HARVARD Kennedy School
CARR CENTER
for Human Rights Policy

SPRING 2022

CARR CENTER FOR HUMAN RIGHTS POLICY
HARVARD KENNEDY SCHOOL

Humanitarian Digital Ethics

A Foresight and Decolonial Governance Approach

Aarathi Krishnan

Carr Center
Discussion Paper

Humanitarian Digital Ethics

A Foresight and Decolonial Governance Approach

Carr Center for Human Rights Policy
Harvard Kennedy School, Harvard University
January 18, 2022

Aarathi Krishnan
Technology and Human Rights Fellow
Carr Center for Human Rights Policy

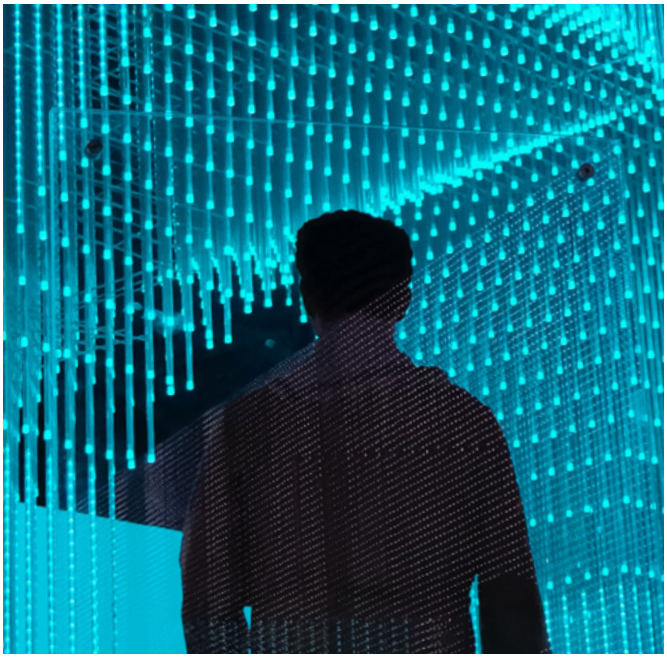
> "Who and what gets fixed in place to enable progress? What social groups are classified, corralled, coerced, and capitalized upon so others are free to tinker, experiment and engineer the future?"
>
> — Ruha Benjamin

Abstract

Just as rights are not static, neither is harm. The humanitarian system has always been critiqued as arguably colonial and patriarchal. As these systems increasingly intersect with Western, capitalist technology systems in the race of 'for good' technology, how do governance systems ethically anticipate harm, not just now but into the future?

Can humanitarian governance systems design mitigation or subversion mechanisms to not lock people into future harm, future inequity, or future indebtedness because of technology design and intervention?

Instead of looking at digital governance in terms of control, weaving in foresight and decolonial approaches might liberate our digital futures so that it is a space of safety and humanity for all, and through this—birth new forms of digital humanism.



Aid Systems and Digital Futures

Our world is at a liminal space—that space between one form of existence and the next. Where our desires, our ideas, our hopes and indeed our dystopias, are imagined in advance, predicted, and laid bare to us, before we can consciously articulate them, and at times, making our decisions for us before we have made it ourselves. Our digital futures are not neutral—nor are those that advocate for its expansion or its mitigation, neutral either. People look at digital futures in awe of their possibilities, how it makes some parts of their lives easier, and recoil in discomfort and horror at how it's used against them, and more pointedly—against those that are disempowered to understand its impact against them. Humanitarians and tech for good advocates analyse statistics of how much of the world is online and what it might mean in terms of who might be 'left behind'. The solving and/or mitigation of the complexities of current and future crisis is relegated to technology—*"it is the instrument that will save the future of humanity and our planet"* as if it's a logical inevitability. As if the very human nature of making sense of and making decisions is now left to the very tools that might be used against those within our social order that don't wield as much influence or power.

The digital futures that lie ahead are uncertain—technology and AI are recasting humanity's understanding of itself and of its place in the world. But it is this very uncertainty that provokes new answers to the question:

What does it mean to be human in these digital futures?

There are complex reasons behind uncertain digital futures. Digital systems exist within complex imperial formations. As communications studies scholar Paula Chakravartty (Gallatin School of Individualized Study, n.d.) suggests in

her studies of new media and racial capitalism, these are all interlocking formations, built on imperial rivalries and a tech worldview that imagines some figures — especially the migrant working classes of the Global South — as outside the world of tech itself. Like the technical architecture of classic colonialism, digital colonialism is rooted in the design of a tech ecosystem for the purposes of profit and plunder. If the railways and maritime trade routes were the "open veins" of the Global South back then, today, digital infrastructure takes on the same role: extraction of data gleaned from the streams of information given up as residents of all countries go online, register for state benefits, and connect with one another through applications whose terms of service demand they give up their personal and private information. As Mohamed, et al. (2020) argue, "the coloniality of power can be observed in digital structures in the form of socio-cultural imaginations, knowledge systems, and way of developing and using technology which are based on systems, institutions and values that persist from the past and remain unquestioned in the present"[1].

The humanitarian aid system is not exempt from the enchantment of digital futures. The term 'humanitarian aid' needs contextualisation for the purposes of this paper. Humanitarian aid exists increasingly in shades of grey. Whilst humanitarian academic research focuses on the classical definitions of 'humanitarian aid' being in response to crisis, 'aid' can cover a plethora of instances — from the pointy end of a crisis, to support during peacetime. Regardless of where in the continuum aid is being provided, it is a truth that over the last two decades, digitalisation and technology have exponentially transformed and is continuing to transform not just how aid is delivered, but *what* aid services are offered and the governing systems that steward aid systems. Digital transformation initiatives have touched on practically every aspect of the aid model:

1. Systems used to support the coordination of aid efforts— for example, biometric digital identities with refugee assistance programs (United Nations High Commissioner for Refugees, n.d.) or earth-observing technology in disaster relief (Borges, 2020).

2. Digital transformation initiatives (The World Bank, n.d.) within developing and least developed countries for digital public infrastructure and e-government services.

3. Technology innovations to support affected populations' access to aid (European Parliament Research Service, 2019) or to have more agency over the aid they receive, or to understand communities and populations—often in the forms of maps, data visualizations or statistical outputs.

4. To tackle, compile and verify information as evidence — for example, ICTs, apps, digital ledgers, digital artefacts

5. To provide more affordable, and faster means to access and share real-time information, connect, and organise.

6. Forecasting trends through predictive analytics (World Economic Forum, 2019), usually related to crisis prevention or early warning.

The diversity of use cases, contexts, stakeholders, and owners does have one common thread, however. As humanitarian organisations do not tend to traditionally have the requisite digital or technology expertise in-house, they partner externally to achieve their aims — predominantly with small communities of public-private, technology partners, and academic institutions from the global north. It is this concentration and self-preservation of powerful interests that raise concern. A humanitarian systems already facing critiques of perpetuating patriarchal, (Chiba & Heinrich, 2019) intersecting with similarly critiqued technology systems risks reinforcing a dominant, hegemonic narrative that assumes:

- The experiences of global civil society and its actors, are homogeneous.
- The singular 'Silicon Valley' values that underpin such digital policies are the only ones that all people aspire to, regardless of where they live or their cultural, societal, economic, geographic bearings.
- Power dynamics will continue to be affirmed in the hands of those that currently hold it without considering the cascading impacts of those policy decisions on those that are most going to be affected by it.

This intersection raises valid concerns on what these impacts might mean for minoritized, historically excluded and oppressed groups that are on the receiving end of aid systems.

"...the relationship between tech industries and those populations who are outside their ambit of power—women, populations in the Global South, including black, Indigenous and Latinx communities in North America, immigrants in Europe—is a colonial one."

— Sareeta Amrute

Perpetuating Harm: An Intertwined Colonial Relationship between Aid and Digital Systems

As Willitts-King et al. (2019) describe, rather than technology being an equalizer of power in the humanitarian system, "fears that [it] will maintain and even further the exclusion" of populations historically overlooked and minoritized are increasing, "particularly [emphasis added] in connection to issues around refugee data and identity... and the trialling of technology on the most vulnerable populations." Examples of drones patrolling the Mediterranean (roborder.eu) to A.I.-powered "lie detectors" (Picheta, 2018), from cell phone tracking (Romm, 2020) to artificially intelligent thermal cameras (Cox, 2020), can all be used against at-risk population groups. Systems that monitor migrants' movements across borders (celebrite.com) don't engender feelings of safety or belonging. Rather, they remove agency from one's physical body as well as from one's digital body. Research (Molnar & Gill, 2018) has shown that technological experiments on refugees and asylum seekers are often discriminatory, breach privacy, and endanger lives—an example seen in how refugees are used to power the work of Silicon Valley without any rights or protection (Jones, 2021). Refugees are often left out of conversations around technological development, and like other marginalized communities, they often become guinea pigs on which to test new surveillance tools before bringing those tools to the wider population (Wright & Verity, 2020). In spaces where automated decision making is used to determine migratory status and resettlement of vulnerable populations, it presents a higher level of risk. This heightened level of risk arises when "(a) the purpose is to target, separate, or distinct a person according to certain population/group characteristics (segregation) in order to automate partially or fully a process for the sake of improving efficiencies, and/or (b) when they completely replace human decision-making processes, and the outcome of these decisions harms directly or indirectly humans" (Coppi et al., 2021).

Wright and Verity (2020) argue that "there are many recent examples of Artificial Intelligence (AI) systems being used for vulnerable people in humanitarian and disaster response contexts, with serious ethical and security-related implications. Impacted populations are put at further risk through biases inherently built into AI systems. There are security concerns regarding their personal information being exposed and even used for persecution purposes. Yet rarely do they have a choice when it comes to the consent of surrendering such information."

As Coppi et al. (2021) note, the accountability gap to affected populations arising from the use of technology solutions has been well known in the humanitarian sector. Recently, the authors put forward the following critical issues in the adoption of AI for principled humanitarian action that present unprecedented risks for humanitarian ethics.

1. Deploying untested technology or AI models (for example biometric technology to prevent fraud in aid claims) on unaware or disempowered individuals who are outside protective legal frameworks as a way of improving digital systems. This approach is also known as 'humanitarian experimentation', a notion that is in direct conflict with the humanitarian principle of Do No Harm.

2 . The opacity that already surrounds Automated Decision Making (ADMs) used in AI models presents unprecedented risks for humanitarian ethics through the possibility of abdication of humanitarian principles in ethical decision-making processes

3. An increase in soft tools in the AI ethics environment hampers the establishment of actual rules and principles that can be mutually agreed to and governed, particularly as ethics are not the only guiding instrument in humanitarian action. Humanitarian actors constantly must make complicated trade-offs between the flexibility of soft ethics guidelines/codes of conduct and the slow rigidity of rights-based frameworks. Without a clear set of agreed to and governable ethics, the question of 'what to follow' becomes malleable.

Despite an abundance of digital device guidelines, data collection methods and soft policy contributions, there is a marked absence of official enforcement, governance or redress policies and standards for harm done to individuals for breaches to their privacy, data protection, or physical injury because of technological failures.

Flawed Humanitarian Digital Democratisation and Ethics

The concerns about the biases of AI and digital systems on populations are arguably slowly getting traction. Some governments are moving to simultaneously limit the power of tech companies with an "urgency and breadth that no single industry has experienced before" (Mozur et al., 2021). Technology firms are working to embed tech ethics (Kinstler, 2020) into their ways of working, though this hasn't been without controversy—as we have recently seen with the firing of Timnit Gebru from Google (Editors, 2020; Ghaffary, 2020)—or notions of ethics washing (Hao, 2019). Campaigns to ban specific technologies that disproportionately harm marginalised and historically oppressed communities are getting traction (i.e., facial recognition technologies in policing; Amnesty International, 2021), **however this hasn't quite extended to the humanitarian space**. This played out in June 2021, when Human Rights Watch published an exposé accusing UNHCR of improperly collecting biometric information for Rohingya refugees in Bangladesh and sharing it with the Myanmar government without the Rohingyas' consent. Zara Rahman (2021) notes that "...data was taken from the bodies of human beings, and shared with those who explicitly want to cause them harm... [this] is a betrayal of their right to self-determination, of dignity, of their very personhood." In August 2021, as Afghanistan fell to the Taliban, it also came to light that biometric data collected by international aid agencies for Afghans were now accessible by the Taliban, raising more concerns of repercussions (Colarossi, 2021).

Where non-traditional stakeholders have ventured into humanitarian work by setting their own principles of humanitarian AI, "humanitarian actors with global or local mandates have been excluded... and paradoxically, considering the humanitarian and ethical principles that should act as framework—the process left out vulnerable populations from the co-design of these new initiatives" (Coppi et al., 2021). This last point was further evidenced in recent research by Nesta (Berditchevskaia et al., 2021) that found that to date, "the practice of participatory AI has been driven by academics... and some major technology companies, they still 'design for' not 'with' users and affected stakeholders. Co-created projects were typically a collaboration between technical team and the frontline users of the model," and there were "almost no examples where communities or stakeholder groups who would be directly impacted by the use of an AI system were involved as 'co-creators' at any stage of model development."

The efforts and ambitions of **digital democratisation in humanitarian aid are complex and require nuance**. As it stands today, digital democratisation in the name of humanitarian aid is fundamentally flawed as it "predicated on a future where soon everyone will have a personal internet connection, a social media account, and therefore will be able to create content and collectively engage online" (Kaurin, 2021). Digital tools for diverse communities are designed based on the assumption that local communities ought to meet on the platforms that international humanitarian organisations are familiar with, rather than localising technology tools to be more easily accessible—i.e., translated into local languages, with UX designed for diversity and use in places with slower Wi-Fi and poorer telecom infrastructures (Kaurin, 2021). As Weitzberg et al. (2021) write, "the use of data technologies in aid interventions is treated by aid organisations and their commercial partners as a straightforward [emphasis added] means of increasing the inclusion, recognition and empowerment of affected populations, often with minimal acknowledgement [emphasis added] of the attendant risks." Supporting this is the argument that scholars and civil society organisations tend to present—the use of data technologies as harm-inducing techno-solutionism or techno-colonialism."

Finally, the **values that the humanitarian system assigns to what determines complex principles such as *rights, fairness, and privacy* do not necessarily consider the cultural contexts** that those values and principles are being applied to. The right to privacy, for example, has been one that has "historically been most difficult to define in a legal framework, due to not only its roots in cultural rituals, but also changing societal and political norms" (Kaurin, 2019). In 2018, Salil Shetty, former Secretary General of Amnesty International, said this of human rights: "human rights often mean different things to different people. And they don't mean anything at all for a good number of people in the developing world." Shetty further argues that "colonialism and early, modern-day human rights fed upon each other... the development and flourishing of the institution of international law itself—with its definition and consolidation of the notions of sovereignty, statehood, trusteeship, and protection—became inextricably linked to the colonial project." By only continuing to draw on normative frameworks for and definitions of these values and principles, it ensures that Western homogeneous definitions of privacy or the rights to be left alone end up being imposed on other cultures.

What this complex confluence of issues and critiques point to is a mosaic of humanitarian digital risks that are not adequately addressed as a whole. Arguably, risks are evolving faster than our capacity to understand them, but what cannot be denied are the harms being imposed on already vulnerable populations. The very current example of biometric data of Afghans that received aid or supported aid programs falling into the hands of the Taliban (Colarossi, 2021) speaks to the fact that despite known technology risks, humanitarians and governments continue to invest in these interventions without necessary consideration of how it might be used against the very people they are trying to protect—not just in the present but also in the future.

Governing Digital Systems in Aid

The appropriateness of digital technologies and AI within aid systems is still being explored and understood. An incredible amount of work has gone into establishing data governance protocols that have not just shifted how humanitarian organisations think about data, but in fact have led to a fundamental shift in data use in humanitarian action. These protocols range from the Handbook on Data Protection in Humanitarian Action by the Brussels Privacy Hub and the International Committee of the Red Cross (2020) to the work being led by the Centre of Humanitarian Data on Data Responsibility (centre.humdata.org/data-responsibility/), the United Nations Development Group Guidance Note on Data Privacy, Ethics and Protection (n.d.), the Signal Code by the Harvard Humanitarian Initiative (Greenwood et al., 2017), and numerous others. Emerging areas of research on AI and Human Rights are calling for a human rights-centred approach to AI Governance (Dawson, 2019). The United Nations Roadmap for Digital Cooperation (2020) goes further into issues of digital public goods, digital cooperation, and digital human rights and points out that "developing countries are largely absent from or not well-represented in most prominent forums on artificial intelligence..." and that "current artificial intelligence-related initiatives lack overall coordination in a way that is easily accessible to other countries outside the existing groupings, other United Nations entities and other stakeholders... and without a broader, more systematic attempt to harness the potential and mitigate the risk of artificial intelligence, opportunities to use it for the public good are being missed." UNESCO's AI Decision Maker's Toolkit (n.d.) aims to elaborate a standard setting instrument on ethics of AI drawing on elements of trends, recommendations, implementation guides, and capability-building resources for the development of a human rights-based and ethical AI.

The commonality across all these protocols as it exists today, is that it **narrows** the focus of governance to issues such as data and AI rather than the broader governance of the use of digital systems **within the systems we intersect with**. Sean McDonald argues in Data Governance's New Clothes (2021) that the issue lies in "problem definition," with two animating frames: (1) data value maximization—"how do we maximise the value of data we are creating"—and (2) solving a problem in context—"can we help solve a problem for a group of people with data or technology." Both are inherently political but perform their interventions differently. Pizzi et al. (2021) also argue that current codes of ethics are limited in that they are "not binding, like law, and hence do not promulgate compliance; they often reflect the values of the organisation that created them, rather than the diversity of those potentially impacted by AI systems; and they are not automatically operationalized by those designing and applying AI tools on a daily basis." In addition to this are questions of **immunity** (traditionally, humanitarian organisations do not go beyond the individual institutional governance mechanisms, which often are bordered by the institution's immunity) and **appropriateness** of the technology innovations we deploy. More urgent is what weaves these protocols together is its focus on the now and its homogeneity—from how we understand harm, protection, human rights, and ethics solutions, and then **what the potential implications might be in the future on impacted populations**.

The fast evolution of AI technology amplifies the risks it brings with it. The confinement and risk to vulnerable people is happening at speed, meaning it is harder to seek justice. The question for humanity then is how we create the architecture of participation around these technologies and balance them with commercial interests for a more equitable, balanced market ideology. It becomes a question of choice. Technologies don't result in change by themselves—they are driven by choices. Once they are in place, they create dynamics on their own. So, what kind of different choices can we make at the outset, where we don't drive an inevitable future where powerful market economies use technology to deepen their self-interests.

Digital technologies and AI mask ideologies of power and are wed to a market ideology of dominance. To intentionally carve a different type of ideology would require governance systems that are informed by different knowledge sources that tangibly **influence** decision making and that **prioritise a focus not just on the firefighting of today but rather the implications on future generations**. That prioritise those that are most impacted by or are on the receiving end of that initiative, rather than only centering donor or aid institutions privileges. How might we embed plurality in governance and ethics that allow us to consider the connected tissue across political, infrastructure and social threads and that embed deeply affected values that frame an intentional equity impact? Kaurin (2021) argues, "digital spaces are representative of ontic spaces; the same challenges regarding gender, age, ableism and discrimination that prevent diverse and representative engagement on the ground are replicated online".

"Digital systems that don't anticipate bad data or adversarial attention typically fail to build the kinds of governance mechanisms and processes necessary to mitigate their impact"

— Sean McDonald

Reimagined Humanitarian Digital Ethics and Governance

This paper posits an *emergent* digital ethics and governance framework for the humanitarian system that specifically aims at interrogating and analysing *context, motives, and impact* of use over a long-term time frame. This focus is to enable safeguards to be built-in to ensure a broader accountability to public interest as well as to ensure that practices and frameworks do not advertently or inadvertently obstruct peoples current and new rights (new rights draw on the argument that rights are neither static nor immutable, and any efforts to design equitable, flourishing futures must consider entirely new forms of harms and rights to be effective [Raman & Schulz, 2020]). As Peter Norvig in *Today's Most Pressing Concerns in AI are Human-Centred* (Lynch, 2021) argues, "exactly what do you want to be optimized? Whose interests are you serving? Are you being fair to everyone? Is the data you collected inclusive? Or is it biased?"

"Ethics is not just being rational and effective. Our choices must also seek to be effective... Acts are, therefore, the ultimate outcome of ethics. The practical field of humanitarian ethics is deliberately known as humanitarian action because of this basic moral insight that ethics without action is nonsensical"

— Hugo Slim

Our current times and futures are complex, uncertain, and continuously evolving. Static frameworks and linear methodologies that are designed on outdated assumptions are not effective in these contexts. Merely saying digital systems ought to be grounded in human rights frameworks is not enough anymore—as the assumptions and models that underpin human rights and traditional ethical frameworks are not equipped for the complexities of our futures. If the fundamental mandate of aid systems is to ensure that those that need aid the most can access it without risk to their lives, their dignity, and their safety, then our understanding of *need, risk, harm, ethics, rights, and empowerment and inclusion* need to evolve beyond current models. Democratic governance of digital systems is required to subvert the current speculative and short-term focus (McDonald, 2021). Such governance is integral at this point not just because it is *the right thing to do* but also because if humanitarian actors don't do so, then they run the risk of contributing to harm and increasing levels of irrelevance. To design more democratic and just future digital systems, humanitarian actors must evolve collectively to optimize the aid model for human well-being, narrow the gap that exists in agony, and provide a "dividend" to historically excluded and oppressed communities (Zolli, 2020).

"What good are human rights frameworks if they prevent you from recognizing the humanity of others?"

— Sabelo Mhlambi

This is a reimagined humanitarian digital ethics and governance approach that de-centres humanitarian coloniality, disrupts the idea of "solutionism," centres justice and equity, and actively works to mitigate harm, now and into the future. It builds on practices and frameworks to date but ensures that the basis of such approaches do not merely "tokenistically include" historically oppressed, excluded, and impacted communities. As Abeba Birhane (2021) calls for, it is an approach that considers "personhood data and justice and everything in between, and places ethics as something that goes above and beyond technical solutions." This approach draws on **foresight and decoloniality** in its design mosaic to ensure mitigation against actors that are incentivised to turn assessment into mere checklist compliance that leaves gaps in resulting accountability regimes (Moss et al., 2021). The framework emphasises these two approaches to promote a *shared responsibility for our collective futures* ***across*** *all actors* and constituents in the aid sector.

"When authority determines meaning, independent thought does not exist, and everything is an extension of the discourse of power"

— Ai Wei Wei

How do we not lock people into future harm, indebtedness, or future inequity?

Currently, digital systems provide greater benefit to the privileged whilst those that are historically excluded, marginalised and oppressed (those most at the receiving end of aid systems) absorb the most harm. Utilising foresight for ethics and governance shifts the focus of current humanitarian digital efforts that prioritise the problem solving of *now*, to one that aims to mitigate future harm and inequity.

Utilising foresight takes it beyond just analysing future technology trends. It aims at not binding or narrowing the digital governance actions of humanitarian actors to merely their institutional legal liabilities and privileges, but rather allowing us to mitigate the systems of harm that pervade and impact our world. Foresight-based ethics builds on and takes further criteria that form the basis of Algorithmic Impact Assessments (AIAs; Moss et al., 2021), ethics-augmented AI Principles (Pizzi et al., 2021), Human Rights Impact Assessments (HRIAs), Forward Engineering Algorithmic Systems (Polack, 2020), and the Intergenerational Fairness Framework (soif.org.uk/igf/), as well as the newly released Participatory AI framework (Berditchevskaia et al., 2021) for humanitarian collective crisis intelligence. It aims to ensure that governance mechanisms for digital systems are fully realized and accountable across a longer period.

In addition, ethics and governance models that are baselined in decolonial theory interrogate patterns of power that shape our intellectual, political, economic, and social worlds. Our models of governance and strategic design require a broader evolution to consider the "sexual, gender, spiritual, epistemic, economic, political, linguistic, aesthetic, pedagogical and racial hierarchies of the "modern/colonial western-centric Christian-centric capitalist/patriarchal world-system" (Grosfoguel, 2011). By embedding a decolonial critical approach within the technical practice of ethical and governance practice, humanitarian digital systems can ensure that the impacts of these systems can amplify impacted populations' ability to flourish in the long term, rather than just to survive in the short term.

The focus on decoloniality means to provide a set of approaches that go beyond diversity, inclusion, and empowerment approaches, which arguably do not necessarily **influence** or **guide** decision making and priority setting. A decolonial approach turns us "towards a pluriversal epistemology of the future that unlike universalism, acknowledges and supports a wider radius of socio-political, ecological, cultural and economic needs" (Mohamed et al., 2020). This shifts the knowledge sources and experiences we draw on in the very **design and decision making** of ethics and governance frames and ensures that we consider the multiplicity of ways in which issues of rights, fairness, privacy, and agency are experienced and understood the world over. Baselining decoloniality allows us to shift current models that are beset with calls for inclusion but still centre decision making on normative and power-centralised status quo.

Foresight and Decolonial Ethics and Governance Framework

Coppi et al. (2021) suggest that the roadmap to humanitarian AI should contain the following: (1) adopting explicability and its proxies as a humanitarian digital tenet; (2) defining a set of metrics for forward engineering a humanitarian AI; (3) enforcing the principle of precaution while building explicable AI systems; and (4) promoting improved legal frameworks. Building on that, this paper puts forward a more granular approach that is firmly grounded on the rights and equity of impacted, minoritized populations. It is aimed at taking owners and designers of digital systems through a series of queries that can be applied to any element of a digital governance system—whether that is related to data, supply chains, investment and resources, legality, legitimacy and authority, due diligence, decision making, and accountability. The framework acts as a compass (not a checkbox) to allow for emergence and relationality to imbue responses.

The framework also draws on the principles as outlined in the AI Decolonial Manifesto (manyfesto.ai/) that posit that notions of "decolonial governances will emerge from community and situated contexts, questioning what currently constitutes hegemonic narratives. Reinventions of AI governance will acknowledge the expertise that comes from lived experience and create new pathways to make it possible for those that have been historically marginalised to have the opportunity to decide and build their own dignified socio-technical futures."

Humanitarian Tech Ethics Assessment Considerations

Foresight-based Ethics	Decolonial-based Governance
What is the **current** and **future** theory of harm that might accompany this implementation of the digital system? Explicitly - which groups of humans will be harmed by this, how, why and when?	What different epistemology on concepts of ethics, privacy, rights, and consent were utilised to inform the design of the digital system? How was this epistemology sourced and from whom?
In what ways might future rights of the target population evolve?	Have impacted users been involved in testing for systems limitations based on their lived experiences and contexts?
Could future impacts of the digital system restrict choice and opportunity for the population?	How are impacted populations inputs into the decision-making design (who, where, what, how), resources and investment allocations weighted?
Who might own the fiduciary and legal accountability to future digital selves of the population?	Who is involved and how is the analysis and interpretation of assessments conducted?
Who would retain advantage and privilege and who would be disadvantaged or dispossessed through the implementation of the digital system, now and into the future?	Has the design of the system considered the wider political, social, economic, and environmental ecosystem in which it will be implemented?
Who might hold responsibility for the absorption of current and future harm on populations?	Has Critical Technical Practice (CTP; Mohamed et al., 2020) been applied to uncover hidden assumptions, privileges, and political biases?
Who might hold responsibility for current and future unintended consequences?	Have definitions of 'fairness' and 'just-ness' been assessed against political and social factors?
What is the plurality of future impacts that this digital system might result in considering an evolved future harm and future rights landscape?	Does the digital system engender feelings of safety or feelings of anxiety by impacted end users?
What possibilities might be foreclosed in the future by implementing this digital system?	Does the governance of the digital system provide pathways for impacted populations to hold institutions to account and to be able to seek redress?
What is the trade off? How many people might benefit from this (and who) versus how many might be harmed (and who)	Have options been provided for local populations to choose which tools are appropriate for them and/or build their own, in their own language? (Kaurin, 2021)?
Transparency in decision making: As answers are uncovered for each query, a weighting of risk or priority is applied to ensure transparency in decision making	
Accountability to all: An external oversight group made up of intentionally diverse constituents to provide independent, principled guidance for the public interest	

What the framework aims, is to ensure the following principles are imbued into the use and deployment of digital systems for humanitarian purposes:

1. **Positionality**: That who holds the responsibility for developing governance, ethics, protocols, and standards of use and deployment consider the positionality of the authorship and decision making within their wider metropolis, and assess the impacts any biases or privileges that it gives rise to.

2. **Future Impacts and Harm Assessment**: The organisations deploying these systems include an expansion on a range of criteria for assessment including plausible, possible, and probable future harms and impacts that might arise on impacted populations and on their future generations. This involves going beyond assessments to ensure that something is used only within its prescribed intent, but rather to assess function creep (Wright & Verity, 2020) current and into the future.

3. **Plurality in Legitimacy**: Utilising a wider range of knowledge sources and experiences to legitimize a multiplicity of conceptual models related to 'ethics' and 'fairness' to ensure assessments and conclusions drawn, intentionally do not replicate an echo-chamber worldview via a limited homogeneous perspective that do not account for normative and cultural realities.

4. **Reverse Accountability against Meaningless Consent**: Ensuring that accountability mechanisms are designed to hold institutions to account so that the burden of harm is not continuously placed on impacted peoples. Specifically, this applies to issues of meaningless consent (Wright & Verity, 2020) where reverse accountability can provide redress and ensures that institutions bear the costs when ethical principles are violated.

5. **Beyond Empowerment and Inclusion**: Moving beyond just 'diversity and inclusion' as a metric for mitigating bias, but ensuring clear, transparent mechanisms that ensure the inclusion of historically excluded, impacted populations can *influence* decision making and provide a re-weighting of decision-making priorities.

6. **Relational Ethics** (Birhane, 2021): Assessing patterns across a wider range of contextual social, technical, economic, and historical systems, norms, and structures to understand *why* rather than merely designing technical solutions and systems blindly.

7. **Transparent Privilege and Dispossession**: Assessing whose rights are privileged and whose are dispossessed in decision making, and the risk of that weighting in the short term and long-term time horizon.

8. **Objective Truth v. Constructed Representation**: Ensuring governance systems are utilised to recognise the context that digital systems will exist in as opposed to a singular representation at a specific point in time (McDonald, 2021), to design malleability and adaptability in changing contexts over time.

> "Our communities are defined by the worst things we permit to happen. What we allow tells the world who we are."
>
> — Anil Dash

Conclusion:

In 2016, Anil Dash recently wrote that "We are accountable for the communities we create." Extending this further, *we are accountable for the futures we create*. Humanitarian actors cannot absolve ourselves of the responsibilities of the intended and unintended consequences of our actions in the short and long term. If our actions enable negative outcomes down the track for the very populations we aim to serve and protect, then our very actions are a fallacy in the name of humanitarian principles. To ensure that we don't continuously relegate speculative responsibilities for 'whatever happens down the line' to impacted populations, current efforts must shift to critical foresight and decolonial approach to understand and shape ongoing advances of digital systems if we want to truly design systems that allow all to flourish.

This framework is not final, it requires further interrogation and validation—and nor should it ever be final. It must continuously evolve and be malleable in its nature, for true anticipatory frameworks cannot be static. Operationalising this approach requires clear pathways to how it influences *decision making and accountability*—as without these, it is merely side-lined to tokenistic gestures of 'inclusion' that continue to affirm power as it looks like today in a status quo that arguably is not fit for the future. As Ahmed Ansari argues "the whole project of democratically-decided AI futures is a cover for essentially continued Anglo-European coloniality and re-asserting global white supremacy because mere representation doesn't necessarily equate to radical alterity" (AI Now Institute, 2021).

Radical alterity, in this sense, is the collective responsibility of all humanitarian actors to not simply expect dehumanizing resilience (Shwaikh, 2021) as a coping mechanism for impacted populations to deal with whatever might come from digital systems in their futures, but rather to radically work towards mitigating and creating systems that don't just say "we leave no one behind" and instead are very intentionally designed through justice, equity and resistance to never do so.

Acknowledgement

I would like to acknowledge the wide range of people across my various tribes who shared their knowledge and wisdom to answer a million questions, patiently stress-test ideas and/or whose thinking has greatly influenced my work. By no means exhaustive but I am grateful to the following: Anasuya Sengupta, Andrew Zolli, Sabelo Mhlambi, Dragana Kaurin, Panthea Lee, Rahul Chandran, Giulio Coppi. Sean McDonald, Nathanial Raymond, Lina Srivastava, Stuart Campo, Caitlin Howarth, Paola Ricaurte, Joana Varon, Chelsea Barabas. Heather Leson, Yves Daccord, David Sangokoya, Chinmayi Arun, and the 2020/21 Fellowship cohort at the Berkman Klein Centre for Internet and Security at Harvard University.

References

AI Now Institute. (2021, July 28). A new AI lexicon: Modernity + coloniality. *Medium*. https://medium.com/a-new-ai-lexicon/a-new-ai-lexicon-modernity-coloniality-7f6979ffbe82

Amnesty International. (2021, January 26). *Ban dangerous facial recognition technology that amplifies racist policing.* https://www.amnesty.org/en/latest/news/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing/

Berditchevskaia, A., Malliaraki, E., & Peach, K. (2021, September). *Participatory AI for humanitarian innovation.* Nesta. https://media.nesta.org.uk/documents/Nesta_Participatory_AI_for_humanitarian_innovation_Final.pdf

Birhane, A. (2021, February 12). Algorithmic injustice: A relational ethics approach. *Elsevier, 2*(2). https://doi.org/10.1016/j.patter.2021.100205

Borges, D. (2020, November 11). How Earth observation can support disaster risk reduction strategies. *PreventionWeb*. https://www.preventionweb.net/blog/how-earth-observation-can-support-disaster-risk-reduction-strategies

Brussels Privacy Hub & International Committee of the Red Cross. (2020, May). *Handbook on data protection in humanitarian action*. https://www.icrc.org/en/data-protection-humanitarian-action-handbook

Chiba, D., & Heinrich, T. (2019). Colonial legacy and foreign aid: Decomposing the colonial bias, international interactions. International Interactions, 45(3), 474–499. https://doi.org/10.1080/03050629.2019.1593834

Colarossi, N. (2021, August 31). Taliban may have access to biometric data used to track Afghans who helped U.S. *Newsweek*. https://www.newsweek.com/taliban-may-have-access-biometric-data-used-track-afghans-who-helped-us-1624666

Coppi, G., Jimenez, R. M., & Kyriazi, S. (2021). *Explicability of humanitarian AI: A matter of principles. Journal of International Humanitarian Action.* https://doi.org/10.1186/s41018-021-00096-6

Cox, J. (2020, March 17). *Surveillance company says it's deploying 'coronavirus-detecting' cameras in US*. Vice. https://www.vice.com/en/article/epg8xe/surveillance-company-deploying-coronavirus-detecting-cameras

Dash, A. (2016, May 27). The immortal myths about online abuse. *Medium*. https://medium.com/humane-tech/the-immortal-myths-about-online-abuse-a156e3370aee

Dawson, P. (2019, November). *Closing the human rights gap in AI governance* [White paper]. Element AI. https://s3.amazonaws.com/element-ai-website-bucket/whitepaper-closing-the-human-rights-gap-in-ai-governance.pdf

The Editors. (2020, December 24). *What Google's firing of researcher Timnit Gebru means for AI ethics*. World Politics Review. https://www.worldpoliticsreview.com/trend-lines/29316/what-google-s-firing-of-researcher-timnit-gebru-means-for-ai-ethics

European Parliament Research Service. (2019, May). *Technological innovation for humanitarian aid and assistance*. https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634411/EPRS_STU(2019)634411_EN.pdf

Gallatin School of Individualized Study. (n.d.). *Paula Chakravartty.* New York University. https://gallatin.nyu.edu/people/faculty/puc1.html

Ghaffary, S. (2020, December 9). *The controversy behind a star Google AI researcher's departure*. Vox. https://www.vox.com/recode/2020/12/4/22153786/google-timnit-gebru-ethical-ai-jeff-dean-controversy-fired

Greenwood, F., Howarth, C., Poole, D. E., Raymond, N. A., & Scarnecchia, D. P. (2017, January). T*he signal code: A human rights approach to information during crisis*. Harvard Humanitarian Initiative. https://hhi.harvard.edu/publications/signal-code-human-rights-approach-information-during-crisis

Grosfoguel, R. (2011). Decolonizing post-colonial studies and paradigms of political-economy: Transmodernity, decolonial thinking, and global coloniality. *TRANSMODERNITY: Journal of Peripheral Cultural Production of the Luso-Hispanic World, 1*(1). http://escholarship.org/uc/item/21k6t3fq

Hao, K. (2019, December 27). In 2020, let's stop AI ethics-washing and actually do something. *MIT Technology Review.* https://www.technologyreview.com/2019/12/27/57/ai-ethics-washing-time-to-act/

Human Rights Watch. (2021, June 15). *UN shared Rohingya data without informed consent*. https://www.hrw.org/news/2021/06/15/un-shared-rohingya-data-without-informed-consent

Jones, P. (2021, September 22). R*efugees help power machine learning advances at Microsoft, Facebook, and Amazon.* Rest of World. https://restofworld.org/2021/refugees-machine-learning-big-tech/

Kaurin, D. (2019, May 15). *Data protection and digital protection for refugees* (Research Paper No. 12). World Refugee Council Research Paper Series. https://www.cigionline.org/publications/data-protection-and-digital-agency-refugees/

Kaurin, D. (2021, September). *Tech localisation: Why the localisation of aid requires the localisation of technology. CDAC Network*. http://www.cdacnetwork.org/tools-and-resources/i/20210930190855-5e47o

Kinstler, L. (2020, February 5). Ethicists were hired to save tech's soul. Will anyone let them? *Protocol*. https://www.protocol.com/ethics-silicon-valley

Lynch, S. (2021, October 11). *Peter Norvig: Today's most pressing questions in AI are human-centered.* Institute for Human-Centered Artificial Intelligence at Stanford University. https://hai.stanford.edu/news/peter-norvig-to-days-most-pressing-questions-ai-are-human-centered

McDonald, S. (2021, July 5). *Data governance's new clothes*. Centre for International Governance Innovation. https://www.cigionline.org/articles/data-governances-new-clothes/

Mohamed, S., Png, MT., & Isaac, W. (2020). Decolonial AI: Decolonial theory as socio-technical foresight in artificial intelligence. *Philosophy and Technology*, 33, 659–684. https://doi.org/10.1007/s13347-020-00405-8

Molnar, P., & Gill, L. (2018). *Bots at the gate: A human rights analysis of automated decision-making in Canada's immigration and refugee system.* International Human Rights Program at Faculty of Law, University of Toronto and the Citizen Lab at Munk School of Global Affairs and Public Policy, University of Toronto. https://ihrp.law.utoronto.ca/sites/default/files/media/IHRP-Automated-Systems-Report-Web.pdf

Moss, E., Watkins, E. A., Singh, R., Elish, M. C., & Metcalf, J. (2021, June). *Assembling accountability.* Data & Society. https://datasociety.net/wp-content/uploads/2021/06/Assembling-Accountability-Policy-Brief_pdf.pdf

Mozur, P., Kang, C., Satariano, A., & McCabe, D. (2021, April 20). A global tipping point for reining in tech has arrived. *The New York Times*. https://www.nytimes.com/2021/04/20/technology/global-tipping-point-tech.html

Picheta, R. (2018, November 2). *Passengers to face AI lie detector tests at EU airports*. CNN. https://edition.cnn.com/travel/article/ai-lie-detector-eu-airports-scli-intl/index.html

Pizzi, M., Romanoff, M., & Engelhardt, T. (2021, March). *AI for humanitarian action: Human rights and ethics* (Article No. 913). International Review of the Red Cross. https://international-review.icrc.org/articles/ai-humanitarian-action-human-rights-ethics-913

Polack, P. (2020, March 20). Beyond algorithmic reformism: Forward engineering the designs of algorithmic systems. *Big Data & Society, 7*(1). https://doi.org/10.1177/2053951720913064

Rahman, Z. (2021, June 21). *The UN's refugee data shame.* The New Humanitarian. https://www.thenewhumanitarian.org/opinion/2021/6/21/rohingya-data-protection-and-UN-betrayal

Raman, S., & Shulz, W. F. (2020). *The coming good society: Why new realities demand new rights.* Harvard University Press.

Romm, T. (2020, March 11). White House asks Silicon Valley for help to combat coronavirus, track its spread and stop misinformation. *Washington Post*. https://www.washingtonpost.com/technology/2020/03/11/white-house-tech-meeting-coronavirus/

Shetty, S. (2018, May 22). *Decolonising human rights* [Speech]. London School of Economics, London, United Kingdom. https://www.amnesty.org/en/latest/news/2018/05/decolonizing-human-rights-salil-shetty/

Shwaikh, M. (2021, May 28). The dehumanizing discourse of resilience. *Progressive Policy Review.* https://ppr.hkspublications.org/2021/05/28/resilience-discourse/

UNESCO. (n.d.). *Building institutional capacity in public policy development in the field - A decision maker's toolkit of AI.* https://en.unesco.org/artificial-intelligence/decision-makers-toolkit

United Nations. (2020, June). *Roadmap for digital cooperation*. https://www.un.org/en/content/digital-cooperation-roadmap/assets/pdf/Roadmap_for_Digital_Cooperation_EN.pdf

United Nations Development Group. (n.d.). *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

United Nations High Commissioner for Refugees. (n.d.). *Biometric Identity Management System*. https://www.unhcr.org/uk/protection/basic/550c304c9/biometric-identity-management-system.html

Weitzberg, K., Cheeseman, M., & Martin, A. (2021, April 1). Between surveillance and recognition: Rethinking digital identity in aid. *Big Data & Society, 8*(1). https://doi.org/10.1177/20539517211006744

Willitts-King, B., Bryant, J., & Holloway, K. (2019, November). *The humanitarian 'digital divide'* [Working paper]. Humanitarian Policy Group at ODI. https://cdn.odi.org/media/documents/The_humanitarian_digital_divide.pdf

The World Bank. (n.d.). *The Digital Economy for Africa Initiative.* https://www.worldbank.org/en/programs/all-africa-digital-transformation/ambition

World Economic Forum. (2019, January). *Civil society in the fourth Industrial Revolution: Preparation and response.* https://www3.weforum.org/docs/WEF_Civil_Society_in_the_Fourth_Industrial_Revolution_Response_and_Innovation.pdf

Wright, J., & Verity, A. (2020, January). A*rtificial intelligence principles for vulnerable populations in humanitarian contexts*. Digital Humanitarian Network. https://www.digitalhumanitarians.com/artificial-intelligence-principles-for-vulnerable-populations-in-humanitarian-contexts/

Zolli, A. (2020, June 10). Humanity and AI: Notes from the field. *Andrew Zolli Blog*. http://andrewzolli.com/humanity-and-ai-notes-from-the-field/

Carr Center for Human Rights Policy
Harvard Kennedy School
79 JFK Street
Cambridge, MA 02138

This publication was published by the Carr Center for Human Rights Policy at the John F. Kennedy School of Government at Harvard University

79 JFK Street
Cambridge, MA 02138

617.495.5819
carrcenter.hks.harvard.edu