly concerning for the deployment of AI systems in the humanitarian context, where the risks for populations affected by conflicts or crises are significant. Accordingly, it is essential to have a clear set of guidelines for implementing AI in the humanitarian context, notably placing the humanitarian imperative of "do no harm" at its core, as discussed in the following section.

AI at the service of humanitarian action: The humanitarian imperative of "do no harm"

As noted earlier, while AI may bring about novel opportunities to strengthen humanitarian action, it also presents significant risks when deployed in the humanitarian context. This section elaborates on the humanitarian imperative of "do no harm" and offers recommendations on making AI work in support of humanitarian action and not to the detriment of populations affected by conflict and crisis.

"Do no harm" in the age of AI

In the face of ever-evolving AI technologies, it is crucial that humanitarians consider the imperative of "do no harm" as paramount to all deployment of AI systems in

---

77 C. Kuner and M. Marelli, above note 58, p. 284; Meg Leta Jones and Elizabeth Edenberg, "Troubleshooting AI and Consent", in Markus D. Dubber, Frank Pasquale and Sunit Das (eds), *The Oxford Handbook of Ethics of AI*, Oxford University Press, Oxford, 2020, p. 366.

78 M. Latonero, above note 19; P. Molnar, above note 20; Pierrick Devidal, "Cashless Cash: Financial Inclusion or Surveillance Humanitarianism?", *Humanitarian Law and Policy Blog*, 2 March 2021, available at: https://blogs.icrc.org/law-and-policy/2021/03/02/cashless-cash/.

79 M. Pizzi, M. Romanoff and T. Engelhardt, above note 2; Linda Kinstler, "Big Tech Firms Are Racing to Track Climate Refugees", *MIT Technology Review*, 17 May 2019, available at: www.technologyreview.com/2019/05/17/103059/big-tech-firms-are-racing-to-track-climate-refugees/.

80 Ziv Carmon, Rom Schrift, Klaus Wertenbroch and Haiyang Yang, "Designing AI Systems that Customers Won't Hate", *MIT Sloan Management Review*, 16 December 2019, available at: https://sloanreview.mit.edu/article/designing-ai-systems-that-customers-wont-hate/.

humanitarian action. This principle of non-maleficence has long been recognized as one of the core principles of bioethics.[81] It was first proposed in the humanitarian context by Mary Anderson;[82] subsequently, various humanitarian organizations have further developed its application.[83] Today, this principle is also commonly invoked in the fields of ethics of technology and AI.[84]

The "do no harm" principle entails that humanitarian actors consider the potential ways in which their actions or omissions may inadvertently cause harm or create new risks for the populations they intend to serve.[85] For example, humanitarian "innovation" may introduce unnecessary risks to already vulnerable populations, such as when technical failures in newly introduced systems lead to delays, disruption or cancellation of aid distribution.[86] Therefore, avoiding or preventing harm and mitigating risks is at the heart of this humanitarian imperative.

Risk analysis and impact assessments may be used to operationalize the "do no harm" principle. Risk analysis can help to identify potential risks arising from humanitarian action and provide a clear avenue for risk mitigation. Impact assessments can provide the means to identify the negative impacts of specific humanitarian programmes and the best ways to avoid or prevent harm. These processes may assist humanitarian organizations as they envisage the utilization of AI technologies for humanitarian action. At times, they may even lead to the conclusion that no technologies should be deployed in a specific context, as these would cause more harm than good to their beneficiaries. On certain occasions, the fact that a technology is available does not mean that it must also be used.

AI technologies present some well-known risks, which ought to be addressed by humanitarian actors before the deployment of AI systems in humanitarian action. For example, humanitarian organizations using data-driven AI systems should identify risks concerning data security breaches that could lead to the disclosure of sensitive information about their staff and their beneficiaries. They should also evaluate whether using AI systems would negatively impact affected populations – for example, by revealing their location while mapping the evolution of a conflict and thereby inadvertently exposing them to persecution. In

---

81 Tom L. Beauchamp and James F. Childress, *Principles of Biomedical Ethics*, 8th ed., Oxford University Press, Oxford, 2019; Luciano Floridi and Josh Cowls, "A Unified Framework of Five Principles for AI in Society", *Harvard Data Science Review*, Vol. 1, No. 1, 2019.

82 Mary B. Anderson, *Do No Harm: How Aid Can Support Peace or War*, Lynne Rienner, Boulder, CO, 1999; Mary B. Anderson, *Options for Aid in Conflict: Lessons from Field Experience*, CDA Collaborative Learning Projects, Cambridge, MA, 2000.

83 ICRC, "ICRC Protection Policy", *International Review of the Red Cross*, Vol. 90, No. 871, 2008, p. 753; Sphere Association, *The Sphere Handbook: Humanitarian Charter and Minimum Standards in Humanitarian Response*, 4th ed., Geneva, 2018.

84 L. Floridi and J. Cowls, above note 81; Luciano Floridi, *The Ethics of Information*, Oxford University Press, Oxford, 2013; C. Kuner and M. Marelli, above note 58.

85 Sphere Association, above note 83, p. 268.

86 Kristin Bergtora Sandvik, Katja Lindskov Jacobsen and Sean Martin McDonald, "Do No Harm: A Taxonomy of the Challenges of Humanitarian Experimentation", *International Review of the Red Cross*, Vol. 99, No. 1, 2017.

sum, the deployment of AI systems should never create additional harm or risks to affected populations.

Accordingly, humanitarian actors must not over-rely on AI technologies, particularly those that remain insufficiently accurate in certain contexts, such as facial recognition technologies.[87] Before adopting AI systems, humanitarian actors should evaluate whether there is a need to deploy these technologies in the field, whether they add value to the humanitarian programmes in question, and whether they can do so in a manner that protects vulnerable populations from additional harm.

Mechanisms for avoiding and mitigating data privacy harms

In the digital age, avoiding or mitigating harm also entails the protection of data privacy. Data privacy should be protected and respected throughout the AI life cycle, from design to development to implementation.

In this regard, "privacy by design" principles provide a good starting point.[88] They offer a proactive (instead of reactive) and preventive (instead of remedial) set of principles based on user-centric approaches. These are valuable tools for building better data privacy protection.[89]

For humanitarian organizations that are subject to EU law, Article 25 of the GDPR imposes a more comprehensive requirement for data protection by design and by default.[90] This provision requires the implementation of appropriate technical and organizational measures aimed at integrating the core data protection principles (enumerated in Article 5 of the GDPR) into the design and development of systems processing personal data. As noted earlier, these core principles are lawfulness, fairness and transparency, along with purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality, and accountability. These are also consistent with the basic data protection principles proposed by the ICRC.[91]

---

87 Davide Castelvecchi, "Is Facial Recognition too Biased to Be Let Loose?", *Nature*, Vol. 587, 2020.
88 These principles were proposed by Ann Cavoukian in 2010, as she occupied the position of information and privacy commissioner of Ontario, Canada. See Ann Cavoukian, "Privacy by Design: The 7 Foundational Principles", Toronto, 2010, available at: www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf. These principles were later endorsed by the International Conference of Data Protection and Privacy Commissioners. See "Resolution on Privacy by Design", 32nd International Conference of Data Protection and Privacy Commissioners, Jerusalem, 27–29 October 2010, available at: http://globalprivacyassembly.org/wp-content/uploads/2015/02/32-Conference-Israel-resolution-on-Privacy-by-Design.pdf. See also Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers*, Washington, DC, 2012, available at: www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.
89 Lina Jasmontaite, Irene Kamara, Gabriela Zanfir-Fortuna and Stefano Leucci, "Data Protection by Design and by Default: Framing Guiding Principles into Legal Obligations in the GDPR", *European Data Protection Law Review*, Vol. 4, No. 2, 2018; Giovanni Buttarelli, *Opinion 5/2018: Preliminary Opinion on Privacy by Design*, 31 May 2018, available at: https://edps.europa.eu/sites/edp/files/publication/18-05-31_preliminary_opinion_on_privacy_by_design_en_0.pdf.
90 Lee Bygrave, "Data Protection by Design and by Default: Deciphering the EU's Legislative Requirements", *Oslo Law Review*, Vol. 4, No. 2, 2017.
91 C. Kuner and M. Marelli, above note 58.

Accordingly, humanitarian organizations designing AI solutions or procuring AI systems from private sector providers should ensure that data protection is implemented by design and by default in these AI systems. For instance, they should ensure that they have obtained consent for processing personal information or that they rely on another legal basis for processing, such as the vital interest of the data subject or of another person, public interest, legitimate interest, performance of a contract, or compliance with a legal obligation.[92] Similarly, data collection should be kept to the minimum needed, storage should be cyber-secure, personal data should be destroyed once it is no longer required, and personal information should only be used for the purpose for which it was collected in the first place.

Moreover, carrying out data protection impact assessments (DPIAs) may also help humanitarian actors understand the potential negative impacts of AI technologies used in humanitarian programmes. A DPIA is a process that identifies the risks for the protection of individuals' data privacy and the ways of mitigating those risks.[93] Humanitarian organizations subject to the GDPR will have an obligation to carry out a DPIA before processing data if there is a high risk of harm to individuals' rights and freedoms (Article 35(1) of the GDPR). DPIAs can add value to humanitarian projects, even if the organizations involved are not legally obliged to carry out such a process. A DPIA can help to provide a clear roadmap for identifying risks, solutions and recommendations concerning data-driven AI systems.

For example, a DPIA can be used to identify situations in which anonymized data used to train AI algorithms may be re-identified, thus becoming personal information again and attracting the application of legal regimes on data protection. Re-identification occurs when data that was initially anonymized is de-anonymized. This can happen when information from different sources is matched to identify individuals from an initially anonymized data set. For instance, a study found that it was possible to match information in order to identify individuals from a list containing the anonymous movie ratings of 500,000 Netflix subscribers, also uncovering their apparent political preferences and other potentially sensitive information.[94] Overall, research demonstrates that individuals have an over 99% chance of being re-identified in certain circumstances, even when data sets were initially anonymized.[95]

In the humanitarian context, anonymization may not be enough to prevent the re-identification of vulnerable populations, and failure to retain information in a cyber-secure manner risks exposing such populations to persecution and harm.

---

92 *Ibid.*, p. 60.
93 *Ibid.*, p. 84.
94 Arvind Narayanan and Vitaly Shmatikov, "Robust De-anonymization of Large Sparse Datasets", *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, 18–22 May 2008.
95 Luc Rocher, Julien M. Hendrickx and Yves-Alexandre de Montjoye, "Estimating the Success of Re-identifications in Incomplete Datasets Using Generative Models", *Nature Communications*, Vol. 10, No. 1, 2019.

A DPIA can help identify other solutions and organizational measures that could prevent re-identification from occurring.

Transparency, accountability and redress

The principle of "do no harm" also implies that humanitarian actors should consider establishing an overarching framework to ensure much-needed transparency and accountability on the uses of AI in humanitarian action.

The term "transparency" is used here to indicate that humanitarian actors should communicate about whether and how they use AI systems in humanitarian action. They should disclose information about the systems they use, even when the way in which these systems work is not fully explainable. In this sense, transparency is a broader concept than the narrower notion of explainability of AI systems.[96]

For example, consider a scenario in which AI systems are used for biometric identity verification of refugees as a condition for distributing aid in refugee camps.[97] In this case, the humanitarian actors using such AI systems should communicate to the refugees that they are doing so. It is equally important that they disclose to those refugees how they are employing the AI systems and what it entails. For instance, they should disclose what type of information will be collected and for what purpose, how long the data will be stored, and who will access it. Similarly, they should communicate which safeguards will be put in place to avoid cyber security breaches.

Accountability is understood as the action of holding someone to account for their actions or omissions.[98] It is a process aimed at assessing whether a person's or an entity's actions or omissions were required or justified and whether that person or entity may be legally responsible or liable for the consequences of their act or omission.[99] Accountability is also a mechanism involving an obligation to explain and justify conduct.[100]

In the humanitarian context, accountability should be enshrined in the relationships between humanitarian actors and their beneficiaries – particularly when AI systems are used to support humanitarian action, due to the risks these technologies may pose to their human rights. For instance, humanitarian actors should inform their beneficiaries of any data security breach that may expose the

---

96 Stefan Larsson and Fredrik Heintz, "Transparency in Artificial Intelligence", *Internet Policy Review*, Vol. 9, No. 2, 2020.

97 Biometrics refers to "the application to biology of the modern methods of statistics" and relates to biometric characteristics or the "biological and behavioural characteristic[s] of an individual from which distinguishing, repeatable biometric features can be extracted for the purpose of biometric recognition", such as fingerprints, iris patterns and facial features. International Organization for Standardization, "Information Technology – Biometrics – Overview and Application", ISO/IEC TR 24741:2018, 2018, available at: www.iso.org/obp/ui/#iso:std:iso-iec:tr:24741:ed-2:v1:en.

98 Richard Mulgan, "'Accountability': An Ever Expanding Concept?", *Public Administration*, Vol. 78, No. 3, 2000.

99 Ivo Giesen and François G. H. Kristen, "Liability, Responsibility and Accountability: Crossing Borders", *Utrecht Law Review*, Vol. 10, No. 3, 2014, p. 6.

100 Mark Bovens, "Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism", *West European Politics*, Vol. 33, No. 5, 2010, p. 951.

beneficiaries' personal information and give an account of the measures taken to remedy the situation. The recent swift response by the ICRC to a data security breach has set an example of good practice in this area. The institution undertook direct and comprehensive efforts to explain the actions taken and inform the affected communities worldwide of the consequences of the cyber security incident.[101]

Finally, individuals should be able to challenge decisions that were either automated or made by humans with the support of AI systems if such decisions adversely impacted those individuals' rights.[102] Grievance mechanisms, either judicial or extra-judicial, could thus provide legal avenues for access to remedy, notably in cases where inadvertent harm was caused to the beneficiaries of humanitarian assistance. Extra-judicial mechanisms such as administrative complaints or alternative dispute resolution could be helpful to individuals who may not be able to afford the costs of judicial proceedings.

Conclusion

Data-driven AI technologies are progressively transforming the humanitarian field. They have the potential to support humanitarian actors as they implement a paradigm shift from reactive to anticipatory approaches to humanitarian action. AI may thus contribute to humanitarian action in its three main dimensions: preparedness, response and recovery.

AI technologies can support humanitarian preparedness. They can do so by analyzing vast amounts of multidimensional data at fast speeds, identifying patterns in the data, making inferences, and providing crucial insights about potential risks before a crisis or humanitarian disaster unfolds. AI technologies can also present opportunities to support effective humanitarian relief responses and promote recovery programmes, notably in protracted conflict situations.

Several AI-based initiatives are currently being deployed and tested by humanitarian organizations. These include AI systems deployed to forecast population movements, map areas affected by humanitarian crises and identify missing individuals, thus informing and facilitating humanitarian action on the ground. Yet, deploying these systems is not without risks. This article has analyzed three main areas of concern: the quality of the data used to train AI algorithms, the existence of algorithmic bias permeating the design and development of AI systems, and the respect for and protection of data privacy.

While these concerns are not exclusive to the humanitarian field, they may significantly affect populations already in a vulnerable situation due to conflict and crisis. Therefore, if AI systems are not to be deployed at the expense of

---

101 ICRC, "Cyber Security Incident: How Could It Affect Me?", 7 February 2022, available at: www.icrc.org/en/document/cyber-security-how-it-affect-me; ICRC, "ICRC Cyber-Attack: Sharing our Analysis", 16 February 2022, available at: www.icrc.org/en/document/icrc-cyber-attack-analysis.
102 M. Pizzi, M. Romanoff and T. Engelhardt, above note 2, p. 179.

humanitarianism, it is vital that humanitarian actors implement these technologies in line with the humanitarian imperative of "do no harm". Risk analysis and impact assessments may help to operationalize the "do no harm" imperative. Both processes may be valuable for mitigating risks and minimizing or avoiding negative impacts on affected populations.

The "do no harm" imperative is especially crucial in situations of armed conflict such as the one currently ravaging Ukraine and prompting the displacement of over 4 million people in Europe.[103] In such contexts, AI technologies can be used in both helpful and damaging ways within and outside the battlefield. For instance, AI can support the analysis of social media data and evaluate the veracity of information,[104] but it can also support the creation of false videos using deepfake technologies, fuelling disinformation campaigns.[105]

As AI systems are not inherently neutral, depending on how they are used, they may introduce new, unnecessary risks to already vulnerable populations. For instance, AI-powered chatbots can help streamline visa applications in the face of large movements of people fleeing conflict,[106] but if these systems are used without proper oversight, they could expose individuals' personal information to needless cyber security risks and potential data breaches. Accordingly, to put AI at the service of humanitarian action, leveraging its benefits while outweighing its risks, humanitarian organizations should be mindful that there is no ready-made, "one-size-fits-all" AI solution applicable to all contexts. They should also evaluate whether AI systems should be deployed at all in certain circumstances, as such systems could cause more harm than good to their beneficiaries. On certain occasions, the fact that technology is available does not mean that it must be used.

Finally, when deploying these technologies, it is crucial that humanitarian organizations establish adequate frameworks to strengthen accountability and transparency in the use of AI in the humanitarian context. Overall, such mechanisms would contribute towards the goal of harnessing the potential of responsible use of AI in humanitarian action.

---

103 See International Organization for Migration, "IOM Ukraine Situation Reports", available at: www.iom.
   int/resources/iom-ukraine-situation-reports.
104 Craig Nazareth, "Technology Is Revolutionizing How Intelligence Is Gathered and Analyzed – and
   Opening a Window onto Russian Military Activity around Ukraine", *The Conversation*, 14 February
   2022, available at: https://theconversation.com/technology-is-revolutionizing-how-intelligence-is-gathered-
   and-analyzed-and-opening-a-window-onto-russian-military-activity-around-ukraine-176446.
105 Hitoshi Nasu, "Deepfake Technology in the Age of Information Warfare", *Articles of War*, 1 March 2022,
   available at: https://lieber.westpoint.edu/deepfake-technology-age-information-warfare/.
106 A. Beduschi and M. McAuliffe, above note 51.