CDA INSIGHTS 2022

Toward Ethical Artificial Intelligence in International Development

GRATIANA FU

with contributions from
Miriam Stankovich and Anand Varghese

JANUARY 2022

CENTER FOR
DIGITAL ACCELERATION

DAI's Center for Digital Acceleration helps our clients integrate digital tools and approaches across their portfolio, especially in emerging markets. We do this by engaging end users, building digital products, and understanding the broader ecosystems that drive the success of technology-based initiatives. Our clients include bilateral and multilateral donors, private sector companies, foundations, and others seeking to drive positive social change across sectors including health, governance, agriculture, education, and economic growth.

Cover Photo: @pressphoto | everypixel

Design: Jennifer Geib, www.jennifergeib.com

CONTENTS

Acronyms

AI	artificial intelligence
BMZ	German Federal Ministry for Economic Cooperation and Development
EU	European Union
FAccT	Fairness, Accountability, and Transparency
GDPR	General Data Protection Regulation
LMICs	lower- and middle-income countries
ML	machine learning
NLP	natural language processing
SDG	Sustainable Development Goal
USAID	U.S. Agency for International Development

Executive Summary

While many artificial intelligence (AI) tools originated in the United States, Europe, and China, the development and adoption of AI in lower- and middle-income countries (LMICs) have been accelerating rapidly. Fueled by the increasing availability of computational power, improved connectivity, and data, AI tools have the potential to help tackle some of the world's most pressing issues by spurring economic growth, improving agricultural systems, enabling higher-quality education, and addressing health and climate challenges. While applications of AI in LMICs are in their early stages, many pilot projects and technology-driven business models demonstrate the potential for AI to benefit underserved populations, better connect local communities and international technology firms, and improve lives.

However, as with other emerging technologies—from cryptocurrency to 5G—AI presents challenges as well as new opportunities, especially as it transitions from Western settings to LMICs. Broadly speaking, these challenges fall into two categories. The first, touched on only briefly in this paper, are the deficiencies in technology capacity and policy making faced by LMICs. The second set of challenges—and the focus of this paper—have to do with deficiencies inherent to the "architecture" of AI systems and how they are developed.

"Artificial Stupidity" and Algorithmic Bias: The first architectural issue involves "artificial stupidity" and algorithmic bias. In the public imagination, AI can often appear to make decisions without the influence of human foibles and misjudgments. However, AI systems are far from infallible. Even a well-designed algorithm must make decisions based on data which is in turn prone to the same flaws or errors we encounter in all spheres if life. And algorithms often make judgment errors when faced with unfamiliar scenarios. This so-called artificial stupidity can extend still further, to the point where AI may make decisions that not only resemble human misjudgment but reproduce human bias and prejudice.

The "Black Box" Problem: A second ethical issue inherent to the architecture of AI systems is the so-called "black box" problem, faced by high-income nations and LMICs alike. Not only is AI prone to error and bias, but the reasons for faulty decisions are not easily accessed or readily understood by humans—and are therefore difficult to question or probe.

Toward Ethical AI in International Development

As is the case with many emerging technologies, the international development community is faced with a conundrum: on the one hand, we recognize the immense potential that AI tools have in solving some of the more complex development challenges facing LMICs. Indeed, the development community is already piloting these tools across various sectors. On the other hand, AI tools appear to have ethical challenges built into their foundations. These intrinsic challenges are likely to have pronounced effect when AI applications are introduced in LMIC settings.

How can we take a balanced approach that moves us toward more ethical uses of AI in international development while still reaping its benefits? To answer this question, we outline four recommendations focused on key areas of future investment by bilateral and multilateral donors:

Develop or adapt an ethical AI framework aligned to country-specific perceptions of ethics:

AI poses new philosophical and ethical questions that experts, policy makers, and societies at large are only beginning to grapple with. In response, AI researchers in the United States and Europe have developed frameworks through which to examine ethical decision making on AI projects and minimize algorithmic harms. We recommend that the international development community adapt these frameworks, beginning with research to determine if they are partially or wholly applicable to LMICs and to understand how ethics are construed in countries of interest. Organizations and individuals from and in LMICs should be meaningfully incorporated into this research agenda.

Diversify data, designers, and decision makers:

A number of architectural issues in AI have their roots in a lack of diversity—especially a lack of diversity in the training data used to develop AI systems and in the backgrounds of people who design AI systems and decide when they're deployed. The international development community can invest in ways to increase diversity in these areas:

- *Data:* AI systems—often developed in Western contexts with Western-centric training data—need access to training data from the Global South. Without this information, AI tools used in the Global South will reinforce the norms and biases of the society in which that source data is collected.

- *Designers:* Much like the AI research community, the community of AI designers and developers is homogeneous—in terms of both technical and identity group background. Adding more diversity in terms of gender, race, and ethnicity would introduce new perspectives to the AI conversation. Achieving ethical AI will also require an interdisciplinary approach, involving a more diverse group of data scientists, software developers, and statisticians, as well as engaging people who specialize in complementary fields such as history, law, and anthropology.

- *Decision makers:* While improvements can be made to AI systems and the processes by which they are developed, issues such as the black box problem stem as much from their application as from their inherent architecture. We need more gatekeepers equipped to make informed decisions about when—and when not—to deploy AI.

Develop metrics to guide ethical AI implementation in LMICs:

Frameworks are the first step, but they can only take AI ethics so far. The development community should develop clear metrics to help AI designers and deployers determine if they are taking adequate steps to counter or mitigate AI bias.

Cultivate partnerships between Global South and Global North:

Adapting ethical frameworks, increasing diversity, and developing clear metrics will demand increased partnership between developed and developing countries. Building on existing AI partnerships, especially North-South and South-South relationships, will create a community and nurture conversations that inform foundational research, data sharing, metrics, and technical assistance for governments and policy makers.

AI in International Development

AI: Definitions

AI refers to the ability of digital tools to perform tasks and act in ways that have historically been associated with human beings, such as seeing and speaking. An umbrella term that encompasses many different methods and abilities to mimic human intelligence, AI uses programs and algorithms that allow computers to attempt to understand languages, speak, see, and recognize images. However, as Meredith Broussard observes in her study, *Artificial Unintelligence*, the reality of what AI can currently achieve differs markedly from the inflated perceptions of the general public, who often believe—erroneously— that AI can do everything a human can do.

AI can be divided into three types: narrow, general, and super.

- *Narrow AI* is currently the only type of AI that exists in the real world. It can complete only discrete tasks that computer scientists have programmed it to do, such as playing a specified song on Spotify after "hearing" a voice command (requiring the AI-driven device to decode a series of syllables) through a virtual assistant such as Alexa or Siri.

- *General AI* is what many people envision when they think of AI—machines that exhibit real human intelligence and can feel, innovate, or emote, like HAL 9000 from the *Space Odyssey* series or Samantha from the film *Her*. Outside such fictional realms, scientists have yet to code human emotions, awareness, and consciousness into machines.

- *Super AI* is a step above general AI and is defined as AI that surpasses human capacity. Again, super AI is a theoretical concept, not yet achieved.

Machine learning (ML) and AI are often used synonymously, though they are not the same. ML is only one branch of AI; the term describes a computer's ability to use data to automatically improve its performance on a given set of tasks without being explicitly programmed to do so. Like all subfields of AI, ML attempts to mimic components of human intelligence—in this case, our ability to learn. Other branches of AI include natural language processing (processing and generating language), computer vision (seeing and understanding images), and expert systems (making the most logical decisions based on previous knowledge).

Data and big data are key to the success of AI tools. Data refers to any information—quantitative or qualitative—that can be used to learn or make decisions. Big data is data that is too great in volume or too complex to be analyzed using traditional methods of data analysis. There are many uses for data in AI, including using data to "train" AI systems (or "training data"), make decisions about the use of AI, or evaluate its effectiveness. To create many automated AI tools, computers are given training data in which they identify patterns that are then used to perform tasks. This underlying mechanism means that computers' ability to perform is heavily dependent on the data they are provided. Theoretically, the more data they are given, the better they perform. But this theory does not always play out in practice. As discussed later in this paper, data can be flawed, and the larger the dataset, the harder it is to identify those flaws.

Applications in Development

Despite AI's current limitations, the international development community has begun to incorporate AI into its programming, albeit in limited and often experimental ways. As Lindsey Andersen concludes in the *Journal of International and Public Affairs*, "Today, most AI initiatives in international development are still in the research, development, and piloting stage. Most rely on a few broadly available data sources such as satellite imagery, mobile phone data, and survey data. These data sources have enabled the development of AI systems in areas such as agriculture and healthcare. The use of AI in international development is likely to become more prevalent now that Amazon, Google, and Microsoft have all introduced cloud-based AI, significantly lowering the cost of running AI systems."[1]

The examples below provide a snapshot of AI technologies being developed and used in LMICs. They illustrate the potential for AI to benefit underserved populations, strengthen relationships between local communities and international technology firms, and improve lives.

AI Use Cases in LMICs and International Development

Agriculture

South African startup Aerobotics[2] uses drones and satellite images to help farmers optimize crop yields in Malawi, Zimbabwe, and Mozambique. The tool provides guidelines to identify areas underperforming due to pests or other deficiencies, manage inventories, and track the impact of farming interventions on crops.

Planet Labs is using satellite images to monitor agricultural development[3] in Kenya, floods in Sri Lanka, and the growth of informal settlements[4] in Dar es Salaam. In Tanzania, the company trained an AI model to identify areas that exhibit expanding footprints in satellite imagery, and combined this information with 3D reconstruction of satellite images to measure building height. This process enabled them to identify population changes and inform urban planning and policy making.

Education

Geekie,[5] an adaptive learning startup in Brazil, is using ML to provide tailored virtual tutoring to students. Its web and mobile applications adapt to the needs of individual students. The tool tracks student progress and uses that information to improve its tutoring. The more a student uses it, the better the tool becomes. Geekie has been accredited by Brazil's Ministry of Education.

Finance

Kudi.ai,[6] a Nigerian chatbot system, allows people to make payments and send money via messaging. The AI-based tool was created to make it easier and cheaper to pay bills and pay others. Built into commonly used messenger systems such as Facebook Messenger and Telegram, Kudi has grown from a chatbot to a full financial services company that offers a variety of banking resources.

Health

Aajoh,[7] another Nigerian product, is developing an AI system for remote medical diagnosis to deal with a severe shortage of doctors in the country. Given a patient's medical condition or symptoms, Aajoh is able to identify the potential disease or health issue. The tool is limited in the number of diseases it can diagnose but could improve over time as it is given more data on more health conditions.

AI has also been deployed to anticipate outbreaks of diseases such as Zika and dengue fever. By partnering with a startup called AIME (Artificial Intelligence in Medical Epidemiology)—which analyzes local government datasets in combination with satellite imagery—Brazilian NGO Viva Rio was able to deliver low-cost quarterly predictions of where diseases may spike. Following its success in Brazil, AIME's low-cost solution was deployed in the Dominican Republic as well.[8]

The dearth of trained lab techs in many low-resource, high-disease-burden areas means that microscopy—the gold standard of disease diagnosis—is not possible. Researchers at Makerere University in Uganda have developed a way to use 3-D printing and computer vision on smartphones to capture and process microscope images. The tool can diagnose malaria in blood smears, tuberculosis in sputum, and intestinal parasite eggs in stool samples.

Environment

The World Wildlife Fund is using AI-powered thermal cameras in Kenya to apprehend wildlife poachers.[9] The cameras can automatically identify people entering their fields of view and notify rangers of potential poaching. National parks and conservation areas around the world—including in Nepal, Zambia, and Tanzania—have begun to adapt similar technology.

Disaster Assistance

The United Nations is using natural language processing to analyze radio content[10] in Uganda, gain insight into public opinion, and assess the effectiveness of UN programs. After collecting radio audio files in two languages (Luganda and Acholi), the team used automatic speech recognition to convert the audio into written text. They then filtered the text to look for keywords and phrases relevant to perceptions of refugees, the impact of disasters on livelihoods and health, and the effectiveness of radio campaigns.

The Geospatial Operations Support Team at the World Bank's Global Facility for Disaster Reduction and Recovery used satellite and drone imagery of three Guatemalan neighborhoods to identify buildings potentially vulnerable to seismic activities and in need of upgrading. They used ML to analyze characteristics such as rooftop material and elevation grades.[11]

Workforce Development

To help the human resources team on USAID's Women in the Economy project in Afghanistan quickly sort the large number of CVs they received, DAI data scientists built an AI tool that uses natural language processing to categorize women's resumes according to thematic areas (law, education, and so forth).[12] Before they had the tool, staff had to review copious documentation and match CVs to relevant job descriptions. By narrowing down their selections, the tool helped the staff more efficiently match applicants with opportunities.

European startup SkillLab has developed an AI-based skill assessment tool to identify and document the professional skills of job seekers quickly and in any language, with a focus on helping refugees find employment. SkillLabs' mobile app helps jobseekers capture their skills, explore careers, and generate job applications.[13]

While these examples demonstrate ongoing sector-specific applications of AI, experts also predict broader economic benefits resulting from the growth in AI capabilities around the world. A 2018 PricewaterhouseCoopers report on "The Macroeconomic Impact of Artificial Intelligence," for example, predicts that AI could contribute up to $15.7 trillion to the global economy in 2030.[14]

In the context of governance, big data combined with AI can enhance decision making and improve accountability. For example, the ability of big data to encompass an entire population, rather than relying on a small sample, enables analysts to mitigate selection bias. New sources of data, new technologies, and new analytical approaches, if applied responsibly, can enable more agile, efficient, and evidence-based decision making in support of the 2030 UN Agenda for Sustainable Development.[15]

AI holds much promise for LMICs. While many AI tools have originated in the United States, Europe, and China, the development and adoption of AI in LMICs have been increasing rapidly. For example, a GSMA study identified more than 3,000 current and potential AI use cases for small and medium-sized enterprises in sectors such as health, education, and agriculture across LMICs in Sub-Saharan Africa, North Africa, and South and Southeast Asia. Fueled by the increasing availability of computational power, improved connectivity, and more data, AI has the potential to help tackle some of the world's most pressing issues by accelerating economic growth, improving agricultural systems, enabling high-quality education, and addressing health and climate challenges.

Ethical Challenges Posed by AI

As with other emerging technologies—from cryptocurrency to 5G—AI presents challenges as well as new opportunities, especially as it transitions from Western settings to LMICs. Broadly speaking, these challenges fall into two categories. The first, touched on only briefly in this paper, are the deficiencies in technology capacity and policy making faced by LMICs. The second set of challenges—and the focus of this paper—have to do with deficiencies inherent to the "architecture" of AI systems and how they are developed.

Capacity- and Policy-Related Challenges

Notwithstanding the opportunities offered by AI, many countries' governments lack the capacity to fully capture relevant data. AI is data-hungry. In many countries, basic access to data remains a challenge, and policies, strategies, and regulations that enable the deployment of AI and data for the public good are lacking. Today, the world is more connected, interdependent, and data-rich than ever before, but we see a growing disparity between countries and populations able to benefit from data analytics in decision making and those left behind. This gap is largely due to the dearth of data related to the poorest and most marginalized people.[16] Poorer countries also may not have the technical capacity and resources to protect themselves against hacking and viruses, or to diagnose manipulations or bugs in systems. AI presents a particularly difficult challenge for traditional public policy making and regulation because it is so technically complex. Most policy makers simply do not understand how AI works. Further, the quality, safety, and precision standards needed for AI deployment tend to be set by the more developed countries. While LMICs' deficiencies in technical and policy-making capacity are extremely important, much has already been written about them. Appendix A lists organizations that have developed resources around these issues.

AI's Architectural Challenges

"Artificial Stupidity" and Algorithmic Bias

In the public imagination, AI can often appear to be beyond the influence of human foibles and misjudgment. However, AI systems are far from infallible. Even a well-designed algorithm must make decisions based on data which is in turn prone to the same flaws or errors we encounter in all spheres of life. And algorithms often make judgment errors when faced with unfamiliar scenarios.

One example of this so-called "**artificial stupidity**" comes from a recent test drive of Tesla's automated driving system, Autopilot, on a public highway.[17] The system registered dozens of traffic lights on its in-car display while it was driving 130 km/h down a road with no traffic lights. An investigation of the incident found that the car was driving behind a truck that was hauling inactive traffic lights, which Autopilot was unable to distinguish from in-service traffic lights. Amazon's voice assistant, Alexa, is also illustrative. Alexa, like all voice assistants, turns on when someone says a keyword. Unlike Google Voice and Apple Siri, which turn on when they hear "Hey Google" and "Hey Siri," Alexa is activated by a common human name: Alexa. Many Alexa users have reported the voice assistant turning on when a character on TV utters the name Alex, or when they talk to a family member who is named Alexa. In short, AI is frequently unable to accurately capture the user's intention.

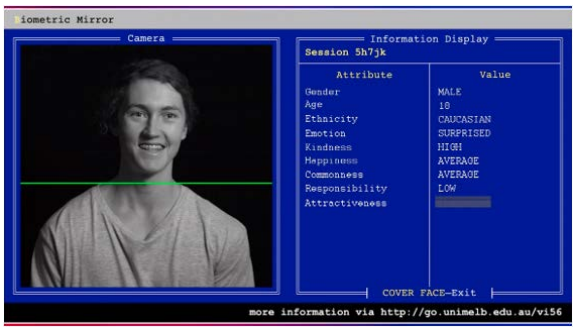
However, artificial stupidity can go beyond resembling human misjudgment to reproducing human **bias**. In 2016, for example, ProPublica analyzed a commercial system—created to help judges make better sentencing decisions—that predicts the likelihood that criminals will re-offend, and concluded that it was biased against people of color.[18] In a perfect world, using algorithms should lead to unbiased and fair decisions. But many algorithms have been found to incorporate biases because of the data on which they are trained. **Training data** can introduce bias into AI systems in two ways: 1) by **excluding** certain groups (such as women or people of color) from the training data to some degree, thus skewing the AI system toward those who are included, and 2) by **"baking in" human bias** into the training data, especially when that data is drawn from human judgments, which are prone to bias.

Creating AI systems also involves **decisions around design, structure, and function**, all of which incorporate the preconceptions of those decision makers. Designers and developers decide which datasets to use, which factors within a dataset to include, and what weight to assign each factor. All of these decisions are affected by the beliefs, norms, and biases of each designer or developer.

For example, a healthcare insurance algorithm used in the United States to identify patients who might need more attention and resources was found to be biased toward white and against Black patients.[19] The algorithm used historical health costs to predict and rank which patients would benefit the most from additional resources. While the developers of the tool did not include race as a variable to consider, the dataset itself was effectively a proxy for race and the racial inequities in U.S. society and U.S. healthcare. Black patients spend approximately $1,800 less in medical costs per year, in comparison to their white counterparts with similar health needs. This difference in spending led the algorithm to believe that Black patients were healthier than white patients when there were other external factors that could have led to these spending differences, such as a lack of trust in the health system[20] (thus lower use of healthcare resources) and higher rates of poverty

in Black communities.[21] The dataset on healthcare costs was assumed to be a proxy for health when, instead, it was a proxy for race. The use of the healthcare costs dataset was a decision made by the system's designers, individuals who we might reasonably infer have encountered relatively few economic and social barriers to accessing healthcare. Had this group of people been more diverse, other datasets or variables might have been considered in designing this tool and the resulting problem might have been avoided.

Another instance of the fallibility of AI is the field of facial recognition technology. Researchers found that three commonly used facial recognition tools were less accurate at identifying female faces and faces with darker skin than they were at identifying male faces and faces with lighter skin.[22] Another group of Australia-based researchers created the Biometric Mirror to demonstrate the ethical issues inherent to facial recognition technology. They asked human volunteers to judge thousands of faces on 14 characteristics, including age, race, and perceived emotions (such as aggressiveness or surprise). They then used this training data to build an AI tool that scans people's faces, analyzes those images, and from that analysis makes conclusions about the person's age, gender, and emotions. However, this analysis was found to often be unreliable, since the AI generates results based on the subjective information provided to it by the initial human volunteers.[23] The Biometric Mirror project highlights the flaws in AI-based facial recognition, especially its ability to accurately identify subtle human traits such as the emotions we display on our faces.



The Biometric Mirror uses an open dataset of thousands of facial images and crowd-sourced evaluations. Image: Sarah Fisher/University of Melbourne

AI systems reinforce what they have been taught from biased training data or biased human designers. These issues in AI architecture have a compounding effect. According to Lindsey Andersen:

> *Unfair AI systems have an unequal impact on different groups of people. This is especially disconcerting when results disproportionately reinforce existing patterns of group marginalization. These unfair systems are the consequence of bias. AI can be biased at the system level and the data level. The biased outputs generate negative feedback loop, where the AI system produces increasingly biased results over time.*[24]

For example, Amazon's recruitment algorithm taught itself to prefer male candidates. The system was trained with job application data collected over a 10-year period mostly from men.[25] The overrepresentation of male candidates in Amazon's hiring pool reflected the existing gender disparity in the technology sector, a disparity reinforced by an automated hiring and recruitment tool. While no one instructed the tool to treat CVs from men and CVs from women differently, it learned from gender disparity in the data that men were seemingly "preferred" over women.

Artificial Stupidity and Algorithmic Bias in LMIC Settings

Artificial stupidity and algorithmic bias are likely to have a pronounced effect when AI applications are introduced in LMIC settings, for several reasons. First, AI systems are largely built on data from Western settings, and the dearth of data in LMIC settings means that these systems are likely to lack the kinds of training data needed to ensure that they are tailored to the needs of LMICs. For example, in its 2018 Gender Gap study, the GSMA found that women in LMICs were 10 percent less likely than men to own a mobile phone, equating to 184 million fewer women than men owning mobile phones in these markets (GSMA 2018). Given that mobile

phone data is one of the few widely available data sources in LMICs, an AI system that uses this data as an input is producing outputs based disproportionately on the habits of men.[26] The effects of this lack of data from LMICs is also illustrated in a study by Facebook's AI lab, which found that object recognition algorithms from large U.S.-based technology companies performed worse at identifying household objects from low-income countries than high-income countries. The algorithms—from Amazon, Microsoft, Clarifai, Google, and IBM—were 15 to 20 percent better at recognizing household items such as soap and spices from the United States than the same items from Somalia and Burundi.[27]

Because many AI applications are developed in high-income countries, AI systems might also depend on algorithms with embedded biases. For example, Google's machine translation tool, Google Translate, is less accurate at translating languages such as Kinyarwanda and Yoruba—languages with fewer available datasets and documented resources—than it is at translating English and French. Many of the translation issues appear to occur in cases where there is a lack of context and the tool interprets each word on its own, rather than each word relative to the entire phrase or sentence.[28] While Google Translate problems may seem innocuous on their face, they have the potential to harm marginalized communities. For example, financial services providers have considered using automated chatbots to serve customers and increase efficiency. But many chatbots depend on automated translation and cannot communicate effectively in local languages.[29] Further, as financial organizations—including those in LMICs—begin to incorporate tools such as automated credit scoring, they risk replicating the bias-driven exclusions of women and minorities that have plagued similar efforts in the West, which could undermine efforts to broaden financial inclusion in LMICs.[30]

Expanding NLP resources

The natural language processing (NLP) sub-field of AI has historically focused on the world's most commonly spoken languages, which has meant that few NLP resources have been expended on the 2,000+ languages spoken across Africa. A grassroots African organization, Masakhane[31], has brought together interested individuals to create benchmarks and training datasets for more than 30 African languages. Their research has implications for international development practitioners interested in providing equitable access to information, for example, or preserving traditional cultures.

Because AI applications have historically focused on languages spoken in the Global North, the BMZ-funded FAIR Forward project pursues the goal of creating more openly available AI training data for African and Asian languages[32]. For example, in Rwanda, FAIR Forward has worked with Rwandan startup Umuganda and web browser developer Mozilla to gather more than 1,200 hours of voice recordings. The collection of open language data will strengthen the local ecosystem of businesses who can develop AI-based digital applications and products such as voice assistants for local consumers.

The "Black Box" Problem

A second ethical issue inherent to AI systems is the so-called "black box" problem. Not only is AI prone to error and bias, but the reasons for AI decisions are not easily accessed or understood by humans—making them difficult to question or probe. AI algorithms often make important decisions, from approving loans to determining diabetes risk, and the complexity of AI decision making means that people have little insight into *why* an AI system decides as it does. A study conducted by the AI Now Institute at NYU confirms that many AI systems are opaque to the citizens over whom they hold power.[33] This black box issue is becoming more salient as countries are increasingly using AI to make truly life-changing and far-reaching decisions—regarding criminal sentencing and enforcement, for example, or the delivery of social services.

AI bias and the black box problem are related. For example, if a woman is refused a bank loan by an AI system that is "biased" against female credit-seekers, the lack of transparency in AI architecture makes it difficult for her to counter this decision. Similarly, companies such as Goldman Sachs and Unilever have used technology developed by the startup HireVue to analyze job candidates' facial expressions and voice to advise hiring managers.34 Without knowing what data HireVue is using to train its facial and voice recognition systems, critics have voiced fears that using this tool will re-create social biases. They worry that training data favors "traditional" candidates (light-skinned men whose first language is English) and may score others—dark-skinned women, for example, people with disabilities, or people for whom English is a second language—less favorably. This problem is compounded by the fact that HireVue's algorithm is proprietary, so there is no way to know whether the critics are correct.[35]

In Western countries, regulators have already begun to enact regulations, known as algorithm accountability laws, that seek to curtail the use of automated decision systems by public agencies. For instance, in 2018 New York City enacted such a law,[36] which created a task force to recommend criteria for identifying automated decisions used by city agencies and a procedure for determining if those decisions disproportionately affect protected groups. However, the law does little to increase transparency—it only permits making technical information about the system publicly available "where appropriate" and states that there is no requirement to disclose any "proprietary information."[37] The end result: though New York City tried to regulate the use of AI, the law did not effectively increase transparency or accountability in the use of AI systems in automated decision making by city agencies.

This lack of transparency is exacerbated by the fact that private commercial developers generally refuse to make their code available for scrutiny because the software is considered intellectual property. While some experts have suggested making algorithms open to public scrutiny, many are not made public because of nondisclosure agreements with the companies that developed them. The EU GDPR requires companies to be able to explain how algorithms use the personal data of customers' work and make decisions—the "right to explanation." While this mandate is an important step forward, it does not provide guidance about the specific aspects of AI systems that companies are obliged to explain to consumers, which leaves much room for interpretation.

The Black Box Problem in LMIC Settings

The effects of the black box problem are accentuated in developing country settings. The same issue outlined above—difficulty understanding, accessing, and explaining the mechanisms behind AI—applies in LMICs, but the impacts are potentially greater. For example, in LMICs, where the gender gap for financial services is wider than in high-income countries,[38] the consequences of women being denied loans are more harmful. The same can be said about the gender gap in employment.[39] Women are more likely to be unemployed than men and less likely to participate in the labor market in LMICs. The impact of a biased, automated employment tool therefore is much greater. Additionally, the lower rates of digital and data literacy among populations in LMICs[40] make it more challenging for people affected by AI to understand it or even recognize if AI is being used.

Toward Ethical AI in International Development

As is the case with many emerging technologies, the international development community is faced with a conundrum: on the one hand, we recognize the immense potential that AI tools have in solving some of the more complex development challenges facing LMICs. Indeed, the community is already piloting these tools across various sectors. On the other hand, AI tools appear to have ethical challenges built into their foundations.

How can we take a balanced approach that moves us toward more ethical uses of AI in international development while still reaping its benefits? To answer this question, we outline a series of recommendations below. While the implementation of these recommendations will require collaboration across a variety of actors—including AI researchers, the technology industry, governments, and advocacy groups—these recommendations are largely focused on **future investment by bilateral and multilateral donors**. They are especially relevant for donors that have already invested in digital development, data literacy, digital government, and digital transformation of business. Already positioned to invest in addressing ethical AI, such donors have both the resources and incentives to get ahead of these problems, especially as AI systems are increasingly rolled out in LMICs.

Develop or Adapt an Ethical AI Framework Aligned to Country-Specific Perceptions of Ethics

AI poses new philosophical and ethical questions that experts, policy makers, and societies at large are only beginning to grapple with. In response to these questions, the AI research community in the United States and Europe has developed frameworks through which to examine ethical decision making on AI projects and minimize algorithmic harms. One example is the Fairness, Accountability, and Transparency (FAccT) framework.[41] Though the FAccT framework is currently the gold standard among ethical AI researchers in the Global North, it may not be applicable to the cultural contexts of Global South countries. Further, while existing toolkits provide actionable steps to minimize discrimination and harm, few address the underlying ethical implications of using AI on donor-funded international development projects.

To advance the discussion of ethical AI in international development and in the Global South, and move toward consensus on what ethical AI principles might look like in the digital development field, the international development community can begin by **researching context-specific definitions of ethics in LMICs.** FAccT has become the standard framework for ethical decision making regarding AI in the United States and Europe. However, FAccT works only when all parties clearly understand the definitions of the principles under discussion. Concepts of fairness in the United States can differ radically from concepts of fairness in India, for example, in the same way that concepts of online privacy and trust vary in different places and cultures.[42] We recommend that the international development community begin by conducting research to determine if the FAccT framework is partially or wholly applicable to LMICs and to understand how ethics are defined and what ethical precepts are followed in countries of interest.

Organizations and individuals from and in LMICs should be meaningfully incorporated into this research agenda. Individuals and organizations in the United States, Europe, and East Asia have to-date conducted most of the research into AI ethics—86 percent of papers presented at AI conferences come from these locations.[43] Findings from these studies are not necessarily generalizable to LMIC contexts. Funding should go to primary investigators from communities that use and are affected by AI.

This research and analysis will provide a foundation upon which experts can develop or adapt existing ethical AI frameworks, such as FAccT, to ensure that they are aligned to country-specific perceptions of ethics. The adaptation of such a framework requires dialogue and discussion among communities to determine what these principles mean in practice and in context. For example, in theory, "fairness" is a universal concept, focused on preventing systems from discriminating against people on the basis of their gender, religion, race, disability status, and so forth. However, definitions of what constitutes fairness are far from universal—they are dependent on the location, specific problem, environment, and cultural context.

Similarly, accountability in terms of AI refers to assigning responsibility for the outcomes of algorithms.[44] It is essential that there are established processes to remedy any adverse effects of—or bias in—algorithmic decision making. Accountability is not a new concept for the development sector; development projects have long sought to improve governments' accountability to their citizens and build mechanisms to hold people and organizations accountable to their mandates. Often, these mechanisms depend on factors such as the existing levels of trust between citizens and government, as well as existing accountability processes. New ethics frameworks will have to work with similar local dynamics and "on-the-ground" realities in order to increase the accountability of AI systems.

Increase Diversity in Data, Designers, and Decision Makers

Because key architectural issues in AI have their roots in a lack of diversity—a lack of diversity in the training data used to develop AI systems and in the backgrounds of those who design AI systems and decide when they're deployed—the international development community should invest to increase diversity in these areas.

Data: AI systems—often developed in Western contexts with Western-centric training data—need access to training data from the Global South. Without it, the AI tools used in the Global South serve only to reinforce the norms and biases of the society in which the data has been collected. The facial recognition tools that perform poorly on Black and female faces in the United States would perform terribly in most of Africa. Voice recognition systems that are used in the United States would likely have difficulty recognizing voices that speak English with a Nigerian or Indonesian accent and fail entirely on voices speaking any language other than English. AI developers therefore need to incorporate more diverse data into their work, and this imperative means going beyond merely collecting *more* data—it means working with communities to improve both the quantity *and quality* of data available to AI developers, and ensuring that developers know how to use this data appropriately.

Designers: Much like the AI research community, the community of AI designers and developers is also homogeneous, both in terms of identity group and technical background. Historically, AI designers have been predominantly white and male.[45] Improving diversity in terms of gender, race, and ethnicity would introduce new perspectives to the conversation. The facial recognition issues outlined above, for example, might have been rectified early on if the design team had included, say, more Black women. During testing, these analysts might have identified that performance was lacking across those racial and gender identities, given their lived experience.

Ethical AI will also require an interdisciplinary approach, requiring more involvement from more diverse data scientists, software developers, and statisticians, as well specialists in complementary fields such as history, law, and anthropology. Non-technical collaborators will also bring value to this field of study by contextualizing the historical and political factors that marginalize certain people.

Decision makers: While improvements can be made to AI systems and the processes by which they are developed, issues around the black box problem, for example, stem as much from their application as from their inherent architecture. To that end, we need more gatekeepers who can make informed decisions about when—and when not—to deploy AI, especially in situations where vulnerable populations such as women and ethnic minorities are likely to be treated unfairly or become victims of bias. Part of the solution will entail increasing the diversity of decision makers and ensuring that vulnerable groups are represented among those that decide when and where to deploy AI.

Develop Clear Metrics to Guide Ethical AI Implementation in LMICs

Ethical frameworks are an essential first step toward ethical AI, but the process does not end there. As Andrew Burt wrote in the *Harvard Business Review*, "[m]any AI ethical frameworks cannot be clearly implemented in practice… there's simply not much technical personnel can do to clearly uphold such high-level guidance."[46] Burt argues that every AI principle in an ethics framework should have clear metrics attached to guide implementation. The international development community should develop metrics to help AI designers and deployers answer questions such as:

- Is the training data used in an AI system adequately diverse? How can we measure diversity in specific countries?
- Does the system's training data accurately reflect the community it is meant to serve?
- Are there inequities in the community that are being reinforced by AI?
- Are adequate levels of transparency built into AI-driven decision making?
- Are mechanisms for redress and accountability adequate to address the harms that AI systems might create? Are these mechanisms actionable?
- Are design and/or decision-maker groups sufficiently diverse?

Cultivate Partnerships between Global South and Global North

Adapting ethical frameworks, increasing diversity, and developing clear metrics will demand increased partnership between developed and developing countries. Broadly, this impetus toward collaboration is aligned to SDG 17, Target 17.6 ("enhance North-South, South-South, and triangular regional and international cooperation on and access to science, technology, and innovation and enhance knowledge sharing on mutually agreed terms"). Building on AI partnerships such as The Partnership on AI[47] and Masakhane[48]—that have forged North-South and South-South relationships—will create a community and foster conversations that inform foundational research, data sharing, metrics, and technical assistance to governments and policy makers worldwide.

CONCLUSION

AI has been seen as a neat solution to many problems in LMICs, without adequate attention to its overall impact, effectiveness, and unintended consequences. While we have certainly seen successful uses of AI tools, their misuse and their pitfalls are also increasingly evident. International donors are well-positioned to build on existing investments in digital development by investing to ensure the protection of those who are entering the digital world. Donors are already investing in mitigating other digital harms, such as misinformation and disinformation, and digital threats to women and children. As AI begins to take hold in LMICs, international donors have an excellent opportunity and an important role to play in improving the design, development, and decision-making processes of AI tools. These investments will go some way toward maximizing AI's benefits and mitigating its risks, especially in the Global South.

ANNEX A:

Organizations and Resources for Further Reading on AI/ML

Resources

USAID, Reflecting the Past, Shaping the Future: Making AI Work for International Development, https://www.usaid.gov/sites/default/files/documents/15396/AI-ML-in-Development.pdf

USAID, Managing Machine Learning Projects in International Development: A Practical Guide, https://www.usaid.gov/sites/default/files/documents/Vital_Wave_USAID-AIML-FieldGuide_FINAL_VERSION_1.pdf

USAID, Exploring Fairness in Machine Learning for International Development, https://d-lab.mit.edu/sites/default/files/inline-files/Exploring_fairness_in_machine_learning_for_international_development_04012020_pages.pdf

FCDO, A guide to using artificial intelligence in the public sector, https://www.gov.uk/government/publications/understanding-artificial-intelligence/a-guide-to-using-artificial-intelligence-in-the-public-sector

IDIA, Artificial Intelligence in International Development, https://static1.squarespace.com/static/5b156e3bf2e6b10bb0788609/t/5e1f0a37e723f0468c1a77c8/1579092542334/AI+and+international+Development_FNL.pdf

Organizations

- **The Alan Turing Institute:** The public policy program of the Alan Turing Institute works with policy makers to solve policy problems and develop ethical foundations for data science and policy making.[49] This program distributes policy papers on the intersection of data science and AI and policy.

- **Center for AI and Digital Policy:** The Center for AI and Digital Policy is a nonprofit education organization that advocates to ensure that AI and digital policies are used to promote social inclusion and a more equitable society.

- **Center for Security and Emerging Technology (CSET)50:** CSET is a policy research organization housed in Georgetown University's Walsh School of Foreign Service. The organization does research on the policy tools that can be used to guide AI development and use, focusing on topics such as hardware, standards, and cybersecurity.

- **Information Technology Industry Council (ITI):** ITI is a policy and advocacy organization whose work spans the wider technology sector, including AI. Its publication, *ITI"s Global AI Policy Recommendations*,[51] dives into policy recommendations such as government responsibilities in regulating AI, investment into research and development, and dedication to improving the AI workforce.

- **OECD AI Policy Observatory52:** A repository of resources from across the OECD and its partners. It includes policy publications on a wide range of thematic areas, including agriculture, health, and trade. The Observatory was launched in February 2020 to help implement the OECD Principles on Artificial Intelligence[53], one of the most widely known and cited set of international AI standards.

ENDNOTES

1 Anderson, Lindsey. 2019. "Artificial Intelligence in International Development: Avoiding Ethical Pitfalls," *Journal of Public International Affairs*, https://jpia.princeton.edu/news/artificial-intelligence-international-development-avoiding-ethical-pitfalls.

2 Aerobotics, https://www.aerobotics.com/.

3 Vilaça, Nuno. 2016. "Financing Smallholder Farms with FarmDrive," *Planet*, https://www.planet.com/pulse/financing-smallholder-farms-with-farmdrive/.

4 O'Shea, Tara. 2018. "New Project From World Bank & Planet Uses Satellite Imagery & Machine Learning to Drive Sustainable Urbanization in Tanzania," *Planet*, https://www.planet.com/pulse/world-bank-planet-sustainable-urbanization-tanzania/.

5 Geekie, https://www.geekie.com.br/.

6 Kudi.ai, https://kudi.co/.

7 Timm, Stephen. 2017. "6 Artificial Intelligence Startups in Africa to Look Out For," *Clevva*, https://clevva.com/press-release/6-artificial-intelligence-startups-africa-look/.

8 Gul, Ehsan. 2019. "Is Artificial Intelligence the Frontier Solution to Global South's Wicked Development Challenges?," *Towards Data Science*, https://towardsdatascience.com/is-artificial-intelligence-the-frontier-solution-to-global-souths-wicked-development-challenges-4206221a3c78.

9 n.a. 2021. "Wildlife Crime Technology Project," *World Wildlife Project*, https://www.worldwildlife.org/projects/wildlife-crime-technology-project.

10 Pulse Lab Kampala. 2017. "Using Machine Learning to Accelerate Sustainable Development Solutions in Uganda," *UN Global Pulse*, https://www.unglobalpulse.org/news/using-machine-learning-accelerate-sustainable-development-solutions-uganda-0.

11 Deparday, Vivien et al. 2018. "Machine Learning for Disaster Risk Management," *GFDRR*, https://www.gfdrr.org/en/publication/machine-learning-disaster-risk-management.

12 DeRiggi, John. 2016. "Machine Learning Will Help Development Projects Achieve Scale," *Digital@DAI*, https://dai-global-digital.com/machine-learning-will-help-development-projects-achieve-scale.html.

13 SkillLab, https://skilllab.io/en-us/solutions.

14 PwC. 2018. "The Macroeconomic Impact of Artificial Intelligence," *PwC*, https://www.pwc.co.uk/economic-services/assets/macroeconomic-impact-of-ai-technical-report-feb-18.pdf; PwC. 2017. "Sizing the Prize: What's the Real Value of AI for your Business and How Can you Capitalise?," *PwC*, https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html.

15 Bamberger, Michael, and York, Peter. 2020. "Measuring Results and Impact in the Age of Big Data: The nexus of evaluation, analytics, and digital technology," *The Rockefeller Foundation*, https://www.rockefellerfoundation.org/report/measuring-results-and-impact-in-the-age-of-big-data-the-nexus-of-evaluation-analytics-and-digital-technology/.

16 Data for Sustainable Development, *United Nations*, https://www.un.org/en/global-issues/big-data-for-sustainable-development.

17 Robitzski, Dan. 2021. "Watch Tesla Autopilot get Bamboozled by a Truck Hauling Traffic Lights," *The Byte*, https://futurism.com/the-byte/tesla-autopilot-bamboozled-truck-traffic-lights.

18 Angwin, Julia et al. 2016. "Machine Bias," *ProPublica*, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

19 Obermeyer, Ziad et al. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science*, 366, 6464: 447-453.

20 Roberts Kennedy, Bernice, Clomus Mathis, Christopher, and Woods, Angela K. 2007. "African Americans and Their Distrust of the Health Care System: Healthcare for Diverse Populations," *J. Cult Divers*, 14, 2: 56-60.

21 Jargowsky, Paul. 2015. "New Data Reveals Huge Increases in Concentrated Poverty Since 2000," *The Century Foundation*, https://tcf.org/content/commentary/new-data-reveals-huge-increases-in-concentrated-poverty-since-2000.

22 Buolamwini, Joy and Gebru, Timnit. 2018. "Gender Shades: Intersectional Accuracy Disparityies in Commercial Gender Classification," *Proceedings of Machine Learning Research*, 81: 1-15.

23 Houser, Kristin. 2018. "The 'Biometric Mirror' Judges You the Way We've Taught it To: With Bias," *The Byte*, https://futurism.com/the-byte/biased-ai-biometric-mirror.

24 Anderson, Lindsey. 2019. "Artificial Intelligence in International Development: Avoiding Ethical Pitfalls," *Journal of Public International Affairs*. https://jpia.princeton.edu/news/artificial-intelligence-international-development-avoiding-ethical-pitfalls.

25 Dastin, Jeffrey. 2018. “Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women,” *Reuters*, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

26 Connected Women. 2018. “Connected Women: The Mobile Gender Gap Report 2018,” *GSMA*, https://www.gsma.com/mobilefordevelopment/resources/the-mobile-gender-gap-report-2018/.

27 De Vries, Terrance et al. 2019. “Does Object Recognition Work for Everyone?,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. 52-59.

28 Adeola, Aremu. Jr. 2020. “Lost in translation: Why Google Translate often gets Yorùbá — and other languages — wrong,” *Global Voices*, https://globalvoices.org/2020/03/03/lost-in-translation-why-google-translate-often-gets-yoruba-and-other-languages-wrong/.

29 n.a. n.d. “Banking for all: Can AI improve financial inclusion?,” *AI for Good*, https://aiforgood.itu.int/banking-for-all-can-ai-improve-financial-inclusion/.

30 Meka, Sushmita and Grasser, Matt. 2019. “Artificial Intelligence: Practical Superpowers – The Case for AI in Financial Services in Africa,” *BFA Global*, https://bfaglobal.com/wp-content/uploads/2019/03/FIBR-Artificial_Intelligence_FINAL_MAY2018-1.pdf.

31 https://www.masakhane.io/home

32 Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages, https://aclanthology.org/2020.findings-emnlp.195/

33 Simonite, Tom. 2017. “AI Experts Want to End ‘Black Box’ Algorithms in Government,” *Wired*, https://www.wired.com/story/ai-experts-want-to-end-black-box-algorithms-in-government/.

34 Chandler, Simon. 2017. “The AI Chatbot Will Hire You Now,” *Wired*, https://www.wired.com/story/the-ai-chatbot-will-hire-you-now/.

35 Chen, Angela. 2019. “The AI hiring industry is under scrutiny—but it’ll be hard to fix,” *MIT Technology Review*, https://www.technologyreview.com/f/614694/hirevue-ai-automated-hiring-discrimination-ftc-epic-bias/.

36 A Local Law in relation to automated decision systems used by agencies, https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0&Options=&Search.

37 Kelly, Ben and Chae, Yoon. 2019. “INSIGHT: AI Regulations Aim at Eliminating Bias,” *Bloomberg Law*, https://news.bloomberglaw.com/tech-and-telecom-law/insight-ai-regulations-aim-at-eliminating-bias.

38 n.a. 2018. “Reducing the Gender Gap in Financial Inclusion,” *Center for Inclusive Growth*, https://www.mastercardcenter.org/insights/reducing-gender-gap-financial-inclusion/.

39 n.a. 2018. “Facts and Figures: Economic Empowerment,” *UN Women*, https://www.unwomen.org/en/what-we-do/economic-empowerment/facts-and-figures.

40 Kumar, Ravi. 2021. “Data Use and Literacy Program,” *The World Bank*, https://www.worldbank.org/en/programs/data-use-and-literacy-program/overview.

41 ACM FAccT Conference, https://facctconference.org/.

42 Roggemann, Kristen, Nurko, Galia. and Tyers-Chowdhury, Alexandra. 2020. “User Perceptions of Trust and Privacy on the Internet,” *DAI Global*, https://www.dai.com/fi-cyber-user-trust.pdf.

43 Perrault, Raymond et al. 2019. “Artificial Intelligence Index 2019 Annual Report,” *HAI Stanford*, https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf.

44 Caplan, Robyn et al. 2019. “Algorithmic Accountability: A Primer,” *Data & Society*, https://datasociety.net/wp-content/uploads/2019/09/DandS_Algorithmic_Accountability.pdf.

45 Carson, Biz and Gould, Skye. 2017. “Uber’s diversity numbers aren’t great, but they’re not the worst either – here’s how they stack up to other tech giants,” *Business Insider*, https://www.businessinsider.com/uber-diversity-report-comparison-google-apple-facebook-microsoft-twitter-2017-3.

46 Burt, Andrew. 2020. “Ethical Frameworks for AI Aren’t Enough,” *Harvard Business Review*, https://hbr.org/2020/11/ethical-frameworks-for-ai-arent-enough.

47 The Partnership on AI, https://www.partnershiponai.org/.

48 Masakhane, https://www.masakhane.io/home.

49 The Alan Turing Institute, https://www.turing.ac.uk/research/research-programmes/public-policy.

50 CSET, https://cset.georgetown.edu.

51 ITI. 2021. “ITI’s Global AI Policy Recommendations,” *ITIC*, https://www.itic.org/documents/artificial-intelligence/ITI_GlobalAIPrinciples_032321_v3.pdf.

52 OECD.AI, https://www.oecd.ai/.

53 n.a. 2021. “What are the OECD Principles on AI?,” *OECD*, https://www.oecd.org/going-digital/ai/principles/.

Acknowledgments

The author wishes to thank the following people for their contributions and support: Greg Maly, Galia Nurko, Steven O'Connor, Inta Plostins, Rob Ryan-Silva, Araba Sapara-Grant, Samantha Weinberg, and the team at DAI's Center for Digital Acceleration.

SHAPING A MORE LIVABLE WORLD.

www.dai.com

@daiglobal