

GLOBAL INFORMATION SOCIETY WATCH 2019

GISWatch 2019
SNEAK PEEK

Artificial intelligence: Human rights, social justice and development



ASSOCIATION FOR PROGRESSIVE COMMUNICATIONS (APC),
ARTICLE 19, AND SWEDISH INTERNATIONAL DEVELOPMENT COOPERATION AGENCY (SIDA)

Table of contents

Introduction	3
Vidushi Marda ARTICLE 19	
Decolonising AI: A transfeminist approach to data and social justice	8
Paz Peña and Joana Varon	
Radicalising the AI governance agenda	13
Anita Gurumurthy and Nandini Chami IT FOR CHANGE	
Country and regional reports introduction	18
Alan Finlay	
Korea, Republic of	22
Korean Progressive Network Jinbo net - MIRU	
South Africa	26
Human Sciences Research Council - PAUL PLANTINGA, RACHEL ADAMS AND SAHIER PARKE	

Introduction

Vidushi Marda¹

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/iss_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 de-

liberations 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 law¹⁵ and labour law¹⁶ 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 Knowl-

edge 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 Intelligence³⁰ 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 AI³¹ 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 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/GuidingprinciplesBusinessshr_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 development³⁷ 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,

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>

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>

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.

Decolonising AI: A transfeminist approach to data and social justice

Paz Peña¹ and Joana Varon²

www.pazpena.com
www.codingrights.org

Introduction

Let's say you have access to a database with information from 12,000 girls and young women between 10 and 19 years old, who are inhabitants of some poor province in South America. Data sets include age, neighbourhood, ethnicity, country of origin, educational level of the household head, physical and mental disabilities, number of people sharing a house, and whether or not they have running hot water among their services. What conclusions would you extract from such a database? Or, maybe the question should be: Is it even desirable to make any conclusion at all? Sometimes, and sadly more often than not, simply the possibility of extracting large amounts of data is a good enough excuse to "make them talk" and, worst of all, make decisions based on that.

The database described above is real. And it is used by public authorities to prevent school drop-outs and teenage pregnancy. "Intelligent algorithms allow us to identify characteristics in people that could end up with these problems and warn the government to work on their prevention,"³ said a Microsoft Azure representative. The company is responsible for the machine-learning system used in the *Plataforma Tecnológica de Intervención Social* (Technological Platform for Social Intervention), set up by the Ministry of Early Childhood in the Province of Salta, Argentina.

"With technology, based on name, surname and address, you can predict five or six years ahead which girl, or future teenager, is 86% predestined to have a teenage pregnancy," declared Juan Manuel

Urtubey, a conservative politician and governor of Salta.⁴ The province's Ministry of Early Childhood worked for years with the anti-abortion NGO Fundación CONIN⁵ to prepare this system.⁶ Urtubey's declaration was made in the middle of a campaign for legal abortion in Argentina in 2018, driven by a social movement for sexual rights that was at the forefront of public discussion locally and received a lot of international attention.⁷ The idea that algorithms can predict teenage pregnancy before it happens is the perfect excuse for anti-women⁸ and anti-sexual and reproductive rights activists to declare abortion laws unnecessary. According to their narratives, if they have enough information from poor families, conservative public policies can be deployed to predict and avoid abortions by poor women. Moreover, there is a belief that, "If it is recommended by an algorithm, it is mathematics, so it must be true and irrefutable."

It is also important to point out that the database used in the platform only has data on females. This specific focus on a particular sex reinforces patriarchal gender roles and, ultimately, blames female teenagers for unwanted pregnancies, as if a child could be conceived without a sperm.

For these reasons, and others, the Plataforma Tecnológica de Intervención Social has received much criticism. Some have called the system a "lie", a "hallucination", and an "intelligence that does not think", and have said that the sensitive data of poor women and children is at risk.⁹ A very complete technical analysis of the system's failures

- 1 Paz Peña is an independent consultant on tech, gender and human rights.
- 2 Joana Varon is the executive director of Coding Rights and an affiliate of the Berkman Klein Center for Internet and Society at Harvard University.
- 3 Microsoft. (2018, 2 April). Avanza el uso de la Inteligencia Artificial en la Argentina con experiencias en el sector público, privado y ONGs. *News Center Microsoft Latinoamérica*. <https://news.microsoft.com/es-xl/avanza-el-uso-de-la-inteligencia-artificial-en-la-argentina-con-experiencias-en-el-sector-publico-privado-y-ongs>

- 4 Sternik, I. (2018, 20 April). La inteligencia que no piensa. *Página 12*. <https://www.pagina12.com.ar/109080-la-inteligencia-que-no-piensa>
- 5 Vallejos, S. (2018, 25 August). Cómo funciona la Fundación Conin, y qué se hace en los cientos de centros que tiene en el país. *Página 12*. <https://www.argentina.indymedia.org/2018/08/25/como-funciona-la-fundacion-conin-y-que-se-hace-en-los-cientos-de-centros-que-tiene-en-el-pais>
- 6 Microsoft. (2018, 2 April). Op. cit.
- 7 Goñi, U. (2018, 9 August). Argentina senate rejects bill to legalise abortion. *The Guardian*. <https://www.theguardian.com/world/2018/aug/09/argentina-senate-rejects-bill-legalise-abortion>
- 8 Cherwitz, R. (2019, 24 May). Anti-Abortion Rhetoric Mislabeled "Pro-Life". *The Washington Spectator*. <https://washingtonspectator.org/cherwitz-anti-abortion-rhetoric>
- 9 Sternik, I. (2018, 20 April). Op. cit.

was published by the Laboratorio de Inteligencia Artificial Aplicada (LIAA) at the University of Buenos Aires.¹⁰ According to LIAA, which analysed the methodology posted on GitHub by a Microsoft engineer,¹¹ the results were overstated due to statistical errors in the methodology. The database was also found to be biased due to the inevitable sensitivities of reporting unwanted pregnancies, and the data inadequate to make reliable predictions.

Despite this, the platform continued to be used. And worse, bad ideas dressed up as innovation spread fast: the system is now being deployed in other Argentinian provinces, such as La Rioja, Tierra del Fuego and Chaco,¹² and has been exported to Colombia and implemented in the municipality of La Guajira.¹³

The Plataforma Tecnológica de Intervención Social is just one very clear example of how artificial intelligence (AI) solutions, which their implementers claim are neutral and objective, have been increasingly deployed in some countries in Latin America to support potentially discriminatory public policies that undermine human rights of unprivileged people. As the platform shows, this includes monitoring and censoring women and their sexual and reproductive rights.

We believe that one of the main causes for such damaging uses of machine learning and other AI technologies is a blind belief in the hype that big data will solve several burning issues faced by humankind. Instead, we propose to build a transfeminist¹⁴ critique and framework that offers not only the potential to analyse the damaging effects of AI, but also a proactive understanding on how to imagine, design and develop an emancipatory AI that undermines consumerist, misogynist, racist, gender binarial and heteropatriarchal societal norms.

Big data as a problem solver or discrimination disguised as math?

AI can be defined in broad terms as technology that makes predictions on the basis of the automatic detection of data patterns.¹⁵ As in the case of the government of Salta, many states around the world are increasingly using algorithmic decision-making tools to determine the distribution of goods and services, including education, public health services, policing and housing, among others. Moreover, anti-poverty programmes are being datafied by governments, and algorithms used to determine social benefits for the poor and unemployed, turning “the lived experience of poverty and vulnerability into machine-readable data, with tangible effects on the lives and livelihoods of the citizens involved.”¹⁶

Cathy O’Neil, analysing the usages of AI in the United States (US), asserts that many AI systems “tend to punish the poor.” She explains:

This is, in part, because they are engineered to evaluate large numbers of people. They specialize in bulk, and they’re cheap. That’s part of their appeal. The wealthy, by contrast, often benefit from personal input. [...] The privileged, we’ll see time and again, are processed more by people, the masses by machines.¹⁷

AI systems are based on models that are abstract representations, universalisations and simplifications of complex realities where much information is being left out according to the judgment of their creators. O’Neil observes:

[M]odels, despite their reputation for impartiality, reflect goals and ideology. [...] Our own values and desires influence our choices, from the data we choose to collect to the questions we ask. Models are opinions embedded in mathematics.¹⁸

In this context, AI will reflect the values of its creators, and thus many critics have concentrated on the necessity of diversity and inclusivity:

So inclusivity matters – from who designs it to who sits on the company boards and which ethical perspectives are included. Otherwise, we

10 Laboratorio de Inteligencia Artificial Aplicada. (2018). *Sobre la predicción automática de embarazos adolescentes*. <https://liaa.dc.uba.ar/es/sobre-la-prediccion-automatica-de-embarazos-adolescentes>

11 Davancens, F. (n.d.). Predicción de Embarazo Adolescente con Machine Learning. <https://github.com/facundod/case-studies/blob/master/Prediccion%20de%20Embarazo%20Adolescente%20con%20Machine%20Learning.md>

12 Ponce Mora, B. (2019, 27 March). “Primera Infancia es el ministerio que defiende a los niños desde su concepción”. *El Tribuno*. <https://www.eltibuno.com/salta/nota/2019-3-27-0-39-0--primera-infancia-es-el-ministerio-que-defiende-a-los-ninos-desde-su-concepcion>

13 Ministerio de la Primera Infancia. (2018, 14 June). Comisión oficial. Departamento de la Guajira, República de Colombia. *Boletín Oficial Salta*. boletinoficialsalta.gob.ar/NewDetalleDecreto.php?nro_decreto=658/18

14 We refer to transfeminism as an epistemological tool that, as Sayak Valencia acknowledges, has as its main objective to re-politicise and de-essentialise global feminist movements that have been used to legitimise policies of exclusion on the basis of gender, migration, miscegenation, race and class. See Valencia, S. (2018). El transfeminismo no es un generismo. *Pléyade (Santiago)*, 22, 27-43. <https://dx.doi.org/10.4067/S0719-36962018000200027>

15 Daly, A., et al. (2019). *Artificial Intelligence Governance and Ethics: Global Perspectives*. The Chinese University of Hong Kong, Faculty of Law. Research Paper No. 2019-15.

16 Masiero, S., & Das, S. (2019). Datafying anti-poverty programmes: implications for data justice. *Information, Communication & Society*, 22(7), 916-933.

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

18 Ibid.

risk constructing machine intelligence that mirrors a narrow and privileged vision of society, with its old, familiar biases and stereotypes.¹⁹

But diversity and inclusivity are not enough to create an emancipatory AI. If we follow Marcuse's ideas that "the technological mode of production is a specific form or set of conditions which our society has taken among other possible conditions, and it is this mode of production which plays the ultimate role in shaping techniques, as well as directing their deployment and proliferation,"²⁰ it is fundamental to dive deeply into what the ruling interests of this historical-social project are. In this sense, theories of data justice have reflected on the necessity to explicitly connect a social justice agenda to the data revolution supported by some states, companies and international agencies in order to achieve fairness in the way people are seen and treated by the state and by the private sector, or when they act together.²¹

For example, as Payal Arora frames it, discourses around big data have an overwhelmingly positive connotation thanks to the neoliberal idea that the exploitation for profit of the poor's data by private companies will only benefit the population.²² This is, in many ways, the sign that two old acquaintances, capitalism and colonialism, are present and healthy every time an AI system strips people of their autonomy and treats them "as mere raw data for processing."²³ Along the same lines, Couldry and Mejias²⁴ consider that the appropriation and exploitation of data for value has deep roots in capitalism and colonialism.

Recently, connecting this critique to the racialisation of citizens and communities through algorithmic decisions, Safiya Umoja Noble has coined the term "technological redlining", which refers to the process of data discrimination that

bolsters inequality and oppression. The term draws on the "redlining" practice in the US by which communities suffered systematic denial of various services either directly or through the selective raising of prices based on their race:

I think people of color will increasingly experience it as a fundamental dimension of generating, sustaining, or deepening racial, ethnic and gender discrimination. This process is centrally tied to the distribution of goods and services in society, like education, housing and other human and civil rights, which are often determined now by software, or algorithmic decision-making tools, which might be popularly described as "artificial intelligence".²⁵

The question is how conscious of this citizens and public authorities who are purchasing, developing and using these systems are. The case of Salta, and many others, show us explicitly that the logic of promoting big data as the solution to an unimaginable array of social problems is being exported to Latin America, amplifying the challenges of decolonisation. This logic not only corners attempts to criticise the status quo in all the realms of power relations, from geopolitics, to gender norms and capitalism, but also makes it more difficult to sustain and promote alternative ways of life.

AI, poverty and stigma

"The future is today." That seems to be the mantra when public authorities eagerly adopt digital technologies without any consideration of critical voices that show their effects are potentially discriminatory. In recent years, for example, the use of big data for predictive policing seems to be a popular tendency in Latin America. In our research we found that different forms of these AI systems have been used (or are meant to be deployed) in countries such as Argentina, Brazil, Chile, Colombia, Mexico and Uruguay, among others.²⁶ The most common model is building predictive maps of crime, but there have also been efforts to develop predictive models of likely perpetrators of crime.²⁷

19 Crawford, K. (2016, 25 June). Artificial Intelligence's White Guy Problem. *The New York Times*. <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>

20 Kidd, M. (2016). Technology and nature: a defence and critique of Marcuse. *POLIS*, 4(14). <https://revistapolis.ro/technology-and-nature-a-defence-and-critique-of-marcuse>

21 Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, July-December, 1-14. <https://journals.sagepub.com/doi/10.1177/2053951717736335>

22 Arora, P. (2016). The Bottom of the Data Pyramid: Big Data and the Global South. *International Journal of Communication*, 10, 1681-1699.

23 Birhane, A. (2019, 18 July). The Algorithmic Colonization of Africa. *Real Life Magazine*. <https://www.reallifemag.com/the-algorithmic-colonization-of-africa>

24 Couldry, N., & Mejias, U. (2019). Data colonialism: rethinking big data's relation to the contemporary subject. *Television and New Media*, 20(4), 336-349.

25 Bulut, E. (2018). Interview with Safiya U. Noble: Algorithms of Oppression, Gender and Race. *Moment Journal*, 5(2), 294-301. <https://dergipark.org.tr/download/article-file/653368>

26 Serrano-Berthet, R. (2018, 10 May). ¿Cómo reducir el delito urbano? Uruguay y el "leap frogging" inteligente. *Sin Miedos*. <https://blogs.iadb.org/seguridad-ciudadana/es/reducir-el-delito-urbano-uruguay/>

27 Van 't Wout, E., et al. (2018). Capítulo II. Big data para la identificación de comportamiento criminal. In I. Irarrázaval et al. (Eds.), *Propuestas para Chile*. Pontificia Universidad Católica de Chile.

As Fieke Jansen suggests:

These predictive models are based on the assumption that when the underlying social and economic conditions remain the same crime spreads as violence will incite other violence, or a perpetrator will likely commit a similar crime in the same area.²⁸

Many critics point to the negative impacts of predictive policing on poorer neighbourhoods and other affected communities, including police abuse,²⁹ stigmatisation, racism and discrimination.³⁰ Moreover, as a result of much of the criticism, in the US, where these systems have been deployed for some time, many police agencies are reassessing the real efficiency of the systems.³¹

The same logic behind predictive policing is found in anti-poverty AI systems that collect data to predict social risks and deploy government programmes. As we have seen, this is the case with the Plataforma Tecnológica de Intervención Social; but it is also present in systems such as *Alerta Infancia* in Chile. Again, in this system, data predictions are applied to minors in poor communities. The system assigns risk scores to communities, generating automated protection alerts, which then allow “preventive” interventions. According to official information,³² this platform defines the risk index by factors such as teenage pregnancy, the problematic use of alcohol and/or drugs, delinquency, chronic psychiatric illness, child labour and commercial sexual exploitation, mistreatment or abuse and dropping out of school. Among much criticism of the system, civil society groups working on child rights declared that, beyond surveillance, the system “constitutes the imposition of a certain form of sociocultural normativity,” as well as “encouraging and socially validating forms of stigmatisation, discrimination and even criminalisation of the cultural diversity existing in Chile.” They stressed:

This especially affects indigenous peoples, migrant populations and those with lower economic incomes, ignoring that a growing cultural diversity demands greater sensitivity, visibility and respect, as well as the inclusion of approaches with cultural relevance to public policies.³³

There are at least three common characteristics in these systems used in Latin America that are especially worrisome given their potential to increase social injustice in the region: one is the identity forced onto poor individuals and populations. This quantification of the self, of bodies (understood as socially constructed) and communities has no room for re-negotiation. In other words, datafication replaces “social identity” with “system identity”.³⁴

Related to this point, there is a second characteristic that reinforces social injustice: the lack of transparency and accountability in these systems. None of them have been developed through a participative process of any type, whether including specialists or, even more important, affected communities. Instead, AI systems seem to reinforce top-down public policies from governments that make people “beneficiaries” or “consumers”: “As Hacking referred to ‘making up people’ with classification, datafication ‘makes’ beneficiaries through census categories that are crystallised through data and made amenable to top-down control.”³⁵

Finally, these systems are developed in what we would call “neoliberal consortiums”, where governments develop or purchase AI systems developed by the private sector or universities. This deserves further investigation, as neoliberal values seem to pervade the way AI systems are designed, not only by companies, but by universities funded by public funds dedicated to “innovation” and improving trade.³⁶

Why a transfeminist framework?

As we have seen, in these examples of the use of these types of technologies, some anti-poverty government programmes in Latin America reflect a positivist framework of thinking, where reality seems to be better understood and changed for good if we

28 Jansen, F. (2018). *Data Driven Policing in the Context of Europe*. <https://www.datajusticeproject.net/wp-content/uploads/sites/30/2019/05/Report-Data-Driven-Policing-EU.pdf>

29 Ortiz Freuler, J., & Iglesias, C. (2018). *Algoritmos e Inteligencia Artificial en Latinoamérica: Un Estudio de implementaciones por parte de Gobiernos en Argentina y Uruguay*. World Wide Web Foundation. https://webfoundation.org/docs/2018/09/WF_AI-in-LA_Report_Spanish_Screen_AW.pdf

30 Crawford, K. (2016, 25 June). Op. cit.

31 Puente, M. (2019, 5 July). Police Leaders Debate Merits of Using Data to Predict Crime. *Government Technology*. <https://www.govtech.com/public-safety/Police-Leaders-Debate-Merits-of-Using-Data-to-Predict-Crime.html>

32 Ministerio de Desarrollo Social. (2018). *Piloto Oficina Local de la Niñez*. www.planderechoshumanos.gob.cl/files/attachment/d41d8cd98f00b204e9800998ecf8427e/phpEfF4QP/original.pdf

33 Sociedad Civil de Chile Defensora de los Derechos Humanos del Niño et al. (2019, 28 January). Día Internacional de la protección de datos. Carta abierta de la Sociedad Civil de Chile Defensora de los Derechos Humanos del Niño. *ONG Emprender con Alas*. <https://www.emprenderconalas.cl/2019/01/28/dia-internacional-de-la-proteccion-de-datos-carta-abierta-de-la-sociedad-civil-de-chile-defensora-de-los-derechos-humanos-del-nin>

34 Arora, P. (2016). Op. cit.

35 Masiero, S., & Das, S. (2019). Op. cit.

36 Esteban, P. (2019, 18 September). Diego Hurtado: “El discurso del científico emprendedor es una falacia”. *Página 12*. <https://www.pagina12.com.ar/218802-diego-hurtado-el-discurso-del-cientifico-emprendedor-es-una->

can quantify every aspect of our life. This logic also promotes the vision that what humans shall seek is “progress”, which is seen as a synonym of augmented production and consumption, and ultimately means exploitation of bodies and territories.

All these numbers and metrics about unprivileged people’s lives are collected, compiled and analysed under the logic of “productivity” to ultimately maintain capitalism, heteropatriarchy, white supremacy and settler colonialism. Even if the narrative of the “quantified self” seems to be focused on the individual, there is no room for recognising all the different layers that human consciousness can reach, nor room for alternative ways of being or fostering community practices.

It is necessary to become conscious of how we create methodological approaches to data processing so that they challenge these positivist frameworks of analysis and the dominance of quantitative methods that seem to be gaining fundamental focus in the development and deployment of today’s algorithms and processes of automated decision making.

As Silvia Rivera Cusicanqui says:

How can the exclusive, ethnocentric “we” be articulated with the inclusive “we” – a homeland for everyone – that envisions decolonization? How have we thought and problematized, in the here and now, the colonized present and its overturning?³⁷

Beyond even a human rights framework, decolonial and transfeminist approaches to technologies are great tools to envision alternative futures and overturn the prevailing logic in which AI systems are being deployed. Transfeminist values need to be embedded in these systems, so advances in the development of technology help us understand and

break what black feminist scholar Patricia Hill Collins calls the “matrix of domination”³⁸ (recognising different layers of oppression caused by race, class, gender, religion and other aspects of intersectionality). This will lead us towards a future that promotes and protects not only human rights, but also social and environmental justice, because both are at the core of decolonial feminist theories.

Re-imagining the future

To push this feminist approach into practice, at Coding Rights, in partnership with MIT’s Co-Design Studio,³⁹ we have been experimenting with a game we call the “Oracle for Transfeminist Futures”.⁴⁰ Through a series of workshops, we have been collectively brainstorming what kind of transfeminist values will inspire and help us envision speculative futures. As Ursula Le Guin once said:

The thing about science fiction is, it isn’t really about the future. It’s about the present. But the future gives us great freedom of imagination. It is like a mirror. You can see the back of your own head.⁴¹

Indeed, tangible proposals for change in the present emerged once we allowed ourselves to imagine the future in the workshops. Over time, values such as agency, accountability, autonomy, social justice, non-binary identities, cooperation, decentralisation, consent, diversity, decoloniality, empathy, security, among others, emerged in the meetings.

Analysing just one or two of these values combined⁴² gives us a tool to assess how a particular AI project or deployment ranks in terms of a decolonial feminist framework of values. Based on this we can propose alternative technologies or practices that are more coherent given the present and the future we want to see.

37 Rivera Cusicanqui, S. (2012). Ch’ixinakax utxiwa: A Reflection on the Practices and Discourses of Decolonization. *The South Atlantic Quarterly*, 111(1), 95-109.

38 Collins, P. H. (2000). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. New York: Routledge.

39 <https://codesign.mit.edu>

40 <https://www.transfeministech.codingrights.org>

41 Le Guin, U. K. (2019). *Ursula K. Le Guin: The Last Interview and Other Conversations*. Melville House.

42 Peña, P., & Varon, J. (2019). *Consent to our Data Bodies: Lessons from feminist theories to enforce data protection*. Privacy International. <https://codingrights.org/docs/ConsentToOurDataBodies.pdf>

Radicalising the AI governance agenda¹

Anita Gurumurthy and Nandini Chami

IT for Change

www.ITforChange.net

What's missing in mainstream global debates on AI governance

Advances in artificial intelligence (AI) present human civilisation with challenges that are unprecedented. As a class of technologies² that simulate human intelligence processes for learning, reasoning and self-correction, AI disrupts the way societies define, organise and use knowledge, thus radically recasting social and economic systems. Understanding and deconstructing AI systems that are self-learning and self-correcting is not easy. In fact, experts in the field have even stated that it is impossible. The widespread diffusion and adoption of AI, even if much of it for now is so-called “narrow AI”,³ is therefore as terrifying as it is exciting – something that Bill Gates has compared to the complexity of nuclear technology. Quite naturally, a vibrant debate on the governance of AI has been gathering momentum, involving governments, multilateral institutions, technology companies, the technical community and global civil society. The search is on for the right combination of legal-regulatory, ethical and technological approaches that constitute effective AI governance.

Mainstream debates on AI governance take note of violations of the human rights considerations of privacy, equality and non-discrimination, uncertain futures of work, and erosion of democracy in the emerging AI paradigm. They do not, however, fully address the entanglement of AI in neoliberal capitalism and what this means for the life-chances of

individuals and communities. Because of this, AI governance debates tend to carry critical blind spots.

Blind spot 1: Collective autonomy and choice in the debate on AI and human rights

Across stakeholders, there is growing acknowledgement of how AI systems could undermine human rights. A systematic mapping of the over 32 sets of influential AI principles/guidelines in existence today by the Cyber Harvard project reveals that informational privacy, equality, fairness and freedom from discrimination are critical concerns shared by all stakeholders involved in the development and deployment of AI technologies: governments, multilateral organisations, advocacy groups and technology companies.⁴ The inscrutability of AI systems means that the subjectivity of their creators can reinforce the very biases that create an unequal society, leading to a due process failure. Inherent biases in input/training data sets as well as in definitions of output parameters produce unfair outcomes.

Institutional and techno-governance mechanisms to address bias in AI are indeed necessary to tackle inequality and discrimination. However, existing proposals in this regard, whether from multilateral agencies (such as the global legal framework mooted by the UN Special Rapporteur on freedom of expression in his 2018 report),⁵ or plurilateral bodies (the OECD Council's Recommendation on Artificial Intelligence),⁶ or governments (the European Commission's Ethics Guidelines for Trustworthy AI),⁷ or civil society (the Toronto Declaration⁸ for protecting equality and non-discrimination in AI systems), or the technical community (such as IEEE's project on evolving an open standard on algorithmic bias), tend to focus exclusively on addressing misrecognition.

1 This report has been adapted from “The Wicked Problem of AI Governance”, which will be published by FES-India in October 2019.

2 Ranging from computer vision, natural language processing, virtual assistants and robotic process automation to advanced machine learning. See: Bowles, J. (2018, 18 September). McKinsey warns that AI will further divide the world economy into winners and losers. *Diginomica*. <https://diginomica.com/mckinsey-warns-that-ai-will-further-divide-the-world-economy-into-winners-and-losers>

3 AI used for a narrowly defined task, as opposed to the more complex general or strong AI.

4 Fjeld, J., et al. (2019, 4 July). Principled Artificial Intelligence: A Map of Ethical and Rights-Based Approaches. *Berkman Klein Center for Internet & Society*. <https://ai-hr.cyber.harvard.edu/primp-viz.html>

5 <https://undocs.org/A/73/348>

6 <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

7 <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

8 <https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems>

They fail to imagine redress to individuals and communities caught in relationships of exploitation that are based on uneven and unfair distribution of intelligence capital. In the AI-led economy, algorithmic intelligence extracted from data resources is the “secret sauce”⁹ that enables the disruption of the economic status quo and the attainment of new levels of efficiency. At present, such “intelligence capital” is concentrated in the hands of a few transnational corporations, which have enclosed valuable data resources in order to cement their market dominance by foreclosing the possibility of competing AI innovations emerging in the future.

Because of their failure to address the unequal distribution of intelligence capital and the resultant inequality in opportunity structures, existing AI and human rights proposals ignore the changing structures of choice. We urgently need framings about equality and non-discrimination in relation to AI that are attentive to “equality of autonomy”¹⁰ – the spread across society of the ability and means of people to choose their life course. Our response to safeguarding human rights in the AI paradigm must move beyond identity-based discrimination, and tackle AI-based economic exploitation through new governance approaches for the AI economy that expand individual and collective choices.

Blind spot 2: Economic self-determination in the debate on AI

In the race towards the “Fourth Industrial Revolution”, an ideology of AI-frontierism is widely evidenced in policy circles. Not wanting to be left behind, developing country governments are caught up in the language of “innovation” and “entrepreneurship”, authoring national plans and road maps for their digital start-up ecosystem and upskilling of workers. These efforts view AI-led development as a simplistic aggregate of individual efficiencies that will somehow magically add up to national productivity gains. They completely ignore the fact that development is a “competitive and global undertaking”, characterised by a sustained and continuing effort to capture opportunities for higher value knowledge and technological capabilities.¹¹ In the current context, strides in development

are possible only for countries that can harness AI at a socio-structural level for higher growth and redistributive gains. Developing countries urgently need to use AI to create and/or deepen national capacity for moving out of low value locations in the global value chain. However, the debate so far¹² seems to flatten the global political economy of development with a broad brush stroke, and even glib prescriptions exhorting countries of the South to build their domestic AI capabilities and upskill their populations.

How can these prescriptions be met if access to and ownership of data and digital intelligence are denied to these countries? The AI-led global order is entrenched firmly in what activists and scholars have argued is a form of neocolonisation.¹³ Today, economic power is a function of how AI technologies are employed in networked systems organised around incessant data processing. As data started flowing on a planetary scale with the advent of the internet, creating and multiplying social and economic connections, predatory capitalism found a new lease of life. The value of the global network of connections has since grown exponentially with the emergence of the platform model, the network-data infrastructures that mediate and organise production and exchange on a global scale. In the emerging global AI economy, competitive advantage is determined by the ability to reach higher levels of efficiency through the intelligence capital generated by processing data.

Moving to the higher value segments of the global economy is, however, inordinately difficult in the current global economic order, where corporations and countries who have enjoyed a first-mover advantage in harvesting data for digital intelligence systematically reinforce their position of dominance. As the United Nations Conference on Trade and Development (UNCTAD) Trade and Development Report¹⁴ cautions, the restructuring of global value chains by the platform business model has

9 Morozov, E. (2018, 28 January). Will tech giants move on from the internet, now we've all been harvested? *The Guardian*. <https://www.theguardian.com/technology/2018/jan/28/morozov-artificial-intelligence-data-technology-online>

10 Sen, A. (2001). *Development as Freedom*. Oxford University Press.

11 Mann, L., & Iazzolino, G. (2019). *See, Nudge, Control and Profit: Digital Platforms as Privatized Epistemic Infrastructures*. IT for Change. https://itforchange.net/platformpolitics/wp-content/uploads/2019/03/Digital-Platforms-as-Privatized-Epistemic-Infrastructures-_5thMarch.pdf

12 Smith, M., & Neupane, S. (2018). *Artificial Intelligence and Human Development: Toward a Research Agenda*. International Development Research Centre. <https://idl-bnc-idrc.dspacedirect.org/handle/10625/56949> and World Economic Forum. (2017). *Accelerating Workforce Reskilling for the Fourth Industrial Revolution: An Agenda for Leaders to Shape the Future of Education, Gender and Work*. www3.weforum.org/docs/WEF_EGW_White_Paper_Reskilling.pdf

13 Avila, R. (2018). *Resisting Digital Colonialism*. Mozilla. <https://internethealthreport.org/2018/resisting-digital-colonialism> and Couldry, N. & Mejias, U. (2018). *Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject*. LSE Research Online. https://eprints.lse.ac.uk/89511/1/Couldry_Data-colonialism_Accepted.pdf

14 UNCTAD. (2018). *Trade and Development Report 2018: Power, Platforms and the Free Trade Delusion*. https://unctad.org/en/PublicationsLibrary/trdr2018_en.pdf

coincided with the appearance in global economic statistics of a “widening gap between a small number of big winners in global value chains and a large collection of participants, both smaller companies and workers, who are being squeezed.”¹⁵

The United States (US) and its allies have also sought to use trade negotiations to assert their advantage and maintain the status quo on unrestricted cross-border data flows to protect US platform monopolies. Similarly, they have been stalling demands of developing countries for disclosure of source code/algorithms by transnational digital corporations, even though such technology transfer conditionalities for market access are currently permissible under the Agreement on Trade Related Investment Measures (TRIMs). Without the sovereign right to control the terms on which the data of their citizens or the data generated in their territories flows across jurisdictions and/or the means to build the digital intelligence capabilities to boost their economies, countries in the developing world cannot create the endogenous conditions for their citizens to reap the AI advantage. They will never be able to create the intelligence capital for reaching higher value knowledge capabilities. On the contrary, their vulnerabilities could potentially be accentuated, as the systematic flight of data from their territories for exogenous AI infrastructure models creates economic and political dependencies.

The terms of the debate therefore need to shift away from individualist solutions to secure the future of the economy towards governance frameworks that invoke the economic right of nation states and communities to have sovereignty over data – which may be seen as “a new form of wealth”¹⁶ – to self-determine their development pathways.

Blind spot 3: The realpolitik of algorithmic scrutiny in the debate on norms for digitally mediated democracy

The early consensus on internet exceptionalism linked to free speech seems to be giving way to a realisation that a hyper-extractive algorithmic regime needs new norms that can hold platform intermediaries accountable for preserving democracy in digitally mediated times. There is thus an increasing

acknowledgement about the need for public scrutiny of the algorithmic tools used by platforms for content curation, user profiling and targeting.¹⁷

In the past year, the European Union (EU) has been at the helm of this debate, with members of the European Parliament calling for an algorithmic audit of the profiling practices of Facebook in October 2018 and the establishment of an EU Committee of Ministers to deliberate on safeguards against algorithmic manipulation by platforms, including digital communication services.¹⁸ While the EU – as a politically powerful and economically relevant bloc – may well be able to create the regulatory structures and enforce accountability mechanisms *vis-à-vis* transnational platform companies within its territory, most countries in the global South lack such clout and the institutional wherewithal for regulatory oversight. As mentioned, the US and its allies have also sought to protect the intellectual property interests of their digital corporations in trade-related negotiations, insisting that no country can make market access contingent on source code/algorithmic disclosure.¹⁹ Most developing countries therefore face a Hobson’s choice: they must give in to opaque and unilateral AI-enabled content governance policies and practices of transnational platform companies in order to have access to the essential communications infrastructure that they depend on the latter to provision.

These geo-economic and geo-political dynamics as well as the absence of a binding international framework on the obligations of transnational corporations render the plausibility of effective regulatory intervention by developing countries moot. Ideas of self-regulation tend to gain currency, furthering a user-centred approach that depoliticises the problem, replacing democratic oversight with corporate largesse.

A two-pronged response is necessary to prevent the degeneration of the digitally mediated public sphere. Firstly, the deleterious consequences of “AI-gone-wrong” for democracy cannot be tackled without a right for all countries to scrutinise the algorithmic apparatus shaping social interactions in

¹⁵ Ibid.

¹⁶ PTI. (2019, 28 June). Data ‘new form of wealth’, needs to take into account developing nations’ needs: India. *New Indian Express*. www.newindianexpress.com/world/2019/jun/28/data-new-form-of-wealth-needs-to-take-into-account-developing-nations-needs-india-1996614.html

¹⁷ Garton Ash, T., Gorwa, R., & Metaxa, D. (2019). *GLASNOST! Nine ways Facebook can make itself a better forum for free speech and democracy*. Reuters Institute for the Study of Journalism and University of Oxford. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-01/Garton_Ash_et_al_Facebook_report_FINAL_o.pdf

¹⁸ Koene, A., et al. (2019). *A Governance Framework for Algorithmic Accountability and Transparency*. European Parliamentary Research Service. [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU\(2019\)624262_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624262/EPRS_STU(2019)624262_EN.pdf)

¹⁹ Ibid.

their territory. The proposed international treaty on business and human rights is a highly pertinent instrument²⁰ through which corporate violations that undercut democracy and human rights can be addressed by governments. Additionally, the health of public spheres in digital times hinges on a global agreement, a binding normative framework on data and AI that prescribes duties of states *vis-à-vis* national and global democracy. A reinterpretation of human rights obligations of state and non-state actors in the age of AI, therefore, is not optional: it is an urgent need. A global normative framework for data and AI must also address the issue of data extractivism, setting limits on individual profiling in the online communications sphere.

A radical agenda for AI governance: Building blocks

Violations of the foundational human rights principle of equality and non-discrimination and the thwarting of political and economic democracy in the AI paradigm are, evidently, a result of data imperialism – the control that algorithmic circuits of digital intelligence confer on the already powerful who own the data. Surprisingly though, this facet of AI is hardly alluded to in the debates on AI governance, which – as demonstrated above – propose liberalist, structural interventions (focusing on correcting misrecognition but not maldistribution) at best and neoliberal, individualistic fixes (that transfer burdens of navigating the digital economy on individuals) at worst. When viewed from this standpoint, the contours of the AI governance debate shift significantly. It becomes apparent that transforming the political economy of data ownership and control that is deepening global development fault lines is the critical missing link. The AI governance agenda therefore needs to be transformed and radicalised, embracing a focus on data and AI constitutionalism.

Two critical steps need to be accomplished for such a radical departure:

(a) Acknowledging data sovereignty as part of the right to development

In the AI paradigm, without a national-level strategy to leverage data resources for inclusive innovation and social transformation, the right and duty of nation states to formulate appropriate

national development policies as envisaged in the Declaration on the Right to Development cannot be realised. For example, in order to safeguard strategic economic interests, countries must be able to build and strengthen public data pools, mandating private firms to relinquish their exclusive rights over data collected and processed as part of their business where such data is assessed to be of national importance. They must also be able to prevent the enclosure and expropriation of cultural/knowledge commons or community data by transnational digital companies. But in a context where the bulk of data resources of developing countries are in the hands of transnational digital companies headquartered elsewhere, such national-level policy measures can be enforced only by re-asserting jurisdictional sovereignty over data resources through the introduction of restrictions and controls on cross-border data transfers, and data localisation measures. It is this policy space that is currently being taken away by advanced AI nations who are utilising trade policy avenues to push for the maintenance of the status quo on unrestricted data flows and protect the interests of their corporations. Such tactics also promote a myth that any national-level conditionalities on data flows are likely to impede global flows of information on the internet.

The sovereign right of nation states to the data on their citizens or collected within their territories needs to be articulated through a binding global normative framework on data and AI. Norms about putting AI to the service of human rights and development justice must embrace the cutting-edge wisdom about the inalienability, indivisibility and interdependence of human rights, with a futuristic outlook for the 21st century. To fulfil their human rights obligations in the AI paradigm, states need to implement various measures, balancing multiple interests and priorities in the national context. A sophisticated governance framework for access to and use and control of data is needed that effectively balances the rights of data principals with the rights of those investing in the resources that enable the creation of digital intelligence, the rights of affected individuals/communities, and the broader public interest.²¹

²⁰ For more details, see Zorob, M. (2019, 30 September). The Lengthy Journey towards a Treaty on Business & Human Rights. *Business & Human Rights Resource Centre*. <https://www.business-humanrights.org/en/the-lengthy-journey-towards-a-treaty-on-business-human-rights>

²¹ British Academy, Royal Society, & techUK. (2018). *Data Ownership, Rights and Controls: Reaching a Common Understanding*. <https://royalsociety.org/-/media/policy/projects/data-governance/data-ownership-rights-and-controls-October-2018.pdf> and Scassa, T. (2018). *Data Ownership*. Centre for International Governance Innovation. https://www.cigionline.org/sites/default/files/documents/Paper%20no.187_2.pdf

(b) Reining in transnational digital corporations

Given that the bulk of AI innovation is currently being spearheaded by transnational corporations, norms and rules at the national level are necessary to protect the interests of domestic businesses and enterprises (across a wide spectrum that includes not-for-profits and cooperatives). Policy measures will need to straddle: FRAND (Fair, Reasonable and Non-Discriminatory Access) provisions in technology patenting to prevent digital corporations from locking in essential building blocks of algorithmic innovation;²² foreign direct investment controls in the digital start-up sector to prevent extractivist investments that cannibalise domestic enterprises;²³ regulation for algorithmic audit and

scrutiny to protect the rights to privacy, equality and non-discrimination; and limits on the use of personally identifiable data for hyper-profiling. But the rapacious greed of digital transnational corporations for data, their opacity about algorithms and brazen non-compliance with domestic regulation are issues that require an international mechanism to enforce corporate accountability. Although some progress has been made in deliberating a legally binding instrument on transnational corporations and business enterprises with respect to human rights, this process has not gathered momentum owing to the clout that transnational corporations enjoy. The need for progress on this front cannot be overemphasised.

22 4iP Council. (2018). *A FRAND Regime for Dominant Digital Platforms? Contribution by 4iP Council to the European Commission's Workshop on Shaping Competition Policy in the Era of Digitisation*. https://ec.europa.eu/competition/information/digitisation_2018/contributions/4ip_council.pdf

23 Ciuriak, D. (2018, 15 November). *Industrial-era Investment Strategies Won't Work in a Data-driven Economy*. *Centre for International Governance Innovation*. <https://www.cigionline.org/articles/industrial-era-investment-strategies-wont-work-data-driven-economy>

Country and regional reports introduction

Alan Finlay

Flawed digital technologies are increasingly at the core of our daily activities, and they interact with us. – Franco Giandana (Creative Commons Argentina/Universidad Nacional de Córdoba)

The 43 reports published here show there are few areas where the potential of artificial intelligence (AI) is not being explored. Even in so-called “least developed countries”, AI experiments and programmes are proliferating. For example, in Rwanda, “innovation companies [are] attracted by [it] being a ‘proof-of-concept’ country where people who are thinking about setting up businesses are offered a place to build and test prototypes before scaling to other countries.” In Benin, among several AI pilots including big data labs, training drones to work in areas such as health, agriculture and conservation, and an annual contest to combine algorithms with local games such as *adji* (dominoes), at least two initiatives in the country focus on empowering women and girls in the use of robotics and AI. “Despite the lack of an enabling environment,” writes Abebe Chekol (Internet Society – Ethiopian Chapter), “the country is becoming a thriving centre for AI research and development.”

The authors take a loose definition of AI, and in doing so cast a relatively wide net on what they consider relevant for discussion. What all of the reports have in common, however, is a focus on when AI – variously defined – meets the intersection of human rights, social justice and development, and “shocks” this intersection; sometimes for the better, but also often raising critical issues that demand the attention of human rights advocates. While the focus in these reports is on perspectives from the global South, reports from countries such as Canada, Germany, Russia, the Republic of Korea and Australia are included, offering a useful counterpoint to countries where the application of AI is only just emerging. Three regional reports are also included: largely the result of authors feeling the need to take a regional perspective on the theme, rather than focusing on developments in a

particular country. Taken together, these reports offer a snapshot of AI-embedded future/s at different stages of development, and a useful opportunity to identify both the positive potential and real threats of AI deployment in diverse contexts.¹

Several reports are concerned with the digitalisation of the workplace, and the impact of AI and automation on worker rights. If predictions of job losses are anything to go by, economies are set to be reshaped entirely. In a country like Ethiopia, for example, about 85% of the workforce is said to be vulnerable to technological replacement, while a similar percentage of those currently employed in Argentina are predicted to need reskilling. In Bangladesh, women working in the ready-made garment sector, “who are at the bottom of the production process and are often engaged in repetitive tasks,” are the mostly likely to suffer the results of automation.

The claim that AI, while shedding menial and repetitive jobs, will create a newly skilled and re-employable workforce currently lacks evidence to support it. This is the “elephant in the room” Deirdre Williams writes in her regional discussion on the Caribbean: “[W]hile there is also insistence that the same new technology will create new jobs, few details are offered and there is no coherent plan to offer appropriate re-training to those who may lose their jobs.” Given the high cost of “retooling” workers, they will instead be “pushed into lower-wage jobs or become unemployed,” writes Chekol. “[I]f the outcome is not mass unemployment, it is likely to be rising inequality.”

In many countries, a reinvigoration of the union movement is necessary. In Argentina, for example, unions report being unprepared to cope with the inevitable changes in the workplace:

Unions are behind in the debate on AI. [They] are disputing basic issues such as salary, health, loss of employment, with no economic stability

¹ Although not usual for GISWatch editorial policy, two country reports were included for India given the number of good proposals we received for that country. We also included a second report on Australia – on AI in the creative industries – because we felt that a focus on AI and the creative sector was a unique consideration not discussed in other country reports.

and pendular changes of government. We started to think in terms of emerging issues such as AI, but suddenly a new government destroyed even the ministry of work.

In that country it was necessary to create a union specifically focused on digital platforms – one that was able to offer collective voice and action for isolated, “on-demand” workers who face new challenges in demanding their rights.

As authors suggest, automation in the workplace is not inherently a bad thing, and can result in meaningful improvements in worker rights, such as assigning robots to do dangerous jobs, or relieving workers from the need to work in unhealthy workspaces. Yet the socioeconomic benefits and costs of workplace change need to be properly understood for their potential impact on society overall – and with the views of workers firmly embedded in policy design and decisions – rather than simply the result of a micro-focus on efficiency and more exact profit, with assumptions made about worker needs.

Authors also show how algorithmic design can perpetuate systemic discrimination – whether due to race, caste, class, gender, or against differently marginalised individuals, groups and communities. In her discussion of automation in the Australian welfare system, Monique Mann calls this a “structural and administrative *violence* [my italics] against those who are socially excluded and financially disenfranchised.” New forms of discrimination are also created (for example, by profiling the unemployed in Poland, what others have called a “double marginalisation” is felt), and the opportunities for discrimination are increased – through, for example, mass surveillance using facial recognition technologies.

Automated facial recognition (AFR) technology receives some attention in these reports, including its use in the persecutory surveillance of the Uyghur ethnic minority in Xinjiang in China, and in Brazilian schools to monitor (and ostensibly improve) attendance. But such a technological response to improve school drop-out rates among lower-income students, Mariana Canto from Instituto de Pesquisa em Direito e Tecnologia do Recife (IP.rec) argues, does not address the structural reasons for this – such as the relevance of the curriculum design, the need for students to work to support their families, and even the levels of crime and violence they are likely to experience on their way to school. Moreover, she adds:

It is important to remember that as systems are being implemented in public schools around the country, much of the peripheral and

vulnerable population is being registered in this “experiment” – that is, data is being collected on vulnerable and marginalised groups.

Mathana Stender from the Centre for the Internet and Human Rights (CIHR) points out in their report on the rise of automated surveillance in Germany that AFR “can [also] lead to automated human rights abuses.” And these abuses are indiscriminate:

With biased assumptions built into training of models, and flawed labelling of training data sets, this class of technologies often do not differentiate between who is surveilled; anyone who passes through their sensor arrays are potential subjects for discrimination.

The implication is that automated surveillance throws the net for potential discrimination wider, increasing the likelihood of global incidences of discrimination being experienced.

Beyond the effect of systemic bias in algorithmic decision making is the question of the quality of the data fed into AI systems. As Malavika Prasad and Vidushi Marda (India) put it, machine learning is “a process of generalising outcomes through examples” and “data sets have a direct and profound impact on *how* an AI system works – it will necessarily perform better for well-represented examples, and poorly for those that it is less exposed to.” For example, census or other socioeconomic data used to train AI or for automated decision making may be varied, and involve questionable methodologies or uneven research processes. This poses challenges for countries where this data is not “clean” or there is a lack of skills and resources to produce the necessary data. In Chile, write Patricia Peña and Jessica Matus from Instituto de la Comunicación e Imagen and Fundación Datos Protegidos, there is a need for “a chain of quality [control] from its collection, capture, use and reuse, especially when it is taken from other databases, so that no bias is generated,” while Ethiopia, “like most other African countries, has the lowest average level of statistical capacity. The lack of data, or faulty data, severely limits the efficacy of AI systems.”

Authors also raise concerns about the access to private data by businesses – especially given that private-public partnerships are seen as necessary to finance much public sector AI development (for example, think of the number of service-level arrangements necessary for smart cities to exist). But questions such as “What access do private companies developing AI technology have to private data?” and “Do they store the data, and for how

long?” largely go unanswered. In the Ponto iD surveillance system set up in Brazilian schools, there is a “lack of information that is included in the company’s privacy policy, or on city halls’ websites.” In its investigation into the introduction of AI in health care in Cameroon, Serge Daho and Emmanuel Biko-bo from PROTEGE QV write:

While patients’ data is collected by the Bonassama hospital and transferred to Sophia Genetics [a company based in the United States and Switzerland] using a secured platform, we could not determine how long this data is stored. [...] Is the confidentiality of Bonassama hospital patients a priority to Sophia Genetics? Hard to answer. Nor have we been able to find out whether or not the patients’ informed consent was requested prior to the data gathering process (the nurses we interviewed could not say).

Korean Progressive Network Jinbonet offers a practical account of policy advocacy in this regard – for example, explaining the legal difference between “pseudonymised” and “anonymised” data – and the litigating temperament necessary from civil society. As it found, not only did guidelines for the de-identification of personal data offer the opportunity for a lively trade in personal data between companies, but the state-run Health Insurance Review and Assessment Service had sold medical data from hospital patients to a life insurance company, and the data of elderly patients to Samsung Life. In Costa Rica, specific legal addenda are needed to oversee and secure the national medical database there, considered “one of the most important information resources in the country.”

The country reports suggest a mixed policy response to AI. A number of countries still do not have adequate data protection laws in place – an essential prerequisite for the roll-out of AI technology. If policies governing AI exist, they are often too broad to account for the real-life implications of the technologies on the rights of people and citizens, or they can become quickly outdated, leaving what Anulekha Nandi from Digital Empowerment Foundation (India) describes as a “governance vacuum over a general-purpose technology with unquantifiable impact on society and the economy.”

In this lacuna, a number of authors (e.g. Rwanda, Pakistan, Jamaica) reference the EU’s General Data Protection Regulation (GDPR) as a template for good governance that can be applied in their own country. Authors point out that a regional perspective on legislation is necessary – but not necessarily easy to achieve. In Latin America, for example,

despite the regional roll-out of Prometea in the judicial system in Buenos Aires, the Constitutional Court in Colombia, and at the Inter-American Court of Human Rights in San José, digitalisation plans in countries like Argentina tend to focus on building a country “brand” as a regional leader in the sector, while being quiet on the need to “[develop] common strategies with other governments in the region.” The result is a regional policy asymmetry, which Raymond Onuoha from the Regional Academic Network on IT Policy (RANITP) at Research ICT Africa argues is detrimental to the global competitive and developmental needs of regions. Moreover, even if regional policy symmetries exist, countries do not necessarily have similar capacities to implement the policies properly:

[M]any African countries are still dealing with basic issues of sustenance like food and housing etc., so technology and technology policy are not at the front burner of critical issues of concern. [...] A harmonised regional data protection policy regime for the continent might impose enforcement liabilities on member countries that lack the required resources for its implementation.

A key policy problem raised by several authors is the question of legal liability in the event of a “wrong” decision by an algorithm (or, in extreme cases, so-called “killer robots”). If this happens, it is unclear whether, for example, the designer or developer of the AI technology, or the intermediary service provider, or the implementing agent (such as a municipality) should be held liable. One solution proposed is that algorithms should be registered as separate legal entities, much like companies, in this way making liability clearer and actionable (a draft bill to this effect was being debated in Estonia – see the Ukraine country report).

Legislation also needs to have a clear view on when and how AI impacts on the current legal framework and rights of citizens. While in Australia, the country’s automated debt-raising programme “reverses the onus of proof onto vulnerable people (and thus overturns the presumption of innocence),” in Turkey, AI is being used in conjunction with copyright law to censor alternative media. Organisational and institutional culture also needs to be addressed in policy – involving significant effort in change management.

A number of authors are critical of the approach to policy design in their countries (in South Korea, for example, the government implements “policies focused on the utilisation rather than protection

of personal data”). They point out that policies often lack inclusivity and context – both essential to understanding the real-life implications on rights when implementing AI technologies. Policy needs to “centre” those most affected by technological changes. In Pune in India – described as one of the “top smart cities” in that country – the city’s smart sanitation project does not address the caste discriminations against the Dalit community, allowing, in effect, unaccountable private sector service providers to “discipline” already marginalised workers engaged in public services.

A useful methodology for better understanding the specific, contextual implications of AI on vulnerabilities and rights – and which can be built into policy design – is “risk sandboxing”. As Digital Empowerment Foundation explains:

Regulatory and data sandboxing are often recommended tools that create a facilitative environment through relaxed regulations and anonymised data to allow innovations to evolve and emerge. However, there also needs to be a concomitant risk sandboxing that allows emerging innovations to evaluate the unintended consequences of their deployment.

Effective policy advocacy may require significant capacity to be built among civil society organisations. For example, in countries like Poland, algorithmic calculations are part of legal and policy documentation. As Jędrzej Niklas writes, “for civil society organisations to successfully advocate for their interests, they must engage in the technical language of algorithms and mathematical formulas.” Reports such as those on the Seychelles and Malawi also show that some work needs to be done in raising public awareness of AI. Better public information on the practical benefits and human rights costs of AI needs to be made available – as well as more detail of the systems that are in place in countries.

Karisma Foundation offers a useful analysis of media coverage of Prometea in Colombia, showing that most reporting offered little understanding of the system: “[T]here was no explanation about what Prometea was, what it does and how it does

it.” When, as in the Ukraine, there appears to be reasonable public awareness of AI and at least some understanding of how it influences their lives, just less than a quarter of people surveyed said AI caused them “anxiety and fear”.

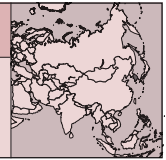
These reports suggest that this fear is not unfounded. Angela Daly (China) points to a global phenomenon of “ethics washing” – or the “gap between stated ethical principles and on-the-ground applications of AI.” While the city of Xinjiang is described as a “‘frontline laboratory’ for data-driven surveillance” in her report, IP.rec suggests “technological advancement” is as much driven by “desire” as anything else; but, “Does this desire turn people into mere guinea pigs for experimentation with new technologies?” For Maria Korolkova from the University of Greenwich, writing on Ukraine, an AI-embedded future risks “dislocating the axis of power in the citizen-state relationship necessary for democracy to function.”

There are several striking examples of the positive use of AI in these reports, and its potential to enable rights in ways that were not possible before. A number of reports focus on the health sector, but promising – although not problem-free – applications are also discussed in areas such as e-government (see South Africa for a useful discussion on this), in “unmasking” forced labour and human trafficking in Thailand, and in combating femicide (see Italy for an example of one of the country’s most advanced data-driven media research projects).

These reports nevertheless also show that an AI-embedded future poses fresh challenges for civil society advocacy – and that purposive action is required. Compromise might not always be possible. Joy Liddicoat, from the New Zealand Law Foundation Artificial Intelligence and Law Project, questions whether the multistakeholder approach to policy design is failing in the wake of the Christchurch terror attacks in her country. Niklas goes further, pointing to the need for a “radical political advocacy”, one that would “not only engage in changes or improvements to algorithms, but also call for the abolition of specific systems that cause harm.”

KOREA, REPUBLIC OF

DATA PROTECTION IN THE AGE OF BIG DATA IN THE REPUBLIC OF KOREA



Korean Progressive Network Jinbonet
Miru
<https://www.jinbo.net>

Introduction

The Korean government is currently focusing on developing emerging technologies, such as artificial intelligence (AI), the “internet of things” (IoT) and “big data”, as part of the so-called Fourth Industrial Revolution. These technologies are interconnected in that deep-learning technology needs big data to train AI, and a vast amount of data, including personal data, is produced through IoT devices. With the development of these technologies, privacy and data protection issues have also been raised. Although the Korean government has recognised data protection as a critical policy issue, the government has continued to implement policies focused on the utilisation rather than protection of personal data.

Policy background and brief history

Personal data protection laws in Korea

Before establishing the Personal Information Protection Act (PIPA)¹ in 2011, there were several acts for regulating personal data in different sectors. The PIPA was enacted to protect personal data covering all areas of society, but even after passing the PIPA, existing acts still remain, such as the Act on Information and Communication Network Utilization (Network Act) and the Credit Information Use and Protection Act (Credit Act). Accordingly, there are several supervisory bodies that govern each act, such as the Ministry of the Interior and Safety (MOIS) which governs the PIPA, the Korea Communications Commission (KCC) which governs the Network Act, and the Financial Service Commission (FSC) which governs the Credit Act, as well as the Personal Information Protection Commission (PIPC) established according to the PIPA. The diffusion of supervisory bodies and acts causes confusion for data subjects and controllers and hinders the establishment of a unified data protection policy. In addition, these

bodies are government ministries, so they have no independence from the government, and the PIPC does not have enforcement powers.²

Guidelines for De-identification of Personal Data³

There has been constant debate in recent years over whether and under what conditions personal data could be processed further beyond the original purpose. Industry keeps requesting permission for utilising personal data for big data analysis and development of AI. As an answer to this, the previous government announced the “Guidelines for De-identification of Personal Data”⁴ in June 2016. According to the guidelines, the de-identification of personal data refers to a “procedure to remove or replace all or part of an individual’s identifiable elements from the data set to prevent the individual from being recognized.”⁵ Because de-identified personal data is no longer considered personal data, it can be processed without the consent of data subjects for purposes other than the original purpose, such as big data analysis, and even provided to third parties. In addition, the guidelines allow companies to combine customers’ de-identified personal data with that of other companies through designated authorities. However, the guidelines were criticised for having no legal basis, because there was no concept of “de-identification” in the PIPA. Moreover, de-identified data is at risk of being re-identified, and as government was aware of these risks, it prohibited disclosing de-identified data to the public.

Since the publication of the guidelines, 20 companies have de-identified customer data and combined the data sets with those of other companies through designated agencies, which amounted to 340 million entries as of August 2017. In opposition to the guidelines, civil society organisations, including the Korean Progressive Network Jinbonet, have laid criminal charges with the prosecutor

¹ www.law.go.kr/lsinfoP.do?lsiSeq=142563&chrClsCd=010203&urlMode=engLsInfoR&viewCls=engLsInfoR#0000

² <https://act.jinbo.net/wp/38733>

³ https://www.kisa.or.kr/public/laws/laws2_View.jsp?cPage=1&mode=view&p_No=282&b_No=282&d_No=3&ST=T&SV=

⁴ https://www.privacy.go.kr/cmm/fms/FileDownload.do?atchFileId=FILE_00000000827254&fileSn=0

⁵ Ibid.

against the relevant companies and designated agencies for violating the PIPA.⁶

Policy hackathon on the use and protection of personal data in the age of big data

In 2018, the current government held a “policy hackathon” – or a multistakeholder discussion forum⁷ – on the use and protection of personal data in the age of big data in order to solve this issue through the amendment of the PIPA. The policy hackathon was attended by stakeholders from industry, civil society, academia and the government. They gathered to reach a social consensus on major issues related to the Fourth Industrial Revolution. Through two hackathon meetings, broad agreements were reached. The participants agreed to use the concepts of personal data, pseudonymised data and anonymised data, borrowed from the European Union’s General Data Protection Regulation (GDPR), instead of the ambiguous concept of de-identification. In this context, pseudonymised data refers to the data processed to make it difficult to directly identify a natural person without combining it with other information. However, it is still personal data because it can be re-identified when combined with other information. On the other hand, anonymised data, such as statistical results, is data processed so that a specific individual can no longer be identified.

Since the hackathon was a place for discussion and interaction, but was not a place to decide policies, there was still a task for government ministries to formulate policies reflecting the hackathon’s agreements and to revise relevant laws in the National Assembly.⁸

Three big data laws

In November 2018, the so-called “three big data laws”,⁹ including the amendments to the PIPA, were proposed in the National Assembly to ease regulation on personal data protection for the purpose of revitalising the big data industry. The three big data laws, however, promote the sale and sharing

of personal data instead of protecting it. In addition, the PIPA amendments undermine the rights of data subjects and reduce the data processor’s obligation to protect personal data. As a result, civil society is against the three big data laws and is again calling for legislation to protect personal data. You can read more detail on this in the section on “Issues around the amendment of the PIPA” below.

Two cases on the use of de-identified data for big data analysis

From 2011 to 2014, The Korea Pharmaceutical Information Center (KPIC) sold the details of 4.7 billion prescriptions for medication to IMS Health Korea¹⁰ for KRW 1.6 billion (USD 138,368).¹¹ KPIC provided the software used for health insurance claims, PM2000, to drugstores. By using PM2000, KPIC collected and sold the information of patients’ diseases and medication claims without permission.¹² No one who received prescription drugs at a drugstore during the period was aware of this.

In 2015, a joint government investigation team on personal data crimes charged IMS Health Korea for violating the personal data of patients. However, the company is claiming innocence. It insists that because the resident registration numbers (RRNs), which can identify specific patients for each prescription, were de-identified through encryption, this data was not personal data.¹³ However, researchers from Harvard University, Latanya Sweeney and Ji Su Yoo, published a paper proving that the encryption method used in the case could be easily decrypted, meaning that individuals could be re-identified.¹⁴

In 2015, the Health Insurance Review and Assessment Service (HIRA), which is run by the state, sold the medical data of 1.1 million hospitalised patients to KB Life Insurance for “insurance product research”. Even prior to this, the HIRA had sold the data of elderly patients to Samsung Life for the purpose of “research” to calculate insurance premiums and develop new insurance products in 2011. Although medical data is considered sensitive data, the HIRA never acquired consent from the patients for using the data. It insisted that the data sets it

6 <https://act.jinbo.net/wp/33555>

7 The policy hackathon was hosted by the Presidential Committee on the Fourth Industrial Revolution and aims to reach an agreement through full-day discussions among stakeholders on critical social issues.

8 Chamsesang. (2018). *A Survey on Data Protection and Human Rights in the Age of the Fourth Industrial Revolution*. National Human Rights Commission of the Republic of Korea. <https://www.humanrights.go.kr/site/program/board/basicboard/view?menuid=001003001004&page-size=10&boardtypeid=16&boardid=7603678>

9 The “three big data laws” mean the PIPA amendments, Credit Act and Network Act. Credit Act: https://elaw.klri.re.kr/kor_service/lawView.do?hseq=46276&lang=ENG; Network Act: https://elaw.klri.re.kr/kor_service/lawView.do?hseq=25446&lang=ENG

10 IMS Health is an international company for data analysis of health care data. The company’s name was recently changed to IQVIA. IMS Health Korea is the Korean branch of the company. <https://www.iqvia.com/about-us>

11 www.monews.co.kr/news/articleView.html?idxno=85001

12 www.hani.co.kr/arti/economy/it/752750.html

13 <https://act.jinbo.net/wp/39218>

14 Sweeney, L, & Yoo, J. S. (2015, 29 September). De-anonymizing South Korean Resident Registration Numbers Shared in Prescription Data. *Technology Science*. <https://techscience.org/a/2015092901>

sold were not personal data because the HIRA de-identified them by encrypting or deleting the RRNs and patient names.¹⁵

Issues around the amendment of the PIPA

The range of use of pseudonymised data

Although hackathon participants agreed to use the concepts of personal data, pseudonymised data and anonymised data instead of the ambiguous concept of de-identification contained in the guidelines, they failed to reach an agreement on the scope of the use of pseudonymised data.¹⁶ Nevertheless, the amendment allows the use and provision of pseudonymised data for “statistics, scientific research and archiving purposes in the public interest” without consent from data subjects (Article 28-2). Here, scientific research includes commercial research. In addition, as with the guidelines for the de-identification of personal data, the amendment allows the combining of data sets from data controllers through designated specialised agencies (Article 28-3).

The Korean government insists that the amendment of the PIPA makes it the equivalent of the GDPR, which also allows further processing of personal data beyond the original purpose of collection under certain conditions for scientific research purposes. However, the amendment allows extensive use of personal data in comparison to the GDPR, while safety measures to protect personal data are meagre.

Firstly, the amendment defines scientific research as “research applying scientific methods such as technological development and demonstration, fundamental research, applied research and private investment research.” Although it borrowed a few phrases from the GDPR,¹⁷ scientific research in the amendment is actually much more widely defined than in the EU. The definition is also somewhat tautological: Is there scientific research that does not apply scientific methods? According to the definition in the amendment, a data controller simply has to claim it is for “scientific research” for pseudonymised personal data to be used and even provided to third parties regardless of the nature of the research.

According to the “reason for proposal” of the amendment, scientific research can include research for “[i]ndustrial purposes, such as the development of new technologies, products and services.”¹⁸

However, civil society insists that the range of scientific research should be limited to research that can contribute to the expansion of a society’s knowledge based on the publication of the research results. Why should the rights of data subjects be restricted for the private interests of companies? Explaining its personal data protection act that reflected the GDPR, the data protection authority in the United Kingdom, the ICO, said that scientific research “does not apply to processing of personal data for commercial research purposes such as market research or customer satisfaction surveys.”¹⁹

Secondly, the GDPR requires that anonymised, not pseudonymised data be provided when research can be carried out with anonymous data, but the government amendment has no such provision to minimise the use of personal data as much as possible.

Finally, the amendment excessively restricts the rights of data subjects. In the case of the GDPR, some rights of data subjects can be derogated only when it is not possible to conduct research without such derogation, but the government’s amendment limits the rights of data subjects comprehensively. For example, in principle, personal data should be discarded when the purpose of the data collection is achieved, but according to the amendments to the PIPA, pseudonymised data provided to a third party in the name of scientific research can be retained by the recipient indefinitely.

The lack of an independent personal data supervisory authority

A personal data supervisory authority should have multiple powers and be independent for effective supervision. The European Court of Justice (ECJ) has emphasised that a completely independent supervisory authority is “‘a guardian’ of rights related to the processing of personal data and an essential component for the protection of personal data.”²⁰ Article 52 (Independence) in the GDPR also states that a “supervisory authority shall act with complete independence in performing its tasks and exercising its powers.”

The PIPC of Korea was established by the enactment of the PIPA in 2011. Korean civil society has demanded the establishment of an independent

15 www.ohmynews.com/NWS_Web/View/at_pg.aspx?CNTN_CD=A0002547315

16 Chamsesang. (2018). Op. cit.

17 GDPR recital 159. <https://gdpr-info.eu/recitals/no-159>

18 law.nanet.go.kr/download/downloadDB.do?dataCode=bbsBasic&dataSid=23941

19 <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/exemptions/>

20 Psygkas, A. (2010, 29 March). ECJ C-518/07 – Commission v. Germany: How “independent” should independent agencies be? *Comparative Administrative Law Blog*. <https://campuspress.yale.edu/compadlaw/2010/03/29/cases-ecj-c-51807-commission-v-germany-how-independent-should-independent-agencies-be>

and fully authorised personal data supervisory authority since before the enactment of the PIPA. However, as mentioned earlier, the Korean supervisory authority, the PIPC, does not have sufficient authority or independence. While it is somewhat positive that the amendment unifies the authorities of the MOIS and KCC into the PIPC, the independence of the integrated PIPC is still limited. This is because the amendment still allows the prime minister to exercise authority to direct and supervise administrative affairs, including the improvement of laws related to the protection of personal data, and the establishment and execution of policies, system and plans. Korean civil society groups are demanding that the PIPC should be guaranteed full independence from the government by excluding the prime minister's authority to supervise.

Conclusion

Civil society fears that if the PIPA amendment is passed as it is, different companies would share, sell and combine customers' data indefinitely. As noted above, companies have consistently sought to combine customers' data with those of other companies. For instance, if this amendment were passed, telecoms could pseudonymise their customers' data and provide this to other companies such as internet service providers and financial companies in the name of research. In this case, the telecom is unlikely to provide the pseudonymised data free of charge, but may require payment or require the other party's personal data sets in return. In addition, through designated public institutions, telecoms and insurance companies would be able to combine pseudonymised customer data. In this way, there is the risk that pseudonymised customer data could be widely shared among numerous companies.

Korean civil society does not oppose the development and utilisation of technologies involving big data, IoT and AI. However, their use should not

justify the violation of the rights to informational self-determination of data subjects.

As can be seen in many international reports, these new technologies could increase the risk of discrimination and surveillance as well as privacy violations. Therefore, for the safe development and utilisation of new technologies, the PIPA needs to be overhauled in response to the era of big data and AI. In addition, it is necessary to establish an independent and fully empowered personal data supervisory authority.

For the development of new technologies such as AI, the data subject needs to trust that his or her personal data will be protected. This is an essential factor if new technologies are to be successfully used in reshaping society. Given the fact that personal data is transferred across borders, this issue is also not just a matter for Korea, but a matter that requires global norms and regulations.

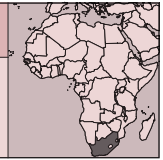
Action steps

The following action steps are suggested for South Korea:

- Launch a campaign to inform the public of the problems in the amendment of the PIPA.
- Convince lawmakers to delete the toxic clause that allows reckless commercial use of personal data in the proposed amendment of the PIPA.
- Urge the government and the national assembly to update the PIPA to include safeguards, such as strengthening the need for a privacy impact assessment, regulating profiling and introducing privacy by design and by default in order to protect personal data that is vulnerable in the era of big data and AI.
- Urge the government and the national assembly to ensure that the PIPC can become an independent and fully empowered authority to protect the rights of data subjects.

SOUTH AFRICA

AI TECHNOLOGIES FOR RESPONSIVE LOCAL GOVERNMENT IN SOUTH AFRICA



Human Sciences Research Council

Paul Plantinga, Rachel Adams and Saahier Parker
www.hsrc.ac.za

Introduction

In 2018, South Africa's Department of Cooperative Governance and Traditional Affairs (CoGTA) partnered with a private company to launch GovChat, an online citizen engagement application designed to promote responsive and accountable local government through the development of an accessible platform for direct messaging between citizens and their local government councillors. The planned pipeline for GovChat includes the integration of artificial intelligence (AI) technologies to boost effectiveness and efficiency.¹ GovChat is one of several applications exploring the use of AI to enhance citizen engagement with local government in South Africa. This country report discusses whether emerging AI-enabled e-government projects, such as GovChat, and associated policies and information legislation are likely to enable a more responsive local government and inclusive development. More specifically, we explore whether these initiatives point to the development of inclusive, "society-in-the-loop"² systems that support the realisation of human rights, including privacy, non-discrimination and access to information.

New directions on poverty, unemployment and inequality

With its recent history of apartheid, South Africa remains saddled with persistently high poverty and unemployment rates as well as stark inequalities, largely along racial lines. Responding to these intersecting crises, the South African government continues to pursue a number of economic and social reforms. A key priority is to build a capable state and responsive public service which is able to

engage with the specific circumstances and capabilities of communities.³

More recently, the government has developed a number of new policies broadly aimed at enhancing the role played by science and technology in supporting more inclusive economic growth, while also re-emphasising the significance of emerging information and communications technologies (ICTs) in an efficient and responsive public service. Among these policy developments are the Draft White Paper on Science, Technology and Innovation,⁴ the National Integrated ICT Policy White Paper,⁵ and South Africa's National e-Strategy Towards a Thriving and Inclusive Digital Future 2017-2030,⁶ all of which fall broadly under South Africa's burgeoning policy discourse on the Fourth Industrial Revolution (4IR).

The recurring emphasis on ICTs comes from a recognition of the impact that the 4IR will have on government, which will "increasingly face pressure to change their current approach to public engagement and policymaking."⁷ To this end, national and subnational government entities have promoted a range of e-governance platforms and policies over the past two decades. The 2018 partnership that saw the launch of GovChat reflects a heightened interest in the role of web, data and social media platforms for improving government service delivery, in this case by CoGTA, the national ministry responsible for ensuring municipalities perform their core service delivery functions.⁸ The increasing prominence of AI in these e-governance plans

1 <https://www.uwc.ac.za/UWCInsight/sholarship@uwc/ColloquiumPresentationsDay1/Govchat%2027%20Oct%202017.pptx>
2 Balaram, B., Greenham, T., & Leonard, J. (2018, 29 May). Artificial Intelligence: real public engagement. *RSA Reports*. <https://medium.com/rsa-reports/artificial-intelligence-real-public-engagement-6bofdo73e2c2>

3 Republic of South Africa. (2018, 20 September). Minister Ayanda Dlodlo: Introducing constitutional values and principles to build a values-driven public service. <https://www.gov.za/speeches/inculcating-constitutional-values-and-principles-including-batho-pele-principles-build>
4 Department of Science and Technology. (2018). *Draft White Paper on Science, Technology and Innovation*. https://www.dst.gov.za/images/2018/Draft-White-paper-on-STI-7_09.pdf
5 Department of Telecommunications and Postal Services. (2016). *National Integrated ICT White Paper*. https://www.dtps.gov.za/images/phocagallery/Popular_Topic_Pictures/National_Integrated_ICT_Policy_White.pdf
6 Department of Telecommunications and Postal Services. (2017). *Digital Society South Africa: South Africa's National e-Strategy towards a thriving and inclusive digital future 2017-2030*. https://www.dtps.gov.za/images/phocagallery/Popular_Topic_Pictures/National-e-strategy.pdf
7 https://www.gov.za/sites/default/files/gcis_document/201812/42078gen764.pdf
8 www.cogta.gov.za/?page_id=253

has significant implications for local government and its relationship with citizens.

AI in South Africa's local government

A 2018 Access Partnership report on “Artificial Intelligence for Africa” compiled by the University of Pretoria identifies examples of where AI can improve citizen interaction, including the use of chatbots, scanning legal documents and classifying citizen petitions. Deeper in the planning and operational activities of public entities, the enhanced predictive capabilities of AI can be used for pre-emptive interventions around the provision of social services and infrastructure maintenance.⁹

GovChat is similarly exploring the use of AI to enhance government efficiency and responsiveness, along the full information processing chain. At its core, GovChat is an online application that allows users to submit queries about public services to councillors and public officials through a variety of electronic channels including websites, WhatsApp and USSD.¹⁰ The South Africa Open Government Partnership (OGP) End-of-Term Report highlights three components of GovChat relevant to citizen engagement:

- A survey tool to rate civil service facilities such as police stations and schools¹¹
- A facility to view service requests
- A donation tool, allowing users to donate blankets, food, clothes and electronics for collection by the local ward councillor.¹²

The expectation from CoGTA is that through GovChat, government will be “instantly accessible to over 16 million people” and “citizens will be able to access over 10,000 public representatives supporting over 30,000 public facilities and services in communities across the country.”¹³ Importantly, the planned pipeline for GovChat includes the integration of “Artificial Intelligence responses”, “Predictive Trend mapping” (in its Version 2 roll-out

in 2019) and “Natural Language query input” (Version 3, 2020).¹⁴

These AI applications dovetail with many of the challenges experienced by local government officials in South Africa. A key concern is improving citizen-government interaction given the large volume of service queries received from citizens on multiple channels. For example, the City of Tshwane 2018 Customer Engagements and Complaints Management Policy expects that AI will be able to proactively “affirm” and consolidate repeat queries.¹⁵ Broadly, under South Africa's constitutional commitments, GovChat and its AI capabilities offer an opportunity to enhance responsive and accountable government,¹⁶ while at the same time fulfilling the state's obligations in terms of the rights of freedom of expression,¹⁷ access to information¹⁸ and just administrative action.¹⁹ Moreover, GovChat is expected to promote access to local government for those segments of the population who may have historically struggled due to physical or social barriers, including women and those with disabilities.²⁰ In this way, GovChat can theoretically contribute to the vision of the Constitution to create a “democratic and open society in which government is based on the will of the people” and all are equal.²¹

While CoGTA's expectations of GovChat seem ambitious, the similar MomConnect initiative has

9 University of Pretoria. (2018). *Artificial Intelligence for Africa: An Opportunity for Growth, Development, and Democratisation*. Access Partnership. https://www.up.ac.za/media/shared/7/ZP_Files/ai-for-africa.zp165664.pdf

10 USSD (unstructured supplementary service data) refers to a mobile communication technology for sending text between a mobile phone device and another application program in the network.

11 Users are able to search for particular facilities and to rate both the service and facilities. Survey results are submitted to contact persons at the relevant facility.

12 Humby, T. (2019). *Open Government Partnership Independent Reporting Mechanism (IRM): South Africa End of Term Report 2016-2018*. https://www.opengovpartnership.org/sites/default/files/South-Africa_EOTR_2016-2018.pdf

13 Republic of South Africa. (2018). Deputy Minister Andries Nel. Launch of Govchat. <https://www.gov.za/speeches/govchat-25-sep-2018-0000>

14 <https://www.uwc.ac.za/UWCInsight/sholarship@uwc/ColloquiumPresentationsDay1/Govchat%2027%20Oct%202017.pptx>

15 City of Tshwane. (2018). Customer Engagements and Complaints Management Policy. www.tshwane.gov.za/PublicParticipation/12.%20Customer%20Engagements%20and%20Complaints%20Management%20Draft%20Policy%20for%20CoT.pdf

16 Under the Constitution of the Republic of the South Africa, Act 108 of 1996, the objectives of local government are set out as follows:

152. (1) The objects of local government are—
- (a) to provide democratic and accountable government for local communities;
 - (b) to ensure the provision of services to communities in a sustainable manner;
 - (c) to promote social and economic development;
 - (d) to promote a safe and healthy environment; and
 - (e) to encourage the involvement of communities and community organisations in the matters of local government.
- (2) A municipality must strive, within its financial and administrative capacity, to achieve the objects set out in subsection (1).

17 Section 16 of the Constitution.

18 Section 32 of the Constitution.

19 Section 33 of the Constitution.

20 DareDisrupt. (2019). *Civic Tech: Smart Use of Civic Tech to Promote Accountability and Transparency*. Danish Church Aid. https://www.danchurchaid.org/content/download/23246/414917/version/1/file/Civic%20tech%20mapping%20final_FEB19_PDFa.pdf

21 Preamble to the Constitution.

registered over two million subscribers.²² MomConnect is a USSD, text and WhatsApp-based maternal health information platform implemented by South Africa's National Department of Health together with various partners. The scale of the programme suggests that AI-supported citizen engagement applications could already reach large audiences across the country. In addition, there has been increasing experimentation with AI methods (mainly machine learning) in the back-end of South Africa's local government operations, such as for planning transport routes,²³ clinic placement²⁴ and electricity management.²⁵ This work builds on a wider base of (typically less adaptive) predictive modelling and automated decision making (ADM) technology already used in South African municipalities.

Ensuring inclusive local governance outcomes

The current and emerging scale of AI and ADM adoption requires urgent reflection on the potential benefits and limitations for local governance, discussed below.

Accessibility

If the benefits of citizen-engagement platforms and AI are to reach all communities equally, we will need to address challenges around the accessibility of GovChat-like applications and associated AI, starting with underlying connectivity. While social media use has increased steeply since 2012, internet penetration in South Africa remains low, particularly in comparison with other African countries.²⁶ Moreover,

internet penetration is especially poor in rural areas of South Africa which would benefit most from remote interaction with local councillors and electronic government applications. Although USSD is a more accessible option for interacting with these services, smart devices enable much richer communication, but with a higher initial device cost as well as the ongoing cost of data. South Africa ranks among the most expensive countries for data services in Africa, especially for prepaid mobile data plans.²⁷ Further, citizen-engagement applications require a particular level of technological know-how and confidence to use and trust the technology, which may be exacerbated by unfamiliar user interfaces and languages, such as current virtual private assistants (VPAs) which are predominantly English-speaking and female.²⁸ Ongoing research around local government's use of AI-supported automated translation and text-to-speech tools is therefore important.²⁹

Privacy and trust

When it comes to government's collection and processing of data through AI-enabled applications, a fundamental concern regarding individual privacy and potential state surveillance is raised. The increased use of social media in South Africa means that governments can mine and analyse comments on public channels, then "agilely respond to citizens' complaints"³⁰ or even influence emerging issues. This raises serious privacy concerns. In South Africa, perhaps the most controversial use of AI technologies by the state has been in predictive policing, such as through "upgrades" to CCTV camera systems in the City of Johannesburg to enable facial recognition³¹ and broader research

22 <https://www.praekelt.org/momconnect>

23 Van Heerden, Q. (2015). *Using Social Media Information in Transport and Urban Planning in South Africa*. Smart and Sustainable Built Environment (SASBE). <https://hdl.handle.net/10204/9871>; and ITU. (2019). WSIS Prizes Contest 2019 Nominee: GoMetro. <https://www.itu.int/net4/wsis/stocktaking/Prizes/2020/DetailsPopup/15434965423625087>

24 Conway, A. (2016). *Optimizing Mobile Clinic Locations using Spatial Data*. Presentation at MIA Meetup at Rise Africa, Cape Town, 27 October. <https://drive.google.com/file/d/0BxzNs-HspAzYSDJoMWpVcdDfYnc/view>

25 <https://dsidweb.github.io/articles/project-matla>

26 Internet penetration in South Africa is currently at 53.7%. Kenya, by way of example, has an internet penetration rate of 83% (see <https://www.internetworldstats.com/stats1.htm>). The government has rolled out free public Wi-Fi access in selected communities and areas, yet the reach of these services is still not sufficient to address the needs of the many millions, particularly those in rural communities. Smartphone applications have, however, found success in selected industries and communities such as small-scale fishers being networked on a smartphone application called ABALOBI that aims to link small-scale fishers to governance processes, thereby increasing profits and limiting time from hook to table. This app helps in retaining good governance structures, compliance, sustainability education and ensures local development through the adoption of fair trade practices. See: <https://abalobi.info>

27 Provisional findings by the Competition Commission highlight South Africa's "anti-poor retail price structures". www.compcom.co.za/wp-content/uploads/2017/09/Data-Services-Inquiry-Report.pdf

28 Ní Loideáin, N., & Adams, R. (2018, 10 October). Gendered AI and the role of data protection law. *talking humanities*. <https://talkinghumanities.blogs.sas.ac.uk/2018/10/10/gendered-ai-and-the-role-of-data-protection-law>

29 <https://www.sadilar.org>; see also Calteaux, K., De Wet, F., Moors, C., Van Niekerk, D., McAllister, B., Grover, A. S., Reid, T., Davel, M., Barnard, E., & Van Heerden, C. (2013). *Lwazi II Final Report: Increasing the impact of speech technologies in South Africa*. Pretoria: Council for Scientific and Industrial Research. <https://hdl.handle.net/10204/7138>

30 Moodley, K. (2016, 5 August). Power of sentiment analysis for public service. *ITWeb*. <https://www.itweb.co.za/content/VKA3Wwqd697rydZ>

31 Swart, H. (2018, 28 September). Joburg's new hi-tech surveillance cameras: A threat to minorities that could see the law targeting thousands of innocents. *Daily Maverick*. <https://www.dailymaverick.co.za/article/2018-09-28-joburgs-new-hi-tech-surveillance-cameras-a-threat-to-minorities-that-could-see-the-law-targeting-thousands-of-innocents>

collaborations with the South African defence and police forces to “Build Safer Communities”.³² Meanwhile, the unauthorised use of data to exploit social grant recipients has undermined already limited trust in IT systems.³³

Concerns about how personal data is going to be used by the state point to a broader challenge of declining trust in government and in South Africa’s local government in particular.³⁴ Mistrust of (and within) local government, including suspicion of and actual corruption, as well as resistance to new technologies which can potentially expose mismanagement or wrong-doing, significantly impedes the possibilities of what emerging technologies could achieve.³⁵ The relatively opaque character of AI risks obscuring transactions and decisions even further.

Practitioners will need to work with elected officials and civil society organisations in using AI to strengthen existing local accountability mechanisms, while building a stronger culture of data protection and safeguards against unnecessary state (and service provider) processing of personal information.

Explainability and accountability

Ensuring that citizens have sufficient understanding about how AI is processing their data is critical for building trust and enabling accountability. However, in local government there are often limited technical skills, which makes it difficult for officials to understand and explain existing data processing in platforms like GovChat, which is likely to be compounded by the introduction of AI features. There is therefore a need to define a reasonable level of understanding and explanation that addresses AI but also the wider spectrum of ADM approaches in use by government.³⁶

The technical complexity and adaptive nature of AI means that it may not be feasible or useful to provide “sufficient information about the underlying logic of the automated processing” as suggested in South Africa’s key data protection law, the Protection of Personal Information Act (POPIA);³⁷ or an extensive “right to explanation”, as debated in the crafting of the European Union’s General Data Protection Regulation (GDPR).³⁸

As a start we may look to define broad principles for “algorithmic accountability” and an acceptable scope of influence for AI and ADM that national and local governments can draw on. For example, the African Union (AU) Convention on Cyber Security and Personal Data Protection defines the limit as:

A person shall not be subject to a decision which produces legal effects concerning him/her or significantly affects him/her to a substantial degree, and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him/her.³⁹

Additional lower level principles may include ensuring that data processing is accurate, does not discriminate, can be audited, and that there are mechanisms for redress and mitigation of negative social impacts.⁴⁰ Moreover, a carefully designed “algorithmic impact assessment” can facilitate broad dialogue about the implications of different AI technologies.⁴¹

Inevitably there will be overlapping layers of global, national and subnational regulation of AI issues. While the AU is seeking to harmonise cybersecurity policy across member states, countries and subnational governments are likely to pursue their own interpretations and legal frameworks governing transparency, accountability and other safeguards in the use of AI. In South Africa, the regulatory body established under POPIA is not

32 Council for Scientific and Industrial Research. (2016). *CSIR Annual Report 2015/16: Our Future Through Science*. https://www.csir.co.za/sites/default/files/Documents/CSIR%20Annual%20Report%202015_16.pdf; Kwet, M. (2017, 27 January). *Cmore: South Africa’s New Smart Policing Surveillance Engine*. *CounterPunch*. <https://www.counterpunch.org/2017/01/27/cmored-south-african-smart-policing-surveillance-engine>; and Ní Loideáin, N. (2017). *Cape Town as a Smart and Safe City: Implications for Governance and Data Privacy*. *International Data Privacy Law*, 7(4), 314-334.

33 The Citizen. (2018, 8 March). *Black Sash back in court over social grants*. *The Citizen*. <https://citizen.co.za/news/1845959/black-sash-back-in-court-over-social-grants>

34 www.hsrc.ac.za/uploads/pageContent/9835/2019-03-28%20DGSD%20Youth%20%20Elections%20Seminar.pdf

35 We are particularly grateful to Caroline Khene, co-director of MobISAM, for her insights in this section of the report. <https://mobisam.net>

36 Algorithm Watch. (2019). *Atlas of Automation: Automated decision-making and participation in Germany*. https://atlas.algorithmwatch.org/wp-content/uploads/2019/04/Atlas_of_Automation_by_AlgorithmWatch.pdf

37 https://www.gov.za/sites/default/files/gcis_document/201409/3706726-11act4of2013protectionofpersonalinformationcorrect.pdf

38 Doshi-Velez, F., Kortz, M., Budish, R., Bavitz, C., Gershman, S., O’Brien, D., Schieber, S., Waldo, J., Weinberger, D., & Wood, A. (2017). *Accountability of AI Under the Law: The Role of Explanation*. Cornell University. <https://arxiv.org/abs/1711.01134>

39 African Union. (2014). *African Union Convention on Cyber Security and Personal Data Protection*. Article 14(5). <https://au.int/en/treaties/african-union-convention-cyber-security-and-personal-data-protection>

40 World Wide Web Foundation. (2017). *Algorithmic Accountability: Applying the concept to different country contexts*. https://webfoundation.org/docs/2017/07/WF_Algorithms.pdf

41 Supergovernance. (2018, 18 March). *A Canadian Algorithmic Impact Assessment*. *Medium*. <https://medium.com/@supergovernance/a-canadian-algorithmic-impact-assessment-128a2b2e7f85>

yet fully functional. However, it is expected to play a crucial role in enforcing compliance with the Act and promoting good data and AI governance in South Africa.

Small data

Globally, AI projects have been affected by the limited availability of training data from many regions and population groups, which has resulted in bias and discrimination in the operation of AI tools.⁴² In local contexts, the relatively small amount of available data can lead to “overfitting” of algorithms and inaccurate predictions.

A further issue is the risk of re-identification of personal data, which is higher in geographic regions with small populations.⁴³ AI-related methods are used to re-identify and link data records across databases, which can be helpful for integrating local government planning or service provision across multiple departments. But it can also result in unauthorised disclosure of private information, which would constitute a violation of POPIA. In these circumstances, data managers may try to ascertain which variables (e.g. town, education, race or gender) increase the likelihood of disclosure and develop masking strategies to reduce the risk, such as in the Google Cloud Data Loss Prevention service.⁴⁴

Beyond these technical issues is a more fundamental question about who data is being collected for and where it is being used. The demand for data in AI (and in national and global data initiatives) creates pressure on local data collection systems to improve the scale and quality of data sourcing, feeding into an extractive local-global pipeline. A “small data” perspective⁴⁵ prioritises more local forms of data collection *and* use, which leads to new questions and possible models for how data is shared and processed within and between individuals and communities. For example, data cooperatives⁴⁶ and

data commons⁴⁷ shift the locus of control to the contributors of the data, while the citizen science community works on vocabularies and ontologies for data sharing between projects.⁴⁸ These activities could provide the conceptual and technical foundations for local government AI projects that are anchored in small data sharing and re-empowered citizens. The AI “black-box” is likely to add to the sense that individuals are losing control over their data⁴⁹ and undermine meaningful, place-based governance processes.

Conclusion

While the planned adoption of AI in GovChat and similar platforms represents an important step forward in the use of AI-related technologies to support the work of government, it also provides a critical opportunity to critique and reflect on the associated social, legal and technological concerns raised by such developments. This report has outlined some of the key concerns in this regard, particularly with regard to accessibility, privacy, trust, explainability, accountability, and the challenges and opportunities associated with small populations and data sets.

A general point is the need to empower both citizens and local government officials to use and benefit from such technologies. Through more inclusive impact assessments, design methods and accountability mechanisms, legislators and system developers can support the development of user-centred AI innovations with higher levels of trust, adoption and impact.

Moreover, in South Africa, as elsewhere, local government is regarded as the “face of government”.⁵⁰ However, the importance of (physical or virtual) proximity and face-to-face interaction in local governance is often underestimated in ICT implementation. This consideration applies to AI-enabled systems which should seek to *enhance* (rather than *replace*) existing, often trusted ways of doing things. In doing so, South Africa can work toward developing its own set of ethical tenets and principles upon which the use of AI in government and elsewhere can be based.

42 Hao, K. (2019, 4 February). This is how AI bias really happens—and why it’s so hard to fix. *MIT Technology Review*. <https://www.technologyreview.com/s/612876/this-is-how-ai-bias-really-happens-and-why-its-so-hard-to-fix>

43 Greenberg, B., & Voshell, L. (1990). Relating risk of disclosure for microdata and geographic area size. *American Statistical Association 1990 Proceedings of the Section on Survey Research Methods*, 450-455. www.asasrms.org/Proceedings/papers/1990_074.pdf

44 <https://cloud.google.com/dlp/docs/concepts-risk-analysis>.

45 See: Data and Sustainable Development: Last Mile Data Enablement and Building Trust in Indicators Data. <https://cs.unu.edu/research/sdgs>

46 Walsh, D. (2019, 8 July). How credit unions could help people make the most of personal data. *MIT Sloan School of Management*. <https://mitsloan.mit.edu/ideas-made-to-matter/how-credit-unions-could-help-people-make-most-personal-data>

47 Baarbé J., Blom, M., & de Beer, J. (2017). *A data commons for food security*. Open AIR. <https://www.openair.org.za/publications/a-data-commons-for-food-security>

48 The Citizen Science COST Action: Working Group 5 – Improve data standardization and interoperability. <https://www.cs-eu.net/wgs/wg5>

49 Thinyane, M. (2018). Towards Informing Human-centric ICT Standardization for Data-driven Societies. *Journal of ICT Standardization*, 6(3), 179-202. <https://dx.doi.org/10.13052/jicts2245-800X.631>

50 https://ossafrika.com/esst/index.php?title=Summary_of_the_Municipal_Systems_Act%2C_no._32_of_2000

Action steps

The following steps are recommended for South Africa:

- Enhance scientific literacy and life-long learning in order to strengthen public understanding of science and technology, including AI, and its potential impact on society.
- Contribute to global, AU and national initiatives on principles for “algorithmic accountability” that local governments can adapt and use.
- Explore what role (sub)national legislatures and independent regulators should play in AI oversight, and build necessary capacity in these entities for supporting government entities with ethical AI implementation in South Africa.
- Run a programme of public engagement and consider a diversity of legal approaches (privacy, competition, criminal) to embed a culture of data protection and formal safeguards against unnecessary state and private sector processing of personal information.
- Design algorithmic impact assessments that can facilitate broad dialogue about the implications of different AI technologies in local government.
- Improve risk assessment and mitigation capabilities among system developers to prevent re-identification and discrimination when building platforms and integrating with local data systems.
- Explore alternative business models and technologies for data collection and sharing to strengthen the role of data contributors in AI systems.
- Support ongoing research into languages/translation and user interfaces for AI implementation in different contexts.

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.

GLOBAL INFORMATION SOCIETY WATCH
2019 Report
www.GISWatch.org

