GLOBAL INFORMATION SOCIETY WATCH 2019

Artificial intelligence:
Human rights, social justice and development

Global Information Society Watch

2019

Global Information Society Watch 2019

Artificial intelligence: Human rights, social justice and development

Disclaimer: The views expressed herein do not necessarily represent those of Sida, ARTICLE 19, APC or its members.

Introduction

Vidushi Marda[1]
ARTICLE 19
www.article19.org

Much has been written about the ways in which artificial intelligence (AI) systems have a part to play in our societies, today and in the future. Given access to huge amounts of data, affordable computational power, and investment in the technology, AI systems can produce decisions, predictions and classifications across a range of sectors. This profoundly affects (positively and negatively) economic development, social justice and the exercise of human rights.

Contrary to popular belief that AI is neutral, infallible and efficient, it is a socio-technical system with significant limitations, and can be flawed. One possible explanation is that the data used to train these systems emerges from a world that is discriminatory and unfair, and so what the algorithm learns as ground truth is problematic to begin with. Another explanation is that the humans building these systems have their unique biases and train systems in a way that is flawed. Another possible explanation is that there is no true understanding of *why* and *how* some systems are flawed – some algorithms are inherently inscrutable and opaque,[2] and/or operate on spurious correlations that make no sense to an observer.[3] But there is a fourth cross-cutting explanation that concerns the global power relations in which these systems are built. AI systems, and the deliberations surrounding AI, are flawed because they amplify some voices at the expense of others, and are built by a few people and

imposed on others. In other words, the design, development, deployment and deliberation around AI systems are profoundly political.

The 2019 edition of GISWatch seeks to engage at the core of this issue – what does the use of AI systems promise in jurisdictions across the world, what do these systems deliver, and what evidence do we have of their actual impact? Given the subjectivity that pervades this field, we focus on jurisdictions that have been hitherto excluded from mainstream conversations and deliberations around this technology, in the hope that we can work towards a well-informed, nuanced and truly global conversation.

The need to address the imbalance in the global narrative

Over 60 years after the term was officially coined, AI is firmly embedded in the fabric of our public and private lives in a variety of ways: from deciding our creditworthiness,[4] to flagging problematic content online,[5] from diagnosis in health care,[6] to assisting law enforcement with the maintenance of law and order.[7] AI systems today use statistical methods to learn from data, and are used primarily for prediction, classification, and identification of patterns. The speed and scale at which these systems function far exceed human capability, and this has captured the imagination of governments, companies, academia and civil society.

AI is broadly defined as the ability of computers to exhibit intelligent behavior.[8] Much of what is re-

---

1   Lawyer and Digital Programme Officer at ARTICLE 19, non-resident research analyst at Carnegie India. Many thanks to Mallory Knodel and Amelia Andersdotter for their excellent feedback on earlier versions of this chapter.

2   Diakopoulos, N. (2014). *Algorithmic Accountability Reporting: On the Investigation of Black Boxes.* New York: Tow Centre for Digital Journalism. https://academiccommons.columbia.edu/doi/10.7916/D8TT536K/download

3   https://www.tylervigen.com/spurious-correlations

4   O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* New York: Crown Publishing Group.

5   Balkin, J. (2018). Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation. *Yale Law School Faculty Scholarship Series.* https://digitalcommons.law.yale.edu/fss_papers/5160

6   Murali, A., & PK, J. (2019, 4 April). India's bid to harness AI for Healthcare. *Factor Daily.* https://factordaily.com/ai-for-healthcare-in-india

7   Wilson, T., & Murgia, M. (2019, 20 August). Uganda confirms use of Huawei facial recognition cameras. *Financial Times.* https://www.ft.com/content/e20580de-c35f-11e9-a8e9-296ca66511c9

8   Elish, M. C., & Hwang, T. (2016). *An AI Pattern Language.* New York: Intelligence and Autonomy Initiative (I&A) Data & Society. https://www.datasociety.net/pubs/ia/AI_Pattern_Language.pdf

ferred to as "AI" in popular media is one particular technique that has garnered significant attention in the last few years – machine learning (ML). As the name suggests, ML is the process by which an algorithm learns and improves performance over time by gaining greater access to data.[9] Given the ability of ML systems to operate at scale and produce data-driven insights, there has been an aggressive embracing of its ability to solve problems and predict outcomes.

While the expected potential public benefits of ML are often conjectural, as this GISWatch shows, its tangible impact on rights is becoming increasingly clear across the world.[10] Yet a historical understanding of AI and its development leads to a systemic approach to explanation and mitigation of its negative impact. The impact of AI on rights, democracy, development and justice is both significant (widespread and general) and bespoke (impacting on individuals in unique ways), depending on the context in which AI systems are deployed, and the purposes for which they are built. It is not simply a matter of ensuring accuracy and perfection in a technical system, but rather a reckoning with the fundamentally imperfect, discriminatory and unfair world from which these systems arise, and the underlying structural and historical legacy in which these systems are applied.

Popular narratives around AI systems have been notoriously lacking in nuance. While on one end, AI is seen as a silver bullet technical solution to complex societal problems,[11] on the other, images of sex robots and superintelligent systems treating humans like "housecats" have been conjured.[12] Global deliberations are also lacking in "global" perspectives. Thought leadership, evidence and deliberation are often concentrated in jurisdictions like the United States, United Kingdom and Europe.[13] The politics of this goes far beyond just regulation and policy – it impacts how we understand, critique, and also build AI systems. The underlying assumptions that guide the design, development and deployment of these systems are context specific, yet globally applied in one direction, from the "global North" towards the "global South". In reality, these systems are far more nascent and the context in which they are deployed significantly more complex.

Complexity of governance frameworks and form

Given the increasingly consequential impact that AI has in societies across the world, there has been a significant push towards articulating the ways in which these systems will be governed, with various frameworks of reference coming to the fore. The extent to which existing regulations in national, regional and international contexts apply to these technologies is unclear, although a closer analysis of data protection regulation,[14] discrimination law[15] and labour law[16] is necessary.

There has been a significant push towards critiquing and regulating these systems on the basis of international human rights standards.[17] Given the impact on privacy, freedom of expression and freedom of assembly, among others, the human rights framework is a minimum requirement to which AI systems must adhere.[18] This can be done by conducting thorough human rights impact assessments of systems prior to deployment,[19] including

9   Surden, S. (2014). Machine Learning and the Law. *Washington Law Review, 89*(1). https://scholar.law.colorado.edu/articles/81

10 For example, image recognition algorithms have shockingly low rates of accuracy for people of colour. See: American Civil Liberties Union Northern California. (2019, 13 August). Facial Recognition Technology Falsely Identifies 26 California Legislators with Mugshots. *American Civil Liberties Union Northern California.* https://www.aclunc.org/news/facial-recognition-technology-falsely-identifies-26-california-legislators-mugshots; AI systems used to screen potential job applicants have also been found to automatically disqualify female candidates. By training a ML algorithm on what successful candidates looked like in the past, the system embeds gender discrimination as a baseline. See: Daston, J. (2018, 10 October). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters.* https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

11 McLendon, K. (2016, 20 August). Artificial Intelligence Could Help End Poverty Worldwide. *Inquisitr.* https://www.inquisitr.com/3436946/artificial-intelligence-could-help-end-poverty-worldwide

12 Solon, O. (2017, 15 February). Elon Musk says humans must become cyborgs to stay relevant. Is he right? *The Guardian.* https://www.theguardian.com/technology/2017/feb/15/elon-musk-cyborgs-robots-artificial-intelligence-is-he-right

13 One just needs to glance through the references to discussions on AI in many high-level documents to see which jurisdictions the evidence backing up claims of AI come from.

14 Wachter, S., & Mittelstadt, B. (2019). A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. *Columbia Business Law Review, 2019*(2). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3248829

15 Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review, 671.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899

16 Rosenblat, A. (2018). *Uberland: How Algorithms are Rewriting the Rules of Work.* University of California Press.

17 ARTICLE 19, & Privacy International. (2018). *Privacy and Freedom of Expression in the Age of Artificial Intelligence.* https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf

18 Kaye, D. (2018). Report of the Special Rapporteur to the General Assembly on AI and its impact on freedom of opinion and expression. https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ReportGA73.aspx

19 Robertson, A. (2019, 10 April). A new bill would force companies to check their algorithms for bias. *The Verge.* https://www.theverge.com/2019/4/10/18304960/congress-algorithmic-accountability-act-wyden-clarke-booker-bill-introduced-house-senate

assessing the legality of these systems against human rights standards, and by industry affirming commitment to the United Nations Guiding Principles on Business and Human Rights.[20]

Social justice is another dominant lens through which AI systems are understood and critiqued. While human rights provide an important minimum requirement for AI systems to adhere to, an ongoing critique of human rights is that they are "focused on securing enough for everyone, are essential – but they are not enough."[21] Social justice advocates are concerned that people are treated in ways consistent with ideals of fairness, accountability, transparency,[22] inclusion, and are free from bias and discrimination. While this is not the appropriate place for an analysis of the relationship between human rights and social justice,[23] suffice to say that in the context of AI, the institutions, frameworks and mechanisms invoked by these two strands of governance are more distinct than they are similar.

A third strand of governance emerges from a development perspective, to have the United Nations' (UN) Sustainable Development Goals (SDGs) guide responsible AI deployment (and in turn use AI to achieve the SDGs),[24] and to leverage AI for economic growth, particularly in countries where technological progress is synonymous with economic progress. There is a pervasive anxiety among countries that they will miss the AI bus, and in turn give up the chance to have unprecedented economic and commercial gain, to "exploit the innovative potential of AI."[25]

The form these various governance frameworks take also varies. Multiple UN mechanisms are currently studying the implications of AI from a human rights and development perspective, including but not limited to the High-level Panel on Digital Cooperation,[26] the Human Rights Council,[27] UNESCO's World Commission on the Ethics of Scientific Knowledge and Technology,[28] and also the International Telecommunication Union's AI for Good Summit.[29] Regional bodies like the European Union High-Level Expert Group on Artificial Intelligence[30] also focus on questions of human rights and principles of social justice like fairness, accountability, bias and exclusion. International private sector bodies like the Partnership on AI[31] and the Institute of Electrical and Electronics Engineers (IEEE)[32] also invoke principles of human rights, social justice and development. All of these offer frameworks that can guide the design, development and deployment of AI by governments, and for companies building AI systems.

Complexity of politics: Power and process

AI systems cannot be studied only on the basis of their deployment. To comprehensively understand the impact of AI in society, we must investigate the processes that precede, influence and underpin deployment, i.e. the process of design and development as well.[33] Who designs these systems, and what contextual reality do these individuals come from? What incentives drive design, and what assumptions guide this stage? Who is being excluded from this stage, and who is overrepresented? What impact does this have on society? On what basis are systems developed and who can peer the process of development? What problems are these technologies built to solve, and who decides and defines the problem? What data is used to train these systems, and who does that data represent?

Much like the models and frameworks of governance that surround AI systems, the process of building AI systems is inherently political. The problem that an algorithm should solve, the data that an algorithm is exposed to, the training that an algorithm goes through, who gets to design and oversee the algorithm's training, the context within which an algorithmic system is built, the context within which an algorithm is deployed, and the ways in which the algorithmic system's findings are applied in imperfect and unequal societies are all political decisions taken by humans.

20 https://www.ohchr.org/documents/publications/ GuidingprinciplesBusinesshr_eN.pdf

21 Moyn, S. (2018). *Not Enough: Human Rights in an Unequal World.* Cambridge: The Belknap Press of Harvard University Press.

22 https://www.fatml.org

23 Lettinga, D. & van Troost, L. (Eds.) (2015). *Can human rights bring social justice?* Amnesty International Netherlands. https://www. amnesty.nl/content/uploads/2015/10/can_human_rights_bring_ social_justice.pdf

24 Chui, M., Chung, R., & van Heteren, A. (2019, 21 January). Using AI to help achieve Sustainable Development Goals. *United Nations Development Programme.* https://www.undp.org/content/undp/ en/home/blog/2019/Using_AI_to_help_achieve_Sustainable_ Development_Goals.html

25 Artificial Intelligence for Development. (2019). Government Artificial Intelligence Readiness Index 2019. https://ai4d.ai/ index2019

26 https://digitalcooperation.org

27 https://www.ohchr.org/en/hrbodies/hrc/pages/home.aspx

28 UNESCO COMEST. (2019). *Preliminary Study on the Ethics of Artificial Intelligence.* https://unesdoc.unesco.org/ark:/48223/ pf0000367823

29 https://aiforgood.itu.int

30 https://ec.europa.eu/digital-single-market/en/high-level-expert- group-artificial-intelligence

31 https://www.partnershiponai.org

32 https://standards.ieee.org/industry-connections/ec/autonomous- systems.html

33 Marda, V. (2018). Artificial Intelligence Policy in India: A Framework for Engaging the Limits of Data-Driven Decision-Making. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2133). https://doi. org/10.1098/rsta.2018.0087

Take, for instance, an algorithmic system that is used to aid law enforcement in allocating resources for policing by studying past patterns of crime. At first glance, this may seem like an efficient solution to a complicated problem that can be applied at scale. However, a closer look will reveal that each step of this process is profoundly political. The data used to train these algorithms is considered ground truth. However, it represents decades of criminal activity defined and institutionalised by humans with their own unique biases. The choice of data sets is also political – training data is rarely representative of the world. It is more often than not selectively built from certain locations and demographics, painting a subjective picture of all crime in a particular area. Data is also not equally available – certain types and demographics are reported and scrutinised more than others.

Drawing from the example of predictive policing, the impact of AI systems redistributes power in visible ways. It is not an overstatement to say that AI fundamentally reorients the power dynamics between individuals, societies, institutions and governments.

It is helpful to lay down the various ways and levels at which power is concentrated, leveraged and imposed by these systems. By producing favourable outcomes for some sections of society, or by having disproportionate impact on certain groups within a society, the ways in which people navigate everyday life is significantly altered. The ways in which governments navigate societal problems is also significantly altered, given the widespread assumption that using AI for development is inherently good. While there is a tremendous opportunity in this regard, it is imperative to be conscientious of the inherent limitations of AI systems, and their imperfect and often harmful overlap with textured and imperfect societies and economies. AI systems are primarily developed by private companies which train and analyse data on the basis of assumptions that are not always legal or ethical, profoundly impacting rights such as privacy and freedom of expression. This essentially makes private entities arbiters of constitutional rights and public functions in the absence of appropriate accountability mechanisms. This link between private companies and public function power was most visibly called out through the #TechWontBuildIt movement, where engineers at the largest technology companies refused to build problematic technology that would be used by governments to undermine human rights and dignity.[34] The design

and development of AI systems is also concentrated in large companies (mostly from the United States and increasingly from China).[35] However, deployment of technology is often *imposed* on jurisdictions in the global South, either on the pretext of pilot projects,[36] or economic development[37] and progress. These jurisdictions are more often than not excluded from the table at stages of design and development, but are the focus of deployment.

Current conversations around AI are overwhelmingly dominated by a multiplicity of efforts and initiatives in developed countries, each coming through with a set of incentives, assumptions and goals in mind. While governance systems and safeguards are built in these jurisdictions, ubiquitous deployment and experimentation occur in others who are not part of the conversation. Yet the social realities and cultural setting in which systems are designed and developed differ significantly from the societies in which they are deployed. Given wide disparity in legal protections, societal values, institutional mechanisms and infrastructural access, this is unacceptable at best and dangerous at worst. There is a growing awareness of the need to understand and include voices from the global South; however, current conversations are deficient for two reasons. First, there is little recognition of the value of conversations that are happening in the global South. And second, there is little, if any, engagement with the nuance of what the "global South" means.

Conclusion

Here, I offer two provocations for researchers in the field, in the hope that they inspire more holistic, constructive and global narratives moving forward:

The global South is not monolithic, and neither are the effects of AI systems. The global South is a complex term. Boaventura de Sousa Santos articulates it in the following manner: The global South is not a geographical concept, even though the great majority of its populations live in countries of the Southern hemisphere. The South is rather a metaphor for the human suffering caused by capitalism and colonialism on the global level, as well as for the resistance to overcoming or minimising such suffering. It is, therefore, an anti-capitalist,

34 O'Donovan, C. (2018, 27 August). Clashes Over Ethics At Major Tech Companies Are Causing Problems For Recruiters. *BuzzFeed News.* https://www.buzzfeednews.com/article/carolineodonovan/silicon-valley-tech-companies-recruiting-protests-ethical

35 See, for example, the country report on China in this edition of GISWatch.

36 Vincent, J. (2018, 6 June). Drones taught to spot violent behavior in crowds using AI. *The Verge.* https://www.theverge.com/2018/6/6/17433482/ai-automated-surveillance-drones-spot-violent-behavior-crowds

37 Entrepreneur. (2019, 25 June). Artificial Intelligence Is Filling The Gaps In Developing Africa. *Entrepreneur South Africa.* https://www.entrepreneur.com/article/337223

anti-colonialist, anti-patriarchal and anti-imperialist South. It is a South that also exists in the geographic North (Europe and North America), in the form of excluded, silenced and marginalised populations, such as undocumented immigrants, the unemployed, ethnic or religious minorities, and victims of sexism, homophobia, racism and Islamophobia.[38]

The "global South" is thus dispersed across geography, demographics and opportunity. It must be afforded the same level of deliberation and nuance as those jurisdictions setting the tone and pace for this conversation. It is incumbent on scholars, researchers, states and companies to understand the ways in which AI systems need to adapt to contexts that are lesser known, in a bottom-up, context-driven way. To continually impose technology on some parts of the world without questioning local needs and nuance, is to perpetuate the institutions of colonialism and racism that we fight so hard to resist. The fact that AI systems need to be situated in context is well understood in current debates. However, "context" necessarily denotes a local, nuanced, granular, bottom-up understanding of the issues at play. Treating the global South "context" as one that is monolithic and generally the opposite of the global North means that we lose valuable learnings and important considerations. A similar shortcoming involves generalising findings about AI systems in one context as ground truth across contexts – which requires a reminder that much like the "global South", AI is not a monolithic sociotechnical system either. The institutional reality within which systems function, along with infrastructural realities, cultural norms, and legal and governance frameworks are rarely, if ever, applicable across contexts.

The governance and politics of AI suffer from fundamental structural inequalities. At present, jurisdictions from the global South do not form part of the evidence base on which AI governance is built. As a result, considerations from the global South are simply added in retrospect to ongoing conversations, if at all. This is an inherent deficiency. Given the invisible yet consequential ways in which AI systems operate, it is crucial to spend time building evidence of what these systems look like in societies across the world. Narratives around AI that inform governance models need to be driven in a bottom-up, local-to-global fashion that looks at different contexts with the same level of granularity in the global South as was afforded to the global North. Much like AI systems operate in societies that have underlying structural inequalities, the deliberation around AI suffers from a similar underlying structural problem. It is incumbent on researchers, policy makers, industry and civil society to engage with the complexities of the global South. Failing this, we risk creating a space that looks very much like the opaque, inscrutable, discriminatory and exclusive systems we aim to improve in our daily work. This edition of GISWatch attempts to start creating an evidence base that nudges conversations away from that risk.

---

38 de Sousa Santos, B. (2016). Epistemologies of the South and the future. *From the European South, 1*, 17-29; also see Arun, C. (2019). AI and the Global South: Designing for Other Worlds. Draft chapter from *Oxford Handbook of Ethics of AI,* forthcoming in 2019.

Artificial intelligence:
Human rights, social justice and development

Artificial intelligence (AI) is now receiving unprecedented global attention as it finds widespread practical application in multiple spheres of activity. But what are the human rights, social justice and development implications of AI when used in areas such as health, education and social services, or in building "smart cities"? How does algorithmic decision making impact on marginalised people and the poor?

This edition of Global Information Society Watch (GISWatch) provides a perspective from the global South on the application of AI to our everyday lives. It includes 40 country reports from countries as diverse as Benin, Argentina, India, Russia and Ukraine, as well as three regional reports. These are framed by eight thematic reports dealing with topics such as data governance, food sovereignty, AI in the workplace, and so-called "killer robots".

While pointing to the positive use of AI to enable rights in ways that were not easily possible before, this edition of GISWatch highlights the real threats that we need to pay attention to if we are going to build an AI-embedded future that enables human dignity.

APC   ARTICLE 19   Sida