



A FRAMEWORK FOR THE ETHICAL USE OF ADVANCED DATA SCIENCE METHODS IN THE HUMANITARIAN SECTOR



DATA SCIENCE & ETHICS GROUP

Executive Summary

What is this document?

This is a framework for applying data science methods for humanitarian outcomes. It aims to provide a set of ethical and practical guidelines for humanitarian data collectors, users, and stakeholders to consider when applying data science for humanitarian work. This work is at the juncture of data science (in particular AI), ethics, responsible data management,¹ humanitarian innovation, and humanitarian principles and standards.

Who is this document for?

This framework provides a practical guide for a broad range of humanitarian data stakeholders ranging from users, collectors and enumerators on the ground to data scientists, project focal points, programme managers and donors. It can also assist private and civil sector actors that work in or are interested in working in humanitarian work, and can be a useful resource for academics, students and policy staff.

How was this document made?

This framework is the first output of the Humanitarian Data Science and Ethics Group (DSEG) - a multi-stakeholder group established in 2018. It is an outcome of a consultative process based on the discussions during DSEG meetings and resulting identified priorities. This document draws perspectives from programme staff from various areas of humanitarian work, such as camp coordination and camp management (CCCM); Cash; Shelter; Protection; Gender Based Violence (GBV) prevention; and Water, Sanitation and Hygiene (WASH). It refers extensively to existing academic and humanitarian documents and frameworks, combines research and interviews with various humanitarian field staff officers, and reflects inputs from a thorough review process involving diverse data stakeholders and experts. This document was compiled after an extensive desk review of documents presented in the Timeline Chart (Figure. 1), as well as employing a bottom-up approach to include stakeholder interviews with data scientists, AI experts, humanitarian programme staff, and ethics advocates. Interviewees include academics, humanitarians, and policymakers, to try and capture the interdisciplinary nature of this rich and evolving discussion.

What are the objectives of this document?

This framework is designed to highlight key ethical considerations and to provide a practical guide for exploring or implementing advanced data science methods to support humanitarian outcomes. While this document is not exhaustive – nor is it possible to highlight all known or unknown associated risks or potential ethical considerations at this nascent stage of data science applications in the humanitarian field – this document aims to provide technical and procedural considerations to mitigate the ethical risks that may arise in humanitarian data science work. Most importantly, in an era of rapidly developing technology, this document seeks to stimulate a much-needed conversation about the continued adherence to humanitarian principles and protection, and consolidates further research into the considerations by the humanitarian sector as technology evolves and new frameworks develop around this topic. The future of advanced data science has the potential to assist humanitarian efforts by making it more efficient, expedient, and potentially anticipatory instead of responsive. However, it is of the highest importance that these methods are used responsibly, ensuring that humanitarian organizations continue to protect, uphold the dignity of, and empower the people they aim to support.

¹ Data responsibility entails the safe, ethical, and effective management of data: Data Responsibility Guidelines: Working Draft, OCHA, 2019. Pp. 7. Available at <https://centre.humdata.org/wp-content/uploads/2019/03/OCHA-DR-Guidelines-working-draft-032019.pdf>.

Data Science and Ethics Group

The Humanitarian Data Science and Ethics Group (DSEG), informally established in June 2018, is an open group consisting of data scientists, humanitarians, and ethics advocates. DSEG convenes diverse voices aiming to create a preliminary shared understanding of the ethical issues arising from humanitarian data. It is designed to lead practical and operational discussions to better understand the potential (including identified but unintended) risks that may accompany the well-intentioned applications of data science methods. This practical approach is reflected in the varied stakeholder engagement of the open group.

The first DSEG meeting in June 2018 led to an early mapping of the most prominent concerns of the different humanitarian data stakeholders, which cemented DSEG's key focus areas:²

- Improving quality control of humanitarian actors' work when applying advanced data science methods, both by peer review and ground truthing theoretical model findings;
- Identifying accountability and to establish safeguarding procedures when using data science methods for humanitarian purposes;
- Supporting ways to better communicate the results of advanced models as well as to enhance the data literacy among end-users, i.e. decision-makers among governments, humanitarians and donors;
- Addressing how to tackle the over reliance on models.

One of the first outcomes of this meeting was the development of a simple peer review mechanism with corresponding guidance notes on rules of engagement for developers, reviewers and facilitators (see DSEG meeting minutes).³ Initially this process was focused on the technical review – i.e. the code of the models – and was orientated towards data scientists aiming to obtain peer review by other data scientists. However, through further discussions, it became clear that the guidance needed to be expanded upon to encapsulate the ethical and practical components as well. While legal considerations are undoubtedly crucial to the discussion, it was determined that due to different geographic legal regulations and status of stakeholders, DSEG is unable to advise on legal issues and we suggest stakeholders engage with legal experts on their own specific activities. Since DSEG's first meeting, the work stream on a peer-review mechanism has been advanced by OCHA's Centre for Humanitarian Data and they have since published "A Peer Review Framework for Predictive Analytics in Humanitarian Response: Draft for Consultation as of September 2019."⁴

Working to solve the wider problem: This work contributes to and overlaps with debates on responsible and ethical data collection and use, data protection, humanitarian innovation, and humanitarian principles and standards. Like previous technological innovations, there is an increased interest to use more advanced methods of data science in humanitarian work, however, this enthusiasm is matched with a parallel urgency to highlight the ethical considerations in a practical and accessible way to support humanitarians to better understand the challenges of introducing new technology or data science methods in the sector. Although most humanitarian data science discussions and projects are largely in the conceptual, exploratory or early pilot phase, we have the opportunity to be pro-active in critically assessing the new technology introduced to the sector and resulting consequences or changes.

2 DSEG Meeting Minutes: <https://www.hum-dseg.org/>.

3 See Action 4 in the DSEG Meeting Minutes: <https://www.hum-dseg.org/>.

4 The Centre for Humanitarian Data, United Nations Office for the Coordination of Humanitarian Affairs (OCHA), A Peer Review Framework for Predictive Analytics in Humanitarian Response: Draft for Consultation as of September 2019. (The Hague, 2019) Available from https://centre.humdata.org/wp-content/uploads/2019/09/predictiveAnalytics_peerReview_updated.pdf.

Acknowledgements

On behalf of the DSEG, this document was first primarily authored by:



The Displacement Tracking Matrix (DTM) is the United Nations International Organization for Migration's (IOM) system to track and monitor displacement and population mobility. It is designed to regularly and systematically capture, process and disseminate information to provide a better understanding of the movements and evolving needs of displaced populations, whether on site or en route.

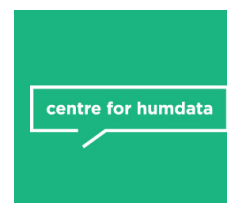


Data Science Initiative: The Data Science Initiative is a project of the City of The Hague. The mission of the Data Science Initiative is to harness the value of data science and artificial intelligence for peace, justice and security.

Authors: Kate Dodgson (Data Science Initiative), Dr. Prithvi Hirani, Rob Trigwell and Gretchen Bueermann (IOM DTM Global Team).

Contributors: Bill Drexel (IOM DTM Global Team), Amar Ashar - Assistant Director of Research & Jenna Sherman - Senior Project Coordinator (Ethics and Governance of AI Project - Harvard Berkman Klein Center for Internet & Society), Jacopo Margutti - Humanitarian Data Scientist (510 Red Cross), Jill Capotosto & Sam Corden (Jackson Institute for Global Affairs at Yale University), UN-OCHA Centre for Humanitarian Data, Translators without Borders, UNICEF, World Food Programme.

Front and back page photo credit: Michael Dziedzic.



With thanks to: The Netherlands Ministry of Foreign Affairs and The Municipality of The Hague



Ministry of Foreign Affairs



Version I:
April 2020

Contents

1.	Introduction	I
2.	The Lifecycle of a Humanitarian Data Science Project	6
2.1	Humanitarian Data Science	7
2.2	STAGE 0: Fundamentals	10
2.2.1	AI Ethics	10
2.2.2	Humanitarian Principles and Ethics	12
2.2.3	Data Responsibility	13
2.2.4	Human Rights	14
2.2.5	Risk Mitigation Guidance	18
2.3	STAGE 1: Problem and Solutions Exploration	20
2.3.1	Issues and Concerns	20
2.3.2	Practical Guidelines	20
2.4	STAGE 2: Data Journey - Collection, Protection and Processing	23
2.4.1	Data Collection	23
2.4.2	Data Processing and Protection	26
2.5	STAGE 3: Algorithm	29
2.6	STAGE 4: Reliance on Outputs	33
2.6.1	False negatives and false positives	33
2.6.2	Accountability to affected populations	34
3.	Current Humanitarian Data Science Use Cases	35
3.1	Natural Language Processing	35
3.2	Predictive Modelling and Forecasting	36
3.3	Computer Vision	39
3.4	Speech/Audio and Facial Recognition	40
3.5	Hybrids	40
4.	Next Steps	41
5.	Glossary	42

I. Introduction

In the past two decades, the “data revolution” has transformed society. We have seen an explosion in the availability and use of big data, the emergence and maturing of data science methods such as machine learning and artificial intelligence, enhanced use of smart phones and the Internet of Things (IoT), the open data movement, the rise of crowdsourcing, and new Information Communication Technologies (ICTs) for data collection.

Advances in computing and data science now make it possible to process and analyse huge quantities of data in real time. New insights gleaned from such data mining can complement official statistics and survey data, adding depth and nuance to information on human behaviours and experiences. The integration of this new data with traditional data has the potential to produce high-quality information that is more detailed, timely and relevant.⁵ The humanitarian sector - not immune to the promises and potential of the data revolution and data science - has begun exploration of its applications. *Chapter 3 elaborates on current humanitarian use cases.*

Greater and wider adoption of data science has consequently attracted more attention and focus on AI ethics (see side bar for explanation of terminology). AI ethics is described by the Alan Turing Institute as “a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies.”⁶ But despite an influx of AI ethics, principles, guidelines and frameworks being published in the private and public sectors, there is an identifiable and urgent gap when it comes to data science specifically for and in the humanitarian sector. The rapid diversification of available data science tools and their use for humanitarian purposes means that the implementation of these technologies risks outpacing the ethical and practical guidance.⁷

Since the beginning of the ‘humanitarian data revolution’ (circa 2009, although there is no official date when this began), the humanitarian data ecosystem has grown exponentially, and continues to do so. Despite numerous humanitarian agencies publishing their own data protection manuals over the last 10 years, there is a recognized lack of cross-organizational and sector-wide endorsed guidance when it comes to applying data science methods.⁸

Terminology: Data Science

Largely regarded as a blend of various tools, mathematical models, algorithms, Machine Learning (ML) and Artificial Intelligence (AI) techniques that aim to discover patterns, correlations and relationships from data. The different “categories” of data science are not definitive, nor universally agreed on, which is why we use the expression “data science” to try and avoid excluding anything. Therefore, for the rest of this document, the term “data science” will refer broadly to advanced mathematical models, algorithms, ML and AI.

However, when referring to ethics, a great majority of publications and sources of information use the expression “AI Ethics”. For ease of reading and citations, we too use this expression.

For more detailed individual breakdowns of relevant terms, please see the glossary.

5 United Nations, Big Data for Sustainable Development, Available from: <https://www.un.org/en/sections/issues-depth/big-data-sustainable-development/index.html>.

6 David Leslie, Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector, (London, Alan Turing Institute, 2019) p.3 Available from: https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf.

7 Nathaniel Raymond and Cassey Harrity, Addressing the ‘doctrine gap’: professionalising the use of Information Communication Technologies in humanitarian action, Humanitarian Innovation: Humanitarian Exchange, (London, Overseas Development Institute, 2016). Available from: <https://odihpn.org/wp-content/uploads/2016/04/HE-66-Web-Final.pdf>.

8 International Organization for Migration and the German Federal Foreign Office (FFO), Workshop Report on Forecasting Human Mobility in Contexts of Crises, (London, 2020). Available from:

The Harvard Humanitarian Initiative's Signal Code,⁹ and other frameworks, such as UN-OCHA's Working Draft Data Responsibility Guidelines,¹⁰ ICRC's Handbook (chapter on Data Analytics and Big Data),¹¹ and USAID's Considerations for Using Data Responsibly,¹² provide fundamental foundations on how to work responsibly with humanitarian data, but do not yet fully cover the relatively new and increasingly popular methods of advanced data science. For specific data activities – such as using AI or collecting biometric data – there is a general lack of established standards such as the Core Humanitarian Standards or SPHERE Standards. Having sector-wide developed and endorsed established standards enables organizations to self-regulate their activities, donors to monitor the quality of work, and establishes and strengthens transparent and accountable standards for the affected populations. A general lack of “digital guidelines” is not only a current and critical gap in humanitarian work, but more importantly, creates an added risk for affected communities and the organizations serving them in a digitalizing world.¹³

In the context of the data revolution, the initiative of measuring the Sustainable Development Goals (SDGs) has to some extent increased stakeholder participation and involvement in handling development data and exploring potential humanitarian data contributions.¹⁴ This also includes new stakeholders from the private sector, such as a recent partnership between the UN and Google, aimed at measuring the SDGs,¹⁵ and Facebook's Disaster Maps which uses methodology from the Internal Displacement Monitoring Centre.¹⁶ These new humanitarian, development and private partnerships are generally seen as positive, fostering innovation through cross-disciplinary learning and adding capacity to humanitarian efforts. However, there are a number of challenges that face such partnerships, including reputational risk, inappropriate technology, data sensitivity and use, uncertainty about new data sources, intellectual property, and dependency and deference.¹⁷ More importantly, humanitarian activities are unique in the sense that humanitarian actors must abide by the

<https://displacement.iom.int/reports/workshop-report-forecasting-human-mobility-contexts-crises>.

9 Campo, S. Nowarth, C. Raymond, N. Scarnecchia, D. The Signal Code: Ethical Obligations for Humanitarian Information Activities, Harvard Humanitarian Initiative, (2018). Available from: <https://hhi.harvard.edu/publications/signal-code-ethical-obligations-humanitarian-information-activities>.

10 The Centre for Humanitarian Data, United Nations Office for the Coordination of Humanitarian Affairs (OCHA), Data Responsibility Guidelines: Working Draft (The Hague, 2019) Available from: <https://centre.humdata.org/wp-content/uploads/2019/03/OCHA-DR-Guidelines-working-draft-032019.pdf>.

11 International Committee of the Red Cross, Handbook on Data Protection in Humanitarian Action, (Geneva, 2017). Available from: https://shop.icrc.org/e-books/handbook-on-data-protection-in-humanitarian-action.html?_ga=2.78584404.153056494.1570181623-874878445.1568374288.

12 USAID, Considerations for Using Data Responsibly (May 2019). Available from: <https://www.usaid.gov/responsibledata>.

13 As stated, The Centre for Humanitarian Data is stewarding a process to develop sector-wide operational guidance on data responsibility in humanitarian action for proposed endorsement by the IASC in late 2020. While this may not address some of the more ‘advanced analytics’ concerns tackled in this document, it signals awareness of some of these issues in the sector.

14 For instance, the data revolution was recognised as an enabler of the Sustainable Development Goals, not only to monitor progress but also to inclusively engage stakeholders at all levels to advance evidence-based policies and programmes and to reach the most vulnerable. The 2030 Agenda asserts that “Quality, accessible, timely and reliable disaggregates data will be needed to help with the measurement of progress (SGDs) and to ensure that no one is left behind. Such data is key to decision making.” See Data Privacy, Ethics and Protection Guidance note on Big Data for Achievement of the 2030 Agenda https://unsdg.un.org/sites/default/files/UNDG_BigData_final_web.pdf.

15 Amy Lieberman, New Google, UN partnership aims to ‘turn on the tap’ to SDG data (New York, DEVEX, 2019) Available from www.devex.com/news/new-google-un-partnership-aims-to-turn-on-the-tap-to-sdg-data-95695 Similarly, WFP and Alibaba have partnered for a Hunger Map Platform which combined information such as weather, population size, conflict, hazards, food security, nutrition and market information – to monitor and predict the global food security situation in near real-time. See www.unglobalpulse.org/news/data-and-ai-progress-during-unga.

16 Dow, A. Giraudy, E. Iyer, S. Maas, P. Pompe, A. Facebook Releases Improved Displacement Maps for Crisis Response. (January 2020) Available from: https://research.fb.com/blog/2020/01/facebook-releases-improved-displacement-maps-crisis-response/?utm_source=social-facebook&utm_medium=fb4d&utm_campaign=organic&utm_content=post-url&utm_offering=research&utm_product=displacement-maps-improvements_012820&eventSource=OrganicSocialFB4D&fbclid=IwAR0xMH52itlRInf4NCVGfSqLI3g9E9sPLUvExz0HwucXoTvyxEbHtao-dy0.

17 The Centre for Humanitarian Data, United Nations Office for the Coordination of Humanitarian Affairs (OCHA), Note #3: Data Responsibility in Public-Private Partnerships. Guidance Note Series, Data Responsibility in Humanitarian Action (The Hague, 2020) Available from: https://centre.humdata.org/wp-content/uploads/2020/02/guidance_note_ppp.pdf

humanitarian principles, unlike the private or academic sector who are driven by different motivations such as profit or knowledge production. In this sense, ensuring dignity while reducing vulnerabilities should not be at the cost of rendering affected communities as “test beds” or “testing sites”. Technologies that would never be applied, nor allowed in countries like the UK or US due to legal and regulatory barriers must not be applied for piloting capacity through humanitarian actions.¹⁸

In January 2019, UN-OCHA’s Centre for Humanitarian Data conducted a global survey to better understand how humanitarians are using data, what their core data skills are, and what kind of support they require to improve their capacity to use data effectively.¹⁹ Notably, 61 per cent of respondents stated that their main motivation for wanting to improve their data skills is to make humanitarian response more effective. When asked what data-related topics they are most interested in learning, the top three responses were big data, predictive analytics, and statistics. The Centre attributes this to the “trendiness” of these topics, but also to a lack of clarity and understanding of them. The survey results therefore support the need for this Framework to exist: humanitarians recognize the importance and potential that data and tech have for humanitarian work, but want and need further clarity on how to use it responsibly and ethically.

Currently, we are witnessing unprecedented rates of data being collected worldwide, a wider pool of stakeholders producing ‘humanitarian’ data, data becoming more machine readable, and data being more accessible via online portals.²⁰ This has enabled an environment for innovation and progress in the sector, and has led to enhanced transparency, informed decision making, and effective humanitarian service delivery. It also paves the way for organizations to reduce duplicate data collection activities, with the endeavor to reduce assessment fatigue for the affected communities, and enabling a “do more with less” approach, to ensure efficient use of resources and value for money to donors.²¹ Despite these advantages, the importance of balancing these benefits with the potential harms cannot be overstated. With much of this new data available online and greater machine-legibility, new and more serious risks are created. We are acutely aware of some of these risks, such as creating or exacerbating bias, de-anonymization, algorithmic discrimination, disinformation, weaponization, which are topics often raised in humanitarian digital discussions. Unfortunately, due to poor tracking and sharing of these occurrences and their consequences, there is a lack of documented evidence of them. In order to generate regulations to help prevent these risks and adverse consequences, evidence to steer discussions is urgently needed. There appears to be a general attitude not to report incidents in the sector for several flawed reasons. That said, some formative work is being undertaken by OCHA’s Centre for Humanitarian Data on guidance around data incident management.²²

How to read this Framework: This Framework covers the ethical issues that arise throughout all the stages of the lifecycle of a humanitarian data science project. However, for the sake of brevity and effectiveness, our chapters are short and contain only the most relevant points, plus resources you can use for more in-depth explorations. There are many comprehensive and useful reports/guides/frameworks available which expand greatly on each stage and issue of data science and ethics and we hope our Framework is useful for identifying, accessing and navigating them.

18 For a reference on humanitarian experimentation, see: McDonal, S. Sandvik, K. Jacobsen, K. From Principle to Practice: Humanitarian Innovation and Experimentation. (December 2017). Available from: https://ssir.org/articles/entry/humanitarian_innovation_and_experimentation#.

19 The Centre for Humanitarian Data, United Nations Office for the Coordination of Humanitarian Affairs (OCHA), We are all Data People: Insights from the Data Literacy Survey, (The Hague, 2019), Available from: <https://centre.humdata.org/we-are-all-data-people-insights-from-the-data-literacy-survey/>.

20 The Centre for Humanitarian Data, United Nations Office for the Coordination of Humanitarian Affairs (OCHA), The State of Open Humanitarian Data: What Data is Available and Missing Across Humanitarian Crises (The Hague, 2020), Available from: <https://centre.humdata.org/the-state-of-open-humanitarian-data-2020/>.

21 Johnson, S. “Minimum Viable Data – When Less Data is More Viable”, Medium, (Jan 2018) Available from: <https://medium.com/hetco-zine/minimum-viable-data-when-less-data-is-more-valuable-d225b5fa9c1b>.

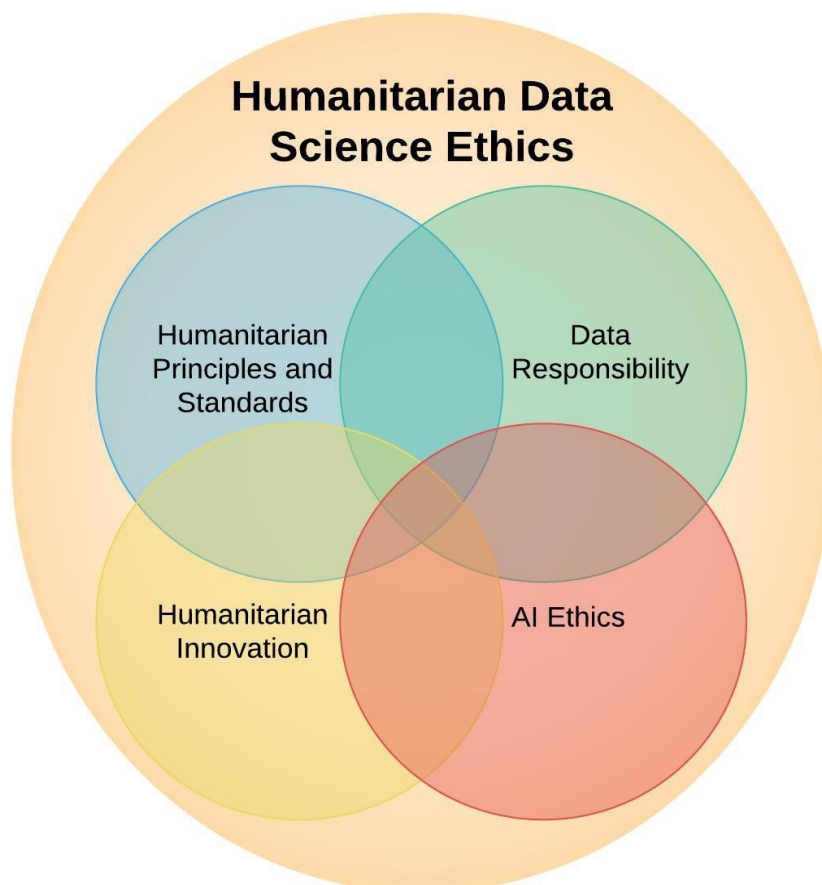
22 The Centre for Humanitarian Data, United Nations Office for the Coordination of Humanitarian Affairs (OCHA), Guidance Note: Data Incident Management (The Hague, 2019) Available from: <https://centre.humdata.org/guidance-note-data-incident-management/>.

Figure 1

Contextualizing DSEG:

How these discussions cannot be done in siloes, but encapsulate multiple topics

This illustration depicts the four thematic areas (Humanitarian Standards and Principles, Data Responsibility, Humanitarian Innovation and AI Ethics) that this ethical framework draws from and combines to create/develop/present the overarching term/theme of humanitarian data science ethics. We recognize that these thematic areas are inter-connected yet separate components for understanding humanitarian data science ethics. Hence, the work of the DSEG begins from the premise that this framework is a combination of these four thematic areas and whilst each have their own “guidance” or “frameworks”, the combination/intersection of them all raises new issues which this framework addresses. This document appreciates that other elements can overlap into this discussion but uses Figure 1 to demonstrate a good entry point to discussions in the application of data science methods to support humanitarian outcomes.



The timeline on the next page shows how AI ethics in the humanitarian sector have progressed over time, drawing on the four main contributing thematic areas illustrated above. Humanitarian standards and principles clearly predate the digital sources, with standards being developed as early as 1965 with the “Fundamental Principles of the International Red Cross and Red Crescent Movement” being proclaimed in Vienna in 1965.²³ After the proliferation of mobile phones and the start of the big data boom, approximately between the years of 2010 and 2015, organizations started developing policies and frameworks to guide responsible data collection and management. Humanitarian innovation got a major boost after it was prioritized at the World Humanitarian Summit (WHS) in 2016, although progress has been relatively slow, with the only comprehensive guide to humanitarian innovation being released in 2019. The Timeline shows how AI ethics emerged fairly recently, in 2016, and gained more traction and importance 2018 onwards.²⁴

23 United Nations Office for the Coordination of Humanitarian Affairs (OCHA), OCHA on Message: Humanitarian Principles, (2012) Available from: https://www.unocha.org/sites/dms/Documents/OOM-humanitarianprinciples_eng_June12.pdf

24 This timeline was compiled through extensive desk research. The Signal Code is in purple as it represents the combination of two thematic areas. Major contributing sources for this timeline include: Berens, J. Mans, U. Verhulst, S. (2016). Mapping and Comparing Responsible Data Approaches. GovLab and Centre for Innovation – Leiden University. Available at: <https://www.thegovlab.org/static/files/publications/ocha.pdf> and: Fjeld, Jessica and Achten, Nele and Hilligoss, Hannah and Nagy, Adam and Srikumar, Madhulika, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for Ai (January 15, 2020). Berkman Klein Center Research Publication No. 2020-1. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482.

Figure 2: Timeline

Humanitarian Principles, Data Responsibility Guidelines, Humanitarian Innovation Guidelines, AI Ethics and Principles.

Pre-2013	2014	2015	2016	2017	2018	2019	2020
<p>1965 Red Cross Fundamental Principles</p> <p>1994 ICRC: Code of Conduct</p> <p>1998 SPHERE: Humanitarian Charter and Minimum Standards in Humanitarian Response</p> <p>2010 IOM: Data Protection Policy</p> <p>2011 UNICEF: Information Disclosure Policy</p> <p>2012 European Union: Draft General Data Protection Regulations</p> <p>2013 MSF: Data Sharing Policy</p>	<p>Core Humanitarian Standards on Quality and Accountability</p> <p>GSMA: Guidelines on the protection of privacy in the use of mobile phone data for responding to the Ebola outbreak</p> <p>USAID: Privacy program</p> <p>Principles for Digital Development: Working Group established to draft Principles</p>	<p>Oxfam: Responsible Program Data Policy</p> <p>UNHCR: Policy on the Protection of Personal Data of Persons of Concern to UNHCR</p> <p>UN Global Pulse: Privacy and Data Protection Principles</p> <p>UN Innovators Network: Formation of informal group</p> <p>University of Oxford: Principles for Ethical Humanitarian Innovation</p>	<p>ICRC: Rules on Personal Data Protection</p> <p>Partnership on AI to Benefit People and Society: Tenets</p> <p>UNDP and UN Global Pulse: A Guide to Data Innovation for Development, from Idea to Proof-of-Concept</p> <p>World Humanitarian Summit: Focus on Innovation and creation of Global Alliance for Humanitarian Innovation (now defunct)</p> <p>MSF: Ethics Framework for Humanitarian Innovation</p>	<p>ICRC: Handbook on Data Protection in Humanitarian Action</p> <p>UNDP: Data Privacy, Ethics and and Protection Guidance Note on Big Data for Achievement of the 2030 Agenda</p> <p>Microsoft: AI Principles</p> <p>Google: AI at Google - Our Principles</p> <p>Future of Life Institute: Asilomar AI Principles</p> <p>The Signal Code: A Human Rights Approach to Information During Crisis</p> <p>The Signal Code: A Human Rights Approach to Information During Crisis ***</p>	<p>SPHERE (2018 edition): Humanitarian Charter and Minimum Standards in Humanitarian Response</p> <p>European Union: GDPR (General Data Protection Regulations)</p> <p>UN: Principles on Personal Data Protection and Privacy</p> <p>Data Science and Ethics Group (DSEG): Creation of partnership between INGOs, UN and universities</p> <p>Amnesty International and Access Now: The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems</p> <p>University of Montreal: The Montreal Declaration for a Responsible Development of Artificial Intelligence</p> <p>European High Level Experts Group on AI : Ethics Guidelines for Trustworthy Artificial Intelligence</p> <p>AI Now: Algorithmic Accountability Policy Tool Kit</p>	<p>USAID: Considerations for using data responsibility at USAID</p> <p>UN-OCHA: Data Responsibility Guidelines Working Draft</p> <p>Elhra Humanitarian Innovation Fund: Humanitarian Innovation Guide</p> <p>UN-OCHA Centre for Humanitarian Data: Predictive Analytics Peer Review Framework</p> <p>The Alan Turing Institute: Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI Systems in the Public Sector</p> <p>The International Development Innovation Alliance: Artificial Intelligence in International Development: A Discussion Paper</p> <p>OECD: AI Principles</p>	<p>UN-OCHA: Survey results on priorities for data responsibility in humanitarian action</p> <p>UN-OCHA Centre for Humanitarian Data: Glossary of Data Terms</p> <p>Harvard Berkman Klein Principled Artificial Intelligence</p> <p>UN-OCHA Centre for Humanitarian Data: Catalogue for Predictive Models in the Humanitarian Sector</p>

2. The Lifecycle of a Humanitarian Data Science Project

This chapter explains the lifecycle of a humanitarian data science project and the ethical concerns involved, whilst offering practical guidance for addressing them. However, it is important to note that this lifecycle is for illustration purposes, and that not all projects will follow this or the same lifecycle. The process is not always linear, nor chronological. This is purely a recommendation based on the stakeholder engagement we undertook through the framework development.

By dividing the lifecycle into stages, it is easier to identify what is involved in a humanitarian data science project, and to run through a checklist process to consider issues and barriers that each stage may raise. While we highlight the need to address ethical considerations as they arise along the data science journey, many of them occur concurrently and/or iteratively. To even consider a humanitarian data science project in the first place requires an appreciation of data science ethics in general as well as humanitarian principles. Therefore, we introduce Fundamentals early in the process as Stage 0.

Stage of Data Science Lifecycle		Description
0	Fundamentals	<ul style="list-style-type: none">• Humanitarian Principles and Ethics• AI Ethics• Data Responsibility• Human Rights• Risk Mitigation
1	Problem and Solution Exploration	<ul style="list-style-type: none">• Problem recognition – breaking down the problem and articulating what you want from a solution;• Search for solutions, ideas and collaborators: is AI the ideal solution?
2	Data Journey	<ul style="list-style-type: none">• Data collection• Data processing and protection
3	Algorithm	<ul style="list-style-type: none">• Ethical principals of algorithmic design
4	Reliance on Outputs	<ul style="list-style-type: none">• False negatives and false positives• Accountability to affected populations

These stages will be expanded upon throughout this chapter.

2.1 Humanitarian Data Science

How and why are humanitarian data science projects different to other data science projects?

As shown on the Timeline (Figure 2), in recent years, there have been numerous reports, principles and guidelines compiled for designing and building ethical AI for various uses in various sectors. So why then have we developed this Framework for humanitarian data science projects? While there are ethical concerns which are generic to data science, the risks and ethical considerations for their humanitarian application are very specific and can have particularly serious ramifications considering the extreme vulnerability of the populations that humanitarian organisations are trying to support. Further, as highlighted previously, the humanitarian sector is a principle-based sector.

Harvard's Berkman Klein Center for Internet and Society analysed the content of 36 prominent AI principles documents from a variety of sectors in the search for common themes. This analysis found eight key themes:

1. Privacy
2. Accountability
3. Safety and security
4. Transparency and explainability
5. Fairness and non-discrimination
6. Human control of technology
7. Professional responsibility
8. Promotion of human values²⁵

These principles are relevant to almost all types of data science projects, however, can be especially relevant and potentially problematic when applied to humanitarian data science projects. The below table highlights how the nature of humanitarian work emphasizes and increases the risk of the above ethical considerations:

General AI Principle	Specific application to the humanitarian sector
Privacy	<p>Consent: In a typical private context, the user of technology is making an active choice to participate and commonly “agrees” to terms and services. However, for vulnerable persons in humanitarian contexts, participation or providing certain types of personal or personally identifiable data may be conditionally tied to aid. In such scenarios, questions of fairness, consent and choice arise, as an individual’s ability to make free decisions about sharing their own data is diminished.</p> <p>Control over the use of data (restriction of processing; right to rectification; right to erasure): Affected populations constantly give up their time and information, often on multiple and frequent occasions. Such populations are unlikely to know their rights regarding their data, nor be in the position to protect them.</p> <p>Anonymity: Under the humanitarian principles of do no harm, the commitment to vulnerable people includes intended and unintended consequences. Anonymity is important to maintain for avoiding risks such as targeting, re-identification, sensitivity while adhering to the right to protect (to name only a few).</p>

25 Fjeld, Jessica and Achten, Nele and Hilligoss, Hannah and Nagy, Adam and Srikumar, Madhulika, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for Ai (January 15, 2020). Berkman Klein Center Research Publication No. 2020-1. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482.

General AI Principle	Specific application to the humanitarian sector
Accountability	<p>Liability and legal responsibility; remedy for automated decision; ability to appeal: Establishing accountability in data science and obtaining remedy is hard enough, without the complications of it applying it to a complex humanitarian situation. Humanitarian subjects that are affected by a data science decision or tool may not even know that it has happened, and even if they do, likely have little recourse.</p> <p>Accountability mechanisms by humanitarians: Greater accountability in the humanitarian system to communities and people affected by crises is a requirement of the Core Humanitarian Standard on Quality and Accountability.</p>
Safety and Security	<p>Security: All sectors must ensure their systems are robust and secure in order to protect privacy, and maintain the integrity and confidentiality of personal data. Humanitarian systems may contain extremely sensitive information either in the form of personal identifiable information, location or demographically identifiable data. For example, a warring party may want access to satellite imagery of a conflict zone in order to unleash more targeted attacks, and they may try to access this data maliciously by attacking an NGO's database. It is therefore imperative that humanitarian tools and platforms are safe and secure.</p>
Transparency and Explainability	<p>For the humanitarian sector, transparency in decision-making and aid delivery are important to improve coordination, efficiency and accountability to beneficiaries. In the growing context of evidence-based decision-making, it is increasingly necessary that processes and rationale for prioritization and targeting are open.</p>
Fairness and Non-Discrimination	<p>Fairness: The humanitarian sector is under an obligation to adhere to the humanitarian principles of humanity, neutrality, independence, and impartiality. The use of data science may create perceptions of partiality/bias. The humanitarian sector should be cautious of ever using (or allowing others to use) vulnerable people as test cases when experimenting with new technologies.</p> <p>Validity (representative and high quality data): Humanitarian data is often difficult to access as it is likely to be affected by natural disaster, conflict, remote areas, lack of identification and political barriers. As such, the data may not be complete or representative and this poses a real threat to the validity of data science outputs.</p> <p>Non-discrimination and ossification:²⁶ Those in need of humanitarian assistance are already in vulnerable circumstances. Many may be suffering from discrimination and persecution, and therefore, the risk of an algorithm perpetuating past bias and discrimination is even more serious.</p> <p>Inclusiveness in design: The 2016 Grand Bargain commits aid organisations and donors to a “Participation Revolution” whereby those receiving aid, have a say in the decisions made about them.²⁷</p>

²⁶ Ossification is the tendency of algorithms to learn and codify the current state of the world and therefore make it harder to change.

²⁷ Interagency Standing Committee, Grand Bargain Participation revolution: work-stream plan of action, Available from: <https://interagencystandingcommittee.org/a-participation-revolution-include-people-receiving-aid-in-making-the-decisions-which-affect-their-lives>.

General AI Principle	Specific application to the humanitarian sector
Human Control of Technology	Some actions/decisions require human interaction and are irreplaceable by technology. Others, could benefit from technology, but still require human-in-the-loop. For example, highly sensitive topics such as gender-based violence (GBV) may require face-to-face interviews to put affected populations/interviewees at ease, as well as demonstrate empathy. However, in some circumstances, technology can provide a platform for people to report on instances of GBV without the challenges of face-to-face interviews (i.e. if the enumerator is a male, or in contexts where women may not feel comfortable/ or have cultural barriers to talk about it, technology can lend them the voice without having to speak out loud). Each context needs to be examined for the suitability of using technology, but it is the case that almost all humanitarian scenarios using technology will still require large amounts of human oversight.
Professional Responsibility	<p>Skills and technology: The humanitarian sector is unlikely to have familiarity with, nor knowledge of the latest cutting-edge technology available in AI/ML. Only few staff members will be able to work with data science tools, and unless they have partnerships with tech companies, will unlikely have access to the computer power or tools suitable for the task.</p> <p>Multistakeholder collaboration: Effective and responsible AI involves collaboration between stakeholders from different sectors (tech, policy, legal etc). As mentioned above, the humanitarian sector is unlikely to have all the necessary expertise internally and therefore will need to rely heavily on partnerships with external parties. Due to the nature of humanitarian work, specific guidance for humanitarian-private partnerships is helpful.²⁸</p> <p>Finance and risk: Humanitarian funding is often project-based and demands short-term or immediate delivery. This impedes the process of taking on more complex long-term projects necessary for developing new technology. The nature of the humanitarian sector also lends itself to a culture of risk aversion, especially in comparison with the private sector which can likely afford to take on high risk projects and commit to longer term strategies.</p>
Promotion of Human Values	Leveraged to benefit society: Incentives to use AI in humanitarian work vary from directly benefitting beneficiaries/end-users (i.e. through forecast-based financing), to indirectly benefitting them (i.e. aid logistics efficiencies mean donations can spread further). The latter poses the risk that operational efficiencies and cost-saving benefits of AI become the priority, and rights concerns such as privacy and discrimination are overshadowed. The purpose of humanitarian action is to protect life and health and ensure respect for human beings, and this should always be the priority over financial or operational improvements.

For these reasons, we believe that humanitarian data science projects warrant their own specific ethical framework.

The remainder of this chapter will break down the stages of a humanitarian data science project, and will include practical “Action Points” you can consider to tackle ethical risks.

²⁸ See: The Centre for Humanitarian Data, United Nations Office for the Coordination of Humanitarian Affairs (OCHA), Note #3: Data Responsibility in Public-Private Partnerships. Guidance Note Series, Data Responsibility in Humanitarian Action (The Hague, 2020) Available from: https://centre.humdata.org/wp-content/uploads/2020/02/guidance_note_ppp.pdf.

2.2 STAGE 0: Fundamentals

From inception, the development of any humanitarian data science project would be incomplete without an understanding and consideration of at least these five fundamentals. They include:

- AI Ethics
- Humanitarian Principles and Ethics
- Data Responsibility
- Human Rights
- Risk Mitigation

2.2.1 AI Ethics

“AI ethics is a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies.”²⁹

- Alan Turing Institute

In 2019, The Alan Turing Institute published a report titled “Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector.” It is one of the most comprehensive guidance pieces on the topic of AI ethics and safety to date. The report’s Executive Summary emphasizes the importance of AI ethics and safety as a top priority, with attention to social and ethical implications of the design and use of AI systems into **every stage** of the delivery of an AI project.

As such, it is important to have a basic understanding of AI ethics before commencing a data science project, as well as reviewing and incorporating them (when relevant), at each stage of the data science lifecycle.

While some major tech companies that are most active in AI have written AI ethical guidelines (such as Microsoft and Google), humanitarians will find more relevant ethical guidelines in documents such as the Alan Turing report, the European High Level Experts Group’s Ethics Guidelines for Trustworthy Artificial Intelligence,³⁰ and the International Development Innovation Alliance’s Discussion Paper on AI.³¹ These publications target projects where financial profit is not the main motive, and where the project owners owe stricter legal and ethical obligations towards those affected by their data science tool.

As mentioned previously, ethical issues and risks can arise throughout the lifecycle of a data science project. Below we have shown how and why the “AI potential harms” (i.e. ethical concerns) as identified by the Alan Turing Institute, can occur throughout a data science lifecycle. This will illustrate the broad and overall considerations you should have before starting work on a humanitarian data science project.

29 David Leslie, Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector, Alan Turing Institute, 2019, pp.6 Available from: https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf.

30 The Ethics Guidelines for Trustworthy Artificial Intelligence, European Commission High Level Expert’s Group, April 2019, Available from: <https://ec.europa.eu/futurium/en/ai-alliance-consultation>.

31 Artificial Intelligence in International Development, The IDIA Working Group on Artificial Intelligence in International Development, June 2019. Available from: https://static1.squarespace.com/static/5b156e3bf2e6b10bb0788609/t/5e1f0a37e723f0468c1a77c8/1579092542334/AI+and+international+Development_FNL.pdf.

AI potential harm	Why it can happen
Bias and Discrimination	<ul style="list-style-type: none"> • Incomplete or unrepresentative data fed into the model • Algorithm designer creates code with bias (“hard-coded bias”)
Denial of Individual Autonomy, Recourse, and Rights	<ul style="list-style-type: none"> • Unclear who to hold accountable for outputs and consequences of the data science tool
Non-transparent, Unexplainable, or Unjustifiable Outcomes	<ul style="list-style-type: none"> • Insufficient transparency and explainability of the algorithm • Discrimination or human rights infringements resulting from use of the tool
Invasions of Privacy	<ul style="list-style-type: none"> • During data collection, processing, if subjects are targeted, profiled or nudged without their consent
Isolation and Disintegration of Social Connection	<ul style="list-style-type: none"> • Excessive automation, hyper-personalization leading to echo chambers
Unreliable, Unsafe, or Poor-Quality Outcomes	<ul style="list-style-type: none"> • Irresponsible data management • Negligent algorithm design • Human rights breaches, discrimination, distrust in the technology and those who use it

Action Points:

- Research and learn from existing work: Familiarize yourself with AI ethics by reading through existing frameworks / guidelines that relate to your kind of work. For example, the public sector will benefit from Alan Turing’s framework and the development sector will benefit from International Development Innovation Alliance’s Discussion Paper on AI. Some of these publications, such as the Alan Turing Institute’s report, contain useful templates for checking the above-mentioned ethical risks.
 - + For a list, analysis and data visualization of major AI Principles publications, visit Harvard Berkman Klein’s Principled Artificial Intelligence.
 - + For a comprehensive list of all available AI principles/standards, see Algorithm Watch’s AI Ethics Guidelines Global Inventory.³²
- Establish an “ethical platform” which will help to determine throughout the lifecycle whether a project still adheres to AI ethics. You can use this platform from inception through to outputs. Include the following questions in your ethical platform:
 - + Ethically permissible: What are the impacts on the wellbeing of affected stakeholders and communities?
 - + Fair and non-discriminatory: What are the potential discriminatory effects on individuals and social groups? How do you mitigate biases that may influence your model’s outputs? How to be aware of the issues surrounding fairness that come into play at every phase of the design and implementation pipeline?
 - + Worthy of public trust: How does the project guarantee to the greatest possible extent the safety, accuracy, reliability, security, and robustness of its product?
 - + Justifiable and transparent: How does the AI project ensure that it is justifiable by prioritizing both the transparency of the process by which the model is designed and implemented, and the transparency and interpretability of its decisions and behaviours?

³² Anna Jobin, Marcello Lenca and Effy Vayena, “The global landscape of AI ethics guidelines”, Nature Machine Intelligence (September 2019) Available from: <https://algorithmwatch.org/en/project/ai-ethics-guidelines-global-inventory/>



2.2.2 Humanitarian Principles and Ethics

All work done by humanitarians needs to uphold the humanitarian principles of humanity, neutrality, impartiality, and independence, which are described as the following:

Humanity	Neutrality	Impartiality	Operational independence
Human suffering must be addressed wherever it is found. The purpose of humanitarian action is to protect life and health and ensure respect for human beings.	Humanitarian actors must not take sides in hostilities or engage in controversies of a political, racial, religious or ideological nature.	Humanitarian action must be carried out on the basis of need alone, giving priority to the most urgent cases of distress and making no distinctions on the basis of nationality, race, gender, religious belief, class or political opinions.	Humanitarian action must be autonomous from the political, economic, military or other objectives that any actor may hold with regard to areas where humanitarian action is being implemented.

*Image source: OCHA on Message: What are Humanitarian Principles?*³³

In addition to humanitarian principles, there are several standards, and code of conducts that agencies use to set benchmarks within their operations. What is useful about these standards, is that most of them are cross-organizational.

- Humanitarian Standards:
 - + Humanitarian Charter and minimum standards (SPHERE Handbook)³⁴
 - + Core Humanitarian Standards³⁵
- Protection Principles:
 - + ICRC and NGO Code of Conduct³⁶

For specific subjects within the humanitarian sector, there are further multi-agency standards – the Education in Emergencies (EIE)³⁷ being one example.

33 United Nations Office for the Coordination of Humanitarian Affairs (OCHA), OCHA on Message: Humanitarian Principles, (2012) Available from: https://www.unocha.org/sites/dms/Documents/OOM-humanitarianprinciples_eng_June12.pdf.

34 The Sphere Handbook 2018, Available from: <https://spherestandards.org/handbook-2018/>.

35 Core Humanitarian Standard, Available from: <https://corehumanitarianstandard.org/the-standard>.

36 Code of Conduct for the International Red Cross and Red Crescent Movement and Non-Governmental Organizations (NGOs) in Disaster Relief, Available from: <https://www.icrc.org/en/doc/resources/documents/publication/p1067.htm>.

37 INEE Minimum Standards for Education: Preparedness, Response, Recovery. Inter-agency Network for Education in Emergencies (INEE), (December 2010), Available from: <https://inee.org/resources/inee-minimum-standards>.

Humanitarians are further encouraged to adhere to humanitarian ethics. Humanitarian ethics, as described in Hugo Slim's seminal book *Humanitarian Ethics: A Guide to the Morality of Aid in War and Disasters*, "... has developed as a principle-based ethics ... grounded in the principles of humanity, impartiality, neutrality and independence that have been developed to guide the provision of humanitarian assistance and protection."³⁸ A recent Guidance Note written by OCHA's Centre for Humanitarian Data looks at how humanitarian ethics can be combined with data ethics, to ensure humanitarian data practices are carried out responsibly. The Guidance Note recommends that organisations invest in three areas to support ethical data practice:

1. Establish clear codes of conduct for ethical data management.
2. Support staff to identify, understand, and debate ethical issues using common tools.
3. Introduce 'ethical audits' as part of standard practice.³⁹

Action Points:

- **Compliance:** Ensure your organization is familiar with the Humanitarian Principles and other relevant standards mention above.
- **Refer:** Read OCHA's Guidance Note on Humanitarian Data Ethics for advice on creating codes of conduct, supporting staff to debate ethical issues, and introduce ethical audits.

2.2.3 Data Responsibility

If an organization is going to use data science methods – and therefore humanitarian data – it is crucial that its data management is compliant with relevant legal and ethical guidelines.⁴⁰ The following documents are sector-leading on humanitarian data responsibility. These guidelines provide direction on issues especially relevant to humanitarian data management such as data minimization, getting informed and meaningful consent, and data rights of the subjects.

- ICRC Handbook on Data Protection in Humanitarian Response⁴¹

This handbook "seeks to raise awareness and assist humanitarian organizations in ensuring that they comply with personal data protection standards when carrying out humanitarian activities, by providing specific guidance on the interpretation of data protection principles for humanitarian action, particularly when new technologies are employed." It does not seek to replace organizations' own data protection guidelines, nor data protection laws.

38 Hugo Slim, *Humanitarian Ethics: A Guide to the Morality of Aid in War and Disaster* (New York, NY: Oxford University Press, 2015), 40.

39 The Centre for Humanitarian Data, United Nations Office for the Coordination of Humanitarian Affairs (OCHA), Note #4: Humanitarian Data Ethics. Guidance Note Series, Data Responsibility in Humanitarian Action (The Hague, 2020) Available from: https://centre.humdata.org/wp-content/uploads/2020/02/guidance_note_ethics.pdf.

40 This document does not cover legal requirements as these are jurisdiction dependent. However, readers should be aware of the General Data Protection Regulations (GDPR) which may apply to their organisation depending on their type and area of operation. You should consult your organisation's legal team for this.

41 International Committee of the Red Cross, *Handbook on Data Protection in Humanitarian Action*, (Geneva, 2017). Available from: https://shop.icrc.org/e-books/handbook-on-data-protection-in-humanitarian-action.html?_ga=2.78584404.153056494.1570181623-874878445.1568374288.

- IOM Data Protection Manual⁴²

This manual is in three parts: the first part outlines the IOM data protection principles as informed by relevant international standards; the second part includes comprehensive guidelines on each principle, consideration boxes and practical examples; and the third part provides generic templates and checklists to ensure that data protection is taken into account when collecting and processing personal data. Although the content of this publication was developed for IOM use, it can be used as a resource tool by other organizations engaging in similar operations.

- UN OCHA: Data Responsibility Guidelines (working document)⁴³

These guidelines “offer a set of principles, processes and tools that support the safe, ethical and effective management of data in humanitarian response.” They are designed for OCHA staff but can be used beyond. They concentrate on sensitive yet non-personal data.

Action Point:

- Guidelines and Frameworks: Ensure your organization either has its own data responsibility practices or adheres to one of the sector leading documents listed above.

2.2.4 Human Rights

Although this report will not cover all of the legal implications of data science projects in the humanitarian sector, it will discuss some of the most relevant human rights issues which may arise from relying on data science methods.

Human rights are more universal and well-defined than ethics principles, and they provide for accountability and redress. They are universal and binding, and are codified in a body of international law including in the International Bill of Human Rights which is made up of the Universal Declaration of Human Rights (UDHR), the International Covenant on Civil and Political Rights (ICCPR), and the International Covenant on Economic, Social and Cultural Rights (ICESCR).⁴⁴ All humanitarian action is subject to human rights law.

Access Now have compiled a report examining the Human Rights Implications of AI,⁴⁵ and recently the Council of Europe published recommendations on human rights impacts of algorithmic systems.⁴⁶

⁴² International Organisation for Migration, *IOM Data Protection Manual*, (Geneva 2010). Available from: <https://publications.iom.int/books/iom-data-protection-manual>.

⁴³ The Centre for Humanitarian Data, United Nations Office for the Coordination of Humanitarian Affairs (OCHA), *Data Responsibility Guidelines: Working Draft* (The Hague, 2019) Available from: <https://centre.humdata.org/wp-content/uploads/2019/03/OCHA-DR-Guidelines-working-draft-032019.pdf>.

⁴⁴ See: UN Human Rights Office of the High Commissioner, *Fact Sheet No.2 (Rev.1)*, *The International Bill of Human Rights*. Available from: <https://www.ohchr.org/Documents/Publications/FactSheet2Rev.1.en.pdf>.

⁴⁵ L. Andersen, “Human Rights in the Age of Artificial Intelligence”, Access Now, 2018. Available from: <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf>.

⁴⁶ Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence, “Addressing the Impacts of Artificial Intelligence on Human Rights: Draft Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems”, 4 November 2019, Available from: <https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-states-on-the-hu/168095eecf>.

Case study

A report titled “Bots at the Gates” looks at the use of AI and automated decision-making by the Canadian immigration and refugee systems through a human rights analysis. It contends that since 2014, the Canadian immigration system has used AI to automate some decisions which would normally be taken by a human immigration officer, such as whether a marriage is ‘genuine’, or if a person is a ‘risk’. The people for whom these decisions are made, for example, refugees - are already vulnerable and likely lack access to legal support. To have a decision made for them which they may not understand due to lack of transparency and explainability of the algorithm, risks further marginalizing them. The legal and ethical grey-area of this raises large concerns about the use of automated decision-making methods in data science, particularly when the subject is a vulnerable or marginalized person. “While efficiency and technological development are valuable, those responsible for human lives should not pursue innovation at the expense of fairness, accountability and oversight. Fundamental human rights must hold a central place in this discussion.”

Source: Molnar, P. Gill, L. (2018). Bots at The Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee Systems. University of Toronto's International Human Rights Program. Available from: <https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf>.

The table below flags some human rights that may be affected, and the possible implications they may have for humanitarian data science projects.

Human Rights	Possible issues for humanitarian data science projects
Rights to life, liberty and security, equality before the courts, a fair trial ⁴⁷	If an individual wants to act against a humanitarian organization to appeal a decision made about them regarding aid provision, or refugee status, how can they do this? What mechanisms are in place?
Rights to privacy and data protection ⁴⁸	If aid provision is tied to data collection, sharing, storage and processing, then how can a person provide freely given informed consent? Are subjects aware that they can access their data? If so, is the humanitarian organization able to provide it and rectify or delete it if necessary?
Right to freedom of movement ⁴⁹	This could be affected by surveillance technologies: facial recognition, satellite imagery, prediction of movements. What if this information falls into the wrong hands? Could governments use this information to close borders? Does this information restrict or confine individuals' mobility?
Rights to equality and non-discrimination ⁵⁰	AI models are designed to sort and filter, whether by ranking search results or categorizing people into buckets. This discrimination can interfere with human rights when it treats different groups of people differently. ⁵¹ The risk that an AI tool does this to already vulnerable and marginalized people is a pressing concern.
Rights to work, an adequate standard of living ⁵²	Could this be at risk if an AI tool determines the level or applicability of aid?
Right to health ⁵³	This could be relevant in AI systems that predict disease outbreaks and recommend responses. What happens when you deploy resources to an area deemed high risk while leaving others without assistance? Human health workers already make this choice, but AI would do this preemptively, and may sometimes get it wrong. ⁵⁴

47 Article 3, 6, 7, 8, and 10 of UDHR, Articles 9 and 14 of the ICCPR.

48 For example: "Everyone has the right to the protection of personal data concerning him or her. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified." - Article 8 of the EU Charter of Fundamental Rights. Also see Article 17 of the ICCPR.

49 Article 12 of the ICCPR.

50 Articles 3, 26 and 27 of the ICCPR.

51 L. Andersen, "Human Rights in the Age of Artificial Intelligence", Access Now, 2018. Available from: <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf>.

52 For example: "The States Parties to the present Covenant recognize the right of everyone to an adequate standard of living for himself and his family, including adequate food, clothing and housing, and to the continuous improvement of living conditions." - Article 11 of the ICESCR.

53 Article 12 of the ICESCR

54 L. Andersen, "Human Rights in the Age of Artificial Intelligence", Access Now, 2018. Available from: <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf>.

Action Points:

- Compliance: The first step to take to prevent an AI project from potentially breaching human rights is to make sure it adheres with data protection laws. The General Data Protection Regulations (GDPR), even if not legally binding on all organizations depending on their type and area of operation, provides a framework which can be used to ensure data is used responsibly.
- Transparency and Explainability: This can be achieved through adherence to open source and open data standards and the publication of meaningful, accessible explanations of how an AI system works so that people can be meaningfully informed about how it may impact them.
- Refer: The Canadian Government is building an Algorithm Impact Assessment tool in order to gauge the effect a tool may have on the rights, health, and economic interests of data subjects. This tool can be used by any organisation/company on any project, and is currently in beta version.⁵⁵
- Conduct a human rights assessment: A practical framework for algorithmic impact assessment is provided for example by the AI Now Institute.⁵⁶ Access Now's Report on Human Rights in AI recommends that any human rights assessment process should include:
 - + Testing and audits by independent experts.
 - + Identifying measures to mitigate identified risks and prevent any rights violations from occurring and measuring compliance and efficacy.
 - + A failsafe to terminate acquisition, deployment, or any continued use if at any point an identified human rights violation is too high or unable to be mitigated.
 - + Identification of any new legal safeguards needed to protect human rights in specific applications of AI tools.
 - + Special determinations of bias.
 - + If a third party is used to develop and/or implement the system, a requirement for the third party to participate in the human rights assessment process.⁵⁷

55 Algorithmic Impact Assessment (AIA), Government of Canada, Available from: <https://www.canada.ca/en/government/system/digital-government/modern-emerging-technologies/responsible-use-ai/algorithmic-impact-assessment.html>.

56 Reisman, D. Schultz, J. Crawford, K. Whittaker, M. "Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability", AI Now Institute, April 2018. Available from: <https://ainowinstitute.org/aiareport2018.pdf>.

57 L. Andersen, "Human Rights in the Age of Artificial Intelligence", Access Now, 2018. Available from: <https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf>.

2.2.5 Risk Mitigation Guidance

Prior to delving into how to mitigate risks, it is important to acknowledge and explore the various types of risk that may arise through the use of data in humanitarian work. On almost every project there will be an element of risk, whether a diversion of resources, or an unintended negative impact on vulnerable populations. Below are some examples of potential risk incidents involving data:

Data weaponization: In 2016, Doctors Without Borders (MSF) stopped communicating data on their medical facilities to Syrian and Russian forces, because instead of safeguarding these facilities, both forces would likely target them.⁵⁸

Data leaks: In 2017, RedRose, which operated a beneficiary management and digital payment system for Catholic Relief Services, announced that a breach of the system occurred by a competitor⁵⁹ and the hackers were able to “access names, photographs, family details, PIN numbers, and map coordinates for more than 8,000 families receiving assistance from the NGO in West Africa.”⁶⁰

Data not being representative: A study on pneumonia⁶¹ found that people with asthma have less fatality rates from pneumonia. However, this was due to being provided health care and treatment faster than non-asthmatic patients and not because asthma lowers the risk of death. If the healthcare industry were to use similar rule-based systems with an AI model then they could risk excluding groups of otherwise vulnerable individuals, such as in this case putting asthmatics as low risk.⁶²

Data revealing private or sensitive information: An internal audit of World Food Program's (WFP) Beneficiary Management (Nov 2017)⁶³ found that a WFP “cooperating partner” had recorded people's religion. While it is not clear where this happened, it could have been in a context such as Myanmar, where a person's religion is a key factor in recent violence there, and therefore counts as very sensitive information.⁶⁴

Data mosaicking: This has been flagged as a potential problem by both UN Global Pulse⁶⁵ and Privacy International.⁶⁶ Both are concerned that with the greater and wider collection of data in the humanitarian sector, there is the risk of re-identification of humanitarian subjects which could put them at security risk.

58 The Guardian, “MSF stops sharing Syria hospital locations after ‘deliberate’ attacks”, Available at <https://www.theguardian.com/world/2016/feb/18/msf-will-not-share-syria-gps-locations-after-deliberate-attacks>.

59 Lisa Cornish, “New security concerns raised for RedRose digital payment systems”, DevEx, Available at <https://www.devex.com/news/new-security-concerns-raised-for-redrose-digital-payment-systems-91619>.

60 The New Humanitarian, “Humanitarian data breaches: the real scandal is our collective inaction”, Available at <https://www.thenewhumanitarian.org/opinion/2017/12/08/humanitarian-data-breaches-real-scandal-our-collective-inaction>.

61 Caruana et al, “Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission”, Columbia University and Microsoft Research, Available at <http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf>.

62 Nogrady, “The real risks of artificial intelligence”, BBC News, Available at <https://www.bbc.com/future/article/20161110-the-real-risks-of-artificial-intelligence>.

63 Internal Audit of Beneficiary Management: Office of the Inspector General Internal Audit Report AR/17/17 Available at: https://docs.wfp.org/api/documents/WFP-0000040084/download/?_ga=2.18686585.1326768420.1516256388-1682848339.1511261484.

64 Parker, B. “Audit exposes UN food agency's poor data-handling” 18 Jan 2018. Available at: <http://www.thenewhumanitarian.org/news/2018/01/18/exclusive-audit-exposes-un-food-agency-s-poor-data-handling>.

65 Big Data for Development and Humanitarian Action', UN Global Pulse, p 14, Available at http://unglobalpulse.org/sites/default/files/Big_Data_for_Development_and_Humanitarian_Action_Report_Final_0.pdf.

66 ‘Development And Humanitarian Aid Initiatives Enable Surveillance In Developing Countries’, Privacy International, Available at: <https://privacyinternational.org/news-analysis/1225/development-and-humanitarian-aid-initiatives-enable-surveillance-developing>.

In order to mitigate risk, the first step is to have a responsible data management process in place (see Section 2.1.3 above on Data Responsibility). Next, utilise Risk Assessment Tools, for example, UN Global Pulse's Risk, Harm and Benefit Assessments Tool which acts as "a data privacy, ethics and data protection compliance mechanism designed to help identify and minimize the risks of harms and maximize the positive impacts of data innovation projects."⁶⁷ This Guide highlights the questions related to the type of data (personal/sensitive), data access and use, proportionality and necessity of data use, legitimacy and fairness of data access and use, due diligence, data minimization, data retention, accuracy and so on.

Action Points:

- Training/workshops: Develop and undertake workshops to educate teams engaged in data science projects and end-users of outputs about the risks of collecting and sharing data.
- Assessments: Always perform risk, harm and benefit assessments prior to undertaking the work.
- Support incident reporting⁶⁸ both internally and externally to document real incidents for awareness to educate the community about potential risk and thus risk mitigation

Now that the Fundamentals have been covered, this Framework will look at the remaining stages of the data science lifecycle: Problem and Solutions exploration, the Data Journey, Algorithm and Reliance on Outputs. Each chapter will be divided into "issues and concerns" followed by "practical guidance" and "action points".

⁶⁷ UN Global Pulse, Risk, Harms and Benefits Assessment Tool, (Jan 2019). Available from: <https://www.unglobalpulse.org/policy/risk-assessment/>.

⁶⁸ The Centre for Humanitarian Data, United Nations Office for the Coordination of Humanitarian Affairs (OCHA), Guidance Note: Data Incident Management (The Hague, 2019) Available from: <https://centre.humdata.org/guidance-note-data-incident-management/>.

2.3 STAGE I: Problem and Solutions Exploration

2.3.1 Issues and Concerns

Humanitarian solutions should be needs-based, not technology-based:

A preliminary ethical consideration of a project using data science is to question whether its use is appropriate, relevant, and necessary to solve the problem at hand. USAID defines this question as “suitability”.⁶⁹ Humanitarian solutions should be needs-based rather than technology-based, as supported by the Signal Code’s Obligations in which the first obligation states that “Humanitarians should ensure that Humanitarian Information Activities are based on the needs of affected populations.”⁷⁰ For this, it is essential to start with a problem from the ground rather than apply a top-down approach.

The temptation to use high-tech to attract donors:

In 2017, amidst the peak of Bitcoin and Blockchain mania, a tea-company changed its name from Long Island Iced Tea to Long Blockchain Corp. Its stock price increased 200% by opening of trading.⁷¹ The humanitarian sector is no exception to this temptation to utilize high-tech for greater visibility or funding. In a push for value for money, donors often see technology as an opportunity to reduce fraud and improve efficiencies. As such, it can be tempting to raise interest in proposed projects by including references to exciting new technologies and buzzwords. However, adding in these buzzwords and seeking to use complex technological applications may be ill-suited for some projects. By adding unnecessary complexity one can run the risk of wasting money and capacity, and at worst, potentially endanger the project either through ineffectiveness, or through mistakes and breaches which could affect vulnerable people.

2.3.2 Practical Guidelines

Those considering data science projects should ensure they have a comprehensive understanding of the problem they are seeking to solve before considering data science methods as a solution. The following steps help guide through the “problem recognition” and “solutions search” steps.

I. Problem Recognition

In humanitarian innovation, the first step of the innovation journey is problem recognition. Elhra’s Humanitarian Innovation Guide describes this process as “identifying a problem or opportunity to respond to, collecting and assessing readily available knowledge on the issue and context, diagnosing root causes, and properly framing the challenge.”⁷² This is the essential first step to finding a solution, as “gaining a deep and insightful understanding of the problem is the key to any successful innovation. The more time spent on understanding the problem, the less the likelihood of developing an inappropriate solution.”⁷³

This rings true with data science projects, as there are situations where data science methods are pushed as a solution regardless. This can occur when a new method is hyped by media as cutting-edge; stakeholders

69 Reflecting the Past, Shaping the Future: Making AI Work for International Development, USAID, 2018, pp.34. Available at <https://www.usaid.gov/sites/default/files/documents/15396/AI-ML-in-Development.pdf>.

70 Campo, S. Nowarth, C. Raymond, N. Scarnecchia, D. The Signal Code: Ethical Obligations for Humanitarian Information Activities, Harvard Humanitarian Initiative, (2018). Available from: <https://hhi.harvard.edu/publications/signal-code-ethical-obligations-humanitarian-information-activities>.

71 Cheng, E. “\$24 Million Ice Tea Company Say it’s Pivoting to the Blockchain, and its stock jumps 200%”, CNBC (21 December 2017). Available from: <https://www.cnbc.com/2017/12/21/long-island-iced-tea-micro-cap-adds-blockchain-to-name-and-stock-soars.html>

72 Humanitarian Innovation Guide, Elhra - Humanitarian Innovation Fund (2019). Available from: <https://higuide.elrha.org/toolkits/recognition/>

73 Ibid.

perceive that a new technology is a possible solution that could address a complex challenge in the humanitarian field, while donors may consider grant submissions more relevant because of the specific technology involved. Additionally, it is worth noting that traditional statistical methods are being rebranded as data science work to catch up with the trend. All such tendencies, whether by internal or external factors must be considered and avoided.

Nethope's AI Suitability Framework describes this problem/solution exploration stage as “defining the opportunity” and urges NGOs to look at the problem they are facing and ask:

“Is your current non-AI/ML solution a repetitive task that would benefit from automation?

What are other benefits you might be looking for in a new solution that are possible with AI? This includes:

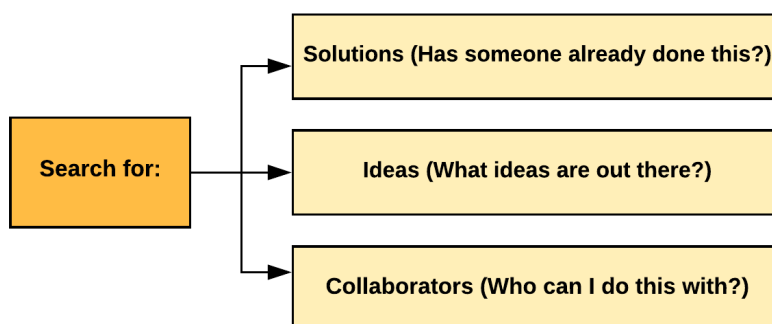
- (a) speed
- (b) accuracy
- (c) cost savings
- (d) massive reach, etc.

It's important to explain why and how current solutions fail to address the identified problem in an expected and adequate way and how AI could help.”⁷⁴

This Suitability Framework contains 32 guiding questions which NGOs should ask themselves throughout a data science project. Their toolkit also offers a workshop plan which can be undertaken at the beginning of a data science project which helps to guide through the problem recognition stage.⁷⁵

II. Search for Solutions

Elhra's Humanitarian Innovation Guide advises that after the problem recognition stage, the next step is to acknowledge that someone else has probably confronted the same issue and a solution might already exist. They advise that one should:



This will enable you to learn from previous experiments/pilots, reduce duplication, and find new potential partners.

Bottom-up innovation and other sources of ideas

As mentioned before, there is a tendency for donors to be attracted to the latest technological solutions – such as exploring the use of AI or blockchain - with little evidence to suggest they may be the best solution to

⁷⁴ Nethope, AI Suitability Framework, (March 2020), Available from: <https://solutionscenter.nethope.org/artificial-intelligence-suitability-toolkit>.

⁷⁵ Ibid.

the identified problem. Such top-down innovation should be questioned. Humanitarian staff who are familiar with the on-the-ground situations as well as the surrounding context and circumstances, are better positioned to identify and explain the challenges and needs required. This is the same for programme staff who are most familiar with internal operations. These staff should be an integral part of breaking down the problem, and therefore determining in consultation with technical experts whether specific data science methods would be a useful and appropriate solution. Humanitarian organisations should also consider (where possible) regularly and actively engaging human rights, data privacy specialists and ethics advisors and other organizational focal points who are actively involved in providing advice or who have expertise related to safeguarding the rights and data of affected populations.

Some larger organizations have departments enabled for data science work, such as WFP's Innovation Labs department and UNHCR Innovation. Whilst these departments can offer a suitable home for new projects involving data science, projects should not be undertaken in siloes.

Problem recognition workshops and structured accelerators, such as those run by the WFP Innovation Lab⁷⁶ and the Dutch Coalition for Humanitarian Innovation,⁷⁷ can offer good environments for the problem recognition stage, which sets up the basis for deciding whether to pursue a specific data science project. Ideally, there should be an open platform to allow for problem sharing, troubleshooting and solution making which could be internal focused or external. For example, events and competitions such as the Humanitarian Action Challenge⁷⁸ and the Hackathon for Good⁷⁹ facilitate such endeavours.

Smaller organisations, or organisations without internal innovation or tech capacities, can either bring in external expertise (consultants) or can try to do the work in-house by following existing guides such as Elhra's Humanitarian Innovation Guide, NetHope's AI Suitability Framework,⁸⁰ or UNDP and UN Global Pulse's "A Guide to Data Innovation for Development: From Idea to Proof-of-Concept."⁸¹

Action Points:

- **Consult your team:** Consultations with your team will help you to better understand the problem you are trying to solve. Speak with your field team, project focal points, programme managers. If possible, speak with legal advisors, data experts and ethicists.
- **Research existing work:** Undertake desk and literature-review and consult with actors who have worked on related projects to identify potential overlaps, avoid duplication and capture potential lessons.
- **Collaborate:** Do stakeholder mapping to identify and strategically combine complementary expertise for humanitarian innovation projects that involve advanced technologies such as AI, and overcome humanitarian sector limitations like lack of funding, domain expertise, technical know-how, managerial support and business acumen.
- **Peer Review:** Check whether any of your peers are willing to review your proposal. You can reach out to organizations that have already undertaken data science projects, or ones with expertise in the domain you will be working in. DSEG can help connect you to them. You can even ask AI companies whether they would be willing to discuss your proposal together.

76 WFP Innovation Lab Accelerator, Available at: <https://innovation.wfp.org/>.

77 Dutch Coalition for Humanitarian Innovation, Humanitarian Innovation Accelerator Program. Available at: <https://dchi.nl/category/humanitarian-accelerator-programme/>.

78 Humanitarian Action Challenge, Impact City, Available at: <https://impactcity.nl/humanitarian-action-challenge/>.

79 Hackathon for Peace, Justice and Security, The Hague, (2018, 2019) www.hackathonforgood.org.

80 NetHope, AI Suitability Framework, (March 2020), Available from: <https://solutionscenter.nethope.org/artificial-intelligence-suitability-toolkit>.

81 UNDP, UN Global Pulse, *A Guide to Data Innovation for Development: From Idea to Proof-of-Concept*, (2016). Available from: https://www.undp.org/content/dam/undp/library/innovation/UNGP_BigDataGuide2016_%20Web.pdf.

2.4 STAGE 2: Data Journey - Collection, Protection and Processing

2.4.1 Data Collection

All data science methods require good data. This section explores some of the most pertinent issues to consider in relation to data collection and practical guidance to avoid data related risks.

2.4.1.1 Issues and Concerns

Obtaining high quality data: AI systems are only as good as the data we put into them.⁸² Obtaining good data in humanitarian contexts can especially be a challenge as data is often scarce, incomplete, not publicly available, expensive and/or faces political barriers to accessing it.⁸³ Data also comes in many shapes and forms including structured and unstructured, digitised and non-digitised. The quality and relevance of data is also subject to its age and frequency of collection. For example, common humanitarian data – household data, is burdensome and time consuming to collect so it isn't frequently updated. Therefore data may be stale and not representative.⁸⁴ Further, some sources of data will come in incompatible formats and structures. For example, there is no universally accepted format for call data records (another common humanitarian data source), so combining datasets from different operators can be very difficult.⁸⁵

In cases of incomplete or inaccessible data, the use of data proxies is common. While proxies can sometimes provide more powerful and equitable alternatives to traditional data sources,⁸⁶ incorrectly applied they will hinder the quality of the outcome and have unintended consequences.

There are also questions regarding the legal and regulatory environment surrounding the data collection, for example, can data collected in-country be stored on servers outside the country? Who owns the data? The local government? Your organization?⁸⁷ These legal questions should be addressed by your legal team.

Bias: If we want AI technology to guide humans in making better humanitarian decisions, then mitigating bias is crucial. Since most AI systems are based on data provided by humans, they can often 'reproduce, reinforce, and amplify patterns of marginalisation, inequality and discrimination that exist in society.'⁸⁸ This is linked to data collection where selection biases and exclusions can occur. The table below outlines some common types of bias that organisations should be aware of. Note that this list is not exhaustive,⁸⁹ and that not all of these biases occur exclusively in the data collection stage – some occur further along the data science lifecycle. Nevertheless, it is useful to have an overview of them in the early stages of the data journey.

82 AI and Bias, IBM Research, Available from: <https://www.research.ibm.com/5-in-5/ai-and-bias/>.

83 Reflecting the Past, Shaping the Future: Making AI work for International Development, USAID, 2019. pp 6. Available at: <https://www.usaid.gov/sites/default/files/documents/15396/AI-ML-in-Development.pdf>.

84 Ibid.

85 Ibid.

86 Ibid.

87 Nethope, AI Suitability Framework, Slide 13 (March 2020), Available from: <https://solutionscenter.nethope.org/artificial-intelligence-suitability-toolkit>.

88 David Leslie, Understanding artificial intelligence ethics and safety A guide for the responsible design and implementation of AI systems in the public sector, Alan Turing Institute, 2019. pp.6 Available at https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf.

89 Two useful resources for more comprehensive descriptions and examples of data science biases include: 1. Crawford, K. The Trouble with Bias - NIPS 2017 Keynote, (2017), Available from: https://www.youtube.com/watch?reload=9&v=fMym_BKWQzk. and 2. Gutttag, J. Suresh H. A Framework for Understanding Unintended Consequences of Machine Learning (2019). Available from: <https://arxiv.org/pdf/1901.10002.pdf>.

Type of Bias	Description	Example
Historical Bias	Captures the world as it is, or as it was. Even if the data is perfectly measured and sampled, the nature of the data (ie. capturing historical discrimination) may produce outcomes that are not wanted. ⁹⁰	If a Google Image search for “CEO” were undertaken, the results would be predominantly male images. This is because historically, CEOs were predominantly male. ⁹¹
Response Bias	Occurs when data is collected from human responses (often online). It causes bias because the responders usually do not represent the full population. ⁹²	Following Hurricane Sandy, researchers used Tweets to try and understand human behaviour post-disaster. However, they found that most Tweets came from Manhattan, and very few came from the severely-hit region of New York. ⁹³
Representation or Selection Bias	Occurs when the full target group is not captured, therefore skewing the dataset.	Data collected by smart-phone responses will only capture information from smart-phone owners which (in humanitarian contexts) is likely to exclude women, elderly and lower-income people. ⁹⁴
Measurement Bias	Caused by human decisions/ categorisations of data which become proxies. This adds another layer of “noise” to the data.	For example, ‘arrested’ is used as a proxy for ‘crime’, or ‘pain medication prescribed by doctor’ is used as a proxy for ‘patient’s pain.’ ⁹⁵
Aggregation Bias	Occurs when a “one-size fits-all” approach is applied to different groups.	Medical applications tend to risk aggregation bias because patients with similar underlying conditions present and progress in different ways. ⁹⁶

⁹⁰ Gutttag, J. Suresh H. A Framework for Understanding Unintended Consequences of Machine Learning (2019). Available from: <https://arxiv.org/pdf/1901.10002.pdf>.

⁹¹ Note that Google has assessed this bias and actively included more female images to try and counter it. See: Gutttag, J. Suresh H. A Framework for Understanding Unintended Consequences of Machine Learning (2019). Available from: <https://arxiv.org/pdf/1901.10002.pdf>.

⁹² Krishnamurthy, P. Understanding Data Bias, Medium (2019). Available from: <https://towardsdatascience.com/survey-d4f168791e57>.

⁹³ Crawford, K. The Hidden Biases in Big Data, Harvard Business Review (2013). Available from: <https://hbr.org/2013/04/the-hidden-biases-in-big-data>.

⁹⁴ Reflecting the Past, Shaping the Future: Making AI work for International Development, USAID, 2019. pp 6. Available at: <https://www.usaid.gov/sites/default/files/documents/15396/AI-ML-in-Development.pdf>.

⁹⁵ Harini Suresh, ‘The Problem with “Biased Data”’, Medium, Available at <https://medium.com/@harinisuresh/the-problem-with-biased-data-5700005e514c>.

⁹⁶ Ibid.

Language Bias	When important data is in a language that cannot be understood or communicated accurately and therefore excluded in a dataset.	In its 2018 Global Report on Internal Displacement, IDMC identified the impact of language bias as affecting the ability to source displacement data comprehensively. ⁹⁷ They could only obtain and analyze information in the languages they speak and read. While their staff and partners speak most of the required languages, they inevitably fail to capture some information, particularly for parts of Asia. These interactions between enumerators and IDPs are not often considered but they can influence data quality and inclusivity. Unless IDPs are able to communicate accurately with those collecting data, there is a risk that they will be mis-counted or misconstrued in reporting. To address this issue, IDMC and Translators without Borders piloted the use of audio recording for verification in multilingual surveys. ⁹⁸
---------------	--	---

2.4.1.2 Practical Guidance

The key to good data is access, adherence to data ethics, representation/inclusivity and completeness.

Access: Accessing good data will likely require stakeholder engagement and negotiation – particularly with governments as they are likely to control the majority of relevant data. To support your request for data, it is important to show that you have robust data protection and management policies and procedures.

Proxies “should be chosen carefully, with an understanding of the local context and the relationship of the proxy to the true outcome of interest.”⁹⁹ For example UNHCR’s Jetson project worked closely with the refugee community to identify a suitable proxy variable (in the case of Somalia the price of goats) to enhance the dataset, recognising that a drop in prices typically reflected large numbers of goats being put up for sale to help fund imminent travel.¹⁰⁰

While proxies may be helpful in some circumstances, it is essential to note that they may weaken the accuracy of your outputs. Projects must be able to recognise when there is insufficient data or data of too-poor quality to input and train a data science tool, and not continue on a project until the data meets this threshold.¹⁰¹

In limited circumstances, data can be repurposed. This means that data collected from previous events or by other organisations or for other purposes can be used again. While this may provide additional and more detailed data, there needs to be thorough checks that it is relevant, suitable and current. Further, if human subjects are involved, consent to re-use the data should be considered.

⁹⁷ Internal Displacement Monitoring Centre, Global Report on Internal Displacement 2018, Methodological Annex p11. Available from: <http://www.internal-displacement.org/global-report/grid2018/downloads/report/2018-GRID-methodological-annex.pdf>.

⁹⁸ When words fail: audio recording for verification in multilingual surveys, Translators Without Borders Blog (2019). Available from: <https://www.translatorswithoutborders.org/blog/when-words-fail/>.

⁹⁹ Reflecting the Past, Shaping the Future: Making AI work for International Development, USAID, 2019. pp 6. Available at: <https://www.usaid.gov/sites/default/files/documents/15396/AI-ML-in-Development.pdf>.

¹⁰⁰ AI and International Development, pp.15 Available at: https://static1.squarespace.com/static/5b156e3bf2e6b10bb0788609/t/5e1f0a37e723f0468c1a77c8/1579092542334/AI+and+international+Development_FNL.pdf

¹⁰¹ Reflecting the Past, Shaping the Future: Making AI work for International Development, USAID, 2019. pp 6. Available at: <https://www.usaid.gov/sites/default/files/documents/15396/AI-ML-in-Development.pdf>.

Bias: There are both technical and non-technical ways to reduce bias in data. Non-technical approaches include critical assessment of the data: run through the above mentioned types of bias to determine whether any might be present in your dataset. Consider:

- How the data was collected;
- The sampling group;
- The collection methodology;
- The age of the data.

If your dataset relates to a specific country, demography or situation, reach out to field-staff or subject experts who have ground-truth knowledge and ask whether the data appears to be fair and representative.

An important aspect of mitigating bias in AI is to ensure that AI models are developed in a multidisciplinary, multicultural, multigender and multi-stakeholder environment so that important issues related to this technology are managed throughout the process. This means inclusion of different experts, field staff and data scientists and other social and cultural groups.¹⁰²

Several organisations provide resources to learn more about assessing bias and fairness, such as the AI Now Institute's annual reports,¹⁰³ the Partnership on AI,¹⁰⁴ and the Alan Turing Institute's Fairness, Transparency, Privacy group.¹⁰⁵

For technical tools to identify and reduce bias, you can use statistical methods to improve training data. Training data should be rigorously checked and tested for bias, the use of data proxies should be justified, and relevant statistical methods should be applied. For example, IBM has developed an AI Fairness 360 Framework tool which includes a comprehensive set of metrics for datasets and models to test for biases, explanations for these metrics, and algorithms to mitigate bias in datasets and models.¹⁰⁶

2.4.2 Data Processing and Protection

2.4.2.1 Issues and Concerns

Once data is collected, the storage and transfer of data raises several ethical questions on privacy, data storage, who has access to it, and the duration of storage. Further still, ethical issues may also arise as humanitarians work in high pressure emergency contexts, where there is not always enough time to be fully trained and fully informed as to the details of technology.¹⁰⁷ This could be problematic when personal and sensitive data is involved and there are potential security breaches.

After data has been collected and stored, it will then likely be processed. Data processing can be understood as any operation performed on data, whether through automated means or not, to produce meaningful outputs.¹⁰⁸

102 AI and Bias, IBM Research, Available from: <https://www.research.ibm.com/5-in-5/ai-and-bias/>.

103 AI Now Institute, Publications, Available from: <https://ainowinstitute.org/reports.html>.

104 Partnership on AI, Research and Publications, Available from: <https://www.partnershiponai.org/research-lander/>.

105 Alan Turing Institute, Fairness, Transparency, Privacy. Available from: <https://www.turing.ac.uk/research/interest-groups/fairness-transparency-privacy>.

106 IBM Research, Trusted AI: AI Fairness 360 Resources. Available from: <http://aif360.mybluemix.net/resources#overview>.

107 Julia Muraszewicz, 'The importance of an Ethics and Privacy Impact Assessment in digital humanitarianism', Hanken University, 2017. Available at <https://blogs.hanken.fi/humlog/2017/05/03/the-importance-of-an-ethics-and-privacy-impact-assessment-in-digital-humanitarianism/>.

108 For more information see GDPR's definition for data processing: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-constitutes-data-processing_en.

In a recent study ‘Making AI Work for International Development’, data processing is divided into three broad categories of sorting, scoring, and discovery, and warns of the possible risks associated such as unfairly excluding or targeting people – in regards to invisible minorities or wrongful predictions.¹⁰⁹ Irresponsible or inappropriate processing of data in humanitarian contexts can place already vulnerable people and communities at greater risk of harm or exploitation, e.g. by exposing their location or identifying a key vulnerability.¹¹⁰ While personal data can categorically be considered sensitive, more complex issues arise from using and processing non-personal data. For example, in conflict settings, the locations of medical facilities can expose patients and staff to risk, even if this data is not personal.¹¹¹ In addition to this, the processing of data for purposes other than those initially specified during collection, and whether those purposes are compatible with the initial purpose is an important ethical question with grave ramifications.¹¹²

A UN Global Pulse report highlights ‘the prominent threat of the ‘mosaic effect,’ which involves matching data-points between separate datasets in order to reidentify an individual or groups of individuals.¹¹³ A major issue is that it is often unknown what other datasets, identifying a certain geographical area or community, are already publicly available, and unintentional linking of datasets may cause significant unforeseeable harms to data subjects.¹¹⁴ These risks, therefore, need to be fully understood before any operations take place and mitigation measures put in place.

2.4.2.2 Practical Guidance

The ICRC Handbook on Data Protection provides useful guidance for how to protect and process humanitarian data. It includes five basic principles: fairness and limited processing, purpose limitation, proportionality, data minimization, and data quality.¹¹⁵

It also emphasizes the relationship between the initial purpose of data collection and the unintended consequences of further processing data subjects. This is directly linked to both the nature of personal data collected, and the safeguards implemented to protect the confidentiality of personal data and the anonymity of the data subject. In the case of humanitarian response, this may involve using open data, crowd-sourcing, social media, and other data points to provide real-time situational awareness in disaster settings. However, the underlying issue with such tools is that they rest on the assumption that there may not be any misuse of this real-time, openly available data.

Another resource which can guide organisations on this topic is the EU Ethics Guidelines for Trustworthy Artificial Intelligence which was published in 2019. It states that principles of technical robustness and safety links directly to how data is stored, and emphasizes that AI systems need to be resilient and secure.¹¹⁶ The need for safety includes “ensuring a fallback plan in case something goes wrong, as well as being accurate,

109 USAID, “Reflecting the Past, Shaping the Future: Making AI Work for International Development”; p. 38.

110 Data Responsibility Guidelines: Working Draft, OCHA, 2019. Pp. 7. Available at <https://centre.humdata.org/wp-content/uploads/2019/03/OCHA-DR-Guidelines-working-draft-032019.pdf>.

111 Ibid.

112 Christopher Kuner and Massimo Marelli, Handbook on Data Protection in Humanitarian Action, International Committee for the Red Cross (ICRC), 2017, I-66. pp. 32. Available at <https://www.icrc.org/en/publication/handbook-data-protection-humanitarian-action>.

113 ‘Big Data for Development and Humanitarian Action’, UN Global Pulse, Available at http://unglobalpulse.org/sites/default/files/Big_Data_for_Development_and_Humanitarian_Action_Report_Final_0.pdf.

114 Ibid.

115 Christopher Kuner and Massimo Marelli, Handbook on Data Protection in Humanitarian Action, International Committee for the Red Cross (ICRC), 2017, I-66. Available at <https://www.icrc.org/en/publication/handbook-data-protection-humanitarian-action>.

116 Ethics Guidelines for Trustworthy Ai, European Commission, 2019. Available at <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.

reliable and reproducible.”¹¹⁷ That is an important way to ensure that any unintentional harm can be minimized and prevented. Another key aspect that this guideline encapsulates is that privacy and data governance should not only ensure full respect for privacy and data protection, but also ensure adequate data governance mechanisms, considering the quality and integrity of the data, and ensuring legitimized access to data.¹¹⁸

Action Points:

- Obtain the best possible data: Demonstrate that the data was collected in a safe and secure manner, with consent. Check for biases through technical and non-technical means. Review data quality by ground-truthing, critically assessing proxies, checking staleness of data.
- Use a Checklist/Assessment tool: If data is available then its use must be dependent on the adherence with operational guidelines for data collection and protection by humanitarian actors such as OCHA, Oxfam, ICRC and Grand Bargain Principles for Coordinated Needs Assessment Ethos.¹¹⁹ Also see The UN Global Pulse Risk Data Innovation Risk Assessment tool which provides a very useful checklist.¹²⁰
- Ensure diversity in teams: Teams working on humanitarian data science projects should come from a multidisciplinary, multicultural, multigender and multi-stakeholder background.
- Data custodian: Develop a clear understanding and working template document of who can access the data and data custodian.
- Internal security measures: Develop data storage and retention policy; clear and foolproof security and safeguarding strategy that is contextually relevant.
- Opting out: Develop a mechanism for individuals to remove themselves from the database, if applicable.
- Clarity: Clearly identify and document intended and possible/unintended consequences of processing this data.
- Test and evaluate: Develop a validation and ground-truthing mechanism to test and evaluate findings.
- Failure mechanism: Develop a strategy that addresses possible inaccuracies, potential harm of data processing as well as further use.

117 Ibid.

118 Ibid.

119 Interagency Standing Committee, Grand Bargain Principles for Coordinated Needs Assessment Ethos, Available from: https://interagencystandingcommittee.org/system/files/ws5_-_collaborative_needs_assessment_ethos.pdf.

120 UN Global Pulse, Risks, Harms and Benefits Assessment Tool, (2018) Available from: <http://unglobalpulse.org/sites/default/files/Privacy%20Assessment%20Tool%20.pdf>

2.5 STAGE 3:Algorithm

To formulate a common set of standards for the ethical use of algorithms in humanitarian contexts, it is both useful and necessary to first define what an algorithm is and what it is not. Definitions used here for algorithms are not all-encompassing, nor is there a universally acknowledged understanding of the definition, but nonetheless, some definition is helpful in terms of contextualizing what falls under the purview of the Framework.

For the purposes of this document, an algorithm can be defined as *a step-by-step, often automated mathematical and mechanical procedure for computing some mathematical function and/or solving a specific problem*. “The word ‘algorithm’ represents an abstract method for computing a function, while a program can be understood as an embodiment of a computational method in some programming language.”¹²¹

Algorithms’ automation power can be useful, but can also alienate human input from processes that affect people. The use or over-use of algorithms can thus pose risks to populations affected by algorithm processes, as human input to such processes is often an important element of protection or rectification for affected groups. Algorithms can often deepen existing inequalities between people or groups, and exacerbate the disenfranchisement of specific vulnerable demographics.¹²² Algorithms, more so than other types of data analysis, have the potential to create harmful feedback loops that can become tautological in nature, and go unchecked due to the very nature of an algorithm’s automation.

In the humanitarian context, algorithms should only be deployed when they constitute a Humanitarian Information Activity (HIA).¹²³ In other words, algorithms should only be employed in needs-based humanitarian situations that directly benefit the population in question. They should be an essential component, part of meeting an identified community need, and should be deployed independently of state, private, and external actor interests. Additionally, use of algorithms should be neutral and impartial.

Directly or indirectly, algorithms can adversely affect data in four primary ways:¹²⁴

1. **Data disparity:** humanitarian actors may have incomplete, inaccurate and/or insufficient data – whether because of a lack of a sharing relationship or because the data may not exist – to make informed decisions about the needs of affected people, which may lead to inaccurate or inequitable responses. This is directly applicable to algorithms in the context of attempting to specify a model or process with too few observations or right hand side variables.
2. **Data deluge:** the amount of data generated by responders and affected communities after a disaster may overwhelm a responders’ ability to make sense of the large flows of information, negatively affecting the ability to make decisions. This often leads to uncoordinated and uncontrolled sharing without accepted protocols for verification and corroboration of data.
3. **Data distortion:** improper and inaccurate analysis of data that either inflates or minimizes the severity of a disaster situation can make responses less effective and waste limited humanitarian resources.

121 Knuth, D. Algorithm and program, information and data: Letter to the editor. Communications of the ACM, (1966), 9(9), 654.

122 Kritikos, M. Scientific Foresight Unit (STOA), European Parliament (Nov 2018). Available from: [https://www.europarl.europa.eu/RegData/etudes/ATAG/2018/624267/EPRS_ATA\(2018\)624267_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2018/624267/EPRS_ATA(2018)624267_EN.pdf).

123 Humanitarian Information Activities (HIAs): Activities and programs that may include the collection, storage, processing, analysis, further use, transmission, and public release of data and other forms of information. HIAs also include the establishment and development of communications capacity and infrastructure by responders and/or populations. These activities occur as part of humanitarian action throughout the response cycle and include, but are not limited to, improving situational awareness; disaster preparedness and mitigation; intervention design and evaluation; connecting populations to response activities and to each other; and supporting ongoing operations, including the delivery of assistance. Campo, S. Nowarth, C. Raymond, N. Scarnecchia, D. The Signal Code: Ethical Obligations for Humanitarian Information Activities, Harvard Humanitarian Initiative, (2018). Available from: <https://hhi.harvard.edu/publications/signal-code-ethical-obligations-humanitarian-information-activities>.

124 Raymond et al. Data Preparedness: connecting data, decision-making and humanitarian response. Available from: https://hhi.harvard.edu/sites/default/files/publications/data_preparedness_update.pdf.

4. **Data damage:** certain uses of data, if not done in a responsible way and in compliance with data privacy and data protection laws, can cause, or be perceived to cause, harm to an organization and also to affected people and their communities. Incidents where data causes damage to affected people can break trust between partners and affected communities.¹²⁵

A 2018 guidance note by the European Parliamentary Research Service considers the following questions when addressing the challenges of including ethical principles into algorithmic design:

- Should the ethical assumptions in the algorithm be transparent and easy for users to identify? (*Legibility of the algorithm*)
- Will people affected by these decisions have any influence over the system? (*Accountability*)
- Even if algorithms were made transparent, how could they be understood by multiple stakeholders of varying technical algorithmic literacy? (*Legibility*)
- Could ethical principles such as fairness and the right to privacy be encoded in the system and, if so, what should those principles be and who should decide upon their choice and weighting? (*Technical Rigour and Ethics*).¹²⁶

For vulnerable populations in humanitarian contexts, accountability and legibility pose acute challenges to be handled with particular care. Unlike civilian contexts, stakeholders in crisis contexts often have the challenges of understanding and influencing algorithms under conditions of distress—making clarity in legibility and accountability all the more necessary for algorithm design in humanitarian contexts.

These questions, and similar questions posed by data scientists globally about the marriage of ethics and algorithmic design give way to a general framework of ethical considerations and concerns that can be broadly placed into four categories, and considered more generally under the umbrella of ‘input’ adjacent concerns and ‘output’ adjacent concerns. Those categories are enumerated below in addition to a complementary categorization from the Oxford Internet Institute:

Category of Ethical Issue and Concern	Methods and Considerations for Mitigating Concerns
Technical Rigour and Properties of the Algorithm (Inputs)	<ul style="list-style-type: none"> • Algorithm validations and testing • Statistical and empirical precedent for the algorithm should exist and be documented • Robustness Checks • Understanding the bias of your algorithm • Clarity on algorithm limitations
Context & Control (Input & Output)	<ul style="list-style-type: none"> • In environments of partnership, government relationships, and donor relations, who controls what part of the algorithm and data lifecycle? • Who has access to the data-input and the output of the algorithm? • Considerations for the “leave-behinds” of the data and algorithm in the affected community • Could the results of the algorithm or the code for the algorithm potentially cause enhanced intelligence capacity for governments or other outside actors? • If the algorithm is being adapted from another context, how is the determination that it is applicable being made?

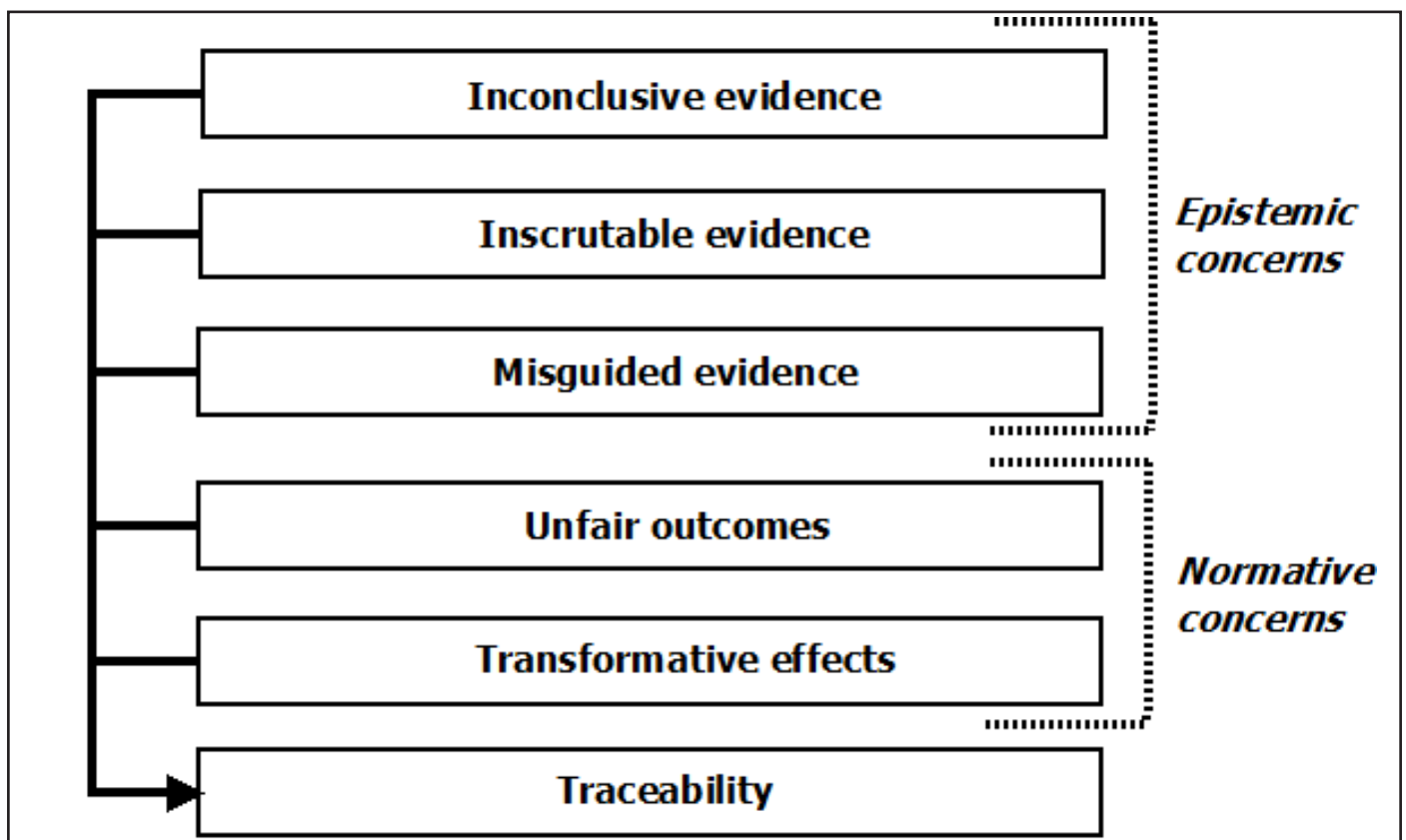
¹²⁵ Raymond et al. Data Preparedness: connecting data, decision-making and humanitarian response. Available from: https://hhi.harvard.edu/sites/default/files/publications/data_preparedness_update.pdf.

¹²⁶ Kritikos, M. Scientific Foresight Unit (STOA), European Parliament (Nov 2018). Available from: [https://www.europarl.europa.eu/RegData/etudes/ATAG/2018/624267/EPRS_ATA\(2018\)624267_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2018/624267/EPRS_ATA(2018)624267_EN.pdf).

Review of Algorithm Pre-Deployment (Inputs & Outputs)	<ul style="list-style-type: none"> • Identification of key stakeholders affected by algorithm • Identify potential algorithm failures and most-affected groups, including detailed analysis of the consequences of a Type I or Type II error (false positive or false negative). • Choose and justify balance between restriction of functional form of the algorithm to better communicate to general audiences, or technically rigorous but opaque algorithm
Communication, Transparency and Accountability (Inputs & Outputs)	<ul style="list-style-type: none"> • Is the algorithm actionable and relevant to the community/ beneficiaries it serves? • Clear communication of what was included and excluded from the algorithm • Replicability of the algorithm • Coherence of code for sharing (even if proprietary, internal sharing and peer review is essential)

The diagram below shows a different but complementary framework of the ethical considerations applicable to algorithmic design. The framework differentiates between technical and epistemic concerns, and normative concerns about the outcomes of the model. Both should be at top of mind for humanitarians.

Source: Mittelstadt et al. “Six Types of Ethical Concerns Raised by Algorithms”



In addition to the above points, it is essential that documentation of the decision-making, testing, and validation process, and all inputs are clear and accessible.

It is also useful here to pause and examine the critical issue of “legibility” in the context of construction and implementation of the algorithm. Legibility refers to the understandability of the algorithm, and the ability of audiences with limited technical capacity to understand the mechanisms and outcomes of the algorithm. There are many ethical issues that arise from a lack of legibility in terms of how we arrive at the conclusions of the algorithm. This should always be considered when weighing technical rigour and functional form against accessibility and legibility.

Ultimately, algorithms and other machine learning-based algorithms have the potential to help humanitarians solve problems more accurately and efficiently. They can be useful if and only if proper consideration is given to ethical implementation and harm reduction throughout the entire lifecycle.

Action Points:

- Validation: Always test the algorithm and run requisite robustness checks. Produce valid outcomes (variables with correct units and inside the proper ranges of validity i.e. number of people cannot be negative number, decimal, or string variables)
- Accuracy: Accuracy should be ensured through establishment of tolerance intervals, rigorous checks for multicollinearity, and use of power calculations (in the case of linear regression analysis) to establish minimum N. A proper 'propagation errors' treatment on the data and their results, given the transformation must be described. Systematized and documented treatment of outliers and missing data, including an understanding of their influence on outcomes, is also an essential component of ensuring accuracy.
- Documentation: Fully document the process of the algorithm development, and make the documentation available when going through the peer review process as well with the presentation of results.
- Peer Review: When possible, run the algorithm through a peer review board like UN-OCHA Centre for Humanitarian Data's Peer Review Framework.
- Legibility: The algorithm should not be opaque to outside stakeholders, and should be clear in its mechanisms and path.
- Communication: Create clear communication points about the objectives, data limitations, algorithm development and limitations so that any subsequent decision making is based on a solid understanding of the results in hand.

2.6 STAGE 4: Reliance on Outputs

Even in this nascent stage of the technology, there is huge potential for data science to support humanitarian operations. While there is no expectation that aid delivery decisions will be based solely on automated analytical processes (and they likely should not be), there is an expectation that there will be an increase in their use to support decision making processes. However, even those decisions which are not entirely automated, i.e. ones with human-in-the-loop, still risk having humans make decisions based on evidence produced by data processed through AI. This will have enormous repercussions for the rights of those affected by the decision. As seen in previous sections, data quality may skew the input data and resulting outputs. The algorithm may ossify past discrimination. Opaqueness may shield bias coded into the system, and legal gray-areas may block an appeal or review of an AI-made decision.

To assess the humanitarian-specific risks that are current or within the foreseeable future, it is useful to reflect on the potential effects of current humanitarian use cases (covered in chapter 3), and question: what could go wrong if the AI doesn't work or produces incorrect/incomplete outputs?¹²⁷

At the extreme, one can perceive:

- A person/community/country not receiving humanitarian aid or receiving the wrong type of/ insufficient/ untimely aid.
- A person's legal status being incorrectly determined (e.g. not receiving refugee status).
- A country shutting its borders due to predicted migration movements.
- A person being placed in danger by data not being protected (e.g. their religious status leaked or analysed satellite imagery of their community falling into the wrong hands).

At a more indirect level, but still detrimental, it may be that aid logistics are adversely affected (which can be both costly and time consuming).

In such worst-case scenarios, not only have the rights of the subject been lost (right to food, water etc.), but the obligations and commitment of the humanitarian organization to accountability have been breached.

2.6.1 False negatives and false positives

“How many errors are acceptable to you?”

While data science methods such as AI are often pitched as being more accurate and efficient than human decision makers, they still make errors. The question above therefore, needs to be addressed.

A false positive is when the AI falsely identifies something in the affirmative. For example, the tool identifies that a community is suffering from desperate food-shortage and requires food aid. If this is not the case, and the community has ample food supplies, then there risks an oversupply of food which would then expire, wasting food and money. A false negative is when the AI identifies something incorrectly in the negative - using our example - the algorithm determines that a different community is not suffering a food shortage and therefore does not receive food aid (or insufficient or untimely aid). The consequences here are far more dire as this could result in sickness and death if the community is not provided with life-saving food. In this example scenario, false positives are far more tolerable than false negatives, and it may be that false negatives are considered completely unacceptable.

¹²⁷ There are clearly many checks and balances in place to prevent such sole reliance on these outputs, however, it is worth considering these worst-case scenarios in order to appreciate the potential adverse consequences of incorrectly applied AI.

2.6.2 Accountability to affected populations

In general, greater accountability in the humanitarian system to communities and people affected by crises is part of the Core Humanitarian Standard on Quality and Accountability (CHS), which sets out nine commitments that organizations and individuals involved in humanitarian response can use to improve the quality and effectiveness of the assistance they provide. In the case of data science methods however, these systems are not morally accountable agents and thus somewhat of an ethical breach in the sphere of applied data science. Targeted principles such as fairness, accountability, sustainability, and transparency are meant to ‘fill the gap’ between the new ‘smart agency’ of machines and their fundamental lack of moral responsibility. The ambiguities surrounding accountability in the use of data science methods raises several questions because these methods may automate cognitive functions that were previously attributable exclusively to accountable human agents. The question of where responsibility falls for an algorithmically generated outcome, a wrong prediction, or prejudiced outcomes therefore becomes unclear. However, the consequences and potential for adverse effects are especially high in the humanitarian context when dealing with some of the most vulnerable populations in complex volatile and insecure environments. As such, the ethical considerations for accountability in the humanitarian system must be linked to accountability of using data science methods for humanitarian work.

Action Points:

- Accountability framework: Identify who is held accountable for the failures of the algorithm across all data science outputs, such as predictions, sorting, and discovery. Address questions of fairness, equality and ethical actions in the project.
- Determine the tolerance of error: How accurate does your tool need to be? How many false positives and false negatives will you accept and what are the costs of both?
- Shared responsibility: Methodology should include discussion on responsibility between the data scientist, authorizing organization, donors and individual authoring and implementation.
- Refer: The Alan Turing Report on Ethical AI contains comprehensive guidance on ensuring responsible delivery of AI outputs, including how to build an ethical implementation platform.¹²⁸

¹²⁸ David Leslie, Understanding artificial intelligence ethics and safety A guide for the responsible design and implementation of AI systems in the public sector, Alan Turing Institute, 2019. pp.6 Available at https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf.

3. Current Humanitarian Data Science Use Cases

The volume of data in the world is increasing exponentially. By some estimates, 90% of the data in the world has been created in the last two years, and it is projected to increase by 40% annually.¹²⁹ Per global discourse, we are therefore seeing greater amounts of humanitarian data being created, collected, processed, and analysed.¹³⁰ For example, in August 2019, the number of datasets available on the Humanitarian Data Exchange (HDX) reached a staggering 10,000, and continues to grow every day.¹³¹ It is now crucial that the humanitarian sector undertake a thorough exploration for the potential of data science to assist its programming and operations with humanitarian ethics as a core concern.

While data science is still relatively new in the humanitarian sector, to illustrate the scope of its application, we draw on a few examples for each of the identified data science methods.¹³²

1. **Natural Language Processing:** text mining and sentiment analysis (such as chatbots)
2. **Prediction and Forecasting:** such as for weather, impact, and finance
3. **Computer Vision:** object detection and classification (such as satellite and drone imagery analysis)
4. **Speech/Audio and Facial Recognition**

3.1 Natural Language Processing

While computers have long been able to process and understand basic text, they traditionally only recognize words with perfect spelling and could only gauge basic relationships between words. The way humans communicate to one another, however, is far from perfect with the use of colloquialisms, words with multiple meanings, tonal nuance, and slurs. This is referred to as natural language, and natural language processing (NLP) is the AI technique used to process and understand natural language.¹³³

NLP is far more advanced than a keyword search: rather than simply identifying and pulling certain words from a document, it can find patterns, and identify relationships between them. It can also process massive quantities of information, far beyond the capacity of a human being.¹³⁴ In the humanitarian sector, NLP has been used for text mining, language translation and sentiment analysis – being able to gauge the emotions and opinions of the audience.

¹²⁹ Big Data for Sustainable Development, United Nations, Available from: <https://www.un.org/en/sections/issues-depth/big-data-sustainable-development/index.html>.

¹³⁰ Centre for Humanitarian Data, UN-OCHA, *The State of Open Humanitarian Data Report*, Available from: <https://centre.humdata.org/the-state-of-open-humanitarian-data-2020/>.

¹³¹ Humanitarian Data Exchange, Centre for Humanitarian Data, UN-OCHA. Available from: <https://data.humdata.org/dataset>.

¹³² The UN's International Telecommunications Union compiled a list of all of the UN's AI activities in 2018. Note that this list includes all UN AI activities – beyond just humanitarian use cases: https://www.itu.int/dms_pub/itu-s/opb/gen/S-GEN-UNACT-2018-1-PDF-E.pdf.

¹³³ Natural Language Processing: Crash Course Computer Science #36. (2017) CrashCourse. Available at: <https://www.youtube.com/watch?v=fOvTtapxa9c>.

¹³⁴ What is Text Mining, Text Analytics and Natural Language Processing? (2019) Linguamatics. Available at: <https://www.linguamatics.com/what-text-mining-text-analytics-and-natural-language-processing>.

Text Mining:

- 510 Red Cross: In order to predict the impact of a future disaster, the Red Cross needs to collect data on the impact of previous disasters i.e. historical data. However, there are few, often incomplete databases with such information, and it is presently scattered throughout hundreds of thousands of documents around the world. 510 uses text mining algorithms to scrape the web, retrieve, and compile relevant information and store it in an internal database to support other projects (such as their forecasting work).¹³⁵
- The Internal Displacement Monitoring Centre (IDMC) uses Idetect, an internal displacement event tagging, extraction and clustering tool which uses web-scraping and NLP to collect and compile human migration data. For example, by analysing anonymized and aggregated data from social media, IDMC can assess the spatial and temporal dimensions of disaster-related displacement.¹³⁶

510 is an initiative of the Netherlands Red Cross which aims to improve speed, quality and cost-effectiveness of humanitarian aid by using and creating data and digital products.

Language Translation:

- Translators without Borders' (TWB) Gamayun project uses NLP and machine learning to build datasets and language technology for marginalised languages, which are then used to enable two-way communications for humanitarian services. TWB's tool incorporates domain-specific terminology and colloquialisms, i.e. expressions that would be used specifically in a humanitarian crisis or by that particular affected community.¹³⁷

Sentiment Analysis:

- Scraping the social media response following a disaster or looking at social media posts in a refugee or IDP camp can help understand public opinion, which in turn can shape humanitarian response.¹³⁸

3.2 Predictive Modelling and Forecasting

While it is not a new field – statistics have used predictive analytics and forecasting methods for a long time – the power and accuracy of these predictions have improved greatly with the advance of computer science. As well as having better access to much larger amounts of data, computer science has introduced algorithms to learn, improve and refine models which process and analyse this data and therefore make more reliable and robust predictions. The humanitarian sector's interest in predictive analytics was brought into the spotlight when senior leadership at the UN acknowledged its potential in humanitarian response financing. UN Under-Secretary General Mark Lowcock stated in March 2018:

“What we need to do is to move from today's approach, where we watch disaster and tragedy build, gradually decide to respond and then mobilize money and organizations to help; to an anticipatory approach where we plan in advance for the next crises, putting the response plans and the money for them in place before they

¹³⁵ Interview with Data Scientist at 510 Red Cross, 26 July 2019.

¹³⁶ Filling the Data Gaps: Innovation at IDMC. (2018) IDMC. Available at: <http://www.internal-displacement.org/innovation>.

¹³⁷ Translators without Borders scales program to develop machine translation for marginalized languages. (2019) Translators Without Borders. Available at: <https://translatorswithoutborders.org/translators-without-borders-scales-program-to-develop-machine-translation-for-marginalized-languages/>.

¹³⁸ See for example, UNHCR's report “From a Refugee Perspective” which portrays the discourse of refugees and migrants and the use of social media: Social Media and Forced Displacement: Big Data Analytics & Machine-Learning, UN Global Pulse and UNHCR Innovation Service, (September 2017), Available from: <https://www.unhcr.org/innovation/wp-content/uploads/2017/09/FINAL-White-Paper.pdf>.

arrive, and releasing the money and mobilizing the response agencies as soon as they are needed.”¹³⁹

This has mobilized the humanitarian sector to further work on predictive analytics, particularly in areas such as humanitarian financing, impact assessment, predicting human migration (including internal displacement and refugees), improving supply chain, identifying poor performers (to allocate appropriate funding) and re-thinking labour costs (by optimizing staffing levels). UN OCHA’s Centre for Humanitarian Data ran a two day workshop in The Hague in April 2019 specifically on Predictive Analytics,¹⁴⁰ and several months after, made Predictive Analytics a core work stream at their Centre for Humanitarian Data.¹⁴¹ Additionally, IOM and the German Federal Foreign Office (FFO) hosted a workshop on forecasting human mobility in the context of crises in October 2019.¹⁴²

In an early webinar on predictive analytics held by OCHA’s Centre for Humanitarian Data in December 2018, the following were listed as potential applications of predictive analytics for OCHA:¹⁴³

1. Early warning: prediction shocks, escalations in current crises or additional crises in complex emergencies;
2. Needs assessment: predicting population affected, migration, and needs by cluster;
3. Anticipatory appeal: predicting length and cost of crises; and
4. Pooled funds allocation: predicting the crises that will be most under-funded.

Early warning - predicting shocks, escalations in current crises or additional crises in complex emergencies:

510 Red Cross have several predictive/forecasting projects underway, aimed at helping with preparedness and early warning/action:

The first involves forecasting the impact of extreme events, such as floods and storms. 510 has developed a typhoon impact forecast, which uses risk data, meteorological data, and historical impact data, to predict the impact distribution of an upcoming typhoon at municipality level. Similar impact forecast models are under development for floods, in Zambia, Uganda, Ethiopia and Kenya. Based on flood forecast models and exposure data, the impact of the flood on people, houses, livestock and crops is predicted.

In the research phase are forecast models for epidemics and drought. For epidemics, the team is researching how the spread of vector borne diseases can be predicted, based on several risk data and dynamic data, such as rainfall, and historical data on mosquito prevalence. The drought modelling work is under research and development for several countries. An initial model was developed for food insecurity in Ethiopia, which again requires identifying the most important variables, and then using these variables to predict future instances of food insecurity.¹⁴⁴

139 Under-Secretary-General for Humanitarian Affairs and Emergency Relief Coordinator, Mark Lowcock: A Casement Lecture: Towards a Better System for Humanitarian Financing, (Dublin 2018), Available from: <https://reliefweb.int/report/world/under-secretary-general-humanitarian-affairs-and-emergency-relief-coordinator-mark-0>.

140 Centre for Humanitarian Data, Workshop Report: Predictive Analytics for Humanitarian Response, (April 2019). Available from: <https://centre.humdata.org/wp-content/uploads/2019/05/Workshop-Report-Predictive-Analytics-April-2019.pdf>.

141 Milano, L. *The Centre’s predictive analytics focus: models, peer review, community*, Centre for Humanitarian Data, (Sept 2019). Available from: <https://centre.humdata.org/the-centres-predictive-analytics-focus-models-peer-review-community/>.

142 International Organization for Migration and the German Federal Foreign Office (FFO), Workshop Report on Forecasting Human Mobility in Contexts of Crises, (London, 2020). Available at: <https://displacement.iom.int/reports/workshop-report-forecasting-human-mobility-contexts-crises>.

143 UN-OCHA webinar on Predictive Analytics, (NYC July 2018), Available from: <https://t.co/XYHSGO6I97>.

144 Interview with Data Scientist at 510 Red Cross, 26 July 2019 and follow up email on 14 October 2019.

Needs Assessment - predicting population affected, migration, and needs by cluster:

World Food Programme's mVAM: In 2013, WFP launched mobile-VAM, or mVAM, to collect high-frequency data from households and key informants. Remote mobile data collection provides a more flexible and efficient way to collect information: surveys are cheaper, faster, and can be conducted even in unstable areas without putting enumerators at risk. Data collection methods are tailored to the needs of each location: the mVAM toolkit includes live calls, Interactive Voice Response (IVR), SMS, online surveys, and two-way communications systems. In 2018, mVAM began transitioning to rolling, near real-time food security monitoring systems to quickly and accurately capture improvements and deteriorations in food security by installing Application Programming Interfaces (APIs) and automated data pipelines.

To address information gaps where mVAM cannot reach, a team of data scientists at WFP have developed a predictive model to estimate or “nowcast” the food security situation in areas where near real-time food security data is not available.¹⁴⁵ Designed based on machine learning algorithms, the predictive model is trained using historical food security data spanning dozens of countries over 13 years. Using information about population density, nightlight intensity, rainfall, vegetation index, conflict, market prices, macroeconomic indicators, undernourishment and past measurements of food security indicators to make predictions. The resulting analysis is displayed on an interactive platform called “Hunger Map LIVE” – WFP’s new global hunger monitoring system that provides near real-time information and estimates on food security in over 90 countries. The platform captures actual data from WFP’s near real-time food security monitoring systems, where available; and predictions from the machine learning model for areas with limited data. Hunger Map LIVE also displays other publicly available datasets related to food security, such as hazards, conflict, vegetation, rainfall, population size, nutrition, undernourishment and macro-economic data.¹⁴⁶ Users can click on a country to drill down on the food security situation at the national and subnational (region/state/province/governorate) levels. These data-driven insights will enable WFP staff, global decision makers and the broader humanitarian community to quickly identify changes in the food security situation and make more informed and timely decisions.¹⁴⁷

Anticipatory Appeal:

The following are some examples of initiatives seeking to predict length and costs of crises:

- IDMC, Disaster Displacement Risk Model: ¹⁴⁸This model is used to calculate displacement risk by looking at the hazard/phenomenon which will cause displacement (i.e. an earthquake) alongside the exposure of the people/buildings in the hazard area, and the vulnerability of how surrounding buildings will react to the hazard. The displacement risk score from this method creates early displacement estimates which aim to improve preparedness and response.¹⁴⁹
- UNHCR, Jetson: seeks to predict forced displacements within and from Somalia into South-Eastern Ethiopia.¹⁵⁰
- World Bank, Artemis: The Famine Action Mechanism (FAM)¹⁵¹ is a global initiative to end famine. It brings

¹⁴⁵ The predictive model makes daily estimations on the number of people with insufficient food intake and the number of people with crisis-level and above coping strategies.

¹⁴⁶ Hunger Map LIVE will be publicly accessible at <https://hungermap.wfp.org/>.

¹⁴⁷ Interview with WFP Programme Manager 21 Oct 2019.

¹⁴⁸ IDMC, Disaster Risk Model, Available from: <https://www.internal-displacement.org/disaster-risk-model>.

¹⁴⁹ For more information on IDMC’s predictive analytics work, view their slides from the UN-OCHA predictive analytics workshop (The Hague, April 2019): https://drive.google.com/file/d/1_BLLyQAgAgL7TgAKAlm4_522oB9kZpFj/view.

¹⁵⁰ For more information on the project, visit their website at: <http://jetson.unhcr.org/#intro>. You can also read blogs from the Jetson team (part of UNHCR Innovation) on their Medium channel: <https://medium.com/unhcr-innovation-service>.

¹⁵¹ World Bank, Famine Action Mechanism, Available from: <https://www.worldbank.org/en/programs/famine-early-action-mechanism>.

together the United Nations, World Bank, International Committee of the Red Cross, Microsoft Corp., Google, and Amazon Web Services. Part of FAM is the Artemis Pilot which is run by the World Bank. Artemis is a model to predict famine and is currently being tested in five countries: Chad, Somalia, South Sudan, Mali, and Niger.¹⁵²

Pooled Funds Allocation – for Anticipatory action/predicting under-funding of crises:

Current examples of such a project include the UN's Central Emergency Response Fund (CERF) who 'by pursuing a more pro-active engagement with country-level actors, using more data and trends analysis, are working to provide funding sooner.'¹⁵³

3.3 Computer Vision

Computer vision is a collection of AI techniques that, among other things, allows machines to identify objects and classify them. In the humanitarian sector, this is currently being used to identify, for example, damaged buildings (post-disaster) or landmarks such as schools and shelters.

Satellite and drone imagery analysis:

- UN Global Pulse and UNOSAT are using AI to detect structures in satellite images during humanitarian crises. The project developed neural network architectures (Mask-RCNN) to detect shelters that were evaluated in images from multiple humanitarian crises in East Africa and the Middle East.¹⁵⁴
- UNICEF's Project Connect uses satellite imagery (plus ML for analysis) to detect and map schools.¹⁵⁵
- Amnesty Decoders is an innovative platform for volunteers around the world to use their computers or phones to help researchers sift through pictures, information and documents.¹⁵⁶ The Decode Darfur Project used digital volunteers to look through satellite images of Darfur villages over time to identify significant changes to buildings and structures. This labelled data was then passed on to machine learning experts to use as training data for a machine learning tool which could analyse the whole of Darfur.¹⁵⁷
- The World Food Program (WFP) also use satellite imagery to support remote sensing work that includes seasonal monitoring, drought prediction, climate analysis, damage assessments, camp dynamics, hot spots analysis and emergency prevention.¹⁵⁸
- Satellite Sentinel Project (SSP) in 2010 Sudan used satellite imagery to provide an early warning system to monitor threats to civilians along the contested border.¹⁵⁹

¹⁵² For technical information about the Artemis Model and overview of its preliminary results, view these slides from the World Bank which were presented at the UN-OCHA predictive analytics workshop (The Hague, April 2019) Available from: <https://drive.google.com/file/d/1KsTlw3pwLVqm8llxcl5Wd8jV-QEam9MU/view>.

¹⁵³ United Nations Central Emergency Response Fund, CERF and Anticipatory Action - Internal Note June 2019. Available from: https://cerf.un.org/sites/default/files/resources/CERF_and_Anticipatory_Action.pdf.

¹⁵⁴ UN Global Pulse, PulseSatellite, Available from: <https://www.unglobalpulse.org/microsite/pulsesatellite/>.

¹⁵⁵ UNICEF, Project Connect. Available from: <https://www.projectconnect.world/>.

¹⁵⁶ Amnesty International, Amnesty Decoders. Available from: <https://decoders.amnesty.org/>.

¹⁵⁷ Amnesty International, Decode Darfur. Available from: <https://decoders.amnesty.org/projects/decode-the-difference>.

¹⁵⁸ Matos, P. Applications of Satellite Imagery in WFP Emergency Operations, World Food Programme (Sep 2018). Available from: <https://docs.wfp.org/api/documents/WFP-0000074334/download/>.

¹⁵⁹ For a critical perspective on the use of satellite imagery, see Raymond, N. et al. While We Watched: Assessing the Impact of the Satellite Sentinel Project, Georgetown Journal of International Affairs (July 2013). Available from: <https://www.georgetownjournalofinternationalaffairs.org/online-edition/while-we-watched-assessing-the-impact-of-the-satellite-sentinel-project-by-nathaniel-a-raymond-et-al>.

3.4 Speech/Audio and Facial Recognition

Speech recognition is a technology which recognizes speech (voice or audio) and converts it into text. The text can then be processed for further application.

- UN Global Pulse developed a Radio Content Analysis Tool which uses AI based speech-recognition technology to convert public discussions on radio broadcasts into text for native languages spoken in Uganda.¹⁶⁰
- In view of establishing complementary communication channels with beneficiaries to collect and share information, World Food Program's (WFP) mVAM utilizes AI to develop chatbots. A chatbot conducts automated conversations with human users via messaging applications to answer users' questions and to disseminate information with communities dealing with food insecurity. Partnering with Leiden University's Centre for Innovation and InSTEDD, mVAM developed and tested two different chatbot builder platforms, AIDA and ChitChat. After the first pilots in Kenya in early 2019, mVAM worked on different use cases and successfully launched chatbots in Nigeria and Peru. Chatbots can be integrated into lightweight websites, simple WFP webpages designed to use up as little data as possible and work in low-bandwidth areas. Such lightweight websites are an ideal platform to disseminate information at minimum data cost and to host the automated chat.¹⁶¹
- Translators without Borders developed Kató Speak, a spoken translation technology, which has far-reaching applications both in humanitarian crises and beyond to address communication issues often faced during a crisis. Kató Speak applies the localization industry's best practices in written translation to voice in order to address the needs of the world's most vulnerable people, many of whom are illiterate. It is also used to collect speech data to build critically needed voice technology in marginalized languages. Currently, it is used in the Rohingya refugee camps as Rohingya is an oral language, with no standardised written script, compounding the need for a spoken translation system.¹⁶²
- Facial recognition and iris scans have been implemented by UNHCR¹⁶³ and WFP for food distribution in camps to use biometric data for effective aid delivery.¹⁶⁴
- Trace the Face, a program implemented by the International Committee of the Red Cross (ICRC), has been reuniting families for decades. With the use of facial recognition, the program has been reuniting families even faster, some dating back to World War 2.¹⁶⁵

3.5 Hybrids

It is also important to note that AI applications often utilize more than one AI tool/technique. For example, UN Global Pulse's Haze Gazer – a crisis analysis tool, is a platform that combines satellite imagery of fire hotspots, census data on population, and real-time data from social media to enhance disaster management efforts by the Indonesian Government. It also allows now-casting of air quality by fusing meteorological data, insights from satellite imagery and photos shared on social media using deep learning.¹⁶⁶

¹⁶⁰ UN Global Pulse, *Analysing Radio Content*. Available from: <https://radio.unglobalpulse.net//uganda/>.

¹⁶¹ Interview with WFP Programme Manager 21 Oct 2019. See also: Centre for Innovation - Leiden University, *ChitChat: A Chatbot to Support WFP*. Available from <https://www.centre4innovation.org/stories/chitchat-chatbot-to-support-world-food-programme-efforts/>.

¹⁶² Translators Without Borders, Kató Speak recognized as a leading innovation 'invader' in the language technology industry, (Dec 2018). Available from: <https://translatorswithoutborders.org/kato-speak-recognized-as-a-leading-innovation-invader-in-the-language-technology-industry/>.

¹⁶³ UNHCR, *Biometric Identity Management System*. Available from: <https://www.unhcr.org/550c304c9.pdf>.

¹⁶⁴ World Food Programme, *Building Blocks*. Available from: <https://innovation.wfp.org/project/building-blocks>.

¹⁶⁵ ICRC Blogs, *Rewards and risks in humanitarian AI: an example*, (Sep 2019). Available from: <https://blogs.icrc.org/inspired/2019/09/06/humanitarian-artificial-intelligence/>.

¹⁶⁶ UN Global Pulse, *Haze-Gazer: a Crisis Analysis Tool* (2016). Available from: <https://www.unglobalpulse.org/2016/03/haze-gazer-a-crisis-analysis-tool/>.

4. Next Steps

There is immense potential for applying data science methods to support humanitarian outcomes if it is used in a careful, controlled and transparent way. It must be recognized that there may be an extreme risk to lives of vulnerable people if data is used inappropriately. We emphasize that any application needs to be problem-driven instead of technologically driven, and that data science often may not be the solution for the problem. Furthermore, when data science can support humanitarian work, there needs to be a critical lens applied to every aspect of the work, in order to fully understand the process, and to mitigate any unintended consequences of the work. Therefore, it needs to be a heavily consulted process, engaging the affected populations, data stakeholders, programme managers and programme staff. Through demonstrated successes (and failures) of various “humanitarian innovations”, projects are typically successful when the problem and potential solutions are identified locally.

In terms of next steps:

- Humanitarian organizations can start using this Framework to contribute to and codify their own ethical guidelines;
- Organizations can document and share lessons learnt by joining the DSEG community and adding to the website (www.hum-dseg.org);
- Interested actors can suggest new tools to be added to the DSEG AI Ethics Toolkit.
- Donors can adapt their funding practices by including ethical parameters as a criterion for funding technologically driven humanitarian projects.

5. Glossary

Our Glossary is compiled from USAID's Making AI Work for International Development and UN Global Pulse's Big Data for Development Reports.¹

Artificial Intelligence (AI)

AI uses computers for automated decision-making that is meant to mimic human-like intelligence. Automated decisions might be directly implemented (e.g., in robotics) or suggested to a human decision-maker (e.g., product recommendations in online shopping); the most important thing for our purpose is that some decision process is being automated. AI often incorporates ML (when using data-driven predictions to make better decisions) but doesn't have to. AI could be called "smart automation."

Algorithm

A systematic procedure for performing a task or solving a problem, often implemented by a computer.

Algorithmic bias

Algorithms make decisions based on data. However, data reflects the social, historical and political conditions in which it was created which can include many biases and discrimination. When AI systems 'learn' based on potentially biased data, the algorithm can reproduce biased, inaccurate, and unfair outcomes.

Bias

Systematically favoring one group relative to another. Bias is always defined in terms of specific categories or attributes (e.g., gender, race, education level). Some types of bias are socially or ethically undesirable.

Big Data

Big Data is an umbrella term referring to the large amounts of digital data continually generated by the global population. Big Data is characterized by the "3 Vs:" greater volume, more variety, and a higher rate of velocity. A fourth V, for value, can account for the potential of Big Data to be utilized for development.

Chatbot

A computational system that engages with human users using natural language. Chatbots typically use text messages or messaging apps (e.g., Facebook Messenger or WhatsApp). Also referred to as "conversational interfaces."

Clean data

The processing of datasets to remove unwanted observations (either duplicated or irrelevant), fix structural errors (housekeeping), handling missing data, and dealing with outliers – avoiding the problem of garbage in-garbage out, and ensuring better data (which beats fancier algorithms).

Data Cleaning

Preparing a dataset for analysis. This may involve standardizing definitions, changing units, removing implausible values, etc.

Deep Learning

A variety of artificial neural networks — models inspired by connections in the animal brain. Simple processing units are connected in layers, with outputs from one layer being input to the next. The number of layers is the "depth" of the network.

¹ Reflecting the Past, Shaping the Future: Making AI work for International Development, USAID, 2019. pp 40. Available at: <https://www.usaid.gov/sites/default/files/documents/15396/AI-ML-in-Development.pdf>; and Big Data for Development: A Primer. Harnessing Big Data for Real-Time Awareness, 2013, UN Global Pulse, 2013, pp. 1. Available at: https://beta.unglobalpulse.org/wp-content/uploads/2013/06/Primer-2013_FINAL-FOR-PRINT.pdf.

Decision Tree

A type of machine learning algorithm in which predictions are made by answering a series of yes/no questions.

Deep Learning

Deep learning is a part of machine learning where the number of layers involved is increased to achieve better results so that it might outperform all models.

Labelling

Providing values of the outcome variable for each instance in a training data set. This may require additional data collection, crowdsourcing, or expert curation.

Machine learning (ML)

Machine learning (ML) is a set of methods for getting computers to recognize patterns in data and use these patterns to make future predictions. ML could be thought of as “data-driven predictions.”

Model

A simplified depiction of reality. ML models consist of an algorithm and parameters that were learned from training data. When an algorithm is combined with training data, we get a predictive model.

Natural Language Processing

Using computers to process a “natural” language spoken and written by humans (e.g., English, French, Arabic).

Predictive analytics

The analysis of data to predict future events, typically to aid in business planning. This incorporates predictive modeling and other techniques. Machine learning might be considered a set of algorithms to help implement predictive analytics. The more business-oriented spin of “predictive analytics” makes it a popular buzz phrase in marketing literature.

Prediction

Guessing an unknown attribute or quality based on known information. ML predictions are not always about the future; they are estimates based on measurable features. Features are often predicted because direct measurement is difficult, dangerous, or expensive.

Proxy

Value that is measured as a substitute for the real quantity of interest. Proxies may be used to make predictions, or as a direct stand-in for things that are hard to quantify (e.g., potential or risk).

Sensitivity

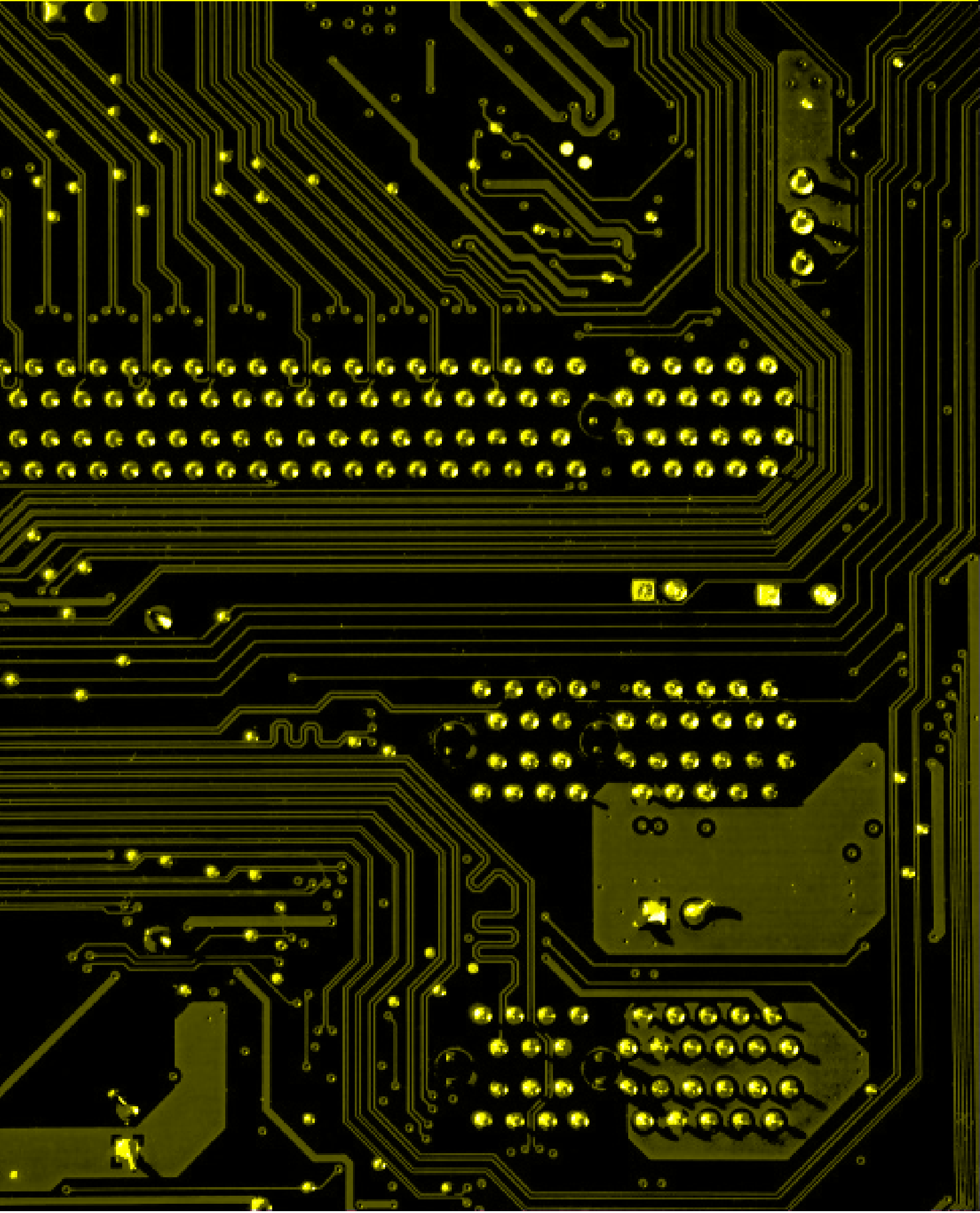
The degree to which outputs change as a single input is changed. Many models will show much higher sensitivity to some features than to others.

Test data

Labeled data that are not used to tune model parameters. Instead, they are set aside in order to evaluate model performance.

Training Data

Data used to develop a ML model. A learning algorithm will find patterns and relationships in training data and use them to define rules for new predictions.



Visit: www.hum-dseg.org