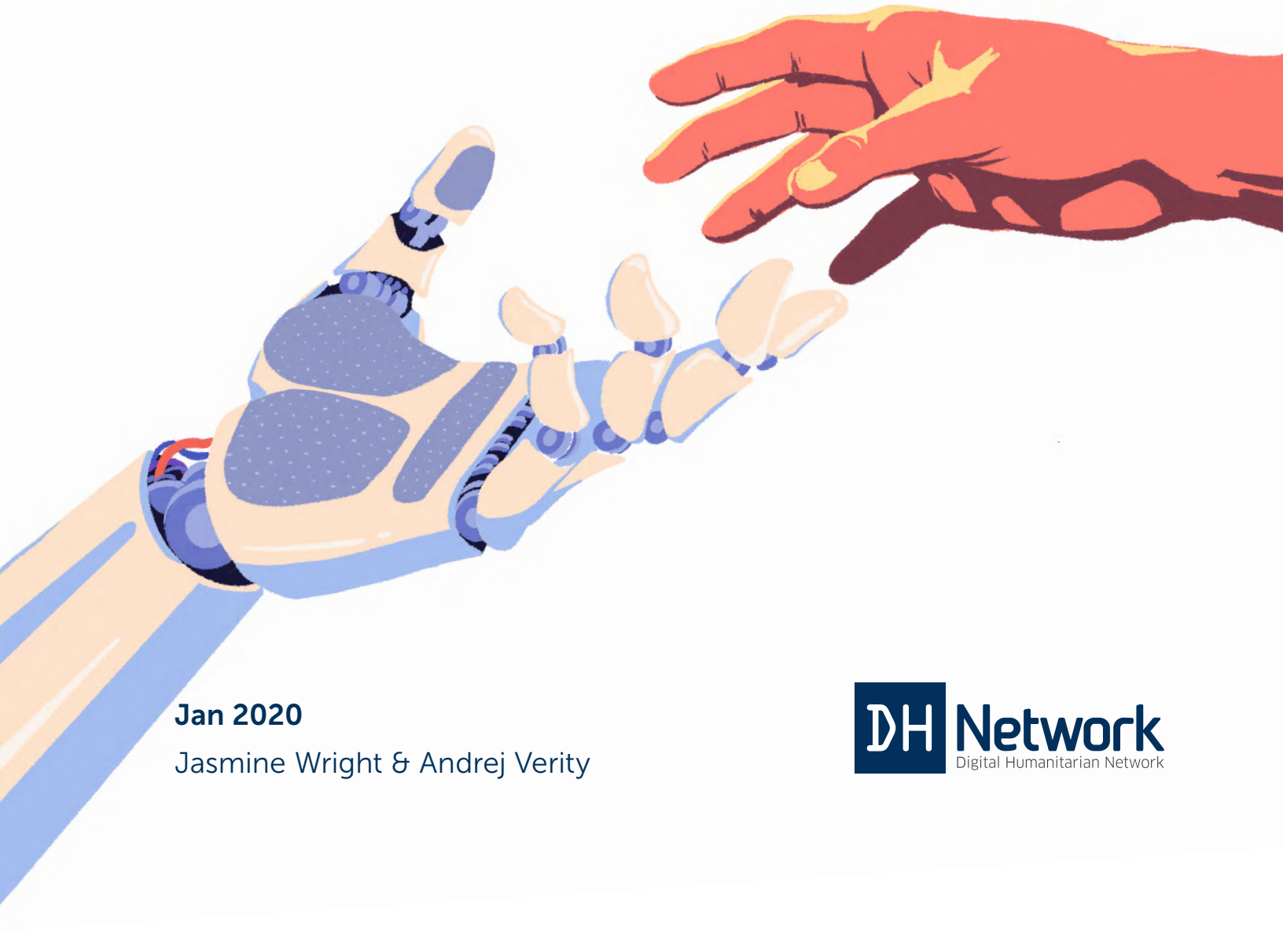
Artificial Intelligence Principles

For Vulnerable Populations in Humanitarian Contexts



Jan 2020

Jasmine Wright & Andrej Verity

DH Network
Digital Humanitarian Network

Licensing Information

Artificial Intelligence Principles

For Vulnerable Populations in Humanitarian Contexts

by

Jasmine Wright (jasmine.beverly.wright@gmail.com)
Master of Global Affairs, Munk School of Global Affairs and Public Policy
University of Toronto

Andrej Verity (verity@un.org | @andrejverity)
Office for the Coordination of Humanitarian Affairs (OCHA)
United Nations

Design

Ignacio G. Rebollo (rebollo@reliefweb.int | igrebollo.com)
M.Des. Ontario College of Arts and Design University (OCAD U)
Office for the Coordination of Humanitarian Affairs (OCHA), United Nations

This document was made possible
with the support of UN-OCHA

OCHA

Contents

Introduction 03

Section One: Examples of AI Systems for Vulnerable Populations 06

Defining AI 06
Karim the Chatbot 07
Hala Systems: An Airstrike Early Warning System 09
Biometrics and Refugees 11

Section Two: Biases 12

Western Cultural Norms Embedded in System Design 12
Biometric Technologies 13
Incomplete Data Sets 15
Diverse Test Cases 16

Section Three: Security Risks 20

Data Breaches and Data Security Practices 20
Function Creep 22
Project Connect and UNICEF 24

Section Four: Issues with Data Consent **26**

Consent and its Relation to Security Risks 26
Meaningful Consent, Captured Consent, and Tradeoffs 27
Parallel AI and Human-based Systems 28
Consent as Meaningless 28

Section Five: AI Principles and Recommendations **32**

AI Principles 32
1. Weigh the benefits versus the risks: Avoid AI if possible 33
2. Use AI systems that are contextually-based 35
3. Empower and include local communities in AI initiatives 36
4. Implement algorithmic auditing systems 37
General AI Recommendations 39
Business Model AI Recommendations 40

Conclusion **42**

Executive Summary

There are many recent examples of Artificial Intelligence (AI) systems being used for vulnerable people in humanitarian and disaster response contexts, with serious ethical and security-related implications. In particular, vulnerable populations are put at further risk through biases inherently built into AI systems. There are security concerns regarding their personal information being exposed and even used for persecution purposes. Yet rarely do they have a choice when it comes to the consent of surrendering such information. Now, as AI adoption grows rapidly, this report aims to develop AI principles and recommendations that would be specific to vulnerable people in the humanitarian field.

This report argues that AI systems with prevalent biases, security risks, and consent issues can undermine the role of humanitarian actors in disaster contexts by leaving aid recipients at further risk of vulnerability. Extending the risk of vulnerability not only impacts vulnerable populations, but also indirectly affects the achievement of the 2030 Sustainable Development Agenda, the maintenance of the Charter of the United Nations, and international human rights laws.

The report proceeds in five sections. The first four detail current examples of AI systems that are used for vulnerable populations, analyze different forms of bias pertaining to AI systems, evaluate security risks, and examine issues of consent over data collection. Based on the points raised in these sections, the fifth section develops AI principles and recommendations specific to vulnerable people in the humanitarian field.

These principles are:

- Weigh the benefits versus the risks: Avoid AI if possible
- Use AI systems that are contextually-based
- Empower and include local communities in AI initiatives
- Implement algorithmic auditing systems

This section continues by making several general suggestions as well as recommendations to align business models with the design and development of more ethical AI systems.

This report is informed by scholarly and news articles, as well as 20 interviews with a diverse range of experts in the AI and humanitarian fields. They work in different branches of the United Nations, in academia, think tanks, non-governmental organizations, and private sector companies.

Introduction

It is March of 2016 in Jordan. X2AI, a Silicon Valley-based Artificial Intelligence (AI) startup, just started testing a psychotherapy chatbot for Syrian refugees in the Za'atari camp on a group of mostly male subjects. With a small supply of therapists available to help Syrians affected by the civil conflict, a psychotherapy chatbot fluent in Arabic would be deployable to anyone with a mobile phone. Though X2AI depicts its chatbot as "therapeutic assistants" rather than replacements for medical professionals, this form of AI would provide some semblance of mental health care to a vulnerable population.[1]

Three years later, it is August of 2019. The World Food Programme (WFP) and Yemen's Houthi rebels reached an agreement to restart food deliveries to Yemeni people after a majority of those helped by the WFP in Sana'a, the country's capital, had aid suspended since June.[2] Aid suspension was caused over the desire to implement a biometric registration system, a type of AI, to better monitor the supply chain process and reduce food diversion from vulnerable populations.[3] [4] The Houthis opposed biometric data collection, stating it is unlawful in Yemen for an international organization to control such data.[5]

These are both recent examples of AI systems being used for vulnerable people. Consequently, they have serious ethical and security-related implications for humanitarianism and disaster response. Vulnerable populations are put at risk through biases inherently built into AI systems. Security regarding their personal information being exposed and used for persecution is a serious concern. Furthermore, these populations often do not have a choice when it comes to the consent of surrendering such information if they want assistance.

---

1 Nick Romeo, "The Chatbot Will See You Now," ***The New Yorker***, December 25, 2016, https://www.newyorker.com/tech/annals-of-technology/the-chatbot-will-see-you-now?reload=true (accessed June 3, 2019).

2 Al Jazeera and News Agencies, "Yemen's Houthis, WFP reach deal to resume food relief," ***Al Jazeera***, August 4, 2019, https://www.aljazeera.com/news/middleeast/2019/08/yemen-houthis-wfp-reach-deal-resume-food-relief-190804133835009.html (accessed August 9, 2019).

3 Al Jazeera and News Agencies, "Yemen's Houthis, WFP reach deal to resume food relief," August 4, 2019.

4 Teresa Welsh, "Biometrics disagreement leads to food aid suspension in Yemen," ***Devex***, June 24, 2019, https://www.devex.com/news/biometrics-disagreement-leads-to-food-aid-suspension-in-yemen-95164 (accessed August 9, 2019).

5 Welsh, "Biometrics disagreement leads to food aid suspension in Yemen," 2019.

As AI adoption grows rapidly, this report aims to develop AI principles and recommendations that would be specific to vulnerable people in the humanitarian field.

Though developing AI-specific principles and recommendations for vulnerable people is a new concept, creating such principles for emerging technologies more broadly is not a unique phenomenon. Such principles have been created by many international organizations (IOs), nongovernmental organizations (NGOs), and large technology companies to advance the goals of the 2030 Sustainable Development Agenda, the Charter of the United Nations (UN) and in the spirit of promoting "tech for social good". For instance, on July 12, 2018 the UN Secretary-General António Guterres created the High-level Panel on Digital Cooperation to discuss how digital technologies can be used inclusively and in accordance to human rights in a multistakeholder setting.[6] This panel was followed by a set of five principles set forth in the Secretary-General's Strategy on New Technologies in September 2018: protect and promote global values, foster inclusion and transparency, work in partnership, build on existing capabilities and mandates, and be humble and continue to learn.[7] Specific branches of the UN are following suit in establishing principles or recommendations related to emerging technologies, including the March 2019 Data Responsibility Guidelines created by the UN Office for the Coordination of Humanitarian Affairs (OCHA).[8]

Similarly, NGOs such as Amnesty International, Access Now, Human Rights Watch, and the Wikimedia Foundation have worked on preparing or endorsing AI principles and recommendations through documents like "The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems" in May 2018. This declaration aims to establish an ethical framework for machine learning and AI systems more broadly in accordance with human rights norms.[9]

---

6 Eleonore Pauwels, "The New Geopolitics of Converging Risks: The UN and Prevention in the Era of AI," ***United Nations University Centre for Policy Research***, April 29, 2019, https://i.unu.edu/media/cpr.unu.edu/attachment/3472/PauwelsAIGeopolitics.pdf (accessed August 9, 2019).

7 United Nations, "Secretary-General's Strategy on New Technologies," ***United Nations***, September 2018, https://www.un.org/en/newtechnologies/images/pdf/SGs-Strategy-on-New-Technologies.pdf (accessed August 9, 2019).

8 The Centre for Humanitarian Data, "Data Responsibility Guidelines (Working Draft)," ***United Nations Office for the Coordination of Humanitarian Affairs***, March 2019, https://centre.humdata.org/wp-content/uploads/2019/03/OCHA-DR-Guidelines-working-draft-032019.pdf (accessed August 9, 2019).

9 Amnesty International and Access Now, "The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems," ***Access Now***, May 16, 2018, https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/ (accessed August 9, 2019).

Furthermore, big tech companies such as Google,[10] Microsoft,[11] and IBM[12] have all developed their own sets of AI principles in recognition of concerns over algorithmic bias, security risks, fairness, and transparency.

Despite the abundance of principles, recommendations, guidelines, and declarations concerning AI systems, what are the implications of this technology for vulnerable populations specifically? There is a danger of having too many broad principles that do not directly address the world's most marginalized communities, especially in situations of humanitarian disaster where many AI systems are used. This report argues that AI systems that have prevalent biases, security risks, and issues with consent can undermine the role of humanitarian actors in disaster contexts by leaving aid recipients at further risk of vulnerability. The increased risk of vulnerability not only impacts vulnerable populations, but also indirectly affects the achievement of the 2030 Sustainable Development Agenda, and the maintenance of the UN Charter and international human rights laws.

This report provides support for the argument by detailing examples of AI systems that are used for vulnerable populations, analyzing different forms of biases pertaining to AI systems, evaluating security risks, discussing issues of consent in respect to disclosing personal information, and compiling AI principles and recommendations for vulnerable people. It is informed by scholarly and news articles, as well as 20 interviews with a diverse range of experts in the AI and humanitarian fields. They work in different branches of the UN, in academia, think tanks, NGOs, and private sector companies. Interviewees were recruited through personal research and recommendations from colleagues and previous interviewees.

In the report, the phrase "AI systems" refers to technological systems that use AI, including conversational agents like chatbots and biometric technologies like facial recognition software. "Vulnerable populations" is also left intentionally broad to refer to marginalized communities in humanitarian and disaster response contexts.

---

10 Google AI, "Artificial Intelligence at Google: Our Principles," ***Google AI***, https://ai.google/principles/ (accessed August 9, 2019).

11 Microsoft, "Microsoft AI principles," ***Microsoft***, https://www.microsoft.com/en-us/ai/our-approach-to-ai (accessed August 9, 2019).

12 IBM, "IBM's Principles for Trust and Transparency," ***IBM***, May 30, 2018, https://www.ibm.com/blogs/policy/trust-principles/#C3 (accessed August 9, 2019).

Section One:

Examples of AI Systems for Vulnerable Populations

There are many recent examples of AI systems that are used on vulnerable populations. Detailing some of them is necessary to demonstrate how AI systems can undermine the role of humanitarian agents by leaving vulnerable people at further risk. This section briefly describes definitions of AI and machine learning before discussing three examples of AI systems for vulnerable people: X2AI's psychotherapy chatbot, Hala Systems' early warning system for airstrikes, and the use of AI technologies on refugees in the European Union (EU).

Defining AI

AI is a broad research field within computer science, with no formal definition.[13] It can generally be described as "the study of the design of intelligent agents", with agents referring to processing units like computers.[14] The narrow form of AI is the type of AI currently in use, and it can execute defined tasks in numerous fields.[15] [16] It is important to note that AI in its current form is reliant on human decision-making, so it cannot think for itself or move beyond extracting patterns in the datasets its algorithms are fed.[17] Machine learning is a type of AI that has become so synonymous with the discipline that people oftentimes use machine learning and AI interchangeably.[18] Machine learning requires having large amounts of data to train an AI system via an iterative process that relies on statistics to create algorithmic models.[19] A specific task is defined, and the algorithmic models detect

13 Pauwels, "The New Geopolitics of Converging Risks," 2019.

14 Frederik Zuiderveen Borgesius, "Discrimination, artificial intelligence, and algorithmic decision making," *Council of Europe*, 2018, https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73 (accessed August 9, 2019).

15 Pauwels, "The New Geopolitics of Converging Risks," 2019.

16 Zuiderveen Borgesius, "Discrimination, artificial intelligence, and algorithmic decision making," 2018.

17 Annette Zimmermann and Bendert Zevenbergen, "AI Ethics: Seven Traps," *Freedom to Tinker: research and expert commentary on digital technologies in public life*, March 25, 2019, https://freedom-to-tinker.com/2019/03/25/ai-ethics-seven-traps/ (accessed August 9, 2019).

18 Zuiderveen Borgesius, "Discrimination, artificial intelligence, and algorithmic decision making," 2018.

19 Pauwels, "The New Geopolitics of Converging Risks," 2019.

patterns in the data set to learn how to most effectively achieve the task without needing to be told exactly what to do.[20] [21] Many AI systems in the humanitarian space use machine learning, including Karim, the psychotherapy chatbot created by X2AI.

Karim the Chatbot

Karim, displayed as a cartoon face with a goatee, is a chatbot, a conversational agent that is programmed to simulate human-to-human interaction through written or auditory methods. X2AI manually created algorithms as well as used machine learning to create their bot, and tested it on a group of 60 mainly male Syrians in the Za'atari camp.[22] Many participants in the pilot phase of Karim had difficulty understanding that they were interacting with an AI entity rather than a person when they received texts on their mobile device. This was exacerbated by the fact that there is no definition of chatbot in Arabic.[23] Moreover, even though X2AI created a secure network in efforts to enhance the security of the text message exchanges, participants remained concerned over the ability of governments and terrorist organizations to monitor these exchanges. There were also concerns over when Karim would alert medical professionals to intervene in situations where Syrian subjects expressed an intent to harm themselves or others; would the algorithms be sophisticated enough to recognize certain cues in oftentimes ambiguous conversational settings?[24]

The experts interviewed for this report expressed deep concerns over the creation of a psychotherapy chatbot itself as well as it being applied and tested on vulnerable populations. One interviewee expressed two concerns with X2AI's bot. First, they pointed out that although some people may say that having a psychotherapy chatbot for vulnerable people is better than having no form of therapy at all, it can conversely reduce the amount of mental health professionals deployed in humanitarian contexts in the future.[25] It can change the framing of mental health support for refugees so the new minimum standard becomes the offerings of psychotherapy chatbots, potentially preventing some refugees from getting more substantive support. In addition, this argument suggests that the good outcomes produced by using this AI system outweigh the bad

20 Pauwels, "The New Geopolitics of Converging Risks," 2019.
21 Zuiderveen Borgesius, "Discrimination, artificial intelligence, and algorithmic decision making," 2018.
22 Romeo, "The Chatbot Will See You Now," 2016.
23 Romeo, "The Chatbot Will See You Now," 2016.
24 Ibid.
25 Os Keyes, interview with author, July 22, 2019.

outcomes. However, bad outcomes are still produced which is arguably unacceptable when the result may be serious harm or death to a vulnerable person.[26] To reduce the amount of bad outcomes and make the chatbot more accurate, additional testing needs to take place, which puts vulnerable people at further risk of harm.

A second concern is that AI systems like psychotherapy chatbots are commonly situated around short term outcomes as measures of success rather than the potential long-term consequences of their interactions with the user. For example, could the use of chatbots actually increase the sense of isolation for these populations?[27]

Concerns were also raised regarding whether these types of chatbots have been clinically tested. Related to the idea that long term consequences are oftentimes not researched is the idea that psychotherapy chatbots may not be clinically tested, which is important for a mental health tool.[28] However, could the lack of clinical testing be deemed acceptable in emergency situations? Perhaps, but such justification could lead to a potential abuse of the circumstance and put the population at further risk.[29]

In addition, research shows that in disaster situations, vulnerable people like human-to-human interactions and like to speak with people about their concerns.[30] Regarding feedback mechanisms, which are ways of giving and receiving feedback, when vulnerable people have complex problems they would like to express them through human conversations and receive feedback from another human rather than texting a chatbot.[31] These points are supported by the Humanitarian Technologies Project, an 18-month ethnographic study of the uses of digital technologies in the aftermath of Typhoon Haiyan that took place in the Philippines in November 2013.[32] [33] This natural disaster sparked an international humanitarian response, providing the necessary circumstances by which

26 Os Keyes, interview with author, July 22, 2019.

27 Os Keyes, interview with author, July 22, 2019.

28 Allison Gardner, interview with author, June 26, 2019.

29 Allison Gardner, interview with author, June 26, 2019.

30 Humanitarian Technologies Project, http://humanitariantechnologies.net (accessed August 9, 2019).

31 Mirca Madianou, interview with author, July 8, 2019.

32 Humanitarian Technologies Project, http://humanitariantechnologies.net (accessed August 9, 2019).

33 Mirca Madianou et al., "The Appearance of Accountability: Communication Technologies and Power Asymmetries in Humanitarian Aid and Disaster Recovery," ***Journal of Communication*** 66, no. 6 (2016): 961.

to investigate how digital technologies are used in disaster contexts. It is important to note that this study did not look at AI systems in particular, but rather focused on the feedback mechanisms associated with vulnerable populations using their mobile devices to send text messages to humanitarian agency "hotlines".[34] However, AI-powered chatbots like X2AI's Karim rely on a short messaging service (SMS) format as well, making the Humanitarian Technologies Project's findings applicable to such conversational agents.

The project found that there is a gap between the assumptions about technology in humanitarian contexts and the actual use and effectiveness of such technology by vulnerable people.[35] This gap is mediated by the cultural contexts of the affected populations. For instance, in the context of the Philippines, disaster relief feedback to humanitarian agencies via SMS was influenced by norms of gratitude that discourage providing negative feedback to aid workers.[36] In the case of this project, there was irony in the fact that though technological platforms were purported as a solution by humanitarian agencies, they ended up contributing to a gap between these agencies and vulnerable populations.[37] These lessons from the Humanitarian Technologies Project are duly applicable to AI systems, in particular psychotherapy chatbots. Though there are resource and funding shortages in humanitarian disasters, the limitations of using conversational agents in lieu of humans should be at the forefront in deciding whether such technologies should be used at all.

Hala Systems: An Airstrike Early Warning System

Another AI system that is used for vulnerable populations is called Hala Systems, a more optimistic form of AI in disaster contexts. Hala Systems, created by two Americans in 2015, is an early warning system for airstrikes in Syria using AI and geospatial technologies.[38] The system relies on the aerial observations of a network of Syrian civilians, alongside acoustic data from remote sensors that help indicate the speed and types of incoming planes.[39] Hala Systems compares newly received data with previous data to form predictions on airstrikes, which are then published on social media sites. This AI-powered system has not only reduced the casualty rate by between 20 to 30 per

34 Madianou et al., "The Appearance of Accountability," 964.
35 Humanitarian Technologies Project, http://humanitariantechnologies.net (accessed August 9, 2019).
36 Madianou et al., "The Appearance of Accountability," 976.
37 Madianou et al., "The Appearance of Accountability," 978.
38 Pauwels, "The New Geopolitics of Converging Risks," 2019.
39 Pauwels, "The New Geopolitics of Converging Risks," 2019.

cent in some Syrian regions in 2018, but it has also provided war crime evidence for 75 events and generated 250 reports on aircraft data and observations to governments, the UN, and NGOs.[40] Hala can be considered a much more successful implementation of an AI system for vulnerable people than X2AI's psychotherapy chatbot. This may be partly due to how these AI systems are structured. Though both AI applications were created by Americans rather than local populations, Hala Systems relies on a bottom-up approach in its implementation since Syrian civilians themselves are involved in making on the ground observations and installing the remote sensors in strategic locations.[41] Conversely, though Karim is fluent in Arabic and coded to use some Arabic slang,[42] it does not require the ongoing involvement of the vulnerable population it is designed to help.

40 Ibid.

41 Eleonore Pauwels, interview with author, July 10, 2019.

42 Romeo, "The Chatbot Will See You Now," 2016.

Biometrics and Refugees

A final example of an AI system used for vulnerable populations more broadly is using AI technologies on refugees in the EU. Chinmayi Arun, summarizing Dragana Kaurin's work on the digital agency of refugees, argues that AI technologies can reduce the agency, or autonomy, of refugees entering the EU and increase their vulnerability.[43] Refugees must provide their personal data when seeking asylum in the EU, often in the form of biometric registration, which is enforced by border control and humanitarian aid agencies alike.[44] Asylum seekers are especially vulnerable, with many arriving at EU borders after fleeing persecution and undertaking harrowing journeys. Even though AI systems, particularly biometric ones, are designed with good intentions, they can put these vulnerable people at further risk. For instance, the International Committee of the Red Cross created the Trace the Face program, which uses facial recognition software, a type of biometric technology, to search for missing people using photos, donated by the person's family, of the missing person themselves or their relatives.[45] However, refugees have mentioned that some may be fleeing from family members or others in their home communities, making facial recognition projects such as Trace the Face potentially harmful.[46] Similar to Karim, the psychotherapy chatbot, vulnerable populations have not been meaningfully involved in the design process of the AI technologies aimed to help them in terms of biometric technologies used at EU borders. This lack of involvement can in turn increase their vulnerability and reduce their digital agency. AI systems in the form of biometrics are a part of the potential biases entrenched in these systems, which is now discussed.

43 Chinmayi Arun, "AI and the Global South: Designing for Other Worlds," in *The Oxford Handbook of Ethics of AI* (forthcoming), eds. Markus D. Dubber, Frank Pasquale, and Sunit Das (Oxford University Press, 2019), 8.

44 Arun, "AI and the Global South," 8.

45 Ibid.

46 Ibid, 9.

Section Two:
Biases

Humanitarian actors using AI systems with prevalent biases is one way that they can undermine their own objectives by leaving vulnerable populations at further risk. This section begins with a discussion of the different forms of biases that may appear in AI systems for vulnerable people. Specifically, the forms of biases examined are: Western cultural norms embedded in system design; biometric technologies, including facial recognition software; and incomplete data sets. Then, this section analyzes varying perspectives of whether there are ways forward in ensuring that diverse test cases are used in training AI systems for vulnerable populations. For an analysis of the different forms of biases in AI systems in general, not only pertaining to vulnerable groups, please refer to Frederik Zuiderveen Borgesius's study "Discrimination, artificial intelligence, and algorithmic decision making" (2018).[47]

Western Cultural Norms Embedded in System Design

One type of bias in AI systems for vulnerable people is that Western cultural norms are embedded in the design of such systems. Referring back to X2AI's psychotherapy chatbot, many assumptions about psychotherapy and mental health more broadly are based on Western cultural norms that may be explicitly or implicitly built into these systems which were designed by Western innovators.[48] As vulnerable populations come from diverse cultural backgrounds, designing AI systems in the West may mean that the system has a high accuracy rate in the testing stage. However, when it is deployed it may have a much lower accuracy rate or lead to discriminatory outcomes because of the way many Western assumptions are inherently built into the system.[49] Systems thought about, designed, and developed with this limited Western perspective can in turn mean that they may be missing other crucial ways these systems should be deployed in the Global South.[50] [51] The root causes of problems

47 Frederik Zuiderveen Borgesius, "Discrimination, artificial intelligence, and algorithmic decision making," *Council of Europe*, 2018, https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73 (accessed August 9, 2019).

48 Os Keyes, interview with author, July 22, 2019.

49 Os Keyes, interview with author, July 22, 2019.

50 Rumman Chowdhury, interview with author, July 16, 2019.

51 Eleonore Pauwels, interview with author, July 10, 2019.

AI systems are designed to improve may be missed because normative and cultural preferences are not accounted for amongst the vulnerable populations themselves.[52]

Biometric Technologies

Biometric technology, including facial recognition software, is another type of AI system that can result in bias. Biometrics uses digital technology to record and analyze a person's biological features like fingerprints, iris scans, voice and facial patterns.[53] The data collected from biometric technology can be used for identification or verification purposes. In the humanitarian field, this technology is mainly used to identify aid recipients and vulnerable groups. Identification procedures however, have a higher error rate than verification processes as they entail checking one biometric profile against an entire database of profiles.[54] Biometric errors can also take place in the registration and processing of the data of vulnerable people, including refugees.[55] Biometric identification processes depend on AI because to make the processes work, artificial neural networks (ANN) use machine learning algorithms to detect patterns in the data inputs.[56] AI is also used in the capturing of biometric data, such as in capturing the image of an iris scan.

An example of biometric technology usage was the International Committee of the Red Cross' Trace the Face program, discussed in Section One. Another example of bias in biometric technology for vulnerable people is the research related to fingerprints and iris scans. The study showed that manual workers, elderly people, people who work in care, health, and beauty industries amongst others, many of which are vulnerable, have lighter fingerprints.[57] The faintness of their fingerprints, which can impact whether they are recognized by biometric systems, is due to the hard, physical labour of their professions. Iris scans, though thought of as a more reliable biometric indicator, can also be less accurate due to a variety of factors including age.[58] These examples show that AI systems are not perfectly accurate.

52 Eleonore Pauwels, interview with author, July 10, 2019.
53 Madianou, "The Biometric Assemblage," 3.
54 Madianou, "The Biometric Assemblage," 4.
55 Ibid, 10.
56 Ibid.
57 Ibid.
58 Ibid.

Perhaps one of the most well-known cases of biases inherently built within biometric technologies is that facial recognition software has a much lower accuracy rate when it comes to non-white populations, blatantly demonstrating that biometrics are biased and discriminate in terms of race amongst other classes like gender.[59] [60] [61] For example, in their study on racial bias within facial recognition technology, Joy Buolamwini and Timnit Gebru (2018) found that females of colour were the most misclassified group, with an error rate of 34.7 percent while white males with a fair complexion had the lowest error rate of 0.8 percent.[62] Overall, facial recognition systems were the most accurate for people with lighter skin tones and worked better on male faces versus female faces.[63]

The startling gap in error rate demonstrates what can happen if facial recognition AI is not trained on diverse test cases rather than mainly images of white men, who make up the majority of AI developers.[64] [65] It is appalling to know that AI systems can be discriminatory, and that they implicitly suggest who is recognized as a person. This bias is even more disturbing in disaster relief and humanitarian contexts when being recognized by biometric technologies may affect whether or not one receives food, shelter, and other basic resources.[66] This problem is amplified by the fact that the populations who have a lower accuracy rate when it comes to their biometrics being recognized are at a greater risk of being vulnerable, including females, people of colour, and children.[67] [68]

Furthermore, when thinking about how biometric technologies can be biased, it is important to note that making these AI systems more accurate is not enough.[69] Improving the demographics of the

59 Mirca Madianou, interview with author, July 8, 2019.

60 Sarah Myers West, interview with author, July 22, 2019.

61 Joy Buolamwini and Timnit Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," ***Proceedings of Machine Learning Research*** 81, (2018): 1.

62 Buolamwini and Gebru, “Gender shades,” 1.

63 Ibid, 8, 12.

64 Zuiderveen Borgesius, “Discrimination, artificial intelligence, and algorithmic decision making,” 2018.

65 Tom Simonite, “AI is the Future - But where are the Women?” ***Wired***, August 17, 2018, https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/ (accessed August 9, 2019).

66 Sarah Myers West, interview with author, July 22, 2019.

67 Sarah Myers West, interview with author, July 22, 2019.

68 Deanna Paul, “A maker of police body cameras won’t use facial recognition yet, for two reasons: Bias and inaccuracy,” ***The Washington Post***, June 28, 2019, https://www.washingtonpost.com/nation/2019/06/29/police-body-cam-maker-wont-use-facial-recognition-yet-two-reasons-bias-inaccuracy/ (accessed August 9, 2019).

69 Meredith Broussard, interview with author, June 13, 2019.

training data may make biometric systems more accurate in recognizing vulnerable populations. But we also need to make sure that this technology and data is not intentionally or unintentionally weaponized and result in the harmful targeting of or lack of awareness of a group in need.[70] [71]

In disaster response situations, biometric errors, which are due to algorithmic models that use training sets that have human biases, can result in the legitimization and reproduction of biases in respect to race, gender, and other types of discrimination.[72] This problem is compounded by the fact that many AI systems are "black boxes", meaning that as a system's machine learning progresses, it becomes increasingly more difficult to ascertain how the AI system makes decisions.[73] It therefore becomes more difficult to know if a vulnerable person becomes the victim of a biased AI technology.

Incomplete Data Sets

A third bias in AI systems for vulnerable people is in terms of incomplete data sets. The effectiveness of AI technology depends on the quality of the data set it is trained on, and inevitably data sets are incomplete, though there is a myth of having one that is complete.[74] In humanitarian and disaster response contexts, data sets will oftentimes have large gaps or temporal biases, as well as a lack of representation of the most vulnerable affected groups.[75] Thus, data that is used in machine learning algorithms that run AI applications can have biases towards these groups, potentially increasing their vulnerability by excluding them and not taking their needs into account.[76] [77] This challenge is exacerbated by the fact that many vulnerable people are not included in data sets because they do not have a digital footprint and thus do not generate data at all.[78] Moreover, they may not even know that they are being excluded, which reduces their agency to respond to a potential bias in the data set.

70 Meredith Broussard, interview with author, June 13, 2019.
71 Cathy O'Neil, interview with author, July 15, 2019.
72 Madianou, "The Biometric Assemblage," 10.
73 Zuiderveen Borgesius, "Discrimination, artificial intelligence, and algorithmic decision making," 2018.
74 Mirca Madianou, interview with author, July 8, 2019.
75 Mirca Madianou, interview with author, July 8, 2019.
76 Ibid.
77 Mila Romanoff, interview with author, July 24, 2019.
78 Interviewee, interview with author, July 16, 2019.

More generally, human decision-making is present throughout the entire development process for AI systems, and these systems will be biased if the data set their algorithmic models are trained on is incomplete.[79] [80] In particular, algorithmic models can be biased because models are representations of reality rather than reality itself.[81] It is how the human who developed the model views reality, so the source of the data, the proportionality of groups within the data, and the variables chosen could be influenced by bias, even if it is unintentional.[82] Regarding proportionality of groups, if there is data on vulnerable groups and it is included in a model, it may need to be reweighted if the sample size is too small in comparison to the other groups in the model. Reweighting the variables for vulnerable groups or prioritizing their needs in the model in some way reinforces the characteristics of this group, which is a form of positive bias.[83] Thus bias, whether positive or negative can occur when developing AI systems that take underrepresented groups into account.

Diverse Test Cases

Considering the fact that there are so many biases that can negatively impact vulnerable populations, including Western cultural norms, biometric technologies, and incomplete data sets, a question posed to interviewees was: *What are ways forward in ensuring that diverse test cases are used in training AI systems?* The results raised varying perspectives.

One perspective is that the concept of diverse test cases should be rejected because it has an underlying assumption that by including sufficiently diverse test cases, it is possible to create a universal algorithm.[84] In turn, the concept of a universal algorithm is related to a solutionist paradigm that assumes AI systems can "solve" certain concerns in vulnerable communities, bolstering the idea that technology can solve societal problems. In other words, when designing an AI system for vulnerable people, even if developers want it to be as robust as possible so they include race, gender, and age diverse datasets to train their algorithmic models, they are unable to include every variable and class to make the test cases completely

79 Allison Gardner, interview with author, June 26, 2019.
80 Miguel Angel Hernandez Rivera, interview with author, July 18, 2019.
81 Miguel Angel Hernandez Rivera, interview with author, July 18, 2019.
82 Allison Gardner, interview with author, June 26, 2019.
83 Miguel Angel Hernandez Rivera, interview with author, July 18, 2019.
84 Os Keyes, interview with author, July 22, 2019.

diverse. For instance, they may exclude employment type or one's willingness to participate in the data collection process.[85]

With these considerations in mind, a way forward for creating more diverse datasets could be to stop thinking of these datasets as a way to create universal algorithms in a solution-driven paradigm. For example, if an AI system is being developed specifically for Syrian refugees in a particular local context, the system would be more adaptable and have the potential to have a much lower error rate than if an AI system is built for all refugees universally.[86] Returning to X2AI's psychotherapy chatbot, rather than Western AI developers building this chatbot and deploying it in the field for testing, local developers could have iterated on the system, testing it amongst the target population and correct it contextually alongside cultural experts.[87] This is not a popular methodology for developing AI systems for vulnerable people because it contradicts a common utopian view that promises the achievability of universal algorithms capable of solving any problem. However, this methodology is not only more realistic, but it reduces the possibility of undue harm towards the vulnerable people these AI systems are designed to help. Ultimately, as Os Keyes states, "It is not about diversifying the datasets so much as it is diversifying the algorithms and making sure that the system you are using is actually designed for the context you are using it in".[88]

In a similar vein, even before thinking about ways forward in diversifying datasets, the developers of AI systems should think about whether there actually is a need for the technology they aim to create, what the potential challenges are, and, considering these challenges, should this technology still be built. Most importantly, these framing questions should include the voices of vulnerable populations, making sure that they have ownership over these systems.[89]

In addition, as previously mentioned, the idea of creating sufficiently diverse test cases is related to a solutionist paradigm that presupposes technological solutions to societal issues. Datasets have biases because there is discrimination in society, so datasets reproduce or amplify them.[90] [91] Though people

85 Os Keyes, interview with author, July 22, 2019.
86 Ibid.
87 Ibid.
88 Ibid.
89 Rumman Chowdhury, interview with author, July 16, 2019.
90 Mirca Madianou, interview with author, July 8, 2019.
91 Zimmermann and Zevenbergen, "AI Ethics: Seven Traps," 2019.

can identify best practices to alleviate such negative outcomes, an awareness needs to be held that AI systems cannot be used to solve social problems[92] Technochauvinism is a term that Meredith Broussard uses to describe the belief that technology can always fix problems. It is something that she rejects as a flawed assumption.[93] This belief is also related to the concept of automation bias - an idea that technological systems are a logical, infallible authority and thus their outputs, such as algorithmic decisions, should be believed and followed.[94]

Though designers of AI systems cannot create test cases that are universally diverse or oftentimes diverse enough, a way forward in thinking about increasing diversity in datasets may be in increasing the diversity of the teams that build them. Many studies on racial bias in biometric technologies, such as the one by Buolamwini and Gebru (2018), suggest that a fundamental problem in building AI systems with diverse test cases is that the teams that build these systems are mostly made up of homogenous males in Western countries who use limited test cases that look mostly like them. Think tanks like Women Leading in AI are looking at building requirements into the regulation process for addressing the lack of diversity and inclusion in the AI field, including in the development of AI systems.[95] If AI system developers have difficulty achieving diverse team requirements, organizations like Women Leading in AI are willing to help.[96]

Legislating for diversity requirements in AI would help demonstrate more ethical AI practices, including more thoughtful design in systems for vulnerable people. However, legislation processes are oftentimes not easy. For example, in New York City the algorithmic accountability bill was enacted in January 2018, establishing a task force to study how agencies working with the city use their algorithms to make decisions that affect the public's life.[97] The task force would also investigate whether the algorithmic systems discriminate against certain groups of people. The bill enacted is a much smaller version of one of the original drafts, which desired to impose a requirement on city

92 Mirca Madianou, interview with author, July 8, 2019.

93 Meredith Broussard, ***Artificial Unintelligence: How Computers Misunderstand the World*** (Cambridge: MIT Press, 2018).

94 Zuiderveen Borgesius, "Discrimination, artificial intelligence, and algorithmic decision making," 2018.

95 Allison Gardner, interview with author, June 26, 2019.

96 Allison Gardner, interview with author, June 26, 2019.

97 Lauren Kirchner, "New York City Moves to Create Accountability for Algorithms," ***ProPublica***, December 18, 2017, https://www.propublica.org/article/new-york-city-moves-to-create-accountability-for-algorithms (accessed August 9, 2019).

agencies to publicize the source code of all the algorithms they use in relation to policing, penalties, or services and make them publicly available for testing.[98] [99] The mandate to publicize the source code of algorithms caused a lot of pushback from policymakers arguing that doing so could create a cybersecurity risk, and from tech companies concerned that such public disclosure could lessen their competitive edge.[100] Due to the pushback, the task force must now rely on voluntary disclosures from city agencies to analyze and audit their algorithms, resulting in little progress.[101] [102] The algorithmic accountability bill exemplifies a main downside to the legislation process: the original bill proposed oftentimes is influenced, shaped, and lobbied by powerful actors in a multistakeholder context, which can result in a less effective bill enacted concerning AI systems.

98 Kirchner, "New York City Moves to Create Accountability for Algorithms," 2017.

99 Julia Powles, "New York City's Bold, Flawed Attempt to Make Algorithms Accountable," ***The New Yorker***, December 20, 2017, https://www.newyorker.com/tech/annals-of-technology/new-york-citys-bold-flawed-attempt-to-make-algorithms-accountable (accessed August 9, 2019).

100 Powles, "New York City's Bold, Flawed Attempt to Make Algorithms Accountable," 2017.

101 Powles, "New York City's Bold, Flawed Attempt to Make Algorithms Accountable," 2017.

102 Allison Gardner, interview with author, June 26, 2019.

Section Three:

Security Risks

Another way that AI systems for vulnerable populations can undermine the role of humanitarian agents is in relation to the security risks associated with these systems. This section discusses the security risks of data breaches and data security practices; function creep; and Project Connect and the UN Children's Fund's (UNICEF) work as an example of how to mitigate potential security risks.

Data Breaches and Data Security Practices

Data breaches and data security practices are two related security risks that can affect AI systems for vulnerable people. Data breaches are not only applicable to AI systems in a humanitarian context, but they can have a particular harmful effect due to the already established vulnerability of the people involved. A data breach concerning vulnerable people's data means that such populations may be at further risk of discrimination, persecution, and forced repatriation amongst other harms.[103] Unfortunately, there are many examples of data breaches in humanitarian and disaster response situations. For instance, in a refugee camp in Greece, an information technology project was undermined by approximately 80,000 malware attacks per week in 2015.[104] In 2017, the cloud server of eleven humanitarian organizations was hacked, leading to thousands of vulnerable people potentially having their personal information revealed.[105] Though cloud computing is not an AI system, the remote storage of data, whether that data is collected and stored by AI systems or not, increases the potential for data breaches.

At the same time, there is also a security risk in respect to data security practices regarding the physical storage of data collected from AI systems in humanitarian contexts. In these situations, data from AI systems is likely to be stored in less controlled environments, especially if a physical copy is kept. For instance, if data is stored on a portable hard drive in a refugee camp, it is much easier to be stolen.[106] The data collection devices themselves can put vulnerable populations at further risk if the

103 Madianou, "The Biometric Assemblage," 10.
104 Madianou, "The Biometric Assemblage," 11.
105 Ibid, 11.
106 Os Keyes, interview with author, July 22, 2019.

sensitive information stored on it gets placed into the wrong hands. It is also important to build AI systems where people can access their data and have it deleted, even if they do not meet atypical Western-designed authentication processes like having an identity document when requesting the deletion of personal data or access to it.[107] Building such systems is hard to achieve though because of poor data security practices in the humanitarian field.

Data security practices can influence whether a data breach occurs. Given that the fundamental elements of digital data are its replicability and retrievability, data breaches can happen if these characteristics are misused due to poor data security practices by humanitarian agencies.[108] Poor practices like leaving computers unattended in public settings can result in the sensitive data of vulnerable people being stolen or compromised.[109] Humanitarian organizations are increasingly taking an interest in addressing data security concerns, seen by the 2019 publication of UN OCHA's Data Responsibility Guidelines. However, a limited capacity in the humanitarian sector still persists regarding the creation of policy and legal frameworks for data security.[110] [111] Reasons behind the lack of strong enforcement of data security practices within humanitarian organizations include a lack of resources, funding, and expertise.[112] There is limited capacity to ensure that there are proper security measures, including proper training modules on security risks and data security in particular. This limited capacity is in part because many humanitarian organizations are now realizing the power of AI systems when these systems were used less in the past. The greater demand is paired with a lack of expertise and capacity to adapt to this rapidly changing industry and the new challenges it poses. Humanitarian workers being unaware of correct data security practices can lead to situations of having too much data stored that produce identifiable data sets on vulnerable populations, which is magnified by the weak protocols in place.[113]

In addition, data security is important to maintain to ensure the privacy of vulnerable people is upheld, especially in contexts where governments and other institutions may pose a danger for these people. Well-curated data security practices can help maintain privacy especially in countries

107 Os Keyes, interview with author, July 22, 2019.
108 Madianou, "The Biometric Assemblage," 11.
109 Madianou, "The Biometric Assemblage," 11.
110 Ibid.
111 Mila Romanoff, interview with author, July 24, 2019.
112 Mila Romanoff, interview with author, July 24, 2019.
113 Ibid.

that do not have strong government support or policies to support this endeavour, or in countries where a lack of data privacy can lead to greater insecurity for marginalized peoples.[114] [115] It is important to remember that AI systems designed in the West usually take the reliability of systems of power, such as a stable government and infrastructure for granted. Vulnerable populations are vulnerable in part because they usually do not have systems of power or existing infrastructure that they can trust.[116] Maintaining better data security practices means that vulnerable populations can have greater trust in the use of AI systems. They can trust that these systems do not manipulate their data for purposes like creating mistrust, political systems that can lead to conflict, and misusing data about voters before an election.[117]

A final important point in relation to data security is that humanitarian organizations should not only be mindful of ensuring strong protocols in the present, but should be aware of the principle of "download now, decode later".[118] This principle refers to the notion that though data security protocols currently centre around the ability to encrypt data, in the long term, or in a time frame when this information is still sensitive, quantum computers may have the computing power to easily decode encrypted data.[119] Quantum computers are machines that rely on quantum physics to solve problems that prove difficult for current computers.[120] If quantum computers with the necessary computing power are built, they would be able to break encrypted data, including data used in AI systems, creating new risks for data security in the future.[121]

Function Creep

Function creep is another security risk that can affect AI systems for vulnerable people. This concept refers to situations where data is collected for one objective, such as combatting the diversion of resources when providing food aid, and then is used for an entirely different goal, such as surveilling

114 Interviewee, interview with author, July 16, 2019.
115 Rumman Chowdhury, interview with author, July 16, 2019.
116 Rumman Chowdhury, interview with author, July 16, 2019.
117 Eleonore Pauwels, interview with author, July 10, 2019.
118 Lambert Hogenhout, interview with author, June 17, 2019.
119 Lambert Hogenhout, interview with author, June 17, 2019.
120 Lily Chen et al., U.S. Department of Commerce, National Institute of Standards and Technology, ***Report on Post-Quantum Cryptography***, April 2016, NISTIR 8105. http://dx.doi.org/10.6028/NIST.IR.8105
121 Chen et al., ***Report on Post-Quantum Cryptography***, 2016.

a vulnerable population.[122] [123] This subsection discusses two types of exchanges that can lead to function creep: exchanges between private vendors and humanitarian agencies, and exchanges between state governments and humanitarian agencies.

Regarding private vendors-humanitarian agency exchanges, data collection for AI systems is often outsourced to private vendors.[124] Data may be shared with private companies because humanitarian agencies oftentimes want to use the technology of said companies to collect data.[125] Private vendors benefit from this relationship because it can provide them with an excellent public relations boost, access to new data and the chance to pilot projects. For example, large tech companies like IBM and Google have started deploying their technologies in the humanitarian space, and in February 2019 the WFP stated they would be partnering with Palantir in a $45 million deal, in efforts to support its Innovation Accelerator.[126] Palantir is a software company often operating within the spaces of intelligence and immigration enforcement that has been rumored to be tied to the Cambridge Analytica scandal.[127] In their partnership, WFP stated Palantir would not have access to data that states beneficiary information, Palantir would not provide or collect data for the WFP, and the company would consider the data confidential.[128] However, this partnership underscores that it is crucial for humanitarian organizations to vet their private vendors wisely. If third party companies perpetuate unjust practices in their AI systems, the partnering humanitarian organizations may unknowingly perpetuate these practices as well.

Regarding state-humanitarian agency exchanges, host governments usually ask humanitarian organizations or try to make it a requirement in their agreements to share the data collected in their countries.[129] For instance, some humanitarian agencies undertake biometric registrations in collaboration with host governments, or support governments in conducting this form of AI

122 Madianou, "The Biometric Assemblage," 8.

123 Mirca Madianou, interview with author, July 8, 2019.

124 Madianou, "The Biometric Assemblage," 7.

125 Mirca Madianou, interview with author, July 8, 2019.

126 Madianou, "The Biometric Assemblage," 7.

127 Madianou, "The Biometric Assemblage," 7.

128 Enrica Porcari, "A statement on the WFP-Palantir partnership," ***Insight by The World Food Programme***, February 7, 2019, https://insight.wfp.org/a-statement-on-the-wfp-palantir-partnership-2bfab806340c (accessed August 9, 2019).

129 Madianou, "The Biometric Assemblage," 8, 11.

technology.[130] These organizations are put in a difficult position because on the one hand, they are operating in the sovereign territory of a state. Yet on the other hand, if permission for data sharing requests is granted, humanitarian agents have no authority over how that data is used, including by future governments.[131] If current or future governments want to cause harm to vulnerable people, function creep via state-humanitarian agency exchanges can provide access to the sensitive information required to do so. For example, certain sexual orientations or religious beliefs could be collected from a vulnerable population, and that information could be detrimental if it is ever shared with a state that intends to persecute people for these characteristics.[132] An example of the potential for function creep related to state-humanitarian agency exchanges is that in 2017 to 2018 the Office of the UN High Commissioner for Refugees (UNHCR) biometrically registered 900,000 Rohingya people in Bangladesh after they left Myanmar as refugees.[133] Biometric registration was done in collaboration with the government of Bangladesh through a private vendor, highlighting potential concerns, like access to sensitive information, regarding function creep.[134]

Project Connect and UNICEF

Given the security risks of data breaches and data security practices as well as function creep, Project Connect and UNICEF's work provides an example of how to potentially treat the data of vulnerable populations collected from AI systems going forward. Project Connect and UNICEF are collaborating to create an online mapping tool of every school worldwide.[135] The tool relies in part on the application of machine learning and advanced data modelling. It includes real-time data on the connectivity of schools, with the ultimate goal being to improve educational development for children. The project recognizes that there are inherent security risks associated with the concept of mapping schools online and thus established a detailed framework on what data is publicly available. Their framework entails ranking countries or regions as a having a high, medium, or low risk regarding the possibility that the data could endanger children if used by malicious actors.[136] Their risk assessment is based on several relevant factors such as global terrorism and fragile states indexes and expert

130 Madianou, "The Biometric Assemblage," 8.
131 Ibid, 11.
132 Chris Butler, interview with author, June 24, 2019.
133 Madianou, "The Biometric Assemblage," 9.
134 Madianou, "The Biometric Assemblage," 9.
135 Project Connect, https://www.projectconnect.world (accessed August 9, 2019).
136 Project Connect, https://www.projectconnect.world (accessed August 9, 2019).

opinion in field offices. Data can be made public for low risk countries, contingent on the provider of that data opting in for its release. For countries deemed medium or high risk, Project Connect and UNICEF have developed thoughtful procedures for determining what data to publicize. For instance, a committee will be set up per country by Project Connect and UNICEF that will for most countries engage host governments and data providers in the decision-making process.[137] In addition, data sharing will abide by country and international level laws in respect to data privacy. This project exemplifies how security risks could be mitigated when working with AI systems for vulnerable populations in the humanitarian field. In particular, they could be mitigated because Project Connect and UNICEF have developed a detailed, accessible framework that takes into account relevant security risks, stakeholders, and legal requirements while other humanitarian actors have not made this information clear or have failed to do so. Interestingly, the framework also explicitly mentions that data a provider asks to remain private will not be shared without their consent.[138]

137 Ibid.

138 Ibid.

Section Four:

Issues with Data Consent

Issues with consent over data collection is a third way that AI systems for vulnerable populations can undermine the role of humanitarian actors by leaving vulnerable populations at further risk of vulnerability. This section examines perspectives on issues of consent in this context. Specifically, it discusses consent and its relation to security risks; the ideas of meaningful consent, captured consent, and tradeoffs; parallel AI and human-based systems; and the concept of consent as meaningless.

Consent and its Relation to Security Risks

First, issues of consent are inherently tied to security risks. It can be very expensive to collect data to train algorithmic models for AI systems. To circumvent the cost issue, function creep can occur where an organization uses data sets from other sources.[139] These secondary uses of data relate to consent because a vulnerable person may be willing to give a humanitarian agency their data for a particular purpose, but they may be unwilling for that data to be used outside of that context since it may put them at risk.[140] Related to this point on function creep, when data is fed into an algorithm, even if certain identifiable markers are made invisible, there is no guarantee that in the data lifecycle the data will not become identifiable.[141] If data sets that are potentially identifiable are then used for secondary purposes, this could exacerbate the problem of function creep. Considering the potential for function creep, it could be said that consent over data collection would only be plausible in the context of a particular intent.[142] For instance, having safeguards to ensure that the data collected for the purpose a person consented to is actually the only purpose it is used for. However, consent may not be useful when looking at the long-term possibility that quantum computers may be able to decode encrypted data or do things with data currently thought to be unimaginable. Moreover, since many AI systems are "black boxes", it is difficult to reconcile the concept of consent with the idea of using data for a particular intent.[143]

139 Sarah Myers West, interview with author, July 22, 2019.
140 Sarah Myers West, interview with author, July 22, 2019.
141 Mila Romanoff, interview with author, July 24, 2019.
142 Lambert Hogenhout, interview with author, June 17, 2019.
143 Lambert Hogenhout, interview with author, June 17, 2019.

Meaningful Consent, Captured Consent, and Tradeoffs

In Global North and humanitarian contexts, do people, whether vulnerable or not, give meaningful consent over their data collection? Is there really a choice in opting out of the use of AI systems? In the Global North, many feel they are captured in an environment where they do not have a choice but to consent to their data being used in order to enjoy AI systems like smartphones in everyday life.[144] The inability to really opt-out of these systems is amplified in humanitarian contexts. Regarding biometric registration systems, if a vulnerable person refuses to register their biometric data they do not receive aid since they need to register to be counted on distribution lists.[145] Since people are usually vulnerable for a reason and rely on humanitarian aid to survive, especially if they have a family to support, there is no meaningful consent, but rather captured or coercive consent in these contexts. In other words, though it can be said that a person needs to have agency to give their consent, vulnerable people do not have much agency due to their vulnerable status and thus cannot freely choose whether to give consent over the collection of their data.[146] Moreover, a UN internal audit report notes that refugees were not given enough information in order to make an informed decision when consenting to the use of their biometric data.[147] If there is no clearly informed consent, there may not be any consent at all.

A lack of informed consent is all the more dangerous since it does not seem like a big trade off to give personal data in order to receive food or cash transfers.[148] In fact, it may seem like a reasonable exchange. Some humanitarian organizations argue that by being able to monitor the supply chain more accurately, AI systems can combat fraud and the diversion of their humanitarian resources from vulnerable populations. Implementing AI systems may be a more efficient method of providing aid, saving organizations money that they can then use to buy further resources for vulnerable people. However, there is a tradeoff here, specifically between aid efficiency and the potential biases, security risks, and lack of consent which put vulnerable populations at further risk of harm. What should we value more?

144 Rumman Chowdhury, interview with author, July 16, 2019.
145 Madianou, "The Biometric Assemblage," 12.
146 Chris Butler, interview with author, June 24, 2019.
147 Madianou, "The Biometric Assemblage," 12.
148 Rumman Chowdhury, interview with author, July 16, 2019.

Parallel AI and Human-based Systems

Having parallel AI and human-based systems may reduce the problems associated with issues of consent. If there are strict protocols in place that specify that a vulnerable person's data would be eliminated within a certain time frame and they can ask for its removal within that time frame, that would be the start of an ethical system of consent.[149] One way of contributing to this potentially more ethical system is ensuring the ability for participants to opt-out and still use the same services.[150] Another is checking with them at multiple points of time to see if they still consent to their data being used,[151] since they are likely to be under duress when asked for consent. If people do not consent to submitting their data, parallel systems can be created that ensure vulnerable people are able to access the same services. Though in the humanitarian space and more broadly there is this push towards automating previously human-run and paper-based systems, human-based systems can provide benefits when AI systems can cause further risks in disaster response contexts. For example, paper-based voting systems may be safer than electronic voting systems because paper is difficult to hack and leaves an audit trail.[152] Thus in some cases, there are fewer security risks from using paper-based systems though on a large scale they may be less environmentally-friendly and more expensive. Similarly, AI systems should be designed with the recognition that some procedures may make vulnerable people feel more secure. In disaster contexts, having an official paper may give someone a sense of security. It could be a kindness.[153] Having parallel systems then could reduce problems associated with consent over data collection. It could ensure people are not in a captured or coercive environment of consent and actually have a choice of opting out or using human-based systems while accessing the same services.

Consent as Meaningless

Another perspective on consent is the concept of consent as meaningless. When people need humanitarian aid, they are almost certainly going to give their consent over the use of their data, making consent a meaningless concept.[154] It may be more insightful to focus on the context in

149 Os Keyes, interview with author, July 22, 2019.
150 Os Keyes, interview with author, July 22, 2019.
151 Sarah Myers West, interview with author, July 22, 2019.
152 Meredith Broussard, interview with author, June 13, 2019.
153 Meredith Broussard, interview with author, June 13, 2019.
154 Cathy O'Neil, interview with author, July 15, 2019.

which an organization is allowed to use data from a vulnerable population rather than looking at if people consented. In other words, there needs to be an awareness that people do not have much negotiating power over the ownership of their data, and the people who do have negotiating power will not be the vulnerable populations who are particularly impacted.[155] Accordingly, "the burden should be on the companies [or humanitarian organizations] that are using data, not the people whose data is being exploited".[156]

Since so much data is already collected on people, vulnerable or not, it may be better to look into enforcing policy and regulatory frameworks to ensure data mining practices or security risks do not obstruct people's rights.[157] Not only should there be a focus on enforcing existing laws, but also on enforcing new laws that have been created in response to issues of consent. For example, there is an Illinois law called the Biometric Information Privacy Act (BIPA) that was enacted in October 2008 to address issues of consent in relation to biometrics.[158] The law states that a private entity using biometric technology must create a publicly accessible written policy outlining how long they will retain a person's biometrics, for up to a period of three years, before they destroy it. The law also claims that private entities cannot profit from a person's biometric information, they cannot disclose a person's biometric information except in select circumstances, and they must establish reasonable safeguards for storing, transmitting, and protecting such information. If consent is violated, either intentionally or negligently, a person is entitled to a small sum of money, commonly $1,000 or $5,000, from the private entity if the person wins their case in a state or federal court.

The BIPA can be a useful model for thinking about consent in relation to AI systems because it aims to allow a person's biometrics to be destroyed after a certain period of time, which reduces the likelihood of function creep and other security risks occurring. It also establishes data privacy and security protocols for the treatment of such sensitive information. However, the law can still foster a captured or coercive consent environment. Even though someone knows when their data will be permanently destroyed, their data is still being collected by a company when they may not have much of a choice to opt-out in the first place. Security breaches can still occur during the period of time in which a person's data is kept. Moreover, vulnerable people may not have access to the same

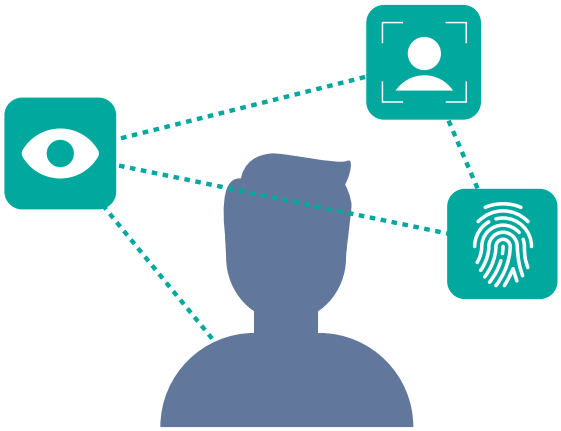
155 Cathy O'Neil, interview with author, July 15, 2019.

156 Ibid.

157 Ibid.

158 Sarah Myers West, interview with author, July 22, 2019.

forms of redress, such as filing a lawsuit, as people in Western contexts if their consent is violated. The downsides to new laws like the BIPA help to reinforce the concept of consent as ultimately meaningless.



Recent examples of AI systems used in humanitarian contexts highlight problems with consent, suggesting that the concept may be meaningless. For instance, the introduction of this report opened with the example of an international organization's desire to implement a biometric registration system in Yemen. The WFP was forced to partially suspend food aid to 850,000 people in Sana'a from June to the beginning of August 2019[159] since the Houthi rebels argued that it was against the law in Yemen to collect biometric data. The biometric systems were intended to more efficiently monitor the supply chain process, ensuring food is not diverted from its targeted recipients and used as a political tool in civil conflicts.[160] Biometrics can easily confirm that the intended recipient received aid. This example underlines an asymmetry in humanitarianism. On the one hand, implementing biometric systems can ensure more vulnerable people get the food they were intended to receive. On the other hand, it could be said that the Houthis did not have much choice but to eventually agree to the proposal because the distribution of aid was threatened.[161] Consent over data control does not really exist in this case because to refuse to register means to refuse aid. Essentially, "the conditions of humanitarian assistance turn consent into coercion in most circumstances".[162]

159 Welsh, "Biometrics disagreement leads to food aid suspension in Yemen," 2019.

160 Welsh, "Biometrics disagreement leads to food aid suspension in Yemen," 2019.

161 Mirca Madianou, interview with author, July 8, 2019.

162 Mirca Madianou, interview with author, July 8, 2019.

Problems with consent are not only visible in Yemen. In 2018, Rohingya refugees living in several refugee camps, including the Unchiprang camp, in the Cox's Bazar district in southeastern Bangladesh participated in a hunger strike over an international organization's desire to implement a biometric registration system.[163] [164] [165] Refugees were concerned about the possibility of their biometric data becoming accessible to the Myanmar government, which would create a digital record that refugees could be verified against. This record may lead to their repatriation or further persecution.[166] [167] In spite of the concerns of refugees, biometric registrations continued and there was little policy change.[168] The persistent use of this AI system suggests that registrations continued in spite of a lack of consent. Thus in both Yemen and Bangladesh, vulnerable populations face coercive consent or a lack of consent completely in relation to having their biometric information taken. These examples support the viewpoint that in practice, consent is becoming meaningless.

163 Welsh, "Biometrics disagreement leads to food aid suspension in Yemen," 2019.

164 Mirca Madianou, interview with author, July 8, 2019.

165 Radio Free Asia, "Rohingya refugees protest, strike against smart ID cards issued in Bangladesh camps," ***refworld***, November 26, 2018, https://www.refworld.org/docid/5c2cc3b011.html (accessed September 23, 2019).

166 Welsh, "Biometrics disagreement leads to food aid suspension in Yemen," 2019.

167 Mirca Madianou, interview with author, July 8, 2019.

168 Mirca Madianou, interview with author, July 8, 2019.

Section Five:

AI Principles and Recommendations

This report has argued that AI systems that have biases, security risks, and issues with data consent can undermine the role of humanitarian agents in disaster contexts by leaving aid recipients at further risk of vulnerability. This final section puts forth AI principles and recommendations to help decision-makers and community members reflect on the use of AI systems for vulnerable people and how we can collectively try to reduce harm. In particular, this section outlines four AI principles, and then discusses further recommendations, including general recommendations as well as those related to changing business models.

AI Principles

Before discussing AI principles specifically, it is important to note that principles in this report do not refer to a definitive checklist suggesting that if one is able to take into account all of these principles at one moment in time, they will not have biases, security risks, or issues of consent in their AI systems afterwards.[169] [170] Though there are so many AI principles already published, principles are only guidelines, meant to highlight certain things to look out for.[171] These principles in particular are meant to inspire reflection on the needs of vulnerable people, reflection being something that is situationally aware and in recognition of the ongoing changes in a situation over time.[172] [173] Though these AI principles may be met at a certain point in time, when the situation changes, these principles may be unfulfilled in the future.[174] Moreover, creating a fixed set of principles risks masking new principles that may be relevant at another time.[175] These principles thus encourage a mindset cognizant of the fact that AI systems must be reflected upon and subject to revision over time.

169 Sarah Myers West, interview with author, July 22, 2019.
170 Zimmermann and Zevenbergen, "AI Ethics: Seven Traps," 2019.
171 Rumman Chowdhury, interview with author, July 16, 2019.
172 Sarah Myers West, interview with author, July 22, 2019.
173 Zimmermann and Zevenbergen, "AI Ethics: Seven Traps," 2019.
174 Sarah Myers West, interview with author, July 22, 2019.
175 Zimmermann and Zevenbergen, "AI Ethics: Seven Traps," 2019.

1. Weigh the benefits versus the risks: Avoid AI if possible

The first principle is avoiding the use of AI systems if possible when dealing with vulnerable populations. There needs to be a better vetting process in place to determine if AI systems are actually needed in humanitarian or disaster response situations before deploying them. This process would entail discussing the positive and negative consequences of deploying such a system as well as how to mitigate the potential risks.[176] If the risks seem to outweigh the supposed benefits, AI systems should not be used. More efficient and less expensive technologies do not necessarily equate with more beneficial outcomes. It would be fairer to create a space where designers and developers of AI systems need to justify why their system is needed and better than non-AI systems. This space would be more just than the current space which entails trying to mitigate biases, prevent security risks, and creating consent mechanisms after the AI systems have already been developed or deployed.[177] In other words, the paradigm should be shifted to insist that the people developing these systems have the burden to prove that their systems are not increasing the risks for vulnerable people, rather than the burden resting on the vulnerable people themselves who are oftentimes being harmed.[178]

To evaluate whether AI is actually needed, a recommendation is creating a group or innovation cell where people using different methodologies in a multi-stakeholder context could better detect issues of bias, security risks, and consent in AI systems for vulnerable people specifically. This group could be one of the multi-stakeholder foresight groups that Eleonore Pauwels discusses in her report, "The New Geopolitics of Converging Risks: The UN and Prevention in the Era of AI" (2019).[179] She recommends the creation of a global foresight observatory for AI convergence, directed by the UN, which would focus on how AI and other related emerging technologies like blockchain could be designed and influenced to "meet the ethical needs of a globalizing world".[180] This observatory would include a "foresight fusion cell", consisting of an in-house team of technology and policy experts that study the implications of using AI technology.

176 Os Keyes, interview with author, July 22, 2019.

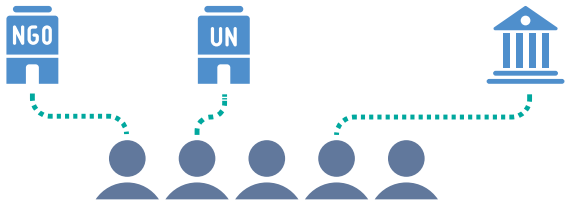
177 Os Keyes, interview with author, July 22, 2019.

178 Cathy O'Neil, interview with author, July 15, 2019.

179 Eleonore Pauwels, "The New Geopolitics of Converging Risks: The UN and Prevention in the Era of AI," *United Nations University Centre for Policy Research*, April 29, 2019, https://i.unu.edu/media/cpr.unu.edu/attachment/3472/PauwelsAIGeopolitics.pdf (accessed August 9, 2019).

180 Pauwels, "The New Geopolitics of Converging Risks," 2019, 53.

The core fusion cell would support a collection of foresight groups, with cross-sectorial actors, including academia, civil society, UN agencies, as well as carefully vetted private companies that could conduct foresight analysis on a variety of topics.[181]



If AI systems for vulnerable populations is one of these multi-stakeholder foresight groups, it could be a more inclusive way of deciding when AI systems are needed in humanitarian contexts. People in the foresight group could include AI developers and security experts who know about designing AI systems and how to mitigate security risks and potential data misuse. They could also include experts on specific vulnerable populations to help determine potential biases, and how to ensure, if possible, informed consent or the guarantee of services if consent is not received.[182] Most importantly, such a group would need to be a safe space for vulnerable populations and actors in innovation ecosystems in the Global South to have their voices heard and their concerns taken seriously.[183] A multi-stakeholder group of this caliber would help counterbalance the uneven distribution of power between the creators of AI systems and their users. To aid in this goal, the increasing number of UN Innovation Labs could collaborate with innovation ecosystems in the Global South specializing in AI to strengthen bottom-up approaches in determining if AI systems are actually needed for vulnerable populations.[184]

A foresight group specific to AI for vulnerable people could also look into establishing or adhering to an engineering Hippocratic Oath, which was one of the suggestions discussed at the UN Secretary-General's recent High-level Panel on Digital Cooperation.[185] The medical community has a Hippocratic Oath for adhering to ethical practices in their profession. Engineering ethics,

181 Pauwels, "The New Geopolitics of Converging Risks," 2019.
182 Eleonore Pauwels, interview with author, July 10, 2019.
183 Eleonore Pauwels, interview with author, July 10, 2019.
184 Pauwels, "The New Geopolitics of Converging Risks," 2019.
185 Interviewee, interview with author, July 16, 2019.

while a discipline, needs to continue to evolve to fully recognize the risks of AI. Enforcing a strong ethical component in the education of engineers as well as in a potential collaborative group for vulnerable people could help reframe AI as a technology used sparingly in humanitarian contexts. Ultimately, as suggested by the ICT4Peace foundation and the Zurich Hub for Ethics and Technology, humans, rather than AI itself, must be the focal point in developing AI systems going forward,[186] which a foresight group centered on AI for vulnerable people can help accomplish.

2. Use AI systems that are contextually-based

Using AI systems that are contextually-based is another principle to adhere to. The ability to effectively minimize biases and security risks, as well as manage issues over data consent rely on designing AI systems for specific vulnerable populations and their needs within a humanitarian context. For example, when analyzing the feedback databases of multiple humanitarian organizations, the Humanitarian Technologies Project found that a majority of the feedback were thank you messages, rather than providing advice on how aid mechanisms could be improved.[187] In the culture of the Philippines, gratitude towards humanitarian agencies is an ingrained concept while more critical feedback is a Western concept, which is why such feedback was rarely received as planned.[188] This finding highlights that AI systems need to be designed with contextuality in mind. Designing generalized systems for all vulnerable populations will probably have a much lower accuracy rate or miss the goals or outcomes of the project. In fact, when AI systems are created without attention towards local contexts, the vulnerability of vulnerable populations can significantly increase, especially in relation to discrimination, bias, security and consent concerns.[189]

Underlying a lack of contextuality in developing AI systems is the belief that technological solutions can solve societal problems. It is deeply problematic for AI developers to think it is acceptable to use an existing AI system for a different vulnerable population, or think that because a system

186 Barbara Weekes, "Digital Human Security 2020: Human security in the age of AI: Securing and empowering individuals," ***ICT4Peace Foundation***, December 21, 2018, https://ict4peace.org/wp-content/uploads/2018/12/Digital-Human-Security-Final-DSmlogos.pdf (accessed August 9, 2019).

187 Humanitarian Technologies Project, http://humanitariantechnologies.net (accessed August 9, 2019).

188 Humanitarian Technologies Project, http://humanitariantechnologies.net (accessed August 9, 2019).

189 Arun, "AI and the Global South," 13.

worked in one environment, it should work in another with minimal need of adjustment.[190] This thought process is based on a belief in technological solutionism, that technology or AI systems specifically can "solve" problems vulnerable populations face, when these are contextually-based social and political problems.[191] Technological solutionism is unsurprising considering the complex nature of humanitarian emergencies and the desire to help vulnerable people as quickly as possible.[192] Solutionism is dangerous however when the need to come up with solutions to help people precedes thorough testing, including analyses of the local intricacies of the situation, which may not require AI systems at all.[193] Outside of the emergency context, technological solutionism is present in the broader technological community, and it includes treating social problems like bias as a statistical issue rather than looking at the wider problem of discrimination in specific contexts.[194] If people in AI and humanitarian communities start treating the problems they are trying to apply AI systems to as social problems, there will be a greater understanding that these systems are going to vary between different situations as well as within a single situation as a result of an AI presence.[195] This contextual and social awareness can create adaptive AI design spaces where vulnerable populations are better served by the systems developed to help them.

3. Empower and include local communities in AI initiatives

A third principle is empowering and including local communities in AI initiatives for vulnerable people. The UN Secretary-General's High-level Panel on Digital Cooperation acknowledged the need for inclusivity when it comes to digital technologies, which involves taking into account local conditions and challenges faced by vulnerable groups.[196] The Panel also highlighted the need for capacity-building, especially in Global South communities, so multi-stakeholder collaborations, like a foresight group, can contribute to decision-making process on emerging technologies.[197] The more capacity-building that takes place, for instance, improving education systems in Global South countries and rural areas to encourage the development of AI and humanitarian experts

190 Rumman Chowdhury, interview with author, July 16, 2019.
191 Os Keyes, interview with author, July 22, 2019.
192 Madianou, "The Biometric Assemblage," 7.
193 Madianou, "The Biometric Assemblage," 7.
194 Sarah Myers West, interview with author, July 22, 2019.
195 Os Keyes, interview with author, July 22, 2019.
196 Arun, "AI and the Global South," 14.
197 Arun, "AI and the Global South," 14.

from broad geographics, the less bias, security risks, and issues of consent are going to be inherent in AI systems.[198]

In other words, "You need local populations to be empowered enough to be able to intervene in the design space of technology. We don't do that well (yet). We need an incentive to move from the exploitation of data to empowerment with data".[199] People can create empowerment by investing in digital literacy and capacity-building, including social capital and civil society organizations more generally.[200] These long-term projects are crucial if local communities are to have a voice in the AI initiatives targeted towards them or vulnerable groups in their communities. However, local communities and vulnerable people do not need to wait to be involved in decision-making processes for AI systems. They can be involved now, and should have their voices heard throughout the design and development process, rather than only in the pilot stage, in ways that are considerate of their time and resources.[201] In particular, community members and vulnerable people should have their voices heard in person rather than through digital platforms. The Humanitarian Technologies Project found that people gain validation and recognition from having face-to-face conversations rather than providing input digitally where they may not receive any acknowledgement or response.[202] Civil society organizations can also encourage local people to make their voices heard in collaborative settings.[203] Empowering and including local communities in all of the processes involved in creating AI systems can reduce the risk of further vulnerability to already vulnerable people.

4. Implement algorithmic auditing systems

A fourth principle is implementing algorithmic auditing systems to serve as a third party check on the potential for biases, security risks, and problems with consent in AI technologies for vulnerable people. Algorithmic auditing systems can open up the "black boxes" of AI algorithms to make them explainable, which provides greater understanding as to how a particular decision

198 Interviewee, interview with author, July 16, 2019.
199 Eleonore Pauwels, interview with author, July 10, 2019.
200 Humanitarian Technologies Project, http://humanitariantechnologies.net (accessed August 9, 2019).
201 Sarah Myers West, interview with author, July 22, 2019.
202 Mirca Madianou, interview with author, July 8, 2019.
203 Humanitarian Technologies Project, http://humanitariantechnologies.net (accessed August 9, 2019).

was reached and what data features were used to get the result when using AI systems.[204] One example of a company that runs algorithmic auditing systems is ORCAA (O'Neil Risk Consulting & Algorithmic Auditing). ORCAA is a consulting company that provides algorithmic auditing services for organizations in recognition of the fact that unaudited algorithmic models are not objective and can harm marginalized people. The company uses a four-step process to audit existing algorithms for "accuracy, bias, consistency, transparency, fairness and timeliness".[205] The process includes examining the design of the algorithm, matching the design to what the algorithm is used for, an execution audit involving testing the algorithm using real world examples, and reporting on the process, including suggestions for improvement.[206] ORCAA wants to help companies and organizations comply with existing laws including anti-discrimination laws in the AI field where there currently is not a lot of regulation.[207] If there is not an active effort to influence the field, one argument is that regulators may create guidelines that do not take a variety of perspectives into account.[208]

Besides ORCAA, there are other organizations in the new space of algorithmic auditing as companies and organizations cross-sectorally realize the risks posed by AI systems and the need to mitigate them. Another such organization is the Algorithmic Justice League, founded by Joy Buolamwini, which highlights and reports algorithmic biases, as well as provides algorithmic auditing services.[209] Furthermore, there have been advancements in addressing the related issue of explainability in AI systems, which can help find and mitigate AI risks like bias. One technique is local interpretable model-agnostic explanations (LIME). LIME examines specific parts of a data set, like the nose and then ears in facial recognition, and records changes in the model's predictions to identify what factors affect the algorithm's decision-making process.[210] Advancements in explainability, while falling short of an algorithmic audit and not being a solution to the oftentimes social problems that make AI systems ineffective, could help identify potential biases and promote greater accountability when using these systems.[211]

---

204 Jake Silberg and James Manyika, "Notes from the AI frontier: Tackling bias in AI (and in humans)," ***McKinsey Global Institute***, June 2019, https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/MGI-Tackling-bias-in-AI-June-2019.ashx (accessed August 9, 2019).

205 ORCAA, http://www.oneilrisk.com (accessed August 9, 2019).

206 ORCAA, http://www.oneilrisk.com (accessed August 9, 2019).

207 Cathy O'Neil, interview with author, July 15, 2019.

208 Cathy O'Neil, interview with author, July 15, 2019.

209 Algorithmic Justice League, https://www.ajlunited.org (accessed August 9, 2019).

210 Silberg and Manyika, "Notes from the AI frontier: Tackling bias in AI (and in humans)," 2019.

211 Silberg and Manyika, "Notes from the AI frontier: Tackling bias in AI (and in humans)," 2019.

General AI Recommendations

Besides specific AI principles to adhere to, there are some general recommendations to consider if using AI systems for vulnerable populations.

Intent. There should be a clear intent when designing AI systems and in data collection. What is the intention of collecting data for this AI system, and even if the technological landscape changes so data can be manipulated in different ways, is the same intent met?[212] The intent of data collection, data analysis, and AI systems more broadly should not change over time because as soon as intent changes, any consent received from the target population becomes inapplicable since they did not agree to the new intent.[213]

Retention schedules. Another related recommendation is having strict record retention schedules, as enforced in the Illinois' BIPA, in order to respect that a person's consent over their data collection may change over time.[214]

Receiving aid is not based on consent. Likewise, regarding consent, a key recommendation is ensuring that "the receipt of the relief [should] not [be] contingent on the consent or use of the data".[215]

Benefits of data collection should outweigh risks. Similarly for security risks, "the collection of that data should not bring on greater risk to the individuals". In other words, the benefits of data collection should far outweigh any potential risks of sensitive information being stolen or used for malicious intent.[216] More generally, in humanitarian contexts there is a saying that "perfect is the enemy of done."[217] What this refers to is that often there is a choice when working in humanitarian or disaster response situations between doing "the best job" and trying to improve a problem, even if the outcome is not perfect.[218]

212 Lambert Hogenhout, interview with author, June 17, 2019.
213 Lambert Hogenhout, interview with author, June 17, 2019.
214 Meredith Broussard, interview with author, June 13, 2019.
215 Kate McBride, interview with author, July 9, 2019.
216 Kate McBride, interview with author, July 9, 2019.
217 Miguel Angel Hernandez Rivera, interview with author, July 18, 2019.
218 Miguel Angel Hernandez Rivera, interview with author, July 18, 2019.

Leave no negative footprint. Humanitarian actors should try to improve a problem without creating another problem or a negative footprint for the future.[219] This recommendation is a humanitarian principle that can easily be applied to AI systems for vulnerable people. Indeed, perhaps core humanitarian principles and codes of conduct should inform the treatment of vulnerable populations if and when using AI systems.[220]

Business Model AI Recommendations

Throughout the interview process, it became clear that private sector companies could help create more ethical AI systems for vulnerable people if they had the incentives to do so. To hear different perspectives on this idea, the question posed to interviewees was: How do you think, if possible, we can change business models to not only be incentivized by profit but by the desire to create fair, unbiased algorithms (to the greatest extent possible) in the humanitarian field?

One perspective on this topic was creating an incentive structure that impacts a company's bottom line, which is influenced by efficiency gains or profitability.[221] Influencing a company's bottom line can be difficult to achieve with the centralization of short term goals in many corporate environments.[222] However, notions of profitability are changing since reputational risks in the current "cancel culture" can make an organization lose customers if they are not viewed as ethical.[223] The point of reputational risks, or even potentially paying a fine, is that these are forms of leverage tied to profitability to influence companies to want to support less biased, more secure AI systems.[224]

To ensure that companies are not driven to use AI systems for profit maximization, there need to be strong government policies that regulate issues of bias, security risks, and consent, and ensure that they can be audited.[225]

There also needs to be advocacy work underlining the potential reputational and thus profitability

219 Ibid.
220 Kate McBride, interview with author, July 9, 2019.
221 Rumman Chowdhury, interview with author, July 16, 2019.
222 Os Keyes, interview with author, July 22, 2019.
223 Rumman Chowdhury, interview with author, July 16, 2019.
224 Cathy O'Neil, interview with author, July 15, 2019.
225 Interviewee, interview with author, July 16, 2019.

risks associated with using poorly designed AI systems, and highlighting the long-term benefits of using well designed systems.[226]

In a humanitarian space specifically, it should theoretically be easier to align business models with ethical AI systems, especially if the AI systems are designed in-house, because the objective is providing humanitarian aid rather than profit maximization.[227] When AI systems are not designed in-house, it is important that humanitarian organizations thoroughly vet the third party companies they work with to ensure they do not unintentionally harm vulnerable populations by using poorly designed systems.[228] In essence, "the best hope we have is that funded initiatives at places like the UN or other kinds of humanitarian relief agencies actually care more about fairness and ethics than they care about profit".[229]

226 Interviewee, interview with author, July 16, 2019.
227 Cathy O'Neil, interview with author, July 15, 2019.
228 Interviewee, interview with author, July 16, 2019.
229 Cathy O'Neil, interview with author, July 15, 2019.

Conclusion

With many recent examples of AI systems being used for vulnerable populations, there is a newfound urgency to reflect on the implications of using this technology in humanitarianism and disaster response. This report argued that AI systems with prevalent biases, security risks, and issues with consent can undermine the role of humanitarian actors in disaster contexts by leaving aid recipients at further risk of vulnerability. Support for the argument was built throughout the first four sections of the report: current examples of AI systems used for vulnerable people in humanitarian contexts; different forms of biases pertaining to AI systems; the security risks of data breaches, data security practices, and function creep; and issues of consent over data collection. Based on the points raised in these sections, the fifth section developed AI principles and recommendations specific to vulnerable people in the humanitarian field. These principles are: avoid AI if possible, use AI systems that are contextually-based, empower and include local communities in AI initiatives, and implement algorithmic auditing systems. While AI systems currently risk leaving vulnerable populations more vulnerable, greater awareness and reflection on these principles can positively impact the measured adoption of AI systems in the humanitarian field in years to come.

Annex I:

List of interviews

This report was largely informed by semi-structured interviews with experts in the AI and humanitarian fields. A heartfelt thank you to all interviewees for their time and research recommendations. The report could not have been completed without them.

Name	Title and Organization
Ronald Ashri	Co-founder of GreenShoot Labs
Meredith Broussard	Associate Professor at New York University; author of ***Artificial Unintelligence: How Computers Misunderstand the World***
Chris Butler	Chief Product Architect, IPsoft Inc.
Rumman Chowdhury	Responsible Artificial Intelligence Lead, Accenture
Allison Gardner	Co-founder of Women Leading in AI; Programme Director of the Data Science Degree Apprenticeship at Keele University
Sanjana Hattotuwa	Special Advisor, ICT4Peace Foundation; PhD candidate at the University of Otago, New Zealand
Lambert Hogenhout	Chief Analytics, Innovation and Partnerships, UN Office for Information and Communications Technology (OICT)
David Michael Kelly	Strategic Planning, UN Executive Office of the Secretary-General (EOSG)
Os Keyes	PhD student at the University of Washington; Ada Lovelace Fellow
Mirca Madianou	Media, Communications and Cultural Studies Professor at the University of London; Principal Investigator for the Humanitarian Technologies Project
Kate McBride	Chief, UN Regional Technology Centre - Africa

Cathy O'Neil	Founder of ORCAA; author of ***Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy***
Jurij Paraszczak	Consultant, IBM Corporate Citizenship/CSR
Eleonore Pauwels	Research Fellow on AI and Emerging Cybertechnologies, UN University Centre for Policy Research; Director of the AI Lab, Woodrow Wilson International Center for Scholars
Miguel Angel Hernandez Rivera	Information Management Officer, UN Office for the Coordination of Humanitarian Affairs (OCHA)
Mila Romanoff	Data Privacy and Data Protection Legal Specialist, UN Global Pulse
Daniel Stauffacher	Founder and President of the ICT4Peace Foundation; President of the Zurich Hub for Ethics and Technology (ZHET)
Regina Surber	Scientific Advisor, ICT4Peace Foundation; PhD candidate at the University of Zurich, Switzerland
Kush Varshney	Principal Research Staff Member and Manager, IBM
Sarah Myers West	Postdoctoral Researcher, AI Now Institute

Annex II:

References

Al Jazeera and News Agencies. "Yemen's Houthis, WFP reach deal to resume food relief." Al Jazeera, August 4, 2019. https://www.aljazeera.com/news/middleeast/2019/08/yemen-houthis-wfp-reach-deal-resume-food-relief-190804133835009.html (accessed August 9, 2019).

Algorithmic Justice League. https://www.ajlunited.org (accessed August 9, 2019).

Amnesty International and Access Now. "The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems." Access Now, May 16, 2018. https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/ (accessed August 9, 2019).

Arun, Chinmayi. "AI and the Global South: Designing for Other Worlds." In The Oxford Handbook of Ethics of AI (forthcoming), edited by Markus D. Dubber, Frank Pasquale, and Sunit Das, 1-15. Oxford University Press, 2019.

Broussard, Meredith. Artificial Unintelligence: How Computers Misunderstand the World. Cambridge, MIT Press, 2018.

Buolamwini, Joy, and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." Proceedings of Machine Learning Research 81, (2018): 1-15.

Chen, Lily, Stephen Jordan, Yi-Kai Liu, Dustin Moody, Rene Peralta, Ray Perlner, and Daniel Smith-Tone. U.S. Department of Commerce, National Institute of Standards and Technology. Report on Post-Quantum Cryptography. April 2016, NISTIR 8105. http://dx.doi.org/10.6028/NIST.IR.8105.

Google AI. "Artificial Intelligence at Google: Our Principles." Google AI. https://ai.google/principles/ (accessed August 9, 2019).

Humanitarian Technologies Project. http://humanitariantechnologies.net (accessed August 9, 2019).

IBM. “IBM’s Principles for Trust and Transparency.” IBM, May 30, 2018. https://www.ibm.com/blogs/policy/trust-principles/#C3 (accessed August 9, 2019).

Kirchner, Lauren. “New York City Moves to Create Accountability for Algorithms.” ProPublica, December 18, 2017. https://www.propublica.org/article/new-york-city-moves-to-create-accountability-for-algorithms (accessed August 9, 2019).

Madianou, Mirca. “The Biometric Assemblage: Surveillance, Experimentation, Profit and the Measuring of Refugee Bodies.” Television and New Media 20 (2019): 1-19. doi: 10.1177/1527476419857682.

Madianou, Mirca, Jonathan Corpus Ong, Liezel Longboan, and Jayeel S. Cornelio. “The Appearance of Accountability: Communication Technologies and Power Asymmetries in Humanitarian Aid and Disaster Recovery.” Journal of Communication 66, no. 6 (2016): 960-981. doi:10.1111/jcom.12258.

Microsoft. “Microsoft AI principles.” Microsoft. https://www.microsoft.com/en-us/ai/our-approach-to-ai (accessed August 9, 2019).

ORCAA. http://www.oneilrisk.com (accessed August 9, 2019).

Paul, Deanna. “A maker of police body cameras won’t use facial recognition yet, for two reasons: Bias and inaccuracy.” The Washington Post, June 28, 2019. https://www.washingtonpost.com/nation/2019/06/29/police-body-cam-maker-wont-use-facial-recognition-yet-two-reasons-bias-inaccuracy/ (accessed August 9, 2019).

Pauwels, Eleonore. “The Ethical Anatomy of Artificial Intelligence.” United Nations University: Centre for Policy Research, July 29, 2018. https://cpr.unu.edu/cpr-voices-the-ethical-anatomy-of-artificial-intelligence.html (accessed August 9, 2019).

Pauwels, Eleonore. “The New Geopolitics of Converging Risks: The UN and Prevention in the Era of AI.” United Nations University Centre for Policy Research, April 29, 2019. https://i.unu.edu/media/cpr.unu.edu/attachment/3472/PauwelsAIGeopolitics.pdf (accessed August 9, 2019).

Porcari, Enrica. “A statement on the WFP-Palantir partnership.” Insight by The World Food Programme, February 7, 2019. https://insight.wfp.org/a-statement-on-the-wfp-palantir-partnership-2bfab806340c (accessed August 9, 2019).

Powles, Julia. “New York City’s Bold, Flawed Attempt to Make Algorithms Accountable.” The New Yorker, December 20, 2017. https://www.newyorker.com/tech/annals-of-technology/new-york-citys-bold-flawed-attempt-to-make-algorithms-accountable (accessed August 9, 2019).

Project Connect. https://www.projectconnect.world (accessed August 9, 2019).

Radio Free Asia. “Rohingya refugees protest, strike against smart ID cards issued in Bangladesh camps.” refworld, November 26, 2018. https://www.refworld.org/docid/5c2cc3b011.html (accessed September 23, 2019).

Romeo, Nick. “The Chatbot Will See You Now.” The New Yorker, December 25, 2016. https://www.newyorker.com/tech/annals-of-technology/the-chatbot-will-see-you-now?reload=true (accessed June 3, 2019).

Silberg, Jake, and James Manyika. “Notes from the AI frontier: Tackling bias in AI (and in humans).” McKinsey Global Institute, June 2019. https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Artificial%20Intelligence/Tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/MGI-Tackling-bias-in-AI-June-2019.ashx (accessed August 9, 2019).

Simonite, Tom. “AI is the Future - But where are the Women?” Wired, August 17, 2018. https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/ (accessed August 9, 2019).

The Centre for Humanitarian Data. “Data Responsibility Guidelines (Working Draft).” United Nations Office for the Coordination of Humanitarian Affairs, March 2019. https://centre.humdata.org/wp-content/uploads/2019/03/OCHA-DR-Guidelines-working-draft-032019.pdf (accessed August 9, 2019).

United Nations. “Secretary-General’s Strategy on New Technologies.” United Nations, September 2018. https://www.un.org/en/newtechnologies/images/pdf/SGs-Strategy-on-New-Technologies.pdf (accessed August 9, 2019).

Weekes, Barbara. “Digital Human Security 2020: Human security in the age of AI: Securing and empowering individuals.” ICT4Peace Foundation, December 21, 2018. https://ict4peace.org/wp-content/uploads/2018/12/Digital-Human-Security-Final-DSmlogos.pdf (accessed August 9, 2019).

Welsh, Teresa. “Biometrics disagreement leads to food aid suspension in Yemen.” Devex, June 24, 2019. https://www.devex.com/news/biometrics-disagreement-leads-to-food-aid-suspension-in-yemen-95164 (accessed August 9, 2019).

Zimmermann, Annette, and Bendert Zevenbergen. “AI Ethics: Seven Traps.” Freedom to Tinker: research and expert commentary on digital technologies in public life, March 25, 2019. https://freedom-to-tinker.com/2019/03/25/ai-ethics-seven-traps/ (accessed August 9, 2019).

Zuiderveen Borgesius, Frederik. “Discrimination, artificial intelligence, and algorithmic decision making.” Council of Europe, 2018. https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73 (accessed August 9, 2019).