ARTIFICIAL INTELLIGENCE AND DEVELOPMENT

Matthew L. Smith, IDRC (Canada)
Ruhiya Kristine Seward, IDRC (Canada/Jordan)

ARTIFICIAL INTELLIGENCE FOR DEVELOPMENT AFRICA

Date of Publication: July 2022
Published by: AI4D Africa

The research presented in this publication was carried out with the financial assistance of Canada's International Development Research Centre. The views expressed herein do not necessarily represent those of IDRC or its Board of Governors.

Authors & Contributors

Matthew L. Smith, International Development Research Centre
Ruhiya K. Seward, International Development Research Centre

"AI and Development", (Forthcoming) in M. Clarke and X. Zhao (eds) *Elgar Encyclopedia of Development*, Edward Elgar Publishers, Cheltenham, UK.

Introduction

Artificial intelligence appears in our news feeds nearly every day, accompanied by a multiplicity of narratives and expectations that are generally hyperbolic - either excited or fearful - and rarely nuanced. With this in mind, it is critical to unpack the reality of what artificial intelligence is and is not, and the implications for our collective future. Artificial intelligence is "the science and engineering of making intelligent machines" where "intelligence is the computational part of the ability to achieve goals in the world" according to John McCarthy, who ran the first-ever gathering on AI in 1956 (McCarthy 2007). Today, artificial intelligence is the science and engineering of computer systems where "intelligence" means having the ability to perform tasks such as visual perception, speech recognition, language translations, and certain types of decision-making.

Since that first AI gathering in 1956, significant advances in AI have been made, although more slowly than early pioneers predicted. It was only in the 2010s that advances in machine learning — a particular approach to AI — started to have real-world impact. Specifically, the introduction of deep learning (LeCun, Bengio, and Hinton 2015), enabled by increasing computational power and data availability, is propelling advances in AI.

In the 2020s, machine learning algorithms are now at the core of many prevalent technologies. They power search engine results, personalize news feeds, enable chatbot conversations, compose music, make medical diagnoses, produce efficient engineering designs, enable real-time facial recognition and surveillance, and inform life-altering decisions about who is eligible for a job interview, a bank loan, and even parole.

The progress made in AI over 60 years becomes clear when we compare an early AI program called Eliza to the 2020 release of the GPT-3 (Generative Pre-Training Transformer) (Brown et al. 2020). Eliza, programmed in the 1960s, was the first chatbot to use early natural language processing to simulate a psychotherapist. The relatively simple system worked by creating rules for Eliza to flip patients' statements around into questions, like this:

Patient: I am feeling stressed out.

Eliza: Do you believe it is normal to be feeling stressed out?

GPT-3, by contrast, does not use a predetermined set of rules. Instead, it is a deep learning language model trained on massive datasets of hundreds of billions of words scraped from Wikipedia and web crawlers. GPT-3 learns the relationships within datasets and uses this learning to generate new outputs from new inputs.

The chasm between Eliza and GPT-3 demonstrates the key shift that has taken place in AI development: We have moved from approaches based on explicit rules to those based on machine learning. This is a critical leap for real-world applications because, in many contexts, it is impossible to explicitly spell out the rules to be followed. For example, how could one possibly write rules to account for all the permutations of visually recognizing a person or navigating a car through a city? AI is not *taught* how to do these things with a set of rules; it learns from experience codified in data.

Despite the incredible power of machine learning, AI models are still narrow in the sense that they can be applied only to the task for which they are trained. They "break" when applied to another task. An AI algorithm trained to play chess cannot play Go. The GPT-3 language model can play neither chess nor Go. Furthermore, these models do not understand in a meaningful sense that they are playing chess. They do not understand the meaning of a sentence as part of a chatbot conversation. The model simply calculates the next move or utterance and chooses based on the move that has the highest probability of being correct.

In short, these algorithms lack a general, humanlike, adaptive intelligence that would enable them to learn and apply learning across domains, situations, and problems. While researchers are actively exploring and developing AI capable of adaptive, general intelligence, such advances fall out of the scope of relevance for AI in development at the time of writing.

AI in Lower- and Middle-Income Countries

The development and deployment of AI applications builds on the ongoing diffusion and adoption of information and communication technologies (ICTs) — from telephones to internet-enabled mobiles to the internet of things. Over the last 40 years, the spread of ICTs has given rise to broad development benefits from digital technologies, improving service delivery, increasing opportunities online and offline, and contributing to economic growth (World Bank 2016). Flowing from this digital infrastructure is the increasing prevalence of data production and collection, which creates new mechanisms for increasing transparency and accountability, making better policy, improving service delivery, and increasing business opportunities (World Bank 2021). The broad application of AI straddles this trend and therefore should only deepen with continued processes of digitization, digitalization, and datafication.

AI leverages digital and data infrastructure in a way that enables it to potentially foster transformative changes at a large scale. AI can turn data into actionable intelligence through data-informed predictions that can inform or lead to concrete decisions. In doing so, AI can augment, or fill in for non-existent expertise. The fact that AI does this while piggybacking on existing digital infrastructure means that provided the data exist, these predictions and decisions can scale easily at minimal cost. Examples abound:

- Automated computer translation of text between languages (machine translation) can contribute to more inclusive governance by making services and information available in a wider array of local languages.
- Optical character recognition techniques can digitize paper-based legal case records, and natural language processing techniques can help lawyers and judges rapidly analyze them.
- Sensor technology can be attached to livestock to monitor vital information, and AI techniques can analyse this data for early detection of diseases or other disorders. This allows for timely medical intervention and contributes to improved livestock health and productivity.

Before AI, many of these activities would either not have been possible or would have been prohibitively time consuming.

While training an AI model may be costly, once trained, running the model typically does not require significant computational power. This means models can be run locally on handheld devices, often without internet connections, which in turn greatly expands their applicability beyond contexts of high connectivity. For example, a farmer can diagnose crop diseases using just a cellphone with a camera. Similarly, rural health workers can diagnose malaria by conducting a microscopic analysis of blood samples using a low-cost mobile device.

Given the continued rapid spread of technology and broad applicability of AI, it is not surprising that AI applications could arguably contribute to achieving 134 of the 169 Sustainable Development Goal targets (Vinuesa et al. 2020).

Despite this potential applicability, practitioners, academics, and policymakers promoting the use of AI to tackle local challenges and improve people's lives face considerable constraints in lower- and lower-middle income countries (LLMICs). These constraints not only limit the impact of AI, but also increase some of the associated risks. While the following is not an exhaustive list, here are three key constraints:

Governance — Like ICTs and data, AI requires sound governance to encourage its use and limit its potential downsides. Currently, many LLMIC governments are limited in their readiness to leverage AI for genuine development aims — although this can also be said for middle- and higher-income countries. This reality is not surprising, as many countries lack the foundations upon which to successfully govern AI. They lack, for example, the institutional capacity to safeguard the rights of citizens (offline or online), and even rudimentary data protection frameworks.

Infrastructure — Despite the rapid spread of ICTs globally, considerable inequities remain in access to digital infrastructure across the world. However, an increasing number of organizations (academic, public, and private sector), typically in large cities, have sufficient computational capacity and access to work with AI tools. Furthermore, access to cloud computing AI services somewhat mitigates the need for local computational capacity (although costs to train can be high). Despite all this, the uneven spread of digital infrastructure has resulted in a lack of contextually relevant, high-quality, labelled data sets for applying AI at a local level. This is amplified by the historical fact that the development of AI (and related datasets) has occurred almost entirely in higher-income countries.

Active lines of research for technical solutions are working to address these infrastructure challenges, which could greatly improve the applicability of AI in LLMIC contexts. For example, transfer learning approaches (i.e. working with partially pretrained models) require significantly less data to achieve high levels of performance.

Human Resources — Many LLMIC countries (such as Kenya, South Africa, and Senegal) already have hubs of AI capacity and well-organized and mobilized AI communities.1 However, gaps remain in the expertise required to advance AI for human development. One considerable gap is in female practitioners — an issue the world over — who are critical for, among other things, bringing female voices and diverse perspectives into the

[1] For example, African researchers and practitioners established Deep Learning Indaba and Data Science Africa to strengthen the machine learning and data science communities. Similarly, Masakahne is a grassroots organization focused on strengthening and spurring natural language processing research on African Languages.

development of applications, and addressing biases in AI development and deployment. Another key gap is a paucity of cross-disciplinary collaborations needed to successfully bridge a development challenge and a relevant, viable, innovative solution.

AI Controversies and Responses

The excitement for AI's potential in development may be equally matched by concerns and controversies. As with all technologies, AI applications can and will have both positive and negative social effects. These concerns have rapidly emerged in the public consciousness, in part due to high-profile incidents of highly biased software, such as when Google photo recognition software produced racist results. These incidents have helped motivate a series of responses by AI companies, and non-governmental, inter-governmental, research, and governmental organizations around the world.

Controversies

AI controversies are rooted in both technical issues and the irresponsible or unethical application of these technologies that may lead to social ills (Bender et al. 2021). Here are a few key concerns:

Bias — Perhaps the most widely discussed controversy stems from AI algorithms that have encoded biases (O'Neil 2016; Smith and Rushtagi 2021). Bias can come from a variety of sources. If an algorithm is trained on a non-representative dataset, the efficacy of an algorithm trained on that data will differ across populations, such as when an algorithm developed to detect melanoma doesn't work well for darker skinned patients. Algorithmic bias can also result from training on data that reflects existing social inequities and biases, such as an AI model developed to predict crime that is trained on data representing past biased policing practices, or a language model trained on language corpuses with sexism biases. Such algorithms effectively learn, and then offer predictions informed by the social biases that drove the social inequities in the first place.

Another source of bias can come from a lack of diversity among those who are designing AI systems. Given the overwhelming prevalence of males working in the computer engineering and AI fields, it is no surprise that biases emerge in the AI applications they create, such as when early voice assistant technologies used almost entirely female voices.

Opaqueness — The GPT-3 encodes its input data in around 175 billion parameters. This non-symbolic representation of the data accounts for both how it can "learn" so much and why it is an opaque black box. Given an input, we can't explain *why* it provides a particular output. In other words, these models lack "explainability" — humans' ability to look inside and discern a clear and understandable explanation of how it arrived at its prediction. While GPT-3 is at the time of writing one of (if not the) largest AI model in the world, the basic mechanism of numerically encoding rules across many nodes remains the same with smaller models. This lack of transparency is arguably a huge challenge both for developing mechanisms of accountability and for building trust in AI systems.

AI snake oil — AI's strong ability to predict outcomes has spurred the development of AI applications that inform potentially life-altering decisions such as who can get parole, a loan, or a job interview. These types of systems are problematic because:

- They are trained on historical data that encode structural biases (socioeconomic, racial, ethnic, gender, among others).
- They involve social predictions for which current AI models have not yet demonstrated sufficiently high levels of accuracy.
- They are high risk, meaning they potentially can have high impact on an individual's life or a community's well-being.

At best, these models don't work very well; at worst, they can cause harm to individuals and communities. These applications are the AI version of "snake oil" (Narayanan n.d.). In a similar vein, well-intentioned but quickly assembled AI models, such as the numerous AI responses to the COVID-19 pandemic, can lead to disappointment, lost resources, and even negative outcomes (The Alan Turing Institute 2021). AI is not needed for many problems and, in most contexts, there are simpler solutions.

Surveillance systems — AI is supercharging our ability to track and profile individuals and communities. For example, police departments across the world are buying and using real-time facial recognition as part of daily policing activities to reduce crime, even though research is demonstrating how such systems can be both biased (resulting in a greater number of errors for darker skinned people, for example) and easily misused. Predatory lenders are scooping up and buying vast amount of personalized data to profile and target poorer populations (O'Neil 2016). Together, the uses of AI for surveillance by the public and private sectors are potentially undermining human and democratic rights, presenting enormous and durable governance challenges for policymakers and innovators alike.

Energy consumption and carbon footprint — Some large AI models can be extremely energy intensive to train. For example, the training of the largest GPT-3 model took several thousand petaflop2 days of compute, with a carbon footprint comparable to a new car driving 703,808 kilometers (Anthony, Kanding, and Selvan 2020).

This list of concerns is not comprehensive, but some of the key issues include:

- AI-based automation can create significant unemployment or underemployment.
- The consolidation of power in a few large corporations is more likely due to the centrality of data and the virtuous cycle of more data leading to better applications.
- The public sphere can become diminished and polarised as a result of filter bubbles and echo-chambers generated from personalized social media and news feeds.

These controversies surrounding AI contribute to the concern that AI applied in LLMICs could work to inhibit the achievement of SDG targets (Vinuesa et al. 2020), and even

2 A petaflop is equal to 1015 floating-point operations per second.

reverse development by contributing to increased inequalities and political instability (Smith and Neupane 2018).

Responses

The well-documented and increasingly well-researched risks of AI have motivated a broad range of responses that seek to balance benefits and potential harms. Below, we discuss some prominent responses to the controversies spelled out above.

Frameworks for ethical, responsible, or trustworthy AI — Since 2015, there has been an explosion of frameworks and principles to guide AI development and deployment in ways that are ethical, that preserve human rights, and that are environmentally sustainable. These frameworks have different names but in general converge on the following five principles: transparency, justice and fairness, non-maleficence, responsibility, and privacy (Jobin, Ienca, and Vayena 2019).

These frameworks, however, have limitations. First, they may have limited applicability, as they have been developed almost entirely within and by people living in "northern" contexts — and therefore may not be appropriate outside these contexts. Second, these high-level principles do not provide guidance on how to put the principles into practice in different contexts. This leaves a significant gap between aspiration and implementation.

Technical tools — AI developers, and particularly those in large multinational companies, have started to develop technical tools to assist innovators and practitioners as they develop AI more responsibly. For example, there are tools for detecting and mitigating bias in training data sets, and tools that report the energy consumption associated with training AI models. There is also an active subset of AI research focused on developing explanations for the operation of an AI application.

National and regional strategies, policies, and regulations — While the pace of technological innovation in AI since the early 2010s has been rapid, governments around the globe are still in the early stages of understanding how to deal with the possibilities and controversies raised by AI. An early response by some governments has been to develop national strategies that focus on a wide variety of activities, including research and development, human capacity development, infrastructure, and the provision of an ethical and legal framework.

Governments are also beginning to explore targeted policies and regulations. For example, governments at different levels of jurisdiction have begun to place moratoriums or outright bans on facial recognition technology. As of 2021, perhaps the most comprehensive approach has been the EU's proposed AI regulation.3 This regulation prohibits, among other things, AI systems that may cause physical or psychological harms and facial recognition in public spaces or by law enforcement. This regulation also seeks to govern "high-risk" AI systems that may undermine fundamental human rights, such as when AI is applied in the administration of justice or to determine eligibility for public benefits, credit scoring, and employment. This includes requiring that these high-risk systems be understandable by users and allow user oversight.

[3] Find details about the legislation.

International collaborations — The transboundary nature of digital infrastructure and data flows means that questions of AI governance require regional and international cooperation and collaboration. One global response is the launch of the Global Partnership on AI (GPAI) in 2020, co-founded by 14 governments and the EU. GPAI is a "multi-stakeholder initiative to bridge the gap between theory and practice on AI by supporting cutting-edge research and applied activities on AI-related priorities."[4] Among key early priorities are the issues of responsible AI and data governance.

Conclusion

AI in development is not just a passing fad. By the time you read this, some key information in this entry on the state of AI will be partially or even embarrassingly out of date. However, the broad applicability of AI and the rapid growth of its underlying infrastructure means that its importance and relevance for AI techniques in development will have increased. While countries around the world are taking notice of the importance of AI in shaping economies and societies, as of 2021, we are still in the early stages of the larger transformation.

The applicability of AI in numerous domains makes it a potentially transformative technology in service of human development. But limited resources exist, and it can be difficult to justify spending resources on AI applications where more fundamental development challenges persist (poverty, health, education, climate change, institutional capacity). AI techniques can, however, be applied in all these domains in ways that cheaply substitute or complement limited inputs (e.g. expertise) and can, at least technically, be scaled relatively cost effectively.

Harnessing AI for human development requires addressing numerous constraints to ensure that governments and civil societies in LLMICs (and globally) have the autonomy, resources, and know-how to govern and responsibly apply AI to achieve local development goals. A reliance on imported AI strategies, regulations, and solutions greatly increases the chances of contextually inappropriate applications that run counter to local value systems — and deteriorate trust as a result. Among other things, local policy and innovation research are needed if we are to better understand the social, political, and economic implications of AI. A deepened understanding and expertise can then inform local implementations and policy, as well as — perhaps equally critically — regional and international discussions, collaborations, and agreements.

[4] https://gpai.ai/

References

Anthony, Lasse F. Wolff, Benjamin Kanding, and Raghavendra Selvan. 2020. “Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models.” *ArXiv:2007.03051 [Cs, Eess, Stat]*.

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” Pp. 610–23 in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. Virtual Event Canada: ACM.

Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. “Language Models Are Few-Shot Learners.” *ArXiv:2005.14165 [Cs]*.

Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. “Artificial Intelligence: The Global Landscape of Ethics Guidelines.” *Nature Machine Intelligence*. doi: 10.1038/s42256-019-0088-2.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. “Deep Learning.” *Nature* 521(7553):436–44. doi: 10.1038/nature14539.

McCarthy, John. 2007. *What Is Artificial Intelligence?* Stanford University.

Narayanan, Arvind. n.d.“How to Recognize AI Snake Oil.”

O’Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishers.

Smith, Genevieve, and Ishita Rushtagi. 2021. “When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity.” *Stanford Social Innovation Review*. Stanford University.

Smith, Matthew L., and Sujaya Neupane. 2018. *Artificial Intelligence and Human Development: Toward a Research Agenda*. Ottawa, Canada: International Development Research Centre.

The Alan Turing Institute. 2021. *Data Science and AI in the Age of COVID-19*.

Vinuesa, Ricardo, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia Dignum, Sami Domisch, Anna Felländer, Simone Daniela Langhans, Max Tegmark, and Francesco Fuso Nerini. 2020. “The Role of Artificial Intelligence in Achieving the Sustainable Development Goals.” *Nature Communications* 11(1):233. doi: 10.1038/s41467-019-14108-y.

World Bank. 2016. *World Development Report 2016: Digital Dividends*. Washington: World Bank.

World Bank (Washington, District of Columbia), ed. 2021. *World Development Report 2021: Data for Better Lives*. Washington: World Bank.