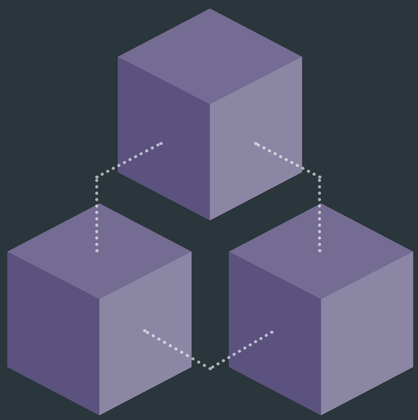


Artificial intelligence and human development

Toward a research agenda

WHITE PAPER

IDRC | CRDI
International Development Research Centre
Centre de recherches pour le développement international

Canada

Researchers and authors
Matthew L. Smith
msmith@idrc.ca
Senior Program Specialist, International Development Research Centre

Sujaya Neupane
Postdoctoral fellow, McGovern Institute for Brain Research,
Department of Brain and Cognitive Sciences, MIT

Editing and research support
Greg Leonard

Design
Claudio Mendonca
ccmdesign.ca

The authors would like to thank Urs Gasser, Amar Ashar, and their colleagues at the Berkman Klein Centre for Internet and Society at Harvard University, Laurent Elder, Maroussia Levesque, Marshall Smith, Pascal Kropf, Sunisha Neupane, Aaron Martin, Raul Zambrano, and Cheryl Chan for their invaluable feedback and insights.

April 2018

The research presented in this publication was partially funded by the UK Department for International Development (DFID). The views expressed herein do not necessarily reflect those of the DFID, IDRC, or IDRC's Board of Governors.

Contents

ABBREVIATIONS

AI	Artificial intelligence
ITS	Intelligent tutoring systems
LMICs	Low- and middle-income countries
ML	Machine Learning
MOOCs	Massive open online courses
RL	Reinforcement learning
SDGs	Sustainable Development Goals

FIGURES AND TABLES

Artificial intelligence and human development

Executive summary

Artificial intelligence (AI) applications will profoundly impact societies in low- and middle-income countries (LMICs), both positively and negatively. AI is an emerging class of technologies upon which other technologies and applications are being built. Fuelled by the increasing availability of computational power, improved connectivity, and data, AI applications offer intriguing possibilities for promoting economic growth and tackling a wide variety of longstanding problems in the Global South. The resulting disruptive impact will be massive and could well be revolutionary.

The introduction of AI applications in the Global South brings tremendous potential for both good and harm. AI makes possible innovative, data-driven, technical innovations to help address pressing social problems. AI can facilitate scientific breakthroughs, improve medical diagnoses, increase agricultural productivity, optimize supply chains, and equalize access to education through highly personalized learning. However, like most new technologies, AI also has the potential to exacerbate existing problems and create new ones. Contributing to social ills, AI might reinforce structural inequalities and bias, perpetuate gender imbalances, threaten jobs, and introduce other unknown risks and unintended consequences.

Consequently, the roll-out of future AI applications requires a healthy critical perspective and an ongoing public dialogue. Concerted efforts are warranted to ensure the equitable and inclusive development and deployment of AI in LMICs. These efforts should focus on enabling these countries to take advantage of AI so they can benefit from the immense value the technology can bring. At the

Photo by Annie Spratt on Unsplash

same time, care and vigilance will be required to mitigate risks and to identify and respond to unintended harmful consequences. Such an approach is critical so as not to exacerbate inequalities and social instability.

This paper proposes a proactive research agenda for the ethical and equitable application of AI in the Global South. Undergirding this agenda is a broad overview of the technologies associated with AI and the opportunities and challenges they present.

WHAT IS AI?

AI is an area of computer science devoted to developing systems that can be taught or learn to make decisions and predictions within specific contexts. AI applications can perform a wide range of *intelligent* behaviours: optimization (e.g., supply chains); pattern recognition and detection (e.g., facial recognition); prediction and hypothesis testing (e.g., predicting disease outbreaks); natural language processing; and machine translation.

AI technologies are poised to have a significant impact on society because they leverage existing infrastructure (the internet, large datasets) to dramatically reduce the costs of activities (both new and old, good and bad) on a large scale.

POTENTIAL BENEFITS OF AI IN DEVELOPING COUNTRIES

Health care: AI can play a crucial role in augmenting health care capacity by filling gaps in human expertise, improving productivity, and enhancing disease surveillance. For example, an NGO in Brazil has partnered with an AI start-up to develop a system to predict upcoming incidences of disease.

Delivery of government services and information: Groups around the world are exploring ways to use AI to help countries improve their e-government efforts by automating complex assessments that take account of a range of technical, organizational, and social factors. For example, a machine learning system has been developed to help predict mass grave locations of Mexican drug cartel victims.

Agriculture: AI is being employed to address the various threats that can compromise a successful harvest. For example, AI systems are being used to support water management in Palestine and drought monitoring in Uganda.

Education: AI can move educational offerings beyond an industrial, one-size-fits-all delivery model toward quality personalized learning opportunities at scale. For example, efforts in India are employing AI to develop intelligent tutoring systems.

Economy and business: AI offers the potential for higher productivity and offers a means of growth in the form of new business development, innovation, and optimization of economic building blocks. For example, several companies are working to extend access to standard financial services to the hundreds of millions of Africans who either do not use them or do not currently have access.

POTENTIAL RISKS

Fairness, bias and accountability: AI systems have the potential to reflect and exacerbate societal bias and produce results that can disadvantage individuals and groups, especially those already marginalized. For example, a computer program used in the U.S. to assess the risk of re-offense by individuals in the criminal justice system flagged black defendants as high risk nearly twice as often as white defendants.

Surveillance and loss of privacy: AI algorithms supercharge surveillance and threaten privacy. For example, AI-powered facial recognition software gives closed-circuit TV systems the capacity to track individuals as they move through the urban landscape. This is concerning both socially and politically, as privacy is key to other fundamental rights such as freedom of expression and association.

Job and tax revenue loss through automation: With the growing use of machine learning and AI systems in nearly all sectors of the economy, widespread automation will extend beyond manufacturing to impact higher-skilled knowledge-based roles. Many of these jobs can be partly or entirely automated, reducing the need for human workers. However, a counter-argument has also been made: AI may shift the nature and scope of work and jobs, for instance through robots complementing human labour, and an increased focus on higher-skilled and higher-paid tasks.

Undermining democracy and political self-determination: In a world increasingly connected and reliant on the free flow of information, misinformation is a genuine and growing threat to stability and democracy. For example, by piggybacking on highly personal data collected on social media campaigns, AI applications facilitate more efficient propaganda and behavioural manipulation campaigns. The 2016 U.S. presidential election has become a notorious example of the role of targeted misinformation over Facebook.

THE FUTURE OF AI IN THE GLOBAL SOUTH

There is little doubt that AI technologies will be transformational. Breathtaking advances will be made, extraordinary wealth will be created, and many of our social and institutional structures will be transformed. However, we must ask: whose lives will be improved (or harmed) by these technologies? A key assertion of this paper is that, if we continue blindly forward, we should expect to see increased inequality alongside economic disruption, social unrest, and in some cases, political instability, with the technologically disadvantaged and underrepresented faring the worst.

This prediction stems from the interweaving of two elements: the nature of AI applications, and projections of the impacts of AI applications *in the current global context*. What is worrisome is the dynamic of how our current set of institutions and cultures shapes the evolution of technologies, and how, in turn, these technologies shape these institutions and cultures. One important element of this context is that it is characterized by what can be called an “AI divide” – that is, a gap between those who have the ability to design and deploy AI applications, and those who do not. Furthermore, both the digital (e.g., infrastructure) and analogue (e.g., regulations) foundations required for an ethical and equitable application of AI technologies in many countries in the Global South are largely absent, and salient power asymmetries persist.

This situation makes it all the more important to address the challenges posed by AI so that we can avoid or mitigate these adverse outcomes while enabling developing countries to take full advantage of AI’s positive potential. The promise of AI is too alluring, and its potential too great to avoid an AI future. The question is whether or not we will be ready.

RECOMMENDATIONS

Based on the conclusions of this paper and the broader literature, we have identified three areas in which action can be taken: policies and regulations, inclusive and ethical AI applications, and infrastructure and skills. Within each area, the paper makes a series of recommendations for research necessary to make concrete progress. Note that this is not intended to be a comprehensive list, but rather an overview of the most pressing interventions.

POLICY AND REGULATORY STRUCTURES

FOSTER THE DESIGN OF POLICIES AND REGULATIONS THAT ENABLE INCLUSIVE AND RIGHTS-BASED AI

Conduct baseline research on the prevalence of AI applications and policies in countries of the Global South. Despite pockets of AI activity in the Global South, there are no systematic overviews of the level of this activity. Baseline data collection should include the sets of AI policies, regulations, applications, existing open datasets, and skill levels. This research should be conducted on an annual or, minimally, a bi-annual basis to support continued activities, policy development, and the research agenda.

Learn about effective regulatory models. Document and assess AI regulatory models developed to deal with the emergence of new AI-driven activities such as predictive policing, autonomous vehicles, and chatbots. Determine whether the potential risks of AI applications are adequately addressed by existing regulation, or if existing regulation needs to be adapted or new regulation developed. Identify regulatory responses to given AI use cases and risk levels that are appropriate for settings with low institutional capacity. While lessons learned in the Global North are useful, it is critical not to directly import institutional and regulatory approaches into Global South contexts where institutional and cultural forms differ.

Track the impact of AI on employment and work. Conduct social and economic policy research to understand the effects of AI on employment, the nature of work, and labour markets. To what extent is AI-enabled automation altering employment patterns and transforming the workplace? What are alternative models of income and resource distribution, education, and job retraining in different contexts?

Explore approaches to addressing liability, accountability, and redress for AI decision-making. Design regulatory systems and frameworks to determine liability and accountability for AI decision-making that is erroneous, biased, or discriminatory, and establish mechanisms for redress. Measures may include policies that stipulate transparency for automated decision-making, evaluative procedures to determine the competency of AI systems, and certification of AI systems that engage in tasks requiring a degree of skill or training. The need for action is particularly urgent in the case of decision-making systems that affect people's well-being or freedom, such as those that involve the use of force or incarceration. Research is critical here to uncover and document which systems for accountability and redress are effective and in what contexts.

Study the impact of AI on human rights. At a broad level, the UN recognizes that offline rights apply online, testifying to the relevance of analog rights in digitally mediated environments. Professional bodies specifically call for full consideration of human rights in the context of AI design and operation. Tailoring impact assessments to the risks of AI would help encourage development programs to incorporate AI technology in ways that respect and promote human rights, including privacy, equality, and freedom of expression.

APPLICATIONS

CATALYZE THE DEVELOPMENT OF INCLUSIVE AND ETHICAL AI APPLICATIONS

Support the development and deployment of innovative AI applications for social good. Invest in developing, deploying, and using applications for education, health, the environment, food security, etc., and and ensure that these applications are ethical and inclusive. As with regulatory innovations, while it is important to draw inspiration from examples around the world, AI applications will often require homegrown solutions to be effective.

Research the social impact of AI innovations. Research is needed to better understand which AI applications work (or do not work), for whom, and in what contexts. We need to know who benefits from AI applications and how, as well as who is excluded or harmed. Special emphasis must be placed on exploring the differential impacts on various groups, particularly those resulting from gender, social and economic status, race, etc. This research should go beyond first-order effects, such as increased efficiencies or accuracy of diagnosis, to include broader social effects. New methodologies for impact assessment and evaluation may be required.

Test and monitor bias in AI applications. AI systems that make or inform decision-making that affects humans' well-being (e.g., medical diagnosis, providing a judge with an assessment of potential recidivism) should be tested and monitored for bias and errors across different contexts and communities, both before release and continuously.

Explore models of participatory design for AI. Conduct research into practices that support the development of inclusive AI applications. What techniques are effective for truly participatory processes that engage diverse populations in the design and deployment of AI applications? How and in what contexts do these practices counter design and learned bias and make AI relevant to marginalized communities? AI stakeholders in the field should release data on diversity of participation in design and development.

Action research to deepen understanding of how to effectively and equitably scale proven AI applications. Research on the process of scaling AI applications in a cost-effective and equitable manner, both vertically to encompass additional functionality and horizontally to expand to new locations, is critical to extending the benefits of these applications. Successful transfer of an AI application across contexts requires an understanding of why an application worked well in a particular context and an appreciation that the application may need to be altered in order to succeed in a new environment. Particular challenges related to data will include scaling beyond the scope of existing datasets and developing means to rapidly generate datasets.

INFRASTRUCTURE AND SKILLS

Build the infrastructure and skills for inclusive and ethical AI

Support programs to build AI expertise in government. Promote AI expertise in all branches and at all levels of government, including regulatory, policy-making, and enforcement entities and, potentially, new advisory bodies.

Foster local capacity to lead the design, development, and deployment of AI applications. Activities might include supporting the growth of multidisciplinary AI centres of excellence in the Global South in order to engage in local development and research and provide evidence-based input into the shaping of national policy and regulatory decisions; building bridges between technology experts and low-income and marginalized communities in the Global South; and supporting South–South collaborations.

Develop and test cost-effective approaches to build relevant AI skills, particularly among women and marginalized populations. Develop and support programs that focus on developing the capacities of women and other marginalized populations to engage in different stages of the design and application of AI technologies. Research should bolster this activity through an exploration of low-cost models for developing AI technology skills and producing and testing effective curriculum and pedagogies.

Expand access to data and computing resources. As much as possible, AI research, tools, and training datasets need to be made freely available. Support the development and sharing of diverse and inclusive datasets that are necessary for AI applications in various contexts.

Study the benefits and risks of open AI. Conduct research on the short- and medium-term risks and benefits of openness in AI (e.g., sharing AI resources, datasets, etc.). Where possible, this research should connect supply-side questions (how best to provide open access to AI algorithms, tools, and datasets) with deepening understanding of the engagement necessary to ensure that open AI resources are available for (re)use and adaptation by diverse populations (and not just by those who are already well-skilled and well-resourced). Special attention should be paid to the issue of balancing the sharing of datasets with the safeguarding of privacy.

Artificial intelligence and human development

Toward a research agenda

Artificial intelligence will profoundly impact societies in so-called developing countries both positively and negatively in the next few decades

Introduction

There is little doubt that artificial intelligence (AI) applications will profoundly impact societies in low- and middle-income countries (LMICs), both positively and negatively (Stone et al. 2016). AI is an emerging class of technologies upon which other technologies and applications will be built.[1] Fuelled by the increasing availability of computational power, improved connectivity, and big data, AI applications offer intriguing possibilities for promoting economic growth and tackling a wide variety of longstanding problems in the Global South. The resulting disruptive impact will be massive and could well be revolutionary.

The promise of AI and its potential to help us reach the Sustainable Development Goals (SDGs)[2] is generating a great deal of interest in the development community. For example, the 2017 AI for Social Good summit, organized by the International Telecommunications Union, explored the many ways that AI could contribute to sustainable development through a broad range of applications, as indicated on the following page.[3]

Photo by Igor Ovsyannykov

Artificial intelligence applications

- Mapping poverty from space to enable real-time resource allocation
- Increasing agricultural productivity through automation and predictive analytics
- Analyzing healthcare data to facilitate scientific breakthroughs
- Revolutionizing classrooms by providing individualized learning pathways and virtual mentors
- Driving balanced hiring practices and spotlighting gender inequity
- Using sensors to predict consumption patterns for efficient and safe water provision
- Improving photovoltaic energy capture, thereby lowering the cost of solar power
- Increasing productivity for economic growth through automation, and enabling more efficient use of resources through optimized supply chains, logistical pathways, and scheduling
- Building a more equal and inclusive society through disability robotics
- Supporting urban planning to make cities smarter and more sustainable
- Reducing waste by predicting and identifying optimal production levels
- Predicting climate disasters through improved climate modelling
- Combating illegal fishing by tracking fishing boat movements

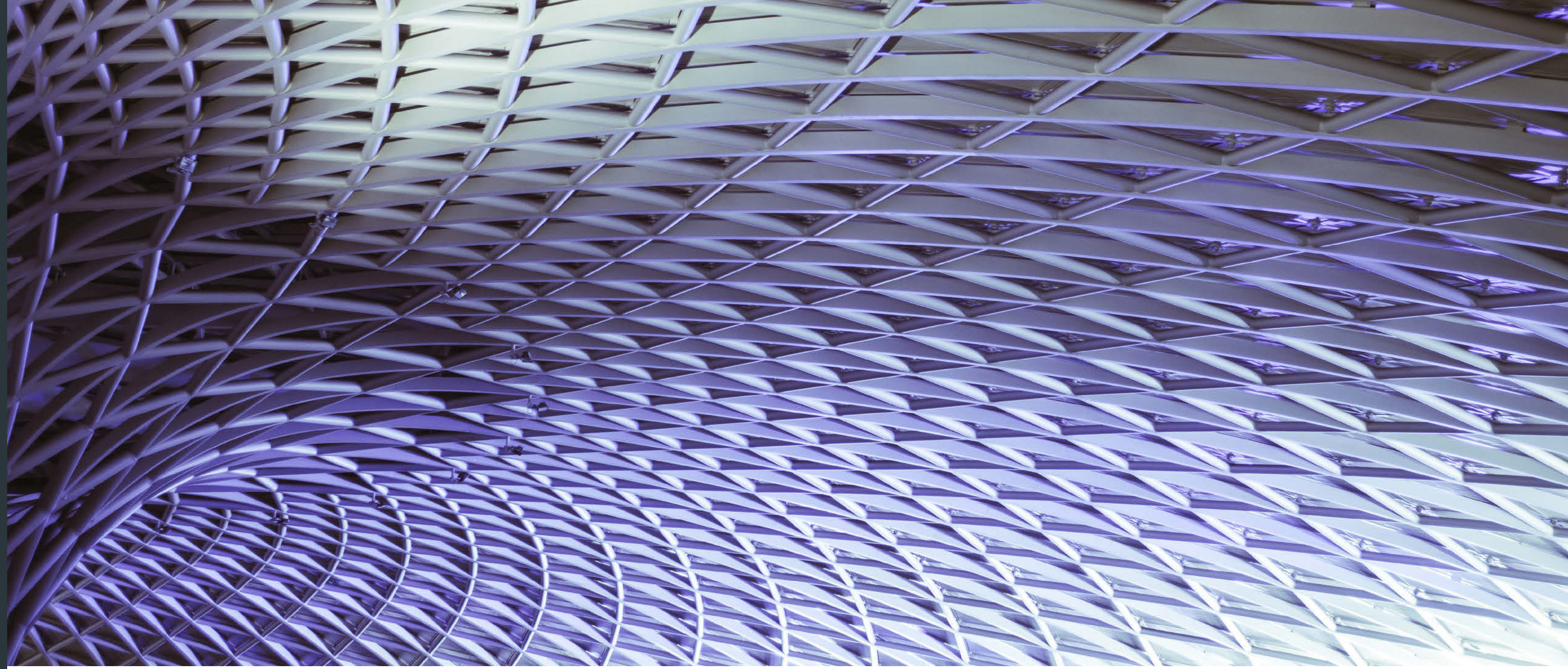


Photo by Clem Onojeghuo on Unsplash

The enthusiasm is similar to that seen for other broadly promising technologies such as the internet and blockchain. The hope is that AI applications will enable radically more effective and efficient approaches to promoting development, which will, in turn, enable us to meet the ambitious SDG targets. In short, AI is seen as holding great potential for innovative, typically data-based, technical solutions to pressing, complex social problems.

However, like most new technologies, AI also has the potential to exacerbate existing problems and create new ones. AI contributes to social ills by, for instance, reinforcing structural inequalities and bias, perpetuating gender imbalances, threatening jobs, and introducing other currently unknown risks and unintended consequences. Recent high-profile examples of AI algorithms gone wrong illustrate some of these risks: a Google algorithm tagging two black people in a photo as gorillas,[4] a driver's licence being revoked when a facial recognition algorithm flagged a photo as probably fraudulent,[5] recidivism prediction algorithms informing life-altering courtroom decisions based on highly biased output,[6] and AI-enabled automation negatively impacting jobs in some sectors and threatening others.[7] However, the risks posed by AI do not only come from biased algorithms. The technology can also be harnessed to promote "negative" social outcomes such as undermining democratic governance and enabling unethical and criminal activity. As James Hendler wrote, "AI can be used for social good. But it can also be used for other types of social impact in which one man's good is another man's evil. We must remain aware of that."[8]

Indeed, the parallel potentials for AI to promote tremendous social good and to cause significant harm are remarkable. From an international development perspective, it is clear that there are and will be numerous life-changing AI applications that can be deployed to great effect in the Global South. However, it is also clear that short- and medium-term risks and potential harms are significant and warrant serious attention. Furthermore, the risks to countries in the Global South are magnified in low- and middle-income settings. This perspective is based on the understanding that the social effects of technologies are shaped by the institutional, political, and economic context in which they are rolled out. Unfortunately, there are fundamental factors typically present in LMICs that will make addressing the challenges and risks of AI more difficult. These include:

- A highly uneven distribution of resources and knowledge required to implement AI techniques (including very large gender and ethnic gaps), and thus an uneven capacity for making decisions about what applications are developed and for whom;
- A large digital divide between the Global North and Global South, as well as a divide between urban and rural settings in countries of the Global South;
- High levels of political and social instability (in some countries);
- High levels of unemployment and low tax bases that limit policy responses to job loss due to automation; and
- Little institutional capacity to protect individuals' rights such as privacy.

Within this context, the risks that stem from AI applications could be particularly harmful in countries with high levels of unemployment and political instability. Given the factors at play, we can anticipate the following general outcomes:

- An unequal distribution of benefits, heavily tilted in favour of the already wealthy;
- The persistence of an AI skills and resources divide that limits the "democratization" of AI-derived benefits and the inclusion of diverse and typically excluded voices in decisions regarding the design, development, and deployment of AI applications;
- The perpetuation and even exacerbation of social marginalization through the automation of unacknowledged biases;
- An expansion of surveillance and threats to privacy, with bad or rogue actors operating with increasing sophistication to foster social discord and political unrest.

These negative effects are self-reinforcing, and disproportionately impact marginalized and economically disadvantaged populations. (We provide more detail on each of these risks in Section 4.2.)

Consequently, the roll-out of future AI applications requires a healthy critical perspective and an ongoing public dialogue. Concerted efforts are warranted to ensure more equitable and inclusive development and deployment of AI in LMICs. These efforts should focus on enabling these countries to take advantage of the potential of AI so they can still benefit from the immense value the technology can bring. At the same time, care and vigilance will be required to mitigate known risks and to identify and respond to unforeseen risks and unintended harmful consequences. Such an approach is critical to prevent the exacerbation of existing inequalities and social instability.

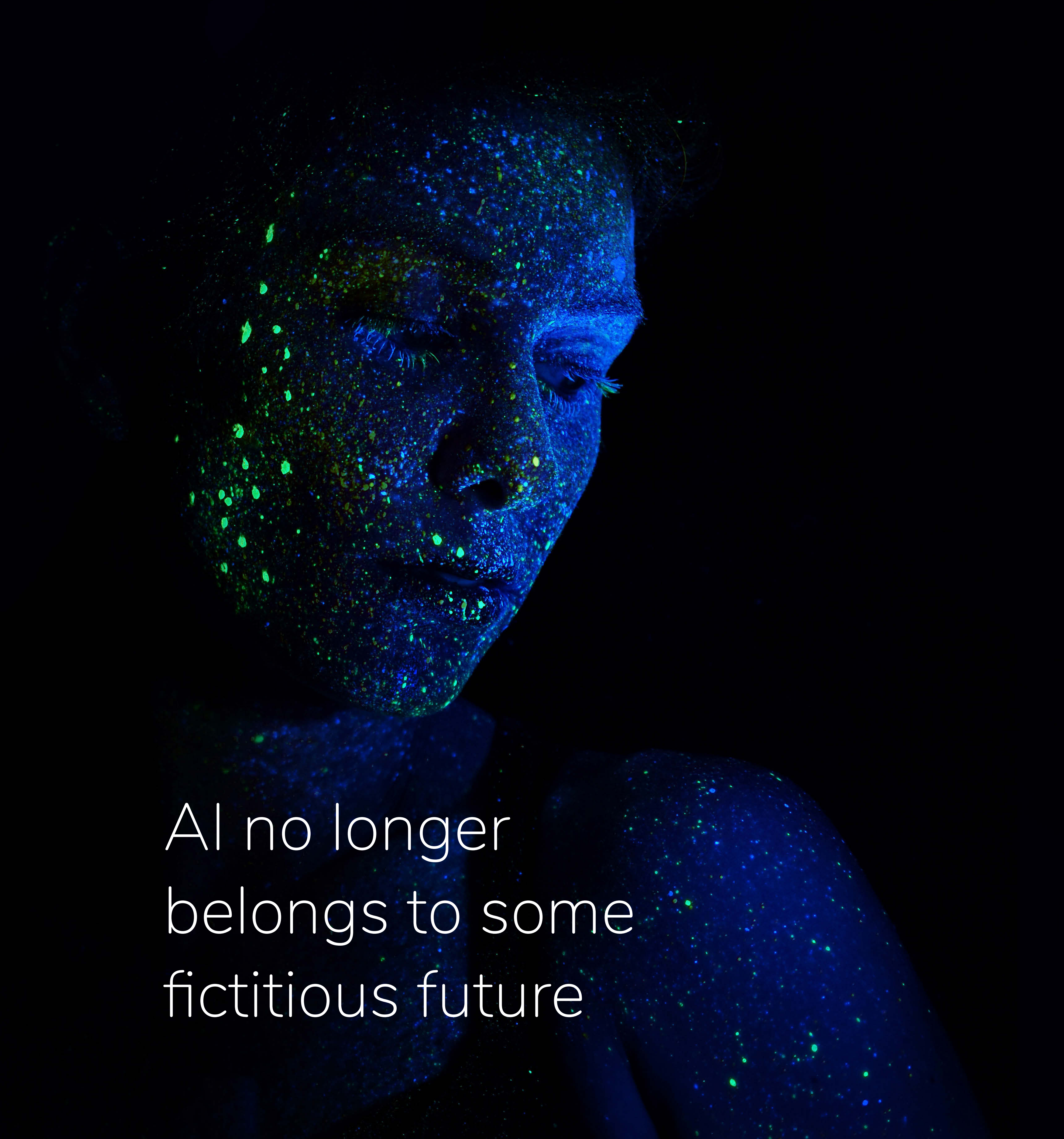
Despite these challenges, we can envision a world where AI innovations are ethically implemented and generate equitable benefits that advance the social condition. AI can even be a means to spur deeper social understanding and overcome longstanding social and cultural biases. The question is, how do we get there?

A solid research grounding will be key to answering this question. Currently, however, precious little empirical research exists to guide the design, development, and deployment of AI in LMICs, or to inform necessary policy and regulatory responses. And open issues abound: How can poor societies take advantage of AI applications to address key development challenges? How should they tackle the tricky ethical, legal, and accountability issues that AI presents? To what extent will AI-enabled automation disrupt economies and potentially eliminate millions of jobs? What are sensible and feasible policy responses?

Consequently, current thinking on the social implications of AI is speculative; much is unknown, and opinions can diverge widely. Given that AI "design and policy decisions made in the near term are likely to have long-lasting influences on the nature and directions of such developments" (Stone et al. 2016: 5), it is imperative that we take steps to fill this knowledge gap as soon as possible.

The goal of this paper is to present a proactive agenda for the ethical and equitable application of AI for development, with a focus on the role of research. Undergirding this agenda is a broad overview of the technologies associated with AI and the opportunities and challenges they present.

In the following section, we offer a working definition of AI and discuss the basics of AI algorithms, which have seen rapid advancement in the last decade. We then identify potential benefits and risks of AI in the context of the Global South. We conclude with an overview of recommendations for advancing the ethical and inclusive development, deployment, and implementation of AI in these contexts.



AI no longer belongs to some fictitious future

What is AI?

Outside the realms of computer science and software development, the concept of AI is familiar to most people as a science fiction trope – often one with threatening or apocalyptic implications. But AI no longer belongs to a fictitious future: it is already part of our world and is poised to have a rapidly growing impact on everyday life.

If we are to critically examine the benefits and risks of AI in the context of development and provide solutions to mitigate the risks, we need to have a clear understanding of what AI is.

Generally speaking, AI is an area of computer science devoted to developing systems that can be taught (e.g., through encoding expert knowledge) or systems that can learn (from data) to make decisions and/or predictions within specific contexts. While the field has existed since the 1950s, the emergence of "big data" (large datasets made possible by the widespread adoption of the internet, mobile phones, and social media) and recent advances in ML techniques, paired with advances in key enabling technologies such as robotics and sensing, have fuelled an explosion of interest in AI and its real-world applications

Photo by H. Heyerlein

What qualifies as “intelligence” is context dependent, and shifts as technology advances

Photo by Alfons Morales on Unsplash

Definitions of artificial intelligence
The phrase “artificial intelligence” (AI) was used for the first time in a 1955 proposal for a study on using computers to “solve kinds of problems now reserved for humans.”[9]
AI is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment.[10]
The field that studies the synthesis and analysis of computational agents that act intelligently.[11]
A constellation of technologies, including machine learning, perception, reasoning, and natural language processing.[12]
A collective term for computer systems that can sense their environment, think, learn, and take action in response to what they’re sensing and their objectives.[13]
Building machines that are intelligent, that can do things that humans can do, and for doing that it needs to have knowledge about the world and then be able to use that knowledge to do useful things.[14]

Table 1. Definitions of AI

One way to think of AI is as an advanced form of data analysis in which a software algorithm can draw conclusions based on the result of the analysis, then act on that conclusion, thus behaving in an “intelligent” fashion. AI algorithms are often used when we “cannot understand a problem well enough to specify it as an [standard] algorithm” (Privacy International 2016).

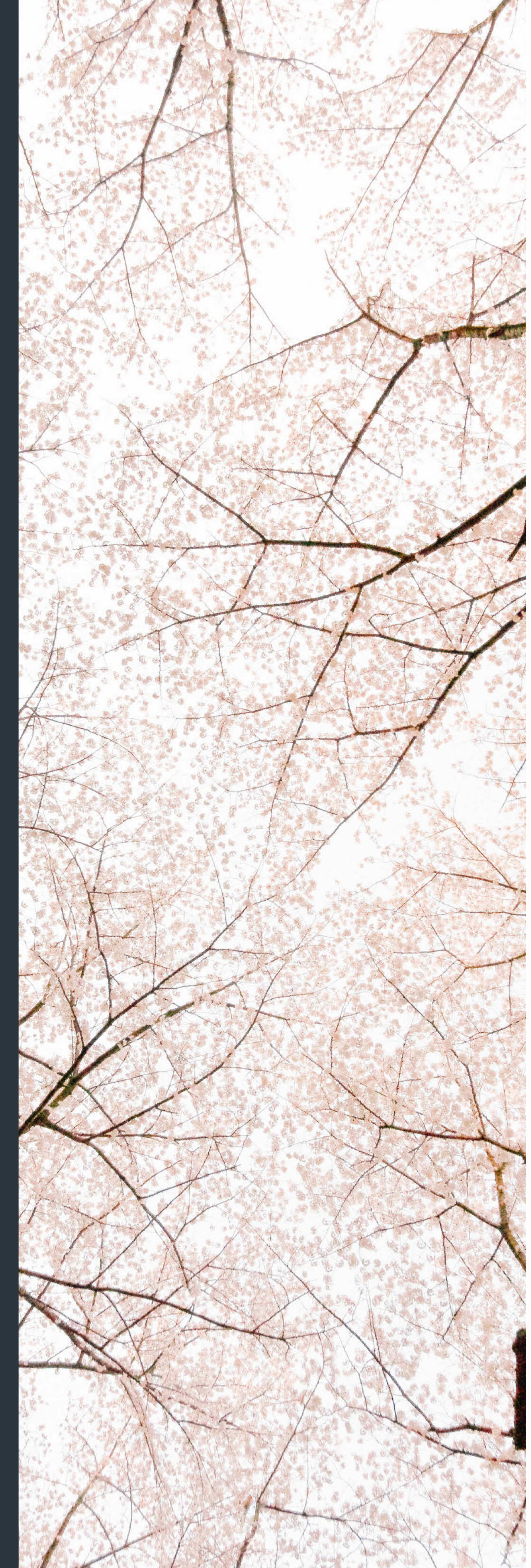
Collecting data on their large online userbases has transformed many companies, such as Google, Facebook, and Amazon, into leaders in big data. The availability of these massive datasets — and the money to be made from mining them — represents a tipping point that has fuelled the growth of AI.

In coming years, another tidal wave of data resulting from the internet of things is set to break. As devices from automobiles to refrigerators to pacemakers become linked to online systems, whole industries will be sitting on large datasets ripe for analysis — and exploitation. Venture capital investments in AI have grown very rapidly since 2013, and the trend appears set to continue.[15] The “big data revolution” is becoming an AI revolution.

As Table 1 shows, there are various definitions of AI, but they are often too vague to draw a clear line between what is considered AI and what is not. This is because what qualifies as “intelligence” is context dependent, and shifts as technology advances. At one time, the calculator was seen as artificially automating human intelligence, but today few people would include a calculator in their working definition of AI.

Intelligent behaviour	**Optimization**
Application	Supply chains, logistical pathways, pricing
Intelligent behaviour	**Pattern recognition/detection**
Application	Face recognition, medical diagnostics, fraud detection
Intelligent behaviour	**Prediction/hypothesis testing**
Application	Flood prediction, disaster prediction, disease outbreak, recidivism
Intelligent behaviour	**Natural language processing**[16]
Application	Voice recognition
Intelligent behaviour	**Machine translation**[17]
Application	Translation of text or speech from one language to another

Table 2. Examples of artificial intelligence behaviours



It is useful to think about AI in two dimensions: what it can do (behaviours) and how it does it (techniques). AI applications can perform a wide range of behaviours, from playing chess to recognizing speech to operating driverless cars. Table 2 gives some examples of behaviours typically considered to fall within the domain of AI. From an AI for development perspective, understanding what AI does (the behaviours) and how they can be applied is key to understanding the range of what is possible.[18]

These behaviours can also be combined to achieve more complex intelligent capabilities. As specific behaviours become increasingly packaged as “skills” or offered as services in the cloud, the development and deployment of these behaviours for different contexts become increasingly available. As various behaviours are combined, the diversity and sophistication of AI applications is increased.

For example, a virtual customer service assistant requires voice recognition and natural language processing capabilities, along with a representation of the expert knowledge of the business. Virtual assistants such as Amazon’s Alexa, Apple’s Siri, and Google’s Home are ever more commonplace, and as their skills improve we expect them to be increasingly deployed in different sectors.

Photo by Robbin Huang

Autonomous vehicles also combine many data-gathering and AI techniques. The driverless car follows traffic rules and navigates obstacles using a combination of hard-coded rules, obstacle-avoidance algorithms, predictive modelling, and "smart" object discrimination (e.g., knowing the difference between a bicycle and a motorcycle). The system responds to its immediate environment using data provided by an array of lasers and other sensors, and achieves the larger goal of reaching its destination using data from navigational systems such as Google Maps.[19]

However, to understand how these behaviours are achieved, we need to examine the various AI techniques on which they are based. The most important of these are ML algorithms and expert systems. We include a brief overview of these techniques as a basis for recognizing their limitations and potential risks, and how to mitigate them. (For more details, see sidebar: *How Does Machine Learning Work?*, and Appendix: *How Does AI Work?)*

Machine learning techniques, such as deep learning and reinforcement learning, enable AI applications to learn by taking in data from existing datasets or through feedback from interaction with their environment. While the availability of big data has made it possible for these algorithms to tackle highly complex problems, their reliance on data has its limitations. First, these algorithms privilege some types of problems and knowledge over others. Not all things can be reduced to numbers and be digitally encoded. Hence, problems that are readily amenable to representation in datasets or have clear desired outcomes are more easily tackled by machine learning (ML). This is one reason AI has advanced so quickly in some areas (e.g., playing games and recognizing images) but not in others (e.g., understanding high-level concepts).[20]

Second, ML techniques that learn by finding patterns in datasets, such as deep learning, have a higher inherent risk of bias and privacy threats. If a training dataset contains biased data, the algorithm will learn this bias. If the algorithm feeds into consequential decision-making processes, it can reinforce or even exacerbate existing social inequalities (see Risk Areas, p. 59). Furthermore, if an algorithm is designed to extract personal information from data, it can jeopardize the privacy of the individuals whose data are in that dataset. Depending on the context, this loss of privacy can seriously undermine fundamental human rights (see Surveillance and Loss of Privacy, p. 68). For more on the limitations of deep learning, see Marcus (2018).

In contrast to ML algorithms, *expert systems* attempt to emulate the problem-solving skills of a human expert by using explicitly encoded knowledge and inference procedures to solve problems. Thus, rather than working with existing datasets, expert systems operate in situations where there is already a corpus of explicit expertise on how to perform a task or solve a particular problem. Expert systems were among the original AI techniques developed before big data, at a time when computational power was limited.

HOW DOES MACHINE LEARNING WORK?

There are three general types of machine learning algorithms: *supervised learning, unsupervised learning, and reinforcement learning.*

Supervised learning algorithms are successively fed example data; for each example, the algorithm provides a response (an output or prediction). The algorithm learns by adjusting its internal parameters such that its response for that particular example will be more accurate the next time.

Unsupervised learning algorithms are fed datasets without any associated outputs; the algorithm extracts patterns (clusters) from the data. Unsupervised learning is especially useful for extracting insights from data.

Reinforcement learning (RL) trains algorithms through positive or negative feedback based on an action taken within a particular environment. RL is appropriate for goal-oriented tasks, such as learning a game (e.g., chess, Go) or teaching a robot a skill (e.g., how to manipulate an object).

See also Appendix: How does AI work?

Development Benefits

SELECT EXAMPLES

Although self-driving cars and speech recognition software are the most media-friendly AI applications, the technology has been applied in many different fields. However, most AI applications have been implemented in the Global North, where the context is more favourable to such applications. So, while there is much interest in AI for development, there are few current applications from which lessons can be drawn.

This section provides a few examples in domains that AI can contribute to social, political, and economic development. This section does not elaborate on the range of possible applications of AI in the Global South; rather, the goal is to provide descriptions of some actual AI applications to give an idea of the current state of AI development and deployment in these contexts.

Photo by Alain Pham

AI can fill gaps in human expertise, improving the productivity of available healthcare workers, and enhancing disease surveillance

HEALTHCARE

Lagging life expectancy and insufficient healthcare resources are ubiquitous concerns in the Global South, where many preventable conditions go undiagnosed and infectious disease outbreaks frequently overwhelm available infrastructure. AI can play a crucial role in augmenting capacity by filling gaps in human expertise, improving the productivity of available healthcare workers, and enhancing disease surveillance (Quinn et al. 2014).

POINT-OF-CARE DIAGNOSTICS

Malaria afflicts more than 200 million people each year, and kills hundreds of thousands. The best way to diagnose the disease is by analyzing blood samples under a microscope, but that analysis requires appropriate technical expertise.

The AI Research Group at Makerere University in Uganda[21] has developed AI software using computer vision techniques, and trained it using malaria samples. Their system outperformed antibody tests, which tend to produce high rates of false positives. Automated systems like this build capacity by supporting the triaging of samples so that healthcare providers in busy centres can work more efficiently, or by extending diagnostic capabilities to rural or remote areas where the expertise is not available (Quinn et al. 2014).

DISEASE SURVEILLANCE

In response to the dengue epidemic of 2011 in Punjab, Pakistan, a disease surveillance system was developed to provide early warning of future outbreaks. The resulting Punjab Intelligent Disease Surveillance System uses statistical learning algorithms to analyze data from a dengue hotline and internet news sources, offering hospitals and government agencies real-time outbreak tracking with a high level of geographic detail (Ahmad et al. 2013).

Photo by Ousa Chea on Unsplash

Photo by Mavis CW on Unsplash

Artificial Intelligence provides tools to help optimize government service delivery

DELIVERY OF GOVERNMENT SERVICES AND INFORMATION

Widespread efforts are underway globally to make government services and information available electronically. The hope is that governments will achieve greater efficiency and enhanced transparency, while citizens will have better access to those services and information through internet and mobile platforms, and find it easier to participate in all aspects of public life. Many countries face a number of challenges on this front, including limited information technology infrastructure, inconsistent electricity delivery, populations that may speak several different languages and dialects, and great disparities in access to internet and mobile services. AI applications have the potential to improve both the understanding and implementation of e-government applications and services. Ultimately, AI provides tools to help optimize government service delivery, ideally maximizing social returns while minimizing financial costs.

ASSESSING E-GOVERNMENT EFFECTIVENESS

Assessing the effectiveness of e-government initiatives is a complex task that must account for a range of technical, organizational, and social factors, such as usability, accessibility, inclusivity, system downtime, upkeep requirements, security, privacy, user satisfaction, and impacts on productivity. Using expert systems and analytical approaches that can accommodate data of varying levels of quantifiability and specificity, groups around the world are exploring ways to use AI to automate e-government assessments to help resource-strapped countries improve their e-government efforts. Examples include research from Bangladesh (Hossain et al. 2015), Taiwan (Yang et al. 2012), and Greece (Magoutas and Mentzas 2010).

IMPROVING GOVERNMENT SERVICES

The specific issue of linguistic diversity can render electronic systems inaccessible to entire groups. South Africa, for example, has 11 official languages. The Centre for Artificial Intelligence Research in that country is working on machine translation approaches to broaden access to government services (World Wide Web Foundation 2017). Similarly, the AI Research Group at Makerere University cited above is working to develop source datasets for some of the dozens of languages spoken in Uganda. Without these datasets, it is not possible to develop the natural language processing necessary for machine translation.

NATURAL DISASTER SUPPORT

One of the earliest applications of AI in the development context was support for planning and mitigation in the event of natural disasters. The hours following a catastrophic event such as an earthquake or a hurricane are chaotic, and the flow of information is overwhelming and difficult to sift through. Two projects represent efforts to capitalize on crowd-sourced support and the real-time deluge of social media. Artificial Intelligence for Disaster Response (AIDR) is an open-source software project that mines, classifies, and tags Twitter feeds during humanitarian crises.[22] To make sense of the flow of information on social media that is beyond the capacity of manual analysis, AIDR uses supervised ML to automate the process, turning the raw tweets into an organized source of information that can improve decision-making and response times.

In 2016, Ecuador experienced a major earthquake. Soon after, a web-based platform for crowd-sourced research called Zooniverse launched a website where volunteers could rapidly analyze hundreds of satellite images in order to target relief work to the areas that most needed immediate assistance. Each contributor's analysis was then verified for reliability by a ML system, and weighted. Two hours after its launch, 1,300 satellite images had been reviewed at least 20 times each, and a heat map of damage was produced two hours later.[23]

On the risk assessment side, researchers in Sweden used expert system techniques similar to the water management example below (see Agriculture, p. 43) to model flood risk in Bangladesh. The system was trained on a complex dataset encompassing a variety of factors related to risk and flood impact. Using real-world data for a specific locality in the country supported the generation of potential scenarios to aid in planning and decision-making.[24]

WILDLIFE CONSERVATION

In Africa, many species are under threat from poachers. Tigers, elephants, rhinos, and other large mammals that are essential to healthy ecosystems and major attractions for tourism are at risk of regional depopulation and outright extinction. To help make the battle against poachers more effective, researchers at the University of Southern California have developed an AI tool for use by ranger patrols. Initially developed in partnership with the Uganda Wildlife Authority, the software uses ML trained on historical data of local poaching activities to produce patrol routes along which poachers are more likely to be found. Datasets include GPS-tagged topographical information, animal sightings, and evidence of poaching such as carcass and snare locations. Trial versions of the system have been tested in Malaysia and in Queen Elizabeth National Park in Uganda.[25]

PUBLIC JUSTICE

Rampant drug violence has plagued Mexico for many years, resulting in the disappearance of more than 30,000 people since 2006. It is not uncommon for the bodies of these *desaparecidos* to eventually be discovered in mass graves, but finding these sites is often a matter of chance. A collaboration among three groups in Mexico and the United States is applying ML to make the process more systematic. Mexico's 2,457 counties are used as geographic units. A detailed sociodemographic profile developed for each county is fed into the system, as are data drawn from media coverage of every grave discovery of a grave. The model uses this input to identify which counties are more likely to have been used as sites for hidden graves. While the collaboration reports a high degree of accuracy (100% for 2014), that accuracy is data dependent, and the collection of media-sourced data is laborious and time consuming.[26]

Photo by Ray Hennessy on Unsplash

AGRICULTURE

In many countries in the Global South, agriculture is an important component of the economy, and much of the population relies on farming as a source of food. However, healthy crops and successful harvests can fall prey to disease, insects, and drought. AI applications can provide critical insights and solutions that can improve the efficiency and quality of agricultural activities.

AI applications can identify disease and infestation in crops by analyzing photos taken with cell phones

Photo by elizabeth lies

CROP DISEASE MONITORING

Monitoring of crop diseases is a time-consuming endeavour often requiring expert knowledge that may not be locally available in a timely fashion. Following are two examples of AI applications that can identify disease and infestation in crop plants by analyzing photos taken with cell phones.

Cassava is a staple food in Africa that is prone to diminished agricultural yield as a result of viral diseases. A comprehensive survey to diagnose diseased crops and map the extent of disease spread can take months and require significant travel by surveyors. The AI researchers at Makerere University have also developed a method to optimize this typically paper-based process by collecting sample images – taken with cell phones – for analysis and classification by an AI system. Images of disease symptoms, such as root damage and white fly accumulation on leaves, are fed into a ML algorithm to provide rapid diagnosis and feedback, and a mapping function can be used to assess and predict disease spread over time. This same group has developed similar systems to assess disease in banana plants using images of the leaves (Quinn 2013).

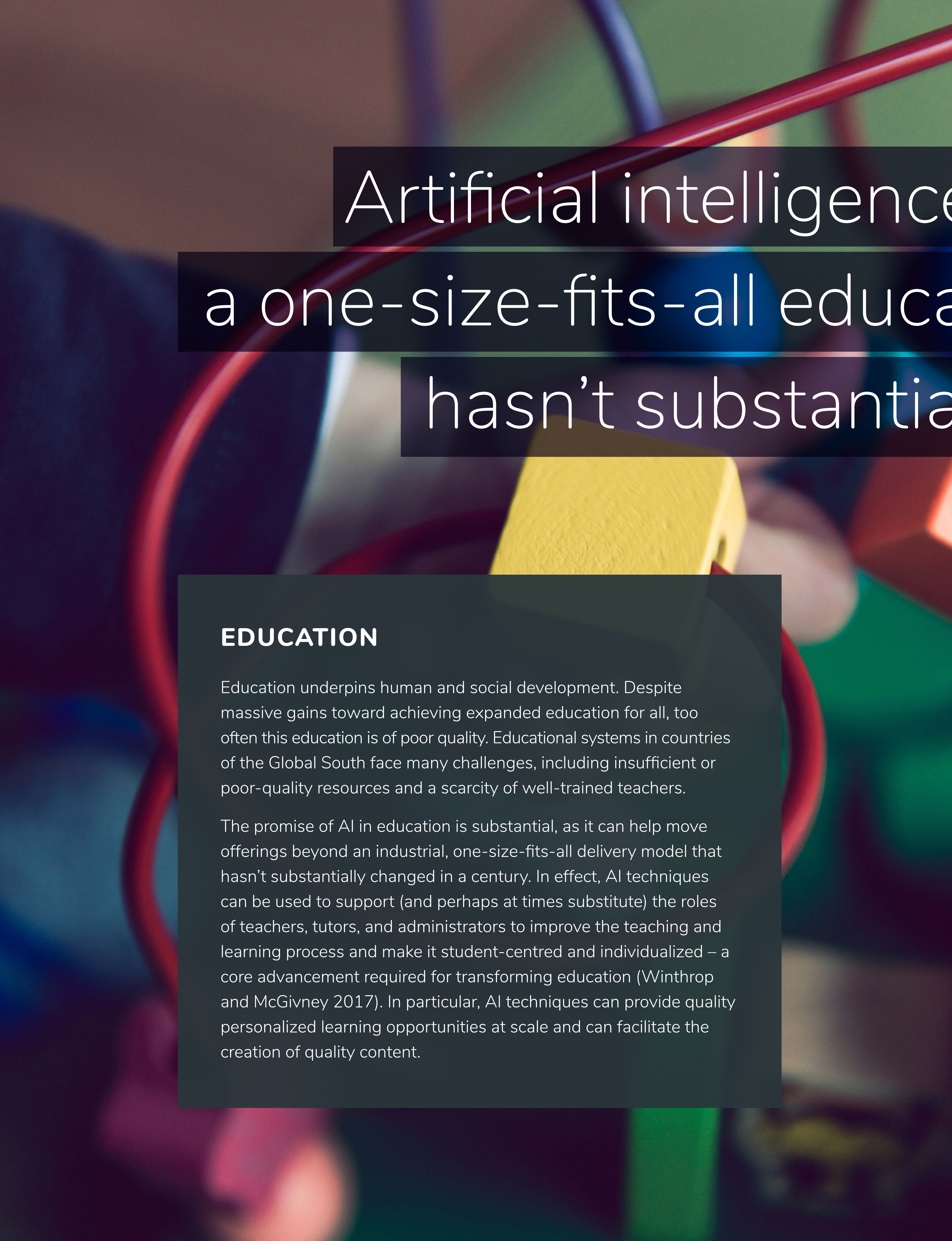
Researchers at Pennsylvania State University and the Swiss Federal Institute of Technology adopted a more generalized approach, developing an AI application that can analyze photographs to identify crops and diseases with nearly 100% accuracy. Using a deep learning approach, they trained the system with a database of more than 53,000 photos of healthy and unhealthy plants from an open-access image archive. The system can recognize 14 crops and 26 diseases in any of 38 combinations. A system such as this can be used by farmers to rapidly identify crop diseases in the field, using photos taken with smartphones.[27]

WATER MANAGEMENT AND DROUGHT MONITORING

In response to water scarcity issues in Palestine, a group in Austria collaborated with Palestinian colleagues to develop an expert system for mitigating loss in water systems. With the goal of optimizing a water loss management strategy, the system assessed a range of specific actions – such as active and passive leak control, water pressure management, metering, and public awareness campaigns – against a set of economic, environmental, technical, and socio-economic evaluation criteria. To address the complex problem of incorporating many variables and a combination of qualitative and quantitative data, the developers employed a type of analytical mathematics to rank available actions appropriate to local conditions.[28]

Access to timely, summarized climate and agricultural information for specific localities can help detect and prevent or respond to catastrophic conditions such as drought. The group at Makerere University in Uganda developed an automated system to produce such reports. Based on manually identified target topics (e.g., water, rainfall, soil conditions) and search criteria, the system scans the web, downloads high-ranking search results, extracts content, and summarizes the information to identify relevant trends for the specified topics (Quinn et al. 2014).

Photo by freddie marriage on Unsplash



Artificial intelligence can help move beyond a one-size-fits-all education delivery model that hasn't substantially changed in a century

EDUCATION

Education underpins human and social development. Despite massive gains toward achieving expanded education for all, too often this education is of poor quality. Educational systems in countries of the Global South face many challenges, including insufficient or poor-quality resources and a scarcity of well-trained teachers.

The promise of AI in education is substantial, as it can help move offerings beyond an industrial, one-size-fits-all delivery model that hasn't substantially changed in a century. In effect, AI techniques can be used to support (and perhaps at times substitute) the roles of teachers, tutors, and administrators to improve the teaching and learning process and make it student-centred and individualized – a core advancement required for transforming education (Winthrop and McGivney 2017). In particular, AI techniques can provide quality personalized learning opportunities at scale and can facilitate the creation of quality content.

PERSONALIZED LEARNING

Perhaps the key move away from a teacher-centred learning process is the development of personalized learning. One way to provide this is through intelligent tutoring systems (ITS), also known as cognitive tutors. An ITS is typically an expert system that attempts to recreate one-on-one instruction by adapting and personalizing the learning experience to the individual learner:

> *An ITS assesses each learner's actions within these interactive environments and develops a model of their knowledge, skills, and expertise. Based on the learner model, it can tailor instructional strategies, in terms of both the content and style, and provides relevant explanations, hints, examples, demonstrations, and practice problems to individual learners (Phobun and Vicheanpanya 2010: 4065).*

As an expert system, an ITS incorporates three types of knowledge models: the expert (representing the domain expertise of the teacher), the learner (models of how the individual learns), and the instructional (for making decisions about instructional tactics). See Figure 1 below.

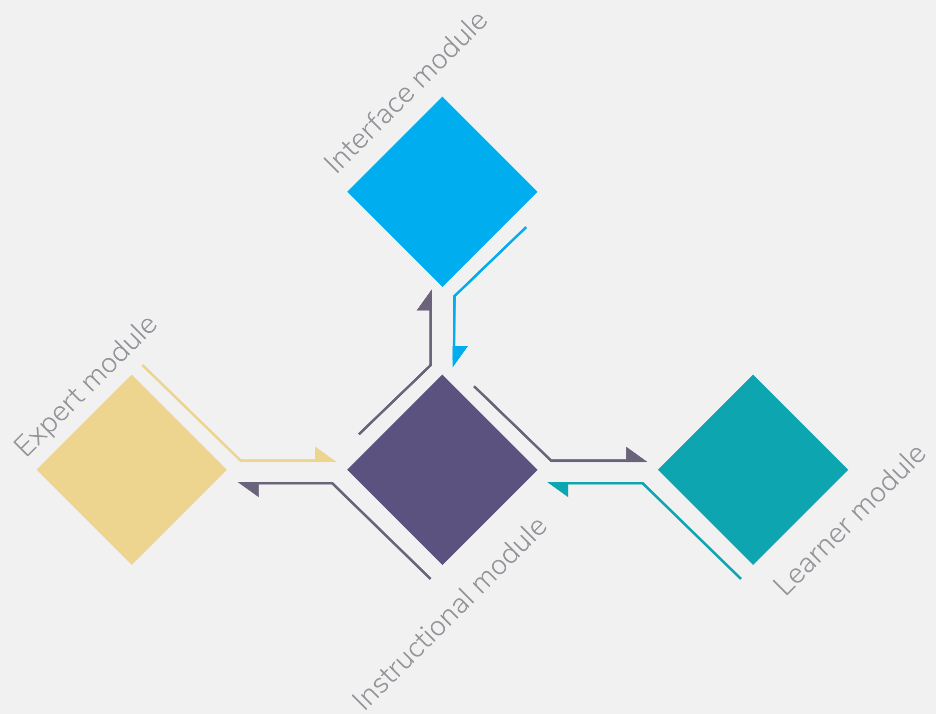
ITS are also beginning to incorporate other AI techniques to enhance instruction. Some ITS — known as affective tutoring systems — also incorporate emotional recognition as a means to enhance the tutoring adaptation to the student (Petrovica et al. 2017). ITS can also apply natural language processing and speech recognition to help identify language errors or interact with students in novel ways.[29]

Photo by markus spiske on Unsplash

There is emerging evidence that personalized instruction by an ITS can be highly effective. A 2016 review of studies on the effectiveness of ITS found substantive improvements in performance in 78% of the 50 studies reviewed (Kulik and Fletcher 2016). There is strong evidence to support the use of ITS in Global South contexts as well, despite known barriers (Nye 2015). Done well, ITS can be a highly inclusive, cost-effective way to improve educational delivery.

Mindspark, an intelligent tutoring platform developed in India, is a case in point. Mindspark identifies the pattern of errors made by student users, and tailors subsequent lessons and tutorials accordingly. A rigorous evaluation of Mindspark found that the program was both cost effective and "successful at targeting instruction precisely to the students' level of achievement and in handling wide variation in the academic preparation in the same grade" (Muralidharan et al. 2017: 23).

FIGURE 1 Structure of an intelligent tutoring system



A typical structure of an intelligent tutoring system incorporating three knowledge models: the domain expert (the teacher), the learner, and instructional (instructional tactics).
From: Phobun and Vicheanpanya 2010: 4066

Photo by Annie Spratt

Learning analytics is "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Long and Siemens 2011: 34). To do so, learning analytics take advantage of large amounts of educational data and ML techniques. For example, applying learning analytics to large amounts of learner data can help to improve the knowledge models in ITS, such as by detecting learning tasks that offer the most effective gains (Lim and Tinio, 2018). Learning analytics can also employ predictive analytics to identify students who are at risk of failing a course.

Early examples from the Global South show some promise. For example, Mwalumbwe and Mtebe (2017) applied learning analytics to data on student engagement collected with a learning management system to predict students' academic performance. Reyes (2018) used similar data to provide input to a student support system at the Open University of the Philippines.

LEARNING AT SCALE

A key challenge to providing quality learning at scale is doing so at low cost. Increasing access to improved connectivity, coupled with online digital learning approaches such as massive open online courses (MOOCs), has improved the capacity to provide a large number of people with quality educational content and experiences. While early evidence on MOOCs shows that people with higher levels of education and socio-economic status tend to benefit disproportionately (Christensen et al. 2013; Hansen and Reich 2015), some research illustrates that certain types of MOOCs – such as those offering job-related training — can benefit users from low- and middle-income populations in the Global South (see, e.g., Garrido et al. 2016).

AI techniques have the potential to build on online learning to achieve the lofty goal of delivering high-quality learning at scale, particularly to marginalized populations. Key to doing so is overcoming the bottlenecks that arise when engaging with large numbers of students — chiefly the lack of human resources to provide individualized feedback, guidance, and assessment of student performance. A combination of AI techniques can help handle high student loads. Intelligent tutors and learning analytics, discussed above, can help provide high-quality personalized learning experiences at a very low cost per additional student (Laurillard et al., in press). Furthermore, developments in automated scoring using natural language processing and other techniques enable the mass grading of quizzes, exams, and essays.

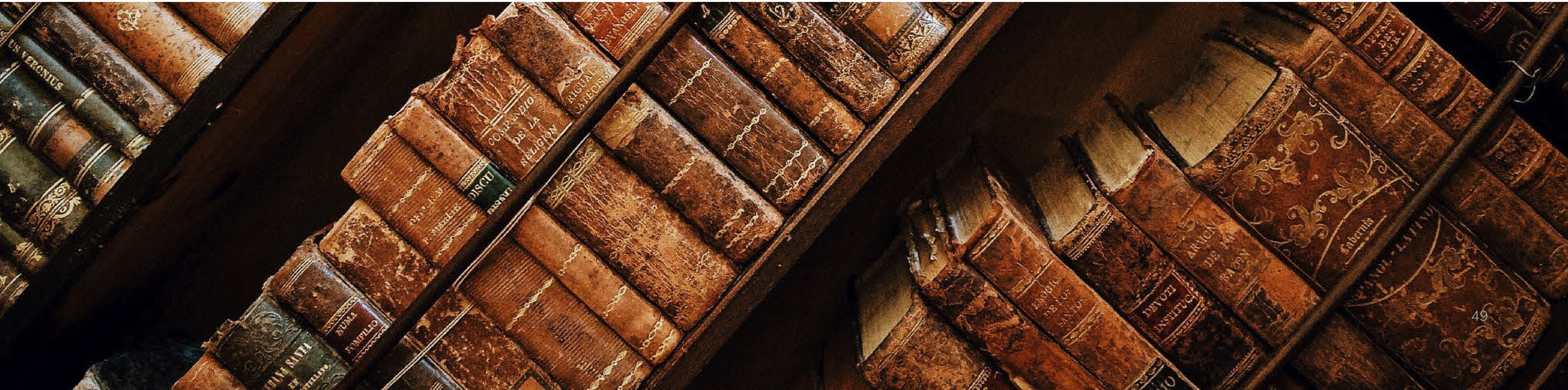
CUSTOMIZING/LOCALIZING CONTENT

AI techniques can also help improve educational content at low cost, for example by providing automated translation of existing works into new languages or leveraging AI to create new content.

For example, the Pratham Books StoryWeaver platform is working with Google.org to leverage Google's AI-powered translation tool to translate children's e-books into as many as 60 different languages.[30] Indeed, the ability to leverage automatic translation should greatly facilitate the localization and use of high-quality open educational resources (freely available, typically digital educational materials) in the Global South (Smith 2013).

Another company, Content Technologies, Inc.,[31] uses ML techniques to create custom textbooks. Educators feed their syllabi and material into the AI engine, and the system creates textbooks and classroom material based on the core concepts it extracts.[32] One can also envision the automated creation of personalized study guides, quizzes, and tests, which would be particularly helpful in massive online learning environments.

Photo by Roman Kraft on Unsplash



Artificial Intelligence could contribute up to $15.7 trillion to the global economy in 2030

PwC 2017: 5

ECONOMY AND BUSINESS

AI can be an important leverage tool for economic development through various means. A 2017 PwC report estimated that “AI could contribute up to $15.7 trillion to the global economy in 2030, more than the current output of China and India combined” (p. 5), stemming from productivity gains through businesses automating processes and augmenting their existing labour force with AI technologies, as well as increased consumer demand for higher-quality, AI-enhanced products and services. AI offers the potential to increase productivity by creating efficiencies, automating work processes, and optimizing key business activities such as supply chains and pricing.[33] However, the PwC report qualifies the potential economic benefits in low- and lower-middle income countries, saying that increases will be “more modest ... due the much lower rates of adoption of AI technologies expected” (PwC 2017: 9).

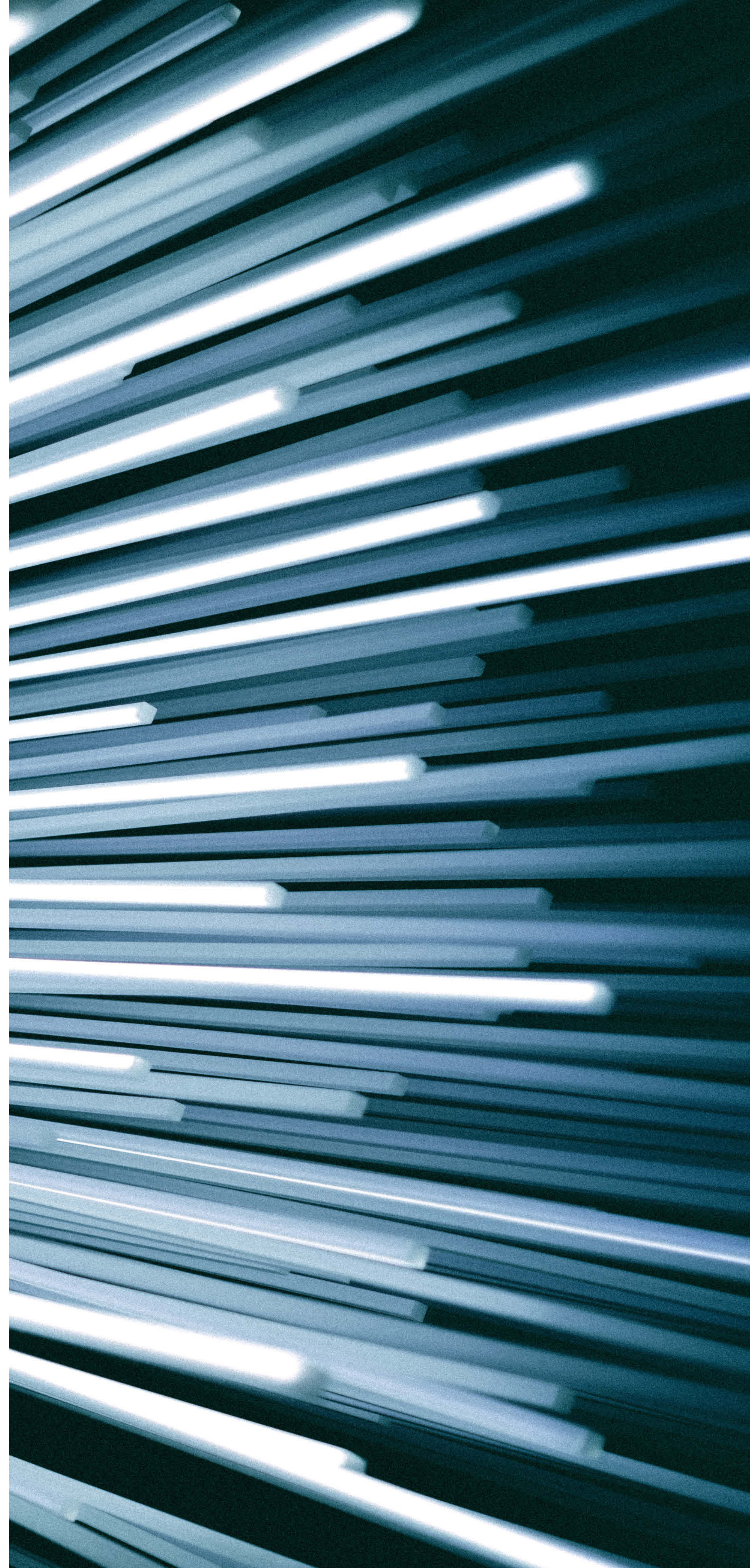
Here we provide examples of two areas where AI is beginning to contribute to growth in the Global South: (1) supporting new businesses and innovation, and (2) optimizing economic building blocks such as financial services.

Photo by Joshua Ness

ENTREPRENEURSHIP AND INNOVATION

The availability of vastly increased computing power for a fraction of what it would have cost a decade ago and the development of open-source software have sparked an explosion in tech start-ups all over the world. Given the many potential applications of AI, there are substantial entrepreneurial opportunities in the Global South, and they will likely increase with advances in infrastructure and new methods of data collection. Evidence of the rise of AI start-ups can already be seen in a number of emerging economies. The following are just a few examples:

- India is home to many start-ups that capitalize on AI. Applications include developer tools for creating AI systems and neural networks, learning platforms, human resources planning, a personal assistant for travellers, delivery and logistics optimization, and remote cardiac diagnosis (Manyika et al. 2017).
- LangBot, a start-up in Addis Ababa, Ethiopia, is developing "gamified AI-powered language teaching chatbots." The early-stage company recently won the Ethiopia portion of the SeedStars start-up competition focused on emerging markets.[34]
- In Pretoria, South Africa, the e-health start-up hearX Group develops digital tools to assess hearing loss and detect ear disease. One product in development is a low-cost, cell phone-connected otoscope used to image the eardrum. The device will use image analysis and AI to generate automated diagnoses for common ear diseases. The group's crowdfunding campaign is oversubscribed.[35]
- The subject of sexual health is taboo in many African settings. This contributes to high rates of sexually transmitted infections (e.g., HIV) and unwanted pregnancies. A Kenyan technology incubator recently supported the developers of Sophie Bot, an AI-driven chatbot that lets users privately obtain accurate information about reproductive and sexual health.[36]



FINANCIAL SERVICES

In the Global South, large segments of the population do not use formal banking or financial services. In Africa, that number is over 325 million people; however, substantial mobile phone penetration in many regions offers a platform for accessing these services. A company called MyBucks is using AI to support the delivery of virtual services in at least nine African countries. The company has also acquired several bricks-and-mortar financial institutions in sub-Saharan Africa to extend its reach to poor and underserved communities. MyBucks uses AI to automate tasks such as credit scoring, fraud detection, and optimization to keep costs low for the microloans, savings accounts, insurance, and transaction options it offers.[37]

While MyBucks was one of the first enterprises to implement AI in financial services in Africa, others are moving in that direction. In 2016, Barclays Africa launched a customer service chatbot that uses AI to simulate "intelligent conversation" for handling simple queries.[38]

In Nigeria, another chatbot, Kudi AI, is a personal banking assistant that uses natural language processing to let users conduct simple financial transactions (such as paying bills or transferring funds) via Facebook Messenger. Since the service is offered via the Facebook Free Basics platform, there are no data charges for using it. Bank-to-bank transfers are free, and other transactions (such as bill payments) cost the equivalent of about 30 cents.[39]

Photo by Christopher Burns

The foundations required for an ethical and egalitarian application of artificial intelligence are largely absent

Challenges and risks

As illustrated in the preceding section, AI applications have the potential to support positive social change – indeed, in some domains their impact could be revolutionary. However, as with any new technology, actually achieving these positive results is challenging and not without risk. In this section, we first lay out some of the contextual factors that are potential obstacles to the effective use of AI for development. We then discuss a series of risks of negative social outcomes stemming from the deployment of AI applications in challenging Global South contexts.

Photo by Yiran Ding on Unsplash

CONTEXTUAL CHALLENGES

As the 2016 World Development Report *Digital Dividends* (World Bank 2016) argues, the ability of countries to achieve a positive and equitable distribution of benefits derived from digital technologies requires the existence of a set of prerequisite "analogue" foundations: regulations, skills, and institutions. We also know from past research and experience that counter-power is required to challenge the inequalities that persist (Benkler 2011). In the case of AI, both digital and analogue foundations required for an ethical and equitable application of the technology in many countries in the Global South are largely absent, and salient power asymmetries remain.

Some of the key contextual challenges include:

Institutional capacity/regulation

Governance and regulation of technology are rudimentary in many countries in the Global South. This leads to opportunities for technological innovations that aren't always possible, or are slow to gain ground, in the more highly regulated Global North. However, with limited regulation comes the potential for greater ethical, social, and political harm. Some experts, for instance, have warned that, from a security perspective, unfettered technological innovation in places like sub-Saharan Africa is a ticking time bomb.[40]

Governments, academia, and civil society have limited capacity and resources to design regulatory frameworks for technological innovation, notably frameworks that find the balance between enabling innovation while protecting privacy, security, or the environment.

The strong push from the private sector in the Global South to promote that sector's specific AI solutions can crowd out homegrown technology development.

Infrastructure

Despite rapid advances since 2010 in the spread of mobile phones and broadband internet, the basic communication and digital service infrastructure is still insufficient in many, particularly lower-income and rural, contexts. Although we anticipate this access divide will decrease substantially over the coming decade, a persistent divide between a digital "fast lane" and "slow lane" will most likely remain (Hilbert 2016), leading to questions about who can realistically innovate in AI for the foreseeable future.

There is a lack of datasets for training AI algorithms specific to countries in the Global South, leading to a "data divide." Poor connectivity and limited data collection technologies in many locales (particularly rural areas) will continue to make data collection highly challenging, if not impossible. Over the next five years, mobile phones and financial services data will be the most likely source of data in these countries.

Skills

Insufficient funding and policies for science and innovation in the Global South hinder the development of local AI expertise and research. As a result, these countries have limited possibilities for organic AI innovation and application development, and must be importers rather than creators of the technology. This also impedes the design of locally relevant AI applications.

There is a general skills shortage for the development and deployment of AI applications, and a well-recognized lack of diversity among those who have the skills. This lack of diversity contributes to bias in AI applications. For example, as in the tech sector generally, there are few women working in AI,[41] and a majority of the men are white.[42] This disparity stems from systemic and cultural influences. In the U.S., boys outnumber girls in computer science advanced placement exams, and in 2013, only 26% of the country's computer professionals were women. We can expect to find a similar pattern of gender disparity across countries in the Global South.

A significant comprehension gap exists between social scientists, policymakers, NGOs, and those with technical understanding of AI. Those who implement AI are generally not aware of its social, political, and ethical implications, and those responsible for regulating or applying the technology for social good have little understanding of how it works.

Photo by Marcus Cramer

An important element of the current global context is that it is characterized by what can be called an "AI divide" — that is, a gap between those who have the ability to design and deploy AI applications, and those who do not. It can be thought of as the gap between those who participate and have voice and agency in shaping which and how AI applications are developed and rolled out, and those who are excluded of the process.

This AI divide transcends geographic, socio-economic, gender, and race boundaries. The infrastructure required for the development of AI applications restricts this activity, for the most part, to locales with sufficient computing power, access to (or resources to collect) relevant data, and the requisite AI skills. The geography of the participation gap is perhaps best illustrated by the relative dominance of a few countries (and a few large tech companies) in the development of AI. A PwC analysis estimated that 70% of the global economic impact of AI will accrue to China and North America (PwC 2017). Thus, for example, most autonomous automobiles are being designed and tested in Global North contexts with relatively high-quality roads, good marking and signage, and orderly traffic. These vehicles would likely not fare well in some Global South contexts. Such technologies need to be at least trained and tested in those contexts if they are to work there.

Beyond geography, the disparity in the AI skills necessary to design and deploy AI applications is reflected in the gender, ethnicity, and socio-economic class of those doing the work. Individuals with AI skills are typically male, white, and based in the Global North, where the majority of AI activity is happening. A World Economic Forum report (2016) highlighted that around the world, only 32% of STEM graduates are women. As a consequence, those who decide what is designed and deployed will have to be cognizant of the limits of their contextual knowledge and of their own potential social biases.

RISK AREAS

In this section, we discuss five types of short- to medium-term risks associated with the application of AI.[43] Two of these risks — biases and lack of transparency in decision-making by AI applications — are inherent to the technology itself. The other three — increased surveillance and loss of privacy, targeted misinformation, and automation leading to job loss — are the result of specific applications of AI in different domains. While these risks apply to any economic or social context (the majority of the examples are from the Global North), the political and institutional contexts of LMICs potentially exacerbate the risks.

FAIRNESS, BIAS, AND ACCOUNTABILITY

Humans are, by nature, biased and prone to making mistakes. Even when we make a conscious effort to act impartially, our prejudices and biases can be so ingrained that we are not even aware of their influence. Justice, for example, is "blind" in theory, but bias is clearly present in judicial systems around the world.

Computers and algorithms, on the other hand, are symbols of efficiency and accuracy. As machines and digital code, they are expected to function in a fair and unbiased way. It would seem logical, then, to design software and AI algorithms to overcome these human limitations, and AI applications have been developed to support judges in the hopes that they will hand down fairer and less biased sentences.[44]

However, our confidence in the impartiality of computer systems is misplaced. Because they are designed and developed by humans and trained on data generated by humans, these systems incorporate the biases of their designers and programmers and of the larger culture in which they are created (Friedman and Nissenbaum 1996; Friedman et al. 1996).

Friedman and Nissenbaum (1996) define biased computer systems as those that "systematically and unfairly discriminate against certain individuals or groups of individuals in favour of others." This definition remains appropriate for AI applications. Indeed, there is already evidence that these applications can systematically and unfairly discriminate in critical areas of people's lives. This is not surprising given AI's potential to have widespread impact on decisions about individual access to services, employment, and financial support. With such extensive influence, biased AI applications can be sources of significant unfairness and injustice in society.

Design bias

The design process is one potential source of bias in AI applications. It is well understood that technologies will reflect many of the often implicit values of those who design and build them (Nissenbaum 2001; van de Poel and Kroes 2014). Much recent media coverage has been devoted to the masculine slant prevalent in Silicon Valley and the software industry generally. This type of bias shapes both the selection of applications for development and the features of those applications. An example of this bias can be seen in the choices made by designers in assigning gender to AI applications (see Figure 2 below). To date, digital assistants are predominantly female, while AI movie characters are primarily male. Admittedly, this is relatively new territory at the commercial level, and the industry seems to be grappling with the best options for chatbots and digital assistants — female, male, a choice of either, or gender neutral.

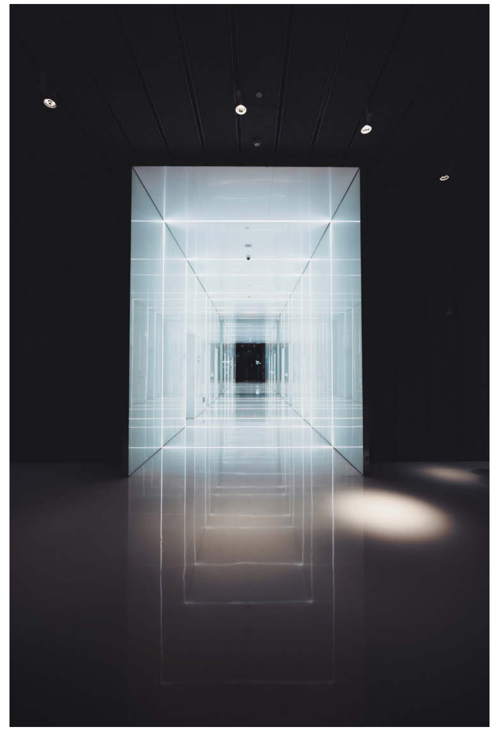


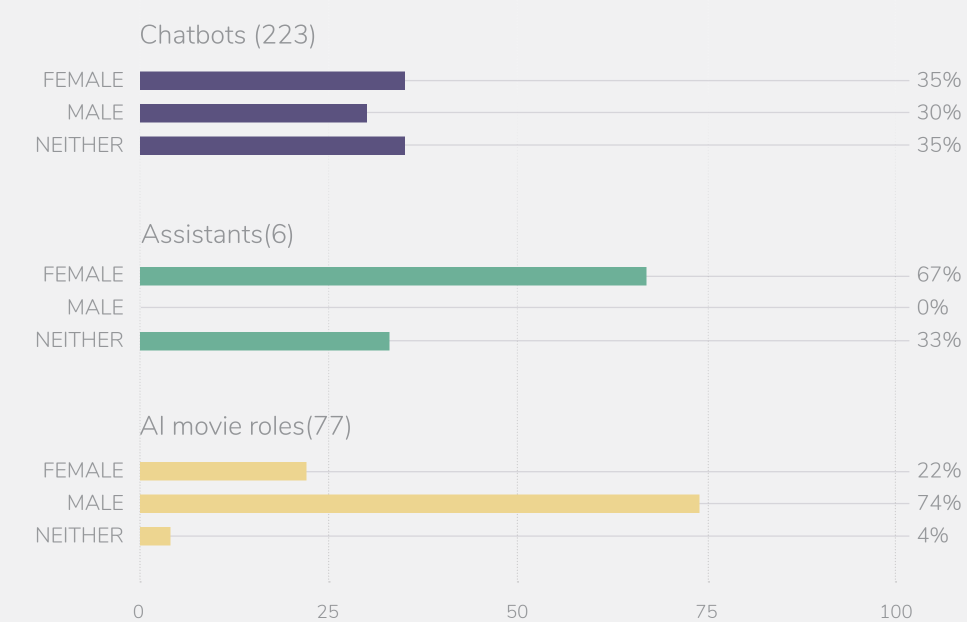
Photo by Alessio Lin

Other biases might include the language in which an algorithm operates, the operating system it is built for, and assumptions about its users. Each of these design aspects entails decision points that may or may not be recognized as value-laden but that can nevertheless introduce biases that shape who can use and benefit from an AI application.

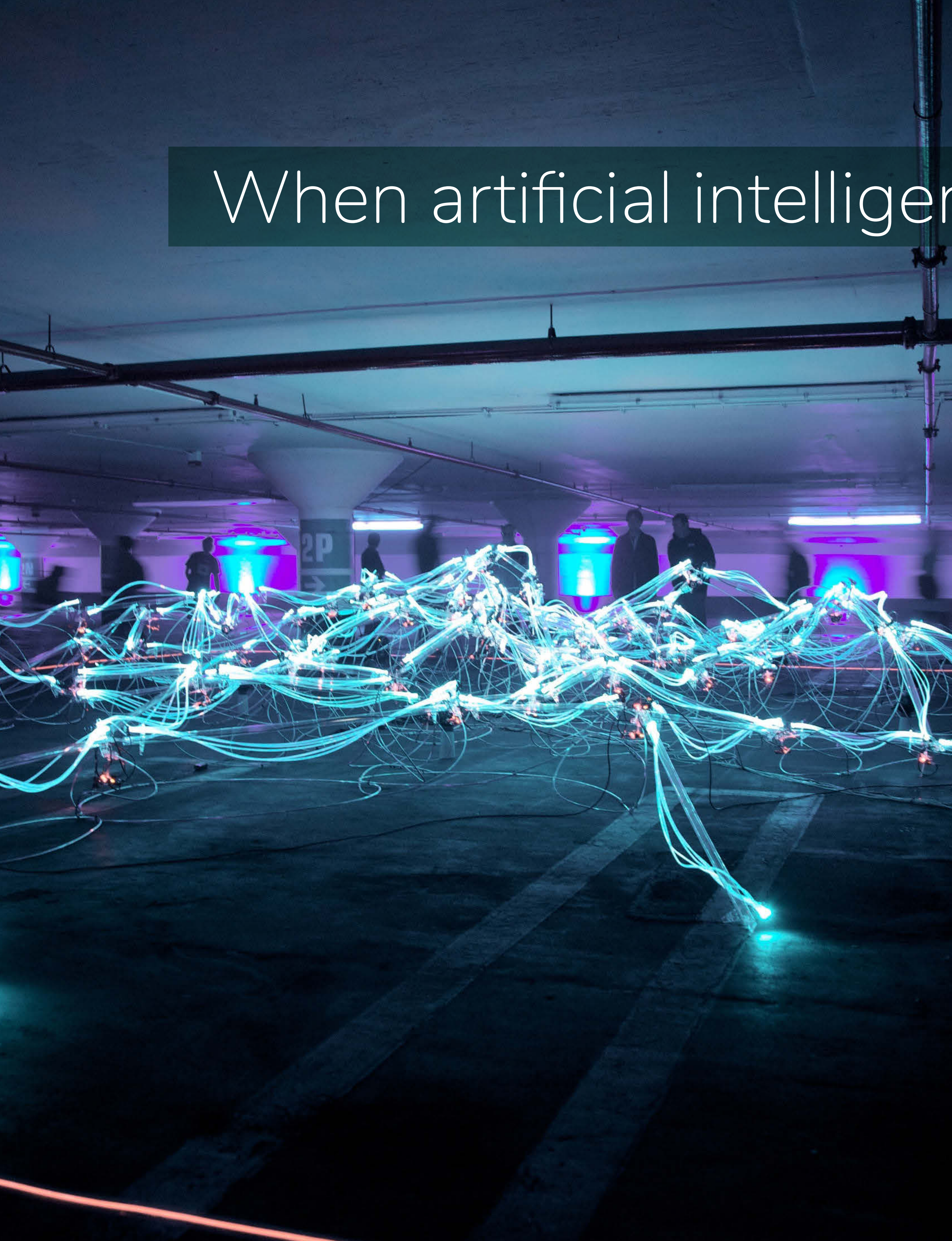
One would imagine that there is also a potential for design bias in expert systems, which are built on knowledge bases that reflect the expertise of human contributors. For that reason, attempts have been made to establish accreditation processes to evaluate the outputs of these systems, much as doctors and teachers are accredited. However, the rest of this section will focus on ML algorithms, as considerably less has been written on the topic of bias in expert systems.

Photo by Mike Kononov on Unsplash

FIGURE 2 Gender distribution in artificial intelligence applications



SOURCE: Crowdflower | bit.ly/2IKzPcr



When artificial intelligence learns from humans, it's bad

James Crowder*

Learned bias

ML algorithms are not biased in and of themselves; they *learn* to be biased. This *algorithmic bias* (Danks and London 2017) has received a great deal of attention. It occurs when the learning algorithm is trained on biased datasets and subsequently "accurately" learns the patterns of bias inherent in the data (see, e.g., Caliskan et al. 2017). In some cases, the learned representations within ML algorithms can even exaggerate these biases (Zhao et al. 2017). Algorithmic bias has two sources: incomplete datasets and datasets that represent biased social phenomena.

Incomplete datasets are those that are not representative of the entire range of potential examples. Consequently, an algorithm trained on an incomplete dataset will perform poorly when given an example that falls outside the scope of the available data. For example, a facial recognition algorithm that is not trained on a wide variety of skin colours might not function accurately for faces of all skin tones. Indeed, one AI researcher with dark skin discovered that an otherwise functional facial recognition algorithm failed to recognize her face unless she put on a white mask[45] . Many similar examples have been reported.[46]

When critical decisions are made based on input from an algorithm trained on a database that is not representative of the entire user population, the results can impact health and well-being. Most clinical trials have highly selective criteria that exclude women (especially pregnant women), the elderly, and those with conditions beyond those being studied. Thus, participants tend to be white males.[47] In some cases, the findings of these studies do not generalize well across the broader population, with outcomes potentially compromised for individuals not represented in the research. As a result, AI algorithms trained on this data are

* bit.ly/2HFetvY

Photo by Marius Masalar

biased and potentially dangerous to those populations that have been excluded. Only within the last few years have funders of medical research begun to require gender parity and further inclusiveness.

Datasets of social phenomena can potentially reflect social biases, and the algorithms trained on them then learn and reproduce those patterns of bias.

For example:

Ad targeting

Researchers have found that women are less likely to receive targeted ads related to high-paying jobs,[48] and online searches of names stereotypically associated with black individuals are 25% more likely to return ads related to criminal and arrest records.[49]

Predictive policing

A computer program used in the U.S. to assess the risk of re-offense by individuals in the criminal justice system incorrectly flagged black defendants as high risk nearly twice as often as white defendants.[50]

Further examples of bias can be found in such areas as pre-selection for job interviews, teacher performance ratings, and loan and mortgage approvals.[51]

In some instances, algorithmic bias can also emerge from the interaction of design choices with biased inputs. A good example of this is Microsoft's infamous Tay chatbot, an autonomous bot that was pulled from Twitter after just 16 hours when it began spewing racist and misogynistic tweets.[52] On one level, it is reasonable to place some blame on the data, as Twitter

Photo by Michal Parzuchowski on Unsplash

Photo by Ryoji Iwata on Unsplash

users were intentionally trying to manipulate the chatbot. However, the fact that the underlying algorithms were neutral to the content and meaning of the tweets taken in as data meant that Tay effectively treated racist or misogynistic expressions simply as further data points to process.

Another example illustrates the precariousness of content-neutral learning algorithms. Using data derived from the content shared by and the online activities of Facebook users, Facebook algorithms automatically create user categories to help advertisers target their ads. Categories based on harshly anti-Semitic phrases such as "Jew hater" enabled advertisers to reach roughly 2,300 users identified as having interest in these topics.[53]

As these examples illustrate, the choice of a "neutral" learning algorithm can have value-laden consequences. In the Tay-bot case, the algorithm was unable to determine whether an utterance was positive or negative from a social values perspective. The designers of such systems regularly make unintentionally value-laden choices that inform how learning algorithms subsequently perform.

Clearly, bias can easily be introduced into and absorbed by AI systems. This speaks to the need for safeguards at the design and operational stages of those systems to ensure that outcomes are fair and unbiased. Additionally, transposing a trained learning algorithm from one context to another could have undesired outcomes: the learned processes of a system trained on North American data, for instance, may produce inappropriate results if deployed in other contexts where the training data is not sufficiently representative. As such, the lack of country-specific datasets is a clear hurdle to generating fair AI systems in the Global South.

AI blackboxes: transparency and accountability

Another barrier to developing fair AI applications (ML algorithms in particular) is that, unlike other software, the choices and actions these programs take can be difficult or even impossible to explain.[54] Looking into the workings of an ML algorithm and identifying encoded bias is not a straightforward activity. The complexity of abstract representations learned through the encoding of patterns into potentially millions of parameters makes it highly challenging for humans to analyze *why* a ML algorithm behaves the way it does. ML has a transparency problem.[55]

Given that AI systems can make decisions that can be considered unfair, inaccurate, or unethical, the difficulty of determining how those decisions are made in particular circumstances raises several questions: By what criteria will decisions be deemed inappropriate? How will liability be determined for the consequences of these decisions? How can those harmed seek redress?

These issues have yet to be addressed in any material way and must be tackled both in policy and in law. The Hundred Year Study report (Stone et al. 2016) frames the question of civil and criminal liability this way:

> *As AI is organized to directly affect the world, even physically, liability for harms caused by AI will increase in salience. The prospect that AI will behave in ways designers do not expect challenges the prevailing assumption within tort law that courts only compensate for foreseeable injuries. Courts might arbitrarily assign liability to a human actor even when liability is better located elsewhere for reasons of fairness or efficiency. Alternatively, courts could refuse to find liability because the defendant before the court did not, and could not, foresee the harm that the AI caused. Liability would then fall by default on the blameless victim. The role of product liability — and the responsibility that falls to companies manufacturing these products — will likely grow when human actors become less responsible for the actions of a machine. (...)*
>
> *As AI applications engage in behaviour that, were it done by a human, would constitute a crime, courts and other legal actors will have to puzzle through whom to hold accountable and on what theory.*

Calo (2017: 12) sums up the broader issue this way:

> *The end game of designing systems that reflect justice and equity will involve very considerable, interdisciplinary efforts and is likely to prove a defining policy issue of our time.*

In order to design fair and equitable AI systems, one approach is to make the inner workings of the algorithms more transparent. Emerging approaches for achieving greater transparency include:

Algorithm auditing

This is a nascent field in which researchers inspect and evaluate algorithms.[56] For ML algorithms, this might involve visualizing what the software has learned.[57]

Explainability

Designers can make at least some of the reasoning behind algorithmic decisions more transparent.[58] When possible, AI systems can be developed with the capacity to provide the reasoning behind a decision.[59]

Legal recourse

Making the software behind critical AI algorithms open source would allow for access that might not be available under protective patents and licensing agreements. This happened recently when a U.S. federal judge ordered a crime lab to make public its software for analyzing DNA evidence so that the code could be analyzed.[60]

Some jurisdictions are already taking concrete steps to legally mandate transparency in the decision-making of automated systems. For example, the European Union now requires, by law, that machine-made decisions be explainable. Penalties for lack of compliance could cost companies billions.[61] However, the language of the mandate is arguably insufficiently precise, rendering it “toothless” (Wachter et al. 2017). This will be an interesting test case for other countries around the world.

Of course, it is not clear that the EU approach is feasible, given the sometimes extreme complexity of ML algorithms. Another approach may be to govern AI decision-making in terms of the optimization of the algorithm rather than the outputs themselves. For example, David Weinberger lays out the following approach: “Self-driving cars might be optimized first for reducing fatalities, then for reducing injuries, then for reducing their environmental impact, then for reducing drive time, and so forth. The exact hierarchies of priorities is something regulators will have to grapple with.”[62]

Surveillance and privacy are two sides of the same coin

SURVEILLANCE AND LOSS OF PRIVACY

Surveillance and privacy are two sides of the same coin. Surveillance is defined as “the processing of personal data for the purposes of care or control, to influence or manage persons and populations” (Lyon 2010: 108), while privacy involves maintaining control over one’s personal information. Surveillance can be conducted for relatively benign purposes, such as discovering personal preferences in order to target online ads and promotions or improve government services. It can also have more sinister purposes, such as tracking the whereabouts and associates of political rivals — purposes that are particularly concerning today, given the rise of state-sponsored surveillance (Deibert 2013) and the potential associated chilling effects on individuals’ freedoms (Penney 2017). Whatever the purpose, the practice is invasive, and individuals have less and less capacity to avoid or prevent it.

From a privacy perspective, AI applications are concerning because powerful pattern recognition algorithms can extract personal details from available data (Calo 2017; Privacy International 2017). The increasing use of surveillance has resulted in vast repositories of raw information about the activities of individuals both online and in the physical world,[63] and AI techniques simplify the process of sorting through the endless data

AI and privacy

In a 2017 submission of evidence to the U.K. House of Lords Select Committee on Artificial Intelligence, Privacy International identified four purposes of AI applications that are particularly concerning in terms of individual privacy:

Identify and track individuals;

Predict or evaluate individuals or groups and their behaviour;

Automatically make or feed into consequential decisions about people or their environment; and

Generate, collect, and share data.

For more, see: privacyinternational.org/node/1525

Photo by Serge Kutuzov on Unsplash

AI techniques boost already significant surveillance capacities

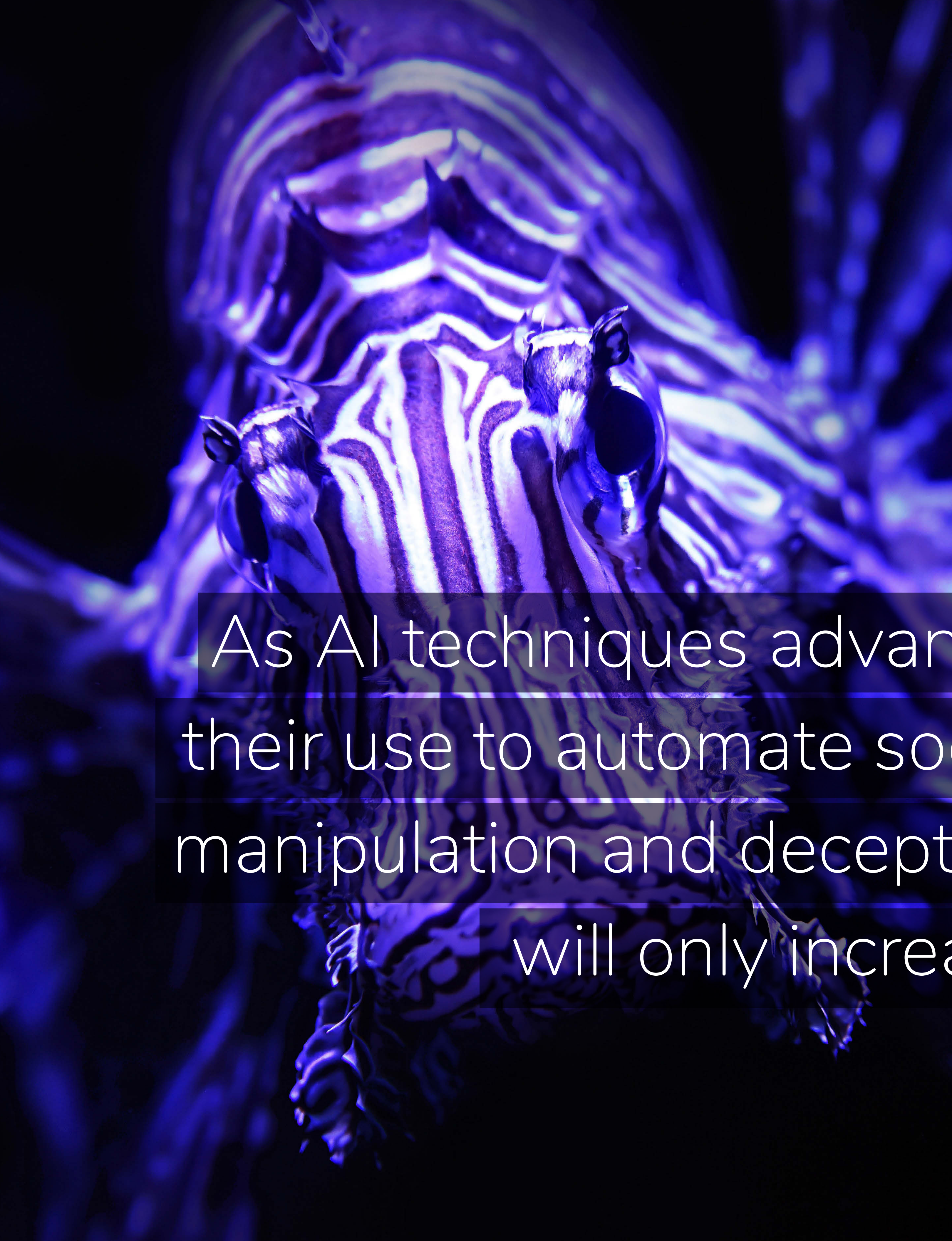
points. Datasets can be mined and cross-referenced, subtle patterns identified, inferences drawn, and conclusions reached, all without active human intervention. AI algorithms supercharge surveillance by processing data faster than previously possible and detecting patterns too subtle for human analysts to uncover.

The result is the erosion of privacy, a concept already considered quaint in some tech circles. This erosion is particularly concerning from a social and political development perspective, as privacy is "the lynchpin of indispensable individual values such as human dignity, personal autonomy, freedom of expression, freedom of association, and freedom of choice" (Privacy International 2017).

Examples of how AI applications can invade individuals' privacy are plentiful. In one case, an analysis of an individual's Facebook friends correctly identified that person's sexual orientation.[64] More recently, analysis of Facebook "likes" using ML identified an individual's race, sexual preference, and political affiliation with a high degree of accuracy.[65]

Governments employ AI techniques to boost their often already significant surveillance capacities. For example, facial recognition software gives CCTV systems the capacity to track individuals as they move through the urban landscape.[66] Facial recognition algorithms can recognize even blurred or pixelated faces, thus defeating some current privacy protection technologies[67]. Similarly, facial recognition algorithms have successfully identified people who have attempted to disguise their identity by wearing a cap, scarf, or glasses (Singh et al. 2017). China recently used this technology to apprehend 25 suspects at a beer festival, including one man who had been on the run from authorities for 10 years.[68] The use of these systems to track criminals or terrorists provides a societal good, but where safeguards against such practices are weak or nonexistent, the same tools can easily be turned on political rivals, business competitors, or others. Authoritarian regimes in LMICs have already invested in powerful surveillance technologies and turned them on dissidents and civil rights activists,[69] among others. We expect this trend to continue, using ever more powerful tools.

The current lack of transparency as to what personal data are being collected from users online and how those data are being used — typically without any form of consent — makes it difficult for individuals to protect their online privacy. As the internet of things becomes a reality, with everyday devices (cars, refrigerators, etc.) connecting digitally, whole new vistas of data will become available for perusal. Issues of control over one's own data and the opportunity for redress in cases of misuse or inappropriate outcomes will become more urgent as AI applications make consequential decisions based on those data.[70]



> As AI techniques advance, their use to automate social manipulation and deception will only increase

UNDERMINING DEMOCRACY AND POLITICAL SELF-DETERMINATION

The democratic process is predicated on a citizen's right to be freely informed and make political choices. The risks of increased surveillance and reduced privacy outlined previously threaten to undermine these fundamental democratic principles. There are at least two other mechanisms through which AI can potentially contribute to undermining democracy: targeted propaganda and manipulation, and social fragmentation.

While these are not new phenomena, the confluence of AI algorithms with other relatively recent changes to the online ecosystem makes misinformation and manipulation much more effective on a large scale than ever before. First is the centralization of attention and activity funnelled through the online platforms of just a few companies, such as Google, Twitter, and Facebook.[71] The situation is similar in China, where companies like Tencent dominate citizens' online experience. According to a 2017 Pew study, almost half of Americans "sometimes" or "often" get their news from social media.[72] These platforms collect large amounts of user data from which AI algorithms extract personal information. The platforms then sell the ability for external actors to reach highly specific categories or types of users with their messages.

Furthermore, recent advances in the study of behavioural manipulation have increased the impact of campaigns designed to influence social media users. The result is a proliferation of highly personalized and effective misinformation and manipulation campaigns (precision propaganda)[73] that represent a real threat to political self-determination.[74]

We can anticipate that as AI techniques advance, their use to automate social manipulation and deception will only increase in sophistication, expanding the range of potential threats (Brundage et al. 2018).

Photo by David Clode on Unsplash

Propaganda and manipulation

AI techniques underpin the emergence of "computational propaganda" (Bolsover and Howard 2017), which can take many forms, including automated chatbots, fake news, and highly targeted misinformation and manipulation. As Howard states, "Algorithms and fake news go hand in hand."[75]

The 2016 U.S. presidential election has become a notorious example of the use of misinformation, whereby AI-enabled Twitter chatbots and fake news were on full display. It is estimated that 33% of pro-Trump tweets and 20% of pro-Clinton tweets during the first presidential debate were created by AI bots,[76] artificially inflating the volume of traffic and appearance of support. An abundance of highly targeted misinformation was also distributed in the form of fake news articles that further influenced people's opinions. Facebook estimated that between January 2015 and August 2017, Russian misinformation may have reached as many as 146 million users.[77] Similar misinformation and propaganda initiatives have been used in attempts to influence other political campaigns.

These tactics can also further other agendas. In Myanmar, for example, it can be argued that Facebook was an essential vector for the transmission of anti-Rohingya sentiment, contributing to the ethnic cleansing there.[78] Similarly, arguably, fake news on social media has contributed to an increase in the number of people in the U.S. who question whether climate change is due to human intervention, and whether tobacco and cancer are linked.[79]

Advances in AI will only make misinformation more compelling. Just as image-editing software makes it relatively simple to create convincing fake images, new AI-enabled software brings this ease of editing to video and audio, potentially turbocharging fake news and misinformation campaigns.[80] For example, the Montreal-based start-up Lyrebird recently released an audio clip on YouTube that apparently features the voices of Barack Obama and Donald Trump discussing the company.[81] The audio clip, entirely fabricated, uses computerized voices to mimic the speaking style of the two men, but a naïve listener might mistake the clip for a recording of an actual conversation processed to sound computerized. With this software, the company has opened the door to the possibility of producing fake video clips using synthesized voices, body language, emotional tone, and so on. In a similar vein, Adobe has developed a tool that lets users edit recorded speech in the same way you edit text. It even allows the insertion of words that weren't spoken.[82] (See also Suwajanakorn et al. 2017.)[83]

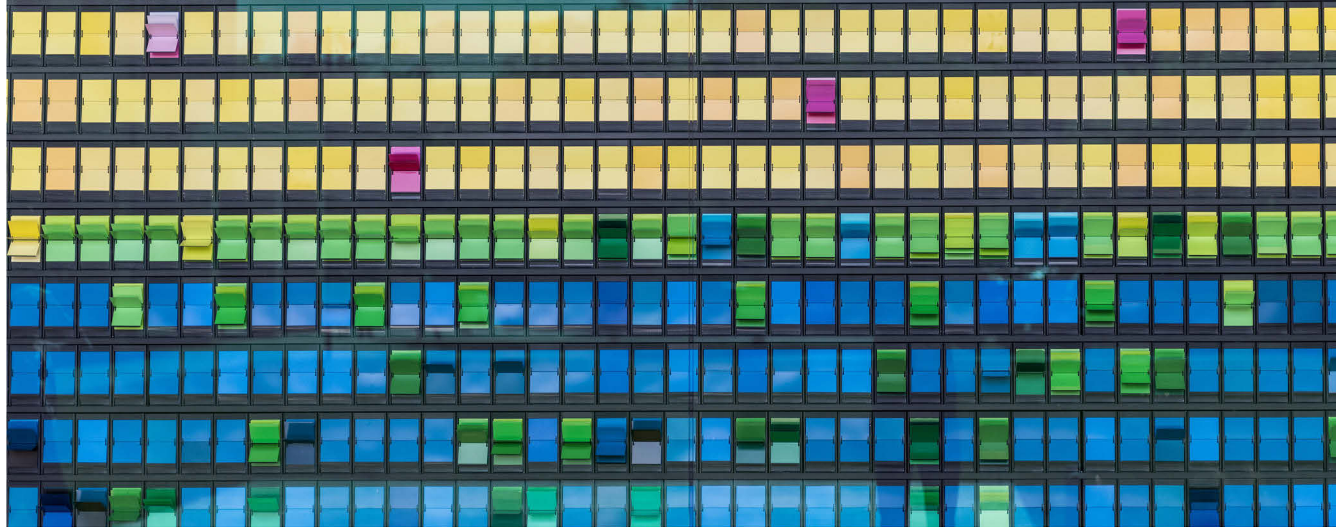


Photo by Al x on Unsplash

Perhaps even more worrisome, there are now freely accessible computer desktop tools to make 'deepfakes'. Using the power of deep learning, this tool allows anyone to seamlessly swap one face for another in a video. The only requirement is a sufficient number and diversity of videos of the person you wish to swap into the video. Given the power of fake stories to move more rapidly amongst people and the susceptibility of humans to forming false memories, particularly those that confirm our biases, the scope for abuse is enormous.[84]

Manipulation of citizens can be achieved not only through propaganda and misinformation, but through the automated consolidation and analysis of the extensive data collected daily on individuals. China's proposed Citizen Score is an extreme but real-world example. Derived from factors such as a person's financial history, online activity, political affiliations, and even the behaviour of the person's friends and acquaintances, this single score will determine that individual's access to loans, housing, jobs, and travel visas. If fully implemented in 2020 as planned, it would be a mechanism for manipulating and controlling the behaviour of an entire populace of billions of people.[85]

Fragmentation

Taken together, efforts to distort and manipulate public opinion, along with the dominance in many regions of one or a few search engines and social media platforms, can limit public discourse. The customization and personalization used to enhance marketing success can also narrow the range of news and opinion users encounter, resulting in an "echo chamber" effect. Political and social views become self-reinforcing, with differing groups having less interaction and common ground. Debate becomes polarized, and collaboration and compromise become difficult. This trend is apparent in U.S. politics. politics. Over time, this process can lead to "fragmentation, possibly even a disintegration, of society."[86]

AUTOMATION, JOB LOSS, AND TAX REVENUE LOSS

Automation is not a new phenomenon: it has contributed to the overall loss of jobs in the manufacturing sector since the 1990s. However, with the growing use of ML and AI systems in nearly every sector of the economy, automation will no longer be limited to manufacturing. Many jobs and careers traditionally viewed as requiring human involvement and expertise will be affected. For example, a competent deep learning algorithm can now detect cancer as effectively as a trained dermatologist can.[87] The upside is the benefit to low-resource communities that lack doctors or trained technicians. The downside is the real threat such automation poses to jobs that are considered highly skilled.

Many of these jobs can be partially or entirely automated, reducing the need for human workers. A combination of deep learning algorithms and natural language processing can potentially do much or all of what tutors, travel agents, tax preparers, health coaches, legal researchers, financial advisors, office assistants, and chauffeurs do. The already low ratio of employment in many countries in the Global South can be affected further by job loss attributable to AI.

Estimates of the extent to which AI-driven automation will impact employment vary widely. An Oxford University study indicated that 47% of jobs could be automated in 10 to 20 years (Frey and Osborne 2013), while the Organisation for Economic Co-operation and Development (OECD) arrived at 9%, based on

Photo by Giovanni Randisi

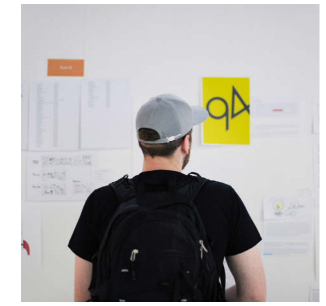
the fact that a substantial number of jobs cannot be completely automated (Arntz et al. 2016). The McKinsey Global Institute projected that "the adaptation of currently demonstrated automation technologies could affect 50% of working hours in the global economy" (Manyika et al. 2017).

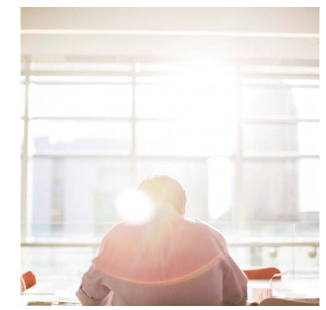
Such seismic change could greatly disadvantage less technologically sophisticated countries. Another report estimated that automation will put 55% of jobs at risk in Uzbekistan and up to 85% in Ethiopia, with emerging economies such as China and India seeing 70% or more of jobs at high risk.[88] In some industries, the potential risk is seen as relatively imminent. In July 2017, fearing massive job losses, the Indian transport ministry banned driverless cars on the country's roads.[89]

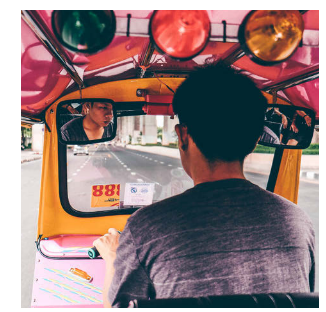
The impact of AI on employment will vary by economic sector, country, and global region, but regardless of specific numbers, many believe that the expansion of AI-based automation will result in widespread disruption of labour markets and a major shift in the very nature of work. For example, the McKinsey report found that with current technologies, "60% of all occupations have at least 30 percent of activities that are technically automatable" (Manyika et al. 2017: 32). Furthermore, they calculate that "the adaptation of currently demonstrated automation technologies could affect 50 percent of the world economy, or 1.2 billion employees and $14.6 trillion in wages" (ibid).

This poses challenges to models of economic development: "While manufacturing productivity has traditionally enabled developing countries to close the gap with richer countries, automation is likely to impact negatively on their ability to do this, and new growth models will be required."[90] We are already seeing instances of companies "reshoring" manufacturing to Europe and North America because the increased productivity and reduced need for labour resulting from automation offset the shipping costs and time lags of transporting goods from factories in Asia.

Additionally, the timeline for looming job losses could be much shorter than seen with previous technological change. An industry report from 2016 projected that 6% of all jobs in the United States could be automated by 2021, with customer service representatives and call centre employees among early casualties.[91] Jack Ma, the founder of Alibaba and a noted innovator in online retail, has said that AI-driven automation will be responsible for "more pain than happiness" in coming decades.[92] As job losses climb the skills ladder, experts foresee an accelerated decline of the middle class.[93] Furthermore, it is







Photos by:
rawpixel.com,
Bethany Legg,
Hanny Naibaho
on Unsplash

anticipated that the burden of expected job losses will fall more heavily on women (World Economic Forum 2016).

The result could well be social instability and upheaval, with Global South countries possibly being the hardest hit. This is in part because many of these countries look to future industrialization as a way to achieve economic growth, and other emerging economies have relied largely on manufacturing for the advances they have made. As already discussed, increased automation is poised to accelerate job losses in this sector. While the future may not be a "deindustrialized" one, it is possible that industrialization will not be the path to success that it once was. As well, populations of unemployed or underemployed young adults can be volatile, as demonstrated by the Arab Spring phenomenon.[94]

Because they lack the social safety nets of most high-income countries, many Global South countries are also less able to soften the blow of substantial job losses.[95] Contracting employment will add further strain due to increased demand for government support and a shrinking income tax base. These outcomes combine to generate even greater risks, and could result in instability within the Global South. The current turmoil in American politics could be a harbinger of things to come.[96]

It is important to note that the foregoing discussion addresses a potential risk, and that its pessimistic outlook is based on estimates and projections without much supporting evidence.[97] Indeed, there is a history of predicting massive job loss due to automation. For example, the introduction of ATMs in the 1970s, it was argued, would eliminate the need for bank tellers. Similarly speculative counter-arguments have been made with regard to AI. For example, rather than causing a net job loss, AI might spur a shift in the nature and scope of work and jobs, for instance through robots complementing and augmenting labour, and an increased focus on higher-skilled and higher-paid tasks (International Federation of Robotics 2017). Overall, while we can expect AI to contribute to social change, the ultimate nature of that change is unknown.

Cyber security is the new arms race

CYBER SECURITY AND CYBER CRIME

Cyber security is a new kind of arms race. Ill-intentioned actors (from criminals to state-sponsored hackers) continually develop and launch new forms of malware and methods of attack to access computer systems in order to steal data, identities, or electronic funds, or to hold those systems and their data ransom. Cyber security professionals, in turn, develop new methods of defence, detection, and encryption to keep computers and data safe. AI won't change this overall scenario; it will be employed by both sides, as a way to improve cyberattacks and as a means to combat them.[98]

AI will increase the sophistication of the ill-intentioned tools available and make criminals faster and more efficient. AI-enhanced malware will be able to break through security roadblocks faster than a human could, and other AI tools will be able to automatically sift through stolen databases of personal information millions of records deep to facilitate identity theft. On the other side, AI-enhanced security software will be better at detection, be more responsive, and learn from previous attacks. Google is currently developing a ML malware detector for its Android operating system.[99]

What is likely is that the advanced capabilities of cyber criminals wielding AI tools will put individuals, organizations, and governments without access to up-to-date security measures at ever greater risk of successful attack. Furthermore, over time these tools will probably lower the costs of engaging in cyberattacks at scale, most likely leading to an expansion of both the set of actors engaging in these attacks and the set of targets (Brundage et al. 2018).

Photo by Patryk Gradys on Unsplash

These predictions add up to a rather dire scenario of increased inequality alongside economic disruption, social unrest

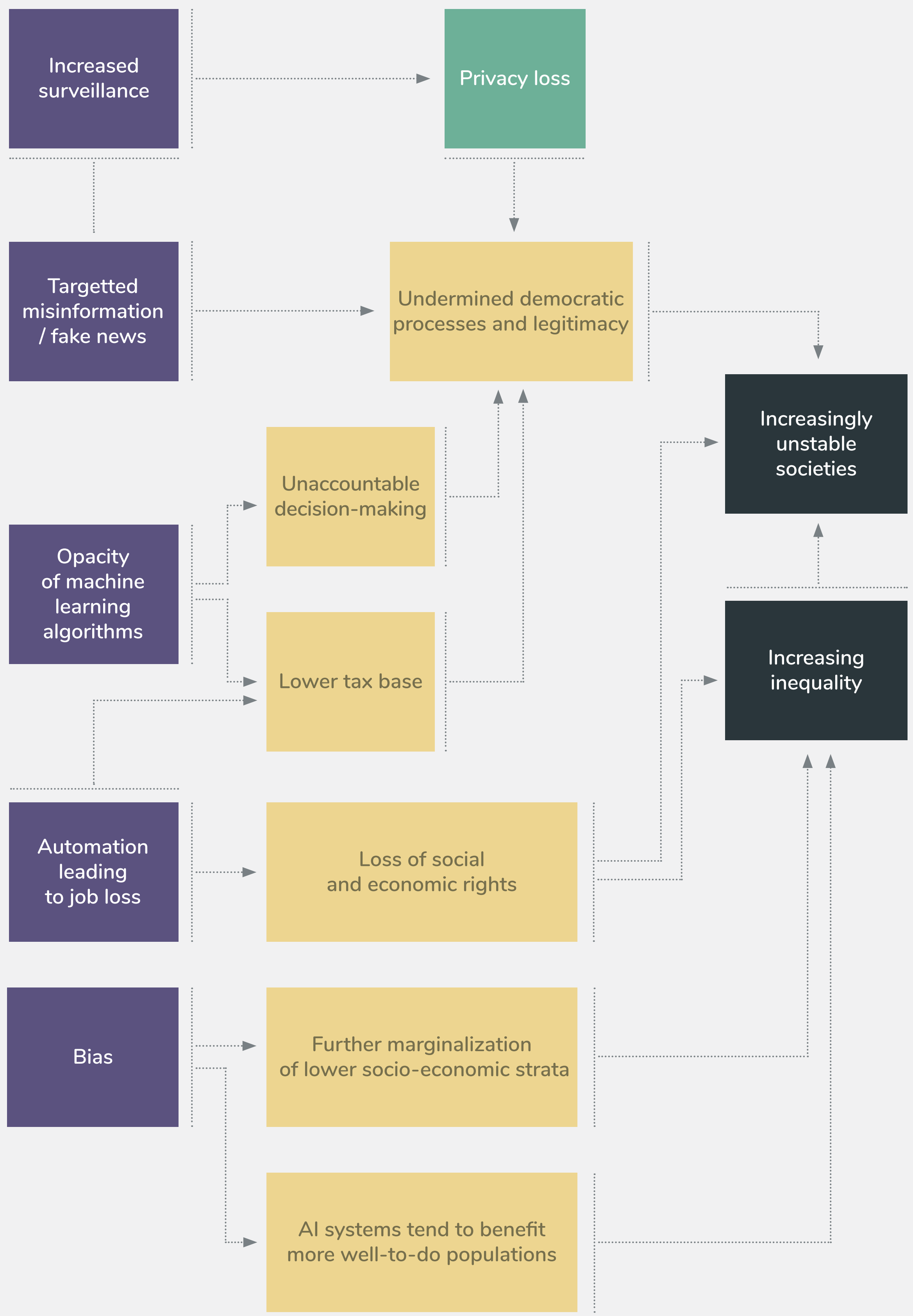
EXTRAPOLATING INTO THE FUTURE: AI RISKS IN THE GLOBAL SOUTH

Despite the potential of AI to help tackle pressing development challenges, the associated risks (discussed previously) and the lessons from over 20 years of research on digital technologies in the Global South provide ample cause for concern regarding the application of AI. Every few years, a new technology comes along that is touted as the "next big thing" in development: computers, the internet, mobile phones, big data, and blockchain, to name a few. Extrapolating from past experience,[100] we can imagine a future for AI in development that plays out much as other technologies have:

- AI applications will bring dramatic social benefits, particularly advances in healthcare, education, and economic efficiencies. However, **these benefits likely will be distributed unequally.**
- **The AI divide will remain a hurdle** to more inclusive AI development and deployment for some time to come, limiting benefits to typically underserved or marginalized communities.
- Some AI applications will be built with unacknowledged or unrecognized **biases** that will **reproduce and aggravate social marginalization.**
- AI will result **in increased surveillance and loss of privacy** due to the activities of both public- and private-sector actors. This will disproportionately impact marginalized and economically disadvantaged populations.
- **Ill-intentioned actors will employ AI techniques with increasing sophistication to foster crime, social discord, and political unrest.**

Photo by The Roaming Platypus

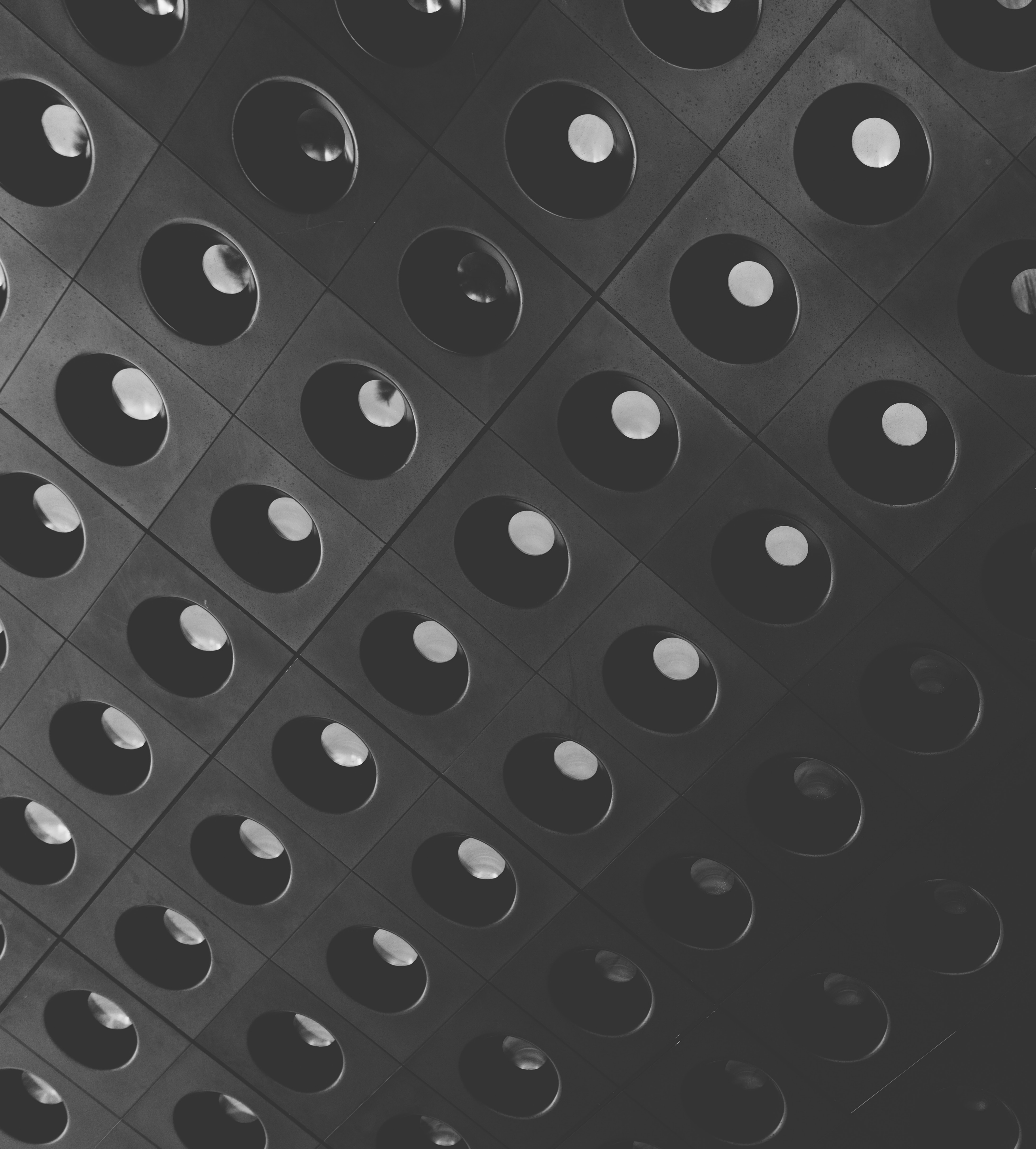
FIGURE 3 Risks from AI and how they may contribute to negative social impacts



On the whole, these predictions add up to a rather dire scenario of **increased inequality alongside economic disruption, social unrest, and, in some cases, political instability** (see Figure 3) unless action is taken. Indeed, it is the dual nature of AI — whereby some of its direct benefits (e.g., economic efficiencies) can simultaneously undermine advances in other areas (e.g., employment and social stability) — that makes it such a disruptive technology. A somewhat dystopian thesis would be based on two assumptions: (1) the foundational nature of AI and its impending integration with other technologies and devices make the potential harms of AI likely to be felt much more broadly and rapidly than we have seen in the past; and (2) the current context in Global South countries makes the harms more likely to occur.

Fortunately, likelihood is not the same as inevitability. If we act to address the challenges posed by AI, we can avoid or mitigate these all-too-predictable adverse outcomes while enabling countries of the Global South to take full advantage of their positive potential. The AI future is already unfolding. The question is whether or not we will be ready.

Photo by Ben Neale on Unsplash



Conclusions and recommendations

Although this white paper appears to take a pessimistic stance on the future of AI in the Global South (or AI for development), it is not intended to be pessimistic or optimistic: rather, it is cautionary and hopeful. There is little doubt that AI technologies will be transformational. Breathtaking advances will be made, extraordinary wealth will be created, and many of our social and institutional structures will be transformed. However, we must ask: *whose* lives will be improved by these technologies? Whose political and economic freedoms will be advanced? For whom will systematic deprivation of these freedoms be alleviated? The conclusion of this paper is that, if we continue blindly forward, we should expect to see increased inequality alongside economic disruption, social unrest, and, in some cases, political instability, with the technologically disadvantaged and underrepresented faring the worst.

Photo by Robert Haverly on Unsplash

This gloomy prediction stems from the interweaving of two elements: the nature of AI applications, and projections of the impacts of AI applications in the current global context. What is worrisome is the dynamic of how our current set of institutions and cultures shape the evolution of technologies, and how, in turn, these technologies shape these institutions and cultures.

As discussed, AI refers to a broad class of various technologies that, for the most part, work in the background. In this way, it is similar to the internet and other foundational technologies that cut across and impact our social, political, and economic lives. The key difference is that AI leverages existing infrastructure (e.g., the internet, large datasets, the ability to draw from increasing numbers of digital sensors and data sources) to dramatically reduce the costs of activities (both new and old, good and bad) on a large scale: for example, personalized learning or health diagnoses, automated care work, automated translation, intelligent chatbots, optimized supply chains, more sophisticated cybercrime, targeted advertisements, and facial recognition surveillance. These powerful new AI technologies are being developed and rolled out in a context of stark inequalities and highly consolidated nodes of power. Why, then, would we expect the results — on the aggregate — to be ethical or equitable? Add to this the vastly increased privacy risks, surveillance potential, and opportunities for misuse by bad actors, and it's easy to be pessimistic.

That being said, we would like to conclude on a more hopeful and constructive note. The argument presented above assumes that we continue on the same trajectory; that societies will be unable to adapt and appropriately regulate AI applications, protect privacy, and rein in abuses; and that AI will not be developed in a more inclusive manner focused on solving pressing social, economic, and environmental issues. However, the past is often a poor predictor of the future, and there are indications of growing concern and interest in constructing a more positive overall outcome.

One indicator is how the emergence of AI in the global discourse seems to have sparked a wider-ranging public discussion about the kinds of societies we want to build going forward. In many ways, it is AI's very potential to transform our economic, political, and social institutions that forces this introspection as it draws attention to aspects of our societies (such as inherent social biases) that were previously ignored or overlooked. It is only when these elements are revealed that we can explore innovative solutions, be they technical, cultural, economic, or political. Furthermore, the imperative to ensure inclusive participation in AI development and deployment and to engage in inclusive design of AI applications, while working hard to mitigate potential risks, offers an opportunity to shift the discourse and practices with respect to the use of technologies within social change projects.

There are many examples:

- The encoding of social biases in AI algorithms, and their subsequent exposure, can open up a space for public discussion about these biases. Indeed, examples of bias in ML algorithms (along with their opacity) have received lots of press. However, this may prove to be a relatively minor critique, and may even be a source for more positive social change. First, this conversation has spurred developers to take steps to avoid gender stereotypes in their AI applications[101] and the new area of research focused on detecting and removing bias in ML algorithms.[102] Removing bias and opacity may be mostly a technical fix. Second, uncovering these social biases opens up a unique possibility for a broader social discussion on the biases themselves, their history and nature, how they might be institutionalized, and what might be done to address them. AI may encode social biases, but it can also be used as a mirror to reflect them back to society.
- The growing recognition of the potential for internet platforms, such as Facebook and Google, to exert significant political and social influence has sparked talk of anti-trust measures and regulatory oversight. Open conversations have also arisen about how the large platforms themselves might address these concerns, with some beginning to take action.[103]
- Predictions of extensive job loss due to AI-enabled automation are driving new policy innovation and experimentation in social protection, taxation, and education.
- Governments are also beginning to experiment with regulation. For example, as mentioned above, the EU is including language around transparency of ML algorithms in its General Data Protection Regulation. In France, President Emmanuel Macron is proposing to introduce a law to ban fake news.[104]

While no one can predict how effective these actions will be, what is clear is that the status quo is unlikely to be maintained.

While there are emerging efforts to confront the challenges posed by AI, much remains to be done, particularly with regard to AI in the Global South, where AI development and policy capacities and resources are comparatively thin, and the potential benefits and risks of AI are magnified. We need research and evidence to inform and shape the development and use of AI applications, and government policy and regulation to ensure fair and appropriate use. We need greater capacity in the Global South to drive this research agenda, develop locally appropriate solutions, and develop policy and regulation that are effective in their local institutional contexts.

Building Southern capacity is just a start; this must be a global effort. The transnational nature of the internet, online platforms, flows of data, and so on, make the ethical and equitable application of AI a collective challenge that no one country can tackle on its own.

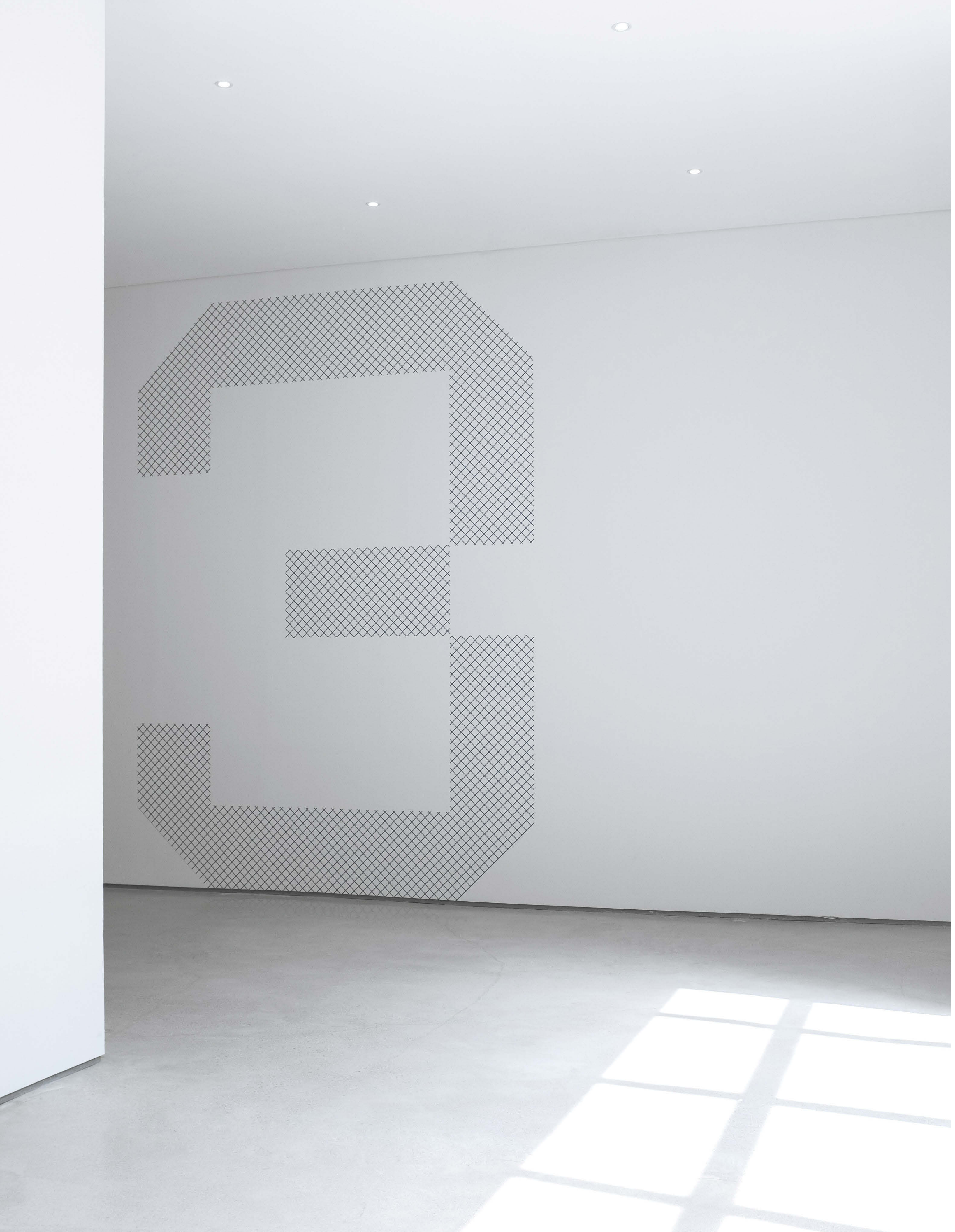
Photo by Matus Kovacovsky on Unsplash

We need to develop global and local values and principles for AI that prioritize inclusion, ethics, and transparency. We need interdisciplinary, international collaborations through which AI researchers focus on real-world problems. And it is critical to expand the self-determination of communities so that they can actively drive discussions around the development and deployment of AI applications that affect their lives.

One potential starting point would be the allocation of significant resources through a global AI for Development fund focused on building the capacities of Southern "AI centres of excellence." These centres would engage in research and provide regional- and national-level, empirically informed policy and regulation advice to governments. They would work not only locally, engaging relevant and affected communities, but would collaborate with research and policy centres around the world to tackle shared challenges and contribute a greater diversity of perspectives and voices at the global level.

The future is unknown, but clearly the time to act is now, before AI dramatically disrupts societies in the Global South. In the next and final section, this paper puts forth a series of recommendations for action and research designed to contribute to the ethical and egalitarian application of AI for development.

The AI future is already unfolding.
The question is whether or not we will be ready.



RECOMMENDATIONS

Based on the conclusions of this paper and the broader literature, we have identified three key areas in which action can be taken: policies and regulations, inclusive and ethical AI applications, and infrastructure and skills. Within each area, we make a series of recommendations for research necessary to make concrete progress.[105] Note that this is not intended to be a comprehensive list, rather, it focuses on the most pressing interventions.

The research needs to be:

- Interdisciplinary: Many of the research questions address the intersection of social and technical factors. Approaching them effectively will require multidisciplinary collaborations across the social (including economic and political), humanistic, and computational sciences.
- Locally conducted: To ensure relevance and usefulness, the research needs to be driven and conducted by researchers in the Global South.
- Designed to support practice and policy: To have real impact, research must be rigorous and generate actionable findings that facilitate the implementation of programs and policy.

We break up the recommendations into the three levels of an AI ecosystem: (1) policy and regulatory structures; (2) applications; and (3) infrastructure and skills. Note that these three layers are interconnected and overlapping, and thus are provided as a useful heuristic rather than discrete categories.

Policy and regulatory layer

Application layer

Infrastructure and skills

Photo by Tobias van Schneider

POLICY AND REGULATORY STRUCTURES

Foster the design of policies and regulations that enable inclusive and rights-based AI

Conduct baseline research on the prevalence of AI applications and policies in the Global South. Despite pockets of AI activity in the Global South, there are no systematic overviews of the level of this activity. Baseline data collection should include the sets of AI policies, regulations, applications, existing open datasets, and skill levels. This research should be conducted on a yearly or, minimally, a biennial basis to support continued activities, policy development, and the research agenda.

Learn about effective regulatory models. Document and assess AI regulatory models developed to deal with the emergence of new AI-driven activities such as predictive policing, autonomous vehicles, and chatbots. Determine whether the potential risks of AI applications are adequately addressed by existing regulation, or if existing regulation needs to be adapted or new regulation developed. Identify regulatory responses to specific AI use cases that are appropriate for settings with low institutional capacity. While lessons learned in the Global North are useful, it is critical not to directly import institutional and regulatory approaches into the Global South where instiutional and cultural contexts differ.

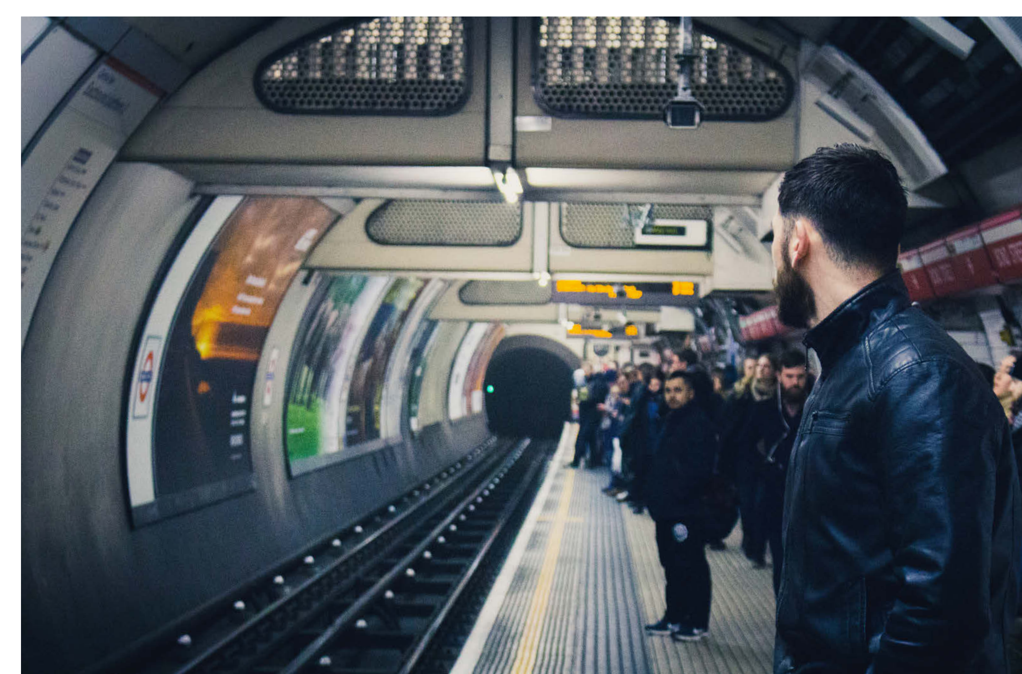
Photo by Richard Loader

Track the impact of AI on employment and work. Conduct social and economic policy research to understand the effects of AI on employment, the nature of work, and labour markets. To what extent is AI-enabled automation altering employment patterns and transforming the workplace? What are alternative models of income and resource distribution, education, and job retraining in different contexts?

Explore approaches to addressing liability, accountability, and redress for AI decision-making. Design regulatory systems and frameworks to determine liability and accountability for AI decision-making that is erroneous, biased, or discriminatory, and establish mechanisms for redress. Measures may include policies that stipulate transparency for automated decision-making, evaluative procedures to determine the competency of AI systems, and certification of AI systems that engage in tasks requiring a degree of skill or training. The need for action is particularly urgent in the case of decision-making systems that affect people's well-being or freedom, such as those that involve the use of force or incarceration. Research is critical here to uncover and document which systems for accountability and redress are effective and in what contexts.

Study the impact of AI on human rights.[106] At a broad level, the UN recognizes that offline rights apply online, testifying to the relevance of analogue rights in digitally mediated environments. Professional bodies specifically call for full consideration of human rights in the context of AI design and operation.[107] Tailoring impact assessments to the risks of AI would help encourage development programs to incorporate AI technology in ways that respect and promote human rights, including privacy, equality, and freedom of expression.

Photo by Hello Lightbulb on Unsplash



APPLICATIONS

Catalyze the development of inclusive and ethical AI applications

Support the development and deployment of innovative AI applications for social good. Invest in developing, deploying, and using applications for education, health, the environment, food security, etc., and in ensuring that these applications are ethical and inclusive. As with regulatory innovations, while it is important to draw inspiration from examples around the world, AI applications will often require homegrown solutions to be effective.

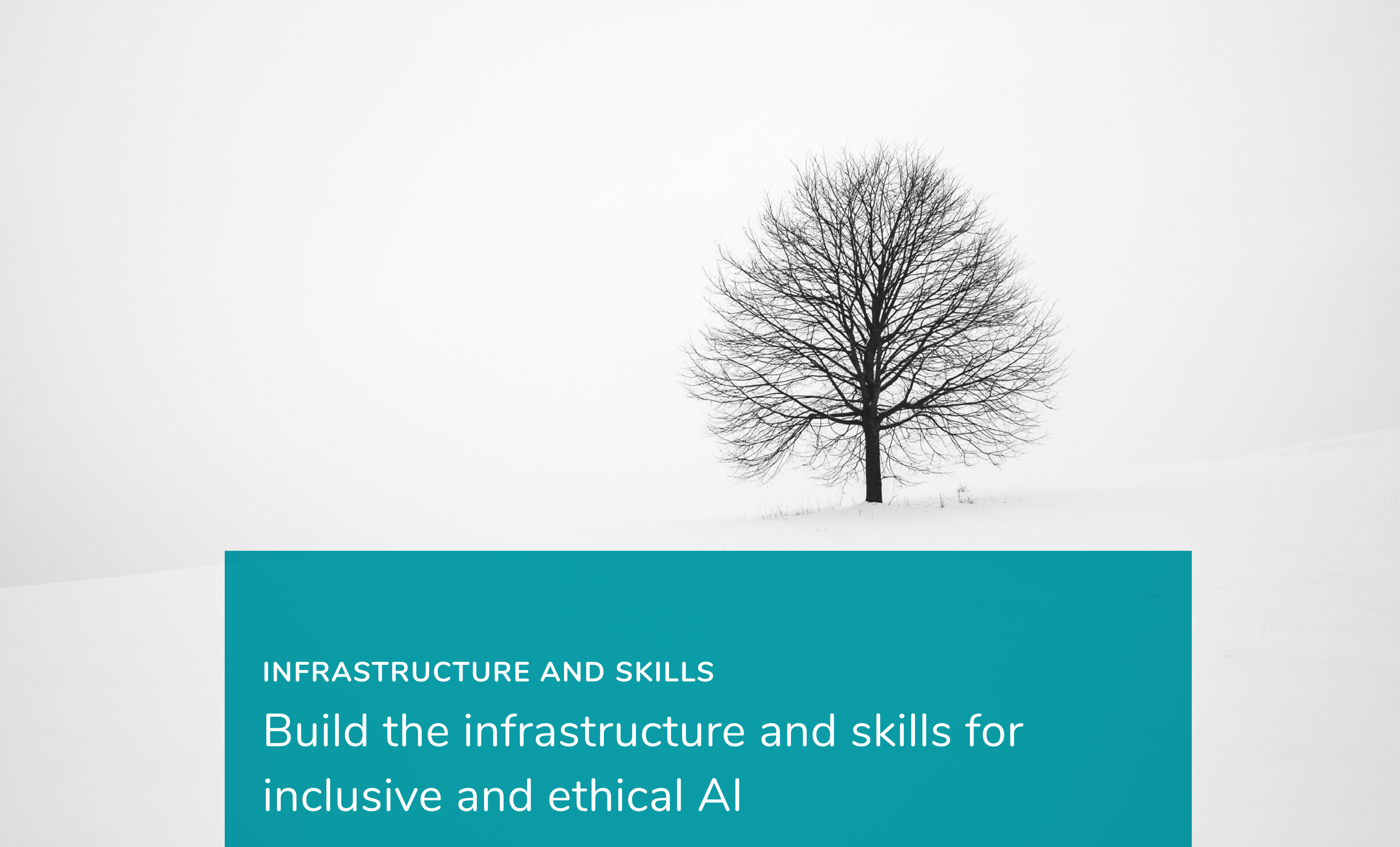
Research the social impact of AI innovations. Research is needed to better understand which AI applications work (or don't work), for whom, and in what contexts. We need to know who benefits from AI applications and how, as well as who is left out or harmed. Special emphasis must be placed on exploring the differential impacts on various groups, particularly those differentials resulting from gender, social and economic status, race, etc. This research should go beyond first order effects, such as increased efficiencies or accuracy of diagnosis, to include broader social effects. New methodologies for impact assessment and evaluation may be required.

Photo by Erico Marcelino on Unsplash

Test and monitor bias in AI applications. AI systems that make or inform decision-making that affects humans' well-being (e.g., medical diagnosis, providing a judge with an assessment of potential recidivism) should be tested and monitored for bias and errors across different contexts and communities, both before release and continuously.

Explore models of participatory design for AI. Conduct research into practices that support the development of inclusive AI applications. What techniques are effective for truly participatory processes that engage diverse populations in the design and deployment of AI applications? How and in what contexts do these practices counter design and learned bias and make AI relevant to marginalized communities? AI stakeholders in the field should release data on diversity of participation in design and development.

Action research to deepen understanding of how to effectively and equitably scale proven AI applications. Demonstrating a successful proof of concept is different from diffusing that application throughout a specified population while maintaining the quality and equity of benefits. Research on the process of scaling AI applications, both vertically to encompass additional functionality and horizontally to expand to new locations, is critical to extending the benefits of these applications. Successful transfer of an AI application across contexts requires an understanding of why an application worked well in a particular context and an appreciation that the application may require altering in order to succeed in a new environment. We need to develop theories behind the implementation of AI applications to more reliably reproduce success with different populations. In this context, the emphasis should be on variation across contexts rather than strict adherence to the fidelity of any one implementation. Particular challenges related to data will include scaling beyond the scope of existing datasets and developing means to rapidly generate datasets.



INFRASTRUCTURE AND SKILLS

Build the infrastructure and skills for inclusive and ethical AI

Support programs to build AI expertise in government. Promote AI expertise in all branches and at all levels of government, including regulatory entities and, potentially, new advisory bodies.

Foster local capacity to lead the design, development, and deployment of AI applications. Activities might include: supporting the growth of interdisciplinary AI centres of excellence in the Global South in order to engage in local development and research and provide evidence-based input into the shaping of national policy and regulatory decisions; building bridges between tech experts and low-income and marginalized communities in the Global South; and supporting South–South collaborations.

Develop and test cost-effective approaches to build relevant AI skills, particularly among women and marginalized populations. Develop and support programs that focus on developing the capacities of women and other marginalized populations to engage in different stages of the development and application of AI technologies. Research should bolster this activity through an exploration of low-cost models for developing AI tech skills and producing and testing effective curriculum and pedagogies.

Photo by Fabrice Villard on Unsplash

Expand access to data and computing resources. As much as possible, AI research, tools, and training datasets need to be made freely available. Support the development and sharing of diverse and inclusive datasets that are necessary for AI applications in different contexts.

Study the benefits and risks of open AI. Conduct research on the short- and medium-term risks and benefits of openness in AI (e.g., sharing AI resources, datasets).[108] Where possible, this research should connect supply-side questions (how best to provide open access to AI algorithms, tools, and datasets) with deepening understanding of the engagement necessary to ensure that open AI resources are available for (re)use and adaptation by diverse populations (and not just by those who are already well-skilled and -resourced). Special attention should be paid to the issue of balancing the sharing of datasets with the safeguarding of privacy.

With eyes wide open and full awareness of both the potential and pitfalls of AI, the world has a unique opportunity and imperative to address the challenges posed by this technology and to build a better future. Concerted, evidence-informed actions such as those recommended above should help the Global South better reap the rewards of an AI-fuelled future while mitigating the risks and harms and preserving human dignity.

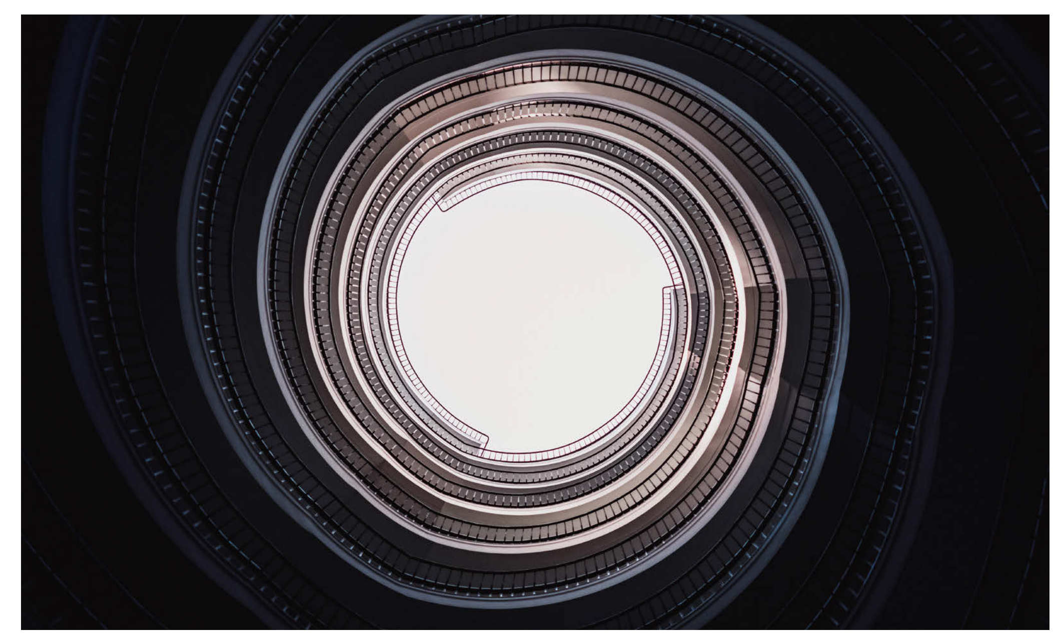


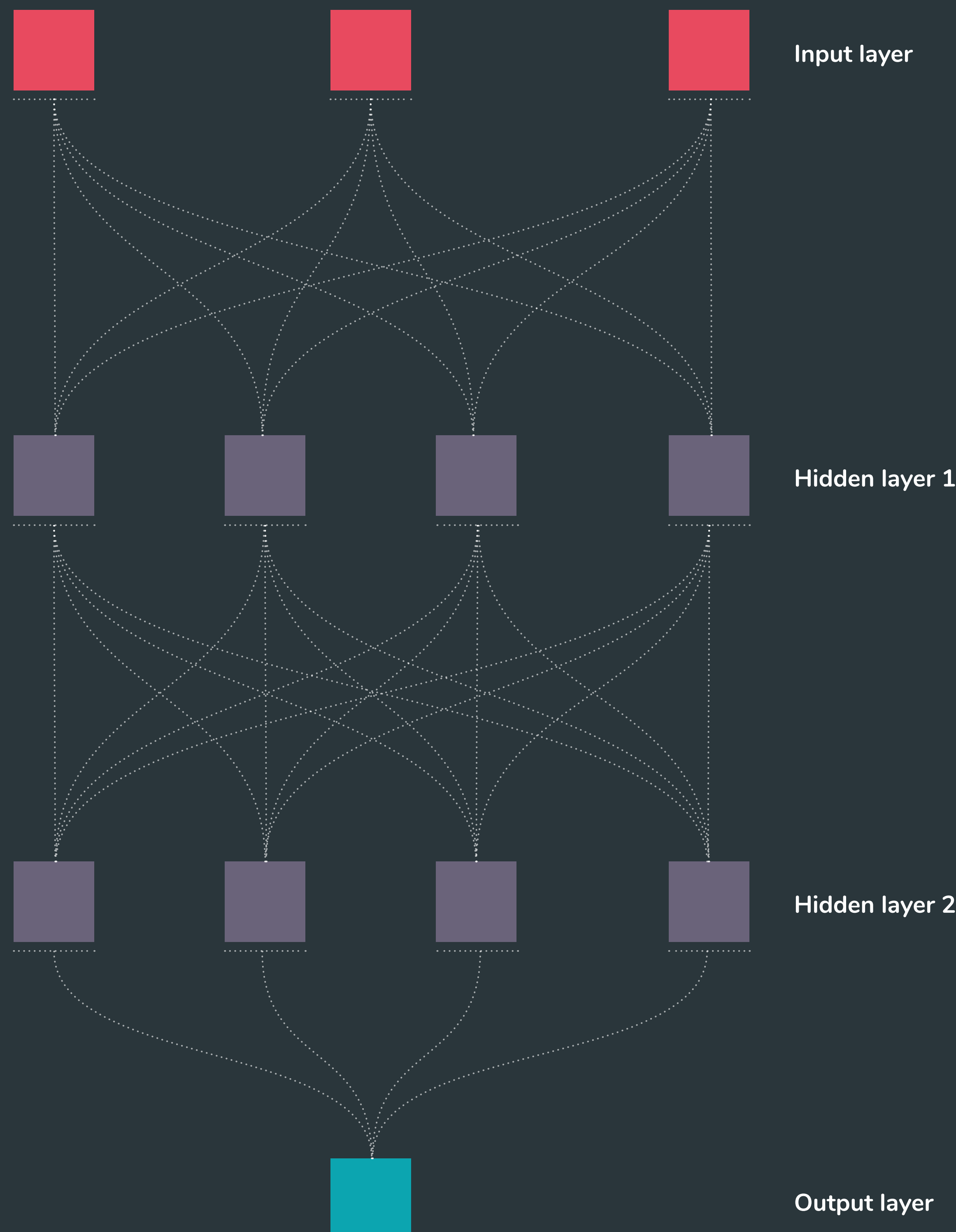
Photo by Samuel Zeller on Unsplash

APPENDIX

How does AI work?

While there are many techniques for developing AI systems, we will focus on two widely used approaches. The first is *machine learning (ML)*, where the system learns to find a solution. The second is *expert systems*, in which the AI follows a set of predetermined and preprogrammed rules and logic designed to produce particular and repeatable behaviours. Expert systems are possible when well-developed prior knowledge exists upon which to base actions.

FIGURE 4 A three-layer neural network



Source: bit.ly/2HZTyna

MACHINE LEARNING

ML has been the most successful and influential approach to developing AI. The recent rise in the use and success of ML is due to two factors: access to more powerful computers as a result of diminishing costs and increasing capacity, and the availability of vast new datasets popularly known as “big data.” In this section, we describe how ML works and the role that datasets play in the functioning of these systems.

There are three main types of ML algorithms: *supervised*, *unsupervised*, and *reinforcement learning*.

Supervised learning

In supervised ML, algorithms learn from data (e.g., digital images of oranges and apples) tagged with metadata indicating correct answers (that a given image is in fact of an orange or an apple). Such a dataset is called *training data*. The algorithm then attempts to extract classification patterns from that dataset in order to classify new, unlabelled data correctly. It extracts these classification patterns by tweaking its own parameters until it has “learned” the patterns in the data. In other words, the pattern in the data becomes represented in the parameters of the ML algorithm. This approach is called *supervised* ML because the algorithm is initially given labelled data to learn from.

One type of ML algorithm is the artificial neural network (ANN). Neural networks were inspired by biological neurons. Data are fed into a layer of neurons that perform a calculation based on a set of parameters. The result of that computation is then transmitted to the next layer(s) until it produces an output. A supervised neural network will then compare that output with the desired output and calculate the “error.” Using this error, it then works back through the network, making small adjustments to the parameters to produce an improved output for that input the next time. This process is repeated many times through for all the data in the training set, until the total sum of the errors is sufficiently low to constitute success (see Figure 4).

Once the network is trained, it can be fed new inputs and will produce a predicted output. In the fruit example above, the algorithm “learns” to represent the salient features of the fruits, such as size and colour, through the adjustment of many parameters. Once it has learned, it effectively constitutes an “orange/apple” detector that will provide its best guess of whether an image is an orange or an apple.

Photo by Aaron Burden on Unsplash

Unsupervised learning

Unsupervised learning algorithms learn patterns in the data without receiving any labelled output data. Unsupervised algorithms are not given the correct "answer"; rather, they extract various features from the dataset and construct clusters of data points with similar features. These can be used to discover groups, such as customer segmentation, or to form associations (or rules) in the data — for example, "when X happens, then Y occurs." In the fruit example, an unsupervised algorithm might classify large and small fruits into two clusters. It could also have the ability to create multidimensional clusters based on other features such as shape and colour.

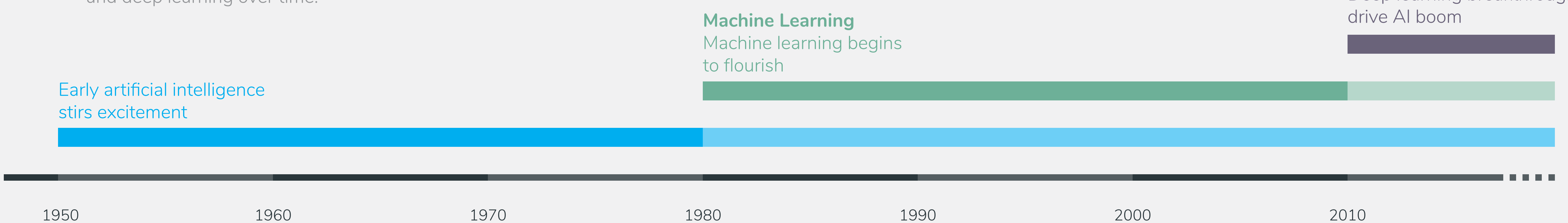
Semi-supervised learning

This approach trains the algorithm on both labelled and unlabelled data. The cost of labelling datasets can be very high, and a fully labelled dataset may not be feasible. The inclusion of labelled data, however, helps the algorithm identify potential clusters or rules in the data, while the unlabelled data enable the algorithm to shape the boundaries of those clusters or rules and even find new ones. Semi-supervised learning works well for things like speech and image recognition.

Deep learning is a type of artificial neural network that uses multiple hidden layers. It typically blends supervised and unsupervised learning, and occasionally semi-supervised learning. While there is no exact number of hidden layers that equates to "deep" learning, deep learning ANNs have been created with more than a billion total connections.[109] The ability to implement deep learning techniques is a driving force behind the renewed interest and advancements in AI (see Figure 5).

The power of deep learning comes from the ability of the multiple layers to extract complex features from the training data. In the example of visually classifying fruits, features such as colour and size would be given greater priority than taste. In the visual cortex of a mammalian brain, the input that enters the eye goes through multiple layers of hierarchical processing that extracts features like edges, contours, and object shapes at multiple visual areas (Hubel and Wiesel 1962; Felleman and Van Essen 1991).[110] (See Figure 6.)

FIGURE 5 The relationship between artificial intelligence, machine learning, and deep learning over time.*

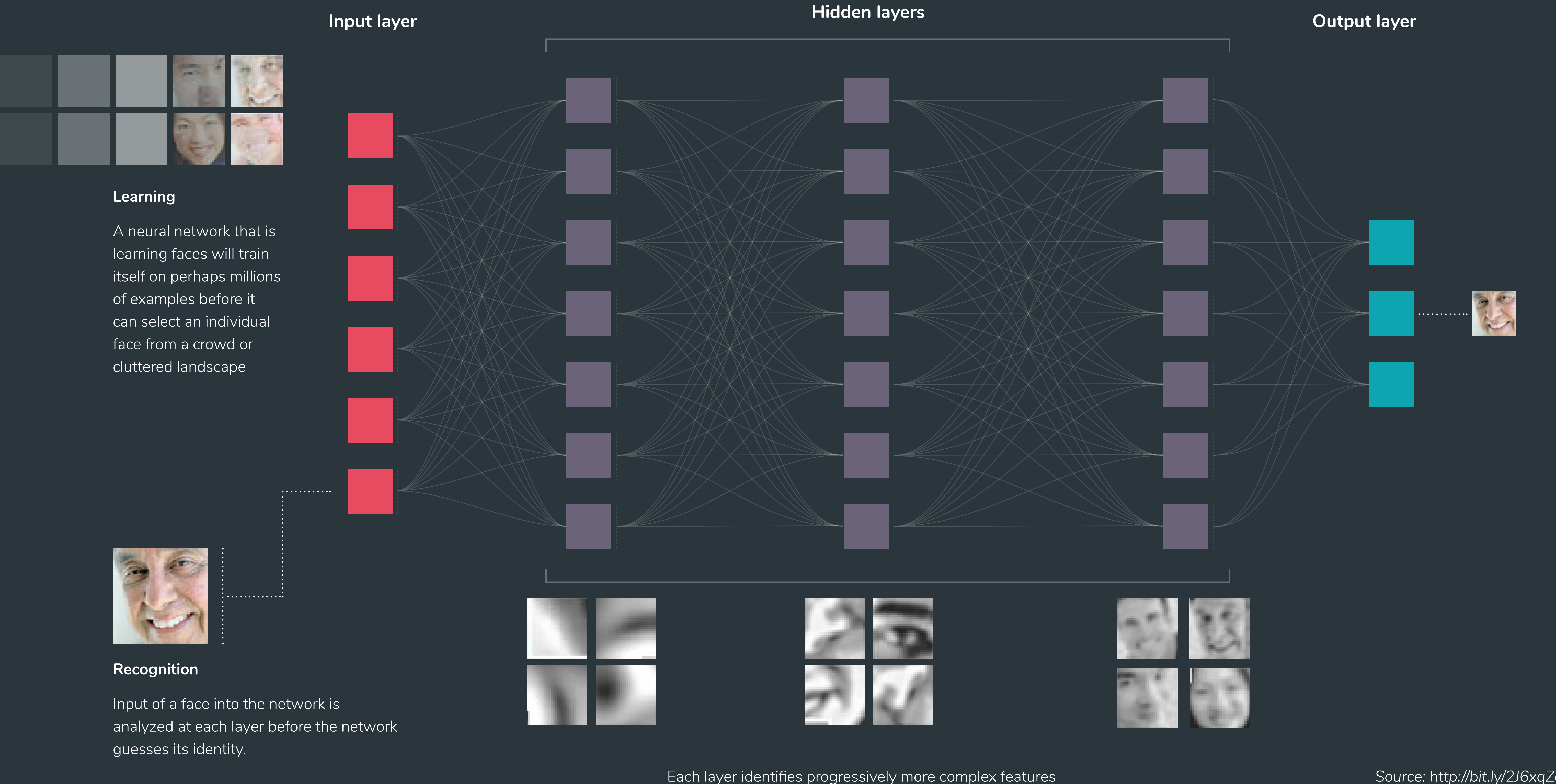


Since the early flush of optimism in the 1950s, smaller subsets of artificial intelligence — first machine learning, then deep learning, a subset of machine learning — have created ever larger disruptions.

* bit.ly/2HXaIVZ

FIGURE 6

Unveiling the hidden layers of deep learning



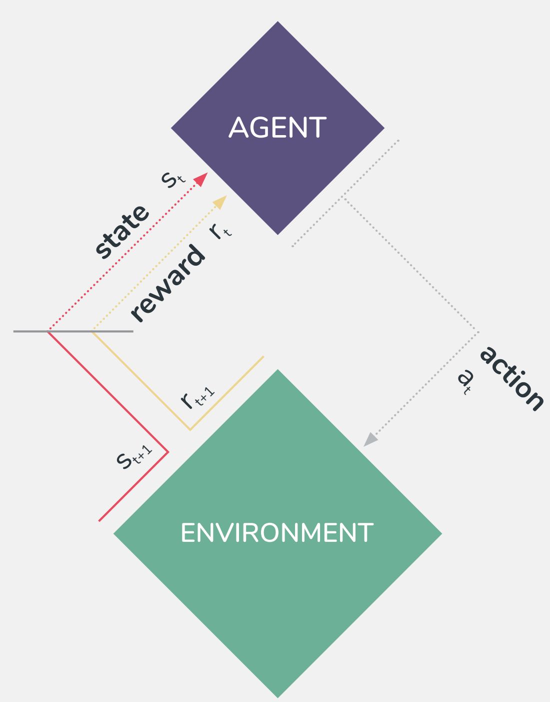
Source: http://bit.ly/2J6xqZq

Reinforcement learning

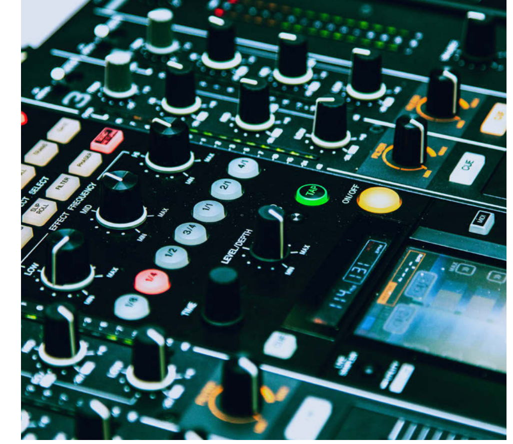
While ML algorithms such as neural networks learn from datasets, reinforcement learning (RL) algorithms learn through receiving positive and negative reinforcement from the environment. As an AI agent with an RL algorithm engages with the world, it is given positive or negative feedback based on its latest action. This is then fed back into the algorithm, which makes corrections based upon the action it chose at that particular state (see Figure 7).

RL is highly powerful for goal-oriented tasks, such as learning a game or teaching a robot a skill. For example, using RL, a robotic arm learned how to flip a pancake[111] and play table tennis,[112] and an autonomous spider learned how to walk.[113] The ability to manipulate objects is critical for robotics within manufacturing. RL has also been used in the financial sector to develop stock trading strategies.[114] RL also enabled the game-playing AI application AlphaGo to develop new and unique strategies to defeat the Go World Champion.[115]

FIGURE 7 Reinforcement Learning & its implementation



Source: bit.ly/2uepiTw



Photos by Hermes Riviera, Chester Alvarez on Unsplash

Final thoughts on ML algorithms

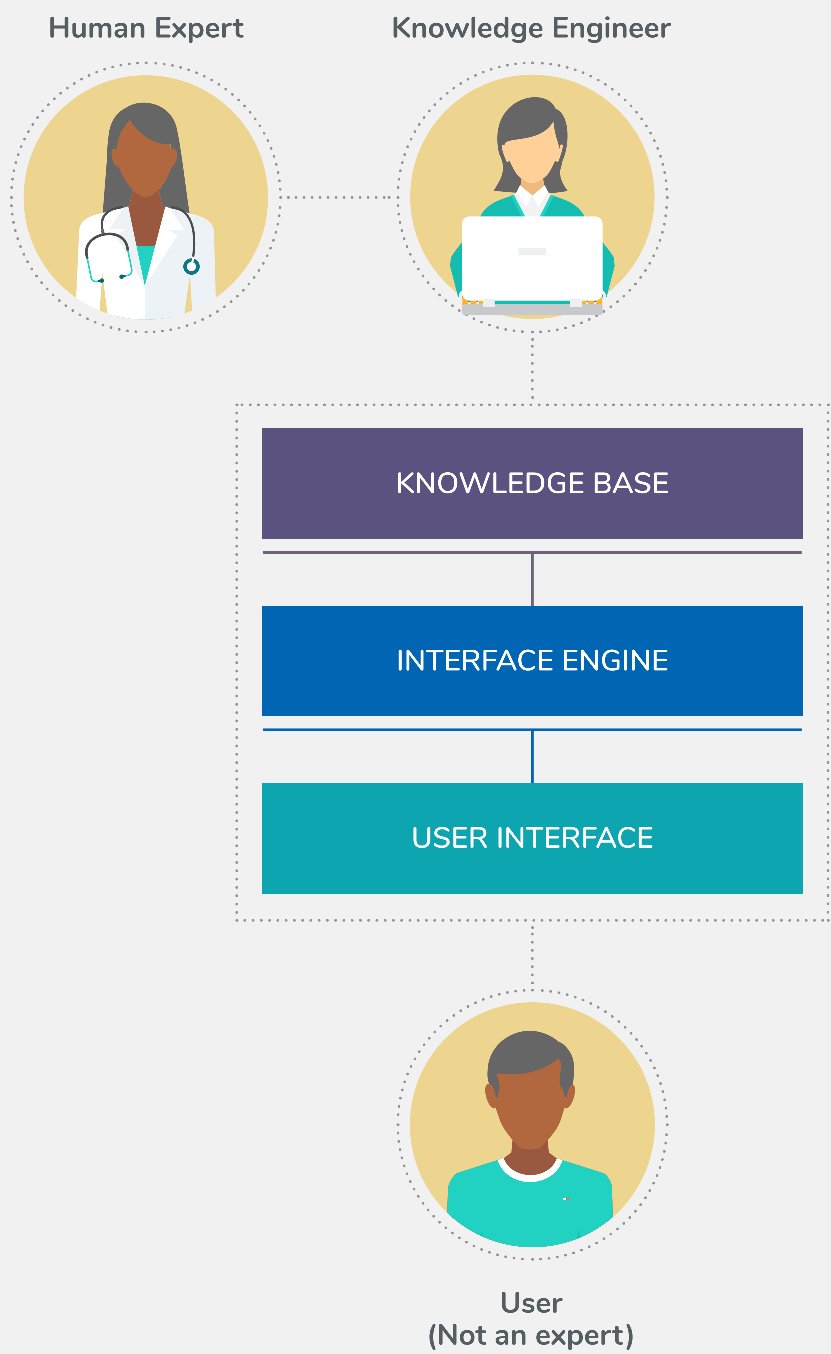
There are also "simpler" learning algorithms that do not require big datasets or extensive computing power. For example, a linear regression — a mathematical technique that develops an equation that "fits" a particular dataset of inputs and an output — is considered a ML algorithm.[116] The algorithm can then "predict" the most likely output given a new data point input. An application of this technique in a higher education context might be to "predict" when a student might drop out based on data such as attendance and participation rates. This particular algorithm doesn't require big datasets, extensive computing power, or highly complex learning algorithms.

EXPERT SYSTEMS

ML algorithms are a "non-symbolic" type of AI. In other words, they are not explicitly programmed to solve a problem using logical instructions (e.g., "if this, do that") or symbolic reasoning; rather, ML algorithms crunch data and learn solutions by adjusting the algorithm's parameters. These techniques were inspired largely by the workings of the human brain, where knowledge was thought to be stored in a mass of neuronal connections that are constantly being adjusted, strengthened, or broken.

The birth of AI, however, was rooted in approaches that used symbols to represent a problem, much as humans might reason through a problem. One such approach was the development of "expert systems" (Harmon and King 1985) that attempt to emulate the problem-solving skills of a human expert (Durkin 1990: 171). Edward Feigenbaum defines an expert system as "an intelligent computer program that uses knowledge and inference procedures to solve problems that are difficult enough to require significant human expertise for their solution."[117] Expert systems were popular at a time when computing power and data availability were more limited. For example, in 1987, over two-thirds of Fortune 1000 companies had expert system projects under development (Durkin 1990).

FIGURE 8 Components of expert systems



Source: bit.ly/2DNtkS9

Key elements of an expert system are:

- Knowledge base consisting of explicitly encoded specialized knowledge (expertise), often encoded as "if … then" rules.
- Working memory: Data entered by users or drawn from other data sources and facts inferred by the system.
- Inference engine: Derivation of new information about the problem using information available in working memory and the knowledge base. This is typically done by either (a) establishing a goal/hypothesis or attempting to verify it through data analysis, or (b) collecting information about a problem and then inferring other information (Durkin 1990).

Figure 8 illustrates the relationship between the expert, the expert system, and the user.

There are many uses for expert systems, from medical diagnostic tools to intelligent tutoring systems, from flying planes in autopilot to running elevators, and from ecological planning to fault diagnosis production and scheduling (Liao 2005). The symbolic representation of knowledge and expert systems techniques are also used concurrently with ML algorithms to achieve more complex behaviours.

REFERENCES

Ahmad, T., Rehman, N. A., Pervaiz, F., Kalyanaraman, S., Safeer, M. B., Chakraborty, S., and Subramanian, L. (2013). Characterizing dengue spread and severity using internet media sources. In *Proceedings of the 3rd ACM Symposium on Computing for Development* (p. 18). ACM. bit.ly/2Fe1y1W

AI Now (2017). *The AI Now 2017 Report*. bit.ly/2FewZcy

AI Now (2016). *The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term*. Summary of the AI Now public symposium hosted by the White House and New York University's Information Law Institute, July 7, 2016. bit.ly/2qX0kDW

Arntz, M., Gregory, T., and Zierahn, U. (2016). *The Risk of Automation for Jobs in OECD Countries* (OECD Social, Employment and Migration Working Papers No. 189). bit.ly/2qT0qM6

Banavar, G. (2016). *Learning to Trust Artificial Intelligence Systems: Accountability, Compliance and Ethics in the Age of Smart Machines.* Somers, NY: IBM. ibm.co/2qXGojE

Benkler, Y. (2011). Networks of power, degrees of freedom. *International Journal of Communication*, *5*, 39.

Bolsover, G., and Howard, P. (2017). Computational propaganda and political big data: Moving toward a more critical research agenda. *Big Data*, *5*(4), 273–276. bit.ly/2JqBLGb

Bostrom, N. (2017). Strategic implications of openness in AI development. *Global Policy*, *8*(2), 135–148. doi:10.1111/1758-5899.12403

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... and Filar, B. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *ArXiv Preprint ArXiv:1802.07228*.

Caliskan, A., Bryson, J. J., and Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186.

Calo, R. (2017). *Artificial Intelligence Policy: A Primer and Roadmap*. SSRN. Retrieved from bit.ly/2qVDegW

Photo by Todd Diemer on Unsplash

Christensen, G., Steinmetz, A., Alcorn, B., Bennett, A., Woods, D., and Emanuel, E. J. (2013). *The MOOC Phenomenon: Who Takes Massive Open Online Courses and Why?* SSRN. Retrieved from bit.ly/2KcBEQ6

Danks, D., and London, A. J. (2017). *Algorithmic Bias in Autonomous Systems*. Presented at the 26th International Joint Conference on Artificial Intelligence (IJCAI). Retrieved from bit.ly/2Kb0sI1

Deibert, R. J. (2013). *Black Code: Inside the Battle for Cyberspace*. Toronto: Signal.

Durkin, J. (1990). Research review: Application of expert systems in the sciences. Ohio Journal of Science, 90(5), December 1990, 171–179. bit.ly/2HnKMU9

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1), January 1991, 1–47. doi:10.1093/cercor/1.1.1

Frey, C. B., and Osborne, M. A. (2013). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, issue C, 254-280. bit.ly/2kP5AGv

Friedman, B., Brok, E., Roth, S. K., and Thomas, J. (1996). Minimizing bias in computer systems. *ACM SIGCHI Bulletin*, *28*(1), 48–51.

Friedman, B., and Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, *14*(3), 330–347.

Garrido, M., Lucas, K., Andersen, S., Mena, A. F., Macapagal, M., and Dalvit, L. (2016). *An Examination of MOOC Usage for Professional Workforce Development Outcomes in Colombia, the Philippines, and South Africa*. Technology and Social Change Group, University of Washington.

Hansen, J. D., and Reich, J. (2015). Democratizing education? Examining access and usage patterns in massive open online courses. *Science*, 350 (6265), 1245–1247.

Harmon, P., and King, D. (1985). *Expert Systems: Artificial Intelligence in Business*. New York: John Wiley & Sons. bit.ly/2HrQ4Ke

Hilbert, M. (2016). The bad news is that the digital access divide is here to stay: Domestically installed bandwidths among 172 countries for 1986–2014. *Telecommunications Policy, 40*(6), 567–581.

Hossain, M. S., Zander, P.-O., Kamal, M. S., and Chowdhury, L. (2015). Belief-rule-based expert systems for evaluation of e-government: a case study. *Expert Systems*, *32*(5), 563–577.

Hubel, D., and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), January 1962. 10.1113/jphysiol.1962.sp006837

Institute of Electrical and Electronics Engineers (IEEE) (2017). *Ethically Aligned Design*, version 2. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. ethicsinaction.ieee.org/

International Federation of Robotics (2017). *The Impact of Robots on Productivity, Employment and Jobs*. bit.ly/2HIPR8Z

Kulik, J. A., and Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, *86*(1), 42–78.

Laurillard, D., Kennedy, E., and Wang, T. (2018). *How Could Digital Learning at Scale Address the Issue of Equity in Education*? Foundation for Information Technology Education and Development, Inc., Quezon City, Philippines.

Liao, S.H. (2005). Expert system methodologies and applications – a decade review from 1995 to 2004. *Expert Systems with Applications*, *28*(1), 93–103. doi:10.1016/j.eswa.2004.08.003

Lim, C. P., and Tinio, V. L. (Eds.). (2018). *Learning analytics for the Global South*. Quezon City, Philippines: Foundation for Information Technology Education and Development. bit.ly/2HVIKr9

Long, P., and Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review*. Retrieved from bit.ly/2JoqoyR

Lyon, D. (2010). Surveillance, Power and Everyday Life. In P. Kalantzis-Cope and K. Gherab-Martin (Eds.), *Emerging Digital Spaces in Contemporary Society: Properties of Technology* (pp. 107–120). Springer.

Magoutas, B., and Mentzas, G. (2010). SALT: A semantic adaptive framework for monitoring citizen satisfaction from e-government services. *Expert Systems with Applications*, 37(6), 4292-4300.

Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., and Dewhurst, M. (2017). *A Future that Works: Automation, Employment, and Productivity*. McKinsey Global Institute.

Marcus, G. (2018). *Deep Learning: A Critical Appraisal. ArXiv Preprint ArXiv:1801.00631*. Retrieved from bit.ly/2HJsBrd

Muralidharan, K., Singh, A., and Ganimian, A. J. (2017). *Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India* (CESifo Working Paper No. 6328). Center for Economic Studies and Ifo Institute.

Mwalumbwe, I., and Mtebe, J. S. (2017). Using learning analytics to predict students' performance in Moodle Learning Management Systems: A case of Mbeya University of Science and Technology. *The Electronic Journal of Information Systems in Developing Countries*, *79*, 1–13. doi:10.1002/j.1681-4835.2017.tb00577.x.

Nilsson, N. J. (2009). *The Quest for Artificial Intelligence: A History of Ideas and Achievements*. New York: Cambridge University Press.

Nissenbaum, H. (2001). How computer systems embody values. *Computer*, *34*(3), 120–119. doi:10.1109/2.910905

Nye, B. D. (2015). Intelligent tutoring systems by and for the developing world: A review of trends and approaches for educational technology in a global context. *International Journal of Artificial Intelligence in Education*, *25*(2), 177–203. doi:10.1007/s40593-014-0028-6

Osoba, O. A., and Welser, W., IV (2017). *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. RAND Corporation. doi:10.7249/RR1744

Penney, J. (2017). Internet surveillance, regulation, and chilling effects online: A comparative case study. *Internet Policy Review*, *6*(2).

Petrovica, S., Anohina-Naumeca, A., and Ekenel, H. K. (2017). Emotion recognition in affective tutoring systems: Collection of ground-truth data. *Procedia Computer Science*, *104*, 437–444. doi:10.1016/j.procs.2017.01.157

Phobun, P., and Vicheanpanya, J. (2010). Adaptive intelligent tutoring systems for e-learning systems. *Procedia – Social and Behavioral Sciences*, *2*(2), 4064–4069. doi:10.1016/j.sbspro.2010.03.641

Press, G. (2017). *Robot Overlords: AI at Facebook, Amazon, Disney and Digital Transformation at GE, DBS, BNY Mellon*. Forbes Online, May 28, 2017. bit.ly/2KaZ9sB

Privacy International (2016). *Algorithms, Intelligence, and Learning Oh My* [blog post]. December 1, 2016. bit.ly/2HuzjOk (2017). *Submission of evidence to the House of Lords Select Committee on Artificial Intelligence*. London, UK, September 6, 2017.

PwC (2017). *Sizing the Prize: What's the Real Value of AI for Your Business and How can you Capitalise?* pwc.to/2vHve8j

Quinn, J. (2013). Computational Techniques for Crop Disease Monitoring in the Developing World. In Tucker, A., Höppner, F., Siebes. A., and Swift, S. (Eds.), *Advances in Intelligent Data Analysis XII*. IDA 2013. *Lecture Notes in Computer Science*, *8207*. Springer. doi:10.1007/978-3-642-41398-8_2

Quinn, J., Frias-Martinez, V., and Subramanian, L. (2014). Computational sustainability and artificial intelligence in the developing world. *AI Magazine*, *35*(3). bit.ly/2HJQA9Q

Reyes, C. T. (2018). Developing a student support system through learning analytics for undergraduates at the University of the Philippines Open University. Quezon City, Philippines: Foundation for Information Technology Education and Development.

Singh, A., Patil, D., Reddy, G. M., and Omkar, S. N. (2017). *Disguised Face identification (DFI) with Facial Keypoints Using Spatial Fusion Convolutional Network. ArXiv Preprint ArXiv:1708.09317*.

Smith, M. S. (2013). Open Educational Resources: Opportunities and Challenges for the Developing World. In Smith, M. L., and Reilly, K.A. (Eds.), *Open Development: Networked Innovations in International Development*. Cambridge, MA and Ottawa: MIT Press and IDRC.

Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., ..., and Teller, A. (2016). Artificial intelligence and life in 2030. *One Hundred Year Study on Artificial Intelligence: Report of the 2015–2016 Study Panel*. Stanford University, Stanford, CA, September 2016. stanford.io/2HYuZrU

Suwajanakorn, S., Seitz, S. M., and Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: learning lip sync from audio. *ACM Transactions on Graphics*, *36*(4), 1–13. bit.ly/2HGwCgj

United Nations General Assembly. 2015. Outcome document of the high-level meeting of the General Assembly on the overall review of the implementation of the outcomes of the World Summit on the Information Society. A/RES/70/125, 70th Sess., 16 December 2015. bit.ly/2vGPtTA

United Nations Human Rights Council. 2016. The promotion, protection and enjoyment of human rights on the Internet. A/HRC/RES/32/13, 32nd Sess., 1 July 2016. bit.ly/2vKayMZ

Van de Poel, I., and Kroes, P. (2014). Can Technology Embody Values? In P. Kroes and P.-P. Verbeek (Eds.), *The Moral Status of Technical Artefacts* (Vol. 17, pp. 103–124). Dordrecht: Springer Netherlands. doi:10.1007/978-94-007-7914-3_7

Wachter, S., Mittelstadt, B., and Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*. Available at SSRN: bit.ly/2vHMiuW

Winthrop, R., and McGivney, E. (2017). *Can We Leapfrog? The Potential of Education Innovations to Rapidly Accelerate Progress*. Center for Universal Education at Brookings.

World Bank (2016). *World Development Report 2016: Digital Dividends*. The World Bank. bit.ly/2JjR8QI

World Economic Forum (2016). *The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution* (Global Challenge Insight Report).

World Wide Web Foundation (2017). *Artificial Intelligence: The Road Ahead in Low and Middle-Income Countries*. bit.ly/2JnldyU

Yang, H.-H., Liu, J.-L., Chang, M.C.S., and Yang, J.-C. (2012). Improvement of e-government service process via a grey relation agent mechanism. *Expert Systems with Applications*, 39(10), 9755-9763.

Zhao, J., Wang, T., Yatskar, M., Ordonex, V., and Chang, K.-W. (2017). *Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-level Constraints. ArXiv Preprint ArXiv:1707.09457*. Retrieved from bit.ly/2HWirkS

ENDNOTES

1 Or as Jeff Bezos called it, "a horizontal enabling layer": bit.ly/2FeQ91P

2 The Sustainable Development Goals (SDGs) are a set of 17 global goals within the UN 2030 Agenda for Sustainable Development. See: bit.ly/2HtQ7cJ

3 From ITU video: bit.ly/2FhfeJA

4 bit.ly/2KfsYZ9

5 bit.ly/2HWYQAR

6 bit.ly/2HLU6R0

7 As highlighted by Eric Horvitz, director of Microsoft Research Labs and a past president of the Association for the Advancement of Artificial Intelligence, in an editorial in Science (July 7, 2017): Excitement about AI has been tempered by concerns about potential downsides … biases buried deep in data sets, leading to unfair and inaccurate inferences … legal and ethical issues regarding decisions made by autonomous systems, difficulties with explaining inferences, threats to civil liberties through new forms of surveillance, precision manipulation aimed at persuasion, criminal uses of AI, destabilizing influences in military applications, and the potential to displace workers from jobs and to amplify inequities in wealth. bit.ly/2qZPmh3

8 ibm.co/2qZcnzX

9 bit.ly/2Ho45wu

10 Nilsson 2009.

11 bit.ly/2JnZt68

12 AI Now 2016: 2.

13 PwC 2017.

14 bit.ly/2vEKkLL

15 bit.ly/2qX1H5Q and bit.ly/2HSnXaz

16 "Natural language processing (NLP) is the ability of a computer program to understand human speech as it is spoken." From: bit.ly/2HsiJCT

17 "Machine translation (MT) is the task of automatically converting one natural language into another, preserving the meaning of the input text, and producing fluent text in the output language." From: nlp.stanford.edu/projects/mt.shtml

18 From a philosophical perspective, it is also one way to determine "intelligence". "If an agent behaves intelligently, it is intelligent. It is only the external behaviour that defines intelligence; acting intelligently is being intelligent." bit.ly/2JnZt68

19 bit.ly/2HEacJ2

20 bit.ly/2vEKkLL

21 air.ug/microscopy/

22 aidr.qcri.org/

23 go.nature.com/2vYDy3u

24 bit.ly/2ra0ilY

25 bit.ly/2jguJc3

26 bit.ly/2Kp8D3t
27 bit.ly/2KnAtwZ
28 bit.ly/2r9ZrYM
29 See, for example: bit.ly/2w0YA1C
30 bit.ly/2r9joik
31 contenttechnologiesinc.com/
32 bit.ly/2r6Z68w
33 mck.co/2HDEwUm
34 bit.ly/2HzrwDk
35 bit.ly/2HzqXtc
36 bit.ly/2FrCkNL
37 bit.ly/2vRxlRs
38 bit.ly/2vSuSvq
39 tcrn.ch/2r7QizT
40 bit.ly/2JAM4aP
41 bit.ly/2vUDO3s
42 nyti.ms/2jeU0n2
43 A 2016 report released by a consortium of experts clearly states that there is "no cause for concern that AI is an imminent threat to humankind" (Stone et al. 2016). However, AI in a video game enabled non-player characters to create super weapons and begin hunting players, which is slightly concerning! See: bit.ly/2KkU0hE
44 bit.ly/2HLU6R0
45 bit.ly/2KnNx5y
46 "But similar errors have emerged in Nikon's camera software, which misread images of Asian people as blinking, and in Hewlett-Packard's web camera software, which had difficulty recognizing people with dark skin tones." Although the group did not build the algorithm to treat light skin as a sign of beauty, the input data effectively led the robot judges to reach that conclusion: bit.ly/2r7Oh6c
47 bit.ly/2jjZQ6V
48 bit.ly/2r9Z4gQ
49 bit.ly/2HC4dV3
50 bit.ly/2KmVisB
51 bit.ly/2r8hZYx and bit.ly/2r6FD7K
52 bit.ly/2HBxL9u
53 bit.ly/2JEAueL
54 bit.ly/2JskYCW
55 bit.ly/2FqRPFQ
56 bit.ly/2JqyAyh
57 bit.ly/2J6xqZq
58 bit.ly/2JskYCW
59 bit.ly/2HwosU7
60 bit.ly/2HPNAbW
61 nyti.ms/2vVoaVR
62 bit.ly/2qZnrOd
63 The risks of big data to privacy have been the subject of discussion and concern for some time, with the topic discussed in detail in a White House report, Big Data and Privacy, published in 2014. bit.ly/2JocxbO
64 bit.ly/2JtxBh9
65 bit.ly/2Ju6CBJ
66 bit.ly/2qYaDYj
67 bit.ly/2IMhrj2
68 bit.ly/2HqRPvo
69 bit.ly/2Kg4bnM
70 Writings on privacy and AI: bit.ly/2Kfvm2a bit.ly/2HrA0bn, bit.ly/2JocxbO, for.tn/2Jqwy17 timreview.ca/article/1067
71 As of Jan 2018, there are over 2 billion active FB users around the world. See bit.ly/2HOhJbz
72 pewrsr.ch/2KcCcVU
73 bit.ly/2JuEb74
74 See Yochai Benkler's argument to this effect: bit.ly/2JoeFjJ
75 bit.ly/2Ke7b41
76 bit.ly/2Kh3pXx
77 econ.st/2qYLjI5
78 nyti.ms/2vLSB0r
79 bit.ly/2Ke7b41
80 read.bi/2JqUZLW
81 lyrebird.ai/demo
82 bit.ly/2HrUosM
83 See an example video: bit.ly/2Hshtf2
84 bit.ly/2rauy6D
85 bit.ly/2KbPRMR
86 bit.ly/2Fe4oUO
87 go.nature.com/2HrH7EQ
88 bit.ly/2JmoXkg
89 bbc.in/2JoJPaF
90 bit.ly/2JmoXkg
91 bit.ly/2HoGawZ
92 bit.ly/2Kb6M2h
93 bit.ly/2Fdb7OI

94 bit.ly/2JmpsuC

95 bit.ly/2JmoXkg

96 bit.ly/2HJRYcy

97 bit.ly/2HXoaHb

98 bit.ly/2Jq3K97

99 bit.ly/2HVJJYn

100 It should be noted that this perspective is heavily influenced by our position in a Global North research funding institution, where we have a history of supporting research on information and communication technologies for development.

101 bit.ly/2IKzPcr

102 For example, see: bit.ly/2xuC2ac and bit.ly/2L9PBxR

103 See, for example: bit.ly/2K7IA1X

104 bit.ly/2Kew2EW

105 These recommendations also draw from other important AI reports (AI Now 2016; 2017; Calo 2017; Privacy International 2017; Stone et al. 2016; World Wide Web Foundation 2017; Brundage et al. 2018).

106 Thanks to Maroussia Levesque for this point.

107 "[T]he same rights that people have offline must also be protected online" (UN Human Rights Council Resolution L13, *The Promotion, Protection and Enjoyment of Human Rights on the Internet*, art. 1); UN General Assembly 2015, art. 43; "[AI] should be designed and operated in a way that both respects and fulfills human rights, freedoms, human dignity, and cultural diversity" (IEEE 2017: 22).

108 For a discussion of the benefits and risks, see Bostrom (2017).

109 bit.ly/2FhONn8

110 David Hubel and Torsten Wiesel were awarded the Nobel Prize in physiology or medicine in 1981 partly for their discovery of visual neurons sensitive to specific features such as oriented bars and gratings. Their work inspired the development of the neocognitron, the first artificial neural network, proposed by Kunihiko Fukushima in 1982.

111 bit.ly/2JtMCPP

112 bit.ly/2HqmXep

113 bit.ly/2KhudHk

114 bit.ly/2Fh4JGf

115 bit.ly/2KfukDd

116 See, for example: bit.ly/2xvJTV2 and bit.ly/2shVF0i

117 bit.ly/2qYHnku

Artificial intelligence and human development

IDRC | CRDI
International Development Research Centre
Centre de recherches pour le développement international

Canada