Toward a More Equal World: The Human Rights Approach to Extending the Benefits of Artificial Intelligence

Elizabeth D. Gibbons
FXB Center for Health and Human Rights,
Harvard T.H. Chan School of Public Health,
Boston, MA 02115 USA

■ **WE ARE ALL** aware of the huge potential for artificial intelligence (AI) to bring massive benefits to under-served populations, advancing equal access to public services such as health, education, social assistance, or public transportation, for example. We are equally aware that AI can drive inequality, concentrating wealth, resources, and decision-making power in the hands of a few countries, companies, or citizens. Artificial intelligence for equity (AI4Eq) [1] as presented in this magazine, calls upon academics, AI developers, civil society, and government policy-makers to work collaboratively toward a technological transformation that increases the benefits to society, reduces inequality, and aims to leave no one behind. A call for equity rests on the human rights principle of equality and nondiscrimination. AI design, development, and deployment (AI-DDD) can and should be harnessed to reduce inequality and increase the share of the world's population that is able to live in dignity and fully realize their human potential. This commentary argues, first, that far preferable to an ethics framework, adopting a human rights framework for AI-DDD offers the potential for a robust and enforceable set of guidelines for the pursuit of AI4Eq. Second, the commentary introduces the work of IEEE in proposing practical recommendations for AI4Eq, so that people living in high-income countries (HICs), low- and middle-income countries (LMICs), alike, share AI applications' widespread benefit to humanity.

One proxy for "benefit to humanity" is *The United Nations Sustainable Development Agenda*, which was adopted by the UN General Assembly in 2015; 193 nations voted in favor of the Agenda, which includes 17 sustainable development goals (SDGs) for the world to achieve by 2030. The Agenda challenges all member states to make concerted efforts toward the SDGs, and thus toward a sustainable, prosperous, and resilient future for people and the planet [2]. It calls for:

> ……. universal respect for human rights and human dignity, the rule of law, justice, equality, and nondiscrimination; of respect for race, ethnicity, and cultural diversity; and of equal opportunity permitting the full realization of human potential and contributing to shared prosperity.[1]

Central to this agenda is a transformation in well-being that "leaves no-one behind" in the

---

[1]Preamble [2].

realization of the 17 SDGs. AI has enormous potential to advance the SDGs, in the content and purpose of its innovations, and through the process it adopts to produce them, and so catalyze this global transformation. Harnessing the power of AI for the achievement of the SDGs could create a more equal world where every person, whether a resident of HIC or LMIC, can live in dignity and realize their potential.

Human rights framework for guiding AI development to benefit humanity

In the early 21st century, the UN developed a human rights-based approach (HRBA) to development [3], by which both the purpose and the process for socio-economic development were firmly rooted in human rights law. This law is expressed through international treaties codifying political, civil, social, economic, and cultural rights, and treaties protecting the rights of certain populations: women, children, indigenous people, workers, and people with disabilities. Every country in the world has ratified and transformed at least one of these treaties into national law; most have ratified several. This means that all countries have established legal recourse for at least some human rights.

The HRBA process for operationalizing these rights, and the laws they rest upon, incorporates key human rights principles:

· equality and nondiscrimination;
· participation of rights holders;
· accountability that includes responsibility, transparency, and remedy.

The HRBA to international development is suitable not only to a pursuit of the SDGs which leaves no-one behind, but to AI4Eq in which benefit to humanity is catalyzed through an approach to AI-DDD which applies the framework in a holistic manner. In this way, AI applications respect the human rights provisions in ratified treaties, their developers take measures to ensure that these applications promote equality and do not discriminate against any group, are designed and monitored through public participation, and provide for an accountability mechanism to remedy any harms resulting from their use.

How can adopting an HRBA framework rooted in law and operationalized through human rights principles help ethical AI-DDD?[2] Given that states have already ratified some human rights treaties and codified them in the national law, elements of AI regulation to protect the human rights of the population and advance equity already exist. In addition, laws are enforceable, while ethics are not.

With respect to the principle of equality and non-discrimination, given the known risk that AI can facilitate discrimination (particularly against minority and marginalized groups), and that unrepresentative data sets can skew machine learning and AI decision-making, this first principle drives the analysis of an AI application to the question of who is left behind in its benefits. Conscious and frequent application of the human rights principle of non-discrimination bring biases to light, so that AI-DDD can address them before they do irrevocable harm to the discriminated or marginalized segments of the society.

Under human rights law, all people are entitled to meaningful, informed participation in the decisions that affect them and their rights. Through participation, rights-holders, including marginalized groups, can hold public and private actors accountable for the impact of AI on their well-being. However, given the asymmetry of knowledge and power between developers and users regarding, for example, sources of training data, algorithms, and other proprietary information, conscious attention to meaningful participation calls on AI developers to address an absence of platforms for the public review of pilot applications. Developing such platforms enables society to have input and oversight of a potentially biased or harmful application before it is deployed, and also to guide AI toward maximum social benefit.

The final operational principle in the human rights framework for ethical AI is accountability. Accountability is the means through which rights are actually realized and consists of three interdependent elements: *responsibility*—who/which institution has the duty to respect, protect, and fulfill rights, and at what standard; *answerability*—a formal process of transparency whereby the public can demand and receive answers to questions about how those in authority reached their decisions; and finally, *enforceability*—when human rights standards are violated and

---

[2] While this commentary presents an HRBA framework grounded in the United Nations' approach to development as a vehicle for AI4Eq, a focus on human rights provisions and/or principles is present in academic articles, as well as in multilateral and civil society guidelines, although not always in the holistic manner of HRBA. Some examples include [4]–[7].

individual or community harm results, a mechanism exists to sanction those responsible *and* provide a remedy to the persons affected [8]. At present, accountability in ethical guidelines for AI is too often limited to transparency: documenting the development process or publishing independent audits. While transparency is necessary, it is not sufficient to ensure accountable AI; HRBA demands accountability for outcomes that prove harmful and requires a mechanism for remedy to population that may have been negatively affected, including the harm that denies them equitable access to AI benefits.[3]

While the human rights framework should be applied in its entirety to AI-DDD, this commentary continues with a focus on practical application of the principle of equality and nondiscrimination, such that populations the world over can enjoy equal benefit to the AI systems' innovations and equal protection from their potential harms.

Promoting AI4Eq: Equal access and sustainable development

Fulfilling AI's potential for good, by catalyzing a more equal world, poses particular challenges when considering the context of AI-DDD for people living in LMIC. In recognition of these challenges, *IEEE's Ethically Aligned Design* (EAD) [10] project (whose purpose is to provide recommendations leading to AI development that benefits humanity), established the Sustainable Development Committee.[4] The committee (whose multidisciplinary members included academics, lawyers, robotics engineers, businessmen and women, and international development experts) was concerned that there be "equal availability" of access to AI's benefits that would, to use the SDG's driving principle, "leave no-one behind." It recognized that while:

A/IS [autonomous and intelligent systems] are among the technologies that can play an important role in the solution of the deep social problems plaguing our global civilization, contributing to the transformation of society away from an unsustainable, unequal socioeconomic system toward one that realizes the vision of universal human dignity, peace, and prosperity [10, p.142],

this social-justice outcome would not happen without a concerted focus on increasing equal access to AI. As things stand today, the vast majority of AI-DDD takes place in HIC within a homogenous, educated subsector of society. A recent article in the *MIT Technology Review* lamented "clear lack of regional diversity in many AI advisory boards, expert panels, and councils," which reflected an AI world where only the "languages, ideas, theories, and challenges from a handful of regions—primarily North America, Western Europe, and East Asia (were considered)… (this lack of diversity is seen in) the current concentration of AI research (pdf): 86% of articles published at AI conferences in 2018 were attributed to authors in East Asia, North America, or Europe. And fewer than 10% of references listed in AI articles published in these regions are to articles from another region. Patents are also highly concentrated: 51% of AI patents published in 2018 were attributed to North America" [11].

The committee, in acknowledgment of this unequal and undesirable situation, elaborated a series of issues and recommendations in the EAD's "A/IS for Sustainable Development" chapter, devoting two of the chapter's five sections to AI4Eq: "A/IS in Service to Sustainable Development to All" and "Equal Availability" [10, pp. 140–168]. The paragraphs which follow are adapted from those sections.

The concentration of AI-DDD in HIC leaves LMIC behind and undermines AI's potential to provide them equal benefit, and thus serve as a motor for reducing global inequality. EAD's Sustainable Development Committee identified several factors [10, pp. 141, 143, and 148] currently mitigating against equal benefit:

- The concentration of AI creator capacity in a few countries, companies, and citizens (from wealthier, more educated enclaves).
- Lack of the human capital and knowledge required to adapt HIC-developed technologies to resolving problems in the LMIC context, or to develop local technological solutions to these problems.
- Difficulty retaining A/IS capacity in LMIC due to globally uncompetitive salaries.
- Inadequate or nonexistent IT infrastructure needed to support AI.
- Lack of internet access for up to 90% of the population in certain countries.

[3]In a forthcoming article (currently under review for publication in *Global Policy*) this definition of a human rights framework is applied to a total of 15 AI ethics guidelines produced by industry, government and civil society." See [9].

[4]The author of this Commentary was the Chair of the EAD's Sustainable Development Committee.

- Reluctance to provide open source licensing, and unavailability of public datasets to facilitate AI research and development in LMIC.
- Poor adaptability of AI to the needs of specific cultures/countries/regions.
- Lack of active participation of populations expected to use technology.
- Insufficient and unrepresentative data production, biasing AI machine learning.
- Lack of organizational and business models for adapting technologies to the specific needs of different regions.
- Lack of political will to allow people to have access to technological resources.
- Existence of oligopolies that hinder new technological development.

To offset some of these structural obstacles, the committee recommended numerous actions, in both the SDG content and process, which could contribute toward increasing equal availability of AI benefits to people in the LMIC, and reduce the risks of marginalization.

In terms of SDG content toward achieving the goals, this EAD Committee recommended that researchers, AI designers, companies, and policy makers:

- Identify, experiment with, document, and promote AI technologies relevant to the SDGs (i.e., Big data for development relevant to agriculture, medical telediagnosis; global information system (GIS) for public service planning; drone delivery of critical health inputs).
- Develop and apply ethical standards for the collection, use, sharing, and disposal of data in fragile social, economic, and political settings (where privacy breaches can lead to death at the hands of competing militias or oppressive governments).
- Cost and propose strategies for universal public provision of internet services, to diminish the gap between the HIC's and LMIC's access to AI's potential benefit.
- Support civil society organizations advocating for marginalized groups' equal AI benefit and data protection.
- Integrate the SDGs into the core of private sector business strategies and adding SDG indicators to companies' key performance indicators, going beyond corporate social responsibility (CSR).

In terms of the process for AI-DDD that would help advance equal access in LMIC, the Committee recommended some practical steps that various stakeholders and constituencies could take:

- Support LMIC in the development of their own AI strategies/applications, and in preventing brain-drain, through retention or return of their own AI talent.
- Deploy A/IS to detect fraud and corruption, to increase the transparency of power structures, to contribute to a favorable investment, governance, and innovation environment.
- Encourage global standardization and open source AI software.
- Develop public databases (protecting personal data), from which LMIC researchers can develop/train locally appropriate applications.
- Provide fora where LMIC stakeholders can shape AI applications to fit their own cultural, economic, and social environments.
- Form collaborative networks between HIC and LMIC developers, including supporting the latter to attend global knowledge-exchange conferences.
- Promote research and support deployment of mobile, lightweight applications readily available in LMIC.
- Prioritize development of AI infrastructure in international assistance as a necessary precondition to "leaving no-one behind" in AI's benefit to humanity through the SDGs.
- Support research on adaptation of AI development methods to scarce data environments; recognizing that existing methods risk bias and discrimination in low data-density environments.
- Strengthen laws for and practice of data protection.

The "A/IS for Sustainable Development" chapter of the EAD also includes individual sections on A/IS and: Employment, Education, and Humanitarian Action. Each of these sections analyses issues and makes practical recommendations that aim to ensure that AI not only leaves no-one behind but creates conditions where everyone, everywhere, has the opportunity to benefit equally from AI's profound transformation of society.

THERE REMAIN MANY challenges to ensuring equal availability to AI's benefit to humanity; IEEE's Sustainable Development Committee made a start on

recommending some practical steps that could advance that goal. Following these recommendations could also help diversify AI research and patent development to better reflect a world where LMICs have equal access to the resources they require and can create innovative AI applications which contribute to achieving the SDGs, leaving no-one behind, and which reflect the unique cultural, language and socio-economic needs of the societies in which they operate.

Implementing the actions that *IEEE's EAD*, 1st Edition layout for AI-DDD, would take steps toward AI4Eq, toward equality, as called for under the human-rights framework, as well as steps away from discrimination by wealth and national origin. Yet, either embedded in or complementary to these actions must be public participation, accessible mechanisms of accountability (which include remedies for harms to users) and protection through the rule of law, which ensures that AI-DDD respects, fulfills and protects human rights, and indeed serves as a benefit to humanity. ∎

■ References

[1] M. S. Pickin, M. Artaso, and E. Gibbons, "AI4Eq: For a true global village not for global pillage," *IEEE Technol. Soc. Mag.,* to be published.

[2] *Transforming Our World: The 2030 Agenda for Sustainable Development (A/RES/70/1)*, United Nations Gen. Assembly, New York, NY, USA, Oct. 2015.

[3] *The United Nations Sustainable Development Group Universal Values Principle One: The Human Rights-Based Approach*. Accessed: Feb. 18, 2021. [Online]. Available: https://unsdg.un.org/2030-agenda/universal-values/human-rights-based-approach

[4] RightsConToronto. (May 2018). *Amnesty International and Access Now: The Toronto Declaration: Protecting the Right to Equality and Non-Discrimination in Machine Learning Systems*. [Online]. Available: https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/

[5] R. Van Est and J. B. A. Gerritsen, with the assistance of L. Kool, "Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality–Expert report written for the Committee on Culture, Science, Education and Media of the Parliamentary Assembly of the Council of Europe (PACE)," The Hague, The Netherlands: Rathenau Instituut, 2017.

[6] M. Latonero. (Oct. 2018). *Governing Artificial Intelligence: Upholding Human Rights& Dignity (Data and Society)*. [Online]. Available: https://datasociety.net/library/governing-artificial-intelligence/

[7] European Commission. (Apr. 2019). *High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI*. [Online]. Available: https://ec.europa.eu/futurium/en/ai-alliance-consultation

[8] *United Nations Office of the High Commissioner for Human Rights. Who Will be Accountable?* UN Office High Commissioner Hum. Rights, Geneva, Switzerland, 2013.

[9] S. Fukuda-Parr and E. Gibbons, "'Artificial intelligence must respect human rights': What does it mean in global health," Global Policy, Tech. Rep., to be pubished.

[10] (Mar. 2019). *Institute for Electrical and Electronics Engineers (IEEE) Ethically Aligned Design (1st Edition): A Vision for Prioritizing Human Well-Being With Autonomous and Intelligent Systems*. [Online]. Available: https://standards.ieee.org/content/dam/ieeestandards/standards/web/documents/other/ead1e.pdf

[11] A. Gupta and V. Heath, "AI ethics groups are repeating one of society's classic mistakes: Too many councils and advisory boards still consist mostly of people based in Europe or the United States," *MIT Technol. Rev.*, 14 Sep. 2020. [Online]. Available: https://www.technologyreview.com/2020/09/14/1008323/ai-ethics-representation-artificial-intelligence-opinion

Elizabeth D. Gibbons is a graduate of Smith College, Northampton, MA, USA, and Columbia University, New York, NY, USA.

She is currently a Senior Fellow with the FXB Center for Health and Human Rights, Harvard T.H. Chan School of Public Health, Boston, MA, USA, where she participates in initiatives which leverage her expertise in advancing the human rights of children and adolescents. These have included the development of a cross-disciplinary child protection curriculum for graduate students, an online HarvardX and Executive Education courses for child protection professionals, which she managed in her capacity as the Director of the Child Protection Certificate Program. Since 2014, she has been engaged in the exploration of artificial intelligence (AI) and its impact on human rights, with particular attention to the potential for these technologies to affect inequality within and between global societies. As the Chair of the Sustainable Development Committee, she led development of a chapter within IEEE's publication *Ethically Aligned Design*, 1st ed., which identifies issues and makes recommendations for ensuring AI

benefits humanity by contributing to the attainment of the UN Sustainable Development Goals. Prior to her academic appointment at Harvard FXB, she enjoyed a lengthy career in international development, primarily with the United Nations Children's Fund (UNICEF), during which she lived and worked in Togo, Kenya, and Zimbabwe, and served as the Head of UNICEF's offices in Haiti and in Guatemala. She also held several positions in UNICEF's New York Headquarters, including Acting Director, Emergency Operations; Chief, Global Policy; and Deputy Director, Division of Policy and Practice. She is fluent in French and Spanish, and the author of *Sanctions in Haiti: Human Rights and Democracy Under Assault,* as well as a contributing author to several other books.

■ Direct questions and comments about this article to Elizabeth D. Gibbons, FXB Center for Health and Human Rights, Harvard T.H. Chan School of Public Health; egibbons@hsph.harvard.edu.