GLOBAL INFORMATION SOCIETY WATCH 2019

Artificial intelligence:
Human rights, social justice and development

Association for Progressive Communications (APC),
Article 19, and Swedish International Development Cooperation Agency (Sida)

Global Information Society Watch

2019

APC Sida ARTICLE 19

Global Information Society Watch 2019

Artificial intelligence: Human rights, social justice and development

Operational team
Valeria Betancourt (APC)
Alan Finlay (APC)
Mallory Knodel (ARTICLE 19)
Vidushi Marda (ARTICLE 19)
Maja Romano (APC)

Project coordination team
Valeria Betancourt (APC)
Cathy Chen (APC)
Flavia Fascendini (APC)
Alan Finlay (APC)
Mallory Knodel (ARTICLE 19)
Vidushi Marda (ARTICLE 19)
Leila Nachawati (APC)
Lori Nordstrom (APC)
Maja Romano (APC)

GISWatch 2019 advisory committee
Namita Aavriti (APC)
Rasha Abdul Rahim (Amnesty International)
Alex Comninos (Research ICT Africa)
Malavika Jayaram (Digital Asia Hub)
J. Carlos Lara (Derechos Digitales - América Latina)
Joy Liddicoat (Centre for Law and Emerging Technologies, University of Otago)
Andrew Lowenthal (EngageMedia)
Micaela Mantegna (Geekylegal/Machine Intelligence Lab, Center for Technology and Society, San Andres University)
Valeria Milanes (Asociación por los Derechos Civiles)

Project coordinator
Maja Romano (APC)

Editor
Alan Finlay (APC)

Assistant editor and proofreading
Lori Nordstrom (APC)

Publication production support
Cathy Chen (APC)

Graphic design
Monocromo

Cover illustration
Matías Bervejillo

We would like to extend a special note of thanks to a number of authors who have made ad honorem contributions to this edition of GISWatch. We gratefully acknowledge the following:

Philip Dawson and Grace Abuhamad (Element AI)
Anita Gurumurthy and Nandini Chami (IT for Change)
Rasha Abdul Rahim (Amnesty International)

Sida ARTICLE 19

APC would like to thank the Swedish International Development Cooperation Agency (Sida) and ARTICLE 19 for their support for Global Information Society Watch 2019.

Published by APC
2019
Printed in USA

Global Information Society Watch 2019 web and e-book
ISBN 978-92-95113-13-8
APC Serial: APC-201910-CIPP-R-EN-DIGITAL-302

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

Table of contents

THEMATIC REPORTS

COUNTRY AND REGIONAL REPORTS

Preface

Valeria Betancourt (APC) and Mallory Knodel (ARTICLE 19)

Sixty years after machines started beating humans at strategy board games, this edition of Global Information Society Watch (GISWatch) focuses on the implications of artificial intelligence (AI) systems on human rights, social justice and sustainable development.

The AI future is here. AI is now finding widespread practical application: from transport to health, agriculture to waste removal; from policing to welfare, and from smart technology in the home to space exploration. Automated decision making is increasingly being used in critical service and infrastructure provision in areas such as employment, housing, access to education, commerce and access to credit, impacting people's lives in profound ways. In this sense, context matters.

Despite the application of AI in geographically diverse contexts, conversations on AI have been driven largely by Western and global North predictions and perspectives. Yet the assumptions, values, incentives and socioeconomic environments within which AI technologies function vary greatly across jurisdictions and, as a consequence, the very real effects of AI are also more diverse.

GISWatch fills this gap between perspective and impact by exploring AI in the local context, with a specific focus on countries in the global South. We have asked: What impact do AI systems have on vulnerable and marginalised populations around the world? How do they impact, positively or negatively, on human rights concerns such as privacy, security, freedom of expression and association, access to information, access to work, to organise and join trade unions? What are the political implications of the widespread use of data in building AI systems? And what are the positive benefits of AI for enabling rights, such as the right to health or education, in making government more accessible to people, or in addressing key social challenges such as forced labour and human trafficking? How is power asymmetry embedded in the ways that AI systems are designed and deployed, and what potential threats or benefits does this have for people in the face of automated decision making? What social values are being transformed by the application of artificial intelligence?

The answers given are contained in the following eight thematic reports, 40 country reports and three regional reports.

This year's GISWatch report has been a fortunate opportunity for collaboration between ARTICLE 19 and APC around an issue of common concern. We believe the joint effort has resulted in valuable knowledge building and context-specific analysis of the impacts of AI. GISWatch is also a network collaboration among all of its contributors and this edition will undoubtedly result in research-based advocacy, influence and shaping of alternative regulatory, technical and policy responses.

AI is ultimately a social phenomenon, and, as such, it is our collective hope that this edition of GISWatch contributes to ensuring that human rights, human dignity, collective and individual agency, social justice and development are not undermined, but rather strengthened by it.

Thematic reports

Introduction

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

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

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

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

The need to address the imbalance in the global narrative

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

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

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

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

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

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

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

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

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

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

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

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

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

Complexity of governance frameworks and form

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Complexity of politics: Power and process

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

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

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

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

[22] https://www.fatml.org

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

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

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

[26] https://digitalcooperation.org

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

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

[29] https://aiforgood.itu.int

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

[31] https://www.partnershiponai.org

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

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

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

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

It is helpful to lay down the various ways and levels at which power is concentrated, leveraged and imposed by these systems. By producing favourable outcomes for some sections of society, or by having disproportionate impact on certain groups within a society, the ways in which people navigate everyday life is significantly altered. The ways in which governments navigate societal problems is also significantly altered, given the widespread assumption that using AI for development is inherently good. While there is a tremendous opportunity in this regard, it is imperative to be conscientious of the inherent limitations of AI systems, and their imperfect and often harmful overlap with textured and imperfect societies and economies. AI systems are primarily developed by private companies which train and analyse data on the basis of assumptions that are not always legal or ethical, profoundly impacting rights such as privacy and freedom of expression. This essentially makes private entities arbiters of constitutional rights and public functions in the absence of appropriate accountability mechanisms. This link between private companies and public function power was most visibly called out through the #TechWontBuildIt movement, where engineers at the largest technology companies refused to build problematic technology that would be used by governments to undermine human rights and dignity.[34] The design and development of AI systems is also concentrated in large companies (mostly from the United States and increasingly from China).[35] However, deployment of technology is often *imposed* on jurisdictions in the global South, either on the pretext of pilot projects,[36] or economic development[37] and progress. These jurisdictions are more often than not excluded from the table at stages of design and development, but are the focus of deployment.

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

Conclusion

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

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

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

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

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

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

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

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

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

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

Towards data governance that empowers the public

Philip Dawson and Grace Abuhamad
Element AI
https://hello.elementai.com/data-trusts.html

Introduction

If the Cambridge Analytica scandal served as the public's "great privacy awakening,"[1] for public policy experts it affirmed several troubling messages about human vulnerability given the current state of the governance of big data and artificial intelligence (AI) systems. What started as a legitimate academic research project quickly became scandalous when a British political consulting firm used the data collected – from up to 87 million people's Facebook profiles – for a different purpose: to influence the 2016 United States (US) election through targeted political advertisements.

The data transfer occurred without consent from Facebook's users or, arguably, even Facebook itself, reinforcing the idea that big data and AI systems pose significant threats not only to the right to privacy, but to the enjoyment of human rights and the integrity of democratic institutions.[2] As the scandal unfolded, and the European Union's General Data Protection Regulation and the Council of Europe's modernised Convention 108+ entered into force, experts cautioned that in the absence of new approaches to data governance, even a new a "bill of data rights" could not check the power imbalances between data controllers and data subjects.[3] Facebook's refusal to attend hearings before the International Grand Committee on Big Data, Privacy and Democracy, even under subpoena, points to the urgency of finding new ways of dealing with the platform's power.

Current approaches to data governance suffer from a lack of transparency and accountability, in part because big data – in combination with AI – continues to make end runs around consent and privacy self-management.[4] AI-enabled methods of analysis can be used by companies to generate, infer and collect sensitive information about people that they have neither provided nor confirmed.[5] Companies have access to an array of data collection methods, some of which circumvent consent without detection: data is now, as the Cambridge Analytica case demonstrated, extracted through online profiling, purchased from third-party brokers, or derived from aggregated data sets. The complexity and opacity of information flows make it virtually impossible for individuals to discern, much less self-manage, the risks or rights they engage when consenting to the use of their personal data.[6]

Another problem is that current approaches to data governance tend to concentrate data in the hands of powerful digital platforms, preventing the public from sharing in its value.[7] In today's digital society, individuals serve as the inputs to AI systems and yet they wield little control over its outputs. While exceptional, scandals like Cambridge Analytica prove the following rule: not only do current approaches to governing data exclude (most) individuals from sharing in its value, but they expose them to human rights abuses, too.

[1] Lapowsky, I. (2019, 17 March). How Cambridge Analytica Sparked the Great Privacy Awakening, *Wired*, https://www.wired.com/story/cambridge-analytica-facebook-privacy-awakening

[2] 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; Privacy International, & ARTICLE 19. (2018). *Privacy and Freedom of Expression in the Age of Artificial Intelligence*. https://privacyinternational.org/report/1752/privacy-and-freedom-expression-age-artificial-intelligence

[3] Tisne, M. (2018, 14 December). It's time for a Bill of Data Rights. *MIT Technology Review*. https://www.technologyreview.com/s/612588/its-time-for-a-bill-of-data-rights; Wylie, B. (2019, 30 January). Why we need data rights: 'Not everything about us should be for sale'. *Financial Post*. https://business.financialpost.com/technology/why-we-need-data-rights-not-everything-about-us-should-be-for-sale

[4] Barocas, S., & Nissenbaum, H. (2014). Computing Ethics: Big Data's End Run Around Procedural Privacy Protections. *Communications of the ACM, 57*(11). https://nissenbaum.tech.cornell.edu/papers/Big%20Datas%20End%20Run%20Around%20Procedural%20Protections.pdf

[5] Kaye, D. (2018). Op. cit.

[6] Solove, D. (2013). Privacy Self-Management and the Consent Dilemma. *Harvard Law Review, 126*(7); Rau, S. (2018, 16 October). Free, Informed and Unambiguous Consent in the Digital Age: Fiction or Possibility? *The Human Rights, Big Data and Technology Project*. https://hrbdt.ac.uk/free-informed-and-unambiguous-consent-in-the-digital-age-fiction-or-possibility

[7] Element AI, & Nesta. (2019). *Data Trusts: A new tool for data governance*. https://hello.elementai.com/rs/024-OAQ-547/images/Data_Trusts_EN_201914.pdf

Accordingly, an increasing proportion of policy workshops, discussions and research over the last year have focused on designing inclusive data governance models that facilitate public accountability and promote a more equitable distribution of data's economic value.[8] While a number of novel approaches have been considered, proposals based on fiduciary models of data governance have garnered significant attention. This report provides a brief overview of current research and policy discussions related to two such proposals – "information fiduciaries", which aim to improve the accountability of online platforms, and "data trusts", a flexible governance tool that is being considered for a range of different purposes – while offering an assessment of their unique value propositions and implementation challenges.

Information fiduciaries

The nature of fiduciary relationships and the precise duties they create is a contested subject.[9] Yet as Richard Whitt notes, "all definitions of fiduciaries share three main elements: (1) the entrustment of property or power; (2) the entrustors' trust of fiduciaries; and (3) the risk to entrustors emanating from the entrustment."[10] As such, the information fiduciary proposal recommends raising the intensity of obligations owed by data controllers (i.e. companies) to data subjects (i.e. individuals) through the imposition of fiduciary duties of care, confidentiality and loyalty, which would transform large data controllers like digital platforms into "information fiduciaries". The intent is to correct the power imbalance between companies and individuals by giving companies duties similar to those held by doctors, lawyers and accountants towards their patients or clients.

While scholars such as Lillian Edwards[11] and Jack M. Balkin[12] may be credited for first considering the imposition of fiduciary obligations, albeit through different methods, their approach has faced some criticism.[13] Whereas Edwards has suggested that fiduciary obligations are "implied"[14] whenever a data subject shares personal data with a data controller, on account of the risk the former exposes her- or himself to, Sylvie Delacroix and Neil Lawrence argue that taking up a duty of loyalty to manage data subjects' rights in their best interests would place data controllers in a conflict of interest with their competing duty to maximise shareholder value.[15]

To resolve this tension, Balkin has proposed that special immunities or financial incentives could help induce data controllers to take up a limited fiduciary obligation that could be defined by statute.[16] Several problems have been identified with this "grand bargain".[17] Special incentives for platforms to behave as fiduciaries may not be enough to nullify the conflict of interest between a platform's duty of loyalty to manage data subjects' rights and its duty of loyalty to shareholders.[18] As Delacroix and Lawrence have put it:

> [T]he "information fiduciary" proposed by Balkin would be placed in a position that is comparable to that of a doctor who gains a commission on particular drug prescriptions or a lawyer who uses a company to provide medical reports for his clients while owning shares in that company.[19]

Mike Godwin has argued that a "professional framework of fiduciary obligations for tech companies" supported by "professional codes of ethical conduct that bind the tech companies that have fiduciary duties to us" could be one way of ensuring platforms take their positions as information fiduciaries seriously.[20]

Balkin's recommendation to restrict digital platforms' fiduciary duties unfairly discounts the intangible vulnerabilities of living in an online world, where privacy and human rights violations routinely go unnoticed or unchallenged. Facebook may not be a doctor or YouTube an accountant,

[8] Delacroix, S., & Lawrence, N. (2019). Bottom-up Data Trusts: disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*. https://doi.org/10.1093/idpl/ipz014; Element AI, & Nesta. (2019). Op. cit.

[9] McDonald, S. (2019, 5 March). Reclaiming Data Trusts. *Centre for International Governance Innovation*. https://www.cigionline.org/articles/reclaiming-data-trusts

[10] Whitt, R. (2019, 26 July). Old School Goes Online: Exploring Fiduciary Obligations of Care and Loyalty in the Platforms Era. *SSRN*. https://ssrn.com/abstract=3427479

[11] Edwards, L. (2004). The Problem with Privacy. *International Review of Law, Computers & Technology*, *18*(3), 263-294.

[12] Balkin, J. M. (2016). Information Fiduciaries and the First Amendment. *UC Davis Law Review,49*(4).

[13] McDonald, S. (2019, 5 March). Op. cit.; Delacroix, S., & Lawrence, N. (2019). Op. cit.; Khan, L., & Pozen, D. E. (2019). A Skeptical View of Information Fiduciaries. *Harvard Law Review, 133*, Forthcoming. https://ssrn.com/abstract=3341661

[14] Edwards, L. (2004). Op. cit.

[15] Delacroix, S., & Lawrence, N. (2019). Op. cit.

[16] Balkin, J. M. (2016). Op. cit.; Zittrain, J. (2013). Engineering an Election. *Harvard Law Review Forum, 127*.

[17] Balkin, J. M., & Zittrain, J. (2016, 3 October). A Grand Bargain to Make Tech Companies Trustworthy. *The Atlanti*c. https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346

[18] Khan, L., & Pozen, D. E. (2019). Op. cit.

[19] Delacroix, S., & Lawrence, N. (2019). Op. cit.

[20] Godwin, M. (2018, 16 November). It's Time to Reframe Our Relationship With Facebook. *Slate*. https://slate.com/technology/2018/11/information-fiduciaries-facebook-google-jack-balkin-data-privacy.html?wpsrc=sh_all_dt_tw_ru; Godwin, M. (2018, 26 November). If Facebook is really at war, the only way to win is to put ethics first. *Washington Post*. https://www.washingtonpost.com/outlook/2018/11/26/if-facebook-is-really-war-only-way-win-is-put-ethics-first

and yet each handles sensitive personal data that has been entrusted to them, and which can expose individuals, as entrustors, to abuse.[21] Balkin's position may be further undermined in his characterisation of data controllers' limited duty of loyalty as a prohibition from "act[ing] like con men"[22] or creating "an unreasonable risk of harm to their end user":[23] corporations already bear a duty to do no harm under tort law duty of care. This aspect of Balkin's proposal has led Lina Khan and David Pozen to question whether Balkin's proposal "is a fiduciary approach in any meaningful sense at all," and, ultimately, to recommend that competition and antitrust policy may be the more productive channel to explore.[24]

While the imposition of fiduciary obligations on data controllers may have been realistic in the past – when business models were less entrenched and operations less complex – today they are unlikely to broker meaningful change. For example, would a fiduciary duty model have prevented the misuse of personal data in the Cambridge Analytica scandal? Moreover, as Whitt observes, the forcible imposition of fiduciary obligations tends to produce suboptimal results, creating relationships of "grudging" care or loyalty.[25] To this end, Whitt posits that the market may one day privilege companies who voluntarily compete around the self-imposition of a positive duty of loyalty.[26] Yet such a prospect would likely depend on a clear signal that the market value of assuming a positive duty of loyalty outweighs the status quo, and on the emergence of alternative service providers. Absent important changes in the competitive landscape surrounding data controllers, neither of these developments is likely to occur, and the information fiduciary proposal will have limited impact on the power asymmetries embedded into current approaches to data governance or problems related to privacy self-management and data concentration. Structural problems such as these may require more than a light touch approach.

Data trusts

Data trusts result from the application of the common law trust to the governance of data or data rights. Trusts begin with an asset, or rights in an asset, that a "settlor" places into a trust.[27] A trust charter stipulates the purpose and terms of the trust, which exists to benefit a group of people, known as the "beneficiary". In more basic terms, a data trust creates a legal way to manage data rights for a purpose that is valuable to a beneficiary.[28]

In a data trust, data subjects would be empowered to pool the rights they hold over their personal data into the legal framework of a trust.[29] A trustee is appointed with a fiduciary obligation to manage the trust's assets in accordance with the trust charter and the interests of its beneficiaries. The trustee is accountable to the beneficiaries for the management of the trust, and has a responsibility to take legal action to protect their rights.

While there is currently no common definition for data trusts, they are often described in reference to the particular problem their proposer is aiming to solve.[30] As outlined below, governments have focused on the potential to use data trusts to promote data sharing and "responsible" innovation. The "civic data trust"[31] has been theorised as a way to protect the public interest in data governance decision-making processes. Perhaps the most expansive vision for data trusts at scale is the concept of the "bottom-up data trust",[32] which has been proposed as a way to return the power that stems from aggregated data to individuals.

To be sure, data trusts are not a governance model in and of themselves, and their effectiveness will depend on the complementary use of other tools that constitute good practice in corporate governance. Rather, data trusts provide a flexible framework that is capable of balancing a constellation of different interests or rights associated with a range of different stakeholders and use cases.

Data trusts as data-sharing vehicles

Governments' interest in data trusts has primarily focused on their potential to facilitate responsible data sharing and innovation in the AI sector. The following is a list of such proposals.

- In 2017, an independent review commissioned by the United Kingdom (UK) recommended "data

[21] Balkin, J. M. (2016). Op. cit.; Balkin, J. M. (2018). Fixing Social Media's Grand Bargain. *Yale Law School*. https://ssrn.com/abstract=3266942

[22] Balkin, J. M. (2018). Op. cit.

[23] Ibid.

[24] Khan, L., & Pozen, D. E. (2019). Op. cit.

[25] Whitt, R. (2019, 26 July). Op. cit.

[26] Ibid.

[27] McDonald, S., & Porcaro, K. (2015, 4 August). The Civic Trust. *Medium*. https://medium.com/@McDapper/the-civic-trust-e674f9aeab43; Wylie, B., & McDonald, S. (2018, 9 October). What is a Data Trust. *Centre for International Governance Innovation*. https://www.cigionline.org/articles/what-data-trust; Delacroix, S., & Lawrence, N. (2019). Op. cit.; Element AI, & Nesta. (2019). Op. cit.

[28] Element AI, & Nesta. (2019). Op. cit.

[29] Delacroix, S., & Lawrence, N. (2019). Op. cit.

[30] McDonald, S. (2019, 5 March). Op. cit.

[31] McDonald, S., & Porcaro, K. (2015, 4 August). Op. cit.

[32] Delacroix, S., & Lawrence, N. (2019). Op. cit.

trusts" as a way to "share data in a fair, safe and equitable way," adding that they would likely play an important role in growing the AI sector.[33]

- In 2018, the Open Data Institute (ODI) announced a partnership with the UK Office for Artificial Intelligence and Innovate UK to run three data trust pilots focusing on tackling illegal wildlife trade, reducing food waste and improving municipal public services.[34]
- In May 2019, the Canadian government announced a new Digital Charter[35] that referenced data trusts as a possible way to facilitate data sharing in a privacy- and security-enhancing manner for research and development purposes in areas such as health, clean technology or agribusiness. The Canadian government also included several recommendations related to data trusts in a discussion paper[36] that accompanied the Digital Charter, outlining proposals for the reform of Canada's federal privacy legislation.
- In May 2019, the Organisation for Economic Co-operation and Development adopted the OECD Principles on Artificial Intelligence (the "OECD Principles"), which recommend data trusts as a way to support the safe, fair, legal and ethical sharing of data.[37]
- In June 2019, the G20 Digital Economy Ministers incorporated the OECD's recommendation on data trusts into their "human-centred AI Principles".[38]
- Finally, that same month, the European Union High-Level Expert Group on AI (AI HLEG) published a series of policy and investment recommendations,[39] which acknowledged the need to foster the creation of "trusted data spaces" for data sharing, referencing the data trust proposals in Canada and the United Kingdom as examples.

Of particular interest in the Canadian proposal is the idea that data trusts could be used to alleviate the burden of consensual exhaustion and privacy self-management for transactions involving de-identified data. Specifically, the proposal states that "de-identified information could be processed without consent when managed by a data trust,"[40] before adding that other protections such as a "prohibition against intentional re-identification or targeting of individuals in data, or re-identification as the result of negligence or recklessness" would need to be put in place.[41] The proposal further stresses that clear linkages between statutory enforcement provisions and the oversight of a data trust would also be necessary. The proposal reflects the belief that a trustee's purpose-driven mandate, fiduciary duties and built-in accountability to beneficiaries could incentivise a level of proactive risk management that could remove the need to seek consent in subsequent data transactions with other trusts.

Civic data trusts

Civic data trusts move beyond the appointment of single trustees to build fiduciary governance structures that manage the use and sharing of rights to data on behalf of beneficiaries.[42] The purpose of a civic data trust is to embed civic values and participation processes into the governance and use of digital technologies.[43] By incorporating civic participation into the trustee organisation, civic trusts could ensure that decisions regarding the governance of data take into account evolving concepts of digital rights and the public good.[44]

Sean McDonald and Keith Porcaro identify at least three ways that a civic data trust is unique:

> [T]heir mission is to define and support the implementation of systems of public participation in decisions about data rights; the trustee organization itself must develop public participation models for its core governance decisions; and [they] can be designed to create reciprocal

[33] Hall, D. W., & Pesenti, J. (2019). *Growing the Artificial Intelligence Industry in the UK*. Government of the United Kingdom. https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk.

[34] Open Data Institute. (2018). *Data trusts: lessons from three pilots*. https://theodi.org/article/odi-data-trusts-report. Interestingly, though the ODI concluded that trust law was not necessary to advance these pilot projects, it has chosen to continue using the term "data trust", risking popular confusion as to whether or not a data trust should always imply the application of trust law, or whether the word "trust" is merely being used as a "marketing tool". See Delacroix, S., & Lawrence, N. (2019). Op. cit.

[35] Innovation, Science and Economic Development Canada. (2019). *Canada's Digital Charter: Trust in a digital world*. https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00108.html

[36] Innovation, Science and Economic Development Canada. (2019). *Strengthening Privacy for the Digital Age*. https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00107.html

[37] https://www.oecd.org/going-digital/ai/principles

[38] https://www.mofa.go.jp/files/000486596.pdf

[39] High-Level Expert Group on Artificial Intelligence. (2019). *Policy and Investment Recommendations for Trustworthy Artificial Intelligence*. https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence

[40] Innovation, Science and Economic Development Canada. (2019). *Strengthening Privacy for the Digital Age*. https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00107.html

[41] Ibid.

[42] McDonald, S., & Porcaro, K. (2015, 4 August). Op. cit.; McDonald, S. (2019, 5 March). Op. cit.; McDonald, S. (2018, 17 October). Toronto, Civic Data, and Trust. *Medium*. https://medium.com/@McDapper/toronto-civic-data-and-trust-ee7ab928fb68; Element AI, & Nesta. (2019). Op. cit.

[43] Ibid.

[44] McDonald, S., & Porcaro, K. (2015, 4 August). Op. cit.

relationships between the public (the trust), technology companies (the licensee), and technology stakeholders.[45]

In 2018, Sidewalk Labs, a subsidiary of Alphabet, proposed to establish an "independent urban data trust"[46] to help manage the data collected as part of its planned smart city development project in the city of Toronto. While Sidewalk Labs drew a lot of attention to data trusts as a concept, their proposal was criticised for its lack of detail, failure to incorporate feedback from community organisations and residents, and for not including fiduciary obligations for the proposed trustee organisation.[47]

Bottom-up data trusts

Delacroix and Lawrence propose a bottom-up[48] approach to data trusts as a way to return the power that stems from aggregated data to individuals. Data subjects would be empowered to pool their data into a trust that would champion a social or economic benefit of their choosing.[49] Professional data trustees would exercise the data rights of beneficiaries on their behalf.[50] The data trustees would act as an independent intermediary that negotiates the terms of data collection and use between data subjects and data collectors.

As more people join a data trust, the trustee's negotiating power over the data controller would grow. In similar fashion, the pooling of data rights could act as a powerful collective action mechanism against abuse by a data controller, as trustees could exercise the right of portability on behalf of all the trust's beneficiaries and withdraw the sum of the trust's data rights en masse.

Delacroix and Lawrence envision an ecosystem of data trusts in which data subjects could choose a trust that reflects their aspirations, and be able to switch trusts when needed. Delacroix and Lawrence explore the application of bottom-up data trusts in several domains, including health care, social media, genetics, financial services and loyalty programmes.[51]

Implementation challenges

Like information fiduciaries, data trusts face their own implementation challenges. First, clarity is needed regarding the legal foundation – whether in property or contract law – that would enable data subjects to pool any rights they may have to the personal data they participate in generating into a data trust of their choosing. Without changes to the current conception of data ownership, this may represent a barrier to the availability of data trusts as a viable model in the context of online platforms.

Second, data trusts would likely require a new class of professional data trustees[52] capable of balancing competing and complex interests related to data access and use. Given the increasing scale of data transactions and potential risks, however, some question whether a single trustee or even a trustee organisation would be able to discharge the trustee's duties, or if technological solutions, such as an ecosystem of "personal AI"[53] trustees, may be necessary.

Core features of trusts, including the nature and scope of fiduciary obligations (but also their governance structures and technical architectures), will need to achieve a level of standardisation for data trusts to be deployed at scale. The Hague Convention on the Law Applicable to Trusts and on their Recognition,[54] which uses a harmonised definition of a trust, and sets conflict rules for resolving problems in the choice of the applicable law, could be a natural starting point for this conversation.

Conversely, civil law jurisdictions – where the reception of trust law is more recent and its features more fluid – may be particularly well suited to the task of adapting fiduciary models of governance to the evolving field of data rights management. The Quebec Civil Code, for instance, conceives of the trust as a universality of rights affected to a particular purpose,[55] which a trustee has positive legal powers to administer on behalf of a trustee or trustee organisation. The fact that neither the settlor, trustee or beneficiary retains rights of ownership in

[45] Ibid.

[46] Harvey Dawson, A. (2018, 15 October). An Update on Data Governance for Sidewalk Toronto. *Sidewalk Labs*. https://www.sidewalklabs.com/blog/an-update-on-data-governance-for-sidewalk-toronto

[47] McFarland, M. (2019, 9 July). Alphabet's plans to track people in its 'smart city' ring alarm bells. *CNN Business*. https://www.cnn.com/2019/07/09/tech/toronto-sidewalk-labs-google-data-trust/index.html; Cecco, L. (2019, 11 September). 'Irrelevant': report pours scorn over Google's ideas for Toronto smart city. *The Guardian*. https://www.theguardian.com/cities/2019/sep/11/irrelevant-panel-pours-scorn-over-googles-ideas-for-toronto-smart-city; Ryan, A. (2019, 24 June). Here's how the Quayside data trust should operate. *The Star*. https://www.thestar.com/opinion/contributors/2019/06/24/heres-how-the-quayside-data-trust-should-operate.html

[48] Delacroix, S., & Lawrence, N. (2019). Op. cit.

[49] Delacroix, S., & Lawrence, N. (2019). Op. cit.; Element AI, & Nesta. (2019). Op. cit.

[50] Delacroix, S., & Lawrence, N. (2019). Op. cit.

[51] Ibid.

[52] Ibid.

[53] Whitt, R. (2019, 26 July). Op. cit.; Wylie, B., & McDonald, S. (2018, 9 October). Op. cit.; McDonald, S. (2019, 5 March). Op. cit.

[54] https://assets.hcch.net/docs/8618ed48-e52f-4d5c-93c1-56d58a610cf5.pdf

[55] Emerich, Y. (2013). The Civil Law Trust: A Modality of Ownership or an Interlude in Ownership? In L. Smith (Ed.), *The Worlds of the Trust*. Cambridge University Press; Smith, L. (Ed.). (2012). *Re-imagining the Trust: Trusts in Civil Law*. Cambridge University Press.

the asset, moreover, may remove legal barriers to the sharing or pooling of data rights.

Other important issues that require further research and testing include the accountability and liability procedures that will need to be developed in the context of data misuse, the integrity of licensing[56] in data supply chains to ensure organisations seeking to use data from a data trust do so in accordance with its terms, and the role of public institutions in defining and conducting oversight of high-level requirements for data trusts in particular sectors, for instance, to ensure data trusts themselves are not manipulated to form new oligopolies of power. Public awareness regarding the opportunity but also the risks associated with the management of their data rights, is another.

Way forward?

It is important to recall that this is not the first time society has successfully devised checks and balances capable of addressing problematic concentrations of power. Good governance in democratic political regimes – the separation of powers, for instance – has helped safeguard individual rights and advance the public good while providing the certainty needed for innovation and economic growth.

The information fiduciary model represents an important first step in recognising that digital platforms should hold obligations towards internet users that are proportionate to the risk of harm they may potentially cause. Nevertheless, practical limitations related to the ability of fiduciaries to manage competing duties to both data subjects and their shareholders may impact their ability to build public trust.

If the data trust agenda appears more ambitious, this is as much an indication of data trusts' promising features as it is a reflection of the public's aspirations for data governance in the digital age: representation, shared rights, accountability and remedy.[57] Not only are these just demands, but meeting them may help create an environment for the digital economy that is sustainable in the long term.[58] So while data trusts may face a number of concrete implementation challenges related to their legal, governance and architectural foundations, they remain an indisputably promising innovation that merits greater investment – narrowly, as tools that could facilitate fair and ethical data sharing to alleviate burdens related to consent and privacy self-management; and broadly, as ways to empower the public to participate in decisions regarding the use of their personal data, and to collectively seek redress in cases of harm.

[56] Benjamin, M., Gagnon, P., Rostamzadeh, N., Pal, C., Bengio, Y., & Shee, A. (2019, 21 March). Towards Standardization of Data Licenses: The Montreal Data License. *arXiv*. https://arxiv.org/abs/1903.12262

[57] Surman, M. (2019, 13 May). Consider this: AI and Internet Health. https://marksurman.commons.ca/2019/05/13/consider-this-ai-and-internet-health

[58] Ibid.

Defending food sovereignty in the digital era

GRAIN
www.grain.org

Introduction

In May 2019, the BBC launched a documentary series titled "Follow the Food"[1] highlighting the current and future challenges of the world food system, from the growing population to the climate crisis to the generational turnover of farmers. According to the documentary, produced in association with Corteva (the new company created from the Dow-DuPont merger),[2] the way to "feed the planet" is by embracing new technology developments. This conclusion is based on the same kind of assumption that has been used over the past 50 years to impose agrochemicals, biotechnology and energy-intensive industrial agriculture on the global food system, while experts and reality on the ground show the contrary: small-scale farms and agroecology practices are able to produce 70% of the food being consumed globally on less than a quarter of all farmland – and are crucial to tackle the climate disaster.[3]

The documentary series shows stories from around the world of high-tech farming, like the use of robotics, digital equipment and blockchain technology. We hear from scientists, technology officers and CEOs in warehouses, corporate headquarters and start-ups. In some scenes, we see an eerie way of producing "food" handled with extra caution: a scientist wearing a hazmat suit to manipulate a vegetable-like form inside a petri dish and a vertical farm with controlled artificial climate and LED light. You won't see many farmers with their hands touching the soil, working in their sunny field growing food that will also be consumed at their dining tables.

Despite often being seen as a backwards economic sector in this era of digitalisation, agriculture in fact often serves as a test-bed for new technology. The implementation of digital technologies in agriculture, such as artificial intelligence (AI), drones and e-commerce, is growing at light speed. Agribusiness corporations are teaming up with digital technology companies, or creating their own digital arms, to create products and services targeting the food supply chain.

But these developments have brought up some important questions. Take the example of AI in agricultural production: is it better to have machines and automation replace farm workers suffering from poor living and working conditions? How will these workers be guaranteed a dignified source of living after their role is digitalised? Or the way e-commerce corporations integrate the vertical supply chain: will this lead to a more concentrated global food supply? Where is the place for small food producers and small vendors in this landscape?

In this report, we aim to stimulate reflection and discussion on the application of digital technologies in the food system using a food sovereignty[4] framework. We explore different areas being impacted by technology, including production, distribution and commercialisation, and highlight the impact on small-scale farmers and local markets. In the race for increased productivity and efficiency, and most of all profits, we point out the ways that this agritech revolution is displacing and disadvantaging the farmers – and the methods and knowledge systems – that actually feed the world.

[1] BBC. (2019). Follow the Food. https://www.bbc.com/reel/playlist/follow-the-food?vpid=p07bj7pv

[2] DowDuPont recently announced it was de-merging, creating three of the largest chemical companies in the world, with Corteva focused on agricultural chemicals. Associated Press. (2019, 3 June). DuPont begins new life after more than 2 centuries. *Business Insider*. https://www.businessinsider.com/dupont-begins-new-life-after-more-than-2-centuries-2019-6

[3] IPCC Working Group III. (2019). *IPCC Special Report on Climate Change and Land*. London: IPCC https://www.ipcc.ch/site/assets/uploads/2019/08/SRCCL-leaflet.pdf and GRAIN. (2014). *Hungry for land: Small farmers feed the world with less than a quarter of all farmland*. https://www.grain.org/en/article/4929-hungry-for-land-small-farmers-feed-the-world-with-less-than-a-quarter-of-all-farmland

[4] "Food sovereignty", a term coined by members of Via Campesina in 1996, asserts that the people who produce, distribute and consume food should control the mechanisms and policies of food production and distribution, rather than the corporations and market institutions they believe have come to dominate the global food system. It also encompasses the right of peoples to healthy and culturally appropriate food produced through ecologically sound and sustainable methods. https://en.wikipedia.org/wiki/Food_sovereignty

Sign in front of Fujitsu vegetable factory. PHOTO: GRAIN, 2016

Agriculture digitalisation

On a hectare of land located on the outskirts of Hanoi in Vietnam stands a large greenhouse. This is not a typical greenhouse: there is a sign that reads "vegetable factory" in front, and inside not one person can be found on a regular day – there are only lettuces growing inside glass-sealed rooms and stems of leafless tomatoes linked to machines that regulate the intake of fertilisers, water and other minerals needed for the growth of the tomato. The greenhouse can be operated from Tokyo, almost 4,000 kilometres away from Hanoi. This "vegetable factory" belongs to Fujitsu, a Japanese company more familiar to many of us for their computer and printer products.

Fujitsu is just one of many information and communications technology (ICT) companies that have jumped on the bandwagon to digitalise agriculture, offering their services to, in their words, "contribute to the further development of [the] agricultural sector."[5] Ever since the Green Revolution led to agricultural industrialisation, efforts have been made to continuously improve management of large-scale industrial farms or plantation areas with a minimum number of workers or human hours working on the farm while maximising yields and profits. Agriculture machinery and technology are evolving from a simple threshing machine to today's robotic farming where everything can be controlled from your computer or phone at a remote location.

Today there is an unprecedented level of investment from digital and ICT companies that traditionally have not worked in the agricultural sector. As vertical and horizontal integration happens along the food supply chain, corporations see the potential of gathering agriculture data, enabled by the development of digital technology, to monopolise the supply chain and maximise their profit.

The factory farm that Fujitsu developed is part of a collaboration with one of Japan's leading food retailers, Aeon Agri Create. This company has established ICT-based farming in Japan and multiple Southeast Asian countries.[6] In Japan, Aeon directly manages 15 farms covering over 200 hectares. Aeon's farms use the Fujitsu Akisai cloud computing service as the basis for daily farm operations and monitoring. The collaboration is aimed at generating more data that Aeon can then use to secure a more stable supply of produce for Aeon Group stores; this way Aeon can ensure their vegetable supply from their contract farms.

AI and digitalisation of agriculture is not just about automation. The transformation of farm operations is further-reaching than replacing farmers with robots to increase profits and decrease the need for human workers. It is about capturing data, valuable information crucial to ensure the continuation of agricultural production.

[5] Fujitsu. (2015, 8 December). Fujitsu and FPT Implement Smart Agriculture in Vietnam. https://www.fujitsu.com/global/about/resources/news/press-releases/2015/1208-01.html

[6] Fujitsu. (2016). *Customer case study: Transforming farming by exploiting ICT*. https://www.fujitsu.com/sk/Images/CS_2016Nov_Aeon_Agri_Create_Co_Ltd_Eng_v.1.pdf

In an OECD paper[7] about the digital transformation of the agriculture and food system, it was explained that the agricultural sector is both an important consumer and supplier of data. Farm data is particularly important to facilitate global value chain integration. The concern is that essential information like soil conditions, climate and water quality that should be publicly accessible to farmers, as it is necessary for agricultural production, can easily be extracted, stored, privatised and monopolised by a handful of agribusiness and digital companies.

A case in point: many were curious when Monsanto, a giant multinational agrochemical company, decided to buy The Climate Corporation, a climate data company, in 2013. But The Climate Corporation, as they claim on their website, has a platform technology that combines weather monitoring, agronomic data modelling and high-resolution weather simulations that can inform about bad weather that may impact farm profits and provide advice for risk management solutions needed.[8] Owning technology like this means profit: as Monsanto said, "In the face of increasingly volatile weather, the global $3 trillion agriculture industry depends on [such] technologies to help stabilize and improve profits."[9] Information that could be very useful for millions of small-scale farmers to produce food and feed the world is being privatised for the sake of profit.

Vertical vegetable farm in Hanoi. PHOTO: GRAIN, 2016

E-commerce taking bigger slices in the food supply chain

> *Hawkers fill the last mile for sale of small products, connecting small producers and farmers to middle and working class consumers. With their micro-reach with data, Walmart and Amazon will also sweep up this link and render millions jobless and reduce consumer choice.* - Shaktimaan Ghosh, General Secretary, India National Hawker Federation[10]

In June 2017, Amazon, the world's third-largest e-commerce company, announced its acquisition of Whole Foods Market for USD 13.7 billion.[11] The deal made Amazon the largest organic food retailer in the United States (US) overnight. Amazon's move follows its biggest competitor, the world's largest e-commerce company, China-owned Alibaba, which invested USD 1.25 billion in the Chinese online food delivery service Ele.me in late 2015, then set up its own fresh produce stores, Hema supermarket. Alibaba invested USD 12.7 billion into physical retail stores.

When it comes to food distribution, 30 global supermarket chains already control a third of the global retail food market.[12] Combined with e-com-

[7] Jouanjean, M. (2019). *Digital Opportunities for Trade in the Agriculture and Food Sectors*. Paris: OECD Publishing. https://www.oecd-ilibrary.org/docserver/91c40e07-en.pdf?expires=1566464711&id=id&accname=guest&checksum=054D28ADB45F2C803EACD0C334B2499C

[8] https://www.climate.com

[9] Monsanto. (2013, 1 November). Monsanto Completes Acquisition of The Climate Corporation. https://monsanto.com/news-releases/monsanto-completes-acquisition-of-the-climate-corporation

[10] Hawkers Joint Action Committee. (2018, 27 August). Traders, farmers, workers and citizens groups decry green-light to Walmart-Flipkart deal – Call on Quit India day for joint struggle against foreign e-commerce. Supermarket Watch Asia. https://www.grain.org/en/article/6029-modern-retail-and-market-concentration-in-thailand

[11] Green, D. (2019, 2 May). How Whole Foods went from a hippie natural foods store to Amazon's $13.7 billion grocery weapon. *Business Insider*. https://www.businessinsider.sg/whole-foods-timeline-from-start-to-amazon-2017-9/?r=US&IR=T

[12] GRAIN. (2014, 17 September). Food sovereignty for sale: supermarkets are undermining people's control over food and farming in Asia. https://www.grain.org/en/article/5010-food-sovereignty-for-sale-supermarkets-are-undermining-people-s-control-over-food-and-farming-in-asia

merce or grocery e-retail, the corporate control gets even more concentrated in this sector. There are only a handful of companies who control regional and even global supply chains, with Walmart, Alibaba and Amazon at the top of the food chain, as it were.[13] This of course brings new challenges to small farmers and traders, especially when their "competitors" are invisible and able to deliver products from anywhere around the world to the doorstep of the consumers.

The growing encroachment of e-commerce in food distribution is happening everywhere, as start-up companies have competed to get their way into this sector over the past few years. The merging of e-commerce with existing grocery stores and supermarket chains is what Alibaba's founder, Jack Ma, describes as the "new retail". But what's behind these investments? Why does e-commerce invest so much money in food distribution and retail?

Compared to other e-commerce purchases, groceries are habitual and frequent. People shop for groceries daily, weekly or monthly. Contrary to predictions made a few years back, there is a growing awareness of the fact that online channels alone can't serve consumers adequately, especially in food retail.[14] Merging offline and online stores covers consumer needs, and also provides greater access to consumer data and purchasing habits. Data helps companies develop vertical integration for their own private brands which, in the long run, will affect how and where these companies source their products in order to get the lowest price.

Furthermore, e-commerce companies like Amazon and Alibaba use automation as a key strategic advantage in their overall grocery strategy, entailing the risk of a major loss of jobs. Alibaba has begun drone-based deliveries to hundreds of customers in three of China's major cities – Beijing, Shanghai and Guangzhou.[15] Meanwhile, after its acquisition of Whole Foods Markets, Amazon plans to add robot workers in Whole Foods Market warehouses to reduce costs. Amazon already uses robots in other business sectors.

Although we can't really tell future scenarios, in-depth analysis already predicts that increasing automation and digital supply chain management of agricultural production, processing and trade will eventually create a profound impact on rural society – with the risk of leaving large numbers out of employment, especially in developing countries.[16] E-commerce is changing traditional value chains and creating a closed environment where inputs, logistics and markets are centrally controlled.[17]

Food sovereignty 2.0? Stepping up to the tech challenge

Whether we like it or not, technology is advancing and influencing many aspects of agriculture and food distribution. Agri-digitalisation is a fast-growing industry that needs to be better understood. Combine this with trade agreements designed to back up this growing industry, and the question remains how to ensure that small-scale food producers, informal traders and local markets are not swept away as collateral damage of the hyper-competition between e-commerce, retail trade and agritech.

The transformation of agriculture and rural economies with digital development, automation and other computing technology brings inequality at different levels and in different ways. New technologies come with big price tags, making much of it inaccessible for smallholder peasants, thereby enforcing the competitive advantage of industrial, corporate-controlled agriculture. Food supply chains, from production to processing to marketing, will be further integrated and concentrated. Smallholders may struggle to engage with these changing systems, or be displaced from their lands and livelihoods completely by drones and robots.

We must continue defending food sovereignty, envisioning and building strategies and alternatives to the industrial food system in this digital era. This could include using ICTs to improve farmer-to-farmer exchanges or connecting small producers with small traders and hawkers in order to move forward in a way that strengthens social, community-based and public food systems, and assures the survival of small-scale food producers and local markets.

[13] GRAIN. (2018, 31 May). Top e-commerce companies move into retail. *Supermarket Watch Asia*. https://www.grain.org/en/article/5957-top-e-commerce-companies-move-into-retail

[14] GRAIN. (2018, 31 May). Online to offline food retailers' transition in China: what's left for farmer's markets? *Supermarket Watch Asia*. https://www.grain.org/en/article/5957-top-e-commerce-companies-move-into-retail

[15] Kelion, L. (2015, 4 February). Alibaba begins drone delivery trials in China. *BBC*. https://www.bbc.com/news/technology-31129804

[16] Norton, A. (2017). *Automation and inequality: The changing world of work in the global South*. IIED. https://pubs.iied.org/pdfs/11506IIED.pdf

[17] Gurumurthy, A., & Chami, N. (2019, 2 April). Why the dominant digital trade paradigm will not work for women in the global South. *IT for Change*. https://medium.com/commentary-itforchange/why-the-dominant-digital-trade-paradigm-will-not-work-for-women-in-the-global-south-d053cd3b470f

AI policing of people, streets and speech

Luis Fernando García Muñoz
R3D: Red en Defensa de los Derechos Digitales
www.r3d.mx

Introduction

An area in which artificial intelligence (AI) systems are producing a direct effect on the enjoyment of human rights today is the use of these systems for the alleged intention of protecting public safety and making justice systems more efficient and objective.

The rapid proliferation of these systems has, however, not only lacked a robust public discussion, but many of its impacts have not been evaluated before implementation. Therefore, it is crucial to review and examine the actual impacts of AI systems used for policing, surveillance and other forms of social control.

Predictive policing

Increasingly, law enforcement agencies have announced the use of AI with the purpose of predicting areas that are more prone to crime or even predicting which persons are more likely to be involved in a crime, both as perpetrators and as victims.[1] These predictions play an important role in decisions such as the deployment of police officers in those areas or a determination on the pre-trial detention of a suspect.

These tools rely on multiple sources of data such as criminal records, crime statistics, the demographics of people or neighbourhoods, and even information obtained from social media.[2]

As numerous reports have demonstrated, many of these data sets are flawed and biased in ways which can reinforce racial and other types of discrimination.[3] Moreover, predictions made by AI systems trained with skewed data are often seen as "neutral" or "objective", further ingraining discriminatory and abusive practices.

Often, predictive policing programmes are implemented without transparency, accountability or community participation in the decisions around their implementation[4] or in the evaluation and oversight of their impacts, further limiting the detection and remedy of undesired outcomes.

Social ranking

Some applications of AI systems are more straightforward in their repressiveness and authoritarianism. Take, for example, China's "social credit system", by which every person receives a score that factors in everyday behaviours, such as shopping habits or online opinions.[5] The score given is then used to determine access to services and jobs or may even prompt questioning or arrest by the police, thus influencing behaviour and social docility.

In some parts of China, massive amounts of data on each person, such as location data, data from ID cards, CCTV footage and even electricity consumption, are gathered, aggregated and processed to identify behaviour and characteristics deemed as suspicious by the state. This may also result in interrogation by the police, and even prolonged detention, often without any explanation given.[6]

[1] Southerland, V. (2018, 9 April). With AI and criminal justice, the devil is in the data. *ACLU*. https://www.aclu.org/issues/privacy-technology/surveillance-technologies/ai-and-criminal-justice-devil-data

[2] Buntin, J. (2013, October). Social Media Transforms the Way Chicago Fights Gang Violence. *Governing*. https://www.governing.com/topics/urban/gov-social-media-transforms-chicago-policing.html

[3] See, for example, Richardson, R., Schultz, J., & Crawford, K. (2019). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. *New York University Law Review Online, Forthcoming*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3333423 and Babuta, A., Oswald, M., & Rinik, C. (2018). *Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges*. London: Royal United Services Institute for Defence and Security Studies. https://rusi.org/sites/default/files/201809_whr_3-18_machine_learning_algorithms.pdf.pdf

[4] Stanley, J. (2018, 15 March). New Orleans Program Offers Lessons In Pitfalls Of Predictive Policing. *ACLU*. https://www.aclu.org/blog/privacy-technology/new-orleans-program-offers-lessons-pitfalls-predictive-policing

[5] Wang, M. (2017, 12 December). China's Chilling 'Social Credit' Blacklist. *Human Rights Watch*. https://www.hrw.org/news/2017/12/12/chinas-chilling-social-credit-blacklist

[6] Wang, M. (2019, 1 May). China's Algorithms of Repression. *Human Rights Watch*. https://www.hrw.org/report/2019/05/01/chinas-algorithms-repression/reverse-engineering-xinjiang-police-mass-surveillance

Facial recognition surveillance

One of the most widespread and fast-growing applications of AI systems for policing is the use of facial recognition software for the surveillance of public spaces. The main capability of these systems is the identification of a person by comparing video images with existing databases, for example, mug shot, driver's licence or ID card databases. In the absence of clear video footage, even sketches or photos of celebrities described as having a resemblance to a suspect have been entered into the databases.[7]

Facial recognition software is usually used to analyse live video feeds captured by CCTV cameras, but it has been found to also be used to analyse recorded video footage. Some systems produce logs that register the historic detection of a person throughout a surveillance system, usually recording the location, time, date and relationships associated with each detection, and some systems claim to be able to even detect emotions such as happy, sad, calm, angry or surprised.[8]

The scale of this surveillance is unprecedented. For example, in the United States (US) it is estimated that approximately half of all residents are captured in the law enforcement facial recognition network. Also, the fact that this surveillance is difficult to escape, since it occupies public spaces, results in a particularly invasive tool with far-reaching consequences for participation in public life.

Facial recognition surveillance often lacks specific and robust regulation detailing the process and requirements to conduct a search through the system or establishing rules with regard to which individuals' faces can be included in the databases used and for how long, among other aspects. This has often led to serious abuse. For example, in the US county of Maricopa, Arizona, the complete driver's licence and mug shot databases of the country of Honduras were included in the database, which clearly indicates an intention to target a group of people with certain ethnic or national characteristics.

However, the potential for abuse is not limited to the arbitrary or discriminatory inclusion of databases in the system. There is a real risk that these tools are used by law enforcement to spy on people for reasons that have nothing to do with public safety. It has been reported that several law enforcement databases have been inappropriately accessed to spy on romantic partners, family members and journalists.[9]

The vulnerability of databases used by these systems adds an important layer of risk, particularly when the data that could be stolen is biometric. Differently from other types of data, like passwords, which can be modified if compromised, the effects of stolen biometric data are far more difficult to remediate. This risk has already materialised on multiple occasions. For example, in 2019, it was reported that the database of a contractor for the US Customs and Border Protection agency was breached, compromising photographs of travellers and licence plates.[10] Also in 2019, the fingerprints of over a million people, as well as facial recognition information, unencrypted usernames and passwords, and personal information of employees were discovered on a publicly accessible database for a company used by the Metropolitan Police, defence contractors and banks in the United Kingdom (UK).[11]

Additionally, facial recognition surveillance has been shown to be highly inaccurate. In the UK, an investigation revealed that implementations of the technology for certain events resulted in more than 90% of the matches being wrong.[12] The proneness of facial recognition surveillance to the misidentification of individuals has already resulted in the detention of innocent people[13] and produced a waste of law enforcement resources that could be allocated to more useful and adequate policing activities.

This technology has been shown to be particularly prone to misidentifying people of colour, women and non-binary individuals. For example, a study of three different kinds of facial analysis software demonstrated that while the error rate in determining the gender of light-skinned men was 0.8%, the error rate for darker-skinned women reached up to 34% in

[7] Garvie, C. (2019, 16 May). Garbage in, garbage out: Face recognition on flawed data. *Georgetown Law Center on Privacy & Technology.* https://www.flawedfacedata.com

[8] Dahua Technology. (2018). *AI Creates Value: Dahua AI Product & Solution Introduction.* https://www.dahuasecurity.com/asset/upload/download/20180327/2018_V1_Artificial-Intelligence%28OP%29.pdf

[9] Gurman, S. (2016, 28 September). AP: Across US, police officers abuse confidential databases. *AP News.* https://apnews.com/699236946e3140659fff8a2362e16f43

[10] Harwell, D., & Fowler, G. A. (2019, 10 June). U.S. Customs and Border Protection says photos of travelers were taken in a data breach. *The Washington Post.* https://www.washingtonpost.com/technology/2019/06/10/us-customs-border-protection-says-photos-travelers-into-out-country-were-recently-taken-data-breach

[11] Taylor, J. (2019, 14 August). Major breach found in biometrics system used by banks, UK police and defence firms. *The Guardian.* https://www.theguardian.com/technology/2019/aug/14/major-breach-found-in-biometrics-system-used-by-banks-uk-police-and-defence-firms

[12] Big Brother Watch. (2019, May). Face Off. https://bigbrotherwatch.org.uk/all-campaigns/face-off-campaign

[13] Todo Noticias. (2019, 31 July). De un DNI mal cargado a una cara parecida: las víctimas del sistema de reconocimiento facial en Buenos Aires. *TN.* https://tn.com.ar/policiales/de-un-dni-mal-cargado-una-cara-parecida-las-victimas-del-sistema-de-reconocimiento-facial-en-buenos_980528

some cases.[14] This gender and racial bias creates an aggravated risk of perpetuating the discriminatory effects that policing and the criminal justice system have been found to be responsible for.

Despite the flaws and risks that facial recognition surveillance poses for the exercise of human rights, this technology is aggressively being pushed around the globe, including in countries with poor human rights records and a lack of robust institutional counterweights, which exacerbates the risk of abuse.

For example, facial recognition surveillance has been introduced or is already operating in Latin American countries like Argentina,[15] Brasil,[16] Chile,[17] Paraguay[18] and México[19] and African countries like Uganda, Kenya and Zimbabwe.[20] Besides the UK, facial recognition applications have been reported in Denmark[21] and Germany.[22]

Some jurisdictions are responding with regulations to limit the rapid proliferation of this technology. For example, it has been reported that the European Commission is preparing regulation[23] and the US cities of San Francisco,[24] Oakland[25] and Somerville[26] have all banned the police from using the technology. However, the vast majority of facial recognition systems remain unregulated and lack meaningful transparency and accountability mechanisms.

Impact on public protest

One strong concern about the use of AI for policing and surveillance of the public space is its impact on the exercise of the right to protest. This impact has recently become more evident, for example, in Hong Kong, where frequent protesting has encountered heavy resistance by the police. One of the tools that the Hong Kong police have used to try to thwart the protests has been the use of facial recognition cameras to attempt to identify the participants.[27]

Protesters in Hong Kong have resorted to multiple tactics to try to resist the heavy surveillance imposed on them – from using masks, certain kinds of makeup and umbrellas to try to cover their faces, to laser pointers aimed at obfuscating the operation of surveillance cameras, to even taking them down and destroying them.[28] The tension has prompted the Hong Kong government to use emergency powers to ban the use of masks[29] so facial recognition surveillance cameras are able to identify and track people participating in the protests. It is quite extraordinary that regulation on what people can wear is so strongly aimed at making an AI system work properly.

While often dismissed, privacy in public spaces is rapidly becoming more recognised as an essential value for the exercise of public protest. For example, the United Nations Human Rights Committee's (HRC) draft general comment on article 21 of the International Covenant on Civil and Political Rights (ICCPR) regarding the right of peaceful assembly[30] makes mention of the importance of the right to express your opinions anonymously, including in public spaces. It points out that even when "anonymous

[14] Hardesty, L. (2018, 11 February). Study finds gender and skin-type bias in commercial artificial-intelligence systems. *MIT News*. https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212

[15] Ucciferri, L. (2019, 23 May). #ConMiCaraNo: Reconocimiento facial en la Ciudad de Buenos Aires. *Asociación por los Derechos Civiles*. https://adc.org.ar/2019/05/23/con-mi-cara-no-reconocimiento-facial-en-la-ciudad-de-buenos-aires

[16] Xinhua (2019, 2 March). Brasil estrena cámaras de reconocimiento facial coincidiendo con inicio del Carnaval. *Xinhua*. spanish.xinhuanet.com/2019-03/02/c_137862459.htm

[17] Garay, V. (2018, 16 November). Sobre la ilegalidad de la implementación de un sistema de reconocimiento facial en Mall Plaza. *Derechos Digitales*. https://www.derechosdigitales.org/12623/sobre-la-ilegalidad-de-la-implementacion-de-un-sistema-de-reconocimiento-facial-en-mall-plaza

[18] ABC Color. (2019, 11 July). Reconocimiento facial: nueva estrategia para combatir la delincuencia. *ABC Color*. https://www.abc.com.py/nacionales/2019/07/11/reconocimiento-facial-nueva-estrategia-para-combatir-la-delincuencia

[19] R3D. (2019, 22 April). Gobierno de Coahuila anuncia compra de cámaras con reconocimiento facial. *Red en Defensa de los Derechos Digitales*. https://r3d.mx/2019/04/22/gobierno-de-coahuila-anuncia-compra-de-camaras-con-reconocimiento-facial

[20] Wilson, T., & Murguía, M. (2019, 20 August). Uganda confirms use of Huawei facial recognition cameras. *Financial Times*. https://www.ft.com/content/e2058ode-c35f-11e9-a8e9-296ca66511c9

[21] Mayhew, S. (2010, 1 July). Danish football stadium deploys Panasonic facial recognition to improve fan safety. *Biometric Update*. https://www.biometricupdate.com/201907/danish-football-stadium-deploys-panasonic-facial-recognition-to-improve-fan-safety

[22] Delcker, J. (2018, 13 September). Big Brother in Berlin. *Politico*. https://www.politico.eu/article/berlin-big-brother-state-surveillance-facial-recognition-technology

[23] Khan, M. (2019, 22 August). EU plans sweeping regulation of facial recognition. *Financial Times*. https://www.ft.com/content/90ce2dce-c413-11e9-a8e9-296ca66511c9

[24] Conger, K., Fausset, R., & Kovaleski, S. (2019, 14 May). San Francisco Bans Facial Recognition Technology. *The New York Times*. https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html

[25] Ravani, S. (2019, 17 July). Oakland bans use of facial recognition technology, citing bias concerns. *San Francisco Chronicle*. https://www.sfchronicle.com/bayarea/article/Oakland-bans-use-of-facial-recognition-14101253.php

[26] Wu, S. (2019, 27 June). Somerville City Council passes facial recognition ban. *Boston Globe*. https://www.bostonglobe.com/metro/2019/06/27/somerville-city-council-passes-facial-recognition-ban/SfaqQ7mG3DGulXonBHSCYK/story.html

[27] Mozur, P. (2019, 26 July). In Hong Kong Protests, Faces Become Weapons. *The New York Times*. https://www.nytimes.com/2019/07/26/technology/hong-kong-protests-facial-recognition-surveillance.html

[28] Abacus. (2019, 30 August). Why are Hong Kong protesters targeting lamp posts? *South China Morning Post*. https://www.scmp.com/tech/big-tech/article/3024997/why-are-hong-kong-protesters-targeting-lamp-posts

[29] Liu, N., Woodhouse, A., Hammond, G., & Meixler, E. (2019, 4 October). Hong Kong invokes emergency powers to ban face masks. *Financial Times*. https://www.ft.com/content/845056ca-e66a-11e9-9743-db5a370481bc

[30] UN Human Rights Committee. (2019). *Draft General Comment No. 37 on Article 21 (Right of Peaceful Assembly) of the International Covenant on Civil and Political Rights*. https://www.ohchr.org/EN/HRBodies/CCPR/Pages/GCArticle21.aspx

participation and the wearing of face masks may present challenges to law enforcement agencies, for example by limiting their ability to identify those who engage in violence," masks or other mechanisms to hide the identity of participants in a protest "should not be the subject of a general ban."

The HRC further justifies the protection of anonymity in the context of a protest by noting that "concerns about identification may deter people with peaceful intentions from participation in demonstrations, or face masks could be part of the chosen form of expression."

It is in this context that the HRC recognises the importance of the protection of privacy in public places from technologies like facial recognition by stating that "the mere fact that participants in assemblies are out in public does not mean that their privacy cannot be infringed, for example, by facial recognition and other technologies that can identify individual participants in mass assemblies."

Content moderation

As online spaces increasingly become essential for deliberation and the formation of public opinion, the power wielded by the biggest internet platforms on deciding what can and cannot be expressed by the users of their services has become more and more relevant.

Increasing pressure for stricter content moderation, for example, with the aim of curbing copyright infringement, child pornography, incitement to violence and other categories of speech, has produced surging investment in the development of AI tools capable of detecting and removing infringing content.

While AI has been touted as a solution to the serious harms that content moderation produces for workers entrusted to carry out this task,[31] the risk of false positives and the increased obstacles for transparency and accountability pose a serious risk for freedom of expression online.

As UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression David Kaye mentioned in a report on the implications of AI technologies for human rights in the information environment, "AI-driven content moderation has several limitations, including the challenge of assessing context and taking into account widespread variation of language cues, meaning and linguistic and cultural particularities."[32] As a result, the use of AI for content moderation is susceptible to making many mistakes when removing content.

Increasing threats of regulation and sanctions for platforms that underperform in removing content deemed as infringing by regulators in different jurisdictions can also lead to incentives for over-blocking as a means of protection against liability.

These risks become exacerbated by the difficulties in detecting the false positives that automated content removals create. As the special rapporteur points out, "AI makes it difficult to scrutinize the logic behind content actions." This is even more so the case when AI is expected to be used to moderate content as it is uploaded to the platforms,[33] without even allowing the content to be published, thus creating less awareness of the removal of content and adding even more opacity and difficulty to remediate errors or abuse caused by the content moderation systems.

The path forward

While AI should not be demonised as a technology, and many applications can contribute to social good, it is important to recognise the impacts that some applications can have on the exercise of human rights.

Policing, criminal justice systems and information flows are already flawed in complex ways, often reproducing systemic injustice against vulnerable groups.

Therefore, it is essential that AI is not deployed without regard of the context, the risks and the ways in which it can not only worsen the discrimination and violence against certain groups, but make these considerably more difficult to reverse.

Until the applications of AI for the attainment of security are informed by evidence, properly designed for human rights compliance and have multiple mechanisms to guarantee transparency and independent oversight, they should not be deployed at the accelerated pace that we see today.

Responsibility must prevail against the politically convenient idea of treating AI as a magical recourse to solve all problems real, perceived or artificially manufactured.

[31] Newton, C. (2019, 25 February). The Trauma Floor: The secret lives of Facebook moderators in America. *The Verge*. https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona

[32] United Nations General Assembly. (2018). *Report prepared by the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, David Kaye, on implications of artificial intelligence technologies for human rights in the information environment, focusing in particular on rights to freedom of opinion and expression, privacy and non-discrimination*. A/73/348. https://undocs.org/A/73/348

[33] Porter, J. (2019, 21 March). Upload filters and one-hour takedowns: The EU's latest fight against terrorism online, explained. *The Verge*. https://www.theverge.com/2019/3/21/18274201/european-terrorist-content-regulation-extremist-terreg-upload-filter-one-hour-takedown-eu

Decolonising AI: A transfeminist approach to data and social justice

Paz Peña[1] and Joana Varon[2]
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 CONIN[5] 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-women[8] 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 transfeminist[14] 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.eltribuno.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. *Boletin 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 Mejias[24] 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/phpEfR4QP/original.pdf

[33] Sociedad Civil de Chile Defensora de los Derechos Humanos del Niño et al. (2019, 28 January). Dia 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 tranfeminist 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 combined[42] 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

Automating informality: On AI and labour in the global South

Noopur Raval[1]

Introduction

Since the publication of Frey and Osborne's important paper in 2013,[2] announcing that approximately 47% of existing jobs in the United States (US) were susceptible to automation, the "future of work" with a focus on questions of employment has become a core research concern among research and policy organisations worldwide. Subsequently, more such regional trend studies[3] have highlighted the impact of automation on employment in different parts of the world, especially in the global North. Various kinds of artificial intelligence (AI)-enabled technologies are already transforming global logistics and supply chains as well as other job domains such as accounting, business processing and others – areas that until now heavily relied upon multiple human agents at each step.

Automated processes, as well as the creation of dynamic, on-demand labour pools in real time, have begun to replace permanent and long-term contract jobs in the global North, leading to serious concerns and considerations about increased precarity (the lack of job security), informalisation, wage theft, granular surveillance and exploitation of human workers in platformised work. Moreover, many developed economies are also facing an ageing workforce,[4] raising concerns with respect to reskilling and care for the elderly once this workforce retires. However, a majority of countries in the global South are witnessing a "youth bulge", where, for instance, the median age of the entire African continent is 19.4 years (2019). India, where this essay's empirical focus lies, is alone home to 600 million young people (between ages 15 and 24).[5] Questions of skilling the youth, combating increasing unemployment and increasing women's participation in work remain big political challenges in developing economies.

Most importantly, the endurance and dominance of informal work remains the biggest distinguishing factor of global South labour markets. While there is no one clear definition for informal work, it includes atypical, non-standard, self-generated and home-based work that is unregistered (or too small to be registered) and is not adequately regulated and/or taxed, often because of the lack of a clear employee-employer relationship.[6] As per a 2018 International Labour Organization (ILO) report, more than 60% of the world's employment happens in the informal economy.[7] While the estimates vary, the informal economy in India still accounts for more than 80% of non-agricultural employment. Widespread informality in labour also makes it hard to enforce minimum wage or decent work standards.

In this sense, the combination of work precarity as the norm as well as the social and cultural constraints on people's abilities to find work demand a *human-centric* orientation to the questions of AI and labour in the global South. In line with this, some existing conversations are already adopting a developmental lens to think about how AI may be harnessed to democratise education, for training and skills development, to provide better health care and to help people find jobs, among other things.

Before moving forward, two points are worth clarifying. First, it is useful to understand that since "artificial intelligence" as a blanket term could refer to varying levels and kinds of big data and algorithmic innovations, in this report I focus specifically on algorithmic platforms as they reshape contemporary work arrangements.

[1] Noopur Raval is a PhD candidate in the Informatics department at the University of California Irvine.

[2] Frey, C. B., & Osborne, M. (2013). *The Future of Employment: How susceptible are jobs to computerisation?* Oxford Martin School. https://www.oxfordmartin.ox.ac.uk/publications/the-future-of-employment

[3] Manyika, J., et al. (2017). *Jobs lost, jobs gained: Workforce transitions in a time of automation*. McKinsey Global Institute.

[4] Christensen, K., et al. (2009). Ageing populations: the challenges ahead. *The Lancet, 374*(9696), 1196-1208.

[5] Betigeri, A. (2018, 18 July). India's demographic timebomb. *The Interpreter*. https://www.lowyinstitute.org/the-interpreter/indias-demographic-timebomb

[6] Hussmanns, R. (2005). *Defining and measuring informal employment*. International Labour Office. https://www.ilo.org/public/english/bureau/stat/download/papers/meas.pdf

[7] International Labour Organization. (2018). *Women and men in the informal economy: A statistical picture*. Third edition. https://www.ilo.org/global/publications/books/WCMS_626831/lang--en/index.htm

The second is that, at least for the near future, especially in the global South but also elsewhere, what we are witnessing is not total and complete automation of/in work but rather what has been called "heteromation"[8] (or a reorganisation in the division of labour between humans and machines). This is also a crucial point as it helps us to "keep humans in the loop" (while building AI) and recentre human work *alongside* machine intelligence. Such a shift also means that we are not necessarily talking about "machines replacing humans" but rather displacing traditional work roles and thus calling for a re-imagination of human work.[9]

This report, then, offers vignettes from the ongoing platformisation and increasing algorithmic management of work in India to give a glimpse of such "heteromated futures" in the global South. With every example, the report also illustrates the socio-technical effects of AI implementation in work, with a focus on prevalent informality and vulnerability as well as social hierarchies of caste, gender and class in India.

Platformisation of blue-collar work

Not surprisingly, a lot of AI-powered productivity technology is marketed to white-collar workers with the promise of "automating banality away", allowing creative and cognitive workers to focus on *real* productivity (versus administrative and repetitive tasks). On the other hand, in blue-collar service and logistics jobs, algorithmic management as well as the use of natural language processing (NLP),[10] facial recognition and biometric attendance have seen a sharp uptake, promising speed, efficiency and standardisation in processes. In India the most prominent examples of algorithmic platforms managing pools of workers in real-time are service intermediaries such as Uber, Olacabs, Swiggy and Zomato (food-delivery platforms), and e-commerce platforms such as Amazon and Walmart-owned Flipkart. Various reports of platform workers in the global South have shown how minutely workers are monitored for work and rest time, how attendance and presence at work is monitored through workers having to take selfies, and how recruitment of new workers happens en masse through artificial conversational agents.

It is worth noting that it is not a coincidence that such technologies are being both deployed on and refined through their use on informal and contract workers, given the overall lack of transparency and monitoring of work conditions among these groups. Further, more work needs to be done to study informal and semi-formal workers' understanding of their rights while their workplaces and processes get increasingly datafied. While in-depth commentary is beyond the scope of this report, the datafication of already vulnerable worker/citizen subjects produces a kind of "double marginalisation",[11] similar to what scholars have been discussing with regard to the datafication of refugees and asylum seekers.

Reintermediation via algorithms at work

As several studies have now shown,[12] app-based ride-hailing drivers are among the largest group of workers currently being algorithmically managed on a granular basis. In India, the intervention and creation of a new "pop-up"[13] labour market due to the arrival of app-based work has resulted in the loosening of traditional local and regional labour markets. Historically, participation in certain occupations (such as cleaning, driving, domestic work, beauty work, etc.) has tightly mapped along the lines of gender, caste, religion and language.[14] Based on the complex hierarchical caste system, only certain caste communities were allowed to teach, trade, farm and so on, while certain caste communities were forced to carry on in a single, stigmatised occupation (such as manual scavenging or sewage cleaning). In pre-algorithmic times, migrating to urban centres to engage in non-agricultural labour required connections, social networks, as well as proofs of belonging (domicile certificates,

[8] Ekbia, H., & Nardi, B. (2014). Heteromation and its (dis)contents: The invisible division of labor between humans and machines. *First Monday, 19*(6). https://journals.uic.edu/ojs/index.php/fm/article/view/5331

[9] For a longer discussion on how machines are not replacing but displacing and reconfiguring human-work, see: https://quote.ucsd.edu/lirani/white-house-nyu-ainow-summit-talk-the-labor-that-makes-ai-magic

[10] NewsVoir. (2018, 18 August). Vahan announces its AI-driven assistant on WhatsApp to automate recruitment. *Deccan Chronicle*. https://www.deccanchronicle.com/business/companies/180818/vahan-announces-its-ai-driven-assistant-on-whatsapp-to-automate-recrui.html

[11] I draw on Dalit feminist writer Bama's articulation of the "double marginalization" of Dalit women under the power of caste and patriarchy here. For a detailed discussion see: Singh, R. (2013). Dalit Women Identity in Bama's Sangati. *The Criterion: An International Journal in English, 4*(V). www.the-criterion.com/V4/n5/Ranjana.pdf

[12] Surie, A., & Koduganti, J. (2016). The Emerging Nature of Work in Platform Economy Companies in Bengaluru, India: The Case of Uber and Ola Cab Drivers. *E-Journal of International and Comparative Labour Studies, 5*(3). ejcls.adapt.it/index.php/ejcls_adapt/article/view/224

[13] "Pop-up" here refers to how algorithmic, big-data platforms like Uber, Ola and others are able to aggregate and show real-time demand and supply within an area, creating lucrative temporary markets in different locations. These markets are not permanent given that – based on the time of the day or special events – some areas have high demands only at certain times (weekends, night, concerts, etc.).

[14] Raval, N., & Pal, J. (2019). Making a "Pro": 'professionalism' after platforms in beauty-work. *Proc.ACM Hum.-Comput. Interact.*, *3*, CSCW, Article 175. https://doi.org/10.1145/3359277

references) to be able to find work in the city.[15] This made it difficult even for migrating workers to cross the social and cultural barriers imposed on employment. Locally, in city-based markets, algorithmic platforms have afforded a way for unskilled and skilled migrant workers to circumvent traditional gatekeeping and at least find temporary work. Mark Graham and colleagues made similar observations about "gig economy"[16] work in the sub-Saharan region, demonstrating how algorithmic platforms are not disintermediating work ("taking humans out"), but rather *reintermediating* work.

AI, temporary labour and social hierarchies

In another study, with app-based female beauty and wellness workers in India who were also migrating from salons and smaller, informal parlour set-ups to doing app-based on-demand beauty work, we discovered how working through app-based platforms allowed women greater flexibility at work.[17] In the AI and work debate, flexibility has been a contested notion.[18] Since the rise of various forms of flexi-work (freelancing through sites like Upwork, micro-tasking through Mechanical Turk and then gig work through Uber, Deliveroo, UberEats, Zomato, etc.), Western scholars have consistently argued that the notion of "flexibility" glosses over the hidden costs[19] and rules of so-called "anytime, anywhere" work. While platform companies claim that dynamic algorithmic matching creates a convenient digital workforce, that is "ready to go, anytime you need work done," it also creates a perpetual "reserve army" of workers who can never log off. Many reports from across the globe have revealed how platform workers end up working long hours in order to make a decent living. In this sense, flexibility at work comes with a heavy price.

Nevertheless, this is not so straightforward when we look at platformised work in the global South. For the women (and men) working through apps, the temporal flexibility to choose or refuse work at certain hours meant that they could attend to social obligations and family needs and build a future of work where formal career pathways are absent.

Often, when asked why workers were choosing to continue working through app-based platforms *despite* seeing and knowing the opacity of algorithmic management, dynamic price determination and the skewed effects of platform rating mechanisms on their ability to work, many workers across platforms reiterated that this was "not permanent work."[20] People were participating for their own reasons – to pay off a loan, to own a vehicle, to set up their own business, to get part-time education and so on. Some also evidenced how working through and with smartphone-based apps was much *cleaner* and more *dignified* than previous work they had done (such as cleaning trucks, manual labour or the same kind of work without tech mediation). Given the long history of technological participation and modernity in various post-colonies of the global South, as well as the role that technologies have been projected to play in developmentalism (from STEM[21] education to cheap laptops to mobile-first internet access), working with technology and technologised work are seen as upward social mobility. In this sense, when we consider the present and near future of AI and labour in India at least, participating in algorithmic work not only appears (to workers) as relatively good temporary work, but often also more dignified work.

AI, surveillance and labour

In the context of AI deployment and the reshaping of work in the global South with a focus on the informality of work, multiple strands emerge regarding surveillance. Continuing with platform work examples, both in service and other e-commerce platforms, the deployment of granular surveillance to track worker movement, worker and customer communications, rest times, and even worker activity while not logged onto company apps is commonplace. Surveillance technologies in the workplace and the resultant metrics have often had dire consequences, such as the reshaping of worker privacy rights or increased performance

[15] Surie, A., & Sharma, L. V. (2019). Climate change, Agrarian distress, and the role of digital labour markets: evidence from Bengaluru, Karnataka. *DECISION*, *46*(2), 127-138; Lalvani, S. (2019, 4 July). Workers' fictive kinship relations in Mumbai app-based food delivery. *CASTAC Blog*. blog.castac.org/2019/07/workers-fictive-kinship-relations-in-mumbai-app-based-food-delivery

[16] Graham, M., Hjorth, I., & Lehdonvirta, V. (2017). Digital labour and development: impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer*, *23*(2), 135-162. https://www.researchgate.net/publication/315321461_Digital_labour_and_development_impacts_of_global_digital_labour_platforms_and_the_gig_economy_on_worker_livelihoods

[17] Raval, N., & Pal, J. (2019). Op. cit.

[18] Kenney, M., & Zysman, J. (2016). The Rise of the Platform Economy. *Issues in Science and Technology, 32*(3). https://issues.org/the-rise-of-the-platform-economy

[19] Lehdonvirta, V. (2018). Flexibility in the gig economy: managing time on three online piecework platforms. *New Technology, Work and Employment,* 33(1), 13-29; De Stefano, V. (2015). The rise of the "just-in-time workforce": On-demand work, crowdwork, and labor protection in the gig-economy. *Comparative Labor Law & Policy Journal*, *37*(3), 471-504.

[20] Rosenblat, A. (2016, 17 November). What Motivates Gig Economy Workers. *Harvard Business Review*. https://www.hbr.org/2016/11/what-motivates-gig-economy-workers

[21] Science, technology, engineering and mathematics.

stress among workers and the concomitant duress on the right to enjoy and be fulfilled by work.

Moreover, surveillance and worker management also have implications with respect to the public "visibility" of a work force when it comes to managing a brand image and the profitability of platform and e-commerce companies. Simply put, the socio-demographics of individual workers can be "better" managed to suit the biases of paying customers. In India, not only have certain castes and tribes, or religious and gendered communities, been confined to specific occupational roles and hence always been viewed and managed suspiciously as "risky subjects", but now, as venture capital-backed platforms seek to pander to the affording middle-class consumers, collecting worker biometric data, tracking their minute-by-minute activity and their location, are seen as desirable for "brand management". So while informal platform work might create a sense of more dignified work for workers, it is worth keeping in mind that the imagined consumers/users of such platforms remain those who can afford to pay for just-in-time services (upper middle-class, upwardly mobile professionals in most cases). In this sense, the social relationships that are found in traditional informal work are not necessarily challenged or reversed in the informal economy produced by platformisation.

This "socio-technicality" is important as it speaks to the unique encounter between informality (a characteristic feature of many developing economies) and algorithmic technologies wherein the "biases" integral to structural arrangements of work are carried over to platform work.

Another strand of AI, work and surveillance in the global South relates to the "hidden ghost work"[22] of data cleaning, image labelling, text processing and content moderation being performed by back-end workers across developing economies. This new phase of back-end work follows the last IT/ITES[23] boom in the early 2000s that became globally visible through the figure of Indian call centre workers. Recently, much has been written about the wage differentials in digital labour and the evidently meagre remuneration being paid to surveilled global South workers doing the "janitorial" labour that keeps digital platforms healthy and productive.[24] Not only this, it has also come to light that many AI assistants are in fact fully powered by real human agents working in developing countries.

Some of this work is done without a clear understanding of the application of the labour, distancing the labourer from the product output. For example, as a part of large, global data-processing chains, many women and men preparing the training data for cutting-edge AI applications under close surveillance may be unwittingly embedded in assembling various policing and surveillance technologies themselves.

Conclusion: Protection for the doubly precarious

As this report has tried to show through different real-world examples, in global South markets such as India where informality is rampant, the combination of financial and socio-cultural precarity as well as the desire to stay *within* the market through technological participation make platform work an attractive option. By offering informality as the dominant metaphor, this report aims to open up space between totalising critiques of AI in the workplace as resulting in bad, exploitative work on the one hand, and AI-embedded futures as automatically empowering and inclusive on the other hand.

Platforms produce complex new realities for work in the global South. While AI-embedded work platforms widen participation for some actors, they have also been known to leverage and reinforce the existing socio-cultural hierarchies that shape certain forms of work themselves. As we see in the examples above, although there is evidence of platformisation empowering informal workers, the algorithmic is likely to re-entrench precarity for informal workers unless there is situated reckoning of the unique historical and economic labour and employment needs of global South geographies, rather than a wholesale embrace of universal (or Western) AI futures. Most importantly, if we are able to hold the evolving data protection, technological innovation and employment conversations with an exclusive focus on informal worker and human-centric issues, only then can we build AI policy that is truly responsive to the needs of workers in the global South.

[22] Gent, E. (2019, 1 September). The 'ghost work' powering tech magic. *BBC*. https://www.bbc.com/worklife/article/20190829-the-ghost-work-powering-tech-magic?ocid=global_worklife

[23] Information technology-enabled services.

[24] Metz, C. (2019, 16 August). A.I. Is Learning From Humans. Many Humans. *The New York Times*. https://www.nytimes.com/2019/08/16/technology/ai-humans.html

Radicalising the AI governance agenda[1]

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 technologies[2] 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 Declaration[8] 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 far[12] 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 Report[14] 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/tdr2018_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_0.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

[19] Ibid.

their territory. The proposed international treaty on business and human rights is a highly pertinent instrument[20] 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

The weaponisation of AI: An existential threat to human rights and dignity

Rasha Abdul Rahim
Amnesty International
www.amnesty.org

Introduction

Over the past decade, there have been extensive advances in artificial intelligence (AI) and other technologies. AI is being incorporated in nearly all aspects of our lives, in sectors as diverse as health care, finance, travel and employment. Another sphere where AI innovation is occurring at a rapid pace is in the military and law enforcement spheres, making possible the development and deployment of fully autonomous weapons systems which, once activated, can select, attack, kill and wound human targets without meaningful human control. These weapons systems are often referred to as Lethal Autonomous Weapons Systems (LAWS) and, more comprehensively, "Autonomous Weapons Systems" (AWS), which encompass both lethal and less-lethal systems.

The rapid development of these weapons systems could not only change the entire nature of warfare, it could also dramatically alter the conduct of law enforcement operations and pose extremely serious human rights risks.[1]

With continuous advances in technology and states such as China, France, Israel, Russia, South Korea, the United States (US) and United Kingdom (UK) heavily investing in and developing weapons with increasing autonomy in the critical functions of selecting and using force on targets, other states are considering how to respond to the automation of warfare and policing. What is clear is that the development and use of AWS raises serious legal, ethical, technological, accountability and security concerns, which is why the Campaign to Stop Killer Robots,[2] of which Amnesty International is a member, is calling for a prohibition on AWS in order to ensure meaningful human control over weapons systems.

Under the auspices of the Convention on Certain Conventional Weapons (CCW), the Campaign has since 2014 been advocating for states to urgently begin negotiations on a legally binding instrument to ensure that meaningful human control is retained over the use of force by prohibiting the development, production, transfer and use of AWS. But while AI weapons technologies race ahead, legal and policy responses to this issue lag woefully behind.

Human rights risks of AWS

AWS can be characterised as weapons capable of selecting and applying force against targets without meaningful control. Autonomy in weapons systems should be understood as a continuum; these systems are not to be confused with unmanned aerial vehicles (UAVs), commonly referred to as drones, which are remotely piloted by a human operator. By contrast, AWS would incorporate software and algorithms which, on their own, would be able to make critical determinations about life and death. Such systems raise important legal, ethical, technological, accountability and security challenges if developed to operate without meaningful control by humans.

The concept of "meaningful human control" was coined by the NGO Article 36,[3] with the aim of setting a normative limit on autonomy in weapons systems by determining the human element required over the use of force.[4] It denotes a level of control which is not purely superficial, for example, a human pressing a button to deploy force against a target that a machine has independently identified.

In situations of armed conflict, the rules of international humanitarian law (IHL) apply alongside human rights law. These require parties to a conflict to distinguish between civilians, who are afforded protection, and combatants, who may be directly attacked. Civilians who are not directly participating in hostilities must never be deliberately targeted. Parties also must distinguish between military

[1] Amnesty International. (2015). *Autonomous Weapons Systems: Five Key Human Rights Issues for Consideration*. https://www.amnesty.org/en/documents/act30/1401/2015/en

[2] https://www.stopkillerrobots.org

[3] www.article36.org

[4] Amnesty International. (2018, 27 August). UN: Decisive action needed to ban killer robots – before it's too late. https://www.amnesty.org/en/latest/news/2018/08/un-decisive-action-needed-to-ban-killer-robots-before-its-too-late

objectives and civilian objects (such as residential buildings, schools and hospitals), and direct attacks only at military objectives. All parties to the conflict must take measures to minimise harm to civilians and civilian objects and must not carry out attacks that fail to distinguish between civilians and combatants, or which cause disproportionate harm to civilians and civilian objects.

Due to the complexity and context-dependent nature of making such assessments in dynamic and cluttered environments, AWS would not be able to comply with IHL, including the requirement to distinguish adequately between combatants and civilians and to evaluate the proportionality of an attack. As former UN Special Rapporteur on extrajudicial, summary or arbitrary executions Christof Heyns argued in his 2013 report to the Human Rights Council, such assessments require intrinsically human qualities and human judgment. They also require:

> [...] common sense, appreciation of the larger picture, understanding of the intentions behind people's actions, and understanding of values and anticipation of the direction in which events are unfolding. Decisions over life and death in armed conflict may require compassion and intuition. Humans – while they are fallible – at least might possess these qualities, whereas robots definitely do not.[5]

Similarly, in law enforcement operations, which are governed by international human rights law (IHRL) alone and elaborated through international policing standards such as the UN Basic Principles on the Use of Force and Firearms (UNBPUFF),[6] the use of lethal and less-lethal AWS without meaningful human control would result in unlawful killings and injuries.

AWS threaten various fundamental human rights, most notably, the right to life which is enshrined in Article 6(1)[7] of the International Covenant on Civil and Political Rights (ICCPR).[8] Under IHRL the use of potentially lethal force is only lawful if it meets the following cumulative requirements: it must have sufficient legal basis in line with international standards; be necessary to protect human life; constitute a last resort; be applied in a manner proportionate to the threat; and law enforcement officers must be held accountable for their use of force.

Under Principle 9 of the UNBPUFF, law enforcement officers may only use lethal force if there is an imminent threat to life or serious injury. This involves a complex assessment of potential or imminent threats, for example, who is posing the threat, identifying and using means other than force, considering whether force is needed to neutralise the threat, deploying different modes of communication to neutralise the threat, deciding on the use of weapons/equipment, etc., and of how best to protect the right to life. These are inherently human skills which cannot be automated, especially given the ever-evolving, dynamic and unpredictable nature of law enforcement operations.

When applying less lethal force, law enforcement officers must apply non-violent means before resorting to use of force, for example, by using techniques including persuasion, negotiation and de-escalation. These techniques require human empathy, negotiation skills, understanding crowd behaviour, and a high level of training and ability to respond to dynamic and unpredictable situations – skills unlikely to be replicated by algorithms.

AWS could also be used to facilitate violations of the right to freedom of peaceful assembly.[9] Indeed, as Heyns has stated:

> [O]n the domestic front, LARs [Lethal Autonomous Robotics] could be used by States to suppress domestic enemies and to terrorize the population at large, suppress demonstrations and fight "wars" against drugs. It has been said that robots do not question their commanders or stage coups d'état.[10]

[5] Heyns, C. (2013). *Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions, Christof Heyns*. www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf; see also General comment No. 36 (2018) on article 6 of the International Covenant on Civil and Political Rights, 30 October 2018, para 65. https://tbinternet.ohchr.org/Treaties/CCPR/Shared%20Documents/1_Global/CCPR_C_GC_36_8785_E.pdf: "For example, the development of autonomous weapon systems lacking in human compassion and judgement raises difficult legal and ethical questions concerning the right to life, including questions relating to legal responsibility for their use. The Committee is therefore of the view that such weapon systems should not be developed and put into operation, either in times of war or in times of peace, unless it has been established that their use conforms with article 6 and other relevant norms of international law."

[6] https://www.ohchr.org/en/professionalinterest/pages/useofforceandfirearms.aspx

[7] Article 6(1), ICCPR: "Every human being has the inherent right to life. This right shall be protected by law. No one shall be arbitrarily deprived of his life."

[8] https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx

[9] Article 21, ICCPR: "The right of peaceful assembly shall be recognized. No restrictions may be placed on the exercise of this right other than those imposed in conformity with the law and which are necessary in a democratic society in the interests of national security or public safety, public order, the protection of public health or morals or the protection of the rights and freedoms of others."

[10] Heyns, C. (2013). Op. cit. The implications of this can be seen in protests in Gaza in 2018, during which Israel deployed semi-autonomous drones to fire tear gas indiscriminately at protesters – it is likely that more autonomous systems will be deployed by law enforcement agencies in the future.

AWS would also undermine the right to privacy (ICCPR article 17) and the right to equality and non-discrimination (ICCPR article 26). Masses of data will need to be collected to train targeting algorithms to profile personal data and create patterns on the basis of which AWS would make decisions on when to use force and against whom. AWS could therefore fuel the bulk collection of data and result in indiscriminate mass surveillance, which is never a proportionate interference with the right to privacy.

The mass collection and profiling of personal data could also have an impact on the right to equality and non-discrimination. Systems employing machine-learning technologies can vastly and rapidly reinforce or change power structures, as the data sets used to teach algorithms contain historical biases which are then reproduced and amplified.[11] For example, in a study by the American Civil Liberties Union, the facial recognition tool called "Rekognition" incorrectly matched 28 members of the US Congress, identifying them as other people who have been arrested for a crime.[12] The false matches were disproportionately of people of colour, including six members of the Congressional Black Caucus. AWS would therefore have the potential to entrench systemic discrimination, with potentially lethal consequences.

Delegating life-and-death decisions to machines

Quite apart from serious concerns as to whether autonomous technologies would be technically capable of conforming to international law, AWS raise numerous important ethical and social concerns – especially since AWS would not be able to refuse orders – about the delegation of human decision-making responsibilities to an autonomous system designed to injure and kill. As Heyns asserts, "[T]here is widespread concern that allowing [autonomous weapons] to kill people may denigrate the value of life itself."[13] Thus the right not just to life, but to a life with dignity, is undermined.[14] Lowering the threshold for the use of force would further depersonalise the use of force, which has already begun through the use of armed drones.

There is also a wider question about the future of our humanity. Is it acceptable to delegate human decision-making responsibilities to use force to a machine? Proponents of AWS argue that removing humans from the equation would increase speed, efficiency, accuracy, stealth and would also cut out emotions – panic, fear, revenge – which can lead to mistakes and unlawful actions. But this is a false dichotomy, as human biases are reflected in algorithms, and therefore neither humans nor machines are infallible. Indeed, human emotions such as empathy can lead to acts of mercy.

Risks to international security

AWS are also vulnerable, as without human oversight they are prone to design failures, errors, hacking, spoofing and manipulation, making them unpredictable. As the complexity of these systems increases, it becomes even more difficult to predict their responses to all possible scenarios, as the number of potential interactions within the system and with its complex external world is simply too large.[15] This would be compounded by autonomous machines interacting with other autonomous machines, posing a risk not only to civilians, but also soldiers and police officers.

The development of AWS would inevitably spark a new high-tech arms race between world superpowers, with each state wanting to keep up with new technologies and seeking to secure them for their arsenals. Given the intangible nature of the software, AWS may also proliferate widely to unscrupulous actors, including non-state actors. In addition, the ease of deploying these weapons may result in an unintended escalation in conflicts.

Therefore, human control and the autonomy of systems should not be viewed as mutually exclusive. The strengths of humans (legal and moral agents, fail-safe) and strengths of machines (data processing, speed, endurance, etc.) should be combined to ensure compliance with the law and predictability, reliability and security.

[11] Lum, K., & Isaac, W. (2016). To predict and serve? *Significance, 13*(5), 14-19. https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x

[12] Eleven of the 28 false matches misidentified people of colour (roughly 39%), including civil rights leader Rep. John Lewis (D-GA) and five other members of the Congressional Black Caucus. Only 20% of current members of Congress are people of colour, which indicates that false-match rates affected members of colour at a significantly higher rate. Snow, J. (2018, 26 July). Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots. *American Civil Liberties Union*. https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28

[13] Heyns, C. (2013). Op. cit.

[14] Article 10, ICCPR: "All persons deprived of their liberty shall be treated with humanity and with respect for the inherent dignity of the human person."

[15] Scharre, P. (2016). *Autonomous Weapons and Operational Risk*. Center for a New American Security. https://www.cnas.org/publications/reports/autonomous-weapons-and-operational-risk

AWS would evade accountability mechanisms

States have legal obligations to prevent and redress human rights violations by their agents, as well as a duty to prevent, investigate, punish and redress the harm caused by human rights abuses by private persons or entities. A failure to investigate an alleged violation of the right to life could in and of itself constitute a breach of this right.[16] In armed conflict, states also have obligations under international humanitarian law to investigate, and where appropriate prosecute, potential war crimes.[17]

Since it is of course not possible to bring machines to justice, who would be responsible for serious violations? Would it be the programmers, commanders, superior officers, political leaders or manufacturers? It would be impossible for any of these actors to reasonably foresee how an AWS will react in any given circumstance, given the countless situations it may face. Furthermore, without meaningful human control, commanders and superior officers would not be in a position to prevent an AWS from carrying out unlawful acts.

This accountability gap would mean victims and families of victims would not be able to access effective remedy. This would mean states' obligation to ensure that victims and families of victims of violations of IHL or IHRL receive full reparation could not be met.

Conclusion

Given the unacceptably high risk that AWS pose to human rights, as well as the ethical, moral and security threats their use would entail, Amnesty International is calling for a legally binding instrument to ensure that meaningful human control is retained over the use of force by prohibiting the development, production, transfer and use of AWS.

Momentum for a ban is steadily growing. Many states, including Austria, Brazil, Mexico, states forming the African Group, and the Non-Aligned Movement, have emphasised the importance of retaining human control over weapons and the use of force. Most states expressed support for developing new international law on AWS, and so far, 29 states[18] have called for them to be banned. These states are largely from the global South, perhaps indicating a well-founded fear that AWS are likely to be used against them.

UN Secretary-General António Guterres also voiced strong support for a ban, describing weapons that can select and attack a target as "morally repugnant".[19] In his Agenda for Disarmament[20] he pledged to support states to elaborate new measures, such as a legally binding instrument. On 12 September 2018 a large majority (82%) in the European Parliament[21] called for an international ban on AWS and for meaningful human control over the critical functions of weapons.

Despite this, a small group of states including Russia, the US, the UK, Australia, Israel, France and Germany are blocking movement towards negotiations for a ban. These are all countries known to be developing AWS.[22] France and Germany have proposed a non-binding political declaration[23] as "a first step" to gather support for the principle of human control over future lethal weapons systems and to ensure they are in full compliance with international law.

[16] General comment No. 36 (2018) on article 6 of the International Covenant on Civil and Political Rights, 30 October 2018, para 27. https://tbinternet.ohchr.org/Treaties/CCPR/Shared%20 Documents/1_Global/CCPR_C_GC_36_8785_E.pdf

[17] International Committee of the Red Cross, Customary International Humanitarian Law, Rule 158.

[18] Campaign to Stop Killer Robots. (2019, 21 August). Country Views on Killer Robots. https://www.stopkillerrobots.org/wp-content/uploads/2019/08/KRC_CountryViews21Aug2019.pdf

[19] Guterres, A. (2018, 25 September). Address to the General Assembly. https://www.un.org/sg/en/content/sg/speeches/2018-09-25/address-73rd-general-assembly

[20] https://www.un.org/disarmament/sg-agenda/en /

[21] https://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2018-0341+0+DOC+XML+V0//EN&language=EN

[22] For example, a recent report revealed that the UK Ministry of Defence and defence contractors are funding dozens of AI programmes for use in conflict, and in November 2018 the UK held exercise "Autonomous Warrior" (https://www.army-technology.com/news/british-autonomous-warrior-experiment), the biggest military robot exercise in British history, testing over 70 prototype unmanned aerial and autonomous ground vehicles. The UK has repeatedly stated that it has no intention of developing or using fully autonomous weapons (https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/673940/doctrine_uk_uas_jdp_0_30_2.pdf). Yet such statements are disingenuous given the UK's overly narrow definition of these technologies ("machines with the ability to understand higher-level intent, being capable of deciding a course of action without depending on human oversight and control"), making it easier for the UK to state that it will not develop such weapons. Although Russia has said it believes the issue of AWS is "extremely premature and speculative", in 2017 Russian arms manufacturer Kalashnikov announced it would be launching a range of "autonomous combat drones" which would be able to identify targets and make decisions without any human involvement; see Gilbert, D. (2017, 13 July). Russian weapons maker Kalashnikov developing killer AI robots. *VICE*. https://news.vice.com/en_us/article/vbzq8y/russian-weapons-maker-kalashnikov-developing-killer-ai-robots

[23] https://www.unog.ch/80256EDD006B8954/(httpAssets)/895931D082ECE219C12582720056F12F/$file/2018_LAWSGeneralExchange_Germany-France.pdf

Encouragingly, momentum has been growing in the private sector. The workforce of tech giants like Amazon, Google and Microsoft have all challenged their employers and voiced ethical concerns about the development of artificial intelligence technologies that can be used for military and policing purposes.[24]

In addition, nearly 250 tech companies, including XPRIZE Foundation, Google DeepMind and Clearpath Robotics, and over 3,200 AI and robotics researchers, engineers and academics have signed a Lethal Autonomous Weapons Pledge[25] committing to neither participate in nor support the development, manufacture, trade or use of autonomous weapons.

This demonstrates widespread support for a legally binding treaty, despite proposals for weaker policy responses. Just as ethical principles have not been effective in holding tech companies to account, non-legally binding principles would fall far short of the robust response needed to effectively address the multiple risks posed by these weapons.

[24] For example, in April 2018 around 3,100 Google staff signed an open letter protesting Google's involvement with Project Maven, a programme which uses machine learning to analyse drone surveillance footage in order to help the US military identify potential targets for drone strikes. Google responded by releasing new AI principles (https://www.blog.google/technology/ai/ai-principles), which included a commitment not to develop AI for use in weapons, and announced it would not renew the Project Maven contract when it expired in 2019. See Wakabayashi, D, & Shane, S. (2018, 1 June). Google Will Not Renew Pentagon Contract That Upset Employees. *The New York Times*. https://www.nytimes.com/2018/06/01/technology/google-pentagon-project-maven.html

[25] https://futureoflife.org/lethal-autonomous-weapons-pledge

Artificial intelligence for sustainable human development

Alex Comninos, Emily Shobana Muller and Grace Mutung'u
Research ICT Africa, Imperial College London, and the Centre for Intellectual Property and Information Technology (Strathmore University)
@alexcomninos @bomu @mathwis_emily on Twitter

Introduction

Artificial intelligence (AI) research is addressing many of the 17 Sustainable Development Goals (SDGs) set by the United Nations General Assembly for the year 2030.[1] The potential of AI to accelerate human development has been acknowledged by international institutions, attracted the focus of civil society and the research community, and resulted in the forging of partnerships between governments, civil society, the technical community and the private sector aimed at implementing AI in pursuit of the SDGs. In the same breath yet at a different pace, AI ethics are being addressed by the private sector and civil society and national policies are being drawn up to mitigate potential injustices and unsustainable externalities. These negative externalities unfold as biased technologies, inequitably deployed technologies, technologies that violate human rights, the marketisation of innovation and the widening of inequality intersect. The potential of technology to both accelerate and hinder human development is not new and the world is rediscovering that the human layers of technological systems are as relevant to the transfer of the technology as the technology itself. This report shows that AI is not only changing the practice of development, but also the structures of power in international development.

This report investigates many of the advances, challenges and changes made by AI for development. It will firstly provide examples of applications of AI for development using the framework of the 17 SDGs, also highlighting the challenges and potential for human harm. It then proceeds to identify two key areas where the impact of AI is shaping development agendas in the global South: the changing power structures in international development in the era of big data, and the widespread introduction of biometric identification (ID).

AI applications for development

AI has a range of applications for achieving the SDGs, or in the new emerging discourse, there are a range of applications of AI for Development (#AI4D) and AI for Good (#AI4Good). Public health is one of the most exciting areas of AI4Good/AI4D. AI can be used to detect malaria outbreaks and track the spread of infectious diseases, monitor preeclampsia[2] in pregnant mothers, catalyse and reduce the costs of drug discovery, assist doctors with decision making and diagnoses, and augment communications with patients through the use of chatbots. AI also has a host of possible applications in agriculture and food security. It can be used to understand diseases affecting crops (for example, cassava, a staple of west African cuisine and a very important plant for food security), to analyse the nutrient composition of soil, and to estimate crop yield. AI could be used in the classroom to assist teachers with their work (grading papers and administration), to create personalised learning assistants that can interact with students and respond to their specialised needs, to translate curricula into different languages, and to augment information and communications technology (ICT) access and usage for those with disabilities. Examples of how AI can help achieve the SDGs are listed in Table 1.

As Table 1 also shows, there are also a wide range of potential societal harms that can be produced by AI.

Responsibility for ensuring that AI4D maximises benefits while minimising risks and harms falls not only on the decision makers, but also the technologists who provide the solutions. It is nevertheless essential that decision makers adopt AI systems in a safe and just manner, guided by legal and regulatory frameworks that protect people from the

[1] For an explanation of the SDGs see https://sustainabledevelopment.un.org/sdgs

[2] Preeclampsia is a pregnancy disorder characterised by high blood pressure which can be treated and managed but can also lead to serious or even fatal complications for mothers and their babies.

TABLE 1

AI and the SDGs: Examples of opportunities and challenges

1. No poverty	*Microfinance:* Machine learning can be used to predict the ability to repay a loan; useful for people with no credit histories.[1] *Digital ID:* Computer vision and biometrics can augment the roll-out of state ID to those without it. Data sources such as imagery and mobile phone records can be used to map poverty for potential interventions.[2]	Digital ID programmes augment opportunities for real-time surveillance by the state as well as by non-state actors. Digital ID is being rolled out in countries in the global South that have weak data protection regimes or no data protection legislation in force. Current research fails to address poverty mapping at small areas which are useful for policy intervention.
2. Zero hunger	*Food security:* Machine learning can be used to better understand plant diseases[3] and AI combined with sensors (e.g. soil sensors) can gather and analyse environmental information in real time.[4]	A lot of this work is being done within the market, providing many job opportunities, but exacerbating inequalities faced by smallholder farm owners who cannot afford technologies.
3. Good health and well-being	AI can be used in health to understand disease outbreaks,[5] monitor conditions like preeclampsia in pregnant mothers,[6] track the spread of infectious diseases,[7] provide decision-making tools for doctors and medical professionals,[8] accelerate drug discovery,[9] and augment communications with patients through the use of chatbots (for example, in the South African MomConnect Maternal Health Platform).[10]	New inequalities could be created through the divide between those with access to AI-driven health and those without. In the global South, AI could replace human staff in an already thinly stretched health care sector facing brain drain. AI is being applied in developing countries often without safeguards for personal and patient information. AI is possibly affecting our mental health in ways that we are only beginning to unpack.
4. Quality education	Personalised learning[11] can be used to augment teaching and curricula for specialised needs and in resource-constrained settings, natural language processing (NLP) can be used to translate curricula into different languages.[12]	AI can introduce new biases into the education system. AI can also be used to surveil students who may give away a lot of personal data to education systems.[13]
5. Gender equality	AI has the potential to make neutral decisions. AI can be used to address gender equality. For example, a design firm released an application which uses machine learning to track gender equality in meetings.[14]	AI has particular problems that can reinforce inequalities including gender bias in algorithms, arising from biased data and a workforce that faces diversity challenges (is disproportionately white and male according to some accounts).
6. Clean water and sanitation	Machine learning and deep learning can be used in water sciences to augment water management.[15]	Smart sanitation projects in countries like India have been shown to perpetuate systemic caste biases.[16]
7. Affordable and clean energy	Machine learning can be used in energy forecasting, helping to smooth the transition to clean and green energy as well as in the implementation of smart grids.	A recent study investigating carbon emissions in the training of some of the cutting-edge national language processing (NLP) models found that the training of one of these models could emit more carbon dioxide than the lifetime emissions of the average American car (including its manufacture).[17]
8. Decent work and economic growth	AI has the potential to catalyse growth in the economy. Through cheaper and more accurate predictions, AI can generate productivity gains.[18]	AI and automation will result in the loss of jobs requiring both skilled and unskilled labour. AI is used in platforms of the gig economy which has threatened the quality of work and sidestepped labour regulations in many countries.
9. Industry, innovation and infrastructure	AI can accelerate innovation and make infrastructure smarter. Online retailers can contribute to a sharing community (such as housing rentals).[19] Low-skilled data cleaning jobs.	AI can result in job losses. The gig economy is radically changing the nature and quality of work while sidestepping regulations. Can contribute to rising inflation of housing costs in areas within a city.[20] Plentiful supply of cheap labour risks turning into a liability.
10. Reduced inequalities	"Data can help bridge inequalities that plague every social, political, and economical sphere."[21] Deep learning was used by researchers at Imperial College London to detect inequalities in four UK cities using official statistics and Google Street View images.[22]	Bias in algorithmic design, and in training data, has been shown to perpetuate prejudice. Algorithms used in calculating social benefits in countries like Poland have been shown to discriminate against groups such as mothers, people with disabilities, and rural citizens, effectively increasing inequality.[23]

SDGs	Opportunities	Challenges and harms
11. Sustainable cities and communities	AI can be used in smart cities and in urban management.[24]	Data privacy regulations are imperative in cities to ensure safe surveillance. Concerns have been raised about the recent use of Zimbabwe state ID data for facial image recognition by Chinese company CloudWalk.[25]
12. Responsible consumption and production	AI can be used in environmental decision support systems to allocate resources more efficiently.[26]	AI is being used by big business to stimulate the unnecessary consumption of goods and services, rather than create a culture of responsible consumption.
13. Climate action	AI can be used in tackling climate change.[27]	The interpretation of big data is open to abuse by climate change denialists. There is also a danger that AI can be used to manage scarce natural resources for the benefit of the powerful few or privileged.
14. Life below water	AI can be used in marine resource management. The Global Fishing Watch Platform (which arose from collaboration between Google, a digital mapping NGO called Skytruth and Oceana) uses neural networks to track fishing activity and activities such as overfishing and human trafficking.[28]	AI can be used by extractive gas and oil industries for underwater drilling at the expense of sea-life ecosystems.
15. Life on land	AI can be hooked up to soil sensors to monitor nutrients in the soil in real time and provide information and decision support to farmers. AI can be used to estimate crop yields.[29]	Factory farms and AI in the agricultural sector and the encroachment of data-driven businesses such as Amazon in the food supply chain are displacing rural farmers and food hawkers, and privatising natural resources such as soil and water.[30]
16. Peace, justice and strong institutions	AI can be used in e-government programmes to create a more responsive government.[31] AI could possibly be used to forecast conflict.[32]	Without the necessary checks and balances, AI can make public institutions more opaque, and undermine rather than enhance participatory governance.[33] AI used in the judicial system has raised concerns regarding due process.[34]
17. Partnerships for the goals	AI is creating synergies between the technical, research and development communities.	AI is transforming power structures in international development, increasing the role of the private sector as a development actor and thus introducing a variety of challenges.

[1] Kostadinov, S. (2019, 30 July). The Future of Lending Money Is Deep Learning. *Towards Data Science*. https://towardsdatascience.com/the-future-of-lending-money-is-deep-learning-61a9e21cf179.

[2] Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science, 350*(6264), 1073-1076. https://science.sciencemag.org/content/350/6264/1073.

[3] Ramcharan, A., et al. (2019). A Mobile-Based Deep Learning Model for Cassava Disease Diagnosis. *Frontiers in Plant Science, 10*; see also https://plantheus.com.ng.

[4] https://mel.cgiar.org/projects/71.

[5] Pindolia, D. K., et al. (2012). Human movement data for malaria control and elimination strategic planning. *Malaria Journal, 11*.

[6] Espinilla, M., Medina, J., García-Fernández, A., Campaña, S., & Londoño, J. (2017). Fuzzy Intelligent System for Patients with Preeclampsia in Wearable Devices." *Mobile Information Systems*. https://doi.org/10.1155/2017/7838464.

[7] Tang, L., Bie, B., Park, S., & Zhi, D. (2019). Social media and outbreaks of emerging infectious diseases: A systematic review of literature. *American Journal of Infection Control, 46*(9).

[8] Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine, 25*, 44-56.

[9] Fleming, N. (2018, 30 May). How Artificial Intelligence Is Changing Drug Discovery. *Nature*. https://www.nature.com/articles/d41586-018-05267-x.

[10] https://www.praekelt.org/momconnect.

[11] Nye, B. D. (2014). Intelligent Tutoring Systems by and for the Developing World: A Review of Trends and Approaches for Educational Technology in a Global Context. *International Journal of Artificial Intelligence in Education, 25*(2).

[12] Abbott, J. Z., & Martinus, L. (2018). Towards Neural Machine Translation for African Languages. *ArXiv*. https://arxiv.org/abs/1811.05467.

[13] Shobana Muller, E., & Marivate, V. (2019). AI in Education. *Proceedings at the Deep Learning Indaba 2019: AI and Fairness II*. https://drive.google.com/file/d/1iyhfxsxvEQ5L3aUjslRB7kF1E5pIn6R1/view.

[14] Ericsson, L. (2016, March). Promoting Gender Equality Using AI. *Doberman IO*. https://doberman.io/thoughts/promoting-gender-equality-using-ai.

[15] Shen, C. (2018). A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resources Research, 54*(11).

[16] See the India country report by ARTICLE 19 in this edition of GISWatch.

[17] Hao, K. (2019, 6 June). Training a Single AI Model Can Emit as Much Carbon as Five Cars in Their Lifetimes. *MIT Technology Review*. https://www.technologyreview.com/s/613630/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes.

[18] OECD. (2019). *Artificial Intelligence in Society*. Paris: OECD Publishing. https://www.oecd-ilibrary.org/science-and-technology/artificial-intelligence-in-society_edfee77-en.

[19] Barron, K., Kung, E., & Proserpio, D. (2018, 29 March). The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb. *SSRN*. https://ssrn.com/abstract=3006832.

[20] Wachsmuth, D., & Weisner, A. (2018). Airbnb and the rent gap: Gentrification through the sharing economy. *Environment and Planning A: Economy and Space, 50*(6), 1147-1170. https://doi.org/10.1177/0308518X18778038.

[21] Birhane, A. (2019, 18 July). The Algorithmic Colonization of Africa. *Real Life*. https://reallifemag.com/the-algorithmic-colonization-of-africa.

[22] Suel, E., Polak, J. W., Bennett, J., & Ezzati, M. (2019). Measuring Social, Environmental and Health Inequalities Using Deep Learning and Street Imagery. *Scientific Reports, 9*. https://www.nature.com/articles/s41598-019-42036-w.

[23] See the Poland country report in this edition of GISWatch.

[24] Woetzel, J., et al. (2018). *Smart Cities: Digital solutions for a more liveable future*. McKinsey Global Institute. https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/smart-cities-digital-solutions-for-a-more-livable-future.

[25] Roussi, A. (2019, 14 May). Chinese investments fuel growth in African science. *Nature*. https://www.nature.com/immersive/d41586-019-01398-x/index.html; Mozer, P. (2019, 14 April). One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority. *The New York Times*. https://nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html.

[26] Cortés, U., et al. (2000). Artificial Intelligence and Environmental Decision Support Systems. *Applied Intelligence, 13*(1), 77-91.

[27] Rolnick, D., et al. (2019). Tackling Climate Change with Machine Learning. *arXiv*. https://arxiv.org/abs/1906.05433v1.

[28] Kourantidou, M. (2019, 11 July). Artificial intelligence makes fishing more sustainable by tracking illegal activity. *The Conversation*. https://theconversation.com/artificial-intelligence-makes-fishing-more-sustainable-by-tracking-illegal-activity-115883.

[29] https://www.stars-project.org/en.

[30] See the thematic report on food sovereignty in this edition of GISWatch.

[31] See the country report for South Africa in this edition of GISWatch.

[32] Card, B., & Baker, I. (2014). A New Forensics: Developing Standard Remote Sensing Methodologies to Detect and Document Mass Atrocities. *Genocide Studies and Prevention, 8*(3), 33-48. https://hhi.harvard.edu/publications/new-forensics-developing-standard-remote-sensing-methodologies-detect-and-document-mass.

[33] See the country report for South Africa in this edition of GISWatch.

[34] See the regional report for Latin America and the country report for Colombia in this edition of GISWatch.

harms of AI. Furthermore, humans who create these systems need to be guided by human rights principles and accord with data protection frameworks. Creating an environment where this is possible will require effective channels of communication between all stakeholders in government, the private sector, the research community, the technical community and civil society.

Data, computational intelligence and development

While development practice has benefited from the new avenues of big data generated by new ICTs as well as from waves of innovation in mathematics and computing, national statistics remain essential for measuring, monitoring and planning for development. Regular, complete, accurate and reliable statistics are essential for equitable development and for the formulation of good economic development policies.[3] The traditional source of information for national development planning has been statistics. These statistics are the result of surveys and censuses often conducted by national statistical offices and international organisations and assisted through international development aid.

Good and informative statistics are collected regularly over long time periods and require human and technical capacity. The production of statistics in the global South (for example, through conducting surveys) can be inhibited by budget, geographic barriers, as well as infrastructural barriers such as transport. The proliferation of mobile phones in the global South has led to a reliance, in both research and policy literature, on supply-side data generated by mobile operators[4] with little demand-side data (e.g. household surveys) to validate this data. More recently, mobile phones and user-generated content/social media platforms have resulted in even more online behaviour and other data being generated, collected and processed by corporations. A promise for many is that big data and AI will augment statistical measurement. However, the ability of supply-side data, which is mostly collected by and controlled by corporations, to represent fairly everyone in the society is questionable and any solution created from this data has the potential to be inequitably deployed.

Changes to power structures in international development

Taylor and Broeders argue that big data is potentially valuable in development research and policy but is also changing power structures within the field and practice of development. One example is a change of where and who donors are going to for data:

> Where previously development donors (governments or international NGOs) worked with LMICs' [low- and middle-income countries] own statistical apparatuses to generate population data, it is becoming increasingly possible and cost-efficient for donors to turn to corporations for consumer-generated data that can proxy for traditional household surveys and other statistics.

The growing "discourse on big data as a resource for development" is another effect of these power shifts and "indicates that a shift is underway from the predominance of state-collected data as a way of defining identities and sorting and categorising individuals, groups and societies to a big-data model where data is primarily collected and processed by corporations and only secondarily accessed by governmental authorities."[5]

Taylor and Broeders investigate two trends arising from this change in power structure. Firstly, they are empowering public-private partnerships in achieving development goals and resulting in "growing the agency of corporations as development actors": "As corporations expand into emerging markets through services which generate digital data, they now find themselves simultaneously expanding into the development field." Secondly, they are creating alternative representations of social phenomena – "data doubles"– that are created in parallel to national data and statistics (for example, when data from mobile phone and internet users are used to create different representations of the same phenomena that are represented in national statistics).[6]

Digital identification

Digital ID in the global South is an example of the application of AI4D. AI technologies like computer vision used in biometrics are being implemented in the roll-out of digital ID in many developing countries. The potential of digital ID has spurred the uptake of official state ID programmes which have been seen as important for development, as well as essential to enabling human rights. Without some form of ID,

[3] Shangodoyin, D. K., & Lasisi, T. A. (2011). The Role of Statistics in National Development with Reference to Botswana and Nigeria Statistical Systems. *Journal of Sustainable Development, 4*(3).

[4] The International Telecommunication Union's data (gathered from mobile and telco operators) is considered a proxy for statistics on mobile phone penetration even though this data often counts inactive SIM cards and does not account for the fact that many users in the developing world own multiple SIM cards as a cost-saving strategy.

[5] Taylor, L., & Broeders, D. (2015). In the Name of Development: Power, Profit and the Datafication of the Global South. *Geoforum, 64*, 229-237.

[6] Ibid.

citizens find it hard to receive government services, to enrol children in schools, and may face obstacles in finding employment or seeking access to loans. Almost 40% of the population over 15 in low-income countries do not have an ID, the poorest are least likely to have an ID, and there are also gender gaps in official ID programmes.[7] SDG target 16.9 calls for provision of "legal identity for all including birth registration" by 2030. The roll-out of digital ID provides an opportunity for reaching this goal.

State-issued identity is essential for social protection. Social protection is defined by the UN Food and Agriculture Organization as "initiatives that provide cash or in-kind transfers to the poor, protect the vulnerable against risks and enhance the social status and rights of the marginalized – all with the overall goal of reducing poverty and economic and social vulnerability,"[8] and includes social assistance ("publicly provided conditional or unconditional cash or in-kind transfers, or public works programmes"), social insurance (like welfare, unemployment benefits, household subsidies and child grants), and labour protection. A third of the population of the developing world (about 2.1 billion people) receive some form of social protection.[9]

However, digital ID programmes present several nuanced challenges.

India's digital ID project, the Unique Identification Authority of India (UIDAI, or Aadhaar), which has been underway since 2008, aims to provide every Indian resident with an ID number linked to their demographic and biometric data. Aadhaar is linked to welfare delivery and widely incorporated into India's social protection schemes. According to Masiero, Aadhaar's "nature as a requirement for social protection has led to concerns of social justice, due to the burden of authentication failure and importantly, to the confinement of social protection entitlements to those residents who 'chose' to enrol."[10] Controversially, an Aadhaar ID is only available to Indian residents who are registered in the National Register of Citizens (NRC) and not to refugees or stateless persons. There is a possibility that over 1.9 million people in the north Indian state of Assam could be excluded from the NRC, and thus from citizenship, Aadhaar ID, state services and social welfare.[11] The criteria for inclusion into the NRC are that an individual, or one of their ancestors, must have had their name on an electoral roll from before 1971.[12] Aaadhaar demonstrates that new ways of counting citizenry will intersect with structures of exclusion, possibly creating new layers of exclusion or amplifying existing ones.

Kenya's digital ID programme, Huduma Namba – which means "service number" in Swahili – is the popular name for the National Integrated Identity Management System (NIIMS). A bill proposed in August 2019 envisages the Huduma Namba system as having three components: a centralised database (or central electronic register), a unique identifier for each person, and a card to be carried for mandatory use in accessing services.[13] Unique identifiers include fingerprints, hand geometry, earlobe geometry, retina and iris patterns and voice waves.[14] DNA and GPS coordinates were originally meant to be collected too, although the collection of this data, together with the mandatory use of the card to access government services, have been temporarily halted by the High Court.[15]

Huduma Namba also has the potential to intersect with existing topologies of societal inclusion and exclusion, ethnic cleavages, and statelessness in Kenya. When governments link digital ID to determination of citizenship, it puts at risk populations who for historical reasons lack primary identification documents. For example, in Kenya, the Nubian, Shona and Makonde communities, which have historically lived in areas that became borders during colonialism, are subjected to long vetting processes before they can acquire identity documents.[16] Huduma Namba also captures ethnicity information which in light of Kenya's recent history concerns many Kenyans. While it does not feed into census data and national statistics, its roll-out was integrated into the recent Kenyan census, also in August 2019.

[7] World Bank Group. (2018). *Identification for Development: 2018 Annual Report*. https://id4d.worldbank.org/sites/id4d.worldbank.org/files/2018_ID4D_Annual_Report.pdf

[8] Food and Agriculture Organisation of the United Nations. (2015). *The State of Food and Agriculture. Social Protection and Agriculture: Breaking the Cycle of Rural Poverty*. www.fao.org/3/i4910e/I4910E.pdf

[9] Ibid.

[10] Masiero, S. (2019, 12 September). A new layer of exclusion? Assam, Aadhaar and the NRC. *South Asia @ LSE*. https://blogs.lse.ac.uk/southasia/2019/09/12/a-new-layer-of-exclusion-assam-aadhaar-and-the-nrc

[11] Ibid.

[12] Ibid.

[13] https://www.hudumanamba.go.ke/wp-content/uploads/2019/07/12-07-2019-The-Huduma-Bill-2019-2-1.pdf

[14] http://kenyalaw.org/kl/fileadmin/pdfdownloads/AmendmentActs/2018/StatuteLawMischellaneousNo18of2018.pdf

[15] Kakah, M. (2019, 2 April). High Court allows Huduma Namba listing but with conditions. *Daily Nation*. https://www.nation.co.ke/news/Huduma-Namba-Govt-barred-from-forced-listing/1056-5051788-t78f1xz/index.html

[16] Kenya National Commission on Human Rights. (2007). *An Identity Crisis? A Study on the Issuance of National Identity Cards In Kenya*. https://www.knchr.org/Portals/0/EcosocReports/KNCHR%20Final%20IDs%20Report.pdf

Digital ID programmes in the global South also often involve international companies. India's Aadhaar is "the largest-scale public-private partnership currently underway in a developing country in terms of its coverage of the population."[17] Officially, Aadhaar is "a quasi-governmental organisation, attached to the national planning authority," but is effectively autonomous with little legal framework governing it, and with little parliamentary oversight. "Besides being non-governmental, UIDAI is also a collaboration between and amongst corporations," is managed and chaired by a co-founder of one of India's largest technology consulting companies, and "the day-to-day work of gathering, processing and storing data is done by private companies."[18]

As countries roll out biometric digital ID programmes, they produce massive amounts of data that may be useful to companies. However, the programmes are being rolled out in the developing world often without any regard for data protection,[19] in many countries with no data protection frameworks in place, and in other countries with weak new data protection regimes. For example, a bilateral agreement between Zimbabwe and China has resulted in the government of Zimbabwe sharing its national ID database with a Chinese company for use in training facial recognition software.[20]

Conclusion

Many of the same challenges that faced ICT for development will face AI4D. Connectivity, literacy, and the price of data remain constraints on the uptake of ICTs and on internet usage and will also be constraints on any benefits from AI.

Regulatory challenges are also important and on the horizon for AI4D. Biased data and the increasing amount of automated decisions that may work themselves into development processes and applications point towards important ethical issues, as well as the need for regulation. In the years and decades to come, a strong regulatory framework must be adopted by all nations to ensure readiness for the widespread use of AI for development.

Data protection regulation and capacity building for data protection enforcement will be very important, especially as so much of the AI in the developing world will be processing data that is possibly private and personally identifying information. The roll-out of digital ID could stimulate the uptake of state identification and ensure that everyone has an ID and access to state services and is equal before the law, and thus act as a stimulus for development. But it is important that digital ID and social protection is rolled out with consideration for privacy and data protection and where possible, after privacy and data protection frameworks are formalised. Data protection need not be a luxury that is considered second to development imperatives, and there need not be any tension perceived between privacy and data protection.

Strengthening the capacity for beneficial AI in the developing world will include strengthening communication between the institutions that house researchers who create AI solutions, and the political bodies able to leverage resources for deployment, with all parties working under a human rights framework. At the same time, as is always the case with development, care is needed to make sure that those most affected are included in the dialogue from the beginning and solutions are not simply being transferred from one nation to the other under the banner of "technological evangelicalism". Efforts are being made in this respect with the creation of the AI4D network[21] in Nairobi this year, and UNESCO's Forum on AI in Africa.[22]

Simultaneously, at the national level, capacity building is essential to get those most passionate about human development working in the space of AI for human development. Huge strides have been taken by the Deep Learning Indaba organisation.[23] This year the Indaba conference hosted 700 African technologists, researchers, innovators and learners in Nairobi with an additional 27 satellite events across the continent. By ensuring policy, people and capacity are in place, the opportunities available will no doubt transform human development and pave the way to a better future.

[17] Taylor, L., & Broeders, D. (2015). Op. cit.

[18] Ibid.

[19] Sepúlveda, M. (2019, 16 April). Data Protection Is Social Protection. *Project Syndicate*. https://www.project-syndicate.org/commentary/social-protection-biometric-data-privacy-by-magdalena-sep-lveda-2019-04.

[20] Chutel, L. (2018, 25 May). China is exporting facial recognition software to Africa, expanding its vast database. *Quartz Africa*. https://qz.com/africa/1287675/china-is-exporting-facial-recognition-to-africa-ensuring-ai-dominance-through-diversity

[21] AI4D. (2019, 8 May). A roadmap for artificial intelligence for development in Africa. https://ai4d.ai/blog-africa-roadmap

[22] https://en.unesco.org/sites/default/files/ai_outcome-statement_africa-forum_en.pdf

[23] The Deep Learning Indaba is the annual gathering for the African AI community, with a mission to strengthen machine learning in Africa. www.deeplearningindaba.com

Country and regional reports

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 Bikobo 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 Jedrzej 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.

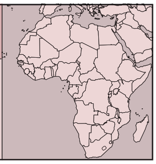
Karisma Foundation offers a useful analysis of media coverage of Prometea in Colombia, showing that most reporting offered little understanding of the system: "[T]here was no explanation about what Prometea was, what it does and how it does it." When, as in the Ukraine, there appears to be reasonable 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."

AFRICA



AI IN AFRICA: REGIONAL DATA PROTECTION AND PRIVACY POLICY HARMONISATION

Raymond Onuoha
Regional Academic Network on IT Policy (RANITP),
Research ICT Africa
https://researchictafrica.net/ranitp

Introduction

Basking in the ubiquitous adoption of mobile technology in Africa, experts in the technology domain prognose a similar upswing in the application of artificial intelligence (AI), especiall y in the communications space, and expect it to help leapfrog critical challenges on the continent.[1] With predictions of significant advancements relying on AI over the next 20 years,[2] there seem yet very sparse collective attempts by regional governments in Africa and the continent as a whole to deal with critical emerging issues. This is especially the case with regard to data protection and privacy, such as government surveillance or corporate influence over customers. Though the challenge of specific AI-related cyberpolicy formulation on the continent may appear unrealistic at this early stage, it is imperative to initiate critical discussions on the context-specific requirements with regard to adapting existing or formulating new regulatory policy as it pertains to AI.

State of play: Regional data protection and privacy policy frameworks in Africa

The adoption and effective implementation of existing data protection and privacy policy frameworks by countries across Africa – even with the limitations on the continent with respect to AI developments – is still a fundamental reference point for ensuring that critical safeguards are in place while we seek to maximise the benefits of AI. The closest continent-wide policy document in this regard – the African Union's (AU) 2014 Malabo Convention on Cyber Security and Personal Data Protection[3] – has been signed by just 11 out of the 55 member countries (these are Benin, Chad, Comoros, Congo, Ghana, Guinea-Bissau, Mozambique, Mauritania, Sierra Leone, Sao Tome and Principe and Zambia), while only three member countries (Guinea, Mauritius and Senegal) have ratified the policy document. While the convention provides "fundamental principles and guidelines to ensure an effective protection of personal data and [seeks to] create a safe digital environment for citizens, security and privacy of individuals' data online," it makes no reference to institutional strategies of mitigating the threats posed specifically by AI deployments on the continent.

At the regional level, the focus has largely been on the policy element of data privacy, with the Economic Community of West African States (ECOWAS) leading the way via the 2010 Supplementary Act on Personal Data Protection within ECOWAS.[4] Similar, albeit non-binding policy instruments have also been developed by the East African Community (EAC) – the 2012 Bill of Rights for the EAC[5] and the 2011 draft EAC Legal Framework for Cyber Laws.[6] In the same regard, the Southern African Development Community (SADC) established the Model Law on Data Protection in 2012,[7] but it is non-binding on member states, making implementation and enforcement difficult.[8] However, disharmony at the regional level with respect to policy formulation generally undermines levels of compliance. This situation demands more continent-level coherence for easier adoption and implementation. If it persists

[1] Bostrom, N., Dafoe, A., & Flynn, C. (2018). *Public Policy and Superintelligent AI: A Vector Field Approach*. Oxford, UK: Governance of AI Program, Future of Humanity Institute, University of Oxford. https://pdfs.semanticscholar.org/9601/74bf6c840bc036ca7c621e9cda20634a51ff.pdf; Dafoe, A. (2018). *AI Governance: A Research Agenda*. Oxford, UK: Governance of AI Program, Future of Humanity Institute, University of Oxford. https://www.fhi.ox.ac.uk/wp-content/uploads/GovAIAgenda.pdf; Gadzala, A. (2018). *Coming to Life: Artificial Intelligence in Africa*. Washington: Atlantic Council. https://www.atlanticcouncil.org/images/publications/Coming-to-Life-Artificial-Intelligence-in-Africa.pdf

[2] Turianskyi, Y. (2018). *Balancing Cyber Security and Internet Freedom in Africa*. South African Institute of International Affairs. https://www.africaportal.org/publications/balancing-cyber-security-and-internet-freedom-africa

[3] https://au.int/sites/default/files/treaties/29560-treaty-0048_-_african_union_convention_on_cyber_security_and_personal_data_protection_e.pdf

[4] www.statewatch.org/news/2013/mar/ecowas-dp-act.pdf

[5] www.eala.org/documents/view/the-eac-human-and-peoples-rights-bill2011

[6] repository.eac.int:8080/bitstream/handle/11671/1815/EAC%20Framework%20for%20Cyberlaws.pdf?seq

[7] www.itu.int/ITU-D/projects/ITU_EC_ACP/hipssa/docs/SA4docs/data%20protection.pdf

[8] Turianskyi, Y. (2018). Op. cit.

as it is, the disharmony inadvertently increases the gap between the frontiers of global technology and mechanisms of local and regional governance that has geopolitical ramifications for the continent.[9]

Institutional challenges for regional/continental data protection policy harmonisation in Africa

The recent and rapid diffusion of the internet across Africa, with the attendant emergence of AI deployments on the continent, is growing ahead of institutional, social and cultural changes. In this purview and in specific relation to data protection and privacy policy institutionalisation, currently only 17 out the 55 member countries of the AU have enacted comprehensive data protection and privacy legislation (these are Angola, Benin, Burkina Faso, Cape Verde, Gabon, Ghana, Ivory Coast, Lesotho, Madagascar, Mali, Mauritius, Morocco, Senegal, Seychelles, South Africa, Tunisia and Western Sahara).[10] This slow pace of data governance policy evolution among AU member countries has been identified as a major hindrance to a harmonised policy framework for data protection and privacy. Of note in this regard, the number of countries in Africa that have enacted comprehensive data protection and privacy legislation is even more than the number that have adopted the continental-level Malabo Convention with respect to data protection and privacy. Most of the national data protection laws also existed before the Malabo Convention and seem to have more specific details with regard to data protection than the AU convention and addressing the issues that have subsequently emerged. This is one concern with the Malabo Convention that has been raised by AU member countries.

A challenge with respect to the continental-level data policy process is that it is a very slow and painstaking process. As a result, although less than half the countries on the continent have implemented policies on data, many have been forced to move ahead without necessarily looking to the region for guidance. In addition, with the largely top-down approach of data policy engagement by the AU and the Regional Economic Communities (RECs), with people just making laws on behalf of countries, regional instruments are bound to run into significant adoption challenges. In this light, research indicates that a top-down regional policy engagement process might only be designed to "serve narrow regime interests at the expense of broader national and collective interests."[11]

Another challenge impacting the adoption of the AU Malabo Convention is the lack of sector or industry-specific considerations with regard to data protection and privacy guidelines akin to the European model laws. This creates unhelpful levels of uncertainty and unpredictability, especially for multinational organisations seeking compliance within national boundaries.[12]

A critical hindrance to data protection and privacy policy cooperation on the continent is the significant variation in cultural and legal diversity, access to technology, and governance capacity for data-related policy making.[13] Compounding this problem is the existing legacy allegiance of the regional blocs within the AU to their former colonial countries in such a manner that sharply divides the policy interests of the Anglophone and Francophone countries, thereby weakening the cohesiveness of the continental body in general with respect to data policy making.

This situation makes Africa's relationship with data governance unclear – a lack of clarity that is compounded by capacity constraints. The policy-making institution in Africa is largely led by a traditionally analogue generation that predates the internet age, making the understanding of data-led digital policy engagements challenging. There is therefore a lack of capacity and understanding of who should take responsibility in the region with regard to data-driven technology and its imperatives with respect to digital rights. This general lack of understanding leads to a lack of policy direction with respect to emerging issues such as AI.

The existing lack of capacity and technical expertise at the policy-making echelon for data governance in Africa poses a significant implementation and process management cost to a harmonised regional policy framework. Further training and

[9] Evanoff, K., & Roberts, M. (2017, 7 September). A Sputnik moment for artificial intelligence geopolitics. *Council on Foreign Relations*. https://www.cfr.org/blog/sputnik-moment-artificial-intelligence-geopolitics

[10] Mabika, V. (2018, 8 May). The Internet Society and African Union Commission Launch Personal Data Protections Guidelines for Africa. *Internet Society*. https://www.internetsociety.org/blog/2018/05/the-internet-society-and-african-union-commission-launch-personal-data-protections-guidelines-for-africa

[11] Söderbaum, F., Skansholm, H., & Brolin, T. (2016). *From top-down to flexible cooperation: Rethinking regional support to Africa*. The Nordic Africa Institute. cris.unu.edu/sites/cris.unu.edu/files/From%20Top%20Down%20to%20Flexible%20Cooperation%20-%20May%202016.pdf

[12] Ridwan, O. (2019, 20 March). The Africa Continental Free Trade Agreement and Cross-Border Data Transfer: Maximising the Trade Deal in the Age of Digital Economy. *African Academic Network on Internet Policy*. https://aanoip.org/the-africa-continental-free-trade-agreement-and-cross-border-data-transfer-maximising-the-trade-deal-in-the-age-of-digital-economy

[13] Mabika, V. (2018, 8 May). Op. cit.

assistance for policy makers may be required; more so as a large number of AU member countries are yet to establish independent data privacy regulatory authorities. Bridging this capacity gap among policy makers within the AU region is imperative, as an unclear understanding of emerging technological developments with respect to data policy might produce the unintended consequences of limiting the region's competitiveness in the AI economy. This is of importance when it comes to issues such as data availability for multinational organisations operating on the continent that collect, process and share data for AI-based applications and services, especially those that are mobile phone based. For example, a forced data localisation regime on the pretext of maintaining national security and sovereignty might restrict cross-border data transfers for such multinational data companies who may choose to move their foreign direct investment to more favourable destinations.

In the final analysis, with regard to priorities, many 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. According to one regional policy expert interviewed for this report, "A government that is still battling [to set up a] school feeding programme in 2019 is not going to be the one to prioritise data and data protection policies with respect to AI." 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.[14] These costs would be in the form of funds necessary for setting up data protection authorities at the governmental levels as well as designated data privacy representatives for private sector players. Furthermore, in as much as a continent-wide data protection and privacy policy framework for Africa will catalyse regional collaboration and cooperation in dealing with the emerging issues and risks posed by AI deployments on the continent, it may however impose costs and raise conflicts with other national data protection and privacy regimes if it is not well harmonised globally. So the Malabo Convention is indeed a good place to start, but then again the AU will need to do a lot more work in promoting its benefits not just regionally, but within an emerging global policy context.

Conclusion

The growing shift towards a more centralised data protection and privacy policy framework in Africa considering the significant cross-border imperatives of AI deployments, as well as cross-sectoral technological developments, comes with critical challenges. Regulators on the continent need to become more innovative and seek to understand emerging AI technologies in order to effectively regulate them. They need to consider AI-related principles that would apply contextually irrespective of the technologies or systems that are deployed. Not all policy needs with respect to AI are complex – some are pretty straightforward to implement. A good example here is driving the adoption of AI-related data policy in the continent by matching AI technology with the socioeconomic needs that are addressable in peculiar contexts within the region. It is nevertheless necessary to have stakeholders with a shared understanding of the policy needs and on the same page with regard to a clear direction of where the continent wants to go, and how its respective countries can benefit from this path. People need to become more aware with respect to the critical issues of transparency and openness. They need to know that there are built-in safeguards to protect their personal data that are collected from misuse, especially in line with the principle of making sure that further processing of their personal data is compatible with the reasons or basis for which they were collected in the first place. These remain fundamental in building trust within the technology ecosystem. Furthermore, and in consideration of the longer term, a coherent regional data policy framework for the region should be technologically neutral with consistencies across multi-industry sectors and services. However, risk assessment models should be built into the regional frameworks in such a way as to reflect accepted privacy principles.

In adapting current regional data protection frameworks in Africa to deal effectively with the emerging challenges of AI, there are many lessons that Africa needs to learn from other regions that have moved forward earlier with policies and practices relevant to data protection and related cyberpolicy.

While the European Union's General Data Protection Regulation (GDPR) is a model for regional data protection policy collaboration, it can be improved on and not just taken as a silver bullet solution for the continent. Nevertheless, many of its requirements are worth adopting. For example, considering the cross-border imperatives of AI systems, regional

[14] Curtiss, T. (2016). Privacy Harmonization and the Developing World: The Impact of the EU's General Data Protection Regulation on Developing Economies. *Washington Journal of Law, Technology & Arts, 12*(1). digital.law.washington.edu/dspace-law/bitstream/handle/1773.1/1654/12WJLTA095.pdf?sequence=4&isAllowed=y

data policy instruments should be framed in such a way that data-handling firms operating in Africa must be made to sign up to the data protection and privacy laws within their operational jurisdictions, whether or not they are registered as a business entity in those jurisdictions. This is of significant importance for Africa considering the fact that critical data-related projects across the continent are handled and processed outside her borders. Some key examples in this regard include the Kenya Digital ID project,[15] which is hosted and processed by a foreign company, and the data collected by Ghana's Electoral Commission, which is not hosted in-country. Moreover, none of the big data firms – Facebook, Google, Amazon and Microsoft – are registered as business entities in any African country.

Action steps

The following action steps are suggested for civil society:

- *Capacity building for effective policy making*: Africa has been saddled with the burden of leaders who are behind technology advancements. Keeping pace with evolving technologies will require policy evolution and adaptations. Civil society can help in bridging these capacity deficits in such a manner that cross-country peculiarities and spillovers are taken into consideration. They can engage in the build-out of AI knowledge centres[16] across the region that will help bridge these critical gaps by encouraging a thorough understanding of the issues involved, and serve as a resource to help understand the policy directions of the various RECs with respect to AI.
- *Pushing for AI-related principles and values in data protection policy:* Data protection laws and frameworks are built on general principles, like most technology laws, which are developed to regulate appropriate behaviour regardless of technology evolution with time. However, AI involves a number of specific issues that need to be addressed. Civil society can advocate for appropriate contextual principles and values around which the regional entities can coordinate on data protection policy relevant to AI. Critical among these principles for Africa is the right to privacy of an individual, which is fundamental for our existence as human beings. Furthermore, people need to become more aware with respect to transparency and openness. Another principal area of concern with regard to AI policy is the issue of bias, as AI is currently being developed in primarily two regions of the world: the West and China/Russia. In each of these regions, there is a paucity of data being fed into AI machines that correlates with the African experience. Furthermore, AI data policy for the continent must not be a one-sided issue; it has to be gender-centric and also take into consideration marginalised groups as well as the diversity of different languages and cultures within the region in order to achieve a broad-based result that engenders equitable technology access.
- *Socioeconomic needs assessment*: Civil society can advocate for the adoption of relevant AI-related data policy by helping to match it to the socioeconomic needs in particular contexts in the region. A needs analysis of countries must be done with respect to AI technology so policy can be linked to economic solutions. AI is useless to African countries if it is not applied in a way that solves their needs.
- *Multistakeholder policy advocacy:* Civil society can contribute to a multistakeholder process that also includes governments, citizens, universities and the private sector to help collaboratively adapt current regulatory frameworks in such a manner that they promote digital innovation while protecting the privacy and security of citizens.

[15] Dahir, A. L. (2019, 21 February). Kenya's plan to store its citizens' DNA is facing massive resistance. *Quartz*. https://qz.com/africa/1555938/kenya-biometric-data-id-not-with-mastercard-but-faces-opposition

[16] Such as the Global Cyber Security Capacity Centre pioneered by the University of Oxford. https://www.oxfordmartin.ox.ac.uk/cyber-security

ARGENTINA

UNION PERCEPTION OF ARTIFICIAL INTELLIGENCE IN ARGENTINA

Nodo TAU
María Florencia Roveri
www.tau.org.ar

Introduction

How will work be reconfigured with the application of artificial intelligence (AI) in the workplace? This is a question that several actors are now asking. Governments, academics, civil society, educationalists and the private sector are all trying to analyse and predict the possible changes in work relations.

Unions are at the forefront of the fight for better labour conditions, and while their battle is mostly to do with the economy, the technological environment has an impact too. In the so-called "Fourth Industrial Revolution", the automation of processes and the use of robotics and algorithms are changing work processes and also impacting on employment statistics.

How are unions addressing this reality in Argentina? Do they consider the impact of AI on labour conditions and rights? Is this issue discussed by business owners? And are unions evaluating the potential positive impact of AI on the welfare of workers?

Context

According to the International Labour Organization, in 2018 Argentina had the third highest unemployment rate in the region, with an indicator of 9.5%, only surpassed by Guyana (12.2%) and Brazil (12.5%). A recent official survey of labour and employment showed a negative movement of 2.8% in the last year, with companies forecasting a worsening of the situation.[1] In this context, the government has implemented institutional reform and restructured ministries. Work and Science are no longer ministries, and the Ministries of Education and Health have been restructured substantially, evidence of the new policy direction of the government.

In late 2018 the government launched the Digital Agenda 2030,[2] coordinated by the Secretariat of Science and Technology, which includes a National Plan for Artificial Intelligence.[3] The agenda deals with issues such as the digital economy, education and infrastructure, which have been discussed in meetings to define frameworks for private companies working in these areas. The objective, according to the government, is the creation of a "country brand" in order for Argentina to become a leader in the region, a role that no country has clearly assumed at the moment. However, the government does not mention developing common strategies with other governments in the region.

Meanwhile, the private sector is looking at AI as strategic for its development. The Centre for the Implementation of Public Policies for Equity and Growth (CIPPEC) was commissioned by Microsoft to analyse the impact of AI on economic growth in Latin America.[4] Its report considered three scenarios with different levels of application of AI, and forecast that Argentina could reach the status of a "developed" country in 20 years if AI is applied extensively. It also warned of economic stagnation if this did not happen.

The report listed "three key qualities" in jobs for the future: "perception and manipulation of complex contexts, creativity and social intelligence."[5] Workers in Argentina whose occupations are intensive in these qualities are 1.9 million out of a total of 11.9 million workers – or 16% of employed workers. The other 84% will require human capital reskilling. "Occupations that have these qualities are related to education, health, psychology and coordination of people," the report states, "and the kinds of workers who will need reskilling are data entry clerks, those doing telesales and machine operators."

From this perspective, the problem is the scarcity of qualified labour. This analysis takes "technological development as an innate process, necessary

[1] Ministerio de Producción y Trabajo. (2019). *Encuesta de indicadores laborales*. www.trabajo.gob.ar/downloads/estadisticas/eil/eil_1905_gacetilla.pdf

[2] Casa Rosada. (2018, 6 November). El gobierno presentó nueva agenda digital. https://www.casarosada.gob.ar/informacion/actividad-oficial/9-noticias/44081-el-gobierno-presento-la-nueva-agenda-digital-2030

[3] Catalano, A. (2019, 22 March). El Gobierno impulsa un plan nacional para liderar el desarrollo de inteligencia artificial en la región. *iProfesional*. https://www.iprofesional.com/tecnologia/288670-banda-ancha-ciberseguridad-computadora-Argentina-busca-liderar-la-inteligencia-artificial-en-la-region

[4] Albrieu, R., Rapetti, M., Brest López, C., Larroulet, P., & Sorrentino, A. (2018). *Inteligencia artificial y crecimiento económico. Oportunidades y desafíos para Argentina*. Buenos Aires. CIPPEC. https://www.cippec.org/wp-content/uploads/2018/11/ADE-ARG-vf.pdf

[5] Ibid.

for economic growth and neutral with respect to the policies that promote it."[6] But there is another point of view which states that "focusing the analysis of AI on its effects, such as job losses and the need to adapt, conceives of technological development as a fatal and unchangeable good." This counter-analysis underlines the "urgency of assessing AI in terms of the socialising nature of work and its moral connotations, not restricting it to quantitative factors to do with growth and not accepting *prima facie* technologies that reduce production costs."[7]

Do unions dream or have nightmares about robots?

Several kinds of work will be replaced by machines in the near future. Arguing that this forecast should be good news for our countries and economies, what changes can we expect in the workplace? And, more precisely, what happens to workers?

Although in Argentina several actors are leading this debate, the voice of unions is not being heard. This report is based on interviews with respondents from the unions representing workers in commerce, public services at the municipal level, media workers and journalists, and the banking sector.[8]

The deputy secretary of the Union of Public Service Workers defines AI as "self-managing tools that require little intervention by people." In 1995, when public information was digitised, the municipal government started to incorporate computers into its operations. "Then workers did not receive adequate training. Nobody should be excluded, so the union worked in that gap for reskilling."

Procedures that used to take time, were manual and involved two or three people, are nowadays solved online. "However, the number of workers was not reduced. Some jobs no longer exist, but there are new ones related to digitisation, for example, in IT or informatics divisions. This change [introducing AI in public services] allows us to dedicate more time to creative work and meeting the needs of citizens. That is what we are pointing out."

This union respondent mentioned "privacy as an issue to take into account in the digitisation of human resources, the control of front-desk assistants, institutional relationships and even ideological differences [that can be monitored on personal digital devices]." They have no complaints at the moment about the potential introduction of AI in public services; however, they do consider it a sensitive field.

The Bank Workers Union is the organisation that fights for the rights of workers in the face of financial power. "Innovation [in the financial sector] is one of its key characteristics, encouraged by public policies that allow for the unfettered adoption of technology." The union says AI will simply result in a reduction of employees. "Our union is not opposed to technology. We can improve the way we work, but we cannot lose our jobs."

ATMs are a good example in this field. The CIPPEC report claims that although ATMs automatised a wide range of tasks, "the number of employees tended to increase due to the reduction of costs in opening branches and freeing employees from transactional tasks, dedicating them to more productive ones." The union representative we interviewed contradicts this statement. "The positions that ATMs replaced, together with those of retiring employees, are not being filled again." She said other innovations in the sector that have resulted in job losses include the automation of administrative checks and controls, cryptocurrency, the use of QR codes, and home banking. "Even tasks related to the processing of information are automatised."

New trends that also affect labour conditions are banks offering 24/7 assistance, and telecommuting. "Some banks implement a work-from-home policy two days a week, providing a computer, an ergonomic chair and a fire extinguisher. We warn about the level of control they will experience, that will result in increased pressure on workers and illness. If work moves out of work, it will be very difficult to take care of working conditions," she said.

"We saw this unhealthy control over productivity in call centres." In that case, she explained:

> We managed to push for regulations, and developed a code of good practice[9] that recommends times for resting, lunch, bathroom breaks. A "callcenterisation" of work may occur due to efforts to optimise time. We think human beings are not prepared for these virtual relations. Perhaps new generations can afford it and even prefer it. But now we are experiencing the transition and we are not prepared for this.

[6] Puyana, A. (2019, 26 March). Inteligencia artificial y trabajo en América Latina. Nuevas pistas de la economía mundial. *América Latina en Movimiento*. https://www.alainet.org/es/articulo/198957

[7] Ibid.

[8] The unions interviewed were the Asociación de Empleados de Comercio de Rosario, Sindicato de Trabajadores Municipales de Rosario, Asociación de Personal del Sistema Legislativo, Sindicato Bancario Rosario and Sindicato de Prensa de Rosario.

[9] Ministerio de Trabajo y Seguridad Social, Provincia de Santa Fe. (2011). Código de Buenas Prácticas para Call Centers. www.santafe.gov.ar/index.php/web/content/download/134922/664629/file/Resolucion318.pdf

According to the general secretary of the Union of Press Workers:

> Unions are behind in the debate on AI. In Argentina unions are disputing basic issues such as salary, health, loss of employment, with no economic stability 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.

He was also critical of media companies. "They try to send one person with a backpack to do the work of a journalist, a photographer and a camera operator. Three people are needed and they argue that half a person is enough." Not every technological process has to reduce jobs:

> If a newspaper covers an event and it needs a video for the web page, the coverage could include a camera operator, a video editor. That is not happening. They diversify their media, and include digital communication, but they do not diversify the workers involved.

The union highlights that in the field of media, the most important issue should be the quality of information, and AI could be an important resource to improve this. "In general, media owners are not interested in that. They are mainly interested in the costs of producing information." They imagine an ideal scenario with owners of newspapers and union representatives discussing the implementation of technology and the skill requirements involved. "We proposed discussions about new processes, new labour categories, but they never sit to talk about this. And neither do they build capacity in the workplace."

In 2018 the union proposed a workshop on coverage using social networks. Employers did not get involved. Their apathy is also a problem. "There is a loss of a willingness, of a motivation to work, that generates apathy. It's like a vicious cycle. Your work depends also on the value the company gives to it."

When speaking to the Union for Workers in the Retail Sector about AI, they point to online retail. The image they fear is of a shop without shoppers. "We are not opposed to technological advancement, but what is under discussion is the essence of work. Instead of promoting better working conditions, technology is worsening them."

They referred to their experience with workers employed by an online delivery service:

> Young workers are colonised by the idea of independent, contractual work; the idea that they are their own bosses, independent in their use of time and in determining their income. This is evidence of another problem we are dealing with: the cultural battle that started with the expansion of neoliberalism, that involves distrust towards unions, and discusses social legitimacy of labour rights.

"We can also see in technology a tool that strengthens our organisation," they added. The union organised a capacity-building workshop on digital communication tools and set up a network of union representatives, "to work on the circulation of information. Information is power and informed representatives can discuss the possibility of better working conditions with employees."

An economy of platforms[10]

At the end of 2018, the first union of workers from digital platforms was created under the name of APP (*Asociación de Personal de Plataformas*), mainly made up by workers from Uber, and the on-demand delivery service offered by Rappi[11] and courier service offered by Glovo.[12] This was after the workers raised issues such as income, security and transparency in the assignment of work. However, they felt that after raising their complaints, they were discriminated against in the assignment of customers and deliveries.[13] In a statement they declared:

> We have to learn from unions. But we also believe that unions have to learn from us. We, the platform workers, have to organise ourselves. If this is the economy of the future, how could it be that we work in such precarious conditions? If this is the future of the economy, we will have to build the unions of the future. If we do not do that, the thousands of workers who will come to work on platforms will have nobody to defend them.[14]

The recent growth of online platforms in the region responds to a variety of reasons: flexibility in online payment systems, a lack of employment, immigration, overpopulation in metropolitan areas, lack of

[10] Madariaga, J., Buenadicha, C., Molina, E. y Ernst, C. (2019). *Economía de plataformas y empleo ¿Cómo es trabajar para una app en Argentina?* CIPPEC-BID-OIT. Buenos Aires https://www.cippec.org/wp-content/uploads/2019/05/Como-es-trabajar-en-una-app-en-Argentina-CIPPEC-BID-LAB-OIT.pdf

[11] https://www.rappi.com.ar

[12] https://glovoapp.com/en

[13] Rumi, M J. (2018, 18 July). Primer conflicto de trabajadores de una app. *La Nación*. https://www.lanacion.com.ar/economia/empleos/primer-conflicto-de-trabajadores-de-una-app-mensajeros-reclaman-cambios-en-las-condiciones-laborales-nid2154219

[14] Zuazo, N. (2018, 10 October). Nace el primer sindicato de plataformas en Argentina. *Política y Tecnología*. https://guerrasdeinternet.com/nace-el-primer-sindicato-de-trabajadores-de-plataforma-de-la-argentina/

public transport and traffic chaos.[15] In 2018, 1% of people employed in Argentina were users or providers of platforms. Workers are mainly young. For every five, four are men. Only 40% receive social security and 90% are taxpayers under a simplified tax regime (*monotributistas*). Almost 90% have finished high school and 37% higher education. For highly qualified workers, platforms may mean an opportunity for professional development.

So while platforms offer formal working conditions, they are not respectful of worker rights. The CEO of the biggest platform in Argentina declared that the digital economy needs certain conditions to optimise its benefits, but pointed to "the problem of regulations and unions that are prejudicial to business."[16] Workers and unions in this environment are, as a result, worried about their rights and the future of their work.

Gender: From "gaps" back to "bias"

The access women have to technology – in all the senses of access – is relevant to understand the gaps that women experience in the exercise of their rights. In the field of AI, gender gaps also appear in the biases that reproduce prevailing concepts about gender, based on past stereotypes and experience, that discriminate against women.

A frequent example is the bias expressed by algorithms that help with the selection of personnel: women are not selected for positions mainly occupied by men. Along the same lines, a platform that offers taxi journeys assigns them more frequently to men than to women due to security concerns. These examples are influencing debates related to biases reproduced by algorithms in general.[17]

Another point to consider is whether women are benefiting from automation or not. Women should be asked this question. According to Becky Faith, a technology researcher, "unpaid care and a lack of digital access and skills are just two of the issues that we need to put on the table to get women into debates about automation."[18] Working at home also calls for a gender perspective. It is more attractive for women than the usual job of having to balance domestic and paid work responsibilities, but working from home can result in an overburdening of both areas of responsibility.

Finally, the respondent from the Public Service Workers Union added another issue related to gender:

> Digitisation not only helps with the organisation of the gender movement, but also provides the movement with useful data sets, both at a macro level and the organisational level. A claim is strengthened if it is validated by data. A reliable record of digitised information allows us to quantify specific needs that support claims and proposals to improve the working conditions for women.

Conclusion

Unions expressed enthusiasm when invited to talk about the impact of AI on the workplace, appreciating the interview as an opportunity to explore the issues involved. The first challenge that emerged in the conversations was the definition of AI.[19] In all the interviews, respondents expressed an uncertainty over misusing the concept. AI is a concept that involves numerous points of reference, making it "a kind of moving target, changing the pace of technological obsolescence, and including computer programs, algorithms or apps."[20] Unions referred to the introduction of computers as the beginning of AI. They also mentioned digitisation and, in rare cases, algorithms and automation. This report does not capture the views of industrial unions, which would refer more specifically to robotics.

Automation effectively replaces jobs but, in some cases, also generates new ones. However, unions insisted on the responsibility of employers in this change. The lack of dialogue to collectively analyse this technological transformation is a reality. As technology for change expert Becky Faith affirms: "If businesses profiting from automation aren't able to understand and mitigate the impacts of their activities we need social dialogue to hold them to account."[21]

It also must be underlined that while the private sector points to the lack of qualifications of workers in an AI-driven workspace, there are new jobs created that are then performed by overqualified workers under precarious working conditions. Regulation, in this sense, is necessary to promote policies and to

[15] Iproup. (2019, 30 January). Rappi, Glovo and Orders Now: how do they divide the business and why they are a "time bomb". https://www.iproup.com/economia-digital/2394-nueva-economia-e-business-e-commerce-Rappi-Glovo-and-Pedidos-Ya-bussiness-division-and-why-they-are-a-time-bomb

[16] Del Río, J. (2018, 14 October). Del B20 a los sindicatos: definiciones en voz alta de Marcos Galperin. *La Nación*. https://www.lanacion.com.ar/economia/del-b20-a-los-sindicatos-definiciones-en-voz-alta-nid2181413

[17] Seminar on Social Impacts of Artificial Intelligence, University of Buenos Aires. https://www.youtube.com/watch?v=p--EKd-D9og

[18] Faith, B. (2017, 24 October). Automation and the future of work: Bringing women into the debate. *GenderIT.org*. https://www.genderit.org/feminist-talk/automation-and-future-work-bringing-women-debate

[19] Puyana, A. (2019, 26 March). Op. cit.

[20] Ibid.

[21] Faith, B. (2017, 24 October). Op. cit.

prevent vulnerable working conditions. Platforms and call centres are good examples of how unregulated spheres can result in the violation of rights.

In this context the role of unions also needs to be raised. Although they are still the main actor in the representation of employees, membership is low. In a context characterised by a cultural battle against arguments that question the legitimacy of union demands, some voices are proposing that unions should create new channels for organising new jobs,[22] search for young people as members, incorporating groups that have decided not to be formally employed and also consider new types of contractual relations. AI and new trends in the workplace are definitely a challenge for unions in these revolutionary times.

Action steps

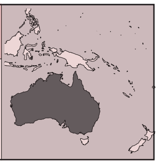
In Argentina there is a need to:

- Properly define and analyse the specific areas of work in different sectors that will be impacted by AI in order to inform policy proposals.
- Promote the debate around training and reskilling. Unions should be consulted regularly about skills and the government should promote training in AI technologies in the education system.
- Unions should create a meeting space to follow the latest debates on AI, to share strategies, evaluate trends and political contexts as well as the potential of AI.
- Promote a regional perspective on how AI is affecting the workplace.
- Work on understanding the potential of AI to improve working conditions, through the analysis of data sets related to labour and AI. "What do unions need AI for?" could be a good starting point.
- Raise public awareness of worker layoffs that are the result of technological innovation, and encourage discussion of the impact of job losses.

[22] Wisskirchen, G., et al. (2017). *Artificial Intelligence and Robotics and Their Impact on the Workplace*. IBA Global Employment Institute. https://www.ibanet.org/Document/Default.aspx?DocumentUid=c06aa1a3-d355-4866-beda-9a3a8779ba6e

AUSTRALIA

SOCIAL *(IN)*SECURITY AND SOCIAL *(IN)*JUSTICE: AUTOMATION IN THE AUSTRALIAN WELFARE SYSTEM



Deakin University School of Humanities and Social Sciences; Australian Privacy Foundation
Monique Mann
https://www.deakin.edu.au/humanities-social-sciences
https://privacy.org.au

Introduction

This report examines automation in social security settings in Australia using the case study of the "Better Management of the Social Welfare System" initiative,[1] also known as the "Online Compliance Intervention" or colloquially as "RoboDebt". Three years ago, the Australian Department of Human Services (DHS), via Centrelink, the agency responsible for administration of social security benefits, launched the automated debt identification programme to detect income reporting discrepancies and the "overpayment" of social security benefits. In practice, the programme aims to achieve savings on social security via the raising of government debts for suspected "overpayment" of welfare benefits, while simultaneously introducing obstacles to contest them.

This programme clearly demonstrates the potential for abuse when big data and automation are deployed against vulnerable individuals. In a wider context of government austerity and cost cutting in both social security and social security services, it reveals the social justice impacts of automated systems that are explicitly designed to target vulnerable individuals. Yet this case study also offers a glimmer of hope regarding the central role that grassroots activism and community-led campaigns can play in countering unjust automated technological systems with human voices. This specifically relates to how conversations about automated technologies can be community focused and inclusive.

Background

In July 2016, the DHS launched an automated debt-raising programme with an algorithm identifying the suspected "overpayment" of government welfare benefits. Request for information letters were sent to welfare recipients requiring them to prove that they did not have a debt, with unconfirmed or unpaid debts subject to an automated recovery process, which included withholding social security payments or outsourcing to private debt collectors.[2] A 10% debt recovery fee was added to the alleged debt.[3]

Between November 2016 and March 2017, 230,000 letters were sent to welfare recipients directing them to an online portal requiring them to prove they did not have a debt. The debt notices were sent out at a rate of approximately 20,000 each week. It was estimated that 20-40% of the debt letters were false positives due to errors in the data used in the system, and the processes that involved averaging annual income and matching it to fortnightly income periods.[4]

Two subsequent inquiries were held: one referred by the Senate Community Affairs References Committee in response to a grassroots community campaign (#NotMyDebt, discussed below), and one initiated by the Commonwealth Ombudsman[5] in response to an increasing number of complaints. The Senate inquiry recommended suspension of the system until issues of procedural fairness were addressed.[6] The Commonwealth Ombudsman recommended assistance and support be provided to vulnerable people, and consultation with stakeholders about the difficulties that vulnerable groups face in interacting with the system.[7]

[1] https://www.humanservices.gov.au/individuals/subjects/compliance-program

[2] Knaus, C. (2017, 11 April). Almost half of all Centrelink robo-debt cases sent to private debt collectors. *The Guardian*. https://www.theguardian.com/australia-news/2017/apr/12/almost-half-of-all-centrelink-robo-debt-notices-sent-to-private-debt-collectors

[3] Commonwealth Ombudsman. (2017). *Centrelink's automated debt raising and recovery system: A report about the Department of Human Services' online compliance intervention system for debt raising and recovery.* https://www.ombudsman.gov.au/__data/assets/pdf_file/0022/43528/Report-Centrelinks-automated-debt-raising-and-recovery-system-April-2017.pdf

[4] Community Affairs References Committee. (2017). Senate inquiry into the design, scope, cost-benefit analysis, contracts awarded and implementation associated with the Better Management of the Social Welfare System initiative. Canberra: Commonwealth of Australia. https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Community_Affairs/SocialWelfareSystem

[5] Commonwealth Ombudsman. (2017). Op. cit.; see also: Commonwealth Ombudsman. (2019). *Centrelink's Automated Debt Raising and Recovery System: Implementation Report.* https://www.ombudsman.gov.au/__data/assets/pdf_file/0025/98314/April-2019-Centrelinks-Automated-Debt-Raising-and-Recovery-System.pdf

[6] Community Affairs References Committee. (2017). Op. cit.

[7] Commonwealth Ombudsman. (2017). Op. cit.

The fiscal context in which the programme was implemented is of central relevance. In 2015-2016, the Australian government forecast it would save AUD 1.7 billion over five years via the identification of welfare overpayments. In 2016-2017, the government then indicated it would achieve AUD 3.7 billion in savings over four years. Within the first six months of the 2016-2017 financial year, the DHS had attempted to recover AUD 300 million of social security "overpayments", and had secured AUD 24 million in repayments.[8] Given the small amounts that the DHS had recovered, it has been questioned whether the government will achieve the forecasted savings via this programme.[9] Casting further doubt on whether the system will achieve any savings, in June 2018 it was reported that the DHS had spent AUD 375 million on the automated debt recovery programme.[10] Despite this, at the time of writing, the system continues to send out "RoboDebts" to meet the financial performance targets set by government.[11]

The "RoboDebt" disaster

This case study has clear implications for inclusivity and social justice involving the explicit targeting of automated technology on vulnerable populations.[12] There are lessons for the design of automated systems to ensure that they produce accurate findings, with the need for clear avenues for review of automated decisions, especially when vulnerable populations are concerned. Importantly, the case study highlights the role that grassroots activism and community-led initiatives can play in foregrounding human impacts of automated systems. There is a need to engage with the community to evaluate social justice and human rights impacts of automated technology deployed in public settings for public service provision *prior* to their implementation. This should involve consultation about whether these types of programmes should be implemented at all, and if so, whether appropriate checks and balances are introduced such as risk, impact, appeal and accountability processes.

Targeting automated technology at vulnerable populations

This is an example of welfare surveillance to further marginalise people who have received social security benefits. It demonstrates that the use of automated technology can serve to perpetuate structural and administrative violence against those who are socially excluded and financially disenfranchised.[13] There are significant social justice issues associated with explicit design of a programme involving automated technologies that target vulnerable individuals facing financial hardship. The automated debt-raising system began sending out debt notices seven weeks before Christmas, already a time where financial pressure is high, especially for those receiving welfare benefits.[14]

The design of the system created issues of administrative justice, procedural fairness, and the rule of law.[15] Numerous challenges were experienced by individuals who attempted to challenge debts, with the onus placed on them to prove they did not have a debt. This reverses the onus of proof onto vulnerable people (and thus overturns the presumption of innocence), which requires them to navigate complex bureaucratic and technical systems to contest alleged debts. According to law, the onus to prove the debt existed technically remained with the DHS: "absent sufficient evidence of an actual debt based on the proper fortnightly data, there can be no *legally sustainable* decision to raise and recover the debt as speculated from averaging."[16] Despite this, the "scheme targets and raises debts in every case where the person cannot *disprove* the possible overpayment."[17] Further, the system is computerised and presumably "objective" so it is harder to argue that it is wrong.

Given that debt notices are being sent out by a government agency, extending six years into the past, when individuals are told by that same agency

[8] Ibid.

[9] Ibid.

[10] Barbaschow, A. (2019, 13 February). Human Services has spent AU$375m on 'robo-debt'. *ZDNet*. https://www.zdnet.com/article/human-services-has-spent-au375m-on-robo-debt/

[11] Henriques-Gomes, L. (2019, 29 May). Centrelink still issuing incorrect robodebts to meet targets, staff claim. *The Guardian*. https://www.theguardian.com/australia-news/2019/may/29/centrelink-still-issuing-incorrect-robodebts-to-meet-targets-staff-claim?CMP=Share_iOSApp_Other

[12] Mann, M., & Daly, A. (2019). (Big) Data and the North-in-South: Australia's Informational Imperialism and Digital Colonialism. *Television and New Media, 20*(4), 379-395.

[13] Ibid.

[14] Community Affairs References Committee. (2017). Op. cit.

[15] See for example: Carney, T. (2018). The New Digital Future for Welfare: Debts Without Legal Proofs or Moral Authority? *UNSW Law Journal Forum, March,* 1-16; Galloway, K. (2017). Big data: A case study of disruption and government power. *Alternative Law Journal, 42*(2), 89-95; Hogan-Doran, D. (2017). Computer says "no": Automation, algorithms and artificial intelligence in Government decision-making. *The Judicial Review, 13,* 1-39; Zalnieriute, M., Bennet-Moses, L., & Williams, G. (2019). The rule of law and automation of government decision-making. *Modern Law Review, 82*(3), 425-455.

[16] Carney, T. (2018). The New Digital Future for Welfare: Debts Without Legal Proofs or Moral Authority? *UNSW Law Journal Forum, March.* www.unswlawjournal.unsw.edu.au/forum_article/new-digital-future-welfare-debts-without-proofs-authority (Emphasis in original.)

[17] Ibid. (Emphasis in original.)

that they are only required to maintain records for six months, those who are "the least literate, least powerful, and most vulnerable" may accept the debt is true and seek to pay it off, even when there is a high probability it is erroneous.[18] This places additional financial burdens on those in receipt of welfare benefits, and "[i]n light of the likely vulnerability of so many Centrelink clients, this is a heavy burden indeed."[19]

Design of the debt-raising system

The debt-raising programme has issues in design and technical considerations, including data matching on the basis of inaccurate data, mismatching data, and the averaging of annual income data that is subsequently matched to fortnightly reporting periods to determine if overpayment has occurred during that period (due to under-reporting of fortnightly income data by the welfare recipient). At the initial design stages, a comprehensive risk assessment and consultation with experts and community groups should have identified such issues. However, no consultation occurred.

The averaging of annual income and matching it to fortnightly income periods involved the extrapolation and creation of an assumed fortnightly income average, when actual fortnightly income may fluctuate.[20] The Senate inquiry recommended that "the department resume full responsibility for calculating verifiable debts (including manual checking) relating to income support overpayments, which are based on actual fortnightly earnings and not an assumed average."[21] This is significant, as 20-40% of the automated debt notices raised by the system were estimated to be erroneous.[22] It also raises the question of whether the DHS took reasonable steps in ensuring the accuracy of the information used for the purposes of debt recovery, whether individuals were able to correct their personal information, or even whether they knew they had a right to do so under Australian privacy law.[23]

Individuals experienced numerous obstacles in speaking to a human in order to correct their personal information, or contest the accuracy of the automatically identified debts. In the two years following the introduction of the programme, millions of calls to Centrelink went unanswered. In 2016-2017, the year immediately following the introduction of RoboDebt, there were 55 million unanswered calls to Centrelink. In 2017-2018, 48 million calls went unanswered.[24] Many individuals waited for hours on hold.[25] The inability to speak with a human undoubtedly created challenges for having false debts resolved. The DHS had a clear conflict of interest as "the harder it is for people to navigate this system and prove their correct income data, the more money the department recoups."[26] So, individuals paid debts, even though they did not believe they owed a debt, because it was "too difficult or too stressful to challenge the purported debt raised against them. Others simply paid the purported debt because they thought the government wouldn't make a mistake."[27]

Fairness and transparency

The use of big data and automated processes allowed for the government to implement a system at national scale with many inherent flaws. There was an absence of fairness in the entire operation of the system including: lack of consultation with stakeholders; no testing or risk assessment processes; the process of automated averaging and matching of data; millions of unanswered calls to DHS; no information being provided to individuals when they sought to challenge a debt; the imposition of a fee for debt recovery; and the fact that the programme extended six years into the past, when individuals are advised they only need to keep records for six months.[28] Accordingly, one of the recommendations arising from the Senate inquiry was that clear and comprehensive advice on reassessment, review rights and processes should be made available to impacted individuals.[29]

There are ongoing issues of explainability and transparency. The debt-raising letters contained no information about how debts were calculated, nor were individuals informed that the debt could be false, or how to contest the calculations.[30] The DHS

[18] Ibid.

[19] Galloway, K. (2017). Big data: A case study of disruption and government power. *Alternative Law Journal, 42*(2), 89-95. https://journals.sagepub.com/doi/full/10.1177/1037969X17710612

[20] Carney, T. (2018). Op. cit.

[21] Community Affairs References Committee. (2017). Op. cit.

[22] Ibid.

[23] Hutchens, G. (2017, 18 May). New privacy code for public servants after Centrelink 'robo-debt' debacle. *The Guardian*. https://www.theguardian.com/australia-news/2017/may/18/new-privacy-code-for-public-servants-after-centrelink-robo-debt-debacle

[24] Dingwall, D. (2018, 30 October). 'No party poppers' for Centrelink's 48 million unanswered calls. *The Sydney Morning Herald*. https://www.smh.com.au/politics/federal/no-party-poppers-for-centrelink-s-48-million-unanswered-calls-20181029-p50c02.html

[25] Community Affairs References Committee. (2017). Op. cit.

[26] Ibid.

[27] Ibid.

[28] Ibid.

[29] Ibid.

[30] Henman, P. (2017, 4 September). The computer says 'DEBT': Towards a critical sociology of algorithms and algorithmic governance. *Zenodo*. https://doi.org/10.5281/zenodo.884117

has refused to release documents that relate to the operation of the system, including risk assessment processes, issues papers, and ICT system reports. In June 2019, the Australian Information Commissioner ruled the DHS must release documents following freedom of information requests first lodged in 2017, yet the DHS later appealed this decision to the Administrative Appeals Tribunal. The DHS challenged the release of these documents claiming that publication may pose security risks and external threats, or as one media article stated: "Human Services claimed people wouldn't pay debts if informed about its IT systems."[31]

Doxing dissenters, chilling critics

In reaction to an individual complaining about the system online, the DHS publicly released their personal information including welfare history to the media.[32] This is an outrageous and reprehensible breach of individual privacy in an attempt to suppress public criticism. In response, a new privacy code is being developed for Australian public servants[33] and it was acknowledged by the Senate inquiry that the system "disempowered people, causing emotional trauma, stress and shame. This was intensified when the Government subsequently publicly released personal information about people who spoke out about the process."[34]

#NotMyDebt: Human voices and community initiatives

Notwithstanding the department's attempts to chill critics, a grassroots campaign known as #NotMyDebt was launched by volunteers led by Lyndsey Jackson, chair of Electronic Frontiers Australia, and united by their "deep concern about the injustice of Centrelink's robo-debt fiasco and the impact it's having on the lives of ordinary citizens."[35] The #NotMyDebt campaign demonstrates avenues for making conversations about artificial intelligence more inclusive and human-centred. The campaign collects, houses and disseminates individual stories of false debts, and provides information and advice to individuals who have received a RoboDebt notice. To date, the #NotMyDebt campaign has collected almost 900 anonymous stories that foreground individual human voices and show the human and social impact of the automated system.

Conclusion

This case study casts light on the failings of big data and automated technology when deployed in public settings for cost-cutting purposes. The RoboDebt programme encapsulates "an error-riddled, unaccountable and politically-driven process."[36] It clearly demonstrates the potential for abuse when big data and automated technology are deployed against vulnerable individuals. Rather than advocating for technical fixes to perfect and optimise the operation of these types of automated systems and technologies, there is a greater need to question whether they should be developed and deployed for certain objectives at all.[37] This involves assessment of the desired aims of the system, and consideration of specific social contexts for their use. In the present case, the aim of the system was to explicitly target vulnerable individuals with debts in an attempt to achieve savings on social security. Given this, it is perhaps unsurprising that the system was designed and operated as intended: it raised large numbers of (erroneous) debts, simultaneously introduced insurmountable obstacles to understand and contest them, and publicly targeted those who spoke out against it.

This case study also demonstrates that automated technology is not merely a technical fix deployed to increase administrative efficiency, but is socially and politically embedded.[38] Automated verdicts have human victims. Recognition that automated technology and systems are socially contingent[39] necessitates proper *a priori* evaluation of the possible social justice and human rights consequences. It requires direct consultation with the community, and involves questioning the type of society that we wish to create, and the role of automated technology within it. As the #NotMyDebt campaign launched in response to RoboDebt has shown, putting human voices back into focus can be an effective strategy of demonstrating not only individual but also wider societal impacts of automated technology.

[31] Stilgherrian. (2019, 7 June). Human Services claimed people wouldn't pay debts if informed about its IT systems. *ZDNet*. https://www.zdnet.com/article/human-services-claimed-people-wouldnt-pay-debts-if-informed-about-its-it-systems

[32] Sadler, D. (2018, 30 May). A 'chilling effect' on free speech. *InnovationAus.com*. https://www.innovationaus.com/2018/05/A-chilling-effect-on-free-speech

[33] Hutchens, G. (2017, 18 May). Op. cit.

[34] Community Affairs References Committee. (2017). Op. cit.

[35] https://www.notmydebt.com.au/about-site

[36] Henman, P. (2017, 4 September). Op. cit.

[37] See for example: Powles, J., & Nissenbaum, H. (2018, 7 December). The Seductive Diversion of 'Solving' Bias in Artificial Intelligence. *Medium*. https://medium.com/s/story/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53

[38] Henman, P. (2017, 4 September). Op. cit.

[39] Ibid.

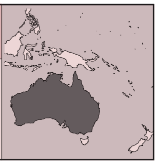
Action steps

The following lessons are suggested by the debt-raising programme:

- Consult the community about the type of society that we wish to create, and the role of automated technology within it.
- Conduct and release risk and impact assessments, including assessment of the impacts for vulnerable individuals and groups, prior to the deployment of automated technology, and periodically during deployment.
- Ensure there are appropriate constraints, checks and balances, and mechanisms for review. This may include approaches such as Article 22 of the EU General Data Protection Regulation[40] that grants a right not to be subject to automated decisions, or transparency provisions such as a right to understand the basis of decisions, and availability of non-automated remedies, for example, the ability to speak to humans during reviews and appeals.
- Support community initiatives and grassroots campaigns that foreground human voices and the human impacts of automated systems.

[40] https://gdpr.algolia.com/gdpr-article-22

AUSTRALIA



SKIRTING THE UNCANNY VALLEY: RISKS, ETHICS, OPPORTUNITIES AND CAUTION FOR CREATIVE INDUSTRIES AND HUMANITY

Faculty of Health, Arts and Design, Department of Media and Communication, Swinburne University
Andrew Garton
www.swinburne.edu.au/health-arts-design/schools-departments/arts-social-sciences-humanities/media-communication

Introduction

The artificial intelligence (AI) technology revolution is set to eclipse the industrial revolution. Where once commonplace trades were dissolved by mass production industries and workers drawn from subsistence living to man-handle factories, AI will replace pretty much every routine job on the planet.[1] The only jobs protected from an AI-augmented workforce are what Kai-Fu Lee, venture capitalist and AI pioneer, describes as jobs of compassion and creativity. “AI,” says Lee, “can optimise, but not create.”[2]

I am a creator, a filmmaker and musician with a background in community media. Toby Walsh, one of Australia’s leading experts in AI, argues that those who raise alarm about an AI achieving consciousness and thereby deciding all humans are expendable are unlikely to be computer scientists and least of all computer scientists who work with and on AI technologies.[3] I am neither. I am neither a computer scientist, nor am I alarmed.

Technology has fascinated me near on as much as nature. Observing both, at times intimately, throughout my career, it is a sad irony that as we create the most incredible of means to advance all facets of knowledge, our rapacious hunger for the Earth’s natural resources is consuming the biosphere all life depends on. Yet we skirt, if not flirt at the edge of technologies that increasingly mimic or resemble human behaviour, more commonly referred to in AI circles as the “uncanny valley”.[4]

Researching this report has been illuminating. There are so many life-affirming AI projects underway, it is bizarre and actually frightening that some governments prefer to invest in AI-assisted surveillance and warfare than direct all efforts towards solving the crisis we are in. That said, there is not a single person whose work I reference that is not aware of the precipice we have neared.

In this report, I contextualise the debates concerning AI in the context of the creative industries in Australia.

I then expand the discussion to outline the policy levers that need to be considered in both the Australian and international contexts. This includes steps one can take to ensure that what some say could be humanity’s greatest or most disastrous creation[5] becomes a transformative technology that works with and not against the best interests of society. Is the world we create with AI, as Walsh hopes for, “the one that we want?”[6]

Background

In 1994, Australia’s first Commonwealth cultural policy document, Creative Nation, was published.[7] Cultural production gave way to the term creative industries, broadening Australians’ understanding of the arts and re-framing culture in economic terms. The language of arts and cultural practice changed, as did what governments chose to fund and what artists would create given the emergence of the internet and new electronic media technologies at our disposal.

The “creative industries” is comprised of film, television, radio, music and the performing arts, publishing and the visual arts. It also embraces creative services such as advertising and marketing, architecture and design, software and all manner of digital content creation and application.[8]

[1] Gallagher, S. (2019, 18 June). The fourth Industrial revolution emerges from AI and the Internet of Things. *Ars Technica*. https://bit.ly/2XXREMb

[2] Lee, K. (2018, 28 August). How AI can save our humanity. *TED*. https://youtu.be/ajGgd9Ld-Wc

[3] Walsh, T. (2018). *2062: The World That AI Made*. La Trobe University Press.

[4] Schwarz, R. (2013, 25 November). 10 Creepy Examples of the Uncanny Valley. *Stranger Dimensions*. https://www.strangerdimensions.com/2013/11/25/10-creepy-examples-uncanny-valley

[5] Hern, A. (2016, 19 October). Stephen Hawking: AI will be ‘either best or worst thing’ for humanity. *The Guardian*. https://www.theguardian.com/science/2016/oct/19/stephen-hawking-ai-best-or-worst-thing-for-humanity-cambridge

[6] Walsh, T. (2018). Op. cit.

[7] Department of Communication and the Arts. (1994). *Creative nation: Commonwealth cultural policy.* https://trove.nla.gov.au/work/16860085

[8] Higgs, P., & Lennon, S. (2014). *Applying the NESTA Dynamic Mapping definition methodology to Australian classifications*. Brisbane: Queensland University of Technology. https://eprints.qut.edu.au/92726

AI technologies are increasingly in use in all facets of creative industries, including to create and curate content. Architects have been using intelligent design systems since the 1990s; EcoDesigner STAR[9] and depthmapX[10] simulate environments measuring the impact the built environment may have on movement and light at any time of any given year.[11] Brian Eno used intelligent systems as early as 1996 to create Generative Music 1,[12] a collection of compositions available on floppy-disk that included the "generative" software required to play his works. Now Australian digital artist, composer and filmmaker David Nerlich works with the latest AI apps that use neural networks "improvising chaotic forms and then selecting and refining what is magical from that chaos," creating what he describes as "digital images that look like paintings."[13]

Spotify's 100 million subscribers,[14] meanwhile, are being monitored by algorithms whether they are aware of it or not. Listeners are served up with customised playlists constantly refined by observing and learning from our listening habits 24/7.

On the near horizon are AIs designed with emotional intelligence, what Rosalind Picard, director of the Affective Computing Research Group, MIT Media Lab, describes as "affective computing".[15] The potential, Picard suggests, will be transformative, and it remains to be seen the impact this might have in the creative industries.

But will we get there? As news agencies replace journalists with AI news aggregators that source, authenticate and write stories to compete within 24/7 news cycles,[16] the climate emergency we are living through, I would argue, is transforming everything at a more rapid pace.

Can AI assist our efforts to adapt how we live in a radically transformed biosphere, or will our planet be so changed that few technologies will survive what Melbourne think tank, the Breakthrough National Centre for Climate Restoration, describes as "a near- to mid-term existential threat to human civilisation" by 2050?[17]

Crossing the uncanny valley

One promise is that AI and its associated technologies will, in performing increasingly sophisticated repetitive tasks, allow creatives more time to create. "Intelligent robots and AI solutions," says Robert Berkeley, co-founder of cloud-based outsourcing service Express KCS, "are their most helpful when used to support human processes rather than take them over."[18] While I am sceptical of the convenience vector inherent in Berkeley's statement, which I will discuss later, what can creatives look forward to?

According to the last Australian census, 5.5% of the Australian workforce are in creative employment. Creative services amounts to three quarters of creative industries, growing jobs within the creative economy twice as fast as the rest of the Australian workforce.[19]

The highest income earners are software and digital content professionals; the lowest but fastest growing workforce can be found in music and the performing arts. While the visual arts saw the lowest income and a declining workforce, its mean income between 2011 and 2016 grew the fastest of the Australian workforce. These figures suggest opportunity, and if Kai-Fu Lee is correct, there will be plenty of roles within creative industries that will involve developing, refining and collaborating with machine learning technologies in the coming years.

But not everyone agrees with Lee's optimism.

After being shown an AI-created animation of a zombie-like humanoid using its head as a leg, Hayao Miyazaki, the creator of internationally acclaimed anime films *Spirited Away* and *My Neighbour Totoro*, said:

> Whoever creates this stuff has no idea what pain is whatsoever. I am utterly disgusted. If you really want to make creepy stuff you can go ahead and do it. [But] I would never wish to incorporate this technology into my work at all. I strongly feel that this is an insult to life itself.[20]

[9] https://www.graphisoft.com/archicad/ecodesigner_star

[10] https://varoudis.github.io/depthmapX

[11] Beqiri, R. (2016, 4 May). A.I. Architecture Intelligence. *Future Architecture*. https://futurearchitectureplatform.org/news/28/ai-architecture-intelligence

[12] https://www.discogs.com/Brian-Eno-Generative-Music-I/release/1452850

[13] https://www.instagram.com/stoch_art

[14] https://www.statista.com/statistics/244995/number-of-paying-spotify-subscribers

[15] https://lexfridman.com/rosalind-picard

[16] Martin, N. (2019, 8 February). Did A Robot Write This? How AI Is Impacting Journalism. *Forbes*. https:/www.forbes.com/sites/nicolemartin1/2019/02/08/did-a-robot-write-this-how-ai-is-impacting-journalism

[17] Spratt, D., & Dunlop, I. (2019, May). *Existential climate-related security risk: A scenario approach*. Melbourne: Breakthrough – National Centre for Climate Restoration. https:/docs.wixstatic.com/ugd/148cb0_90dc2a2637f348edae45943a88da04d4.pdf

[18] Berkeley, R. (2017, 7 September). The Role of AI in Creative Industries. *IT Pro Portal*. https://www.itproportal.com/features/the-role-of-ai-in-creative-industries

[19] Cunningham, S., & McCutcheon, M. (2018). Innovation driving Australia's creative economy boom. *QUT*. https:/www.qut.edu.au/news?news-id=128711

[20] Humphries, M. (2016, 12 December). Studio Ghibli Founder 'Utterly Disgusted' By AI Animation. *PCMag*. https:/au.pcmag.com/software/45342/studio-ghibli-founder-utterly-disgusted-by-ai-animation

The team from Japanese telecommunications and media company Dwango, who had worked on the AI model, were stunned at Miyazaki's response. Disappointed their idol had not embraced their efforts, they struggled to describe that their desire is to create images humans could not imagine, to also "make a machine that can draw like humans do." To that Miyazaki responded, "I feel like we are nearing to the end of times. We humans are losing faith in ourselves."[21]

Dwango's AI team had focused their skills on replacing our humanity, not on rediscovering it. Toby Walsh argues that the AI revolution "will be about rediscovering the things that make us human." By focusing on our social and emotional intelligence and our arts practices, Walsh concludes that "our *technological* future will not be about technology, but about our *humanity* […] the jobs of the future are the most human ones."[22]

However, the Australian Digital Alliance's Elliott Bledsoe[23] reminds us that for all the repetitive tasks AI will free artists from, how a considerable number of artists will support themselves will be a challenge:

> Income insecurity and housing affordability are realities for many artists. Artists' incomes are potentially jeopardised by new technologies. For example, many artists draw part of their income from non-arts sources and some of these non-arts income sources come from industries that may be displaced by AI and automation.

Digital artist Chris Rodley, also working with neural networks creating what he describes as "algorithmic horror", suggests that some of this displacement might be good:

> What I think we're going to see with AI is perhaps a gradual erosion of this idea that artists have this absolutely unique insight that really puts them on this other plane from the rest of us.[24]

Author and futurist Arthur C. Clarke imagined an entirely different outcome. "The goal of the future," he said, "is full unemployment."[25] Clarke foresaw a fully automated future that digital economist and writer Nick Srnicek and sociologist Alex Williams argue is top of their list of minimum demands towards a post-capitalist world without work. Their demands include the reduction of the working week, the provision of a basic income, and diminishment of the work ethic, a world where the latest technologies would "liberate humanity from the drudgery of work while simultaneously producing increasing amounts of wealth."[26]

While Walsh may agree with Srnicek and Williams, he argues for a re-imagining of the work ethic, one predisposed to persistent reinvention of ourselves, to lifelong learning:

> Humans will instead need strong analytical skills. They will need emotional and social intelligence. And they will need all the other traits that make us human – creativity, resilience, determination and curiosity. These skills are what will keep us ahead of the machines.[27]

In the meantime, if Dwango's attempts to create a human that could learn how to walk using its head as a leg had failed, the developers behind FakeApp, DeepFaceLab, FaceSwap and MyFakeApp have successfully mimicked real people doing things they did not do nor say. By synthesising speech and fine-grained movement, anyone can be re-represented on video to say and/or do anything. Although there are considerable advantages for media makers, such as correcting an actor's dialogue in films and significant improvement in foreign language voice-dubbing, "deepfakes", as they are known, are problematic.

Veteran multidisciplinary artist David Nerlich (aka Stoch)[28] is concerned by AI's "ability to deceive us":

> Deepfakes warn us we can no longer believe our eyes. It's often possible to spot fakes, but may be just as easy not to. Photography is increasingly less viable as evidence.

And yet satellite imagery comprised of billions of pixels can be analysed by an AI to interpret spectral bands evolving over time, determining the vegetation of any given area on Earth. Machine learning researcher and computer scientist François Petitjean teaches computers to recognise "whether the evolution of colours of a particular pixel corresponds to a gumtree forest or some grassland." Petitjean and his team have created a detailed vegetation map of Victoria, Australia, by understanding the complex information available within billions of pixels that comprise these pictures taken over time.[29]

[21] https:/youtu.be/ngZoK3lWKRc

[22] Walsh, T. (2018). Op. cit.

[23] Interviewed for this report.

[24] Reich, H. (2018, 1 September). Digital artist Chris Rodley says artificial intelligence could spell death of the artist. *ABC*. https:/www.abc.net.au/news/2018-09-01/artificial-intelligence-chris-rodley-on-changing-role-of-artist/10188746

[25] Youngblood, G. (1969, 25 April). Interview: A. C. Clarke author of '2001'. *Los Angeles Free Press*.

[26] Srnicek, N., & Williams, A. (2015). *Inventing the Future: Postcapitalism and a World Without Work*. Verso Books.

[27] Walsh, T. (2018). Op. cit.

[28] Interviewed for this report.

[29] Pelletier, C., Webb, G., & Petitjean, F. (2019). Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series. *Remote Sensing, 11*(5). https://doi.org/10.3390/rs11050523

> Vegetation, in particular, can be tracked like that because plants reflect infrared when they grow and are healthy – you can then track when things grow, how fast, maybe when they're harvested and that tells you the type of vegetation that might be.

Petitjean's map and the data drawn from the Victorian Land Use Information System he and his team utilised are both open data projects available for use under a mix of Creative Commons licences. While such data is available to all creative industries, Bledsoe is concerned that copyright enabled by technology "may have potentially negative impacts on artists":

> Digital rights management, automated scripts to issue notice and take-down requests and AI-based automatic copyright detection software are some of the types of technologies enabling assertion of copyright in the digital environment. While identifying and stopping infringements is important, the same technology can also be used to block legitimate uses of copyright material. Criticism and review has long been a fair dealing exception in Australian copyright law and, since the Copyright Amendment Act 2006, we have had a fair dealing exception for parody and satire. These and other exceptions create circumstances in which reuse of copyright protected material is legitimate without the permission of the copyright owner.

With the emergence of automated copyright infringement notices and "robo-takedown" requests, what Bledsoe describes as "the preferred cease and desist of the internet age", no humans are reviewing these notices to verify their accuracy.

I have myself received a copyright notice from YouTube stating that "[c]opyrighted content was found in your video." Fortunately, it also said that "[t]he claimant is allowing their content to be used in your YouTube video." The music is my own composition used in a short film of my own making. The claimant turned out to be a third party collecting royalties on my behalf. But not all such scenarios turn out so well, as Bledsoe describes:

> On the flip side, the volume of requests that ISPs [internet service providers] receive coupled with the time frame in which they must do something about them necessarily has prompted a move to streamline the removal of content identified in takedown requests (Google's Trusted Copyright Removal Program, TCRP, is an example of this in practice). This automation at both ends of the request's life cycle means that no person is involved in any part of the decision-making process (i.e. what content to target with a takedown request and what content to remove as a result of a takedown request). [...] This can lead to a number of concerning situations where content filtering can lead to unsubstantiated copyright infringement claims, suppression of marginalised voices and political speech and unintended claims of copyright over material in the public domain.

What happens when such systems fail? When legitimate actors are caught up in, for example, Australia's robo-debt crisis,[30] what Nerlich describes as "the veneer or perception of [impartiality and] autonomy attributable to automated debt collection agents" becomes evident. "Malicious policy makers," he alleges, "are hiding behind the supposed impartiality of robots that make frequently inaccurate debt demands of welfare recipients":

> The makers and programmers are hiding behind these robo-agents they've designed, attempting to place responsibility for these decisions at arms length from themselves. The minister didn't do it, the department staff didn't do it, the robot did it. The AI did it. Perhaps even the notion that no one did it becomes permissible.

Self-described long-term optimist and open government geek Pia Andrews[31] proposes that AI may give artists a "novel way of exploring what it means to be or mirror being human. A way to explore what truly makes us human, and what augmented humanity could look like." Andrews imagines "many ways both direct and philosophically that AI could inspire or enable new forms of art."

We will no doubt find many opportunities as we skirt the uncanny valley, but the risks are evident there too. "All the worst human behaviours," says Andrews, "if rewarded financially, become exponentially worse and harder to disrupt with AI."

The government's response to AI

The Australian Government Department of Industry Innovation and Science is developing an ethics framework for AI in Australia.[32] The discussion paper

[30] See the country report by Monique Mann, also on Australia, in this edition of GISWatch.

[31] Interviewed for this report. Pia Andrews is executive director of Data, Insights and Transformation, Department of Customer Service, NSW Government.

[32] https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework

it funded, *Artificial Intelligence: Australia's Ethics Framework*, distilled AI risks down to three issues:

1. *Data governance:* AI-enabled technologies rely on data. Lots of it. Where is this data drawn from, who owns it and how will it be used to develop AI?
2. *Using AI fairly:* How and where will AI be used? Will it be used fairly and will the public be aware of its use?
3. *Automated decisions:* Can we rely on AI to be totally autonomous? In what kind of scenarios would human decision making continue to be relied on?

The authors proposed an ethics plan identifying eight principles supporting the ethical use of AI and its development in Australia.[33] These are:

1. *Generates net-benefits:* The AI system must generate benefits for people that are greater than the costs.
2. *Do no harm:* Civilian AI systems must not be designed to harm or deceive people and should be implemented in ways that minimise any negative outcomes.
3. *Regulatory and legal compliance:* The AI system must comply with all relevant international and Australian local, state/territory and federal government obligations, regulations and laws.
4. *Privacy protection:* Any system, including AI systems, must ensure people's private data is protected and kept confidential plus prevent data breaches which could cause reputational, psychological, financial, professional or other types of harm.
5. *Fairness:* The development or use of the AI system must not result in unfair discrimination against individuals, communities or groups. This requires particular attention to ensure the "training data" is free from bias or characteristics which may cause the algorithm to behave unfairly.
6. *Transparency and explainability:* People must be informed when an algorithm is being used that impacts them and they should be provided with information about what information the algorithm uses to make decisions.
7. *Contestability:* When an algorithm impacts a person there must be an efficient process to allow that person to challenge the use or output of the algorithm.
8. *Accountability:* People and organisations responsible for the creation and implementation of AI algorithms should be identifiable and accountable for the impacts of that algorithm, even if the impacts are unintended.

Public submissions to the national consultation closed on 31 May 2019. However, some state governments in Australia are developing their own AI governance approaches. Pia Andrews describes the outcome of user testing with representatives from the New South Wales government, pointing to the important factor of building public trust in technology:

> We are also exploring how to normalise and make consistent explainable AI approaches, especially where citizens/customers should have line of sight to decision making about them. We have recognised that without explainability, you don't have accountability, appealability or traceability, and ultimately you will lose trust. We are exploring what the trust infrastructure of the 21st century looks like, which doesn't stop at AI or ML [machine learning], but extends to identity, and transparency of government service delivery.

Conclusion

For all our concerns around intelligent systems gaining control over us, Rosalind Picard argues that we have a long way to go before any form of AI becomes self-aware, let alone able to comprehend who we are and how we differ from any of the tasks it will have been assigned to perform.[34] Google's AlphaGo had impressively beaten internationally renowned South Korean Go player Lee Sedol in four out of five games. Its algorithms were designed to predict probability, training itself in days, but at no time was it aware that it had been learning, let alone playing Go.[35] This is an important point.

Jaan Tallinn, co-founder of the Cambridge Centre for the Study of Existential Risk (CSER), is unimpressed with the narrow way in which AIs such as AlphaGo work. He is adamant that we need to program limits to what an AI can do. But if we teach an AI to adhere to human values, it is unlikely to know what a human is. So do we provide these protections

[33] Dawson, D., Schleiger, E., Horton, J., McLaughlin, J., Robinson, C., Quezada, G., Scowcroft, J., & Hajkowicz, S. (2019). *Artificial Intelligence: Australia's Ethics Framework*. Data61 CSIRO. https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/supporting_documents/ArtificialIntelligenceethicsframeworkdiscussionpaper.pdf

[34] https://lexfridman.com/rosalind-picard

[35] Gibney, E. (2017, 18 October). Self-taught AI is best yet at strategy game Go. *Nature*. https://www.nature.com/news/self-taught-ai-is-best-yet-at-strategy-game-go-1.22858

anyway? "It's a frontier," says Nerlich, "and there is an unknown extent of undiscovered territory, and unknown possibilities for actors in that territory."

Senior writer for *CreativeFuture*, Justin Sanders, in summarising his thoughts on how AI might impact copyright industries, quotes Thomas Edison: "Genius is 1% inspiration and 99% perspiration." He asks, "What if machines could take the burden of some of that perspiration, leaving more room for inspiration?"[36] I would argue that it is the 99% that leads us to the mystical 1% genius.

Convenience does not lead to startling, life-changing innovations. It's the hard work of getting there does. If all the effort were to be replaced by an augmented intelligence, if the journey is no longer part of the experience of arriving, how will we learn from all that it takes to reach our destination? A solution to a creative problem arrives through numerous processes. It takes effort. Yes, some repetitive processes can be replaced and we may well learn entirely unique ways of approaching our problems, but I would caution that we do not innovate ourselves out of a meaningful existence.

Picard reminds us that "there is a critical piece missing in AI. That critical piece is us, it's humans. It's human connection."[37] As we skirt the uncanny valley, what kind of world do we want to create there? Will we survive the climate crisis and biomass collapse to know? Andrews offers the long-term optimistic view:

> I think AI is both a threat and an opportunity for every industry, but also for society as a whole. It challenges a lot of 20th century and before paradigms, and we must take a little time to reflect on what sort of society, values and quality of life we want, if we are to have any hope of not reinventing the past with shiny new things. AI, machine learning and indeed emerging tech and social trends of all kinds are not new things to react to. They are all part of a broader paradigm shift that is moving us away from a centralised, secretive, analog and scarcity economy towards one that is highly distributed, open, digital and surplus. Every sector, every discipline, and indeed every human needs to consider how we want to live better, and then use this opportunity in time to build better futures. Then all emerging tech becomes used in the pursuit of something better, rather than the naïve whack-a-mole game of trying to tackle an exponentially growing backlog with linear strategies.

Action steps

While public submissions to Australia's AI Ethics Framework closed in May,[38] we still need to ensure that the framework meets the expectations of public and professional concerns. To achieve this I would recommend:

- Considering the legal and ethical implications of AI within your arts practice. Be informed:
  - Participate in open dialogue with artists and researchers on moderated mailing lists such as YASMIN.[39]
  - Listen to informed discussion, research and interviews on podcasts like Lex Fridman's Artificial Intelligence Podcast[40] and This Week in Machine Learning and AI.[41]
- A deep reading of the *Artificial Intelligence: Australia's Ethics Framework* discussion paper.
- Developing an awareness of legislative proposals and technical options such as:
  - Elimination of bias in algorithms and machine learning tools.
  - Transparency in terms of public knowledge of what services are governed or augmented by AI.
  - Client-side AI functionality ensuring personal data never leaves one's device.
- Participating in public awareness campaigns.
- Engaging with and lobbying local representatives.
- Supporting the work of advocacy groups such as Digital Rights Watch[42] and Electronic Frontiers Australia.[43]
- Supporting the work of computer science and machine learning researchers through organisations such as the Commonwealth Scientific and Industrial Research Organisation (CSIRO), which hosts Australia's premier science, technology and innovation event, Data61.

[36] Sanders, J. (2018, 19 September). Can AI Be Creative? Here's How Artificial Intelligence Might Impact the Core Copyright Industries. *CreativeFuture*. https://creativefuture.org/ai-creativity

[37] Picard, R. (2018, 17 December). Why build AI? *TEDxBeaconStreet*. https://youtu.be/itikdtdbevU

[38] https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework

[39] YASMIN is a network of artists, scientists, engineers, theoreticians and institutions promoting collaboration in art, science and technology around the Mediterranean Rim and beyond. https://ntlab.gr/mailman/listinfo/yasmin_discussions_ntlab.gr

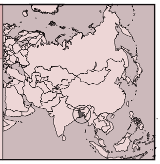
[40] https://lexfridman.com/ai

[41] https://twimlai.com

[42] https://digitalrightswatch.org.au

[43] https://www.efa.org.au

BANGLADESH



ARTIFICIAL INTELLIGENCE AND LABOUR RIGHTS IN THE READY-MADE GARMENTS SECTOR IN BANGLADESH

Bytesforall Bangladesh
Munir Hasan, Fayaz Ahmed and Partha Sarker
https://bytesforall.net

Introduction

Artificial intelligence (AI) can enhance productivity through automation, machine learning and precision. But to automate tasks through an intelligent system that can reduce time, costs and the need for labour also has the potential to diminish jobs. This is the case in the ready-made garments (RMG) sector in Bangladesh. As commentators have put it:

> Changes to the garment industry's business model are threatening the livelihoods of millions of people in low- and middle-income countries, and how these economies adapt will have far-reaching implications.[1]

In particular, female workers, who are at the bottom of the production process and are often engaged in repetitive tasks, are at the greatest risk of losing their jobs in this sector.

Bangladesh is the world's second-largest exporter of garments (after China). The RMG sector in Bangladesh employs 2.5 million female workers in around 4,000 factories across the country[2] and generates 80% of total exports in Bangladesh,[3] contributing approximately 16% to the GDP.[4] As a result, any technology disruption in this sector will have a long-term impact, not only on the sector's production processes and outputs, but also on the workforce employed in the sector. This report explores the impact of AI on the RMG sector in Bangladesh from a rights and development perspective.

Context

Female workers in the RMG sector are mostly involved in the labour-intensive tasks of product development, such as marking and spreading fabric, cutting, sewing, ironing, finishing, tagging and packaging. A study done by the Manusher Jonno Foundation[5] on garment workers in Dhaka and Gazipur found that 72.7% of them do not have any signed contract with their employer. The same study also found that more than 50% of the workers do not even get a salary on time (by the 7th of each month, as per the labour code of Bangladesh).[6] RMG factories in Bangladesh are also known for their unsafe working environments. The collapse of buildings, factory fires, and other incidents in the sector claimed at least 1,512 workers' lives between 2005 and 2013, while 1,691 workers were killed between 1990 and 2013.[7]

Different RMG factories in Bangladesh currently use different advanced technologies for production and supply chain processes. For example, computer-aided design (CAD) tools are being used to expedite product design and development processes, to maximise the use of fabric and to reduce laundry costs. CO_2 dyeing technology uses reclaimed CO_2 in the dyeing of fabrics, eradicating the need for chemicals and water and resulting in a massive reduction in the consumption of energy.[8]

Around 70% of the total manufacturing cost of an apparel product is fabric. The use of tools such as CAD can bring the production cost down by almost 10%.[9]

Technology adoption is directly related to factory size. According to a study by the Centre for Policy Dialogue, the use of advanced technologies in the

[1] Deegahawathura, H. (2018, 8 June). The Garment Industry's Technology Challenge. *CPD RMG Study.* https://rmg-study.cpd.org.bd/the-garment-industrys-technology-challenge

[2] Uddin, M. (2018, 10 September). Artificial Intelligence in RMG: What's in store for Bangladesh? *The Daily Star*. https://www.thedailystar.net/opinion/perspective/news/whats-store-bangladesh-1631515

[3] LightCastle Partners. (2018, 27 September). RMG and textile sectors in Bangladesh. *DATABD.CO*. https://databd.co/stories/rmg-and-textiles-sector-in-bangladesh-526

[4] Uddin, M. (2018, 10 September). Op. cit.

[5] www.manusherjonno.org

[6] The Daily Star. (2018, 17 July). The situation of women workers in the RMG sector in Bangladesh. *The Daily Star*. https://www.thedailystar.net/round-tables/the-situation-women-workers-the-rmg-sector-bangladesh-1606447

[7] Mobarok, F. (2014, 24 April). 1,841 workers killed in 12 yrs. *The Daily Star*. https://www.thedailystar.net/1-841-workers-killed-in-12-yrs-19973

[8] Uddin, M. (2019, 23 March). Technology and Sustainability in RMG. *Dhaka Tribune*. https://www.dhakatribune.com/opinion/op-ed/2019/03/23/technology-and-sustainability-in-rmg

[9] Ovi, I. H. (2017, 27 November). RMG factories turn to technology to maintain competitive advantage. *Dhaka Tribune*. https://www.dhakatribune.com/feature/tech/2017/11/27/rmg-sector-software-cutting-technology

RMG sector is prevalent in 47% of large-scale RMG enterprises, compared to 25% of medium-scale RMG enterprises.[10]

AI and automation technologies are also slowly being adopted by RMG factories, not only because of AI's time-, cost- and resource-saving properties, but also because of its game-changing nature in the RMG business model, as many garment factory owners have said. For example, as a representative of the RMG sector explained:

> AI's powerful vision and image recognition system can easily help in identifying and grading textile fibres. AI can enable accurate performance during cutting, spreading and sewing. CAD and pattern-making, production planning and control, shop floor control systems – all these can be performed at a much higher level of productivity and accuracy with AI technology.[11]

Artificial neural networks (ANN) are also used in different garment manufacturing facilities, including for the prediction of fabric mechanical properties, the classification and grading of fabrics, and the identification and analysis of faults.

Labour rights and the impact of AI

Given that the productivity benefits of AI are becoming evident, it is important to see how it is likely to affect the existing labour force and will have far-reaching implications for the Bangladesh economy. A study done by Access to Information (a2i) in Bangladesh and the International Labour Organization (ILO) estimated that more than two million or 60% of jobs in the RMG sector in Bangladesh may disappear by 2040 due to automation.[12] This study identified the RMG sector as one of the most affected sectors due to AI and automation. Another study from a think tank in Bangladesh, the Centre for Policy Dialogue, shows that the automation of manufacturing reduced the participation ratio of female workers in the garment sector to 60.8% in 2016 from 64% in 2015. The study also shows that the factory owners think female workers are not able to handle modern machinery properly.[13]

Bangladesh still does not have any formal policy on AI implementation in any sector, let alone the RMG sector. Recently, the government's Information and Communications Technology (ICT) Division took steps towards formulating a national strategy for AI, and a draft outline of that strategy has been released. The strategy[14] has identified eight strategic pillars for the future implementation of AI in Bangladesh, from 2020 until 2025: AI in the government; the industrialisation of AI technologies; data and digital infrastructure; skilling the AI workforce; research and development; funding and accelerating the AI ecosystem; inclusive and diverse AI; and ethics, data privacy and security. The strategy tries to explain key issues raised by the implementation of AI in different areas, including AI in citizen service delivery, AI in manufacturing, AI in agriculture, AI for smart mobility and transportation, AI for skills and education, AI for finance and trade, AI for health, and AI for the environment, water and renewable energy. Interestingly, the strategy does not touch upon the issue of job losses and how and when the existing workforce who may lose their jobs due to the implementation of AI would be compensated or redeployed through training or skills development. Apart from this, aspects such as data privacy, security, and the need for regulations identified in the strategy are crucial for AI.

There are different opinions in the country with regard to AI and automation. For example, Dr. Atiur Rahman, ex-governor of Bangladesh Bank, the central regulatory bank in Bangladesh, says that the "collaboration of humans and robots could integrate technological innovation into the apparel industry in Bangladesh."[15] He thinks AI is a pathway to integrate a great number of unskilled production workers into a structurally difficult labour market that depends on foreign investment.

Some also argue that any new technology advancement, while it may make some jobs redundant, may also bring in new types of jobs that did not exist before which nevertheless may require new skill sets. As traditional factory jobs evolve, technology-servicing roles will become more important. Just as sewing machines break and need calibration, so, too, will the apparel printers and packaging systems of the future.[16] There is also an emphasis on technology companies

[10] Saadat, S. Y. (2019, 19 January). Employment in the age of automation. *The Daily Star*. https://www.thedailystar.net/opinion/perspective/news/employment-the-age-automation-1689487

[11] Uddin, M. (2018, 10 September). Op. cit.

[12] The Daily Star. (2018, 21 December). Automation to cut 53.8 lakh jobs by 2041: study. *The Daily Star*. https://www.thedailystar.net/business/news/automation-cut-538-lakh-jobs-2041-study-1676683

[13] The Daily Star. (2018, 4 March). Women losing more jobs to automation: CPD study. *The Daily Star*. https://www.thedailystar.net/business/women-losing-more-jobs-automation-cpd-study-1543222

[14] Alam, S. (2019, 12 April). Development of National Strategy for Artificial Intelligence, Bangladesh. *Medium*. https://medium.com/@shofiulalam/development-of-national-strategy-for-artificial-intelligence-bangladesh-57bcf09ccfaf

[15] Fibre2Fashion. (2018, 12 February). Slowly adopt AI in Bangla RMG sector: ex-central bank chief. *Fibre2Fashion*. https://www.fibre2fashion.com/news/apparel-news/slowly-adopt-ai-in-bangla-rmg-sector-ex-central-bank-chief-240555-newsdetails.htm

[16] Deegahawathura, H. (2018, 8 June). Op. cit.

working more collaboratively with the apparel industries to manage AI and automation platforms. The question is obviously to understand whether new jobs will match the rate of worker lay-offs and whether or not the country can keep pace with the skills gaps and training required to acquire new skills.[17]

Labour standards and labour rights at the RMG factories in Bangladesh paint a very dismal picture. There are three core concerns: wages, access to unions and workplace safety. According to a 2016 ILO survey, the highest number of suppliers who felt compelled to accept orders below production costs came from Bangladesh. This directly affects workers' wages and other working conditions.[18] Union membership among Bangladeshi garment workers is low – estimations range from 5% to 10% – while unionisation is banned entirely in the few export processing zones.[19]

The amended Labour Act of 2013[20] in Bangladesh incorporated provisions including the right for workers to form trade unions without getting the approval of owners, establishing safety measures for employees in the workplace, creating safety committees, establishing workplace health centres, and enabling inspectors to visit factories to assess compliance and to penalise owners if necessary.[21] With the introduction of this amended law, more than 300 trade unions were registered in the RMG sector by April 2015.

The government created another amendment to the Labour Act in 2018, which has already been approved by cabinet. It reduces the requirement of worker participation to set up a trade union from 30% to 20%. It also suggests that the working hours of labourers in general should not be more than 10 hours a day, excluding meal and rest time, and introduces mandatory maternity leave and other leave for expectant mothers.[22]

In an official circular, the Ministry for Labour and Employment announced the increase of minimum wages for garments workers to 8,000 taka (about USD 95) a month – it was 5,300 taka (USD 63) previously. This is also far less than the 16,000 taka (USD 200) that the workers' unions have been asking for.[23] The 2006 Labour Act – which as discussed above was amended in 2013 and 2018 – deals with issues such as job termination, lay-offs and the responsibilities of workers in the workplace. These provisions may be important in the context of technology in the workplace.

As per this Act, workers have some rights in cases where they are laid off by the employer. Whenever a worker whose name is on an employer's master roll of workers, and who has completed at least one year of continuous service, is laid off, they need to be paid compensation by the employer for the days when they are without work. The amount of compensation needs to be equal to half of the total of the basic wage received plus dearness allowance[24] and ad hoc or interim pay if any. Any worker whose name is not on the master roll would not receive anything. Re-employment of retrenched workers is also possible depending on whether the worker is asked to do the same job or a job in the same category within a year. Retrenched workers, according to their length of service, will have higher priority for re-employment over other workers, such as non-contractual and casual workers.[25]

Recent research[26] into the sector found that RMG factories provide few training and development opportunities for their workers outside of the immediate demands of their job. Only 5% of respondents confirmed that they have training facilities for their employees. Training for supervisors is nevertheless common (70%). Worker development depends to a great extent on learning on the job. General workers are recruited as non-skilled workers and learn the necessary work techniques under close supervision of the floor supervisors. As a result, the companies use the supervisors to develop the workers and ensure quality. Skill training institutes for garment workers are rare and, as a result, workers in many cases feel frustrated: 86% of worker respondents reported that there are no training institutes available for developing their knowledge and capability.[27]

[17] Uddin, M. (2018, 10 September). Op. cit.

[18] Kashyap, A. (2019, 1 May). Rights makeover overdue in Bangladesh garment industry. *The Daily Star*. https://www.thedailystar.net/opinion/human-rights/news/rights-makeover-overdue-bangladesh-garment-industry-1737301

[19] Ashraf, H., & Prentice, R. (2019). Beyond factory safety: labor unions, militant protest, and the accelerated ambitions of Bangladesh's export garment industry. *Dialectical Anthropology, 43*(1), 93-107. https://doi.org/10.1007/s10624-018-9539-0

[20] https://www.ilo.org/dyn/natlex/natlex4.detail?p_lang=en&p_isn=94286&p_classification=01.02

[21] Mausumi, N. (2016, 15 December). Recent developments in 'labor rights' and 'safety at workplace' of Bangladesh RMG industries. *Textile Today*. https://www.textiletoday.com.bd/recent-developments-labor-rights-safety-workplace-bangladesh-rmg-industries

[22] The Daily Star. (2018, 4 September). Govt eases trade union rules. *The Daily Star*. https://www.thedailystar.net/news/city/bangladesh-labour-act-amendment-2018-cabinet-approved-banning-child-labour-1628449

[23] Paul, R. (2018, 13 September). Bangladesh raises wages for garment workers. *Reuters*. https://www.reuters.com/article/us-bangladesh-garments/bangladesh-raises-wages-for-garment-workers-idUSKCN1LT2UR

[24] https://en.wikipedia.org/wiki/Dearness_allowance

[25] https://www.ilo.org/dyn/natlex/natlex4.detail?p_lang=en&p_isn=94286&p_classification=01.02

[26] Akter, S., & Alam, R. (2016). HR Practices in Selected RMG Factories in Dhaka. *Journal of Business, 1*(4), 39-42. https://www.researchgate.net/publication/316027572_HR_Practices_in_Selected_RMG_Factories_in_Dhaka

[27] Ibid.

Conclusion

Regardless of potential benefits, AI and automation pose a great risk of job losses for workers, particularly female workers who are at the bottom of the production process. The RMG sector in Bangladesh is already known for its precarious working environment, wage discrimination, lack of formal contracts, massive workload, etc. Even the attempts to unionise workers are made harder by legal restrictions (despite the recent amendments to the Labour Act), intimidation against organisers, and the threat of corporate flight; but local unions also struggle to meet the complex needs of workers employed in the garments sector.[28] AI and automation are one of these complex things that labour unions in Bangladesh are yet to comprehend.

In many countries, trade unions remain one of the major voices to share worker concerns, demand rights and negotiate on job losses, but this is not the case in Bangladesh's RMG sector. This is particularly important because AI will certainly create new types of jobs (for example, machines will require operators, repair technicians, etc.) to which the existing workforce, with appropriate training, can be redeployed and accommodated as a priority.

Bangladesh policy makers and private sector bodies also need to understand that the dependence on cheap labour to gain competitive advantage will not work any longer, as in the longer term this advantage is offset by lower productivity due to lack of skills development. This approach is clearly reflected in the hiring of workers without raising wages in Bangladesh. Available data indicates that real wages in the manufacturing sector in Bangladesh have not increased over the last decade and a half.[29] Education policy and skills development are two important areas that need attention in the context of a game-changing technology such as AI being implemented in the RMG sector, as in other sectors in the country.

Action steps

The following steps are necessary in Bangladesh:

- Civil society in Bangladesh is yet to take into cognisance the emergence and effects of AI, particularly in the RMG sector, a sector where workers are among the most vulnerable and which includes a large number of women in the workforce. A national conversation on the issue is already overdue, as AI technologies have already started to be used in some factories. There is no evidence-based research available with data, facts or figures that could inform the policy-making process. Trade unions in the RMG sector are weak and often are not allowed to operate, let alone make demands for workers' rights. Therefore, civil society has a role to sensitise policy makers about the topic and bring about a change of attitude in addressing the issue. A more human, public good approach to AI implementation will be a win-win situation for everyone.
- A research and evidence-based policy intervention in education and skills development policy is an imperative to minimise the risks posed by AI and automation. As traditional factory jobs evolve, technology-servicing roles will become more important. Just as sewing machines break and need calibration, so, too, will the apparel printers and packaging systems of the future. To help ease the transition from manual to modern manufacturing, businesses and governments must begin improving the tech literacy of current employees. If today's workforces are to remain relevant in the economies of tomorrow, employees will need the skills to contribute.[30] All trade and labour policy may also need to be revisited in order to accommodate AI-specific shocks to society. For example, policy may need to dictate trade agreements that cushion the impact when manufacturing jobs are lost, while laying the groundwork for the transition to more tech-heavy industries.[31]
- Civil society also needs to start paying attention to other issues to do with AI including ethics, data privacy, security, and best practices in regulatory frameworks. For example, the European Union has unveiled ethics guidelines that illustrate that any AI technologies that are to be implemented on an industrial scale need to be accountable, explainable and unbiased. Around 30 countries in the world have already created or drafted an AI framework for the generic application of AI, as well as dealing with specific issues.[32] Bangladesh civil society should formulate a position paper on AI, and this research from Bytesforall Bangladesh can be an important first step in that direction.

[28] Ashraf, H., & Prentice, R. (2019). Op. cit.

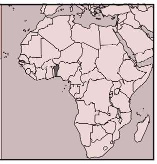
[29] Mahmood, M. (2017, 15 November). Industrial automation: a challenge for Bangladesh's manufacturing future. *CPD RMG Study*. https://rmg-study.cpd.org.bd/industrial-automation-challenge-bangladeshs-manufacturing-future

[30] Deegahawathura, H. (2018, 8 June). Op. cit.

[31] Ibid.

[32] Adib, A. (2019, 13 May). Artificial intelligence, real progress. *Dhaka Tribune*. https://www.dhakatribune.com/opinion/op-ed/2019/05/13/artificial-intelligence-real-progress

BENIN



GIVING GIRLS A HEAD START IN ARTIFICIAL INTELLIGENCE IN BENIN

POPDEV Bénin
Sênoudé Pacôme Tomètissi
tometissi@gmail.com

Introduction

Benin's digital sector is a dynamic one, with an environment that includes a wide diversity of start-ups. Although artificial intelligence (AI) seems new and unclear to some people (many link it only to the police as their use of drones and robots to track criminals), several start-ups invest their resources in the subsector. Some of them, for example femCoders, promote the better engagement of women in this still male-dominated domain.

To help bridge both the digital and gender gap in the country, and to mobilise women around the digital economy, femCoders – an award-winning female-led initiative with a focus on holding free after-school programmes to expose girls to digital technologies – runs a robotics training programme with the support of the United States Embassy in Benin. Through this programme, hundreds of young people, mostly girls, are trained on robot assembling and programming.

Digital equality is at the centre of almost all discussions in Benin. This includes the country's digital policy, with its six key government projects in the information and communications technology (ICT) sector (see below), and the annual Digital Weeks,[1] which aim to highlight and develop the digital ecosystem in Benin through a series of activities including a forum (Afro Tech International Forum), an exhibition (Benin Start-Up Week), training (Benin Digital Tours) and contests on digital solutions and digital art.

Encouraging women's engagement in AI is also an imperative in the country. A Girls in ICT Day,[2] plus a National Network of Women in ICTs that gathers all women, from the minister of digital economy to farmers to students, highlight the importance of gender in the digital economy.[3] However, the sector remains male-dominated, which makes the femCoders initiative all the more impressive. To address inequality in the ICT sector, the country needs more gender-sensitive policies and strategies that will not only support women-led start-ups in the AI field and protect the rights to equality and non-discrimination, but also initiatives that give girls a head start.

The policy context in the digital sector

Benin's sector strategy in the field of ICTs aims to transform Benin into West Africa's digital services hub for accelerated growth and social inclusion by 2021. According to the Sectoral Policy Declaration,[4] the development of the digital economy is an undeniable means for social inclusion and well-being and will help reduce unemployment, catalyse economic development in other sectors, and create transparency and accountability in the public administration. Under the strategic guidelines of the Declaration, by 2021, the coverage rate of broadband internet services should reach 80%, the fixed telephone line penetration rate 40%, and mobile penetration 60%. The size of the market, meanwhile, is expected to reach nearly USD 1 billion and to create 90,000 jobs. As a result, Benin is expected to enter the top 100 of the Networked Readiness Index (NRI) and achieve a UN E-Government Development Index of 0.5.

As part of its Programme of Action (*Programme d' action du gouvernement*),[5] the government is implementing 45 flagship projects, including six in the digital sector. These six projects include the deployment of high and very high speed internet throughout the country, the transition to Digital Terrestrial Television (DTT), the implementation of smart public service (Smart Gov), the roll-out of e-commerce, e-education and e-training, and the promotion and development of digital multimedia content.[6]

In addition to the six core digital projects implemented by the government, the country through its

[1] https://semainedunumerique.bj/fr

[2] Assogbadjo, M. (2018, 7 May). Journée internationale des jeunes filles dans les Tic: Encourager la cible à évoluer dans le secteur. *La Nation*. https://www.lanationbenin.info/index.php/actus/159-actualites/16075-journee-internationale-des-jeunes-filles-dans-les-tic-encourager-la-cible-a-evoluer-dans-le-secteur

[3] Registration form to join the network: https://tinyurl.com/forumRenafetic

[4] https://numerique.gouv.bj/images/DPS.pdf

[5] https://www.presidence.bj/benin-revele/download

[6] www.revealingbenin.com/en/programme-dactions/programme/digital-economy

parliament passed the Digital Code,[7] a collection of laws comprising 647 articles that frame activities related to electronic communication. The Code defines both rights and duties of citizens and service providers, but also the institutional and legal frameworks for activities in the digital sector.

The AI environment in Benin

The digital policy does not formally include AI as a subsector, and the concept is still new to the public. Nevertheless, the ICT environment in Benin is very dynamic, with a wide variety of actors. In addition to mobile operators and other internet service providers, a large number of start-ups in the commercial space offer a variety of digital solutions. Digital Week[8] and Girls in ICT Day,[9] two annual opportunities to share knowledge and promote activities in the sector, help to give greater visibility to start-ups and innovative initiatives. According to Minister of Digital Economy Aurélie Adam Soule Zoumarou, it is the government's duty to encourage students to enter the engineering, computer science and mathematics streams, and to increase the involvement of girls and women in ICTs.[10]

Among the numerous digital actors in the commercial field, some invest in AI even if the subsector is still almost unknown to the general public:

- RINTIO is a digital service start-up specialised in the development of digital solutions, including using big data and AI in its data lab.[11]
- Machine Intelligence For You (MIFY)[12] organises an annual contest on AI combining algorithms with local games (e.g. dominos or *adji* in the local language).[13]
- RightCom is a start-up in the field of AI. Its main product, RightCapture, is software connected to cameras that helps collect data through a real-time video stream analysis. RightCapture can detect movements as well as gender, age and emotions of a human being.[14]
- SmartCo promotes women's initiatives in the digital sector including in AI. For its first annual training, 100 young women aged 18 to 35 were selected and trained on the development of web and mobile applications.[15]
- KEA Medicals is an internet database on people's medical information created by Arielle Ahouansou, a former participant in the Tony Elumelu Entrepreneurship Programme for people under 30 years of age. The platform, an electronic medical log, includes details of an individual's medical history and family contacts. The aim of KEA Medicals is to create an e-identity in the form of a QR code printed on a bracelet or a patch to stick on a smartphone. A scan of the QR code provides access to the patient's medical record.[16]
- Benin Flying Labs is a hub that provides hardware and software training for various drone platforms in Benin. They have projects in a wide range of sectors such as in health, agriculture, conservation and development.[17]

femCoders and its robotics programme

femCoders is a community of women who train their peers in robotics and programming such as logic and languages, Blockly[18] and Scratch games,[19] drawing and 3D printing, and creating websites using HTML. They strengthen the technical capacity of women in an environment where they can share their knowledge with their peers.[20] Following her mentorship as part of Benin's young robot builders' team at the *FIRST* Global Challenge in Washington in 2017[21] – where the team was ranked seventh out of 163[22] – Rachael Orumor founded her start-up "femCoders" and received a grant from the US Embassy in Benin to implement its robotics training programme focused on girls. For its first training programme, about 400 students, most of them, but not all, girls, were drawn from six Cotonou schools.

They were trained on programming three robots, namely Vex Edr Robot, Vex IQ Robot and Benin Bot Official. These robots were trained for various tasks, notably to sing Benin's national anthem (*l'Aube nationale*), or a lullaby to a baby who wakes up in the absence of its mother, to simulate the transport of a patient to a health centre, or to create a labyrinth full of obstacles.[23] The training lasted six weeks for each

[7] https://sgg.gouv.bj/doc/loi-2017-20

[8] https://semainedunumerique.bj/fr

[9] Assogbadjo, M. (2018, 7 May). Op. cit.

[10] Ibid.

[11] https://twitter.com/rintiogroup?lang=en

[12] https://mify-ai.com

[13] https://maic.mify-ai.com

[14] https://right-com.com/fr/la-ministre-beninoise-de-leconomie-numerique-visite-rightcom-a-cotonou

[15] https://labinnovation-sg.com/smartco

[16] https://www.keamedical.net

[17] https://flyinglabs.org/benin

[18] https://developers.google.com/blockly

[19] https://scratch.mit.edu/explore/projects/games

[20] www.femcoders.info/index.php

[21] https://first.global/fr/2017-nations/team-benin

[22] Africa Times. (2017, 19 July). No. 7 Team Benin leads the pack of Africa's FIRST Global robotics winners. *Africa Times*. https://africatimes.com/2017/07/19/no-7-team-benin-leads-the-pack-of-africas-firstglobal-robotics-winners

[23] Meton, A. (2018, 24 August). Projet "Girls focused robotics training programme": Des élèves initiés à la technologique robotique. *La Nation*. https://www.lanationbenin.info/index.php/societe-2/146-societe/17522-projet-girls-focused-robotics-training-programme-des-eleves-inities-a-la-technologique-robotique

school. Seven learning projects in robotics science were created at the schools, and each school benefited from a donation of a robotics kit in addition to the training manuals offered to each student.

Above all, the initiative aims to give girls the opportunity to consider and even embrace a career in robotics at a young age. "We understand that computer science and robotics are replacing people in various tasks today to the point where there is the problem of unemployment, much of which is due to the fact that the tasks are performed by computers instead," said Orumor. "And if this continues, in a few years, we will not be able to continue working in Africa because we do not have the necessary skills in this area," she stressed.[24] "This training allowed me to learn a lot. It was a great experience where we learned the rigour required from the work and team building," said Daniella Bossa, one of the learners, during a demonstration session.[25]

Giving girls a head start in robotics

femCoders and other initiatives – whether by the government or non-state actors – that give girls a head start in AI will improve the participation of women in the digital sector, but will not be enough to bridge the digital divide and the gender gap in the country. There is a need for a digital equality policy that includes women with a clear programme to support gender empowerment.

One aspect of such a revised digital policy would be adopting the UN Development Programme (UNDP) Gender Equality Seal,[26] a ten-step process developed to support companies to close their gender gap by assessing policies and implementation strategies. It guides public and private enterprises in meeting gender standards and promotes empowerment and equality in the business sector. The programme is an excellent opportunity given that most of the actors in the AI field in Benin are start-ups. In February 2019, the Ministry of Social Affairs and Microfinance, with the technical and financial support of the UNDP, organised a training session in Cotonou on certification using the Gender Equality Seal and the process involved. The training gathered officials from the departments of planning and programming as well as gender focal points of ministries and civil society organisations.

"If we are able to close the gaps between men and women in the labour market, in education, in health care, we could substantially reduce this loss of economic gain for our country," said Adama Bocar Soko of UNDP Benin during the training.[27]

Using the UNDP's Gender Equality Seal, especially in the digital and AI sector in Benin, will help increase women's leadership in decision making in the sector and increase their participation. To help this happen, the government can add a Gender Equality Seal certificate for all start-ups in Benin as a requirement for accessing the newly created Digital Fund, which was set up by the government to support digital entrepreneurship. This will encourage start-ups to develop and adopt gender equality policies and action plans to address inequalities. This will also increase the number of women-owned start-ups.

Supporting start-ups by women and introducing a quota at universities for girls in technological studies should contribute significantly to closing the digital gap so that what is currently perceived as a threat is turned into an opportunity. Women make up 52% of the population of Benin. It would therefore be fair to grant them at least parity in scholarships to universities, especially in sectors dominated by men. Creating an AI stream at the National University of Engineering and Mathematical Technologies of Abomey, for example, could pave the way for empowered engagement by women in this subsector of the digital economy.

Another idea would be to introduce the principle of gender rotation in the exhibition stands for start-ups during the Digital Weeks, which are organised by the government in all provinces each year. By prioritising start-ups by women every two years, the Digital Weeks would give the start-ups greater visibility and attract investors. Parity for start-ups in exhibition stands during the Digital Week is another possibility. But a Digital Week dedicated to women in the digital sector is likely to give much more visibility to start-ups by women, which is not necessarily guaranteed with the parity option.

[24] Amoussou, G. (2018, 6 December). La robotique, l'expérience béninoise avec Rachael. *24 Heures au Bénin*. https://www.24haubenin.info/?La-robotique-l-experience-beninoise-avec-Michael

[25] Ibid.

[26] "The GES is a tool for public and private enterprises to come together and contribute towards the achievement of the Sustainable Development Goals (in particular, SDGs 5, 8, 10 and 17) by reducing gender gaps and promoting gender equality and competitiveness simultaneously, for a fair, inclusive and sustainable growth. [...] The programme has created a dynamic partnership between the private sector, public sector, trade unions and UNDP with a tool to develop public policy, foster constructive dialogue, invite companies to go from commitment to action and provide hard-evidence of gender mainstreaming efforts to tackle the most pressing gender inequalities." https://www.genderequalityseal.org/programme

[27] www.bj.undp.org/content/benin/fr/home/presscenter/pressreleases/promotion-de-l_egalite-de-genre-dans-le-secteur-public-au-benin.html

Conclusion

This is a call for innovative approaches to address the issue of gender equality in the digital sector, in particular in the AI subsector. Developing a gender-sensitive policy for AI can help increase the engagement of women in the subsector, but also encourage gender-sensitive strategies and a human rights-based approach in business life.

In its report released in July 2019 entitled *The West Africa Inequality Crisis: How West African governments are failing to reduce inequality, and what should be done about it*, Oxfam ranked Benin 12th out of 16 countries with a score of 0.19.[28] This ranking shows the low commitment of leaders to reduce inequalities. In Benin, as in many other low- and middle-income countries, women are subject to double discrimination, due to both their gender and their socioeconomic context. The fact that AI is a new field in the digital economy offers an opportunity – albeit a small one – to correct some of these imbalances.

Action steps

The following steps are suggested for Benin:

- Revise and adapt the sectoral policy declaration to make it more gender-sensitive.
- Develop and implement a gender-sensitive policy in the AI subsector.
- Include criteria relating to the UNDP's Gender Equality Seal certification for start-ups accessing the Digital Fund, especially those in the AI subsector.
- Introduce the principle of gender rotation in the exhibition stands for start-ups during the Digital Weeks.
- Grant young women at least parity in scholarships to universities, especially in the sectors dominated by men.

[28] Hallum, C., & Obeng, K. W. (2019). *The West Africa Inequality Crisis: How West African governments are failing to reduce inequality, and what should be done about it*. Oxfam International. https://oxfamilibrary.openrepository.com/bitstream/handle/10546/620837/bp-west-africa-inequality-crisis-090719-en.pdf

BRAZIL

WE DON'T NEED NO OBSERVATION: THE USE AND REGULATION OF FACIAL RECOGNITION IN BRAZILIAN PUBLIC SCHOOLS

Instituto de Pesquisa em Direito e Tecnologia do Recife (IP.rec)
Mariana Canto
https://www.ip.rec.br

Introduction

The use of facial recognition technology in schools around the world in countries such as China and the United States has encouraged the similar use of this technology by other countries, including Brazil. However, it has also raised questions and concerns about the privacy of students. Because of this, analyses of the nature and consequences of the use of facial recognition technology in diverse scenarios are necessary.

This report presents a brief reflection on the use of facial recognition technologies in Brazilian public schools, including in the state of Pernambuco, where IP.rec is based, and considers their implications for citizens' rights to privacy, as well as the possibility of the technology being regulated by existing laws.

Background

Artificial intelligence (AI), algorithms, the "internet of things", smart cities, facial recognition, biometrics, profiling, big data. When one tries to imagine the future of big cities, it is impossible not to think about these terms. But is the desire to make cities "smarter" jeopardising the privacy of Brazilian citizens? Does this desire turn people into mere guinea pigs for experimentation with new technologies in a laboratory of continental proportions?

The use of facial recognition technologies in Brazilian public schools has already been implemented in several cities such as Jaboatão dos Guararapes (in the state of Pernambuco), Cabo de Santo Agostinho (Pernambuco), Arapiraca (Alagoas), Cravinhos (São Paulo), Tarumã (São Paulo), Potia (São Paulo), Paranavaí (Paraná), Guaíra (Paraná), Viana (Espírito Santo), Anápolis (Goiás), Senador Canedo (Goiás) and Vila Bela da Santíssima Trindade (Mato Grosso).[1] Among the features provided by the so-called Ponto iD[2] system is the monitoring of the attendance of students at school without the need to take roll call. The system also aims to help optimise class time, as the time spent on the roll call is saved; help manage school meals, as cooks are notified of the exact number of students in class as soon as the gates close; and decrease the school drop-out rate, as guardians receive, through an app, notifications that their child is in the school. The last is noted as a primary social consequence of using the technology. To implement the system, the city of Jaboatão dos Guararapes, for example, has spent BRL 3,000 (USD 780) per month per school.

The technology provider's webpage states that the solution is designed in an integrated way, linking government departments. Because of this, a diverse range of public institutions can share information with each other. For example, according to the government of the city of Jaboatão dos Guararapes, if a student is absent for more than five days, the Guardianship Council is notified as the system also shares students' information with that body.[3]

In 2015, the service provider also stated that the system would be connected to the *Bolsa Família* programme,[4] which is a direct income transfer programme aimed at families living in poverty and extreme poverty throughout the country, and intended to help them out of their vulnerable situation. In Brazil, more than 13.9 million families are served by Bolsa Família.[5] The receipt of the benefits is conditioned, among other duties of a student whose family is a beneficiary of the programme, to a minimum school attendance of 85% for children and adolescents from six to 15 years old and 75% for adolescents 16 and 17 years old.[6]

Privacy? Absent. Risks? Present

As previously observed by several scholars, digital technologies not only make behaviour easier to monitor, but also make behaviour more traceable.[7]

[1] www.pontoid.com.br/home

[2] https://www.youtube.com/watch?v=YRMihVhocew&t=24s

[3] Folha de Pernambuco. (2017, 19 April). Tecnologia para registrar presença nas escolas de Jaboatão. *Folha de Pernambuco Folha de Pernambuco.* https://www.folhape.com.br/noticias/noticias/cotidiano/2017/04/19/NWS,24738,70,449,NOTICIAS,2190-TECNOLOGIA-PARA-REGISTRAR-PRESENCA-NAS-ESCOLAS-JABOATAO.aspx

[4] https://www.youtube.com/watch?v=YRMihVhocew&t=24s

[5] https://en.wikipedia.org/wiki/Bolsa_Fam%C3%ADlia

[6] https://www.caixa.gov.br/programas-sociais/bolsa-familia/Paginas/default.aspx

[7] Lessig, L. (2006). *Code: Version 2.0.* New York: Basic Books. Monitoring is mostly related to observation in real time and tracking can be done afterwards based on certain information.

Given this potential, a number of critical issues in relation to the application of facial recognition systems in educational contexts were identified.

According to the service provider, the system works off a platform that uses cloud computing capabilities, but it was not possible to identify any information regarding the level of security related to the storage of collected data in the company's privacy policy available on its official website. Despite this, among the said benefits offered by the technology implemented are not only the monitoring of students' attendance and their school performance, but the possibility of monitoring students' personal health data.

The mayor of the city of Jaboatão dos Guararapes[8] states that in addition to facial recognition, the software also offers, through the collection of other information, the possibility of better planning the number of daily school meals. As soon as the school's gates are closed, the cooks receive via SMS[9] the exact number of students who are in the classrooms. Meanwhile, according to the secretary of education of Jaboatão dos Guararapes, Ivaneide Dantas, at some point even the health problems of students will be identified and parents informed using the system.

However, the lack of information that is included in the company's privacy policy,[10] or on city halls' websites, constitutes a critical point in the relationship between students and the education system. The problem becomes all the more obvious since the solution involves sensitive data – biometric data – of minors.

The text of the Brazilian General Data Protection Law (LGPD),[11] recently approved unanimously in congress after almost 10 years of discussions and two years of proceedings, has undergone major changes due to the vetoes of former president Michel Temer at the time of its sanction and more recently by President Jair Bolsonaro. The changes to the text of the LGPD through Provisional Measure 869/2018[12] have resulted in the impairment of the effectiveness of the law as well as a series of setbacks. These setbacks have ignored issues considered already decided on in discussions that had popular participation, as well as the input of members of the executive and legislative branches. As argued in a report released by the joint committee set up to investigate the possible impacts caused by the Provisional Measure,[13] the revised act put at risk the effectiveness of the guarantees reached by the previous text.

The public sector, especially the police authorities, are using new technologies to monitor the population without the social consequences being considered or even measured. The adoption of facial recognition systems for public security purposes is already a reality in several Brazilian cities. Recently, the increase in the use of the tool in public and private spheres has led to the establishment of Civil Public Inquiry No. 08190.052289/18-94[14] by the Personal Data Protection Commission of the Public Prosecutor's Office of the Federal District and Territories (MPDFT), as well as a public hearing held on 16 April 2019.[15] The hearing sought not only to promote debates about the use of facial recognition tools by businesses and the government, but also to function as an open space for the participation of NGOs and civil society.

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. As researchers have pointed out, biases are being increasingly incorporated into the variety of new technological tools.[16] These digital tools can act more quickly, on a larger scale, and with actions that can be hidden by a greater complexity, such as, for example, through profiling systems that use biased machine learning algorithms that police, profile and punish minorities. The struggle is still against the perpetuation of the same problem: the deprivation of civil rights of certain groups of society as a result of social inequalities and power relations in society.

It is not uncommon for technology to be used in a way in Brazil that suggests the possibility of a future Orwellian dystopia. The use of facial recognition technology during the Carnival of 2019 in the cities of Rio de Janeiro and Salvador, resulting in a number of arrests, drew the attention of

[8] Santos, N. (2017, 18 April). Jaboatão inicia reconhecimento facial nas escolas. *LeiaJá*. https://m.leiaja.com/carreiras/2017/04/18/jaboatao-inicia-reconhecimento-facial-nas-escolas

[9] https://www.youtube.com/watch?v=YRMihVhocew&t=24s

[10] www.pontoid.com.br/politica_privacidade_education.jsp

[11] IAPP. (2018). *Brazil's General Data Protection Law (English translation)*. https://iapp.org/resources/article/brazils-general-data-protection-law-english-translation

[12] Medida Provisória nº 869, de 27 de dezembro de 2018. www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/Mpv/mpv869.htm

[13] https://legis.senado.leg.br/sdleg-getter/documento?dm=7945369&ts=1556207205600&disposition=inline

[14] Inquérito Civil Público n.08190.052289/18-94 Reconhecimento Facial. www.mpdft.mp.br/portal/pdf/noticias/março_2019/Despacho_Audiencia_Publica_2.pdf

[15] https://www.youtube.com/watch?v=pmzvXcevJr4

[16] Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*. New York: St Martin's Press.

Brazilians.[17] The apparent relativism of fundamental rights and the relaxation of laws that limit surveillance due to the magnitude of events,[18] and the lack of official information on the effectiveness of the new automated measures, as well as on the storage and sharing of biometric data, were just some of the various problems identified.

The need to manage large events with thousands of participants is not the only reason why surveillance technologies are used in Brazilian urban centres. As Marcelo Souza from the Federal University of Rio de Janeiro (UFRJ) explains,[19] the increase in punitive policies and the militarisation of the police are the main reasons behind the increasing use of dystopian and invasive AI devices and technologies, typical of combat zones.[20] Even after a series of changes that have taken place in Brazil since the promulgation of the 1988 Constitution, public security institutions have not been significantly modified. The culture of war against the "internal enemy", for example, remains as present as in the days of the military dictatorship.[21]

Given that the recently approved LGPD does not apply to data processing for public security purposes, it would be possible for authorities to argue that the biometric database from the facial recognition system in schools can be used to identify suspects and improve public security. This would place its use even further outside the remit of current legislation. As the LGPD states, data processing for public security purposes "shall be governed by specific legislation, which shall provide for proportionate and strictly necessary measures to serve the public interest" (Article 4, III, § 1, LGPD).[22]

However, a specific law does not exist so far. According to the Brazilian lawyer and privacy advocate Rafael Zanatta,[23] the civic battle in Brazil will be around the shared definition of "proportional measures" and "public interest". The solution proposed by some researchers and activists for the time being is the protection offered by the constitution, such as the presumption of innocence, and the general principles of the LGPD itself, which guard against the improper use of collected data. On the other hand, it is believed that a tremendous effort will be needed to consolidate jurisprudence where these principles are applied in cases of state surveillance.

There is also a lack of easy access to public documents and information on the use of surveillance technologies. Often information related to the operation of these technologies depends on ad hoc statements made by public agents or private companies, whistleblowers or, when granted, requests for access to public information made possible by the Access to Information Law (LAI).

The discussion on the regulation of AI systems in Brazil is still very new. According to Brazilian researchers Bruno Bioni and Mariana Rielli,[24] the general debate on data protection around the globe has different degrees of maturity depending on the region and the specific issues faced. In addition, a paradigm shift has been observed through the transition from a focus on the individual's right to self-determination with regards to his or her private information to a model of risk prevention and management with respect to data-processing activities.

However, the use of AI for the collection and processing of sensitive personal data, such as the biometric data in question, makes these risks difficult to measure. In part this is because the general population does not have sufficient knowledge of the technology to recognise the extent of its impact on their personal lives.

In this way, an imbalance of power with regard to public awareness and the use of the technology has been created in Brazil.

The need for impact reports on the protection of personal data is an important requirement that has been gaining prominence in legislation, such as European legislation and the recently approved LGPD. However, Bioni and Rielli draw attention to the few requirements placed on developers of AI technologies in Brazil, as well as on the consumer who buys and implements the technology. In particular, there is no law in relation to the purchase and use of facial recognition devices for public service and public safety purposes in Brazil, unlike similar public projects elsewhere that seek an informed public debate and the inclusion of citizens' interests in

[17] Távora, F., Araújo, G., & Sousa, J. (2019, 11 March). Scanner facial abre alas e ninguém mais se perde no Carnaval (e fora dele). *Agência Data Labe*. https://tab.uol.com.br/noticias/redacao/2019/03/11/carnaval-abre-alas-para-o-escaner-facial-reconhece-milhoes-e-prende-seis.html

[18] Graham, S. (2011). *Cities Under Siege: The New Military Urbanism*. New York: Verso Books.

[19] Souza, M. (2008). *Fobópole*. Rio de Janeiro: Bertrand Brasil

[20] Kayyali, D. (2016, 13 June). The Olympics Are Turning Rio into a Military State. *Vice*. https://www.vice.com/en_us/article/wnxgpw/the-olympics-are-turning-rio-into-a-military-state

[21] Machado, R. (2019, 19 February). Militarização no Brasil: a perpetuação da guerra ao inimigo interno. Entrevista especial com Maria Alice Rezende de Carvalho. *Instituto Humanitas Unisinos*. www.ihu.unisinos.br/159-noticias/entrevistas/586763-militarizacao-no-brasil-a-perpetuacao-da-guerra-ao-inimigo-interno-entrevista-especial-com-maria-alice-rezende-de-carvalho

[22] IAPP. (2018). Op. cit.

[23] https://twitter.com/rafa_zanatta/status/1085583399186767875

[24] Bioni, B., & Rielli, M. (2019). *Audiência Pública: uso de ferramentas de reconhecimento facial por parte de empresas e governos*. Data Privacy Brasil. https://dataprivacy.com.br/wp-content/uploads/2019/04/Contribui%C3%A7%C3%A3o-AP-reconhecimento-facial-final.pdf

decision-making processes (e.g. the ordinance on the acquisition of surveillance technology recently adopted in San Francisco, in the United States).[25]

Conclusion

A decrease in school drop-out rates: this is the main advantage of facial recognition technology in schools, according to the company that has developed the technology. But are we assigning to technology something that would be the responsibility of society and the state?

As the Brazilian educator and philosopher Paulo Freire[26] showed, many of the most common practices in the Brazilian educational system are dictated by the Brazilian elite. This includes the use of educational content that does not correspond to the reality of students from lower classes, but instead inhibits their ability to think critically of their reality, and in the end discourages them from attending school.

Easy and safe access to school is also an important consideration impacting on the student's educational performance. The Brazilian Statute of the Child and Adolescent, in chapter IV, art. 53, inc. V,[27] states that one of the rights of the child and adolescent is access to public and free schools near his or her home. However, when distance is not an impediment to school attendance, issues related to the safety of the student's route to school should also be considered. For example, in some communities in Rio de Janeiro,[28] there are frequent incidents of armed confrontation between police and drug traffickers, visible abuses of power by the authorities, stray bullets and other incidents endangering the lives of passers-by, and even street executions, all of which are daily threats to residents. In addition, the reality faced by marginalised populations in Brazil raises another important question: the need for children and adolescents from low-income families to leave school in order to work and help support the household.

The problem of the school drop-out rate in Brazil is neither a new issue nor something that can be solved only with the implementation of a new school attendance system using AI. It is necessary for society to seek more meaningful ways to mitigate the present crisis facing the educational system.

In addition to projects that raise public awareness about issues related to new technologies in Brazilian urban centres, there is also a need to strengthen legislation that governs how sensitive data is shared and used in public projects in order to maintain the quality of public services. When the fundamental rights and guarantees of citizens are understood and respected, a relationship of trust is established.

Cities in Brazil should strengthen privacy protection instruments and create legal certainty through the establishment of fundamental principles, rights and duties for the operation of an ever-changing array of technologies in society. Regulating the use of personal data in the public sphere has the power to promote the conscious, transparent and legitimate use of this information.

Action steps

The following advocacy priorities are suggested for Brazil:

- Formulate regional policies and strategies for the use of AI in Latin America and the Caribbean.
- Develop legally enforceable safeguards, including robust transparency and accountability measures, before any facial recognition technology is deployed.
- Promote national campaigns and public debate over surveillance technology given the impact such technologies may have on civil rights and civil liberties.
- Include a social, cultural and political understanding of the needs of vulnerable groups in the country's education strategies to make the learning environment more attractive for these groups.
- Develop open source AI systems that enable wider community use, with appropriate privacy protections.

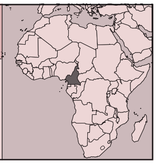
[25] Johnson, K. (2019, 14 May). San Francisco supervisors vote to ban facial recognition software. *VentureBeat*. https://venturebeat.com/2019/05/14/san-francisco-first-in-nation-to-ban-facial-recognition-software

[26] Freire, P. (1976). *Education: The Practice of Freedom*. London: Writers and Readers Publishing Cooperative.

[27] www.planalto.gov.br/ccivil_03/leis/l8069.htm

[28] Brito, R. (2017, 2 October). Rio's kids are dying in the crossfire of a wave of violence. *AP News*. https://www.apnews.com/efeeaed43c7b47a0ae4a6cfaa8b871e2

CAMEROON



AT A SNAIL'S PACE: THE INTRODUCTION OF ARTIFICIAL INTELLIGENCE IN HEALTH CARE IN CAMEROON

PROTEGE QV
Serge Daho and Emmanuel Bikobo
https://www.protegeqv.org

Introduction

Cameroon is a country located in west-central Africa, and is bordered by six countries, Nigeria, Chad, Gabon, the Central African Republic, the Republic of Congo and Equatorial Guinea. It has 25,869,160 inhabitants,[1] the bulk of whom still live under the poverty line, with uneven access to health care services between rural and urban areas.[2]

While artificial intelligence (AI) offers the potential to enhance health care throughout the country and to overcome an acute shortage of medical personnel in rural areas, Cameroon has been slow in its adoption.

This report discusses some of the challenges to the use of AI at the Bonassama district hospital, a governmental hospital in Douala, one of two health care facilities in the country that are experimenting with AI in health care.[3]

Policy and political background

Unlike other fundamental rights such as those of freedom of expression, communication, or the right to security, the Cameroonian constitution does not provide for the right to health as a fundamental right. The health sector in our country is also rather poor when it comes to both legislation and regulation. However, several laws, strategies and plans are worth mentioning:

- Law No. 96/03 of 4 January 1996, which establishes the general framework for action by the state in the field of health, in particular through the national health policy which focuses on universal access to quality basic health care services through the development of district hospitals, the promotion and protection of the health of vulnerable and underserved populations, and a needs-based priority agenda for health regarding the fight against great pandemics such as HIV.
- Decree No. 2013/093 of 3 April 2013, on the organisation of the Ministry of Public Health. The Ministry is responsible for the formulation and implementation of the government's public health policy.
- The National Health Development Plan (NHDP). Headed by the Ministry of Public Health, this plan was developed in 2015 by a national inter-sectoral technical committee. It is supposed to provide the country with a universal health care coverage scheme.[4]
- The Health Sector Strategy 2016-2027.[5] The goal of this document is to protect and improve on health care for Cameroonian citizens, based on measurable targets.

Marked by insufficient funding[6] by the state, the public health sector of the country has a centralised system of administration running from the Ministry, through the intermediary regional levels, and cumulating at the district levels.[7] Three different levels of

[1] www.worldpopulationreview.com

[2] Dr Tetanye Ekwe, the vice-chairman of Cameroon's National Order of Doctors made it clear during an interview with the Voice of America: the doctor-patient ratio stands at one doctor per 50,000 inhabitants in rural areas, instead of the one doctor per 10,000 inhabitants recommended by the World Health Organization (WHO). In the country's two major cities, Douala and Yaounde, the doctor-patient ratio does meet the WHO standard. Kindzeka, M. E. (2018, 25 November). Cameroon Doctors Overwhelmed with Patients. *Voice of America*. https://www.voanews.com/africa/cameroon-doctors-overwhelmed-patients

[3] The other hospital using AI in health care is the HSPC, a private facility located in Kumba. It uses a software application that is meant to help health care professionals at the hospital get a more complete picture of the health of their community as it pertains to non-communicable diseases. The laboratory at the Pasteur Centre in Yaounde also uses a mobile application based on AI to diagnose malaria.

[4] The launch of the first phase of the Universal Health Coverage (UHC) scheme was announced in Yaounde on 5 June 2018.

[5] The Health Sector Strategy (HSS) in Cameroon is a document serving as the roadmap for the Ministry of Public Health for the years 2016 to 2027. The HSS focuses on almost the whole health sector of the country, including the organisation of the sector, and deals with communicable and non-communicable diseases, maternal, neonatal and infant health, and health finance. Ministry of Public Health. (2016). *Health Sector Strategy 2016-2027*. https://www.minsante.cm/site/sites/default/files/HSS_english_0.pdf

[6] The proportion of the national budget allocated to the Ministry of Public Health varies between 5% and 5.5%, far from the Abuja Declaration prescription. In April 2001, the African Union countries met in Abuja and pledged to set a target of allocating at least 15% of their budget to improve the health sector and urged donor countries to scale up support.

[7] Each of the country's ten regions has at least one regional hospital and 189 district hospitals are registered in Cameroon.

health care delivery also exist in Cameroon: tertiary, secondary and primary services.[8]

According to the Health Sector Strategy (HSS), public health facilities are more accessible to the rich.[9] As a result, in this poverty-stricken country, universal health coverage is more than welcome and will introduce equity in the country's health system.

Challenges of AI in health care in Cameroon

The district hospital in Bonassama[10] is a public health facility located in the city of Douala. It made a significant step in leveraging AI to improve analyses of patients and to provide them with more precise diagnoses in a shorter turn-around time. It is worth pointing out that the HSS, which is the Ministry of Public Health's roadmap for 12 years, does not even mention AI as part of its concerns. At most, a vague reference is made to "[a] more complex specialized care approach", linked to the implementation of the head of state's Triennial Emergency Plan 2015-2018.[11]

Far away from the uncertainty surrounding the government plans about AI, Sophia Genetics, based in Lausanne (Switzerland) and in Boston (United States), announced on 24 March 2017 the list of the seven African hospitals, including the district hospital of Bonassama, that will begin using Sophia. Sophia is an AI-based platform that helps hospitals to diagnose patients better and faster in five areas: oncology, metabolism, paediatrics, cardiology and hereditary cancers. It uses a technology that enables predictive analysis by feeding and training AI algorithms intended to enhance the reading and analysis of DNA sequencing. This software uses statistical inference pattern recognition and machine learning to analyse both genomics and radiomics (medical imagery) data.[12] Specific conditions and tests covered by the agreement[13] with Sophia Genetics include BRCA1 and BRCA2,[14] HNPCC,[15] clinical exome sequencing[16] and HCS.[17]

The use of AI is expected to decrease medical costs as there will be more accuracy in diagnoses, better treatment plans, as well as more prevention of disease. However, several challenges with the Sophia system have been identified.

Although AI technologies are hailed for their innovative, infinite applications in medical care and in medical research, their real-life implementation is still facing obstacles. In Cameroon, the first hurdle involves regulations. As mentioned, current regulations governing the health sector are poor, a fact also outlined in the HSS,[18] which decries the absence of a public health code, which would establish a set of standards for health care, among other legal gaps.

Another major impediment in Cameroon concerns the country's poverty levels. According to the HSS, in 2010, 70% of the population was in a situation of global underemployment – that is, they involuntarily worked less than the minimum working week of 35 hours, or earned less than hourly minimum wage. Yet, citizens bear the brunt of the financial burden for health care here. Because of this, the more money you have, the more likely you are to have qualified professional assistance. AI in health care in our country is likely to perpetuate or even accentuate this divide between the rich and the poor, raising the issue of social justice in terms of the distribution of wealth, opportunities and privileges within our society.

During our meeting with Dr. Esther Dina Bell, the head manager of the Bonassama district hospital, she disclosed that 45% of the hospital patients were not wage earners, highlighting the relatively limited number of patients who have access to AI for their health care issues. While the average monthly wage in Cameroon is USD 173, Sophia services cost

[8] The difference lies with the technical facilities offered at each level, coupled with the distribution of the health personnel at each of the three levels. Generally, tertiary level services are delivered in the country's two big cities, Douala and Yaounde. The best medical doctors are also often found in these cities.

[9] The share of the richest quintile that consulted a public medical doctor was close to 43% in 2007, and about 3% for the poorest. Ministry of Public Health. (2016). Op. cit.

[10] The Bonassama district hospital is one of the Cameroon's 189 district hospitals and as such is ranked at the bottom level of the health care system in the country.

[11] The Triennial Emergency Plan was meant to accelerate the country's economic growth. The plan (2015-2018) covered the entire national territory in the following sectors: health, urban development, animal industries, water, energy, roads, agriculture, regional development and security.

[12] https://en.wikipedia.org/wiki/Sophia_Genetics

[13] This is an agreement solely between the Bonassama district hospital and Sophia Genetics. The state is not in any way involved.

[14] The name BRCA is an abbreviation for "BReast CAncer Gene". BRCA1 and BRCA2 are two different genes that have been found to impact a person's chances of developing breast cancer. https://www.nationalbreastcancer.org/what-is-brca

[15] HNPCC stands for hereditary nonpolyposis colorectal cancer. It is a condition in which the tendency to develop colorectal cancer is inherited. https://www.hopkinsmedicine.org/gastroenterology_hepatology/diseases_conditions/small_large_intestine/hereditary_nonpolyposis_colorectal_cancer.html

[16] Clinical exome sequencing is a highly complex molecular test that analyses the exons (or coding regions) of thousands of genes from a small sample of blood, by next generation sequencing techniques. The purpose of this test is to identify the underlying molecular cause of a genetic disorder in an affected individual. https://geneticscenter.com/test-menu/exome-sequencing

[17] HCS stands for high content screening. It is a set of analytical methods using automated microscopy, multi-parameter image processing and visualisation tools to extract quantitative data from cell populations. https://www.thermofisher.com/uy/en/home/life-science/cell-analysis/cellular-imaging/high-content-screening.html

[18] Ministry of Public Health. (2016). Op. cit.

between USD 499 and USD 1,288 for patients at the Bonassama hospital. This confirms the early warnings issued by Margaret Chan,[19] the World Health Organization (WHO) Director-General, who spoke of the need to ensure that medical AI applications work for the both poorest and the richest.

Concerning the attitudes of nurses working at the Bonassama district hospital toward the adoption of AI in their workplace, most of the nurses we spoke to are not even aware of its existence. And if AI is a threat to their job security, they have little to fear. The slow pace of the introduction of AI in hospitals in Cameroon, and the relatively high cost of services, means that nurses are unlikely to be replaced by machines in the foreseeable future.

One pressing issue that requires more investigation is the issue of privacy.[20] While patients' data is collected by the Bonassama hospital and transferred to Sophia Genetics using a secured platform, we could not determine how long this data is stored by Sophia Genetics – we believe at least a year or two. The analysis of data using AI systems may reveal private information about individuals, and should be treated as sensitive. 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). It was, for us, also impossible to find out if patients have access to their personal data or the right to amend or delete any information they have previously provided.

Conclusion

AI is gradually gaining ground and even superseding human ability in accurately diagnosing diseases, and many countries are leveraging this potential to improve their health care systems. Nevertheless, in Cameroon, the adoption of new technologies in health care is surprisingly slow – a situation that is not only a question of finance, but also of a lack of awareness (for example, AI does not feature in the health ministry's roadmap for 2016-2027). This does not only relate to illnesses such as heart disease – cardiovascular diseases were the second cause of mortality in Cameroon in 2013[21] – or breast cancer. As a tropical nation, Cameroon is afflicted by common diseases such as malaria that countries like Uganda[22] are tackling using AI.

As we learned at the Bonassama district hospital, the cost for accessing AI-driven health services is very high for the average Cameroonian citizen. It is also disturbing to notice that the Ministry of Public Health does not even mention AI in its 12-year strategic plan. In a country classified by the WHO as having a critical shortage of health personnel,[23] mainly in rural areas, the country could have taken advantage of AI in these understaffed and under-resourced areas to enhance health care, given the ability of AI to be trained to recognise and diagnose certain medical conditions in the absence of doctors.

Action steps

Since any major technological innovation brings potential to advance or damage society, we suggest the following recommendations:

- Cameroon should increase the health budget so that it meets the Abuja Declaration requirement of 15% being allocated yearly to the health sector.
- The government should subsidise citizens' access to AI, and fund its use nationwide.
- Citizens should be informed of the existence of AI and its capacities to perform early diagnosis of various health conditions.
- Citizens should also be made aware of their right to give informed consent for their personal data to be gathered for use in AI services.
- The private sector should be encouraged to create affordable AI for the health care sector, with a specific focus on cancers,[24] cardiovascular diseases,[25] as well as malaria,[26] which is a widespread disease in Cameroon.

[19] Margaret Chan was speaking at the AI for Good Global Summit, 2017, in Geneva. AI for Good is the leading UN platform for global and inclusive dialogue on AI. https://www.who.int/dg/speeches/2017/artificial-intelligence-summit/en

[20] In Cameroon, privacy violations are crimes prohibited and punishable under articles 43 and 44 of the December 2010 Law on Cybersecurity.

[21] Ministry of Public Health. (2016). Op. cit.

[22] Uganda's first AI lab, at Makerere University, has developed a way to diagnose blood samples for diseases like malaria using a cell phone. Lewton, T., & McCool, A. (2018, 14 December). This app tells your doctor if you have malaria. *CNN*. https://edition.cnn.com/2018/12/14/health/ugandas-first-ai-lab-develops-malaria-detection-app-intl/index.html

[23] Tandi, T. E., et al. (2015). Cameroon public health sector: shortage and inequalities in geographical distribution of health personnel. *International Journal for Equity in Health, 14*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4440287

[24] In 2012, 14,000 new cases of cancer were diagnosed and about 25,000 persons lived with cancer in Cameroon. More than 80% of persons affected were tested at a very late stage of the disease and the majority died within 12 months after they had been diagnosed. Ministry of Public Health. (2016). Op. cit.

[25] There are still many inadequacies in the capacity of the health system to effectively control cardiovascular diseases (low-skilled human resources and insufficient equipment for quality management of the cases). Ibid.

[26] Generally, out of 19,727 deaths recorded in health facilities in 2013, 22.4% were related to malaria. Ibid.

CANADA

FEMINIST OR NOT? CANADA'S CHALLENGES AS IT RACES TO BECOME A LEADER IN ARTIFICIAL INTELLIGENCE

Institute for Gender and the Economy, Rotman School of Management, University of Toronto; International Human Rights Program, Faculty of Law, University of Toronto; The Citizen Lab, Munk School of Global Affairs, University of Toronto
Victoria Heath, Petra Molnar and Irene Poetranto
https://www.gendereconomy.org; https://ihrp.law.utoronto.ca; https://citizenlab.ca

Introduction

The Government of Canada, led by Prime Minister Justin Trudeau, declared itself a "feminist government" in 2015,[1] pursuing policies and funding initiatives intended to promote gender equality and protect human rights like the Feminist International Assistance Policy[2] and a CAD 160-million investment into women's and indigenous organisations.[3] Canada also committed itself to being a key player in artificial intelligence (AI), publishing the first national strategy for AI in 2017,[4] which was backed by CAD 125 million in federal funding.[5]

The government is clearly committed to the development of the AI field, and because of this AI's use in the public sector is expected to increase. Canada has been using automated decision-making experiments in its immigration mechanisms since 2014, while the Toronto Police Service has been using facial recognition technology since 2018. Yet the negative implications of AI-backed systems and tools are only beginning to surface, with research showing that the use of AI in government decision making can violate human rights and entrench social biases, especially for women.[6] This report discusses the challenges that must be confronted to ensure that the implementation of AI, especially in immigration and policing, does not contradict Canada's commitment to human rights and gender equality.

Context

Following the civil rights movement in the 1960s, the "rights culture" in Canada "evolved from simply prohibiting overt acts of discrimination to ensuring substantive equality."[7] This led previous Canadian governments to implement a range of policies and initiatives, including the establishment of the Status of Women Canada[8] (1976) and the Human Rights Commission[9] (1977), as well as the passage of the Human Rights Act[10] (1977) and the Charter of Human Rights and Freedoms[11] (1982). Canada has also ratified international human rights treaties, including the Convention on the Elimination of all Forms of Discrimination against Women[12] (CEDAW), the International Covenant on Civil and Political Rights[13] (ICCPR), and the Universal Declaration of Human Rights[14] (UDHR), among others. However, the government has failed to protect the rights of historically marginalised groups, particularly indigenous women, and struggles with eliminating systemic inequality and discrimination of minority groups.[15]

[1] www.canadabeyond150.ca/reports/feminist-government.html

[2] Global Affairs Canada. (2017). *Canada's Feminist International Assistance Policy*. https://www.international.gc.ca/world-monde/issues_development-enjeux_developpement/priorities-priorites/policy-politique.aspx?lang=eng

[3] Women and Gender Equality Canada. (2019, 12 June). Government of Canada announces investment in women's organizations in Ottawa. https://www.canada.ca/en/status-women/news/2019/06/government-of-canada-announces-investment-in-womens-organizations-in-ottawao.html

[4] Natural Sciences Sector. (2018, 22 November). Canada first to adopt strategy for artificial intelligence. *United Nations Educational, Scientific and Cultural Organization (UNESCO)*. www.unesco.org/new/en/media-services/single-view/news/canada_first_to_adopt_strategy_for_artificial_intelligence

[5] Canadian Institute for Advanced Research. (2017, 22 March). Canada funds $125 million Pan-Canadian Artificial Intelligence Strategy. *Cision*. https://www.newswire.ca/news-releases/canada-funds-125-million-pan-canadian-artificial-intelligence-strategy-616876434.html

[6] Molnar, P., & Gill, L. (2018). *Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System*. University of Toronto. https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf

[7] Clément, D., Silver, W., & Trottier, D. (2012). *The Evolution of Human Rights in Canada*. Canadian Human Rights Commission. https://www.chrc-ccdp.gc.ca/eng/content/evolution-human-rights-canada

[8] https://cfc-swc.gc.ca/abu-ans/who-qui/index-en.html

[9] https://www.chrc-ccdp.gc.ca/eng

[10] https://www.chrc-ccdp.gc.ca/eng/content/human-rights-in-canada

[11] Ibid.

[12] https://www.ohchr.org/en/hrbodies/cedaw/pages/cedawindex.aspx

[13] https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx

[14] https://www.ohchr.org/en/udhr/documents/udhr_translations/eng.pdf

[15] National Inquiry into Missing and Murdered Indigenous Women and Girls. (2019). *Reclaiming Power and Place: The Final Report of the National Inquiry into Missing and Murdered Indigenous Women and Girls*. https://www.mmiwg-ffada.ca/final-report

The government has actively supported the development and deployment of AI, despite the risk that technologies may exacerbate pre-existing disparities and can lead to rights violations. For example, it invested CAD 950 million in the Innovation Supercluster Initiative in early 2018,[16] which included the SCALE.AI Supercluster. It also launched a fast-track visa programme for tech talent in 2019,[17] and established the Advisory Council on Artificial Intelligence, composed of researchers, academics and business executives, to build on Canada's AI strengths and identify opportunities.[18] Canada is also engaged in AI efforts globally. It is a party to the G7's Charlevoix Common Vision for the Future of Artificial Intelligence, which highlights the importance of supporting gender equality and preventing human rights abuses by involving "women, underrepresented populations and marginalized individuals" at all stages of AI applications.[19] While these steps, along with investments from companies such as Uber, Google and Microsoft, indicate Canada's growing leadership in AI, the government has been criticised for not doing enough to address the gender diversity problem.[20]

Women account for only 25% of all technology jobs in Canada and about 28% of all scientific occupations, despite representing half of the workforce.[21] Women also continue to have poor representation among the executive teams and boards of Canadian technology companies.[22] Canada's Minister of Science Kirsty Duncan has said that getting more women into high-ranking scientific positions is a priority,[23] but more work needs to be done to ensure gender diversity in science and technology, which will have a tremendous effect on how they impact human rights and equality.

Human rights challenges in the age of AI: A look at immigration and policing

The lack of diversity in science and technology is exceptionally pronounced in the AI field, which is predominantly white and male. For instance, more than 80% of AI professors are men.[24] The gender pay gap also compounds the diversity problem. Women working in the Canadian tech industry with a bachelor's degree or higher, typically earn nearly CAD 20,000 less a year than their male counterparts, and this pay gap is higher for visible minorities. Black tech workers in particular not only have the lowest participation rates in tech occupations, but also experience a significant pay gap relative to white and non-indigenous tech workers in Canada.[25]

A study by the AI Now Institute found that lack of diversity results in the creation of AI systems and tools with built-in biases and power imbalances. It is consistent with the feminist critique of technology, which posits that existing social relations and power dynamics are manifested in and perpetuated by technology.[26] This is particularly true of gender relations, which are "materialised in technology, and masculinity and femininity in turn acquire their meaning and character through their enrolment and embeddedness in working machines."[27] An example of this is AI-powered virtual assistants, such as Apple's Siri, which are predominantly modelled on feminine likeness and stereotypical characteristics.[28] These products have been criticised as products of sexism, due to the fact that they replicate stereotypes of what are considered to be "women's work" (e.g.

[16] Innovation, Science and Economic Development Canada. (2018, 15 February). Canada's new superclusters. *Government of Canada*. www.ic.gc.ca/eic/site/093.nsf/eng/00008.html

[17] Employment and Social Development Canada. (2017, 6 December). Global Skills Strategy. *Government of Canada*. https://www.canada.ca/en/employment-social-development/campaigns/global-skills-strategy.html

[18] Innovation, Science and Economic Development Canada. (2019, 14 May). Government of Canada creates Advisory Council on Artificial Intelligence. *Cision*. https://www.newswire.ca/news-releases/government-of-canada-creates-advisory-council-on-artificial-intelligence-838598005.html

[19] https://www.international.gc.ca/world-monde/international_relations-relations_internationales/g7/documents/2018-06-09-artificial-intelligence-artificielle.aspx?lang=eng

[20] PwC Canada, #movethedial, & MaRS. (2017). *Where's the Dial Now? Benchmark Report 2017*. https://www.pwc.com/ca/en/industries/technology/where-is-the-dial-now.html

[21] Information and Communications Technology Council. (2018). *Quarterly Monitor of Canada's ICT Labour Market*. https://www.ictc-ctic.ca/wp-content/uploads/2019/01/ICTC_Quarterly-Monitor_2018_Q2_English_.pdf

[22] Ontario Centres of Excellence. (2017, 15 November). Bridging the female leadership gap in Canada's technology sector — The Word on the Street in the World of Innovation. *Medium*. https://blog.oce-ontario.org/bridging-the-female-leadership-gap-in-canadas-technology-sector-9923bf90babf

[23] Mortillaro, N. (2018, 8 March). Women encouraged to pursue STEM careers, but many not staying. *CBC News*. https://www.cbc.ca/news/technology/women-in-stem-1.4564384

[24] Paul, K. (2019, 17 April). 'Disastrous' lack of diversity in AI industry perpetuates bias, study finds. *The Guardian*. https://www.theguardian.com/technology/2019/apr/16/artificial-intelligence-lack-diversity-new-york-university-study

[25] Lamb, G., Vu, V., & Zafar, A. (2019). *Who Are Canada's Tech Workers?* Brookfield Institute for Innovation and Entrepreneurship. https://brookfieldinstitute.ca/wp-content/uploads/FINAL-Tech-Workers-ONLINE.pdf

[26] Wajcman, J. (2010). Feminist theories of technology. *Cambridge Journal of Economics, 34*(1), 148-150.

[27] Ibid.

[28] Sternberg, I. (2018, 8 October). Female AI: The Intersection Between Gender and Contemporary Artificial Intelligence. *Hackernoon*. https://hackernoon.com/female-ai-the-intersection-between-gender-and-contemporary-artificial-intelligence-6e098d10ea77

service-oriented roles) and behaviour (e.g. servile, helpful, submissive, etc.).[29]

Even if diversifying the technology field leads us to build AI systems and tools that are more neutral, the data that are fed into them might still contain bias. For example, research shows that deploying facial recognition algorithms using mug shot photos resulted in racial bias, as it derived an incorrect relationship between skin colour and the rate of incarceration.[30] These algorithms are also more likely to fail in correctly identifying dark-skinned women than light-skinned women.[31] Due to indications that AI systems may replicate patterns of racial and gender bias, and solidify and/or justify historical inequalities, the deployment of these tools should be concerning. This is particularly the case for any government, like Canada's, that has made commitments to protecting human rights and achieving gender equality.

AI's deployment in Canada is well underway. Canada has been using automated decision-making experiments in its immigration mechanisms since 2014.[32] The federal government has more recently been in the process of developing a system of "predictive analytics" to automate certain activities conducted by immigration officials, and to support the evaluation of some immigrant and visitor applications. Immigration, Refugees and Citizenship Canada (IRCC) confirmed that the federal department launched two pilot projects in 2018 using algorithms to identify routine Temporary Resident Visa applications from China and India for faster processing. The Canadian Border Services Agency (CBSA) has also implemented automated passport processing that relies on facial recognition software, in lieu of initial screenings performed by a CBSA officer.[33] In May 2019, an investigation revealed that the Toronto Police Service (TPS) has been using facial recognition technology that compares images of potential suspects captured on public or private cameras to approximately 1.5 million mug shots in TPS's internal database. Privacy advocates have been critical of TPS's use of facial recognition technology due to concerns with potential discrimination, as well as infringements of privacy and civil liberties.[34]

Canada is not alone in using AI technologies for migration management. Evidence shows that national governments and intergovernmental organisations (IGOs) are turning to AI to manage complex migration crises. For example, big data is being used by the United Nations High Commissioner for Refugees (UNHCR) to predict population movement in the Mediterranean,[35] while retinal scanning is being used in Jordanian refugee camps for identification.[36] At the US-Mexico border, a tweak to the Immigration and Customs Enforcement's (ICE) "risk assessment" software has led to a stark increase in detentions, which indicates that decision making in immigration is also increasingly relying on technology.[37] A 2018 report by the University of Toronto's Citizen Lab and International Human Rights Program shows that despite the risks associated with utilising AI technology for decision making in immigration and policing, these experiments often have little oversight and accountability, and because of this could lead to human rights violations.[38]

Research from as early as 2013 shows that algorithms may lead to discriminatory and biased results. The outcome from a Google search, for example, may yield discriminatory ads targeted on the basis of racially associated personal names, display lower-paying job opportunities to women,[39] and perpetuate stereotypes based on appearance (e.g. by associating "woman" with "kitchen").[40] Claims have also been made that facial-detection technol-

[29] Chambers, A. (2018, 13 August). There's a reason Siri, Alexa and AI are imagined as female – sexism. *The Conversation*. https://theconversation.com/theres-a-reason-siri-alexa-and-ai-are-imagined-as-female-sexism-96430

[30] Garvie, C., & Frankle, J. (2016, 7 April). Facial-Recognition Software Might Have a Racial Bias Problem. *The Atlantic*. https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991

[31] Meyers West, S., Whittaker, M., & Crawford, K. (2019). *Discriminating Systems: Gender, Race, and Power in AI*. AI Now Institute. https://ainowinstitute.org/discriminatingsystems.pdf

[32] Keung, N. (2017, 5 January). Canadian immigration applications could soon be assessed by computers. *The Toronto Star*. https://www.thestar.com/news/immigration/2017/01/05/immigration-applications-could-soon-be-assessed-by-computers.html

[33] Dyer, E. (2019, 24 April). Bias at the border? CBSA study finds travellers from some countries face more delays. *CBC News*. https://www.cbc.ca/news/politics/cbsa-screening-discrimination-passports-1.5104385

[34] Lee-Shanok, P. (2019, 30 May). Privacy advocates sound warning on Toronto police use of facial recognition technology. *CBC News*. https://www.cbc.ca/news/canada/toronto/privacy-civil-rights-concern-about-toronto-police-use-of-facial-recognition-1.5156581

[35] Petronzio, M. (2018, 24 April). How the U.N. Refugee Agency will use big data to find smarter solutions. *Mashable*. https://mashable.com/2018/04/24/big-data-refugees/#DNN5.AOwfiqQ

[36] Staton, B. (2016, 18 May). Eye spy: biometric aid system trials in Jordan. *The New Humanitarian*. www.thenewhumanitarian.org/analysis/2016/05/18/eye-spy-biometric-aid-system-trials-jordan

[37] Oberhaus, D. (2018, 26 June). ICE Modified Its 'Risk Assessment' Software So It Automatically Recommends Detention. *Motherboard, Tech by VICE*. https://www.vice.com/en_us/article/evk3kw/ice-modified-its-risk-assessment-software-so-it-automatically-recommends-detention

[38] Molnar, P., & Gill, L. (2018). Op. cit.

[39] Sweeny, L. (2013). *Discrimination in Online Ad Delivery*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2208240

[40] Simonite, T. (2017, 21 August). Machines taught by photos learn a sexist view of women. *WIRED*. https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women

ogy has the ability to discern sexual orientation.[41] If these biases are embedded in emerging technologies used experimentally in immigration, then they could have far-reaching impacts. In airports across Hungary, Latvia and Greece, a new pilot project called *iBorderCtrl* introduced an AI-powered lie detector at border checkpoints. Passengers' faces are monitored for signs of lying and if the system becomes more "sceptical" through a series of increasingly complicated questions, the person will be selected for further screening by a human officer.[42] Canada has tested a similar lie detector, relying on biometric markers such as eye movements or changes in voice, posture and facial gestures as indicators of untruthfulness.[43] The deployment of such a technology raises a number of important questions: When a refugee claimant interacts with these systems, can an automated decision-making system account for their trauma and its effect on memory or for cultural, age and gender differences in communication? How would a person challenge a decision made by AI-powered lie detectors? These questions are important to consider given that negative inferences due to AI will impact the final decision made by a human official.

Migrants who identify as women or gender non-binary also confront challenges that are not yet adequately understood by AI technologies. For example, women and children generally have a different expectation of privacy than adult men, given the risks to their personal safety if their data is shared with repressive governments. It is clear that the use of AI-backed technologies in immigration results in a number of concerns, such as infringing on domestically and internationally protected human rights, including freedom of movement, freedom from discrimination, as well as the individual right to privacy, life, liberty and security.[44] Other concerns include an inappropriate reliance on the private sector to develop and deploy these technologies; an unequal distribution of technological innovation, exacerbating the lack of access to the justice system for marginalised communities; and an overall lack of transparency and oversight mechanisms. Finally, affected migrant communities, such as refugees and asylum seekers, are not sufficiently included in conversations around the aforementioned development and adoption of AI technology.[45]

Women, in particular, fear the prevalence of AI technology because it can be weaponised against them. "Smart home" technologies that automate various facets of household management (e.g. appliances, temperature, home security, etc.), for example, have enabled gender-based violence (GBV) and domestic abuse. In more than 30 interviews with *The New York Times*, domestic abuse survivors told stories of how abusers would remotely control everyday objects in their homes, sometimes to just watch or listen, but other times to demonstrate power over them.[46] As technologies become affordable and internet connectivity becomes more ubiquitous, technology-facilitated violence or abuse is likely to continue. In Canada, immigrant women have claimed refugee status or asylum due to GBV, while indigenous women, many of whom live in large urban population centres,[47] experience violence at a rate six times higher than non-indigenous women.[48]

Conclusion

Technology is neither inherently neutral nor democratic, but is a product of existing social relations and power dynamics, and because of this it can also perpetuate them.[49] Therefore, when technological experiments are introduced into the provision of public services, border management, or the criminal justice system, they could potentially exacerbate social divisions, strengthen unequal power relations, and result in far-reaching rights infringements. These concerns are particularly acute if diverse representation and human rights impact analyses are missing from AI's deployment. Without oversight to ensure diversity and proper impact assessments, the benefits of new technologies like AI may not accrue equally.

[41] Murphy, H. (2017, 9 October). Why Stanford Researchers Tried to Create a 'Gaydar' Machine. *The New York Times*. https://www.nytimes.com/2017/10/09/science/stanford-sexual-orientation-study.html

[42] Picheta, R. (2018, 2 November). Passengers to face AI lie detector tests at EU airports. *CNN*. https://edition.cnn.com/travel/article/ai-lie-detector-eu-airports-scli-intl/index.html

[43] Daniels, J. (2018, 15 May). Lie-detecting computer kiosks equipped with artificial intelligence look like the future of border security. *CNBC*. https://www.cnbc.com/2018/05/15/lie-detectors-with-artificial-intelligence-are-future-of-border-security.html

[44] Meyers West, S., Whittaker, M., & Crawford, K. (2019). Op. cit.

[45] Molnar, P. (2018, 14 December). The Contested Technologies That Manage Migration. *Centre for International Governance Innovation*. https://www.cigionline.org/articles/contested-technologies-manage-migration

[46] Bowles, N. (2018, 23 June). Thermostats, Locks and Lights: Digital Tools of Domestic Abuse. *The New York Times*. https://www.nytimes.com/2018/06/23/technology/smart-home-devices-domestic-abuse.html

[47] Arriagada, P. (2016, 23 February). First Nations, Métis and Inuit Women. *Statistics Canada*. https://www150.statcan.gc.ca/n1/pub/89-503-x/2015001/article/14313-eng.htm

[48] https://www.canadianwomen.org/the-facts/gender-based-violence

[49] Wajcman, J. (2010). Op. cit.

Canada must ensure that its use of AI-backed technologies is in accordance with its domestic and international human rights obligations, which is especially important given that Canada aims to be one of the world's leaders in AI development. As such, Canada's decision to implement particular technologies, whether they are developed in the private or the public sector, can set a new standard for other countries to follow. The concern for many human rights advocates and researchers is that AI-backed technologies will be adopted by countries with poor human rights track records and weak rule of law, who would be more willing to experiment and to infringe on the rights of vulnerable groups. For example, China's mandatory social credit system – which aims to rank all of its citizens according to their behaviour by 2020 – punishes individuals by "blacklisting" them, and therefore creates "second-class citizens".[50]

If Canada intends to be a leader in AI innovation, while maintaining its commitment to human rights and advancing gender equality as a "feminist government", then it must confront the challenges associated with the development and implementation of AI. Some of the challenges outlined include the gender and racial imbalance in science, technology, engineering and mathematics (STEM) education and employment, as well as the lack of accountability and transparency in the government's use of emerging technologies across the full life cycle of a human services case, including in immigration and policing. Steps that can be taken include ensuring that the Advisory Council on Artificial Intelligence is diverse and representative of Canadian society,[51] allowing for civil society organisations that work on behalf of citizens to conduct oversight over current and future uses of AI by the government, and supporting further research and education to help citizens better understand the current and prospective impacts of emerging technologies (e.g. AI-backed facial recognition or lie detectors) on human rights and the public interest.

Action steps

The following action steps are suggested for Canada:

- Push for better diversity and representation in the Advisory Council on Artificial Intelligence.
- Develop a code of ethics for the deployment of AI technologies to ensure that algorithms do not violate basic principles of equality and non-discrimination. The Toronto Declaration, for example, can serve as guidance for governments, researchers and tech companies dealing with these issues.[52]
- Advocate for accountability and transparency on the government's use of AI, including algorithmic transparency, for example, through creating legislation similar to Article 22 of the European Union's General Data Protection Regulation (GDPR) on automated individual decision making, which gives individuals the right to object to any profiling that is being performed.[53]
- Work with affected communities, such as refugees and asylum claimants, to understand the purpose and effects of the government's use of AI.
- Push for digital literacy and AI-specific education to ensure that Canadians understand new technologies and their impact.
- Utilise gender-based analysis plus (GBA+)[54] and other human rights impact assessments to evaluate AI tools.
- Conduct oversight to ensure that technological experimentation by the government complies with domestic and internationally protected human rights.

[50] Ma, A. (2018, 29 October). China has started ranking citizens with a creepy 'social credit' system — here's what you can do wrong, and the embarrassing, demeaning ways they can punish you. *Business Insider*. https://www.businessinsider.com/china-social-credit-system-punishments-and-rewards-explained-2018-4

[51] Poetranto, I., Heath, V, & Molnar, P. (2019, 29 May). Canada's Advisory Council on AI lacks diversity. *Toronto Star*. https://www.thestar.com/opinion/contributors/2019/05/29/canadas-advisory-council-on-ai-lacks-diversity.html

[52] Access Now. (2018, 16 May). The Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems. https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems

[53] European Union. (2018, 25 May). Art. 22 GDPR Automated individual decision-making, including profiling. https://gdpr-info.eu/art-22-gdpr

[54] "GBA+ is an analytical process used to assess how diverse groups of women, men and non-binary people may experience policies, programs and initiatives. The 'plus' in GBA+ acknowledges that GBA goes beyond biological (sex) and socio-cultural (gender) differences." Source: Gender Based Analysis Plus (GBA+), Status of Women Canada. https://cfc-swc.gc.ca/gba-acs/index-en.html

THE CARIBBEAN

"IT'S NOT JUST ABOUT PUTTING AN APP IN THE APP STORE"[1]

Deirdre Williams
williams.deirdre@gmail.com

Introduction

Is the advent of artificial intelligence (AI) the panacea for many of the ills of the developing world, or is AI a Trojan horse to facilitate invasion, the smallpox blankets of the new colonialism? The stories of the horse and the blankets exist as a part of the human story. The Trojan horse was built by the invading Greeks, and filled with their soldiers. Then the Greeks simply waited for the Trojans to come for the horse and drag it into their besieged city. Victory from inside the city walls was easy. The blankets, infected with smallpox, and offered as gifts (as the Trojan horse had been), assisted in the colonisation of the so-called "New World" by removing indigenous inhabitants who objected to the invasion. The fact that we no longer remember our history makes us vulnerable, as what happened before can easily happen again.

This report, which is based on interviews with Caribbean people, some of whom are experts in the field of information and communications technologies (ICTs) in the Caribbean,[2] will consider "artificial intelligence" in its broadest terms as the manipulation of complex databases. AI is a tool; as such it is neither good nor evil. The issues are – for what is it used? How? Why? In the Caribbean as elsewhere we need to be vigilant to answer these questions.

Informed decision making?

Since the appointment of the current secretary general, Bernadette Lewis, in 2003, the Caribbean Telecommunications Union (CTU)[3] has distinguished itself by its ability to listen and to include. Back in 2005, it was the first institutional agency in the world to hold an Internet Governance Forum. More recently, in 2016, it was instrumental in setting up the Caribbean ICT Collaboration Forum, open to all, together with the Caribbean ICT Collaboration Committee (CICC), with representatives from government, regulators, operators and civil society.

Earlier this year (2019) the CTU celebrated the 30th anniversary[4] of its founding by the Caribbean Community (CARICOM).[5] Among other activities, the CTU arranged Caribbean FutureScape, a "hands-on" exercise to expose participants to actual experience of a possible future facilitated by digital means, with the objectives, among others:

- To demonstrate the transformative potential of ICT in a futuristic "Caribbean Single ICT Space".
- To encourage Caribbean prime ministers to commit to "21st Century Government" and national and regional digital transformation.[6]

At more or less the same time came the call from GISWatch with its theme of AI.

The juxtaposition of these events created the thought that many of the experiences that FutureScape demonstrated would, in the real world, require the use of some form of AI. It also raised the questions: how much awareness exists in the Caribbean of the potential positives and negatives of AI; and are we buying into something that may be more negative than positive?

The Caribbean, in common with the developing world, has its share of problems. Many records are still generated, and stored, on paper. One of the informants for this article deplored the contribution of the continuing use of paper to poor air quality and unsatisfactory conditions of work for employees. He hoped that digitisation might solve that problem.[7] Different aspects of the citizen's life are stored in different formats in different data sets which do not communicate directly with one another, making tasks like getting a passport or starting a business onerous and frustrating.

[1] Lynda Chin, vice-chancellor and chief innovation officer at the University of Texas System, quoted in AFP. (2017, 21 June). Artificial intelligence and the coming health revolution. *Jamaica Observer*. www.jamaicaobserver.com/business-observer/artificial-intelligence-and-the-coming-health-revolution_102379?profile=1056

[2] I thank everyone who took time to respond to my questions. All of your answers shaped this report, although you are not all cited directly. Hopefully the discussion will continue and spread.

[3] https://www.ctu.int

[4] https://www.ctu.int/event/30th-anniversary-celebration

[5] https://www.caricom.org

[6] https://www.ctu.int/wp-content/uploads/2019/04/FutureScape_Overview-Issue-3.pdf

[7] Interview with Claude Paul, General Secretary, Saint Lucia Civil Service Association, 26 June 2019.

In this situation, the possibility of "21st Century Government" envisaged by the CTU sounds like a very attractive solution. However, are there downsides, disadvantages? Are we aware of the whole picture? In fact, are the decision makers in government sufficiently aware to be able to create good policies, and to put in place the legislation and regulation necessary to make 21st Century Government truly effective?

How high a price should we pay for convenience?

Most Caribbean people agree about the advantages of AI, about the conveniences that it can provide, and that its use is inevitable. There is much less cognisance of the risks involved, although some local ICT experts are beginning to express concern. Over the last few months have come several warnings to the developing world about the need for vigilance in their approach to AI. One of the most recent was published in April this year by Oxford University Professor of Globalisation and Development Ian Goldin: "The clock is ticking and the risks posed by AI to development have never been higher. Policymakers everywhere should be listening carefully and thinking hard about how to respond."[8]

In the Caribbean, the emphasis is still on the possibility of using AI to solve developmental and other social problems. One informant suggested that a failure to apply the technology appropriately will leave us so far behind that we would drop out of sight. The same informant insisted that we must apply the technology as its masters and developers so as to avoid "forever being subscribers." We must also resist the "potential erosion of human agency."[9] We must run the technology, rather than allowing the technology to run us.

The main areas where informants perceive an intersection between "government" and AI are: data collection and management; management of the public service; energy and telecommunications; education; health care; tourism; and agriculture.

Data collection and management

Most informants commented on this issue. One response pointed to the new variations on the traditional questionnaire for collecting statistics, with the possibility of using audio recordings and photographs.[10] There was concern over the data being collected and monetised by global tech companies like Google and Facebook. The developed world is beginning to recognise its vulnerability and build protections for itself, for example, the European Union's (EU) General Data Protection Regulation (GDPR); the Caribbean is more concerned with making itself compliant with those regulations than with protecting itself.

A more subtle difficulty, presented in an article discussing AI and health care concerns in the United States and the EU,[11] is that the data used to create the AI technology may not be representative of the population where it is being applied. In a medical context this may lead to a false diagnosis. In any context a non-representative data set is likely to lead to inaccuracies. The data necessary for the particular technology may not have been collected at all, or in sufficient quantity, in the Caribbean, and in many cases the data set from elsewhere will not enable the technology to return correct results.

Most Caribbean countries already have, or are working on, some form of ICT policy. However, AI involves particular issues, especially concerning data. Lodewijk Smets, senior country economist, and Zubin Deyal, research assistant, at the Inter-American Development Bank (IDB), Trinidad and Tobago, stated:

> AI technology comes with risks; it can potentially worsen inequality by eroding jobs through automation and redirecting profits to capital owners. Appropriate economic and social policies are therefore needed to prevent this. [...] With the systematic and coordinated adoption of AI technology and the rules surrounding its uses, the people of the Caribbean would be able to benefit from this unique opportunity.[12]

Policy is needed to address data collection, to ensure that it is accurate and adequate for its purpose. One informant pointed out that "people have learned to accept that Google or Facebook make decisions (and money) on their personal data."[13]

[8] Goldin, I. (2019, 18 April). Will AI kill developing world growth? *BBC*. https://www.bbc.com/news/business-47852589

[9] Email from Organisation of Eastern Caribbean States Director General Didacus Jules, 13 June 2019.

[10] Email from Edwin St. Catherine, Director of Statistics, Government of Saint Lucia (Retired), 13 June 2019.

[11] Henderson, R., & Ross, D. (2019, 25 April). Artificial Intelligence in Healthcare: Can It Work? Perspectives from the United States and European Union. *Business Law Today.* https://businesslawtoday.org/2019/04/artificial-intelligence-healthcare-can-work-perspectives-united-states-european-union

[12] Smets, L., & Deyal, Z. (2018, 20 November). Artificial intelligence and the Caribbean. *Caribbean DevTrends*. https://blogs.iadb.org/caribbean-dev-trends/en/9397

[13] Email from Daniel Pimienta, former president of the Networks and Development Foundation (FUNREDES), 24 June 2019.

The people of the Caribbean also need to have their data protected, as well as to be compliant with the regulations in other parts of the world.

Management of the public service

The largest employer in many countries of the Caribbean is the government. While digitisation and AI may lead to less long-term storage of paper and a cleaner work environment, 21st Century Government tends to create redundancy. The "elephant in the room" at public discussions of the new technology is the threat of unemployment. And while there is also insistence that the same new technology will create new jobs, few details are volunteered and there is no coherent plan to offer appropriate re-training to those who may lose their jobs.

One approach to the current period of fiscal constraint is the public-private partnership. The public sector needs to have policies in place to ensure that such an arrangement does not accidentally make the private sector privy to personal information which the citizen is required to share with government agencies. One danger area for this is the concept of "smart cities", where an investment by a technology company that streamlines things like traffic, parking and rubbish collection contrives to skim off large quantities of data in the process. The Caribbean needs to begin to be aware of the hidden costs of convenience.

An area of concern for Caribbean people is the problem of political patronage. Jobs are sometimes awarded on the basis of "the party you vote for" rather than on qualifications. Assuming that the AI being used for decision making has been programmed to consider the local cultural context, using AI could benefit the Caribbean by basing employment appointments on objective values.

Supporting the benefit of AI in a customer service environment, one informant pointed to the fact that "[h]umans can sometimes over-complicate, dramatise or confuse scenarios with inaccurate explanations."[14]

The important task for governments is to do the research that will allow them to adopt the aspects of the technology that provide a public good while protecting citizens from aspects that may threaten them.

Energy and telecommunications

One aspect of AI that seems regularly to be forgotten is the dependence of the system on energy and telecommunications. The energy system still depends largely on imported fuel to keep running even if the generating company is itself local. For most Caribbean countries the telecommunications system is foreign owned, at least as regards its connection to the outside world. Most countries in the Caribbean are now sovereign states; rather than the previous dependency on a foreign government we are now dependent on the reliability of broadband and electricity, which itself relies on an external source. When the internet is down or there is a power cut many people are unable to work. This is another issue that needs to inform government policy – that AI might entail an increased dependency on technology for everyday essential functions and services, which are in turn vulnerable to the stability of both energy and telecommunications infrastructure.

Education, health care, tourism, agriculture

AI cannot and should not be relied upon to do everything. The decision makers need to distinguish between the different functions that are required, applying the technology in areas where it will be most effective.

In the area of education it is critical that local content displaying local values should be included in what is taught. However, AI has the capability to be endlessly patient with the minority of students who need extra support and explanation in a crowded classroom, and is particularly gifted at providing training to re-skill adults for other employment if they have been made redundant by automation, and in specialist training which may be unavailable locally. It also has the potential to greatly improve record keeping and data collection. The Cabinet of the Government of Saint Lucia ratified its "ICT in Education Policy and Strategy for Saint Lucia 2019-2022"[15] in February 2019. It is of some concern that this policy should make so little mention of the specific benefits and concerns of AI. In 2010 the current Minister for Education in Saint Lucia, Gale T. C. Rigobert, wrote about ICT proposals:

> Governments […] can be inclined to fall for these schemes […] not being sufficiently aware of the handicaps they suffer and ought to correct should they wish to benefit from the trumpeted promises.[16]

As mentioned, a main concern about using AI for health care is that it should be tailored for the

[14] Email from Dwight Thomas, Business Intelligence Developer, 28 June 2019.

[15] https://camdu.edu.lc/wp-content/uploads/2019/04/ICTE-Policy-Final-2019_2022-Web.pdf

[16] Rigobert, G. (2010). *Bridging the Digital Divide? Prospects for Caribbean development in the new techno-economic paradigm*. Brighton: World Association for Sustainable Development (WASD).

population it is being used for. Health care also generates a great deal of highly personal data. The citizen reasonably expects the state to put in place measures to protect that data, especially since "it has been noted that medical data is now three times more valuable than credit card details in illegal markets."[17]

Tourism is adopting AI enthusiastically but with caution. The Royal Caribbean cruise line has begun to use an AI facial recognition system to speed up the process of monitoring the hundreds of passengers moving on and off their ships in the different cruise ports. The company claims:

> With this, as with its other AI initiatives, Royal Caribbean follows a model of carefully monitored, small-scale trial deployments, before individual initiatives are put into organization-wide use.[18]

Agriculture was once the main employer and income earner in the region. It was an early user of AI in the form of expert systems to communicate highly specialised information about crops to illiterate farmers. It should be noted that in the IDB Caribbean DevTrends blog post cited earlier,[19] the advantages of AI are proposed as things that could happen, rather than as things already in place.

In 2019 the Food and Agriculture Organization of the United Nations (FAO) and the Caribbean Development Bank (CDB) published the "Study on the State of Agriculture in the Caribbean".[20] David Jessop, consultant to the Caribbean Council,[21] and reviewing the report in the Cayman Compass newsletter on 16 June 2019, comments:

> [T]he report […] says the sector as a whole has great potential for the creation of stronger market linkages with sectors such as tourism if support is provided to farmers, fisherfolk and agri-food businesses to adopt current international best practice and technologies.[22]

Attention should be paid to the conditional "if".

The absence of ICTs (including AI) among recommendations for improved communication of information and more efficient networking between the tourism and agriculture sectors suggests that the use of technologies in the different sectors is still highly compartmentalised.

However, there is reason for cautious optimism in the agricultural sector. In a guest article, "Caribbean farmers could be going digital", published by SciTech Europa,[23] Ken Lohento of the Technical Centre for Agricultural and Rural Cooperation ACP-EU (CTA),[24] describes three new pilot projects using blockchain that will help to "ensure that the Fourth Industrial Revolution does not leave agriculture, and small-scale farmers in particular, behind."

Conclusion

The people of the Caribbean are a resilient people, but they are also expert at "making do". They may complain among themselves very fluently, but they are somehow accustomed to being imposed on by outsiders.

They are being told that AI is overall a good and useful technology.[25] There is very little information available about the negative side of the argument, although the Inter-American Development Bank[26] and the World Bank are beginning to advocate caution. David McKenzie, in a World Bank blog post, stresses the need to "beware of the hype" and asks, "Are we learning about enough failures?" He speaks of the "high ratio of pretty pictures to demonstrated impact," and adds: "We need to be better about also making clear when these methods do not offer improvements (or when they do worse) than current methods."[27]

People in the Caribbean, as people everywhere, have a right to be told the truth, to be offered a clear picture which will allow them to make informed decisions. They also have the right to be protected by their governments.

[17] Henderson, R., & Ross, D. (2019, 25 April). Op. cit.

[18] Marr, B. (2019, 10 May). AI On Cruise Ships: The Fascinating Ways Royal Caribbean Uses Facial Recognition And Machine Vision. *Forbes. https://www.forbes.com/sites/bernardmarr/2019/05/10/the-fascinating-ways-royal-caribbean-uses-facial-recognition-and-machine-vision/#2fc51e1524bf*

[19] Smets, L., & Deyal, Z. (2018, 20 November). Op. cit.

[20] FAO & CDB. (2019). *Study on the State of Agriculture in the Caribbean*. Rome: FAO & CDB. www.fao.org/3/ca4726en/ca4726en.pdf?eloutlink=imf2fao

[21] https://www.caribbean-council.org

[22] The idea is that agriculture should organise itself to supply the foodstuffs that tourism requires. Jessop, D. (2019, 16 June). Jessop: Restoring the central role of Caribbean agriculture. *Cayman Compass*. https://www.caymancompass.com/2019/06/16/jessop-restoring-the-central-role-of-caribbean-agriculture

[23] Lohento, K. (2019, 13 August). Caribbean farmers could be going digital. *SciTech Europa*. https://www.scitecheuropa.eu/caribbean-farmers-could-be-going-digital/96550

[24] https://www.cta.int/en

[25] Ammachchi, N. (2019, 10 April). Artificial Intelligence Can Become a Game-Changer for Caribbean Economies. *Nearshore Americas*. https://www.nearshoreamericas.com/artificial-intelligence-ai-become-game-changer-caribbean-economies

[26] Smets, L., & Deyal, Z. (2018, 20 November). Op. cit.

[27] McKenzie, D. (2018, 5 March). How can machine learning and artificial intelligence be used in development interventions and impact evaluations? *World Bank Blogs*. https://blogs.worldbank.org/impactevaluations/how-can-machine-learning-and-artificial-intelligence-be-used-development-interventions-and-impact

They have a right to the protection of their private information. To understand how this works in the online world, they have a right to an explanation of how their privacy may be threatened.

They have a right to elect a government that works for their well-being. They have an obligation to "supervise" the work that their government is doing for them.

The suggestion has been raised by more than one informant that the relationship between citizen and state should be reviewed. The big tech companies are pushing AI. At the same time, they require the data that the technology will generate for them. If the citizen/state relationship could be improved, then there might be a chance of restoring a more equitable balance.

Action steps

The following is necessary in the Caribbean:

- Many organisations are largely unaware of the extent to which they do not know about the issues surrounding AI. It is important for those who do know to be proactive about sharing their knowledge.
- Since the people have empowered the government for their joint social good, they should ensure that the government is aware of their concerns.
- A government/civil society alliance is necessary to counterbalance the thrust of the "big tech" companies.
- Be vigilant! There seem to be areas in which AI offers a solution to the problem; however, there is little awareness among those with the problem of this possibility.

CHILE

BIG DATA IN PUBLIC EDUCATION IN CHILE: BUILDING A PLATFORM FOR EQUAL EDUCATION

Instituto de la Comunicación e Imagen, Universidad de Chile; Fundación Datos Protegidos
Patricia Peña and Jessica Matus
www.icei.uchile.cl; www.datosprotegidos.org

Introduction: The digital transformation of the public sector

Chile, like most countries in the Latin American region, is moving towards the so-called "Fourth Industrial Revolution" without a specific public policy to guide the process that takes into account the social and cultural implications and challenges of this transformation. As a result, artificial intelligence (AI) technology is a challenge for both the private and public sector looking to develop new services and innovations. This is particularly the case when it comes to the implementation of technologies using big data systems.

This report discusses the challenges of a big data platform that has been designed and developed for the public education system in Chile. It aims to identify social inequalities that occur in different regions and at the local level. The platform has been set up by the Ministry of Education in order to strengthen decisions made in line with the New Public Education (NEP) policy that was implemented recently. The task of improving the quality of and access to education for the more than one million students in the public education system in Chile is one of the biggest challenges facing the current administration of Sebastián Piñera.

Transforming public education in Chile

Speaking at the VI Annual Meeting of the Chilean Society of Public Policies in 2018, Pasi Sahlberg, a Finnish academic and researcher specialising in educational policies, had this to say about the Chilean education system:

> Chile is an international example of a widely privatised system that operates according to free market principles, which brings with it a decreasing equity in learning outcomes, a lower than expected quality of general education, and a growing dissatisfaction of parents towards the educational system.[1]

Since the military dictatorship (1973-1990), Chile has had a mixed system of education, with three different models for schools administered by the public sector: municipal or public schools which are funded with public money; schools subsidised through public funds, but where fees are also paid by the parents; and private schools, where the fees are paid in full by the parents or families. The so-called "municipalisation of public education" – under which schools were managed by municipalities rather than the Ministry of Education – was one of the most drastic changes made by Pinochet's dictatorship. The return to democracy did not modify this system, and several national and international reports[2] and measurements have identified critical weaknesses and shortcomings, ranging from the quality of learning to the need to increase the role of the public sector in the country's education system.[3] Some of this data indicates that the budget for primary education in Chile is one of the lowest among member countries of the Organisation for Economic Co-operation and Development (OECD), along with Mexico, and that between 17% and 16% of students in Chile do not complete schooling.[4]

The NEP is an unprecedented strategy for the Chilean education system. As mandated by Law No. 21,040 (2019), the publicly run kindergartens, primary schools and secondary schools in 345 municipalities will now be run by 70 new Local Public Education Services (SLEPs). These services will replace the municipalities in the administration of public education at the local level, not only in relation to administration and financial or infrastructure management, but also in relation to pedagogical and technical advice on improving the quality of learning.

[1] https://www.eldinamo.cl/educacion/2018/01/10/desigual-segregadora-y-con-poca-calidad-academico-de-harvard-hace-lapidario-analisis-de-la-educacion-chilena

[2] See, for example, OECD. (2018). *Education at a Glance 2018: OECD Indicators*. https://www.oecd-ilibrary.org/education/education-at-a-glance-2018_eag-2018-en (Chile has been a member of the OECD since 2010); UNICEF. (2018). *An Unfair Start: Inequality in Children's Education in Rich Countries*. https://www.unicef.org/publications/index_103355.html; and the reports of the OECD's Programme for International Student Assessment (PISA), which tests 15-year-old students from all over the world in reading, mathematics and science every three years. https://www.oecd.org/pisa

[3] The role of the public sector is relatively weaker than the role of the private sector in primary, secondary and university education.

[4] OECD. (2018). Op. cit.

Pilots have been launched in two communities, which will allow for the impact of the implementation process to be assessed. The roll-out of the new system will then continue until 2025, with the long-term objective of encouraging participation and freeing resources to develop locally relevant projects that improve the quality of education and provide opportunities for students.

The digital platform used for this new education system is called the "Public Education Information, Monitoring and Evaluation System" (*Sistema de información, seguimiento y evaluación de la educación pública*). The aim of the platform is to improve decision making regarding the implementation of the new public policy in education by providing objective, timely and up-to-date information. The system was developed by the Centre for Advanced Research in Education (*Centro de Investigación Avanzada en Educación* – CIAE) at the Universidad de Chile and by the Territorial Intelligence Centre (*Centro de Inteligencia Territorial* – CIT) at the Universidad Adolfo Ibáñez. Its development was funded by the National Commission on Science and Technology. The system was delivered a year ago, in 2018, to the Public Education Division of the Ministry of Education.

A big data platform for improved decision making in public education

How many schoolchildren in Chile travel between communes (the smallest administrative subdivision in Chile) to access their education? What is the sort of coverage public schools offer local areas? How many girls and how many boys drop out of the system and what are their educational trajectories? And what relation do these variables have with the educational experience that Chilean children have and will have in their future? These are some key questions that needed to be answered in the implementation of the NEP system.

"The education system does not operate in a vacuum, and the characteristics of the local territories and of the students who study there, or how the public provision of education in a given territory is, condition what we can do from the public policy point of view," explained Patricio Rodríguez and Luis Valenzuela, academics from the CIAE and CIT, when the digital platform was presented to the media in 2018.[5] "These factors are usually invisible to the design and implementation of public policies. For this reason, the platform provides objective, timely and updated information for correct decision making regarding the implementation of public policies in education," they said.

In this way, the platform seeks to obtain answers to strategic questions that guide public policies in education, including information on the local context, socio-demographic data, or statistics on the school drop-out rate.

The design and development of the platform was carried out in several stages, and drew on evidence collected in two previous projects focused on the management of schools. For example, in regions such as the Santiago Metropolitan Region, where Santiago, the capital of Chile is located, children from the peripheral communes, which are the poorest, have to travel a long distance to their schools because for many families, there are no good-quality public education institutions in their areas. Rodríguez explained:

> The idea is that with this platform, decision making and the elaboration of strategic plans for the new services have information about what is happening in their territories, because urban and rural areas are not the same, and the most important thing is that decisions are made based on evidence. What we hope is that this platform can enhance and strengthen the promise of equity of public education in the territory, which ultimately means that the authorities invest more resources and budgets where there are more needs.

The challenge of designing big data with the right questions

The phrase "data is the new oil" points to the value that social and economic development gives to data that is already collected and mined for platforms and systems that are set up to solve diverse needs. A great challenge faced by big data systems, however, is the need to draw on different data sets created by systems set up in the public and private sectors.

In this case, the platform uses available data from various sources of public information, such as the Population and Housing Census, student enrolments from the Ministry of Education, and the results of the Quality of Education Measurement System (*Sistema De Medición de la Calidad de la Educación* – SIMCE)[6] tests conducted by the Education Quality Agency, among others, which it

[5] www.ciae.uchile.cl/index.php?page=view_noticias&id=1370&langSite=es

[6] SIMCE was first implemented in 1988, and evaluates learning achievements in the subjects of language and communication (reading and writing comprehension), mathematics, natural sciences, history, geography, social sciences and English. The SIMCE tests are conducted with students in grades 2, 4, 6, 8 (primary) and II and III (secondary) in Chile.

analyses and visualises using interactive graphs and maps. As Rodríguez highlighted:

> "What data do we need to make decisions?" will also be a key question to build this evidence-based approach to public sector decision making. There must be what is called a chain of quality from its collection, capture, use and reuse, especially when it is taken from other databases, so that no bias is generated.

In this context, he pointed out that the skills and capacities of state employees implementing the system are critical:

> The most important thing has to do with [their] competencies. [...] They must make decisions based on the evidence that these systems deliver. Thinking about an automatic or automated decision-making process is risky [...] if we treat it like a black box and do not know how machine learning works in the processing of this data. There are ethical and technical problems that can occur.

This point will be key in the political decisions on how and why to use this data to make decisions that have an impact on citizens.

The protection of personal data in Chile in the face of the big data challenge

Big data solutions and platforms are generally presented as the solution to diverse challenges faced in the delivery of public and private services. However, a key issue is that citizens are not aware of the data that the systems have on them, and are also typically unaware of the laws and regulations that protect their personal data.

According to a recent report on so-called "data trusts":

> Individuals have little control over their data – how it is collected, who collects it, and for what it is used. For many, the common experience with online platforms, mobile apps and other digital services is blindly accepting whatever demands they make of our data, which are often a necessary condition of use. Yet a new public awareness has grown amid news of scandals around the misuse of data and major data breaches, and it is clear that the private sector has failed to protect individual privacy rights through self-regulation.[7]

It is only since 2018 that data protection has been constitutionally recognised in Chile.[8] This guarantee is the power to control one's own information as opposed to its automated processing. However, we are governed by an old data law (Law Nº19.628), passed in 1999. Although this law has been modified, it is still not specific enough to comply with international standards, nor to offer due protection of the rights and freedoms of individuals.

For example, the law does not allow for independent oversight of the processing of data, such as an independent public body that supervises those who process data. This body should not only have powers of intervention, investigation, inspection and sanction, as has been established in current legislation, but also of promotion, dissemination and assistance. The definitions, principles and basic rules of data processing contained in the law, and the self-regulation of the private sector, are not sufficient to protect individual privacy rights.

A new personal data law is currently being discussed in the Chilean Congress, which will also be key in the development of technologies that will use data as the basis for development.

The challenge of new approaches to big data: Diversity and citizenship

The case presented in this report is one of the first public sector initiatives in Chile that proposes a solution based on big data. It will be important to understand its impact, as this is relevant to the future adoption of AI in the public sector. Although Chile does not yet have a specific national policy or plan for AI, a commission of politicians and scientists has been set up to develop a proposal for a National AI Strategy.

In this context, the so-called "datafication" of life needs an informed and critical public debate and analysis. There is a need for transparency on how these platforms and systems are designed and configured, and how the algorithms that will be used to make key decisions in a person's life will be programmed and built. It is not only about fundamental human rights (such as privacy), the security of sensitive data, or the promise of the benefits of "digital transformation" using AI in the development of services and systems with social impact. It is also about the structural transformation of society.

[7] Element AI, & Nesta. (2019). *Data Trusts: A new tool for data governance.* https://hello.elementai.com/rs/024-OAQ-547/images/Data_Trusts_EN_201914.pdf

[8] Article 19.4 of Chile's Constitution enshrines respect for and protection of the privacy and honour of individuals and their families, as well as the protection of their personal data. The processing and protection of this personal information is regulated by a law which establishes the manner and conditions in which these are carried out.

A perspective along these lines was proposed by Milan and Treré[9] after the Big Data from the South Initiative invited academics, researchers, practitioners, activists and civil society organisations from a variety of areas and regions of the world to discuss the issue. The discussion raised uncomfortable questions about the consequences of big data for social justice, such as new forms of surveillance of people who use public systems. It is also a challenge for academics and researchers who seek to design new solutions using big data and AI, and emphasises the need to promote participation at the local level, with citizens not as users or consumers of digital solutions, but as intrinsically a part of the construction of societies and democracies.

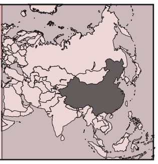
Action steps

The following steps are necessary in Chile:

- Advocate for the inclusion of the civil society sector, representing different groups, interests and minorities, in the National AI Strategy discussion process. They should not only be consulted right at the end of the development of the strategy.
- Build the knowledge, understanding and competency of civil society actors on issues to do with datafication and digital transformation in an alliance with academia, considering the diversity of disciplines and specialties in the academic sector. This is important to understand the social impacts and challenges these imply, particularly in relation to human rights, privacy, ethics, equality and inclusion.
- Share the knowledge and experiences from other countries that have developed strategies for AI.
- There is also a need to share the experiences of civil society in research and activism, such as designing new methodologies for research that create "data activists" at the grassroots level.

[9] Milan, S., & Treré, E. (2017, 16 October). Big Data from the South: The beginning of a conversation we must have. *DATACTIVE*. https://data-activism.net/2017/10/bigdatasur

CHINA



ALGORITHMIC OPPRESSION WITH CHINESE CHARACTERISTICS: AI AGAINST XINJIANG'S UYGHURS

**Chinese University of Hong Kong Faculty of Law/
Strathclyde University Law School**
Angela Daly
www.law.cuhk.edu.hk/en

Introduction

The ways in which artificial intelligence (AI), in particular facial recognition technology, is being used by the Chinese state against the Uyghur ethnic minority demonstrate how big data gathering, analysis and AI have become ubiquitous surveillance mechanisms in China. These actual uses of facial recognition will be compared with the rhetoric on AI ethics which is beginning to emerge from public and private actors in China. Implications include the mismatch between rhetoric and practice with regards to AI in China; a more global understanding of algorithmic discrimination, which in China explicitly targets and categorises Uyghur people and other ethnic minorities; and a greater awareness of AI technologies developed and used in China which may then be exported to other states, including supposed liberal democracies, and used in similar ways.

Context

Digitisation in China and China's digital industries are now of global importance, given the huge market of over a billion people, high levels of connectivity and use of digital services such as mobile payments (which outstrips take-up in Western markets), and development of a home-grown internet service industry centred on Baidu, Alibaba and Tencent to rival Silicon Valley's Google, Amazon, Facebook and Apple. Digitisation in China has performed important roles in socioeconomic development, with private sector actors' activities aligned with the Chinese Communist Party (CCP) government's goals, and in compliance with government censorship rules.[1]

The next phases of digitisation are rapidly being implemented in China, notably the full roll-out of the Social Credit system by 2020. The system includes a number of data-gathering and analysis techniques including facial recognition applications in public places and algorithmic decision making. AI is viewed as a highly strategic area of development by the Chinese government, and China rivals only the United States (US) in its AI technology research and development, and also implementation.[2]

However, digitisation and the implementation of AI so far in China have exhibited serious concerns for human rights. Similarly to the ways in which AI and algorithms in the US reinforce existing racial and gender inequalities,[3] and how surveillance and other forms of data gathering are specifically targeted at racial and religious minorities across the West,[4] intensified practices of data gathering, analysis and AI implementations are being directed at the majority-Muslim Uyghur people and other minorities in the Xinjiang Uyghur Autonomous Region in northwest China (also known as East Turkestan), "amplify[ing] systems of inequality and oppression."[5]

The Uyghurs, a Turkic-language speaking group who predominantly reside in Xinjiang, have been subject to repression from the Chinese state, including the internment of up to a million people in detention camps.[6] While the Uyghurs socially, religiously and culturally have strong affinities with their Central Asian neighbours, politically and economically they have been connected to China since the Qing Dynasty annexation of their territory in 1755, aside from two brief periods of independence in the 20th century, culminating in the incorporation of Xinjiang into the People's Republic of China in 1949.[7] The territory is geographically located close to other politically vola-

[1] Arora, P. (2019). Benign dataveillance? Examining novel data-driven governance systems in India and China. *First Monday*, *24*(4).

[2] Lee, K. F. (2018). *AI Superpowers: China, Silicon Valley and the New World Order*. Boston: Houghton Mifflin Harcourt.

[3] Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

[4] See, for example, Mann, M., & Daly, A. (2019). (Big) Data and the North-in-South: Australia's Informational Imperialism and Digital Colonialism. *Television and New Media*, *20*(4), 379-395. https://eprints.qut.edu.au/123774/1/North-In-South.pdf

[5] Arora, P. (2019). Op. cit.

[6] Zenz, A. (2018). New Evidence for China's Political Re-Education Campaign in Xinjiang. *China Brief*, *18*(10). https://jamestown.org/program/evidence-for-chinas-political-re-education-campaign-in-xinjiang; Mozur, P. (2019, 14 April). One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority. *The New York Times*. https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html

[7] Aktas, I. (2015). Uighur Separatism and Human Rights: A Contextual Analysis. In M. Kosmala-Mozlowska (Ed.), *Democracy and Human Rights in East Asia and Beyond – Critical Essays*. Warsaw: Collegium Civitas Press.

tile regions such as Tibet, Afghanistan and Kashmir, thereby exposing it to the influence of "wider Asian power politics", and is also endowed with natural resources.[8] Xinjiang is also a key connection point for China's Belt and Road Initiative (BRI), a trade policy aimed at strengthening "Beijing's economic leadership through a vast programme of infrastructure building throughout China's neighbouring regions."[9] A Uyghur separatist or pro-independence movement exists, and for some time has been under Chinese state surveillance. The security situation has been heightened by sectarian riots in Urumqi (the capital of Xinjiang) in 2009 between Han Chinese and Uyghurs, and a series of attacks (mainly involving knives and vehicles) on Han Chinese perpetrated by Uyghurs since 2000.[10]

AI and surveillance in China

Government surveillance activities in China have existed at least since the birth of the current state in 1949. However, since 2013, when Xi Jinping came to power in China, censorship, surveillance and monitoring of electronic communications, as well as the gathering of big data and analysis of citizens' communications and activities, have intensified as a means of shoring up the CCP party-state against challenges, [11] "rapidly" turning China into "an information or surveillance state".[12]

Surveillance cameras, especially used by the government, are now prevalent in China, and increasingly incorporate facial recognition and "intelligence analysis" (flagging objects or "events of interest").[13] China's very large population and the facial data generated from it via these cameras, coupled with government support for industry endeavours in this area, have fuelled research into machine learning and the implementation of algorithm-powered facial recognition technology.[14] In 2017 it was reported that the Chinese government's "Skynet" video surveillance initiative (accompanied by the more recent "Sharp Eyes" programme) had been completed. This is the largest initiative in the world to monitor public spaces and details such as the "gender, clothing, and height" of people coming into the surveillance cameras' vision.[15]

Surveillance in Xinjiang

While these technologies and other surveillance programmes have been rolled out throughout the country, the Uyghur minority in Xinjiang has been specifically targeted:

> China's western Xinjiang region, home to the Uyghurs, has effectively become "a 'frontline laboratory' for data-driven surveillance." Cameras are ubiquitous in Xinjiang, and their view extends well outside urban centers. The methods employed in this province may well foreshadow the nationwide implementation of similar "predictive-policing tactics" in the months to come. Xinjiang is also the place in which DNA-collection efforts have taken their most extreme form.[16]

This use of surveillance technology, in particular AI-enabled facial recognition, takes place amidst increasing repression of Uyghurs and other minority groups in Xinjiang. Since 2017, a major "re-education campaign" has been taking place in the territory, involving the large-scale "extrajudicial" internment of at least tens of thousands of people in re-education camps, with the ostensible aim of "de-extremification" and ideological "assimilation" of the Uyghur and other minorities to a *de facto* Han Chinese/ atheist CCP norm.[17] The CCP has "constructed a sophisticated multi-layered network of mass surveillance in Xinjiang" which "includes both covert and overt monitoring as well as the categorization, exhortation and disciplining of its population in the name of safety, civility and progress."[18]

Leibold notes that the surveillance assemblage in Xinjiang includes both "machine and human-driven systems that do not always fit together nor form a coherent whole."[19] Despite the government's wish for a seamless and ubiquitous form of monitoring and control, there are still practical obstacles including "poor technological integration and human coordination", and

[8] Ibid.

[9] Cai, P. (2017). *Understanding China's Belt and Road Initiative*. Lowy Institute for International Policy. https://hdl.handle.net/11540/6810

[10] Aktas, I. (2015). Op cit.

[11] Qiang, X. (2019). The Road to Digital Unfreedom: President Xi's Surveillance State. *Journal of Democracy, (30)*1, 53-67.

[12] Leibold, J. (2019). Surveillance in China's Xinjiang Region: Ethnic Sorting, Coercion, and Inducement. *Journal of Contemporary China* (forthcoming).

[13] Qiang, X. (2019). Op. cit.

[14] Ibid.

[15] Ibid.

[16] Ibid.

[17] Zenz, A. (2019). 'Thoroughly reforming them towards a healthy heart attitude': China's political re-education campaign in Xinjiang. *Central Asian Survey*, *(38)*1, 102-128.

[18] Leibold, J. (2019). Op. cit. See also Human Rights Watch's work reverse engineering an app used by police in Xinjiang to collect information about individuals and communicate that information with the authorities' Integrated Joint Operations Platform which aggregates this data and "flags" individuals who are deemed to be potentially threatening. Human Rights Watch. (2019, 1 May). China's Algorithms of Repression: Reverse Engineering a Xinjiang Police Mass Surveillance App. https://www.hrw.org/report/2019/05/01/chinas-algorithms-repression/reverse-engineering-xinjiang-police-mass-surveillance

[19] Leibold, J. (2019). Op. cit.

the cost and technical difficulty in updating and maintaining surveillance equipment in Xinjiang given its "harsh arid climate, where surveillance systems remain susceptible to decay, sabotage and obsolescence."[20] Furthermore, the facial recognition technologies developed in China also exhibit errors and inaccuracies like facial recognition systems developed elsewhere, including in identifying individuals and in particular in categorising them by ethnic group.

Chinese AI and corporate involvement

That being said, Xinjiang retains a "laboratory" status for the trial of Chinese-made surveillance technologies including those that use AI methods, in particular facial recognition technology. The Urumqi train station was the first to use fully automated gates incorporating facial recognition technology in 2016, and various companies such as Taisau (based in Shenzhen) provide cutting-edge technology for smart-gates implemented throughout the region in public spaces.[21] Facial recognition is also used in smart cameras throughout Xinjiang, including in mosques, provided by companies such as Hikvision (based in Hangzhou).[22] Hikvision has received contracts to provide surveillance equipment, including with facial recognition capacity, in Xinjiang, totalling over USD 290 million.[23] It was reported that Hikvision previously offered options to identify minorities but this was phased out in 2018.[24]

Facial recognition start-ups SenseTime and Megvii (Face++) are also reported to be providing their systems to surveillance operations in Xinjiang.[25] *New York Times* reporters were shown a database provided by SenseNets which "contained facial recognition records and ID scans for about 2.5 million people, mostly in Urumqi, a city with a population of about 3.5 million."[26] Meanwhile, local technology companies have been benefiting from the heightened surveillance activity in Xinjiang, such as Leon Technologies based in Urumqi, which in 2017 saw a huge increase in earnings.[27] Beijing-based CloudWalk has also advertised facial recognition technology which it claims can recognise "sensitive" groups of people,[28] and university researchers in Xinjiang have conducted research into "ethnic" aspects of facial recognition templates distinguishing "Uyghur" features.[29]

Some of these companies have received funding from larger, including foreign, investors such as Qualcomm, which has invested in SenseTime, while Kai-Fu Lee's Sinovation Ventures has invested in Megvii.[30]

It is not only in Xinjiang where facial recognition is being used against Uyghur people. It was reported that Chinese East Coast cities are using facial recognition cameras to detect Uyghurs in these locations as well.[31] Police in other Chinese provinces, including in the prosperous Guangdong Province in the south of the country, have meanwhile expressed interest in the surveillance technologies and applications tested and implemented in Xinjiang.[32]

Export of Chinese surveillance technology

Products and services developed in China have started to be exported to other countries, including by companies which have been active in providing surveillance technology and facial recognition capabilities in Xinjiang. Such recipient countries include Ecuador (with its notable ECU-911 system), Zimbabwe, Pakistan and Germany.[33]

In the context of the US-China trade war, at the time of writing the US is considering adding several surveillance companies including Hikvision, Megvii and Dahua to a blacklist, with their participation in Xinjiang facial recognition surveillance activities being part of the justification.[34]

[20] Ibid.

[21] Ibid.

[22] Ibid.

[23] Rollet, C. (2018, 23 April). Dahua and Hikvision Win Over $1 Billion In Government-Backed Projects In Xinjiang. *IPVM*. https://ipvm.com/reports/xinjiang-dahua-hikvision

[24] Mozur, P. (2019, 14 April). Op. cit.

[25] Ding, J. (2018, 24 September). ChinAI Newsletter #29: Complicit - China's AI Unicorns and the Securitization of Xinjiang. https://chinai.substack.com/p/chinai-newsletter-29-complicit-chinas-ai-unicorns-and-the-securitization-of-xinjiang

[26] Buckley, C., & Mozur, P. (2019, 22 May). How China Uses High-Tech Surveillance to Subdue Minorities. *The New York Times*. https://www.nytimes.com/2019/05/22/world/asia/china-surveillance-xinjiang.html

[27] Rajagopalan, M. (2017, 17 October). This Is What a 21st Century Police State Really Looks Like. *BuzzFeed News*. https://www.buzzfeednews.com/article/meghara/the-police-state-of-the-future-is-already-here

[28] Mozur, P. (2019, 14 April). Op. cit.

[29] Zuo, H., Wang, L., & Qin, J. (2017). XJU1: A Chinese Ethnic Minorities Face Database. Paper presented at IEEE International Conference on Machine Vision and Information Technology (CMVIT). https://ieeexplore.ieee.org/abstract/document/7878646

[30] Mozur, P. (2019, 14 April). Op. cit.

[31] Ibid.

[32] Buckley, C., & Mozur, P. (2019, 22 May). Op. cit.

[33] Mozur, P., Kessel, J. M., & Chan, M. (2019, 24 April). Made in China, Exported to the World: The Surveillance State. *The New York Times*. https://www.nytimes.com/2019/04/24/technology/ecuador-surveillance-cameras-police-government.html

[34] Bloomberg. (2019, 23 May). U.S. weighs blacklisting five Chinese video surveillance firms over treatment of Uighurs. *Japan Times*. https://www.japantimes.co.jp/news/2019/05/23/asia-pacific/u-s-weighs-blacklisting-chinese-surveillance-firms/#.XXgjTWbQ_IU; Kharpal, A. (2019, 26 May). US takes aim at Chinese surveillance as the trade war becomes a tech war. *CNBC*. https://www.cnbc.com/2019/05/27/china-mass-surveillance-state-technology-at-center.html

AI, law and ethics in China

In principle there are legal protections for the Uyghurs as a recognised minority "nationality" in the Constitution of the People's Republic of China. Article 4 guarantees the equality of all nationalities, protects against discrimination and also protects their rights to use their languages and practise their customs. Freedom of religious belief is also guaranteed by Article 36 of the Constitution. In the case of the targeted use of facial recognition against Uyghurs and other minorities in Xinjiang, "[p]olicies and administrative decisions on both central and provincial levels, however, often contradict the legal protection."[35] Furthermore, individuals are not able to enforce constitutional rights through the court system in China if the rights concerned are not also prescribed in civil laws.[36]

Ironically, Chinese government agencies, companies and universities have been active recently in the global trend towards formulating and issuing statements on AI ethics.[37] Yet the discriminatory ways in which state organs, companies and academics have researched, developed and implemented facial recognition in China would seem not to comply with Article 3 ("Fair and just") of the recent Artificial Intelligence Industry Alliance (AIIA) draft Joint Pledge on Artificial Intelligence Industry Self-Discipline, nor Principle 3 ("Fairness and justice") of the National Governance Committee for the New Generation Artificial Intelligence's Governance Principles for the New Generation Artificial Intelligence.[38]

This gap between stated ethical principles and on-the-ground applications of AI is not unique to China and can be observed in many other countries, including supposed liberal democracies in the West. However, this gap does demonstrate the weakness of unenforceable ethics statements and suggests that "ethics washing" is not a phenomenon confined to the West.[39] In any case, China's ambitions to become the world leader in AI by 2030 and also the leading role it is taking, along with the European Union, in formulating AI ethics initiatives, should be viewed critically given these highly unethical uses of facial recognition domestically.

Conclusion

The uses of cutting-edge AI, especially facial recognition, and other digitised technologies to keep Uyghur and other ethno-religious minorities in Xinjiang under the eyes of a watchful state can be viewed as a particularly acute and racist form of "digital social control" in the context of the increasingly authoritarian rule of Xi Jinping.[40] There is the potential for such monitoring techniques to be rolled out to the broader Chinese population and also beyond through the export of facial recognition technology developed in the Xinjiang "laboratory" worldwide. While so far China's Christian minority has not been subject to the same level of repression as Uyghurs, who tend to be predominantly Muslim, there are reports that Chinese authorities have also tried to install cameras in Christian churches.[41]

The extreme surveillance and re-education measures affecting large swathes of the Uyghur population in Xinjiang are not only disproportionate but also unethical. In particular, the ethnically and religiously targeted use of facial recognition technology against Uyghur people and other minorities in Xinjiang demonstrates the way in which, despite the official rhetoric on AI ethics, AI technologies are being used in China and in different parts of the world to reinforce and at times also exacerbate existing inequalities. In this sense, "algorithmic oppression" is not a phenomenon confined to the US or the West, but is also taking shape in other locations, notably China. The fact that China's AI industry rivals only the US, and the fact that other countries, including Germany, are importing surveillance technologies from China, should give all of us cause for concern.

[35] Aktas, I. (2015). Op. cit.

[36] While specific provisions of the PRC Constitution may be cited and mentioned in court reasoning, they cannot be used as a direct reason for a ruling.

[37] Daly, A., Hagendorff, T., Li, H., Mann, M., Marda, V., Wagner, B., Wang, W., & Witteborn, S. (2019). *Artificial Intelligence, Governance and Ethics: Global Perspectives.* The Chinese University of Hong Kong Faculty of Law Research Paper No. 2019-15. https://ssrn.com/abstract=3414805

[38] Webster, G. (2019, 17 June). Translation: Chinese AI Alliance Drafts Self-Discipline 'Joint Pledge'. *New America.* https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-ai-alliance-drafts-self-discipline-joint-pledge; National Governance Committee for the New Generation Artificial Intelligence. (2019, 17 June). Governance Principles for the New Generation Artificial Intelligence – Developing Responsible Artificial Intelligence. *China Daily*. https://www.chinadaily.com.cn/a/201906/17/WS5d07486ba3103dbf14328ab7.html?from=groupmessage&isappinstalled=0

[39] Wagner, B. (2018). Ethics as an escape from regulation: From ethics-washing to ethics-shopping? In E. Bayamlioglu, I. Baraliuc, L. A. W. Janssens, & M. Hildebrandt (Eds.), *Being Profiled: Cogitas ergo sum.* Amsterdam: Amsterdam University Press. https://www.aup.nl/en/book/9789463722124/being-profiled-cogitas-ergo-sum

[40] Shan, W. (2018). Social Control in China: Towards a "Smart" and Sophisticated System. *East Asian Policy*, *10*(1), 47-55.

[41] Ma, A. (2018, 11 December). China rounded up 100 Christians in coordinated raids and banned people from talking about it on social media. *Business Insider.* https://www.businessinsider.com/china-rounds-up-christians-bans-discussion-2018-12; Yan, A. (2017, 3 April). In 'China's Jerusalem', 'anti-terror cameras' the new cross for churches to bear. *South China Morning Post.* https://www.scmp.com/news/china/policies-politics/article/2084169/chinas-jerusalem-anti-terror-cameras-new-cross-churches

Action steps

The following advocacy priorities are necessary in China:

- Mass surveillance and data gathering targeted at Uyghur people and other minorities in Xinjiang by Chinese authorities and companies should cease immediately, along with the more general repression of these minority groups including internment in re-education camps.
- Chinese authorities and companies developing and implementing facial recognition technologies and other AI applications should be held to account for the unethical ways in which they may be used as a tool of oppression. In particular, they should be judged against the AI ethics principles which have begun to proliferate in China from companies and academic researchers to expose the mismatch between rhetoric and practice.
- The export and use of AI and surveillance technologies from China which may have been developed in the Xinjiang laboratory should be blocked by other countries, and campaigned against by civil society internationally.
- Campaigns against unethical AI should not hesitate to call out unethical developments and uses of AI wherever that may be, in the US, Europe, China, India and elsewhere. Legally enforceable ethical standards for AI must be implemented everywhere. The problem of unethical AI is global. There should be a worldwide ban on the use of facial recognition technologies.

COLOMBIA

ENTHUSIASM AND COMPLEXITY: LEARNING FROM THE "PROMETEA" PILOT IN COLOMBIA'S JUDICIAL SYSTEM

Karisma Foundation
Lucía Camacho Gutiérrez, Juan Diego Castañeda Gómez and Víctor Práxedes Saavedra Rionda
https://karisma.org.co

Introduction

At the start of 2019, the arrival of a so-called artificial intelligence (AI) solution for the Colombian Constitutional Court was announced in the media. Many questions were raised by civil society and academia regarding the likely impact of "Prometea" – the name of the system – including how it worked and the decision-making process that led to its adoption. There was, however, little information forthcoming. Prometea ended up being a pilot that is currently on hold.

The arrival of Prometea presented us with two possible levels of analysis: 1) the impact of the AI system itself, and 2) the impact of the adoption of an AI system in a society. Both levels are at risk of being overlooked due to the "charismatic halo" surrounding technologies and the false perception of technological neutrality and its infinite potential to solve pressing problems.

By considering the case of Prometea, we aim to identify some of the possible social and human rights impacts that are the result of the adoption of AI-based solutions in particular contexts. For this purpose, we reviewed how the media covered Prometea and conducted a series of interviews with different actors involved and interested in the issue.[1]

Background

In 2011, the Colombian Congress enacted the Code of Administrative Litigation and Procedures.[2] This code required (in article 186) the Supreme Judiciary Council to digitise its records within five years. This time has now expired with no significant advances in this regard, despite the fact that several other laws,[3] some going back as far as 1996, have also required the creation of digital judicial records.

Previous to the creation of the Code of Administrative Litigation and Procedures, the Colombian judicial system had been troubled by delays in resolving cases in part due to legal processes being paper-based. Among the risks of paper-based records were the possibility of losing the records, the vulnerability of the records to being destroyed, for example by fire, and their deterioration over time. Processing cases on paper also affected the rights of plaintiffs because it took time to process the large numbers of paper-based records needed in each legal action.

According to statistics published by the Private Council for Competition in 2018,[4] the Colombian judicial system commonly takes between 385 to 956 days to settle a case, and depending on the issue at stake it may take even twice as long as that. One of the main recommendations the Council gave to the national government was to digitise judicial records as a means of achieving not only efficiency in justice, but also timeliness.

In December 2018, it was announced by the national press[5] that an agreement had been reached by the highest national courts and government to digitise judicial records. This agreement was nothing different from the previous year's promises and provisions in the law on the same issue that remained unattended. But an announcement made in

[1] In particular, we have interviewed journalists, scholars and constitutional court workers. Given the size of the piece, we have decided to agglutinate the opinions on the main issues raised instead of directly quoting individuals.

[2] Código de Procedimiento Administrativo y de lo Contencioso Administrativo, Ley 1347 de 2011. www.secretariasenado.gov.co/senado/basedoc/ley_1437_2011.html

[3] Ley 270 de 1996, Ley 1437 de 2011, art. 186, Ley 1564 de 2012, Ley 1709 de 2014, Ley 1743 de 2014, Decreto 272 de 2015, Decreto 1069 de 2015, Decreto 1482 de 2018.

[4] Consejo Privado de Competitividad. (2017). *Informe Nacional de Competitividad 2017-2018*. https://compite.com.co/informe/informe-nacional-de-competitividad-2017-2018/

[5] Ámbito Jurídico. (2018, 13 December). En estos cinco procesos se implementará el expediente electrónico. https://www.ambitojuridico.com/noticias/tecnologia/procesal-y-disciplinario/en-estos-cinco-procesos-se-implementara-el-expediente; Redacción Tecnósfera (2018, 12 December). Arranca piloto para implementar el Expediente Digital Electrónico. *El Tiempo*. https://www.eltiempo.com/tecnosfera/novedades-tecnologia/gobierno-anuncia-plan-piloto-de-expediente-electronico-digital-304654; Redacción Judicial. (2018, 28 November). Consejo Superior de la Judicatura aboga por una justicia digital. *El Espectador*. https://www.elespectador.com/noticias/judicial/consejo-superior-de-la-judicatura-aboga-por-una-justicia-digital-articulo-826076

the press[6] during February 2019, about an AI solution that would be applied to the national judicial system – specifically in the Colombian Constitutional Court – caught everyone's attention.

Prometea was presented as the main means of resolving the inefficiency and delays within the Constitutional Court. Its design, according to the media coverage, would impact the selection and revision process involving the writ of protection of constitutional rights[7] that every citizen has the right to file to protect his or her own fundamental rights without any cumbersome legal formality.

The court receives more than 2,000 writs of protection daily coming from all the judiciary benches across the country.[8] The Constitutional Court only has nine judges and under 200 employees who serve the entire country. In the selection process, the court decides which cases, due to their relevance and novelty, may need to be reviewed by one of these judges to decide whether or not to protect the allegedly violated fundamental rights.

Media coverage and social impact of Prometea

The media reported on Prometea in early 2019 when the Constitutional Court announced the conclusion of its pilot to help assistants and court personnel sort, read and retrieve key information from the hundreds of cases received to be reviewed. The coverage of Prometea was scarce, but had a significant impact.

There was a radio interview with the president of the Constitutional Court,[9] a very short segment about the system in a well-established TV news slot,[10] a piece in a major newspaper,[11] and two articles in a specialised magazine covering judicial and legal issues.[12] The most nuanced and critical pieces about Prometea appeared in the specialised magazine. They considered the needs of the court and alternatives to the proposed system, such as the redesign of the selection process, and noted that there was not much information available publicly about Prometea.[13] In contrast, the rest of the media was uncritical and plain when covering the pilot of Prometea.

Several points can be made about the media coverage of Prometea. First, the view that the selection process needed to be improved was shared by many actors and was in most cases the angle of the news stories about the system. However, there was no explanation about what Prometea was, what it does and how it does it. Second, in every instance it was presented as an "artificial intelligence" solution, the operational readiness of the system was highlighted, and the fact that it was employed in a pilot that only focused on the right to health was hardly mentioned. "With Prometea, the Constitutional Court finally enters the world of the highest informatics technology," said one TV report on the pilot.

The charismatic effect of AI systems was felt in the comments on Prometea by some law scholars. Grenfieth Sierra Cadena, head of public law research at the Universidad del Rosario, emphasised the importance of the project because it was the first time AI had been applied "in an executive and active way by a supreme court."[14] In the same piece, Cadena stated that "the court improved the management of writs of protection of constitutional rights related to the right to health by 900%," a number that apparently is calculated using the time that it takes the system to create documents, but does not include time saved in the process of selection. This is striking, as several parts of the software (i.e. the search functions and the automatic document generator) seem to be confused and mixed instead of clearly discerned for the purposes of evaluating its efficiency.

The media coverage showed two key characteristics: limited information and excessive enthusiasm. The news was clear enough: the Constitutional Court had adopted an AI-based solution

[6] Redacción Judicial. (2019, 5 February). Prometea, la nueva tecnología para la selección de tutelas en la Corte Constitucional. *El Espectador*. https://www.elespectador.com/noticias/judicial/prometea-la-nueva-tecnologia-para-seleccion-de-tutelas-en-la-corte-constitucional-articulo-838034

[7] The Spanish legal term used in Colombia is *acción de tutela*. It refers to the right to ask a judge to protect a fundamental right from an imminent threat when no other legal recourses are available. It is similar to what is known in other Latin American countries as a writ of *amparo*. See: https://en.wikipedia.org/wiki/Recurso_de_amparo

[8] https://twitter.com/CConstitucional/status/1141011549387153408; https://www.youtube.com/watch?v=r1ifDdWuW-k&feature=youtu.be; Corte Constitucional. (2019, 17 June). La Corte Constitucional está al día en la radicación de tutelas. www.corteconstitucional.gov.co/noticia.php?La-Corte-Constitucional-esta-al-dia-en-la-radicacion-de-tutelas.-8741

[9] laud.udistrital.edu.co/content/llega-inteligencia-artificial-para-agilizar-tutelas-en-salud

[10] https://youtu.be/DsmVL_Xybjo?t=2563

[11] Redacción Judicial. (2019, 5 February). Op. cit.

[12] Giraldo Gómez, J. (2019, 12 April). Prometea: ¿debe rediseñarse el proceso de selección de tutelas en la Corte Constitucional? *Ámbito Jurídico*. https://www.ambitojuridico.com/noticias/informe/constitucional-y-derechos-humanos/prometea-debe-redisenarse-el-proceso-de; Rivadaneira, J. C. (2019, 22 March). Prometea, inteligencia artificial para la revisión de tutelas en la Corte Constitucional. *Ámbito Jurídico*. https://www.ambitojuridico.com/noticias/informe/constitucional-y-derechos-humanos/prometea-inteligencia-artificial-para-la

[13] Giraldo Gómez, J. (2019, 12 April). Op. cit.

[14] Rivadaneira, J. C. (2019, 22 March). Op.cit.

to overcome the bottlenecks in the selection process. However, the lack of any further information or investigation by the media seems to be for three possible reasons: 1) a blind trust in information coming from the Constitutional Court, 2) concerns that looking deeper into the issue may be read as mistrust of the Court, and 3) the complexity of the issue, both for the journalist and for the reader.

This situation had a double-edged effect – muting the possibility of debate, and at the same time, stimulating the need for more information. While any possibility of public debate among the general population was defused, concerns were raised by groups interested in AI, mainly in civil society and universities. This increased the demand for further information and for opening a public debate about Prometea. The few events that took place in universities to try mitigate these demands only increased them. Cadena, acting as the promoter of Prometea at these events,[15] towed the same line taken by the media, and did not provide more technical information on the system.

This led to social and human rights concerns on three levels: 1) general concerns regarding AI, 2) concerns regarding the relationship between digital technology-based solutions and the Colombian context, and 3) and concerns regarding the selection process at the Constitutional Court. These concerns are related to transparency, privacy and data protection, and include issues to do with labour rights among other secondary concerns.

Prometea impacts transparency in two ways: first, regarding how a technological developer is selected in terms of the framework for public contracting, and second, how stakeholders, scholars and civil society organisations are allowed to participate in the process of discussing and deciding on technological solutions intended for the judicial system. As pointed out by scholars, a key factor in this is understanding exactly what the proposed technological solution is, and how it works.

It is also important to know the legal rationale behind the decision made by the Constitutional Court to implement the system, and how use of Prometea can be reconciled with normal protocols for motivating changes in the judicial system. The lack of transparency with regards to the AI decision-making process and how it may affect a citizen's right to due process is also worrisome. Furthermore, due process allows the writ of protection plaintiff to file an appeal in case he or she is not favoured in the Constitutional Court decision to review their case. A technological solution may put in danger the exercise of this right because it would be impossible for the plaintiff to argue against a machine's decision irrespective of how the algorithm and system work.

In terms of privacy and data protection, the main concern has to do with sensitive data being shared with third parties, such as a software developer. The fact that minors are involved in some cases, or others are to do with sexual crimes, among other situations that may require the anonymity of victims and their personal information or data, is considered critical. It is a breach of confidentiality for someone other than the judge and the parties involved in the processing of the case to access this information or data. It is especially worrying that a possible leak of personal data to the media or other third parties with an interest in the case can occur given the system's vulnerability in this respect, with irreversible consequences in terms of the protection of privacy for those involved in the case.

On labour rights, Prometea was seen as a replacement for those doing basic clerical activities, but by no means affecting more specialised work such as that done by judges or even judges' assistants. At the Constitutional Court this impacts on the so-called *ad honorem* system, training and experience for law students who help with writ of protection selection: reading files, writing abstracts, and classifying cases. This may also have an indirect impact on the future staffing at the court, as *ad honorem* is not only a training ground for future constitutionalists, but also a first step into more important positions at the court, such as judge's assistants.

Conclusion

After researching Prometea, we believe there are three areas of consideration when adopting AI-based solutions. First, the enthusiasm in adopting these systems overshadows the need to evaluate the human factors involved, both with regard to the skills needed to use the technologies, and issues to do with a reluctance to change. Second, the limitations of the technical infrastructure need to be considered; and third, the flexibility of processes from a legal point of view are important – or how they can be adapted properly to a technological platform.

[15] On March 12, a public event was held at Los Andes University to critically debate about the pilot due to the concerns of some professors there. The general characteristics of Prometea were presented (it was an AI solution, it included blockchain technology) but all questions, coming from both the public and the participants on the roundtable, regarding technical details of the pilot (dataset structure, type of algorithm, possible bias in the design or criteria selection, etc.) were eluded. A report on the event is available online in Spanish at: https://gecti.uniandes.edu.co/images/pdf/PROMETEA_EVENTO.pdf

With respect to how we adopt a technological solution in the Colombian judicial system, Prometea showed that the discussion is led by lawyers with few to no computer scientists or technical people involved. This inevitably results in a shallowness in the discussions due to the complexity of AI systems. The decision-making process focused on big and vaguely defined problems with the aim of solving these with an "AI solution".

Such concerns lead to questions that are still without clear answers. For instance, should the information which feeds an AI system be included in the concept of "data" according to the Colombian data protection law? Is it time to reassess or "update" our legal vocabulary and frameworks related to digital technologies? Should we think about other, more primary technological solutions such as the digitisation of judicial records instead of more questionable solutions such as AI systems?

As an alternative, some initiatives in Colombia like the Legal Design Lab at the Universidad de Los Andes[16] advocate for a more detailed definition of problems in order to address them with tailored actions.

Finally, when it comes to an AI solution being applied to a specific issue, the usual concerns relating to the use of technologies (the possibility of bias, the opaqueness of the technology, privacy concerns regarding data sets, etc.) were also raised in the case of Prometea. Due to the scarcity of information on Prometea, these concerns were largely speculative, given that it was not possible to understand if these issues were in fact present in this particular case.

Bearing this in mind, we believe that there are several factors that need to be taken into consideration when adopting AI-based solutions:

- Lawyers should be more involved in technical discussions and should advocate for multidisciplinary spaces for discussing the proposed technology.
- A "big problems – big solutions" enthusiasm in the approach should be replaced with a more grounded methodology based on detailed and complex definitions of problems and smaller, tailored proposals. Such methodologies should include an analysis of alternatives: AI systems are just one possibility among others.
- General contextual assessments such as a baseline study should be done, including the capacity of the staff to appropriate the technology, the actual capabilities of the technology, the flexibility of the legal framework to accommodate the technology, the need for technological training for legal practitioners, and the general technological culture and awareness of technology in society.
- More information is needed both for the general public and third parties such as other judicial systems in Latin America, including the Inter-American Commission on Human Rights, where Prometea is being promoted. More information is also needed to inform public policy decision-making processes that affect the administration of justice at the national level.

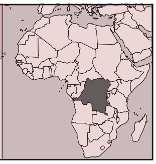
Action steps

The following advocacy steps are suggested for Colombia:

- Build interdisciplinary networks ideally involving universities and civil society organisations to enrich the debates around AI initiatives in Colombia.
- Promote transparency in the form of participatory processes as a prerequisite for any AI public initiative.
- Advocate for a detailed and down-to-earth definition of a problem to counter enthusiasm for ready-made AI solutions and to contribute to informed public debate on an issue.
- Request the presence of computer scientists, data scientists or similar experts in conjunction with scholars from human sciences working on technology issues in debates involving the use of AI in public initiatives.
- Fight the preconceptions of technological neutrality and technological determinism that prevent the critical analysis of any solution based on digital technologies.

[16] Legal Design Lab, Universidad de los Andes. Official Twitter account: https://twitter.com/legaldlab?lang=en

CONGO, DEMOCRATIC REPUBLIC OF



3D PRINTING GIVES HOPE TO AMPUTEES LIVING IN POVERTY IN THE DRC

Mesh Bukavu Network
Pacifique Zikomangane
www.meshbukavu.org

Introduction

Despite the presence in its soil of almost all the minerals used in the new technology industry, the Democratic Republic of Congo (DRC) is far behind in terms of new technology, particularly in the field of robotics, which is very advanced elsewhere in the world.

To this first paradox is added a second. Despite the existence of several universities in the country at which computer science studies are offered, students from these institutions are clearly out of step with several recent developments in their field of study. For example, this is the case with 3D printing technology, which is very popular today in the medical, aerospace and automotive sectors.

In this report, we will talk about the "3D Prosthesis Project",[1] which is a project to manufacture artificial limbs using a 3D printer for people who have had their limbs amputated. Given the level of automation and reprogrammability in 3D printing, many consider it a growing field of robotics engineering.[2] This is a project by the Institut Français de Bukavu (French Institute of Bukavu) in partnership with Ciriri Hospital in the same city. This is a first in the country not only in the field of robotics, but also in the medical field. The members of the Congolese team in charge of producing these prostheses were trained by Sano Celo,[3] a French start-up promoting 3D printing for health and development in Africa.

A high-risk environment

In 2013 it was estimated that there are 10.5 million people with disabilities in the DRC, or nearly 15% of the population.[4] This high statistic, which includes both physical and intellectual disabilities, is partly due to the fact that the Congolese population lives in a high-risk environment, characterised by armed conflict in parts of the country, regular earthquakes, disease, poorly controlled and dangerous road traffic, and even unregulated construction sites, all of which carry the potential risk of losing limbs. This risk is worsened by the deterioration of the health system throughout the country.

According to the National Road Accident Prevention Commission in the DRC, of the 12,500 accidents that occurred between 2007 and 2009, there were 1,400 deaths and 4,402 injuries.[5] While earthquakes are prevalent for those living in the provinces of North and South Kivu in the east of the country, regular reports also emerge of people being injured when houses or walls collapse on ad hoc construction sites that are largely uncontrolled by authorities.

Even if there is officially no civil war,[6] the country continues to record deaths and injuries linked to the activities of armed groups that are very present in the eastern part of its territory. People living in this part of the country, for example, face the threats of being shot or stepping on anti-personnel landmines. In 2014, the International Committee of the Red Cross working in the DRC said it was caring for 576 people with lower-limb amputations, all a result of the armed conflict.[7]

On 30 September 2015, the DRC acceded to the Convention on the Rights of Persons with Disabilities.[8] The purpose of this convention is to promote, protect and ensure the dignity, equality before the law, human rights and fundamental freedoms of people with disabilities of all kinds. Yet despite the work by groups such as the Red Cross, widespread poverty, coupled with a deterioration of the health system in the country, means that the care of patients poses many problems. People living with disabilities are abandoned to their plight and have enormous difficulties in caring for themselves.

[1] https://cd.ambafrance.org/L-Institut-francais-de-Bukavu-Halle-des-Grands-Lacs

[2] Harrop, J. (2015, 21 July). Are 3D printers robots? *IDTechEx*. https://www.idtechex.com/en/research-article/are-3d-printers-robots/8118

[3] https://www.facebook.com/profile.php?id=100011551902360

[4] http://www.adry.up.ac.za/index.php/2013-1-section-b-country-reports/republique-democratique-du-congo-rdc

[5] http://business-et-finances.com/les-accidents-de-circulation-en-baisse

[6] The DRC has experienced several wars since 1996, so it is difficult today to talk about a post-conflict situation.

[7] https://www.icrc.org/fr/document/republique-democratique-du-congo-5-000-personnes-handicapees-prises-en-charge-depuis-1998

[8] https://treaties.un.org/doc/Publication/CTC/Ch_IV_15.pdf

There is no national policy to support them, including in finding meaningful work or when it comes to other needs, such as public transport. For example, the construction of buildings in the country does not take into account people with reduced mobility. In these circumstances, equipping amputees with prostheses manufactured by a 3D printer can allow them to perform certain tasks they were unable to do in the past, and help to secure their independence.

From a simple competition to a concrete project

The 3D Prosthesis Project is a revolution in the fitting of people with lower or upper limb amputations with prostheses in the DRC. While 3D printers have been used in the country for some time, this is a first for the medical field. For example, in Kinshasa, the capital of the country, there is the Lisungi FabLab[9] digital manufacturing laboratory, where there is state-of-the-art technology and prototyping equipment to facilitate the implementation of ideas and promote the acquisition of skills and knowledge through practice, using digital technology.[10] However, this laboratory has not done work linked to the medical field.

Where did the idea come from to manufacture the prostheses using a 3D printer? It all began in October 2017 with a robotic "hackathon" organised by the French Institute of Goma, which brought together 85 young Congolese engineers, computer scientists, doctors and designers together with specialists from Sano Celo. By the end of the hackathon, a hand prosthesis had been produced using 3D printing. This prosthesis is now used by a young person in Goma who was the victim of a traffic accident.[11]

The success of the robotic hackathon quickly fuelled new ambitions. In 2018, the French Institute of Bukavu set up the 3D Prothesis Project in collaboration with Ciriri Hospital and Sano Celo. The objective of this project is to provide patients (amputees) with prosthetic limbs at a lower cost.

It is these three partners who have given shape to the project, each with a specific role. The French Institute offers a training framework and laboratory as well as equipment including printers and PLA[12] and PVA[13] from France. Sano Celo provides training on the manufacturing process of producing prostheses using the 3D printer. Ciriri Hospital sends amputee patients in need of prostheses to the laboratory.

The doctor-patient-laboratory process

The surgeons at Ciriri Hospital identify the patients, then send the measurements of the limbs that have been amputated to the laboratory, which in turn designs the prostheses. All the post-operation activities, from manufacturing to fitting, are carried out in the laboratory.

It is important to note that centres for people with disabilities in the DRC have existed since 1960, where traditional prostheses are manufactured for patients who have lost their upper or lower limbs. Among these centres are the Centre for the Physically Disabled, "Shirika la Umoja", in Goma[14] and "Heri Kwetu" in Bukavu.[15] The prostheses manufactured in these centres are relatively expensive compared to the means of the patients, most of whom come from poor families. They are also static, i.e. they do not make it easy for the wearer to move when using them.

Although the prostheses manufactured in the laboratory at the French Institute of Bukavu are not robotic prostheses, their design allows wearers more mobility. "With a hand prosthesis made using the 3D printer, the patient can lift a key ring or a bottle, and even if [he or she] cannot lift heavy objects, [he or she] can make movements," according to Dr. Flory Cubaka,[16] a surgeon at Ciriri Hospital. They are also currently given free of charge to patients.

According to Charles Bulabula,[17] a computer scientist at the laboratory, three types of prostheses are manufactured there, namely hand, leg and finger prostheses. The average time for the manufacture of a prosthesis using the 3D printer is 48 hours. The manufacture of finger prostheses is a first in the DRC. Until now, the various centres for people with disabilities in the country have only manufactured leg and hand prostheses, but not finger prostheses. In other words, the manufacture of finger prostheses is another particularity of this project, because now people whose fingers have been amputated can benefit from artificial fingers, which was not possible in the country before the project started.

Scaling up the 3D Prosthesis Project

Currently the project only treats patients from Ciriri Hospital, but it plans to expand to other hospitals

[9] https://www.lisungifablab.org

[10] http://www.onerdc.net/?navigation=ar&id=1524

[11] https://www.impactmag.info/un-hackathon-robotique-a-goma

[12] PLA (polylactic acid) is a plastic that is fully biodegradable under industrial conditions and particularly popular in the fields of food packaging, plastic bags and 3D printing.

[13] PVA (polyvinyl alcohol) is a water-soluble synthetic resin used in 3D printing.

[14] https://www.umoja.be/umoja/le-centre-handicap%C3%A9s-phys-chp

[15] http://www.herikwetu.org/fr/qui-nous-sommes

[16] Interviewed for this report on 22 June, 2019 in Bukavu.

[17] IT consultant and intern at the 3D Prothesis Project of the French Institute of Bukavu.

in the near future. This is why an application has already been set up to improve project management. The so-called “3D Application” is a platform designed by Charles Bulabula that allows surgeons from different hospitals to send the measurements of the prostheses to the laboratory, follow up with the patients who receive prostheses, and get feedback from the patients.

Users of the application will be hospital surgeons and patients registered on the platform, as well as members of the project team. To access it they will have to have a PC or a mobile phone connected to the internet.

Because of the poor quality of the internet and its low penetration rate in the DRC, Bulabula is currently working on the possibility of providing access to this platform by 2020 even if users are not connected to the internet and do not have smartphones. Users will only need to have a simple mobile phone and be in an area covered by the cellular network to connect to the platform. Because of this, patients and surgeons from village hospitals will have access to the platform and will be able to benefit from the project.

Despite the hopes raised by the 3D Prothesis Project, Blaise Bulonza, who is disabled and a coordinator of the Initiative for a Better Future (INAM),[18] believes that for this project to be sustainable, the Congolese government must be involved in subsidising the manufacture of the prostheses. His remark is nothing more than a simple call to the Congolese state to respect the constitutional spirit according to which people with disabilities are entitled to specific protective measures in relation to their physical and intellectual needs and rights.[19]

Conclusion

The arrival and use of the 3D printer in the field of health in the DRC is a major advance in that it now allows patients with amputated hands, arms or fingers to access the benefits of modern technology, allowing them more mobility. However, it should be pointed out that the impact of this project is still minimal in relation to the needs of amputees. At Ciriri Hospital alone, at least one to two patients each week have a limb amputated.[20] At the same time, the environment conducive to accidents and even diseases that lead to the amputation of patients’ limbs is far from being low-risk. Entire villages in eastern DRC still contain anti-personnel mines that continue to claim victims. Quality health care is not yet accessible to the majority of the Congolese population, who, due to a lack of resources, do not go to hospital or go there late, increasing the risk of having a limb amputated.

How the project has been set up also does not assure its longevity. The manufacture of prostheses using the 3D printer is entirely handled by the French Institute of Bukavu, yet each project has a beginning and an end, which means that the day the institute cuts its support, the project could come to a standstill. The training of Congolese surgeons and young computer scientists by the French Institute on the use of the 3D printer in the health field is a good thing for the sustainability of the laboratory, but it is not enough, because the raw materials for printing prostheses come from France via the French Institute.

Although the majority of the Congolese population is poor, there is reason to doubt the project’s ability to continue to offer the prostheses to all patients free of charge, especially when the project scales up to include many other hospitals.

Action steps

The following steps are recommended for the 3D Prothesis Project:

- The French Institute of Bukavu should involve the Congolese government in the project through the provincial health division of South Kivu, and request, for example, the exemption of taxes and tariffs on imported products and parts used for the manufacture of prostheses for amputees.
- The laboratory at the French Institute of Bukavu should also collaborate with Congolese universities that have computer departments, so that their students can be trained in the use of the 3D printer, especially in the health field. While this could help with the sustainability of the project, trained students will also be able to continue this project in other forms and exploit other opportunities that the 3D printer offers.
- The institute should offer the 3D Prothesis Project team the possibility of carrying out study trips to France, for example, to see how French start-ups operating in the field of artificial intelligence are oriented towards health work.
- The project team should pay particular attention to feedback from patients who are already equipped with its prostheses to understand their experiences and how they adapt to their new prostheses, in order to improve them if necessary.

[18] INAM is a human rights organisation for people with disabilities based in Bukavu.

[19] Article 49 of the DRC’s constitution.

[20] Dr Flory Cubaka, surgeon at Ciriri Hospital, during an interview with us on 22 June 2019 in Bukavu.

COSTA RICA

AI APPLICATIONS TO HEALTH DATA AND CHALLENGES FOR THE RIGHT TO HEALTH IN COSTA RICA

Cooperativa Sulá Batsú
Kemly Camacho and Christian Hidalgo
www.sulabatsu.com; edus.ccss.sa.cr

Introduction

Costa Rica has a public universal health care system run by the *Caja Costarricense del Seguro Social* (Costa Rican Social Security Fund – CCSS), an institution that has been serving the country's citizens for more than 75 years. It has a large structure with a total of 55,000 employees covering the entire Costa Rican territory.

Eight years ago, the CCSS began to develop the Unified Digital Health Record (*Expediente Digital Único en Salud* – EDUS). It consists of a set of applications and services that allows the automation of all health service delivery processes. The objective was to improve the quality of integrated health care for Costa Ricans by providing health professionals with easy access to all of a patient's medical information and history.

We must understand that EDUS is not only a technical system but also a development programme that impacts positively on society in relation to the right to health. Using technology, it transforms and develops health service delivery institutions and processes with public funds. The richness of its data, which includes the medical histories of individuals and their families, in our view makes it one of the most important information resources in the country.

As the data collected by EDUS is expanded, artificial intelligence (AI) tools could play an important role in the exercise of the right to health in Costa Rica. However, the use of AI can introduce many risks too.

Context

To understand the potential of AI applications when it comes to EDUS, is important to highlight the following:

- EDUS applications are developed by the public sector, with public funds, and tailored to the Costa Rican population's needs. In other words, the Unified Digital Health Record can be considered a public good for people in Costa Rica.
- The Costa Rican health system is universal. This means that the databases managed by EDUS would contain all the information and health data of all inhabitants who have used the health system. Practically, this is nearly everyone in Costa Rica, including migrants of various nationalities. It is important to note that these databases are maintained in a private data centre.
- EDUS was fully implemented at the time we wrote this article. Currently its functionality includes digital filing systems, family health history and an appointment system, while additional functionalities include applications for emergency admissions or hospitalisation, links to pharmacies, and sharing medical images such as x-rays, ultrasounds and tomographies. This information is crucial in understanding the population's health profile. EDUS has data on medical records, prescriptions, laboratory tests, medical images, and family health histories, among others.
- Costa Rica has a strong National Law for the Protection of Inhabitants' Personal Data[1] and a national data protection agency.[2]

In previous reports, including our report for Global Information Society Watch (GISWatch) 2014,[3] we analysed the risks of the EDUS platform in terms of data security. As the EDUS databases grow, the platform is becoming more and more interesting for private actors in the health sector (insurers, private clinics, pharmaceutical companies, etc.). Because of this, there is an urgent need to have better data privacy strategies and policies for the EDUS system.

[1] https://www.prodhab.go.cr//reformas

[2] www.prodhab.go.cr

[3] Camacho, K., & Sánchez, A. (2014). Universal health data in Costa Rica: The potential for surveillance from a human rights perspective. In A. Finlay (Ed.), *Global Information Society Watch 2014: Communications surveillance in the digital age.* APC & Hivos. https://www.giswatch.org/en/country-report/communications-surveillance/costa-rica

The right to health and AI

The use of AI on the EDUS platform can make an important contribution to improving universal health care as envisaged by the CCSS. According to an EDUS staff member:

> With data and AI applications, I believe that for the first time we can generate population profiles that can give value and real information to the medical sector, not only based on studies from other countries, but based on our reality. For example, we can predict the onset of a disease, its evolution, and the relationship between this and the habits of our population, etc.[4]

In this regard, over the next few years, priority will be given to actions aimed at the prevention and early care of chronic diseases that affect a majority of the Costa Rican population, as well as the decision-making process to strengthen the health infrastructure that responds to citizen health profiles. For this reason, one of the most important lines of work using AI on the EDUS platform will be the prediction of future health profiles from the analysis of existing data.

This has already started with the use of AI to analyse case studies, such as those documenting Type 2 diabetes mellitus. Through the application of AI, the future behaviour of this chronic disease can be predicted using existing historical information gathered over the last four years through the EDUS platform. The analysis seeks to identify groups of people who are at risk but are not yet sick, and promote strategies that allow prevention in the present, so that the disease does not manifest in these people in the future as predictions have indicated.

Another example where AI has been used in conjunction with EDUS is in the area of infrastructure. From the application of AI, it has been possible to identify potential public health challenges that will need attention over the next five years. For example, we are able to understand what population groups, given specific characteristics, face the challenge of teenage pregnancies, drug use, or mental illnesses, among others. This allows health authorities to build the necessary infrastructure and spaces to meet the future needs of these challenges, even before they have become evident.

For this, CCSS personnel have been trained in exploratory and predictive methods, building capacity in public health institutions for data mining and AI-enabled analysis techniques.

However, there are challenges to applying AI to the EDUS platform. The first one is related to the National Law for the Protection of Inhabitants' Personal Data, which effectively qualifies EDUS databases as containing sensitive data, and therefore regulates and restricts their use. For this reason, the full potential of EDUS cannot be exploited for disease prediction using AI applications. This is an important area for national discussion: whether or not to create a special legal status for EDUS data and the use of AI applications in the health system without violating the individual rights and privacy of those using the system.

Secondly, although EDUS offers access to various datasets, the quality of this data needs to be ensured. This cannot yet be guaranteed since it is necessary to strengthen the processes and procedures that guarantee the quality of data. It is extremely risky to work with AI-based predictions if the algorithms use data that has not passed a rigorous quality process, because this can generate inaccurate predictions, result in wrong decisions, waste public health investments and present huge health risks.

Given the appropriate training processes and regulation of personal data respectful of personal rights, and with databases that guarantee the quality of the data, AI applications using EDUS envisaged in the very near future include "smart pharmacies".

The development of smart pharmacies, where the entire process of preparation and distribution of medicines within a clinic or hospital would be automated, is based on the prescriptions issued to patients over time. Automated distribution of medicines to hospitalised persons using robots is also being contemplated.

The first stage has already begun to be developed at the Hospital de Heredia (Heredia is one of the provinces of the country) in a pilot project using a system called E-flow that has reduced delivery times. The medications likely to be required by a patient are automatically determined by means of predictive AI applications.

E-flow also analyses the operations of the public health pharmacy system, so that the headquarters can make decisions to transfer, at any given time, human resources to those tasks that had not previously been prioritised due to a lack of staff. For example, these might include educational and information campaigns on some of the topics that the hospital has defined as a priority for its patients. However, even if the need for resources is known, this may not always be possible. As Dr. Bastos, who works with EDUS, explained:

> Artificial intelligence in relation to the prognosis of diseases can generate a breakthrough in making decisions in advance of the event and

[4] Interview with EDUS staff member conducted for this report.

> being able to prepare the health system before it happens. However, the questions to discuss would be: If I know how many and who are the future patients and I do not have enough resources, what decision should I make? Should all diseases be analysed or should I decide that some diseases should not be analysed? With the data I have, how will I respond to create real changes in public health? How do I follow up on patients?[5]

Bastos highlights a fundamental aspect of AI: the relationship between the results it determines, the decisions that need to be made by public institutions, and the availability of resources to respond to these needs that have been determined by the application. In terms of the right to health, this discussion is very important: the solutions proposed by AI applications may require decision making and action at a speed greater than the capacity of public institutions. This is a huge challenge and an important risk to highlight.

Another risk is an over-reliance on AI. Bastos told us:

> In relation to robotics and AI, the challenge is to include processes that are not going to risk patient care. It should not be overlooked that every robot can potentially replace a human process. But robotics should rather strengthen the capacities and abilities of human beings to improve health care, as well as open up the new possibilities of employment or new positions that this can generate.

She raises several highly relevant aspects related to the exercise of the right to health when it comes to the human-robot relationship that will undoubtedly become more pressing questions over time. For example, which tasks should be replaced by AI applications and robots, and which should not? Is it possible to do this in a rights-based framework? What new rights need to be defined when humans interact with robots? These questions are already necessary with the use of chatbots for automated health consultations. For instance, are bots limiting the exercise of the right to health by having a limited number of preset responses for their users?

At the same time, while an important transformation in the health sector is seen through the incorporation of AI, this raises further important issues for health workers and practising professionals, including with regard to their training needs.

How will the work of health practitioners change given the use of AI and robots? Will some jobs in the health sector become precarious? What new forms of training are required? Moreover, what responsibilities will be assumed in terms of health decisions? Who will assume responsibility when mistakes are made by robots or algorithms? Will the companies who made the robots or the developers of the AI be liable? Or will the staff interacting with the AI be responsible?

When EDUS was integrated into the CCSS, an important process of change management had to be carried out with its personnel. The transformations that AI and robotics can produce in the health sector are even greater, so the adaptation of the public and universal health system and its staff to this new reality is something that should be reflected and acted upon now.

Conclusions

The case of EDUS in Costa Rica and the potential applications of AI are very interesting.

EDUS is an excellent system for the development of AI applications that strengthen the provision of health services.

These are AI applications that will be developed for the health sector – and therefore be a public good – using a database that is a public good as well. This is positive for the exercise of the right to health and is an example for the health systems in other countries.

At the same time, this implies the need for public action so that the applications of AI in health using EDUS have a legal basis, and include processes and procedures that guarantee not only the right to health, but also the public custody of the results generated by AI and the protection of national health data.

It is necessary to prepare the public health system for the incorporation of AI. It must be considered as a public management strategy with all that this implies. This should not simply be considered an issue of technological development, but also as the development of the health system as a whole, including changes to training and to work profiles, and how and under what conditions health services are delivered.

As the EDUS team indicates, "as long as the technology is used for the improvement of the health of the population and with the consent of the users, it can be a fundamental tool for the exercise of the right to health that has been promoted since the Costa Rican constitution was written."

[5] Interview conducted for this report.

Action steps

The following action steps are important in Costa Rica:

- It is necessary to incorporate key actors and citizen organisations in the discussion about the advantages and challenges of incorporating AI into public health. It is an extremely important issue for the exercise of the right to health that should not be treated only as a technological development.
- We propose that constructive, collective spaces must be established including a diversity of actors working on the right to health. These should feed into national decision-making forums deliberating public investment and the application of AI in the health sector.
- It is also necessary to demystify the technology and to make it understandable to the general population and to health personnel.
- It is important to continue developing actions to ensure the security of EDUS data, especially data that has already been collected for more than four years. It is currently in the custody of the CCSS, but with few digital security standards.

ECUADOR

THE USE OF SOCIAL MEDIA AND AI TO SHAPE DEMOCRACY IN ECUADOR

Pontifical Catholic University of Ecuador
María José Calderón
www.puce.edu.ec

Introduction

Latin America has a long history of surveillance states supported by weak democracies. Strategies for authoritarian governance have always included spying on civilians and political opponents. While technology has helped society become part of an interconnected world, it has also led to more intensive surveillance practices.[1]

This report discusses the use of artificial intelligence (AI) for profiling the electorate and tailoring campaign messages using social media. Taking the 2019 elections in the capital Quito as an example, it argues that AI is being used to polarise society as a political strategy to get elected. There is, as a result, a lack of real information available to inform the voting public.

This potential to influence election outcomes is increased when voting is mandatory, and voter turn-out is therefore largely predictable. By simultaneously drawing on government data gathered about citizens, AI becomes a powerful influencer in shaping democracy. However, this undermines the constitutional right to free and fair elections, as the rule of law is overthrown by the rule of Twitter, Facebook and WhatsApp.

A surveillance state

The International Principles on the Application of Human Rights to Communications Surveillance[2] (adopted in 2014) define the concept of "communications surveillance" as the process of "monitoring, intercepting, collecting, obtaining, analysing, using, preserving, retaining, interfering with, accessing or similar actions taken with regard to information that includes, reflects, arises from or is about a person's communications in the past, present, or future."[3] The problem starts when surveillance technologies are available and used when the state's basic survival is at risk. Ecuador was one of the first countries in the region to use surveillance technology, and was the first in the region to adopt facial recognition technologies.[4] In a context where social media platforms like Facebook and Twitter have become tools for social and political control, Ecuador also has one of the many governments that spend significant resources and employ large numbers of people to generate content online in an attempt to shape the opinions of both local and foreign audiences.

What is known as "public information" involves private data, which is collected and stored on an ongoing basis by state agencies. The state now has an accurate x-ray of the population in Ecuador – almost 18 million people as of 2018 – including, through the monitoring of social media, of their desires.

There is a dramatic increase of the surveillance of communications by states. This is happening without adequate transparency, nor information being made available to people about the surveillance or use of the data. Communications metadata and regular content displayed on a citizen's social media profiles create a detailed picture of an individual's life, including medical conditions, political and religious viewpoints, associations, interactions and interests.

Like most new technologies, AI has the potential to increase existing problems, reinforce structural inequalities, and superimpose biases. There is also potential for good. However, what we see in countries that experience constant political protests and unrest, which are part of democracy, is a great potential for intrusion into a citizen's life and the chilling effects this has on how democracy is enacted.

[1] As argued, there is a long tradition of surveillance as part of a government strategy to ensure its own legitimacy. Foucault's concepts of governmentality and biopolitics shape every aspect of our lives. See, for example, Hope, A. (2015). Governmentality and the 'Selling' of School Surveillance Devices. *The Sociological Review, 63*(4), 840-857. https://doi.org/10.1111/1467-954X.12279; Holmer Nadesan, M. (2010). *Governmentality, Biopower, and Everyday Life*. New York: Routledge; and Rodríguez, K. (2017, 2 January). Surveillance in Latin America: 2016 in Review. *Electronic Frontier Foundation*. https://www.eff.org/deeplinks/2016/12/surveillance-latin-america

[2] https://necessaryandproportionate.org/principles

[3] Ibid.

[4] Human Rights Watch has issued several reports on the dangers of surveillance technologies such as face recognition. See, for example, the joint letter to Google published on 15 January 2019 at: https://www.hrw.org/news/2019/01/15/letter-google-face-surveillance-technology

The Ecuadorian government purchased the latest surveillance technology from China in 2014.[5] While, as mentioned, Ecuador was the first Andean country to have a facial recognition system in place, the extent of its surveillance only become apparent when several newspaper reports were published, particularly in *The New York Times*.[6]

The use of social media for propaganda: A threat to democracy

During Rafael Correa's government (2007-2017), trolling became a standard practice on social media and became a part of the public institutional design.[7] There are reports on state-sponsored troll farms that were initially fostered by Correa's government. They revealed efforts to skew public opinion in favour of its policies and actions for more than a decade. Correa and his supporters even continued this practice after Lenin Moreno's election in 2017. For example, the Secretariat of Communication said that institutional Twitter and Facebook accounts created by the former government, such as Enlace Ciudadano (Citizen's Link), were used to disseminate information that was not authorised by the current administration.[8] Political campaigns in Ecuador are tightly controlled by the Electoral Council that oversees all political campaigning using traditional media. The law governing elections and political campaigning requires political actors to stop campaigning two days before an election.[9]

However, social media and the internet are excluded from these regulations. The exclusion of social media is important for freedom of expression, but it also allows hate speech among candidates to proliferate and the dissemination of disinformation. This is an important deficiency in the law in a country where internet penetration is around 70%, and about 90% of Ecuadorians use a mobile phone.

"Cyber troops" or troll farms are government, military or political party sponsored individuals or groups committed to manipulating public opinion using social media. Added to this is the potential to surveil social media and to build maps of the political and other preferences of the electorate, based on where and how they live and their socioeconomic demographics.

These tactics were also used in elections in countries such as Brazil and the United States with some success by right-wing politicians who managed to fuel the socio-political differences in their countries. The amount of data compiled on social media has inevitably worked to the detriment of civil rights and is now dangerously available without any control or civil supervision. Pattern recognition technology has made it easy for government officials to respond quickly to political crises, or to election result forecasts, in a targeted and effective way.

The 2019 mayoral elections in Quito

With some exceptions, voting is mandatory in Ecuador.[10] Since 2009, electoral districts have been clearly demarcated, together with citizen registration information. Although gerrymandering is not legal, the demographic divisions of these districts allow for focused campaigning and influencing of the electorate – a potential which is now substantially greater given the surveillance of social media. Social media has become an invaluable open source resource for researchers and political advisors.

Quito, the country's capital, is one of the largest electoral districts, with over two million registered voters. What happened in the 2019 mayoral elections in the city is the latest evidence of how the mixture of data mining and machine learning can determine the outcome of an election, given mandatory voting and the detailed knowledge of the demographics and desires of voters in districts. Table 1 presents data from the 2019 elections in Quito, where the electoral outcome shows a narrow percentage of votes among all mayoral candidates.

[5] Rollet, C. (2018, 9 August). Ecuador's All-Seeing Eye Is Made in China. *Foreign Policy*. https://foreignpolicy.com/2018/08/09/ecuadors-all-seeing-eye-is-made-in-china

[6] *The New York Times* has written several pieces of investigative journalism about surveillance in Ecuador – see, for example, Kessel, J. M. (2019, 26 April). In a Secret Bunker in the Andes, a Wall That Was Really a Window. *The New York Times*. https://www.nytimes.com/2019/04/26/reader-center/ecuador-china-surveillance-spying.html; Buzzfeed also shared a report about Correa's surveillance practices: Gray, R., & Carrasquillo, A. (2013, 25 June). Exclusive: Documents Illuminate Ecuador's Spying Practices. *BuzzFeed News*. https://www.buzzfeednews.com/article/rosiegray/exclusive-documents-illuminate-ecuadors-spying-practices

[7] http://milhojas.is/612261-troll-center-derroche-y-acoso-desde-las-redes-sociales.html

[8] Details about trolling can be found in the Freedom on the Net report for Ecuador by Freedom House. The information dates back to 2015, when these reports were initially issued: https://freedomhouse.org/report/freedom-net/2018/ecuador; see also Nyst, C., & Monaco, N. (2018). *State-Sponsored Trolling: How Governments Are Deploying Disinformation as Part of Broader Digital Harassment Campaigns*. Palo Alto: Institute for the Future. https://www.iftf.org/fileadmin/user_upload/images/DigIntel/IFTF_State_sponsored_trolling_report.pdf

[9] The Organic Electoral Law states: "Forty-eight hours before the day of the elections and until 5:00 pm on the day of voting, the dissemination of any type of information provided by public institutions is prohibited." This includes publicity, opinions or images published by traditional media outlets that encourage voters to favour a particular party. Failure to comply with these provisions results in penalties according to article 277 of the same law. https://docs.ecuador.justia.com/nacionales/leyes/ley-electoral.pdf

[10] Voting is optional for Ecuadorians between 16 and 18 years of age or over 65, people living abroad, members of the armed forces and national police in active service, people with disabilities, illiterate people, and foreigners aged 16 and up who have resided legally in the country for at least five years and are registered to vote (Article 207, Organic Electoral Law).

TABLE 1.

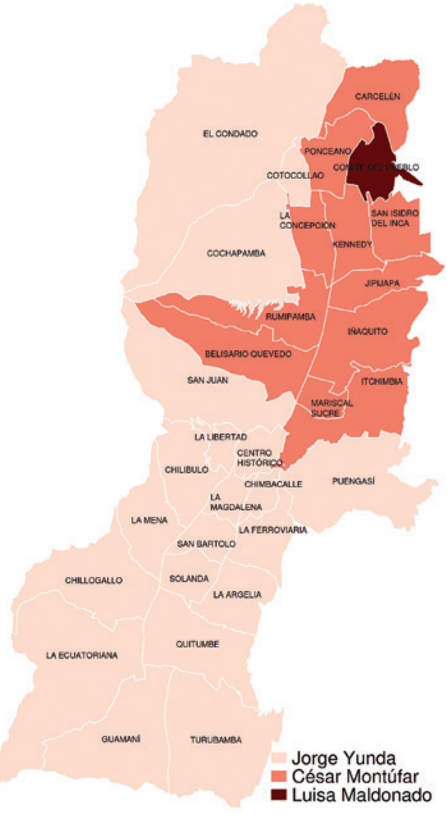
Results of the 2019 mayoral election in Quito

Candidate	Number of votes	Percentage
Jorge Yunda	296,096	21.39%
Luisa Maldonado	255,007	18.42%
Cesar Montúfar	234,442	16.93%
Paco Moncayo	246,142 (higher turnout due to the electorate from rural parishes)	17.78%

These percentages are represented in Figure 1, which shows how the different neighbourhoods in Quito ended up voting along economic divides in the city. For example, the main north and central district is known as a middle- and upper-class neighbourhood, while the south and the northernmost part of Quito are known as working-class neighbourhoods with some areas that have distinctive poverty issues. All of the neighbourhoods were led by the three major candidates.

FIGURE 1.

Electoral turnout in the city of Quito by neighbourhood



Jorge Yunda was initially a candidate from the ruling party (Alianza País) and then ran with another party (Unión Ecuatoriana) that supported Alianza País. Eventually, he won the election with just 21.39% of the vote overall. While this can suggest problems with the electoral divisions, and the fact that many candidates were in the race for the position, social media information dissemination strategies by the candidates made it easier for the contenders to succeed by a very narrow margin. The clever use of social media for political campaigning proved effective for Yunda, who benefited from a very fragmented electorate.

Micro segmentation of audiences has been a long-time political strategy when it comes to online media – and there are numerous applications that make this possible.[11] Segmentation offers a targeted profile of the type of person to whom an advertisement can be directed. This allows for focused campaigning in the context of a widely fragmented electorate. Data mining is easy when there are several different kinds of databases available to political actors. There is a wide array of user-friendly data mining apps available to amateur political advisers and even community managers of politicians. Most of them offer a varied range of information regarding the public's use of social media. Both segmentation and data mining are used effectively when cross-referenced with mandatory voting where voter turnout can be largely predicted.[12]

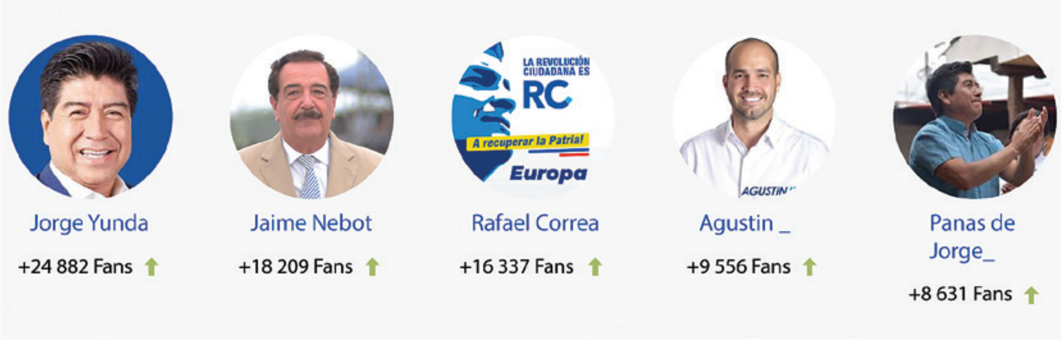
The election of Yunda as the mayor with only 296,096 votes is a clear indicator that a clever combination of data mining and machine learning can determine the outcome of an election. The question is raised, however, whether this results in a legitimate election.

[11] Pedro-Carañana, J., Broudy, D., & Klaehn, J. (Eds.) (2018). *The Propaganda Model Today: Filtering Perception and Awareness.* University of Westminster Press.

[12] Apps such as TweetDeck, Twitonomy and others. Each of them offers a paid option with more features. AI and machine learning are not as easy, but most training algorithms are available to the public with the use of R, a programming language and environment used for statistical computing, and its numerous support communities.

FIGURE 2.

Fastest-growing politics pages on Facebook during the second week of March 2019



Jorge Yunda	+24 882 Fans ↑
Jaime Nebot	+18 209 Fans ↑
Rafael Correa	+16 337 Fans ↑
Agustin _	+9 556 Fans ↑
Panas de Jorge_	+8 631 Fans ↑

Source: Facebook

Conclusion

Mandatory voting combined with the intensive use of machine learning and data mining has become fundamental to political campaigning in Ecuador. It has made voting patterns more predictable, and therefore more open to manipulation. The Quito mayoral elections are a cautionary tale of how the source of citizens' information is used and the dangers of having it moulded according to someone else's agenda.

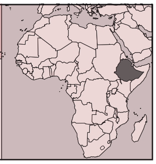
Action steps

Appropriate data laws are necessary for citizens to take back control of their data, and to control the use of AI and data mining that impacts on their everyday lives. In particular, laws that govern the use of social media and personal online data for electoral campaigns need to be developed. Here the International Principles on the Application of Human Rights to Communications Surveillance, which lay the foundation for a framework for privacy, should be drawn on.[13] Unfortunately, they are inadequate due to the immense amount of data harnessed by the state.

Awareness needs to be raised so that citizens understand how critical political discourse is shaped by political figures, particularly during an election. As it is, the system allows for citizens to be manipulated, and requires us to be constantly vigilant in order to overcome the attempts to control and shape our own political viewpoints, and ultimately to control democracy.

[13] https://necessaryandproportionate.org/principles

ETHIOPIA



THE THRIVING AI LANDSCAPE IN ETHIOPIA: ITS IMPLICATIONS FOR HUMAN RIGHTS, SOCIAL JUSTICE AND DEVELOPMENT

Internet Society – Ethiopia Chapter (under formation)
Abebe Chekol
abechekol@yahoo.com

Introduction

Ethiopia, with a population of over 105 million people (2017), is the second most populous nation in Africa after Nigeria, and the fastest growing economy in the region. It is, however, also one of the poorest countries, with a per capita income of USD 783, while aiming to reach lower-middle-income status by 2025. This entails significant investment in, among others, energy, transport, quality basic services, and accelerating agro-based industrialisation by expanding the role of the private sector. To this effect, the role of information and communications technologies (ICTs) is important. However, access to ICTs is one of the lowest, with 37.2% penetration of mobile-cellular subscriptions, and 13.9% and 18.6% penetration of mobile broadband and individuals using the internet, respectively.[1]

Despite the lack of an enabling environment, the artificial intelligence (AI) sector has been a growing phenomenon in Ethiopia over the last five years, with various reports indicating the country is becoming a thriving centre for AI research and development including robotics. This is evident from various initiatives happening, ranging from AI-enabled applications and services to the development of AI-powered robots. While there are promising signs of interest in AI development towards positively contributing to socioeconomic development, there has also been evidence of using the capabilities of AI adversely to violate human rights. In this context, this report tries to explore the thriving AI landscape in Ethiopia to identify its positive contribution to building a better society, as well as how it adversely affects privacy, data protection and social justice issues.

Setting the scene

Being an area of computer science, AI is devoted to developing systems that can be taught or learn to make decisions and predictions within specific contexts.[2] The AI currently in use across the world is broadly categorised as "narrow AI". Narrow AI involves single-task applications for uses such as image recognition, language translation and autonomous vehicles. In contrast, what is known as "artificial general intelligence" refers to systems that exhibit intelligent behaviour across a range of cognitive tasks that are not anticipated to be achieved for at least decades.[3]

AI is creating an increasing range of new services, products and value-adds in various sectors. AI applications can perform a wide number of intelligent behaviours: optimisation (e.g. in supply chains); pattern recognition and detection (e.g. facial recognition); prediction and hypothesis testing (e.g. predicting disease outbreaks); natural language processing; and machine translation.[4]

For AI to take root, the digital components supporting the AI landscape such as the internet of things (IoT), cloud computing, broadband and connectivity, and big data need to be developed in countries like Ethiopia. While there is growing potential for large data sets from commerce, social media, science and other sources to become available, 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. In this regard, the government of Ethiopia has recognised the importance of data and commissioned the drafting of the National Open Data Policy for the Government of Ethiopia[5] in January 2018, which is yet to be approved by the parliament.

Ethiopia has also recognised privacy throughout its constitutional history. The most comprehensive privacy safeguards statement was introduced in the constitution of 1995[6] which protects the priva-

[1] https://www.itu.int/net4/itu-d/icteye/CountryProfileReport.aspx?countryID=77

[2] Smith, M., & Neupane, S. (2018). *Artificial intelligence and human development: Toward a research agenda*. Ottawa: IDRC. https://idl-bnc-idrc.dspacedirect.org/handle/10625/56949

[3] Access Now. (2018). *Human Rights in the Age of Artificial Intelligence*. https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf

[4] Ibid.

[5] Ministry of Communication and Information Technology (MCIT). (2018). Consultation on the Recommendations and Working Text of the National Open Data Policy of the Government of Ethiopia. Addis Ababa: MCIT. www.mcit.gov.et/web/guest/-/draft-open-data-policy-and-guideline

[6] https://www.wipo.int/edocs/lexdocs/laws/en/et/et007en.pdf

cy of persons, their homes and correspondence in a detailed manner. However, this has been undermined in the past almost two decades due to the introduction of some unfriendly laws such as the anti-terrorism law, the freedom of mass media and information law, and the computer crime proclamation. Since April 2018, however, with the change of leadership in the Ethiopian People's Revolutionary Democratic Front (the ruling party) and the new administration of Prime Minister Abiy Ahmed, many positive reforms, including lifting bans on some media outlets and unblocking over 250 websites, together with initiatives to review and revise the legislation that led to these restrictions, have been seen.

Seizing the AI opportunity in Ethiopia and its implications

Ethiopia is one of the few African countries to deploy AI solutions at scale. It started small, unlike countries in Africa that attracted global technology giants such as Google's AI research lab in Accra, Ghana, IBM's AI-oriented research labs in Kenya and South Africa, and Facebook's African technology hub in Lagos, Nigeria. However, young Ethiopian AI pioneers are still making their mark in global projects.[7] One such initiative is the iCog Labs Software Consultancy,[8] which is an Addis Ababa-based research and development company collaborating with international AI research groups and providing services to customers around the world. The core speciality of iCog Labs includes machine learning-based data analytics, computational linguistics, computer vision, mobile robots and cognitive robotics, and cognitive architecture, and it has a vision that looks towards the widespread introduction of artificial general intelligence.

iCog Labs was launched in 2013 with USD 50,000 capital and four programmers, including the founder and chief executive officer of SinguarityNET,[9] a global AI marketplace, and chief scientist of Hanson Robotics, the Hong Kong-based engineering and robotics company known for its development of the interactive humanoid robot "Sophia".[10] One of the achievements of this ambitious company is its involvement in more than half of the software programming of Sophia.[11] Furthermore, one of its flagship projects is Solve IT[12] (a pun on "solve it") which is a nationwide competition that runs for seven months each year. It includes teaching young Ethiopians about computer coding and IT hardware and entrepreneurship, and challenges them to find technology-based solutions to community problems. Organised by the US Embassy in Addis Ababa in partnership with iCog Labs and Humanity+,[13] Solve IT showcases the work of enthusiastic young inventors, and the potential of using technology to creatively solve pressing social challenges faced by vulnerable groups and communities.[14]

Other labs are blooming in the country, laying a foundation for AI developers to develop, test and incubate ideas for products and services that address real community needs. For example, EthioCloud[15] allows AI developers to work in Ethiopia's native Amharic language, creating advanced Amharic programming code. It runs on Microsoft's .NET and C# platforms, and converts Amharic paper documents into editable text, and includes an Amharic text-to-speech conversion system and Amharic translator. There are also other hubs engaged in one way or another in AI-related activities, including iceaddis[16] and blueMoon,[17] and other established technology-led ventures such as Gebeya,[18] an online marketplace for young talent in the IT sector. Ethio Robo Robotics[19] is another recent AI initiative in Ethiopia. It aims to transform access to robotics training in the country by focusing on children to promote the early adoption of AI technologies. It works in partnership with VEX Robotics,[20] a US-based company whose mission is to create tools that educators and mentors can use to shape the learners of today into the problem-solving leaders of tomorrow.

All these developments in the AI sector are met

[7] Gadzala, A. (2018). *Coming to Life: Artificial Intelligence in Africa*. Washington: Atlantic Council. https://www.atlanticcouncil.org/images/publications/Coming-to-Life-Artificial-Intelligence-in-Africa.pdf

[8] https://icog-labs.com

[9] SingularityNET is a global decentralised AI network that lets anyone create, share and monetise AI services at scale. https://singularitynet.io

[10] Lewton, T. (2018, 13 June). Futurists in Ethiopia are betting on artificial intelligence to drive development. *Quartz*. https://qz.com/africa/1301231/ethiopias-futurists-want-artificial-intelligence-to-drive-the-countrys-development

[11] Sophia is a social humanoid robot developed by Hong Kong-based company Hanson Robotics. It is powered by artificial AI and capable of over 60 different facial mechanisms to create natural-looking expressions. It was activated and made its first public appearance in March 2016 with the ability to display more than 50 facial expressions. Since then Sophia has been covered by media around the world and has participated in many high-profile interviews. https://en.wikipedia.org/wiki/Sophia_(robot)

[12] https://icog-labs.com/solveit

[13] https://humanityplus.org

[14] Abdu, B. (2018, 12 October). Optimism amid challenges for IT innovators in Ethiopia. *iCog Labs*. https://icog-labs.com/optimism-amid-challenges-for-it-innovators-in-ethiopia

[15] https://www.ethiocloud.com

[16] http://www.iceaddis.com

[17] https://www.bluemoonethiopia.com

[18] https://www.gebeya.com

[19] https://ethioroborobotics.com

[20] https://www.vexrobotics.com

with what has been described as a minimal interest in investing in innovative ideas by the Ethiopian private sector. The young innovators complain that local investors would prefer to build an asset than invest in innovation. On the positive side, the government has invested 87 million euros[21] in a technology park called Ethio ICT Village[22] with the ambition of it becoming a centre of excellence for scientific and technological research. Furthermore, the government has also given priority attention to ICT and innovation driving its transformation agenda, through, for example, imposing a quota requiring 70% of students in universities to study in the fields of science, technology, engineering and mathematics (STEM). At least two universities have devoted themselves to the field of AI. Meanwhile the Artificial Intelligence and Robotics Center of Excellence,[23] promoted by the ministry of science and technology and established under the aegis of the Addis Ababa Science and Technology University, has been set up to create a close collaboration between academia and industry in the fields of AI and robotics.

It is well acknowledged that AI has a tremendous impact on economies and businesses and has the potential to revolutionise societies. However, as with any scientific or technological advancement, there is a real risk that the use of new tools by states or corporations will have a negative impact on human rights.[24] With the Ethiopian industrial parks and integrated agro-industrial parks proliferating across the country, the Ethiopian government aims to enable the manufacturing sector to contribute to 20% of Ethiopia's GDP and 50% of the export volume by 2025. The question is, with the industrial application of AI gaining momentum, what will be the scale of industrial job losses due to automation?

According to World Bank Development Report 2016 estimates, two-thirds of all jobs are susceptible to automation in the developing world, and the share of jobs at risk of being lost to automation and advanced technologies is about 85% for Ethiopia.[25] The implication is that in the absence of adequate policies, many workers are likely to be pushed into lower-wage jobs or become unemployed, even if temporarily. While technologies create new opportunities and enhance productivity, given the high cost of retooling workers for the future world of work, if the outcome is not mass unemployment, it is likely to be rising inequality. Given the projected effect of automation on jobs[26] and Ethiopia's vision of becoming a lower-middle-income economy by 2025, there is a need to address the impact of automation through education policy, especially now while the policy[27] itself is currently under revision.

In addition to the potential risk of job losses through the adoption of AI in Ethiopia, there are also concerns about privacy and personal data protection. While the CEO of iCog Labs, who was interviewed for this report, confirmed that they had used videos available freely online in their development of the various expressions they developed for the Sophia robot, it does point to the need for AI practitioners to respect privacy and use data responsibly. Ethiopia faces challenges that threaten privacy and data protection like most other African countries. Among others, one of these threats is the absence of adequate legal, regulatory and policy frameworks given the collection of large amounts of personal data by government entities.[28]

For example, the Proclamation on the Registration of Vital Events and National Identity Card[29] allows the collection of personal data and the transfer of this data to various institutions including intelligence authorities without the consent of data subjects. Without regulatory safeguards, the law also authorises the storage of sensitive data in a central database. Likewise, the incumbent telecom operator, Ethio Telecom, collects a lot of personal information to register SIM cards. A customer needs to provide detailed information including name and address, a photo ID, a photograph, and a signature before one can purchase a SIM card.[30] Another concern is the use of surveillance technologies by government law enforcement agencies to gather personal data without putting in place regulatory mechanisms to protect personal data.

[21] Karas-Delcourt, M. (2016, 28 January). The Ethiopian AI Geeks Building Cutting-Edge Robots. *iCog Labs*. https://icog-labs.com/the-ethiopian-ai-geeks-building-cutting-edge-robots/#more-1295

[22] ethioictvillage.gov.et/index.php/eng

[23] www.aastu.edu.et/research-and-technology-transfer-vpresident/the-artificial-intelligence-robotics-center-of-excellence

[24] Privacy International & ARTICLE 19. (2018). *Privacy and Freedom of Expression in the Age of Artificial Intelligence*. https://privacyinternational.org/report/1752/privacy-and-freedom-expression-age-artificial-intelligence

[25] World Bank Group. (2016). *World Development Report 2016: Digital Dividends*. https://www.worldbank.org/en/publication/wdr2016

[26] Ibid.

[27] Teferra, T., et al. (2018). *Ethiopian Education Development Roadmap (2018-30): An Integrated Executive Summary – Draft for Discussion*. Addis Ababa: Ministry of Education, Education Strategy Center. https://planipolis.iiep.unesco.org/sites/planipolis/files/ressources/ethiopia_education_development_roadmap_2018-2030.pdf

[28] Yilma, K. M. (2015). Data privacy law and practice in Ethiopia. *International Data Privacy Law, 5*(3), 177-189; see also Enyew, A. B. (2016). Towards Data Protection Law in Ethiopia, in A. B. Makulilo (Ed.), *African Data Privacy Laws*. Springer International Publishing.

[29] Federal Democratic Republic of Ethiopia. (2012). A Proclamation on the Registration of Vital Events and National Identity Card, Proclamation No. 760/2012. https://chilot.files.wordpress.com/2013/04/proclamation-no-760-2012-registration-of-vital-events-and-national-identity-card-proclamation.pdf

[30] Taye, B., & Teshome, R. (2018). *Privacy and Personal Data Protection in Ethiopia*. CIPESA. https://cipesa.org/?wpfb_dl=301

Furthermore, AI also has the potential to impact negatively on freedom of expression. Using irresponsible social media activism and fake news that has recently catalysed ethnic tension and violence in the country as a pretext, the government has proposed to pass a new law on hate speech.[31] Although hate speech is a growing concern, it also has to be handled with care given the potential of new technologies such as AI to manipulate video, audio and images. For example, "deepfake" technology uses machine learning to help users edit videos and add, delete or change the words coming right out of somebody's mouth.[32] Such emerging technologies can exacerbate the potential risk of AI and its implications for human rights and social justice if not responsibly used by the public, corporations, the state and other stakeholders.

Conclusion

Ethiopia's AI landscape is surrounded by both optimism and fear; optimism as to the potential that AI has for economic and social development, and fear of its human rights implications. While increasing youth enrolment in STEM fields and supporting AI innovation brings about economic benefits through the creation of new job streams, AI also has negative consequences with regard to the susceptibility of the country to job losses due to automation. Given the impact on future jobs, there will be demand for retooling, and its associated cost will contribute to rising inequality. While the government is accelerating agriculture-led industrialisation through establishing industrial parks across the country, the anticipated impact on unemployment is likely to be reduced given the take-up of automation in the manufacturing industries. With unemployment ranging at different times between 16% and 26%, the negative impact of AI on jobs would be significant. Furthermore, the increasing rate of unemployment in Ethiopia, particularly among young, college-educated people, presents a challenge to economic development, placing the onus on the government and universities to develop future-ready skills for the technology sector.

Governments have the main responsibility to protect human rights. They act as the primary guarantors of these rights, and should be held accountable when rights are not realised. In this regard, they have both a positive and negative role to protect and to refrain from interfering in the citizens' exercise of their rights and freedoms. In many ways, the internet – and particularly emerging technologies – have opened ways for the exercise of many rights and freedoms while at the same time challenging them. Examples include online hate speech, fake news, surveillance, or privacy issues. In this regard, the main role of the government is to ensure the balance between freedom and protection, rights and responsibilities. This maintenance of the right balance requires the government to engage the private sector, civil society and other stakeholders in respecting human rights, including in the design, development and delivery of its own digital services.[33]

Action steps

Ethiopia can reap the benefits of AI if the Ethiopian government, investors and other stakeholders can equip workers with 21st century skills, and reform laws and education to meet the demands of the digital economy. To this end, the following actions are proposed:

- Ensure that the education development roadmap currently under discussion is forward-looking so that it integrates AI studies in the educational system and meets the demands of tomorrow's economy.
- Review the national ICT policy so that it embraces new emerging technologies including AI, big data, IoT and cloud computing.
- Put in place legal frameworks with respect to privacy and data protection, taking into account the African Union Convention on Cyber Security and Personal Data Protection.[34]
- Build a sound statistical system that adapts to the emerging data revolution.
- Promote the thriving innovation labs so that they can leverage innovations across industrial parks in the country.
- Manage the industrialisation process in order to ensure gender-responsive outcomes that benefit women and girls,[35] including society as a whole, through the removal of barriers to equity.

[31] Tsegaye, Y. (2018, 23 November). Ethiopia Preparing New Bill to Curb Hate Speech. *Addis Standard*. https://addisstandard.com/news-ethiopia-preparing-new-bill-to-curb-hate-speech

[32] Vincent, J. (2019, 10 June) AI deepfakes are now as simple as typing whatever you want your subject to say. *The Verge*. https://www.theverge.com/2019/6/10/18659432/deepfake-ai-fakes-tech-edit-video-by-typing-new-words

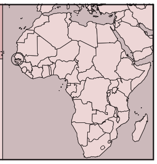
[33] DiploFoundation. (2018). *Mapping the challenges and opportunities of artificial intelligence for the conduct of diplomacy*. Diplo AI Lab and Ministry of Foreign Affairs of Finland. https://www.diplomacy.edu/AI-diplo-report

[34] https://au.int/sites/default/files/treaties/29560-treaty-0048_-_african_union_convention_on_cyber_security_and_personal_data_protection_e.pdf

[35] United Nations Development Programme. (2018). *Ethiopia National Human Development Report 2018: Industrialization with a Human Face*. Addis Ababa: UNDP. hdr.undp.org/sites/default/files/ethiopia_national_human_development_report_2018.pdf

THE GAMBIA

THE NEED TO USE AI TO DEAL WITH ONLINE HARASSMENT IN THE GAMBIA



SHOAW Gambia
Rameesha Qazi and Anna Anet Sambou
www.shoawgambia.org

Introduction

This report argues for the use of natural language processing to protect women and girls from being harassed online. By tracking how people communicate with each other, we will be able to limit the ways in which messages of harassment are sent and who they reach. To protect the greatest number of at-risk users of social media, civil society organisations and companies like Facebook (which also owns WhatsApp and Instagram) need to work together to protect human rights that are currently being violated.

Context

Women and girls all over the world are targets of online harassment, stalking, and so-called "revenge porn". In extreme cases, harassment that started online has turned into murder. These kinds of behaviour need to be stopped in the quickest and most straightforward way possible.

The Gambia is no exception – young girls are falling victim to many forms of online harassment and abuse. Recently, a video of a high school girl wearing her school uniform and dancing "inappropriately" with a boy went viral. Subsequently the girl was expelled from school. When human rights activists reached out to the school to review their decision, one of their best teachers said if the girl came back to school he would quit. The Ministry of Education failed to address this. This young girl had a bright future ahead of her and now she will carry around the shame and trauma of this incident for the rest of her life, effectively limiting her future opportunities. We believe that if used in the right way, artificial intelligence (AI) has the power to limit these stories.

Social media in The Gambia is not as widely used as it is in the West, but almost everyone uses WhatsApp (which was acquired by Facebook in 2014). WhatsApp will be the focus of this report, but the impacts of online harassment can be far reaching across all platforms around the world.

Politically, the time is right in The Gambia for something like this to be taken seriously and pursued. Currently the country has no policies or regulations in place dealing with online harassment, nor does it have any legal consequences for perpetrators. This is something that was brought up at the national Internet Governance Forum (IGF) and West African IGF that happened locally in July 2019.

Gender bias in AI: How women's needs are neglected

The most basic indicator of diversity is gender, and AI is a male-dominated field. According to the World Economic Forum's latest Global Gender Gap Report, only 22% of AI professionals globally are female compared to 78% who are male.[1] The biggest problem with this is that when male developers create their systems, they incorporate, often in an unconscious way, their own biases in the different stages of their creation.[2]

There are already many examples of biases in AI that have been found to be impeding the success of women in a variety of fields:

- Several reports have found that voice and speech recognition systems performed worse for women than for men.
- Face recognition systems have been found to make more errors with female faces.
- Recruiting tools based on text mining can inherit gender bias from the data they are trained on.[3].

AI is impacting on the lives of millions of people around the world, from Netflix predicting what movies people should watch to corporations, governments and law enforcement deciding who gets a loan, a job or immigration status. When AI systems make biased, unjust decisions, it has real-world consequences for people – very often impacting

[1] Teigland, J. L. (2019, 2 April). Why we need to solve the issue of gender bias before AI makes it worse. *EY.com*. https://www.ey.com/en_gl/wef/why-we-need-to-solve-the-issue-of-gender-bias-before-ai-makes-it

[2] Gomez, E. (2019, 11 March). Women in Artificial Intelligence: mitigating the gender bias. *JRC Science Hub Communities*. https://ec.europa.eu/jrc/communities/en/community/humaint/news/women-artificial-intelligence-mitigating-gender-bias

[3] Ibid.

on women and people of colour.[4] Researchers have found that AI systems will spit out biased decisions when they have "learned" how to solve problems using data that is exclusive and homogeneous – and those mistakes disproportionately affect women, people of colour, and low-income communities.[5]

There are also many layers of biases in AI. One is the "unknown unknowns"[6] which appear after the AI system is complete. Karen Hao writes:

> The introduction of bias isn't always obvious during a model's construction because you may not realize the downstream impacts of your data and choices until much later. Once you do, it's hard to retroactively identify where that bias came from and then figure out how to get rid of it. In Amazon's case, when the engineers initially discovered that its tool was penalizing female candidates, they reprogrammed it to ignore explicitly gendered words like "women's". They soon discovered that the revised system was still picking up on implicitly gendered words – verbs that were highly correlated with men over women, such as "executed" and "captured" – and using that to make its decisions.[7]

Amazon's system taught itself that male candidates were preferable for many jobs based on the data it had been built on.

Dealing with the "unknown unknown" – or predicting online harassment

For the preparation of this report, meetings were arranged with victims of online abuse and harassment. These meetings were with students from age 13 to 16. The aim was to understand how they are being targeted and what kinds of things are happening to them. These students were representative of various socioeconomic backgrounds and ranged from public, private and international schools.

All the students concurred that if they had a circle of 10 friends, more than half would either have been or are being targeted online by men, many of whom they do not know, and some who have positions of power in the victim's life. All the girls said that they do not know how to address these kinds of problems as there is a culture of silence in The Gambia, and when girls do speak out the blame falls on their shoulders. All the girls who were present at these meetings said that they were targeted on WhatsApp and Facebook – both platforms that use AI to manage messages[8] and track behaviour.

The issue is not that AI is being used; the issue is how the AI is built and the gaps that exist in its learning. As suggested, the risk exists in the fact that the learning mimics that of its creator, which can make the problem of gender inequality worse by not addressing it consciously when the AI is developed. So, if a male developer does not see online harassment as a key problem with the internet, it is unlikely to receive much attention in the development of the AI system.

The problems presented here can be easily rectified if civil society organisations and governments come together to gather stories from victims that can be used in the design of AI systems. For example, the creators of the AI can use the messages and behaviours from these interactions to teach the AI which messages to block and how to ensure users who are attempting to harass women and girls are tracked and managed properly.

Julie Teigland writes:

> Women need to be builders and end users of the AI-enabled products and services of the future. By shifting the perception, and role, of women within society, we can correct the digital bugs that perpetuate existing bias and make the AI lifecycle more trustworthy. Technology can do many great things, but it cannot solve all our problems for us. If we are not careful, it could end up making our problems worse – by institutionalizing bias and exacerbating inequality.[9]

The problem of online harassment is clearly manageable and solvable. Monitoring text communication using AI can be done, and natural language processing can be introduced to AI systems that already exist to filter voice messages. Natural language refers to language that is spoken and written by people, and natural language processing attempts to extract information from the spoken and written word using algorithms.[10]

A typical human-computer interaction based on natural language processing happens in this order: (1) the human says something to the computer,

[4] Gullo, K. (2019, 28 March). Meet the Bay Area Women in Tech Fighting Bias in AI. *Seismic Sisters*. https://www.seismicsisters.com/newsletter/women-in-tech-fighting-bias-in-ai

[5] Ibid.

[6] 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-happensand-why-its-so-hard-to-fix

[7] Ibid.

[8] For example, certain messages cannot be sent on Facebook Messenger, such as links where you can download music or movies illegally.

[9] Teigland, J. L. (2019, 2 April). Op. cit.

[10] Nicholson, C. (n.d.). A Beginner's Guide to Natural Language Processing (NLP). *Skymind*. https://skymind.ai/wiki/natural-language-processing-nlp

(2) the computer captures the audio, (3) the captured audio is converted to text, (4) the text's data is processed, (5) the processed data is converted to audio, and (6) the computer plays an audio file in response to the human. But there are many other uses for natural language processing:

- Chatbots use natural language processing to understand human queries and respond.
- Google's search tries to interpret your question or statement.
- Auto-correct is a lot less frustrating these days thanks to deep learning and natural language processing.[11]

By applying the existing text and voice management technology to the serious human rights problem of online harassment, we will be able to protect every vulnerable user from being harassed, abused, and even murdered.

Conclusion

At the Gambian IGF and the West African IGF, both held in The Gambia in July 2019, these topics were addressed. It was concluded that the internet space needed to become safer for all users and this issue will be raised again at the global IGF.

By implementing natural language processing in AI systems used around the world, one glaring problem arises, which is that the kinds of messages sent to victims could be similar to messages between consenting adults – for example, through sexually explicit flirting, or "sexting". If the AI has learned how to protect victims of harassment and abuse, then it will block consensual messages as well. But, in our view, the well-being of millions of young girls and women should be protected over the ability of some to be able to send explicit messages.

Civil society organisations and governments need to come together to ensure that proper policies on online safety exist, proper consequences for perpetrators of harassment are in place, and that stories of harassment are collected as they happen so that social media platforms can better teach their AI systems, and install proper reporting tools. AI is a new domain that needs new policies and procedures, but it is critical that all countries have them in place – and soon, as these issues are happening every day to society's most vulnerable groups, and they need to be protected.

Everyone has conscious and unconscious biases about a variety of things, and AI has the potential to overcome but also to inherit/perpetuate biases. We need more women in AI to make sure AI systems are developed *by* women and *for* women's welfare.[12] A gender-responsive approach to innovation will help to rectify the bias that is already built into the system; however, it requires thinking about how we can better leverage AI to protect the most vulnerable users globally.[13] Once we can secure the most basic human rights online, women and girls around the world can use the power of the internet and social media to empower themselves and their communities beyond what can be imagined today.

Action steps

The following action steps are necessary in The Gambia:

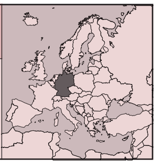
- Collect stories and messages from victims of online harassment to feed into the design of AI systems. While text- or image-based harassment can be easily monitored, AI can also learn speech patterns to block voice messages.
- Civil society organisations need to pressure social media platforms to adapt the AI used in their systems so that they block attempts at harassment of women and girls online. Reporting tools need to be implemented that monitor online harassment more strictly, with real consequences for the perpetrators of these behaviours.
- Civil society organisations need to work together to ensure that victims of online harassment have a safe place to report incidents and get the help they need.

[11] Greene, T. (2018, 25 July). A beginner's guide to AI: Natural language processing. *TNW*. https://thenextweb.com/artificial-intelligence/2018/07/25/a-beginners-guide-to-ai-natural-language-processing

[12] Gomez, E. (2019, 11 March). Op. cit.

[13] Teigland, J. L. (2019, 2 April). Op. cit.

GERMANY



BIOMETRICS AND BODY POLITIK: THE RISE OF AUTOMATED SURVEILLANCE IN GERMANY

Centre for the Internet and Human Rights (CIHR),
European University Viadrina
Mathana Stender
https://twitter.com/StenderWorld

Machine-readable humans

As technology has made our lives easier, it has also numbed us to the potential threats that can be posed by new forms of "convenience". Putting one's iPhone in front of your face to unlock it with FaceID, instead of entering a short string of numbers, is an alluring prospect for some, but the impact of the development and proliferation of human-sensing technologies has broad societal repercussions, particularly for society's most vulnerable.

We are subjected to distributed-yet-persistent digital surveillance, where sensors that are able to discern minute details from our bodies have been installed all around us. Because no two sets of fingerprints or irises or voices are exactly the same, each human body contains physical identifiers that are unique to the individual. Using computer vision technology combined with algorithmic decision making, biometric identity management and access control systems capture and analyse our permanent, immutable characteristics, and are being increasingly used for automated surveillance.[1] These systems, while claiming to differentiate, classify and categorise people, are however flawed due to biased data and technical limitations. Though constrained by various limitations, the systems have become increasingly adopted by both companies and governments to streamline surveillance.

With biased assumptions built into training of models, and flawed labelling of training data sets,[2] this class of technologies often do not differentiate between who is surveilled; anyone who passes through their sensor arrays are potential subjects for discrimination.

Surveillance does not happen in a vacuum. The use of biometric and automated access control mechanisms is increasing globally at an alarmingly high rate: India's compulsory Aadhaar biometric ID system has records from well over a billion individuals, while 23 million people were blocked from travelling in China last year due to their low social credit scores.[3] Biometric access has now become a gatekeeper to both basic services and the freedom of movement.

Automated facial recognition in action

Biometric surveillance and access control involves the computational analysis of parts of a person's physical attributes – fingerprints, facial features, retina or voice. In order for a system to recognise one's immutable characteristics, they must compare input against a database to identify key features. When the result of a facial feature map leads to a particular conclusion about the subject being surveilled, which is dynamically generated by an algorithm, this is known as "automated facial recognition" (AFR).

Sometimes this AFR takes the form of verification: is a person who they say they are (or, more accurately, do the physical characteristics being "scanned" by a system match the system's records for this person)? In other cases, however, biometrics are used to analyse or extrapolate an aspect of a person's identity. An automated biometric system might be looking to see if a person is a child or adult, or seek to classify one's ethnicity and age. Other, more invasive forms of AFR might be assessing someone's sexual orientation,[4] or assigning a score pertaining to the likelihood that they will commit a crime.[5]

Over the past few years, both the German government and the European Union (EU) have turned

[1] Ohrvik-Stott, J., & Miller, C. (2019). *Responsible Facial Recognition Technologies*. Doteveryone. https://doteveryone.org.uk/wp-content/uploads/2019/06/Doteveryone-Perspective_Facial-Recognition-1.pdf

[2] "Many companies report high accuracies using a data set called Labeled Faces in the Wild, but this data set only contains 5,171 people. Most large cities are in the millions. What works for 5,000 doesn't necessarily work for 5 million." Interview with AFR researcher Adam Harvey, 11 June 2019.

[3] Kou, L. (2019, 1 March). China bans 23m from buying travel tickets as part of 'social credit' system. *The Guardian*. https://www.theguardian.com/world/2019/mar/01/china-bans-23m-discredited-citizens-from-buying-travel-tickets-social-credit-system

[4] Gutierrez, C. (2019, 29 June). Unregulated facial recognition technology presents unique risks for the LGBTQ+ community. *TechCruch*. https://techcrunch.com/2019/06/29/unregulated-facial-recognition-technology-presents-unique-risks-for-the-lgbtq-community

[5] Du, L., & Maki, A. (2019, 24 March). These Cameras Can Spot Shoplifters Even Before They Steal. *Bloomberg*. https://www.bloomberg.com/news/articles/2019-03-04/the-ai-cameras-that-can-spot-shoplifters-even-before-they-steal

to biometric systems as they deal with a humanitarian crisis. Political instability brought about by the Syrian civil war and the rise of ISIS threatened the lives of millions of people, triggering one of the largest mass migrations of externally displaced persons in the 21st century. As millions of people fled war-torn lands, Europe witnessed a massive influx of refugees and asylum seekers. In 2015, the German government granted asylum to one million refugees,[6] a decision that would further galvanise xenophobia through political propaganda, and, within two years, propel a far-right political party into a strong position of political power.[7] The government's response to the crisis was to create a new biometric identity system for refugees, a system that was eventually integrated into a new EU-wide biometric identity management surveillance system that went far beyond its original intent. By 2019, biometric surveillance and algorithmic policing had become normalised to the extent that members of the European Parliament (MEPs) voted in favour of the creation of a biometric database that would centralise law enforcement, immigration and other information on over 350 million people.

Surveillance-as-a-service: Bodies and borders

Biometric sensors often use new algorithmic processes to leverage existing infrastructure. In Germany, where there are 6,000 CCTV cameras[8] scattered throughout the country's roughly 900 train and metro stations alone, existing capacity for a widespread surveillance network is already in place. Unlike in the past, however, when there were not enough humans to watch all the recorded video, meaning that much footage was seen only in cases where evidence of crime was needed, recent advancements in computer-vision AI can now "watch" CCTV footage in real time and automatically notify authorities when something "suspicious" has occurred.[9]

Sometimes, seemingly benign uses of new technologies can show how ill conceived technology implementation can be. Algorithmic-driven biometric systems are inherently problematic not only because there is so much cultural and ethnic diversity in humanity, but also because, in a day and age where characteristics like gender are fluid, such systems may be built on data sets that are developed using binary parameters and rudimentary perceptions of performativity-based gender analysis.

Prevailing winds

The origins of Germany's biometric identity registries coincided with the large uptick in refugees and asylum seekers into Germany and the EU in 2013[10] that had been triggered by the war in Syria and other instability. An EU system centralised the identity of those seeking asylum in the EU along with their fingerprints into a unified database called the Eurodac system. The system, and its Automated Fingerprint Identification System (AFIS), were created to facilitate the "Dublin Regulation", which stipulated that refugees apply for protection in (and only in) the first European country that they arrive at. All asylum seekers, regardless of the location of their asylum claim, would now also have their fingerprints and photos aggregated in the Eurodac.[11]

While EU legislation was rolling out its biometric registry for refugees, Germany was developing its own plans. In December 2015, the German cabinet approved a measure establishing the creation of identity cards for refugees.[12] Former German Interior Minister Thomas De Maiziere, who oversaw the issuance of the new identity card, was a proponent of a form of social engineering. He alarmed advocates when he spoke about the need for "Leitkultur", the idea of instilling dominant (German) cultural values in refugee seekers.[13]

The implementation of biometrics would soon reach German citizens. Documents revealed by the German media in 2016 showed a draft plan by the Interior Ministry to deploy AFR in areas ranging

[6] Werber, C. (2015, 26 August). Germany is the first European country to free Syrian refugees from a draconian bureaucratic "trap". *Quartz*. https://qz.com/488413/germany-is-the-first-european-country-to-free-syrian-refugees-from-a-draconian-bureaucratic-trap

[7] Clarke, C. (2017, 25 September). German elections 2017: Full results. *The Guardian*. https://www.theguardian.com/world/ng-interactive/2017/sep/24/german-elections-2017-latest-results-live-merkel-bundestag-afd

[8] Global Rail News. (2017, 2 August). Facial recognition technology to be trialled at Berlin railway station. *Global Rail News*. www.globalrailnews.com/2017/08/02/facial-recognition-technology-to-be-trialled-at-berlin-railway-station

[9] Glaser, A. (2019, 24 June). Humans Can't Watch All the Surveillance Cameras Out There, So Computers Are. *Slate*. https://slate.com/technology/2019/06/video-surveillance-analytics-software-artificial-intelligence-dangerous.html

[10] OECD. (2015, 22 September). Comprehensive and co-ordinated international response needed to tackle refugee crisis. https://www.oecd.org/migration/comprehensive-and-co-ordinated-international-response-needed-to-tackle-refugee-crisis.htm

[11] https://ec.europa.eu/home-affairs/what-we-do/policies/asylum/examination-of-applicants_en

[12] Copley, C. (2015, 9 December). German cabinet approves identity card for refugees. *Reuters*. https://www.reuters.com/article/us-europe-migrants-germany-idUSKBN0TS1K620151209#ODikvJ2zgKwilx4w.97

[13] DW. (2017, 30 April). German interior minister speaks out in favor of 'Leitkultur' for immigrants. *DW*. https://www.dw.com/en/german-interior-minister-speaks-out-in-favor-of-leitkultur-for-immigrants/a-38643836

from shopping malls to train stations and airports.[14] The plan received some push-back at the time from opposition parties and the media. In June the following year, a resolution from the Conference of Independent Data Protection Authorities of Federal and State Governments (DSK) stated the threat to society posed by AFR technology in no uncertain terms: the "use of video cameras with biometric facial recognition can completely destroy the freedom to remain anonymous in public spaces. It's practically impossible to evade this kind of surveillance, let alone to control it."[15] Despite such a warning, the technology was not put on hold.

One high-profile case from 2017 that shocked the nation saw a German army officer publicly named Franco A. charged with a terror-related plot to assassinate German officials all the while posing as a Syrian refugee (the charges were later dropped for lack of evidence).[16] Prosecutors disclosed that the man was looking to frame refugees in a "false flag" attack and thus further degrade public opinion toward asylum seekers.[17] Because the army officer and would-be terrorist had been able to enrol for asylum seeker services posing as a Syrian, the government decided that biometrics could prevent a repeat of the incident. Yet another biometric regime was implemented for refugees and asylum seekers, as the government proclaimed "no more Franco A.s".[18] This system saw the roll-out of speech analysis, which the government claimed could analyse linguistic dialects to verify a place of origin.[19]

In 2017, an AFR pilot project deployed by the Interior Ministry, German federal police departments and Germany's Deutsche Bahn rail company[20] was introduced into one of the capital's sprawling metro systems. Known as Safety Station Südkreuz, the programme enrolled 300 individuals who volunteered to have their faces used to help train the system in exchange for a EUR 25 Amazon voucher.[21].

In other parts of Germany, individual states have implemented their own AFR schemes. Section 59 of a 2019 Saxon police law, titled "Use of Technical Means in Order to Prevent Serious Cross-border Crime", created a new security zone 30 kilometres into Saxony from the Czech and Polish borders. The civil society group Digital Courage[22] warned that "the planned border surveillance places large parts of Saxony under some sort of state of emergency," adding that this was a "statement of distrust toward our Czech and Polish neighbours"[23] and "by implementing these changes, the Saxon Judiciary and Police will take on characteristics of a preventive state."[24]

In line with the DSK's 2017 statement, certain elements of the German government seem to be coming around to the idea that algorithmic governance is a pressing issue. In June 2019, while speaking at an AI conference in Dresden, German Chancellor Angela Merkel again addressed the need for automated decision-making technologies to be deployed under more formal governance oversight mechanisms: "We need [regulation], I'm convinced of that. Much of that should be European regulation."[25] Unclear, however, is how Germany will balance privacy, the politicisation of domestic security issues and EU data-sharing regulations.

By January 2019, the Schengen Information System (SIS II) alone contained nearly 240,000 fingerprints,[26] a further expansion of AFIS.[27] In April 2019, MEPs passed legislation establishing the creation of the Common Identity Repository (CIR). A shared Biometric Matching Service will provide "fingerprint and facial image search services to

[14] Knight, B. (2016, 26 October). Germany planning facial recognition surveillance. *DW*. https://www.dw.com/en/germany-planning-facial-recognition-surveillance/a-36163150

[15] Data Protection Conference (DSK). (2017, 30 March). Einsatz von Videokameras zur biometrischen Gesichtserkennung birgt erhebliche Risiken [The use of video cameras for biometric facial recognition poses considerable risks]. https://www.datenschutzkonferenz-online.de/media/en/20170330_en_gesichtserkennung.pdf

[16] DW. (2018, 7 June). German court throws out terrorism charges against soldier. *DW*. https://www.dw.com/en/german-court-throws-out-terrorism-charges-against-soldier/a-44116741

[17] DW. (2018, 12 December). German soldier charged with plotting to kill politicians while posing as refugee. *DW*. https://www.dw.com/en/german-soldier-charged-with-plotting-to-kill-politicians-while-posing-as-refugee/a-41766093

[18] Chase, J. (2017, 27 July). German refugee agency unveils new asylum identity technology. *DW*. https://www.dw.com/en/german-refugee-agency-unveils-new-asylum-identity-technology/a-39857345

[19] Ibid.

[20] Global Rail News. (2017, 2 August). Op. cit.

[21] Delcker, J. (2019, 19 April). Big Brother in Berlin. *Politico*. https://www.politico.eu/articleberlin-big-brother-state-surveillance-facial-recognition-technology

[22] https://digitalcourage.de

[23] van der Veen, M., & Lisken, S. (2019, 22 January). Police Laws in Saxony: Czech, Polish and German Criticism on Plans for Facial Recognition in the Border Region. *Digital Courage*. https://digitalcourage.de/blog/2019/police-laws-in-saxony

[24] Interview with Friedemann Ebelt of Digital Courage, 5 July 2019.

[25] Delcker, J. (2019, 24 June). AI experts call to curb mass surveillance. *Politico*. https://www.politico.eu/article/eu-experts-want-curtailing-of-ai-enabled-mass-monitoring-of-citizens

[26] Bundesministerium des Innern, für Bau und Heimat. (2019, 23 January). Zahlen zu Speicherungen in polizeilichen EU-Datenbanken (2018). https://andrej-hunko.de/start/download/dokumente/1287-speicherungen-polizeiliche-eu-datenbanken-2018/file [note: written response from the President of the German Parliament to Andrej Hunko, a member of Germany's Die Link party]; Sánchez-Monedero, J. (2018). *The datafication of borders and management of refugees in the context of Europe*. Data Justice Project. https://datajusticeproject.net/wp-content/uploads/sites/30/2018/11/wp-refugees-borders.pdf

[27] https://ec.europa.eu/home-affairs/content/automated-fingerprint-identification-system-afis_en

cross-match biometric data present on all central systems".[28] ZDnet, quoting EU officials, reported how the CIR will create new rules for data sharing and "would include the Schengen Information System, Eurodac, the Visa Information System (VIS) and three new systems: the European Criminal Records System for Third Country Nationals (ECRIS-TCN), the Entry/Exit System (EES) and the European Travel Information and Authorisation System (ETIAS),"[29] giving law enforcement agencies unprecedented access to personal information such as names and biometric data. The CIR will combine law enforcement, immigration data and other data into a searchable database containing the records of 350 million people who live in and travel to Europe.

Throughout the same time period when the new biometric identity management systems were created with the aim of policing groups like refugees along Germany's eastern border, a new and highly troubling domestic threat began to emerge. The rise of well-trained, highly organised, overtly violent far-right groups have evaded surveillance and in certain cases – like that of Franco A. – actually attempted to carry out atrocities while on the government's payroll. In fact, more than 400 cases of right-wing extremism in the German army alone were under investigation as of April 2018,[30] and by mid-2019, Germany's domestic intelligence service, the Federal Office for the Protection of the Constitution (BfV), was aware of 24,100 right-wing extremists in Germany, more than half of whom were thought to be "violence-oriented".[31] While refugee dialects were being analysed, the messages of the violent right-wing movement were finding new, home-grown adherents.

Blindspots and camouflage

Originally, fingerprints of asylum seekers and visa applicants in Europe were entered into a highly restricted database, searchable by only certain law enforcement agencies under specific protocols.[32] Over time, however, biometric records were repurposed for more general security screening. The infusion of biometric surveillance in German society has taken a few years and in many cases occurred in a piecemeal fashion. What was once billed as a system to verify refugee identity and status has built up an invasive capacity; the ability to police the most vulnerable has given way to a European-wide access control mechanism based on immutable physical characteristics.

When policies are reactive to hyperbolic rhetoric, the result can be turning a blind eye to actual threats to a peaceful society. On 2 June 2019, Germany was shocked by a politically motivated assassination of an outspoken pro-immigration politician by a man who, despite a long history of anti-immigrant violence, was not present on the "watch list" of the BfV.[33] Walter Lübcke, a regional leader from Merkel's CDU party, was shot in the head by a handgun at close range outside his home in Kassel in central Germany's Hesse. The same month, the BfV reported that a group of 30 extremists, most of whom were associated with Germany's police or military, had used police databases to compile a list of the names and addresses of 25,000 people, most of whom were active in various political parties and, according to Deutsche Welle (DW), Germany's public international broadcaster, supported "pro-refugee" policies.[34]

Despite calls from policy makers and AI experts, public figures must do more to educate themselves and the public regarding the shortcomings of this new technology. Some of this education involves a critical re-evaluation of what AFR fundamentally is. "Decision makers need to highlight policy around data sourcing and consent," said Adam Harvey,[35] a Berlin-based researcher and artist who studies AFR. "They need to understand that AI products are data-driven products and therefore data sets are part of the product, not an externality."

Conclusion

Despite the push by some of Germany's leaders to increase the cultural assimilation of refugees, the worrying prospect of a society-wide surveillance state powered by biometric access control mechanisms now looms large over the entire German society. In a

[28] European Commission. (2018). *EU Interoperability Framework For Border Management Systems: Secure, safe and resilient societies*. https://www.securityresearch-cou.eu/sites/default/files/02.Rinkens.Secure%20safe%20societies_EU%20interoperability_4-3_v1.0.pdf

[29] Cimpanu, C. (2019, 22 April). EU votes to create gigantic biometrics database. *ZDNet*. https://www.zdnet.com/article/eu-votes-to-create-gigantic-biometrics-database

[30] DW. (2018, 12 April). Cases of far-right extremism on the rise in German military. *DW*. https://www.dw.com/en/cases-of-far-right-extremism-on-the-rise-in-german-military/a-43352572

[31] Knight, B. (2019, 27 June). Germany records small uptick in far-right extremist violence. *DW*. https://www.dw.com/en/germany-records-small-uptick-in-far-right-extremist-violence/a-49379510

[32] Monroy, M. (2019, 23 January). Significantly more fingerprints stored in the Schengen Information System. *digit.site36.net*. https://digit.site36.net/2019/01/23/significantly-more-fingerprints-stored-in-the-schengen-information-system

[33] Knight, B. (2019, 26 June). Suspect in German politician's murder confesses. *DW*. https://www.dw.com/en/suspect-in-german-politicians-murder-confesses/a-49357904

[34] Knight, B. (2019, 29 June). German neo-Nazi doomsday prepper network 'ordered body bags, made kill lists'. *DW*. https://www.dw.com/en/german-neo-nazi-doomsday-prepper-network-ordered-body-bags-made-kill-lists/a-49410494

[35] Interviewed by the author. Disclosure: the author was a contributing researcher to megapixels.cc, a project co-founded by Harvey.

country where 20th century atrocities still loom large, evident in discussions on education,[36] and in the recent offer of reparations to Holocaust survivors,[37] the country's approach to how it deals with the world's most vulnerable[38] is a new test of a nation's resilient openness. With the rise of an automated infrastructure, individuals and advocates must be vigilant to safeguard human rights protections. Depending on the system design for algorithmic decision making, certain attributes – like being a police officer or member of the military – may lower the risk score of an individual, yet the number of enlisted extremists is mind-blowingly high.

The non-unified approach by German states and the federal government to both domestic laws and EU obligations put the most vulnerable at risk of having their rights eroded by algorithmic bias and automated discrimination. As biometric systems enter our lives, we also run the risk of normalising invasive surveillance. AFR can also lead to automated human rights abuses, or at least can take humans out of the loop in safeguarding against decisions to ensure that human rights are upheld. In 2018, a record number of refugees were deported from Germany to other EU countries.[39] If this trend is exacerbated by xenophobic policy making or reliance on biased data sets, innocent people may be deported or denied entry into Germany.

German activists working on such issues have warned that "in a free democracy, there is no place for mass surveillance."[40] Ubiquitous AFR can also have a chilling effect on people's actions, as the prospect of "always being watched" by the state can "nudge" our actions for fear of reprisal. Such systems do not appear overnight, however, and the slippery slope from "security" to draconian social control is often paved with seemingly mundane technological steps. Yet once biometric databases like the CIR are accessible by enforcement agencies, regulatory oversight is needed to protect (and deter) against adversarial and unsanctioned actions.

Advocates should see this as a local, regional, national and international issue. Once a person's biometrics are entered in a database, they in many ways are at the mercy of automated systems. Perhaps the only way for someone to be completely safeguarded from automated biometric bias is for the systems to not exist at all.

Action steps

While biometrics are increasingly being used for surveillance, identity management and access control, such a deployment entails cooperation of a wide range of actors. For activists, this means pressurising companies, appealing to governments and lobbying members of parliament.

- *Transparency:* "There may be many more data sets that we don't yet know about that are private," noted Adam Harvey, the AFR researcher.[41]
- *Direct advocacy:* Activists can put pressure on private companies who may seek to sell their biometric access control technologies to governments. Advocacy, geared towards boycott calls, labour-based organising such as employee walk-outs, and other direct-action campaigns have been shown to be effective in some cases.[42]
- *Research:* By unmasking the origins of data sets and procurement practices for the data contained in the data set,[43] advocates can learn more about potential biases in data procurement, labelling and use.
- *Legislative lobbying:* Create model legislation and replicate strategies used by cities like San Francisco[44] to ban AFR use by municipalities.
- *Strategic litigation:* Contest the constitutionality of regulations by governments to prevent the aggregation of various aspects of their surveillance infrastructure.

[36] PBS Frontline. (2005, 31 May). Holocaust Education in Germany: An Interview. *PBS.* https://www.pbs.org/wgbh/pages/frontline/shows/germans/germans/education.html

[37] Der Spiegel. (2013, 29 May). Germany to Pay 772 Million Euros to Survivors. *Der Spiegel.* https://www.spiegel.de/international/germany/germany-to-pay-772-million-euros-in-reparations-to-holocaust-survivors-a-902528.html

[38] Werber, C. (2015, 26 August). Op. cit.

[39] The Local. (2019, 21 January). Germany deported record number of refugees in 2018 to EU countries: report. *The Local.* https://www.thelocal.de/20190121/germany-deported-record-number-of-refugees-in-2018-report

[40] Interview with Friedemann Ebelt of Digital Courage, 5 July 2019.

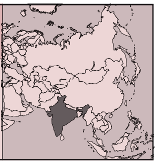
[41] Interview with AFR researcher Adam Harvey, 11 June 2019.

[42] Fang, L. (2019, 1 March). Google Hedges on Promise to End Controversial Involvement in Military Drone Contract. *The Intercept.* https://theintercept.com/2019/03/01/google-project-maven-contract

[43] Murgia, M. (2019, 6 June). Microsoft quietly deletes largest public face recognition data set. *Financial Times.* https://www.ft.com/content/7d3e0d6a-87a0-11e9-a028-86cea8523dc2

[44] Sheard, N. (2019, 14 May). San Francisco Takes a Historic Step Forward in the Fight for Privacy. *Electronic Frontier Foundation.* https://www.eff.org/deeplinks/2019/05/san-francisco-takes-historic-step-forward-fight-privacy

INDIA



ARTIFICIAL INTELLIGENCE IN EDUCATION IN INDIA: QUESTIONING JUSTICE AND INCLUSION

Digital Empowerment Foundation
Anulekha Nandi
www.defindia.org

Introduction

The National Strategy for Artificial Intelligence, released by NITI Aayog (the Indian government's policy think tank) in June 2018, underscores the government's policy intent to mainstream artificial intelligence (AI) in critical social and economic infrastructures.[1] Out of the 10 focus areas identified by the Artificial Intelligence Task Force, constituted by the Ministry of Commerce and Industry,[2] the education sector has seen the most successful public-private partnerships (PPPs) to deal with some of the institutional gaps plaguing the sector.[3]

With AI being a data-hungry technology, it becomes increasingly problematic when it is trained on the sensitive personal information of marginalised populations through service delivery in key social infrastructures like education. This is especially concerning given the current lack of a data protection regulation in India and the concomitant carve-outs for state functions in public service delivery. Moreover, the draft data protection bill which is currently tabled before the parliament does not contain explicit provisions on algorithmic decision making, including the right to be informed of its existence and the right to opt out, unlike the European Union's General Data Protection Regulation (GDPR).[4]

The impetus behind the deployment of AI has outstripped legal and regulatory development in the area, leaving a governance vacuum over a general-purpose technology with unquantifiable impact on society and economy. Given the multidimensional and cross-cutting risks and opportunities that this poses, along with complex and dynamic ethical challenges, it becomes imperative to study and understand use cases to inform and work towards context-sensitive AI governance frameworks.

Institutional AI in the Indian technology and education paradox

The use of technology in education in India traverses the unequal realities of two facets of the country. On one hand there is the segment of the population with access to digital, social and economic resources and on the other there is the vast majority for whom even basic institutions of social infrastructure offer rudimentary support at best. Private investment in education technology – or EdTech – is a burgeoning industry which clocked a valuation of USD 4.5 billion globally in 2015.[5] As per data from the research firm Tracxn, out of the 300 Indian start-ups that use AI as a core product offering, 11% are based in the education sphere.[6] India's digital learning market was valued at USD 2 billion in 2016 and is projected to grow at a compound annual growth rate (CAGR) of 30% to reach USD 5.7 billion by 2020.[7] However, the product offerings that result from these significant investments either aim to offer tutoring services, improve learning outcomes, or provide customised learning, all of which serve to leverage and augment the agency of the first segment of the population. The uptake, adoption and usage of these services proceed through the notice and consent protocols of informed consent due to service requirements of digital distribution platforms like Android's Google Play Store or Apple's Apple Store.

However, institutional applications of AI are based on public data collected by the government through service delivery, especially with regard to the social protection of marginalised, underserved and vulnerable populations, and are not undergirded by the need for adherence to data protection principles. Moreover, a joint reading of the draft data protection bill and judgement of the Supreme

[1] NITI Aayog. (2018). *National Strategy for Artificial Intelligence*. New Delhi: NITI Aayog. https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf

[2] https://www.aitf.org.in

[3] NITI Aayog. (2018). Op. cit.

[4] Das, S. (2018, 30 July). 8 differences between Indian data protection bill and GDPR. *CIO & Leader*. https://www.cioandleader.com/article/2018/07/30/8-differences-between-indian-data-protection-bill-and-gdpr

[5] NITI Aayog. (2018). Op. cit.

[6] Khera, S. (2019, 21 January). Artificial intelligence in education in India, and how it's impacting Indian students. *The News Minute*. https://www.thenewsminute.com/article/artificial-intelligence-education-and-how-its-impacting-indian-students-95389

[7] NITI Aayog. (2018). Op. cit.

Court on the use of Aadhaar biometric information[8] for exercising state functions of public service delivery highlights the exemptions from informed consent or other complementary data protection protocols for state functions aimed at social and economic inclusion. This begs the question as to how the sensitive personal data of citizens, especially of the marginalised, are to be protected within institutional applications of AI through PPP arrangements which lack transparency on the commercial service commitments, data protection protocols and safeguards, data-sharing arrangements and processes of labelling and annotation. These are further compounded by the gaps in explainability, framing, deployment and application.

One of the most pervasive problems within the Indian public education system has been low retention rates beyond primary education, with even lower rates for girls.[9] In one of the first institutional applications of AI in the social sector in the country, the Government of Andhra Pradesh,[10] in partnership with Microsoft, implemented machine learning and analytics through its Azure cloud platform to predict and prevent public school drop-outs. This report aims to use this partnership as a case study to throw into sharp relief the contextual parameters and questions that must be taken into account when evaluating institutional applications of AI in society and developing ethical governance frameworks that can answer to contextual nuances of the application taking cognisance of the actual incidence of its impact. It might also help highlight issues that would be helpful for future deployments in the sector to address, given that the NITI Aayog plans to scale up this project with Microsoft on the basis of the Andhra Pradesh experience.[11]

Identifying parameters of algorithmic decision making and its implications for justice and inclusion

Literacy levels in Andhra Pradesh have been the second lowest in the country, with one of the highest percentages of school drop-outs, most of whom come from farming families or those involved in agriculture.[12] It also topped the list of the highest number of female school drop-outs, with seven out of 10 girls dropping out of school before they reach the 10th standard.[13] In a partnership with the Government of Andhra Pradesh, Microsoft offered its Azure cloud computing platform with machine-learning and analytics capabilities as a part of the overall Cortana Analytics Suite (CAS)[14] to develop a predictive model for identifying school drop-outs in the state. The project commenced with a pilot of a little over 1,000 schools and 50,000 students and has now been rolled out to all 13 districts covering 10,000 schools and five million children. The aim of the project was for the information gathered and analysed to be made available to district education officers and school principals who could then deploy targeted interventions and customised counselling.[15]

Data was triangulated from three databases in order to build the data pipeline for the project. This included the Unified District Information System for Education (U-DISE), containing school infrastructure information and the data on teachers and their work experience, education assessment data from multiple sources, and socioeconomic data from the UIDAI[16] Aadhaar system.[17] By aggregating these multiple data points from different sources, the aim of the project is to track the students' journey through

[8] The Aadhaar is a 12-digit unique identification system based on biometric information and demographics issued to an Indian resident. It is governed by the Aadhaar (Targeted Delivery of Financial and Other Subsidies, Benefits and Services) Act, 2016. It became controversial when mobile phone service providers and banks started asking for the card as a condition for using their services. More problematically, it became conditional for the delivery of critical social protection schemes like midday meals to underserved students, availing rationed food items, pension schemes, etc., in some cases with people denied these services dying. The card was the subject of cases filed before the Supreme Court of India which challenged its constitutional validity due to its privacy infringing features and that it was being required to access private sector services. Though the Supreme Court upheld the fundamental right to privacy, in September 2018 it also upheld the constitutional validity of the identification system in that it allowed Aadhaar-based authentication for establishing the identity of an individual for receipt of a subsidy, benefit or service provided by the government by retaining section 7 of the Aadhaar Act that allows for welfare to be made contingent on the production of Aadhaar.

[9] Taneja, A. (2018, 31 January). The high drop out rates of girls in India. *Live Mint*. https://www.livemint.com/Opinion/iXWvKng7uU4L8vo5XbDn9I/The-high-dropout-rate-of-girls-in-India.html

[10] Andhra Pradesh is a state in southern India.

[11] Agha, E., & Gunjan, R. K. (2018, 28 April). NITI Aayog, Microsoft Partner Up to Predict School Dropouts Using Artificial Intelligence. *News18*. https://www.news18.com/news/india/niti-aayog-microsoft-partner-up-to-predict-school-dropouts-via-artificial-intelligence-1732251.html

[12] India Today. (2016, 23 April). Education survey shows the poor state of Telengana, Andhra Pradesh. *India Today*. https://www.indiatoday.in/education-today/news/story/andhra-pradesh-education-319526-2016-04-23

[13] Baseerat, B. (2013, 10 October). Andhra tops in girl school dropouts: Activists. *Times of India*. https://timesofindia.indiatimes.com/city/hyderabad/Andhra-tops-in-girl-school-dropouts-Activists/articleshow/23937897.cms

[14] Cortana Analytics Suite is the fully managed big data and advanced analytics suite.

[15] Srivas, A. (2016, 10 May). Aadhaar in Andhra: Chandrababu Naidu, Microsoft have a plan for curbing school dropouts. *The Wire*. https://thewire.in/politics/aadhaar-in-andhra-chandrababu-naidu-microsoft-have-a-plan-for-curbing-school-dropouts

[16] The Unique Identification Authority of India is the entity mandated to issue the 12-digit Aadhaar number and manage the Aadhaar database.

[17] Srivas, A. (2016, 10 May). Op. cit.

the education system by providing a 360-degree view of students after mapping close to 100 variables. Initial results from the project reaffirmed longstanding notions behind school drop-outs. These include girls being more likely to drop out in the absence of adequate toilet facilities, higher drop-out rates among students failing to score well in key subjects like English and mathematics, which reduces their faith in formal education, along with the role of the socioeconomic status of the family and the wider community to which the student belongs.[18] In a study based on the National Family and Health Survey-3, it was found that drop-outs tended to be higher among children belonging to minority Muslim families, scheduled castes and scheduled tribes.[19] Further, children belonging to illiterate parents were four times more likely to drop out than those belonging to literate parents. The possibility of children of non-working parents dropping out is also relatively high.[20]

Anil Bhansali, the managing director of Microsoft Research and Development, had told the online news outlet *The Wire* in 2016 that the CAS suite deployed in the project "can provide a lot of useful insight as long as you pump in the data and the right modelling,"[21] with "right modelling" being the operational phrase. With algorithmic decision making coming to play an increasingly significant role in institutionalising individual and systemic bias and discrimination within social systems, it becomes important to evaluate the processes through which these are pervasively deployed.

Data choices: The pilot project was restricted to students of the 10th standard. This is because, according to Bhansali, the 10th standard represents one of the few inflexion points when one takes their first standardised tests and after which a reasonable number of students drop out on their way to 11th standard. Another likely reason is that 10th standard results are already online and the education department has access to gender and subject grading data through examination hall tickets.[22] Educational assessment information for lower classes entails the herculean task of having to be digitised in order to be of use in a machine-learning system.[23]

However, it is also the case that the drop-out rates are the highest in secondary education (standards 9 and 10),[24] coinciding with the completion of standard 8 after which midday meals are no longer provided, which are a major factor driving school attendance.[25] Those who continue beyond standard 8 to reach standard 10 show a comparative degree of resilience to the non-provision of these sorts of interventions aimed at ameliorating the disadvantageous socioeconomic conditions behind school drop-outs. Therefore, using such data as a training model for the system misplaces the inflexion point and thereby undermines other structural and intersectional socioeconomic issues driving high rates of school drop-outs during the transition to secondary education from standard 8 to standard 9. This leads to elision of the structural socioeconomic parameters that constrain equitable access to resources.

In addition, the U-DISE database containing information about teachers' work experience does not necessarily map the effectiveness or efficacy of a given teacher and their contribution to better learning outcomes well.

Modelling and inferences: Data choices are not the only criteria determining the questions on inclusion and justice. Decision making regarding the input processes that develop the statistical models and inferences made are equally significant in determining the incidences of impact that a given machine-learning project is likely to have in the areas of its intervention.[26] Given the lack of transparency on the decision-making process, the insights gained from news reports on the subject show that the input process in developing the model involved a combination of existing knowledge, beliefs, and findings about the factors driving school drop-outs, coupled with the convenience of digitised data.[27] Since the extent to which these data were interpreted with bias during the input

[18] Ibid.

[19] Scheduled Castes and Scheduled Tribes are officially designated historically marginalised groups in India recognised in the Constitution of India.

[20] M., Sateesh, & Sekher, T. V. (2014). Factors Leading to School Dropouts in India: An Analysis of National Family Health Survey-3 Data. *International Journal of Research & Method in Education, 4*(6), 75-83. https://www.researchgate.net/publication/269932850_Factors_Leading_to_School_Dropouts_in_India_An_Analysis_of_National_Family_Health_Survey-3_Data

[21] Srivas, A. (2016, 10 May). Op. cit.

[22] Examination hall tickets offer rights of admission to a test taker during state or national-level examinations. They contain details of the student like the identity number assigned to the student for the examination, a photograph and a signature along with details of the examination such as location and room (where applicable). Sometimes they also contain the student's name and date of birth. There is no standard format for examination hall tickets and they differ from examination to examination.

[23] Srivas, A. (2016, 10 May). Op. cit.

[24] PRS Legislative Research. (2017, 2 October). Trends in school enrolment and drop-out levels. *Live Mint*. https://www.livemint.com/Education/k1ANVHwheaCFWCupY3jkFP/Trends-in-school-enrolment-and-dropout-levels.html

[25] Jayaraman, R., Simroth, D., & de Vericourt, F. (n.d.). The Impact of School Lunches on Primary School Enrollment: Evidence from India's Midday Meal Scheme. *Indian Statistical Institute, Delhi Centre*. www.isid.ac.in/~pu/conference/dec_10_conf/Papers/RajiJayaraman.pdf

[26] Algorithm Watch. (2019, 6 February). 'Trustworthy AI' is not an appropriate framework. https://algorithmwatch.org/en/trustworthy-ai-is-not-an-appropriate-framework

[27] Srivas, A. (2016, 10 May). Op. cit.

process is unclear, the extent to which biased socioeconomic profiles based on caste, gender and religion played a role in determining the drop-out rate, versus structural and institutional barriers, is also unclear. Moreover, it is not unlikely that such models can then influence seat allocations within higher education and government services based on such profiles, undermining India's constitutionally guaranteed affirmative action protections for marginalised and vulnerable groups.

This highlights the problem of using existing knowledge and statistics in an ahistorical and acontextual manner without duly quantifying the structural and institutional indicators that produce such inequalities in the first place. For example, if the model shows that a Scheduled Tribe girl from Jharkhand is more likely to drop out of school in the absence of a targeted intervention, could this lead to fewer seats allocated in higher education, and reservations in government services for women from the community? Moreover, coupled with data choices, it is unclear to what extent the data trained on standard 10 would be effective in predicting drop-out rates in the transition phase from upper primary (standard 8) to secondary school (standards 9 and 10) where arguably the driving factors are more structural and institutional as compared to performance in a given set of subjects.

Service agreements and data protection: Data sharing within PPPs is unclear due to a lack of transparency, especially in a country like India, which is yet to have its own data protection law but harbours high aspirations of becoming the world leader in AI adoption, deployment and innovation.[28] Bhansali has said that the data is stored in data centres located in India and is tied to the Andhra Pradesh government's account, and that Microsoft cannot own it or repurpose it.[29] While Microsoft might not own or repurpose the data, it is unclear whether it does or does not have the same rights over the insights generated out of processing of such data. It is also not clear what bespoke – or customised – data protection safeguards were incorporated, if at all, within the public-private agreements. An evaluation study of Microsoft's cloud computing shows that irrespective of the geographical location that a customer selects to locate their data, Microsoft warns that customers' data, including personal data, may be backed up in the United States (US) by default. Moreover, if any beta or pre-release Microsoft software was used or there was back-up of web or worker role software[30] in any of its cloud services, data would be stored or replicated in the US.[31] The movement and replication of data increases the attack surface. These fears are not allayed given that Microsoft is the second most targeted entity after the Pentagon,[32] and Andhra Pradesh leads in the leakages of sensitive personal data of its constituents.[33]

Conclusion

Given that AI systems are increasingly aiding state institutions in the allocation of resources, it becomes imperative that they align with principles of non-discrimination rather than perpetuate existing misallocations by creating pervasive systems of privilege being trained on unrepresentative data sets and models.

Significant international multilateral and multistakeholder attention has been diverted towards developing ethical governance frameworks for AI. This includes the OECD Principles on AI,[34] the European Commission Ethical Guidelines for Trustworthy AI,[35] as well as the Toronto Declaration: Protecting the Right to Equality and Non-Discrimination in Machine Learning Systems,[36] along with attention in other digital policy global initiatives like the United Nations High-Level Panel on Digital Cooperation.

However, these provide broad-based principles without adequately applied examples, which delimit

[28] NITI Aayog. (2018). Op. cit.

[29] Srivas, A. (2016, 10 May). Op. cit.

[30] "Web Role is a Cloud Service role in Azure that is configured and customized to run web applications developed on programming languages/technologies that are supported by Internet Information Services (IIS), such as ASP.NET, PHP, Windows Communication Foundation and Fast CGI. Worker Role is any role in Azure that runs applications and services level tasks, which generally do not require IIS. In Worker Roles, IIS is not installed by default. They are mainly used to perform supporting background processes along with Web Roles and do tasks such as automatically compressing uploaded images, run scripts when something changes in the database, get new messages from queue and process and more." Source: https://cloudmonix.com/blog/what-is-web-and-worker-role-in-microsoft-azure

[31] Calligo. (n.d.). *Microsoft Azure and Data Privacy*. https://calligo.cloud/wp-content/uploads/Azure-Data-Privacy-Stack.pdf?utm_campaign=Hybrid%20Azure&utm_source=hs_automation&utm_medium=email&utm_content=69182393&_hsenc=p2ANqtz-8e2YnyrNnNjouMWxrv80aYaLLlso_vi8apwlbq3HTVVRqgl2WY94jmAKBStWuDTwC-U1F_NPubB4SltezcA43mZl1YTw&_hsmi=69182393

[32] Ibid.

[33] See, for example: MediaNama. (2019, 29 May). Andhra Pradesh exposes Aadhaar of farmers - once again. *MediaNama*. https://www.medianama.com/2019/05/223-andhra-pradesh-exposes-aadhaar-of-farmers-once-again; Jalan, T. (2018, 27 August). CCE-Andhra Pradesh leaks students' gender, caste, quota, Aadhaar data on website. *MediaNama*. https://www.medianama.com/2018/08/223-apcce-students-aadhaar-exposed; Tutika, K. (2018, 20 March). Aadhaar data leak of Andhra Pradesh women raises security concerns. *The New Indian Express*. www.newindianexpress.com/states/andhra-pradesh/2018/mar/20/aadhaar-data-leak-of-andhra-pradesh-women-raises-security-concerns-1789648.html

[34] https://www.oecd.org/going-digital/ai/principles

[35] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

[36] https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems

their uptake and applicability and serve to act as an "alternative or preamble to regulation", thereby diluting "state accountability and rights-based obligations."[37] These also serve to act as light-touch non-discrimination norms that provide the leeway for businesses to not actually engage with non-discrimination principles within data choices, modelling, design and application, thereby ending up entrenching discrimination by making inequalities institutionally pervasive.

A second approach, which is a technical approach, aims to ensure fairness, accountability and transparency (FAT) in AI systems. However, the FAT approach fails to identify structural socioeconomic indicators to contextualise the principles of non-discrimination within systems design.[38]

It has been argued that multilateral commitments to universally agreed human rights principles with regard to AI would serve to strengthen the intended application of both these approaches.[39] However, all approaches must be accompanied with evidence-based case studies to develop principled processes like algorithm impact assessments, explainability, transparency of commercial contracts, etc. with a clear understanding of learnings from use cases, and the role of different stakeholders within the process, rather than principled outcomes like trustworthy AI and fair and ethical machine-learning systems.

Action steps

The following advocacy steps are suggested for India:

- *Risk sandboxing*: Regulatory sandboxing[40] and data sandboxing[41] 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. Risk sandboxing is envisaged as a natural progression from regulatory sandboxing in which the product is tested for its decision-making impact on vulnerable and marginalised populations on the basis of non-discrimination principles.

- *First stage process-based transparency*: While there has been much discussion about the need for explainable AI to counter the black-boxing phenomenon underlying AI's opaque decision-making process, there needs to be a first-level transparency with respect to the inputs into the model development process, data choices, and platform capabilities and jurisdictions.

- *Disclosure of service agreements:* There need to be disclosure of service agreements within PPPs deploying AI technologies to understand the data protection commitments and data-sharing practices.

- *Mapping contextual parameters of knowledge used in modelling:* Studies that constitute knowledge about a given subject area are the result of divergent research objectives which should be evaluated for their relevance and bearing to the machine-learning system being deployed before they are factored into predictive modelling. Moreover, socioeconomic and structural indicators – such as caste, gender and family income in conjunction with how that caste group fares overall in the economy – must be identified and mapped into the model along with transparency on the decision making that maps these indicators to train the machine learning system.

- *Representative data choices:* The data on which a machine-learning system is trained must be representative of the population in which it is to be deployed.

- *Recommendations structured on non-discrimination principles:* Recommendations must be structured on non-discrimination principles. This should be done, for example, to avoid instances when an AI system recommends fewer STEM (science, technology, engineering and mathematics) courses for women because the data shows women are less likely to take up STEM subjects.

[37] ARTICLE 19. (2019). *Governance with teeth: How human rights can strengthen FAT and ethics initiatives on artificial intelligence.* London: ARTICLE 19. https://www.article19.org/wp-content/uploads/2019/04/Governance-with-teeth_A19_April_2019.pdf

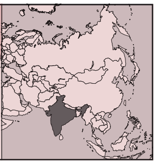
[38] Ibid.

[39] Ibid.

[40] Regulatory sandboxing allows for a controlled environment with relaxed regulations that allows a product or innovation to be thoroughly tested out before being released for public use. It involves a set of rules that allow innovators to test their products within a limited legal environment subject to pre-defined restrictions like limitation on exposure, time-limited testing, pre-defined exemptions, and testing under regulatory supervision. Source: https://cis-india.org/internet-governance/files/ai-in-india-a-policy-agenda/view; pubdocs.worldbank.org/en/770171476811898530/Session-4-Pavel-Shoust-Regulatory-Sandboxes-21-09-2016.pdf

[41] Data sandboxes allow companies to access large anonymised data sets under controlled circumstances to enable them to test their products and innovations while keeping in mind privacy and security compliance requirements.

INDIA



INTERROGATING "SMARTNESS": A CASE STUDY ON THE CASTE AND GENDER BLIND SPOTS OF THE SMART SANITATION PROJECT IN PUNE, INDIA

Independent; ARTICLE 19
Malavika Prasad and Vidushi Marda
www.article19.org

Introduction

Pune, a city of 6.4 million people in western India, is home to a smart sanitation project that aims at building resilient and sustainable sanitation solutions in the city. Artificial intelligence (AI) is one of the potentially transformative technologies currently being considered in this project.[1] Real-time collection and monitoring of data through sensors, as well as analytics and insights at scale, means that the use of AI systems can be beneficial.

Current data collection through sensors will impact the design and development of AI systems in the future. Questions of *how* this data is created, *where* it arises from, *what* types of analytics and insights are being recorded, *which* areas of work are made more efficient, and *who* benefits from smart sanitation remain to be answered.

This report aims to highlight the importance of situating AI systems in context by analysing how Pune's smart sanitation project interacts with the societal fabric within which it is being developed. We explore two pointed questions. First, what AI systems are planned to be deployed and how are they being designed and developed? And second, how does this impact caste and gendered systems of labour undergirding sanitation work in India?

Context

Pune has consistently ranked as one of India's top smart cities since the launch of the Smart Cities Mission (SCM).[2] In 2017, the Pune Smart City Development Corporation signed a memorandum of understanding with the University of Toronto and the Indian Institute of Technology (IIT) Bombay to use AI to make it a truly smart city.[3] This is in line with the National Strategy for AI, published by the National Institution for Transforming India (NITI Aayog) in June 2018, which identifies smart cities as a key area for AI intervention in India.[4] In August 2017, Pune's Municipal Corporation announced plans to make Pune the world's first smart sanitation city.[5]

Pune's smart sanitation project lies in the intersection of five key imperatives of the Indian government. The first is the SCM, a centrally sponsored policy seeking to enhance 100 cities by improving the "quality of life" in a "sustainable environment" using "smart solutions",[6] while driving economic growth.[7] Smart solutions envisaged under this policy range from e-governance solutions such as e-service delivery and video crime monitoring, energy management solutions such as smart meters and harnessing renewable energy, and waste management solutions such as the circular use of waste as energy and compost and the reuse and recycling of water.[8] The second is the Swacch Bharat Mission (SBM), also a centrally sponsored policy, which aims to achieve safe, sustainable, cost-effective and "universal sanitation coverage", making India "open defecation free"[9] by October 2019. The third is Digital India, an initiative that is closely related to SCM,[10] and is geared towards providing digital infrastructure to every citizen as a core utility, and

[1] Toilet Board Coalition. (2018). *Smart Sanitation City.* https://www.toiletboard.org/media/45-TBC_2018PuneReport_11202018.pdf?v=1.0.1

[2] Ministry of Housing and Urban Affairs, Government of India. (2016). *City Challenge.* http://smartcities.gov.in/content/city_challenge.php?page=winning-city-proposals-in-round-1-of-city-challenge.php; also see http://smartcities.gov.in/upload/city_challenge/58dfa4cb13064582318f5d6d8eRankingofSmartCities(1).pdf

[3] Smart Cities Council. (2017, 5 December). AI to make Pune a truly "Smart City". https://india.smartcitiescouncil.com/article/ai-make-pune-truly-smart-city

[4] NITI Aayog. (2018). National Strategy for Artificial Intelligence #AIFORALL. https://www.niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf

[5] Express News Service. (2017, 1 September). PMC partners with TBC to become 'Smart Sanitation City'. *The Indian Express.* https://indianexpress.com/article/india/pmc-partners-with-tbc-to-become-smart-sanitation-city-4823309

[6] Guideline 2.3, Ministry of Urban Development, Smart City Mission Statement and Guidelines, 2015 ("SCM Guidelines").

[7] SCM Guideline 2.6.

[8] SCM Guideline 2.5.

[9] The term was defined as follows: "ODF is the termination of faecal-oral transmission, defined by: 1) no visible faeces found in the environment/village; and 2) every household as well as public/community institutions using safe technology option for disposal of faeces." See Government of India Letter no.@-11011/3/2015-SBM dated 9 June 2015.

[10] IANS. (2017, 16 March). Digital India initiatives playing major role in smart cities mission. *The Financial Express.* https://www.financialexpress.com/india-news/digital-india-initiatives-playing-major-role-in-smart-cities-mission/590381

transforming India into a digitally empowered society and knowledge economy.[11] The fourth is the National Urban Innovation Stack (NUIS), a central government initiative using layered digital infrastructure to provide all stakeholders with "digital tools and platforms, standards, specifications and certifications, and enable greater coordination and integration amongst them."[12] The fifth is the National Strategy for Artificial Intelligence,[13] which contemplates the need for AI to propel inclusive economic growth and social development, with smart cities as a key area of AI intervention.

The government views the convergence of these policy imperatives as desirable.[14]

Our report focuses on sanitation workers – largely women from Dalit communities[15] – who are hired by the municipality as labourers through a contractor and therefore often fall outside of labour laws.[16] Sanitation work refers broadly to all *safai kam* or "cleaning work" ranging from sweeping streets, collecting and transporting garbage, to cleaning sewers.[17] Cleaning workers are considered polluted and impure by upper castes, due to which Dalit communities,[18] treated as ordained by birth for such labour, are excluded from other occupational opportunities.[19] That sanitation workers are largely Dalit women means that they are marginalised in an altogether qualitatively different manner: by upper-caste women and men on account of being Dalit,[20] and by Dalit men on account of being women.[21]

Some states have introduced government employment for sanitation work, open to persons from all castes. Non-Dalit candidates have filled about half of these posts[22] because of the economic benefits and job security in government posts.[23] However, some upper-caste candidates use these posts as a stepping stone to permanent posting, after which they transfer to other governmental departments.[24] As succour to Dalit workers who have been dispossessed by non-Dalits, some states have begun to mandate that Dalit communities shall be prioritised for sanitation posts.[25] Nevertheless, governmental intervention in this realm has further marginalised Dalits, without undoing the stigma attached to cleaning labour. The key issue that states have failed to address is that Dalit workers who used to earn a living from sanitation work are unable to access alternative opportunities for work.[26]

Indian law forbids employment and even contractual engagement for one type of

[11] Press Information Bureau, Government of India. (2014, 20 August). *Digital India – A programme to transform India into digital empowered society and knowledge economy.* https://pib.gov.in/newsite/PrintRelease.aspx?relid=108926

[12] Ministry of Housing and Urban Affairs. (2019). *National Urban Innovation Stack.* https://smartnet.niua.org/sites/default/files/resources/national_urban_innovation_stack_web_version.pdf

[13] NITI Aayog. (2018). Op. cit.

[14] SCM Guideline 14.2.

[15] Kadlak, H., Salve, P. S., & Karwade, P. (2019, 19 March). Intersectionality of Caste, Gender and Occupation: A Study of *Safai Karamchari* Women in Maharashtra. *Contemporary Voice of Dalit.*

[16] Yadavar, S. (2017, 17 June). Sanitation Workers Clean Our Cities But They Are Denied Even Minimum Wage. *India Spend.* https://archive.indiaspend.com/indias-great-challenge-health-sanitation/sanitation-workers-clean-our-cities-but-they-are-denied-even-minimum-wage-72329; Fernandes, S. (2019, 4 March). How Mumbai's Sanitation Workers Fought the Municipal Corporation - and Won. *The Scroll.* https://scroll.in/article/914733/how-mumbais-sanitation-workers-fought-the-municipal-corporation-and-won

[17] Kadlak, H., Salve, P. S., & Karwade, P. (2019, 19 March). Op. cit.

[18] The Constitution of India abolishes the practice of untouchability and requires parliament to punish the enforcement of disabilities arising from untouchability (Article 17). To this end, the Scheduled Caste and Scheduled Tribe (Prevention of Atrocities) Act, 1989, criminally punishes practices arising from the system of untouchability. Nonetheless, untouchability manifests in new ways.

[19] B. R. Ambedkar, political theorist, chairman of the Drafting Committee of the Indian Constitution and a towering caste scholar has this to say: "[I]t is clear that according to the Hindu Shastras and the Hindu notions, even if a Brahmin did scavenging, he would never be subject to the disabilities of one who is born a scavenger. In India, a man is not a scavenger because of his work. He is a scavenger because of his birth irrespective of the question whether he does scavenging or not." Ambedkar, B. R. (1945). *What Congress and Gandhi Have Done to the Untouchables.* Bombay: Thacker and Co.

[20] The caste system is built on the subjugation of all women, as Ambedkar had argued as early as 1916. However, the feminist movement in India has not successfully reckoned with this reality and has been criticised for engaging in advocacy using an upper-caste lens. See Ambedkar, B. R. (1917). Castes in India: Their Mechanism, Genesis and Development. *Indian Antiquary, XLI.* http://www.columbia.edu/itc/mealac/pritchett/00ambedkar/txt_ambedkar_castes.html

[21] The separability of caste and sex is easy for those who are subordinated on the basis of one marker and privileged by the other (such as upper-caste women), but is virtually impossible for those who are either privileged (upper-caste men) or subordinated by both markers (Dalit women). See MacKinnon, C. (2016). *Sex Equality*. (Third edition). Foundation Press.

[22] Tripathi, T. (2012). Safai Karmi Scheme of Uttar Pradesh: Caste Dominance Continues. *Economic and Political Weekly, 47*(37). https://www.epw.in/journal/2012/37/commentary/safai-karmi-scheme-uttar-pradesh.html

[23] Tripathi, T. (2015). Safai Karmis of Uttar Pradesh. *Economic and Political Weekly, 50*(6). https://www.epw.in/journal/2015/6/reports-states-web-exclusives/safai-karmis-uttar-pradesh.html

[24] Roytalukdar, R. (2019, 28 January). #Republic of Grit: Who gets Coveted Government Sanitation Jobs in Rajasthan? *The Wire.* https://thewire.in/caste/republic-of-grit-who-gets-coveted-government-sanitation-jobs-in-rajasthan

[25] Jain, S. (2019, 29 April). Manual Scavengers in Rajasthan Struggle to Be Recruited as Govt Sanitation Workers. *The Wire.* https://thewire.in/labour/manual-scavengers-in-rajasthan-struggle-to-be-recruited-as-govt-sanitation-workers

[26] There is some evidence that the State of Mizoram has successfully dislodged the stigma attached to cleaning work by hiring tier four government employees who are required to serve in turns as sanitation workers, peons, office assistants, etc. Pisharoty, S. B. (2019, 20 March). For Sanitation Workers in Aizawl, Stigma Isn't a Problem. *The Wire.* https://thewire.in/rights/for-sanitation-workers-in-aizwal-stigma-isnt-a-problem

sanitation work called "manual scavenging"[27] – the manual cleaning of faecal matter – owing to the caste-based origin of confining this labour to only certain Dalit communities.[28] However, the prohibition on such hiring is lifted if protective gear or devices are provided to the workers.[29] From a human rights perspective, as Special Rapporteur Leo Heller reported to the 39th session of the UN Human Rights Council,[30] protective gear does not eliminate the stigma associated with manual scavenging or cleaning labour, which continues to be the only occupational opportunity available to 1.3 million Dalits in India.[31] While "eradication of manual scavenging" was stated as a policy imperative in the SCM guidelines,[32] further details were not forthcoming until 2017.[33] The scant guidance in the revised 2017 guidelines does not have anything to say on the varieties of sanitation work outside manual scavenging.[34]

Designing and developing Pune's smart sanitation: A closer look

The Union Ministry of Urban Development stated in May 2016 that it did not plan to privatise the management of basic utilities such as sewage treatment in smart cities.[35] However, private sector involvement is key to the financing of a smart city plan under the SCM.[36] In Pune, the "Smart Sanitation Economy" was the vision[37] of the Toilet Board Coalition (TBC) – a business-led partnership[38] that is offering technical assistance to the Pune Smart City and the Pune Municipal Corporation (PMC).[39] The "Smart Sanitation Economy" aims to lower costs of delivering sanitation by tapping a yet untapped market of health care and other services around the sanitation system.[40] The other two mutually reinforcing "economies" of interest for the TBC are the "Toilet Economy", comprising businesses looking to buy and sell toilet products and services, and the "Circular Sanitation Economy", which replaces traditional waste management practices, and comprises businesses capturing toilet resources such as urine and faecal matter for producing fuel, reusable water, compost and organic fertilisers, bio-plastics, etc.[41]

Our research and interviews indicate that in the "Toilet Economy", toilets will be privately established and operated. Sanitation workers will thus be hired by private entities, either as employees or as contract labourers, and it is unclear what legal regime will regulate their hire.[42]

Democratic accountability and responsiveness

The question of democratic accountability in smart cities generally is vexing. In the Pune case, the Pune Smart City is a special purpose vehicle (SPV) – a limited company held by the state and the PMC[43] – to whom rights, obligations and powers of the PMC are to be delegated.[44] However, this requirement does not convey the exact terms of the relationship between the PMC and the SPV, or delineate

[27] Section 5 of the Prohibition of Employment as Manual Scavengers and their Rehabilitation Act, 2013 (2013 Act).

[28] "Manual scavenging" is defined as the manual "cleaning, carrying, disposing of, or otherwise handling in any manner, human excreta" in any insanitary latrine or open drain or pit or railway track or other such spaces notified by the States or Central Government, under Section 2(g) of the 2013 Act .

[29] See Explanation (b) of Section 2(g) of the 2013 Act. It is worth noting that those sanitation workers hired for manual scavenging with protective gear, if hired as contract labourers, continue to remain outside labour law protections.

[30] Report of the Special Rapporteur on the human rights to safe drinking water and sanitation on his mission to India from 27 October to 10 November 2017, 6 Jul 2018, A/HRC/39/55/Add.1, Paragraph 25.

[31] International Dalit Solidarity Network, Manual Scavenging. https://idsn.org/key-issues/manual-scavenging/

[32] Guidelines for SBM (Gramin) 2014.

[33] The Revised Guidelines for SBM (Urban), 2017 state in one line: "The State Governments shall pursue the following: i. All manual scavengers in urban areas are identified, insanitary toilets linked to their employment are ii. upgraded to sanitary toilets, and the manual scavengers are adequately rehabilitated." Para 6.4.14 of the Revised Guidelines for SBM (Gramin), 2017, merely forbid the construction of insanitary latrines and mandate conversion of existing ones to sanitary latrines. See also Updated Guidelines for SBM (Gramin), 2019.

[34] See also Interview of Bezwada Wilson, National Convenor of the Safai Karmachari Andolan and Beena Pallical, Chair of the National Campaign for Dalit Human Rights by Newslaundry, November 2017. https://www.youtube.com/watch?v=GjqT7rwwtcY&t=625s

[35] Lok Sabha Unstarred Question No. 2964, to be answered on 11 May 2016.

[36] SCM Guideline 9.1.2.

[37] Toilet Board Coalition. (2018). Op. cit.

[38] The TBC is a "business-led partnership and platform" connecting "private, public and non-profit sectors" towards achieving Sustainable Development Goal 6 for universal access to sanitation. https://www.toiletboard.org/about

[39] Guideline 6.3.2 of the SCM Guidelines defines Handholding Agencies for the proposal stage of the Smart City Mission. The Toilet Board Coalition was engaged well into Pune's proposal being selected in the SCM.

[40] Toilet Board Coalition India Roundtable, 15 November 2017. https://www.youtube.com/watch?v=VUKB1WDJBjQ

[41] Toilet Board Coalition. (2018). Op. cit.

[42] Since the regulation of contract workers is less onerous than that of workers hired directly as employees, most establishments prefer to hire contract workers. If more than 20 workers are hired on contract, the Contract Labour Act, 1970, will regulate their hire. However, in cities such as Hyderabad, contractors have evaded this regulation by offering less than 20 workers for hire. Other cities have followed suit. See Yadavar, S. (2017, 17 June). Op. cit.

[43] SCM Guideline 10.2 requires that the SPV be held in a 50:50 ratio by the State and the ULB in question.

[44] Para 4, Annexure 5, SCM Guidelines.

the hierarchy of authorities.[45] Members of the Pune Smart City are largely bureaucrats, and to a lesser extent politicians, assisted by consultants for technical support, monitoring and evaluation, fund-raising from the market, as well as procuring implementation agencies.[46] Such conferral of decision-making powers in unelected bureaucrats raises questions about the democratic legitimacy of the smart city.[47]

Scholars have also expressed concerns about the class of citizens who will be consulted in the design of smart cities.[48] What demographics are represented at Pune Smart City's "Citizen Engagement meetings" remains to be studied. At least two such meetings for discussing smart city projects were conducted in housing-societies in the local area chosen for smart city development,[49] which comprise citizens with home-owning or renting capabilities. We are unable to find evidence of the attendance of citizen labourers servicing the smart city. The absence of participatory planning, failure to account for all citizens' needs or conduct social audits has effaced the "citizen-government interface" even in past urban-renewal missions in India.[50] India needs to reckon with this deficit in the democratic responsiveness of smart cities to citizen residents.

The deployment of "smart" solutions such as sensors in internet of things (IoT)-enabled toilets as a business use case – designed to maximise efficiency – as opposed to a public sector use case that accounts for human costs, further reduces democratic responsiveness in smart cities. The central concern is that privately deployed efficiency maximising systems need not reckon with human rights baselines that public sector systems do, such as, in this case, the dignity of workers who are confined to such labour owing to their caste identity.

The TBC's approach is concerned with the citizen toilet users at its core, because it is a business-led coalition attempting to tap a customer base of people without toilet access.[51] However, the state (i.e. the Pune Smart City) is constitutionally required to be accountable to all citizens, including citizen labourers. What instead appears to be happening is a dispersion of governmental power across multiple private, democratically unaccountable actors – from toilet operators to the toilet business in question – who manage and discipline sanitation workers to provide more efficient maintenance.[52] Since the Pune SCM's toilets will be privately operated, it is unclear how the toilet businesses will be held to norms of constitutional accountability. For instance, what legal recourse exists for sanitation workers who have been displaced from their previous jobs in the municipality's public toilets, in areas that now have privately operated "smart toilets"? Will toilet businesses running smart toilets be held to the constitutional obligation not to entrench caste-based sanitation work, or to ensure sanitary work conditions? Will the Pune SCM be truly democratically responsive to all its citizens, including citizen labourers from marginalised groups?

Sensors: Governmentality and unrepresentative training data

Sensors[53] form a fundamental building block of Pune's smart sanitation project by "enabling the collection of new data, feeding new insights, and

[45] Anand, A., Sreevatsan, A., Taraporevala, P. (2018). *An Overview of the Smart Cities Mission in India*. New Delhi: Centre for Policy Research. https://cprindia.org/system/tdf/policy-briefs/SCM%20POLICY%20BRIEF%2028th%20Aug.pdf?file=1%26type=node%26id=7162; also see Taraporevala, P. (2017, 6 September). How Smart Cities Mission can help municipalities to improve governance. *The Hindustan Times*. https://www.hindustantimes.com/opinion/how-smart-cities-mission-can-help-municipalities-to-improve-governance/story-mTV2uXWofuiVlgL7A9cXPK.html

[46] https://punesmartcity.in/about-pscdcl

[47] Anand, A., Sreevatsan, A., Taraporevala, P. (2018). Op cit.

[48] Ibid. Also see Hoelscher, K. (2016). The evolution of the smart cities agenda in India. *International Area Studies Review, 19*(1), 28-44.

[49] Pune Smart City. (2018, 19 December). Pune Smart City continues with its Citizen Engagement Program campaign in Housing Societies to build awareness regarding the mission. https://punesmartcity.in/सटिीझन-एंगेजमेंट-कार्यक्/; Pune Smart City. (2018, 19 November). Pune Reaches Out To Housing Societies for Awareness Through Citizens Engagement Program, https://punesmartcity.in/पुणे-स्मार्ट-सटिीच्या-प्/

[50] Roy, S. (2016). The Smart City Paradigm in India: Issues and Challenges of Sustainability and Inclusiveness. *Social Scientist, 44*(5-6), 29-48.

[51] Cheryl Hicks, executive director of the TBC, stated: "We are a corporate-led coalition started because companies saw a business opportunity in sanitation. The 2.3 billion people without toilets are customers that they don't have." When asked about the large numbers of the urban poor, she responded: "That is why the companies formed the coalition, because the people without toilets, the poor, are customers they don't have." Civil Society. (n.d.). 'Sanitation is the fortune at the bottom of the toilet'. https://www.civilsocietyonline.com/business/sanitation-is-the-fortune-at-the-bottom-of-the-toilet

[52] Foucault, M. (1978). Lectures Three and Four. In M. Senellart (Ed.), *Security, Territory, Population: Lectures at the Collège de France 1977-1978*. Palgrave Macmillan. See also other Foucauldian critiques of private sector technological solutions in smart cities such as Vanolo, A. (2014). Smartmentality: The Smart City as Disciplinary Strategy. *Urban Studies, 51*(5), 883-898; Klauser, F., Paasche, T., & Söderström, O. (2014). Michel Foucault and the Smart City: Power Dynamics Inherent in Contemporary Governing through Code. *Environment and Planning D: Society and Space, 32*(5), 869-885.

[53] A sensor is a machine that observes the environment and converts physical quantity into signals. The physical quantities that can be measured are numerous – including air quality, blockages, temperature, movement, humidity, etc. For a more detailed explanation on sensors, see: 3Bplus. (2018, 4 November). How do smart devices work: sensors, IoT, Big Data and AI. https://3bplus.nl/how-do-smart-devices-work-sensors-iot-big-data-and-ai

creating Sanitation Intelligence."[54] At least one smart solution in the Pune case uses sensor data for building self-cleaning toilets. The GARV toilets,[55] some of which are installed in Pune, have sensors triggering floor and toilet-pan washing mechanisms.[56] However, other smart solutions appear to rely on transferring sensor data to a control centre to enable local municipal authorities that take care of urban planning and resource management – usually called Urban Local Bodies (ULBs) or the PMC in this case – to make data-driven decisions. Regardless of the solution, one must ask what data is being collected, what decisions will be made based on such data, and by whom.

When data from sensors is assimilated in the command and control centre, the centre becomes a one-stop-shop for data acquisition, assimilation and analytics. The end result is a digital map of the entire city.[57] According to the TBC Interim Report on the progress in the Pune smart sanitation project, as of November 2018, "sensors" include footfall sensors to understand user numbers, sensors in treatment plants to understand flow and quality of toilet resources, and sensors within a toilet to detect pathogens and other indicators.[58] Data from these sensors will be used to improve efficiency and management of operations, including managing peak usage times, and disease prevalence in communities.

The TBC Interim Report suggests that smart public toilets will produce a layer of "city intelligence" on "operational status of toilets triggering maintenance and cleaning", which will be used by ULBs to "adjust stockage and deploy maintenance." Likewise, smart public toilets will also produce a layer of "business intelligence" on "scheduling maintenance and customer usage patterns", which will be used by the toilet-operating businesses to improve "consumer communications" as well as by the on-ground operators to "optimize service levels."[59]

This is ostensibly a laudable development for citizen toilet users, especially considering that the TBC's aim in Pune is to develop "new models" in products, services and infrastructure, to "scale up access, usage and behaviour change" in India.[60] A closer look, however, indicates that citizen labourers do not figure anywhere in this scheme – an issue of utmost gravity given the caste and gendered nature of sanitation work in India.

As mentioned, sensor data used as envisaged aids the ULB, the business or on-ground operators to monitor the condition of toilets for either deploying maintenance or optimising service delivery. We worry that this will amount to a round-the-clock surveillance of sanitation workers tasked with the maintenance and cleaning of toilets.[61] The dispersion of governmental power across multiple private, democratically unaccountable actors – from toilet operators to the toilet business in question – for managing and disciplining sanitation workers[62] to provide more efficient maintenance, further multiplies the avenues for surveilling the sanitation workers.[63]

In addition to the worry that sanitation workers will be subjected to a top-down surveillance[64] aimed at more efficient management of the smart toilets, we also worry that their needs will not be captured in the data sets that will be created using

[54] Toilet Board Coalition. (2018). Op. cit.

[55] https://www.garvtoilets.com/about.html

[56] Dietvorst, C. (2016, 6 October). Indestructible and smart: public toilet innovation in India. *IRC Wash Systems.* https://www.ircwash.org/blog/indestructible-and-smart-public-toilet-innovation-india

[57] Ministry of Housing and Urban Affairs. (2018). *ICCC Maturity Assessment Framework Toolkit.* https://smartnet.niua.org/sites/default/files/resources/iccc_maturity_assessment_framework_toolkit_vf211218.pdf.

[58] Toilet Board Coalition. (2018). Op. cit.

[59] Ibid.

[60] Civil Society. (n.d.). Op. cit.

[61] The SBM has come under heavy criticism similarly for attempting to enforce more stringent cleaning obligations on already exploited sanitation labourers from the Dalit communities, by valorising the service they provide without either ensuring alternate occupational opportunities for Dalits or destigmatising sanitation work by attracting an upper-caste work force. See Teltumbde, A. (2014). No Swacch Bharat Without Annihilation of Caste. *Economic and Political Weekly, 49*(45). https://www.epw.in/journal/2014/45/margin-speak/no-swachh-bharat-without-annihilation-caste.html; Gatade, S. (2015). Silencing Caste, Sanitising Oppression: Understanding Swachh Bharat Abhiyan. *Economic and Political Weekly, 50*(44). https://www.epw.in/journal/2015/44/perspectives/silencing-caste-sanitising-oppression.html

[62] Foucault, M. (1978). Op. cit. See also other Foucauldian critiques of private sector technological solutions in smart cities such as Vanolo, A. (2014). Op. cit.; Klauser, F., Paasche, T., & Söderström, O. (2014). Op. cit.

[63] See Levy, K. (2015). The Contexts of Control: Information, Power, and Truck-Driving Work. *The Information Society*, 31(2), 160-174, in which she finds that "[t]ruckers, a spatially dispersed group of workers with a traditionally independent culture and a high degree of autonomy, are increasingly subjected to performance monitoring via fleet management systems that record and transmit fine-grained data about their location and behaviors" and further that "managers make use of electronic monitoring (and the data it generates) to control workers by making their day-to-day practices more visible and measurable." Noopur Raval notes regarding Levy's findings that the worker is "no longer accountable to a single or limited set of entities (human boss, customer) and is now, instead, constantly being monitored" and that such "fine-grained surveillance of truckers' behaviour [...] also causes a fundamental shift in how efficiency is determined and who gets to determine how to structure the organization of daily work in order to maximize overall work done." Raval, N. (2019). Advances in Computing and the Future of Work. (Unpublished draft.) https://www.academia.edu/38459868/TLSDraftPaper_Raval.docx

[64] Zuboff, S. (2019). *The Age of Surveillance Capitalism.* London: Profile Books.

sensors (i.e. data collection from the ground up). Sensors at present are collecting data only on the use of public toilets (including footfall, frequency of usage, air quality, etc.) and the state of sewer water treatment plants (to manage flow, blockages and leaks). Sanitation workers are conspicuously absent from these data sets, and consequently, from the digital map of the smart city. In other words, no data is collected for the potential of alleviating the work load of sanitation workers, minimising sanitary hazards by providing protective equipment and gear, or enhancing the sanitary conditions of their labour.

The digital map of the city is thus selectively built, with dark spots in data sets overlapping almost entirely with marginalised groups. As data sets embed prevalent intersectional caste, class and gender biases,[65] the digital map of Pune's smart city will ignore and thus doubly marginalise already underrepresented groups. Sensors, being one of the primary points of data collection within the smart sanitation system, will feed data directly into the design and development of potential AI systems. AI applications that focus on gaining insights from data will use machine learning – a process of generalising outcomes through examples.[66] This means that 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.[67] The marginalising of sanitation workers no doubt predates the use of technology in sanitation. It is precisely that marginalisation that will be further normalised and solidified through dark spots in data sets.[68] Communities at the margins of data collection are often those who need to be counted most – making the nature of data set creation crucial.

This seems to be a pervasive blind spot across government initiatives. The NUIS, for instance, contemplates empowering field-level employees by providing tools and knowledge for their everyday work, and providing city administrators with data and intelligence. It is not clear who field-level employees are. Given that they are expected to have formalised tasks and deliverables,[69] it appears that they are operators of infrastructure, and not the workers hired by the operators. Consequently, we worry that while improving citizens' ease of living is a purported aim of the NUIS, citizen labourers including sanitation workers and their needs are not contemplated in the NUIS' core data infrastructure.

Conclusions

Smart cities are currently envisaged and framed as a business opportunity. In the process, we have lost sight of the responsibility of the state in smart cities. In studying publicly available documents on the SCM, SBM, Digital India, NUIS and the national strategy for AI as applied to the Pune smart sanitation project, we find that the needs and perspectives of only citizen toilet users and citizen residents are reflected in the development of the smart sanitation project. The design of "smart" solutions, that are pervasive and IoT-based, indicates a dispersion of governmental power across several private, democratically unaccountable actors, to better discipline and manage citizen workers servicing the smart city.

Consequently, while the rights and ease of living of citizens are purportedly a central concern of smart cities, not *all* citizens' needs and perspectives are reflected in designing "smart" solutions. The turn to technological solutions that are designed, developed and deployed by private actors, as a substitute for public services and obligations, benefits the citizen consumer but further marginalises the citizen worker, thus aggravating extant worries on democratic accountability and responsiveness of smart cities in India. The SCM generally also bears testimony to this fact, by permitting area-based development that encompasses less than 10% of the Indian population.[70] Studies have found that lower-income groups are suspicious of smart city initiatives as they view it as a vehicle for gentrification and driving them away from certain areas in cities.[71]

Examining the crucial data stage of Pune's smart sanitation project, we learn that the design of data sets at the bedrock of the Pune Smart City

[65] 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*. https://doi.org/10.1098/rsta.2018.0087

[66] Surden, S. (2014). Machine Learning and the Law. *Washington Law Review, 89*(1).

[67] Lerman, J. (2013). Big Data and Its Exclusions. *Stanford Law Review Online*. https://www.stanfordlawreview.org/online/privacy-and-big-data-big-data-and-its-exclusions

[68] Crawford, K. (2013, 10 May). Think Again: Big Data. *Foreign Policy*. https://www.foreignpolicy.com/articles/2013/05/09/think_again_big_data.

[69] Ministry of Housing and Urban Affairs. (2019). Op. cit.

[70] Housing and Land Rights Network. (2018). *India's Smart Cities Mission: Smart for Whom? Cities for Whom?* New Delhi: Housing and Land Rights Network.

[71] Anand, T., et al. (2018). *Smart City Mission in Pune.* On file with authors.

are also likely to be unrepresentative and embedded with bias. Decisions made at this stage undergird how systems function, define their goals and parameters for success, and influence the ways in which systems can be used. It is important for data collection and creation efforts within the Pune Smart City, therefore, to work on being inclusive and non-discriminatory throughout the process of building "smart" systems, not just at the stage of outcomes.

One overarching learning from studying the disparate impact of the above policies on Dalit women sanitation workers is that the condition of the most marginalised citizen groups ought to be at the centre when designing policies and initiatives. While the potential of AI systems to transform Pune's sanitation crisis is deliberated upon aggressively by a variety of stakeholders, current debates must also focus on where systems are deployed, how systems impact the marginalised, and what axes of power these systems entrench.

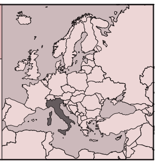
Action steps

The following lessons can be drawn from the smart sanitation project in Pune:

- Centre the perspectives of those persons and communities lying at the intersection of many axes of disadvantage, such as caste, class, gender and ability, prior to and while designing policy, instead of studying social impact after the fact.
- Be critical of AI as a socio-technical and not just technical system, i.e. do not only worry about accuracy, but critique the very existence of systems, their placement, and their beneficiaries.
- Focus on bringing about transparency in government procurement of privately developed technology for public service delivery. This is one of the only ways in which the constant obsession with the "business case" for programmes will be combated.

ITALY

CHALLENGES FOR ACTIVISTS PUSHING FOR AN ETHICAL APPROACH TO AI IN ITALY



Eurovisioni
Giacomo Mazzone
www.eurovisioni.eu

Introduction

Italy is the second biggest manufacturing country after Germany in the European Union (EU), and one of the main university and research centres. Because of this, the use and application of artificial intelligence (AI) in the country has advanced and spread. While this is most notable in robotics applied to the industrial sector, there are a number of experimental projects using AI in the country, including in the media sector. Recently, another area that has been developing fast has to do with building so-called "smart cities", with a vast number of start-ups specialised in making services more accessible to citizens in various fields (health, education, social assistance, unemployment benefits, etc.). There is also strong and lively research activity on the ethical consequences of AI, including decision-making processes assisted by AI, and on related issues such as the regulation and self-regulation of media and social media, especially in the context of elections, hate speech, fake news, and child protection online.

Governments in the country, over the years, have had different approaches to and interest in AI. The centre-left government that was in power from 2013 until the beginning of 2018 was pushing hard (both in the country and in international forums such as the G7) to develop common policies on AI. In 2017, the Ministry of Economic Development launched a special fund (called "Industria 4.0") to finance the introduction of AI and digital innovation in the traditional industrial sectors.[1] The plan was a huge success, and increased investment, especially in the automotive industry, but also in the pharmaceutical and energy sectors.[2]

Outlining some AI initiatives in the media sector in Italy, this report argues that the current political climate in the country and a fragmented civil society are making the ethical regulation of AI very difficult. It nevertheless identifies two potential policy windows for activists to push for a human-rights approach to AI implementation.

Laboratories for innovation in the media sector

In Italy, AI applications in general (and particularly in the media sector) have been mainly developed in areas where there are a lot of repetitive tasks and where using a human for a task adds little value to the job at hand. The two AI-based services of particular social relevance in the media sector have been developed in the areas of increasing accessibility for people with disabilities and the automatic archiving of digital multimedia files.[3]

For example, Radiotelevisione Italiana (RAI, the national public service media), through its research centre based in Turin (CRITS), is working on numerous AI applications aimed at making accessible radio and TV programmes for the elderly and people with disabilities. This work is the result of a new 2018-2027 framework agreement between the government and RAI, which aims to increase the percentage of programmes for people who are hard of hearing and those with reduced vision, and even deaf and blind people.[4]

RAI has developed three prototypes for new services based on AI that are now in the testing phase: Virtual LIS (short for Lingua dei Segni Italiana, Italian Sign Language), a virtual weather forecast using LIS, and Stretch TV.

Virtual LIS is a system that automatically converts voice audio into sign language using a 3D virtual interpreter. The first tests using LIS have been for weather forecasts.

[1] https://www.mise.gov.it/images/stories/documenti/investimenti_impresa_40_eng.pdf

[2] The media and cultural sectors were not allowed to access these incentives, which were reserved for the industrial sector, but nevertheless many research initiatives have been launched since then with private and EU funding.

[3] www.crit.rai.it/CritPortal/wp-content/uploads/2018/11/2018-10-09_TheExperienceOfRaiDigimasterProjectInProgress_sitow.pdf

[4] Obligations are mentioned in the new service's contract with RAI for 2018-2022, published in January 2018. The services to increase accessibility are covered in Article 25 of the contract. www.rai.it/dl/doc/1521036887269_Contratto%202018%20testo%20finale.pdf

Stretch TV is a system that can slow down the speed of TV programmes – both images and audio – by 10% to 20%, so that even the hard-of-hearing can easily follow dialogue. As part of EU-funded projects, RAI has also developed an archiving system for video and film using AI.[5]

There are many of these kinds of activities in the country, often implemented as part of EU research consortia.

One particularly interesting project involving media research is called "Femicide Storytelling", run by the University of Turin (within the PRIN project).[6] Using AI, the project gathers data on cases of femicide and violence against women that are reported in newspapers and on radio and TV. These are then categorised and mapped against existing socio-demographic data in order to create a framework for predicting crimes against women. This is one of the most advanced experiments in big data in the country. While preventing crimes against women, it can be used to raise public awareness, and produce meaningful stories with a social impact.[7]

The most important EU project currently ongoing in the country is SoBigData, of which the Istituto Superiore Sant'Anna, based in Pisa, is one of the major actors and coordinators.[8] This project also has applications in the field of media research, particularly online.

The idea behind the project is very simple: according to the dominant narrative, big data is the oil of future growth. According to the research project, this narrative is true, but is currently being undermined by the misappropriation of personal data by some big companies that use personal data to make profits.

SoBigData proposes to create the "Social Mining and Big Data Ecosystem", a research infrastructure that would allow the use of data in a protected environment, and on a collaborative basis, aiming to create a totally new approach to the use of big data. The aim of the project is to create a research community that could use the research infrastructure as a "secure digital wind-tunnel" for large-scale "social data analysis and simulation experiments." Establishing common ethical principles on the use of algorithms, AI and big data is the pre-condition of making such an ecosystem workable, reliable and sustainable.

The areas of potential applications are enormous. The project has identified some of them: from smart cities to health, from media and understanding the influence of social media, to migration, to sports. For instance, the consortium has produced an analysis of mobility within a metropolitan area and even a map of accessibility for people in wheelchairs for a town, based on the data collected through an app for mobile phones. Another ongoing test involves the correlation between the use of social media and the polarisation of political and societal debates (analysed through algorithms checking and monitoring social media debates). The consortium publishes a newsletter presenting some of its experiments and achievements.[9]

The context for policy advocacy on AI

Italy is a country where the debate about social media and its misuse (hate speech, discrimination, sexual harassment and so on) has been widespread in media and society for many years. The proliferation of local media (radio and TV) in the last decades of the 20th century has raised awareness of media-related issues and led to the establishment of specific jurisprudence, as well as a media-related authority set up in 1987. The Autorità per le Garanzie nelle Comunicazioni (AGCOM, the Authority for the Control of Media) is a multisectoral authority, now with a very broad mandate: it is in charge of regulating printed media, electronic media, telecoms and (by extension) social media.

In an already highly regulated media ecosystem, the arrival of social media was nevertheless a shock, because some of the protections previously ensured by the traditional media regulations proved to be inefficient when it came to the new online forms of communication. As a result, calls to regulate social media came from various sectors (the judiciary, human rights activists, the media, etc.) – calls that the two parties governing the country in a coalition since June 2018 have not supported at all.

On the contrary, the Five Star Movement, led by comedian Beppe Grillo, and the League, led by Matteo Salvini – the two political parties forming the government that collapsed at the end of August 2019 – have based a lot of their success on the use of social media. For example, Salvini, whose Facebook page has more than 3.75 million fans, prefers to talk to his supporters live on Facebook instead of going into Parliament, holding press conferences or being interviewed on TV.[10]

[5] CRITS and RAI Teche presented the archiving system at the FIAT/IFTA World Conference in Venice on 9 October 2018. www.crit.rai.it/CritPortal/?notizia=fiat-ifta-world-conference-2018&lang=en

[6] https://cercauniversita.cineca.it/php5/prin/cerca.php?codice=2015BBLK7M

[7] www.crit.rai.it/CritPortal/?notizia=1553

[8] The project receives funding from the EU Horizon 2020 research and innovation programme. See sobigdata.eu/index for more information; a complete description of the project can be found here: www.sobigdata.eu/sites/default/files/www%202018.pdf

[9] sobigdata.eu/newsletter

[10] According to the last available statistics on 4 August: https://www.socialbakers.com/statistics/facebook/pages/detail/252306033154-matteo-salvini

The Five Star Movement makes most of its decisions using a voting platform called Rousseau,[11] where registered members are called to vote on the selection of candidates for elections, on main legislative proposals and even on internal reform of the party's bylaws.[12]

Consequently, both parties had no great interest to intervene on the matter of the regulation of the internet, because they position themselves as anti-establishment forces, and consider traditional media to be controlled by the main economic and political interest groups in the country. This refusal, however, creates a growing tension between government, civil society, opposition parties and traditional media, but also within civil society, and the same two ruling parties.

This also makes pushing for the ethical regulation of AI difficult in the country. Nevertheless, two potential policy windows remain.

Policy windows to advocate for the ethical regulation of AI

The push for regulating social media

Despite the government's reluctance to regulate social media, civil society pressured AGCOM to approve a proposed regulation against online hate speech on 15 May 2019. The proposed regulation is called the "Regolamento recante disposizioni in materia di rispetto della dignità umana e del principio di non discriminazione e di contrasto all'hate speech" (Regulation containing provisions regarding respect for human dignity and the principle of non-discrimination and combatting hate speech).[13]

In Article 9 of the Regulation, AGCOM requires that internet platforms distributing video content:

- Adopt, publish and respect formal codes of self-regulation in which they commit themselves to remove hate speech content and other content violating human rights online.
- Provide quarterly monitoring reports on content identified, removed and penalised.
- Launch awareness campaigns among their users to promote diversity and to fight any kind of online discrimination.

The regulation was submitted for public consultation before it was adopted. Some of the consulted entities have argued that the media and telecom authority has no mandate to regulate (or even co-regulate) social media and the internet. AGCOM has submitted the proposed regulation officially to the Italian Parliament, which will now have to decide whether to enshrine the regulation as law and formally confirm AGCOM's de facto mandate on social media.[14]

This proposal potentially opens up the door for further public discussion on ethical principles with respect to the use and implementation of AI online.

The multistakeholder roundtable

Another potential avenue where activists could push for the regulation of AI is in a multistakeholder roundtable set up by AGCOM, which contributed to the development of the regulation approved in May. In November 2017, AGCOM created the roundtable in an attempt to guarantee pluralism and to ensure reliable news on digital platforms.[15] This informal, voluntary group has grown over the years, and today gathers around 30 companies, including Google, Facebook and Wikipedia (Twitter has not joined the roundtable yet), institutions and civil society organisations. The job of the roundtable is to come up with solutions to problems that are raised, based on in-depth research and knowledge of the Italian media market.[16]

Since January 2019, it has produced a regular newsletter (called the Observatory on Online Disinformation) and all relevant communication between the authority and the parties involved, such as an internet platform, are discussed publicly and published on its website. For instance, in March 2018 the secretariat of the roundtable addressed a detailed request for explanations from Facebook on the impact of the Cambridge Analytica case in Italy. Another area of concern is the impact of fake news and hate speech on the online world.

[11] https://rousseau.movimento5stelle.it/sso_home.php

[12] The platform has around 150,000 registered participants, but data is not certified. The Italian authority on privacy has recently (4 April 2019) criticised the platform for not being protected against piracy and third-party manipulation. See: https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9101974

[13] https://www.agcom.it/documentazione/documento?p_p_auth=fLw7zRht&p_p_id=101_INSTANCE_ls3TZlzsKohm&p_p_lifecycle=0&p_p_col_id=column-1&p_p_col_count=1&_101_INSTANCE_ls3TZlzsKohm_struts_action=%2Fasset_publisher%2Fview_content&_101_INSTANCE_ls3TZlzsKohm_assetEntryId=15055471&_101_INSTANCE_ls3TZlzsKohm_type=document

[14] In particular, Article 1.31 of the law that instituted AGCOM (Legge 31 luglio 1997, n. 249 – Istituzione dell'Autorità per le garanzie nelle comunicazioni e norme sui sistemi delle telecomunicazioni e radiotelevisivo) grants it the authority to impose penalties on those who do not comply with its regulations, ranging from fines to revoking their licence to operate in Italy. https://temi.camera.it/leg18/post/OCD15-54606/approvato-dall-agcom-regolamento-contrasto-alle-espressioni-d-odio-hate-speech.html

[15] AGCOM. (2017). Tavolo tecnico per la garanzia del pluralismo e della correttezza dell'informazione sulle piattaforme digitali. https://www.agcom.it/documents/10179/8860814/Comunicato+stampa+16-11-2017/7977d222-b9bd-4763-b1e6-9ae7c960a0ee?version=1.0

[16] https://www.agcom.it/gli-studi-e-le-indagini-conoscitive

During the campaigns for the Italian national elections in March 2018 and for European elections in May 2019, the roundtable agreed on a special strategy to counter possible cyber interference,[17] while some guidelines to guarantee equal access for all candidates to the online platforms were issued.[18]

Conclusion

In the current political turmoil through which Italy is passing, civil society organisations are in a trap. On the one hand they are very involved with institutions in the debate about AI and algorithms, on their implications for society (as seen in some of the projects discussed here) and in the development of regulations. On the other hand, they are paralysed in their advocacy because criticism of how internet platforms operate is considered as an attack on both of the ruling parties (the Five Star Movement and the League), both of which use the internet as a bedrock for their campaigns, including in their attacks on traditional media.

Civil society in Italy is also currently fragmented, with different groups concentrating on their own problems: political parties are worried about trolls and elections; feminist movements are worried about femicide; the child protection movement is worried about the risks of social media for children; minorities and refugee organisations are worried about hate speech and its consequences for their constituencies.

Nevertheless, two avenues for pushing for regulations on AI remain: the current advocacy for the regulation of hate speech online, and the multi-stakeholder roundtable set up by AGCOM.

A further policy window has presented itself with the sudden and unexpected collapse in August 2019 of the populist and right-wing parties that make up the current government. This has ushered in a change of alliances (now populist and left-wing parties), and opened up new perspectives for civil society activism, and for the Parliament. However, what exactly these advocacy opportunities will be is still too early to predict.

Action steps

If the fragmented civil society movement is able to free itself from its current paralysis, then a lot could be achieved with respect to AI. They need to collectively start to push for ethical principles to be incorporated into AI, such as privacy by design in the algorithms that are developed. This could be a successful way to unify movements that at the moment are not cooperating and are working separately, focused on their own battles. In this sense, the Italian tradition of fighting for civil liberties – which resulted in a very detailed and in-depth juridical framework and a modern constitution – could contribute significantly to the European and global debate on AI.

[17] AGCOM. (2019). Osservatorio sulla disinformazione online n. 4/2019. https://www.agcom.it/visualizza-documento/721fe550-fcb0-4a05-b3d9-30012a98c616

[18] AGCOM. (2018). Linee guida per la parità di accesso alle piattaforme online durante la campagna elettorale per le elezioni politiche 2018. https://www.agcom.it/documentazione/documento?p_p_auth=fLw7zRht&p_p_id=101_INSTANCE_2fsZcpGr12AO&p_p_lifecycle=0&p_p_col_id=column-1&p_p_col_count=1&_101_INSTANCE_2fsZcpGr12AO_struts_action=%2Fasset_publisher%2Fview_content&_101_INSTANCE_2fsZcpGr12AO_assetEntryId=9478309&_101_INSTANCE_2fsZcpGr12AO_type=document

JAMAICA

ARTIFICIAL INTELLIGENCE AND "CONSENT OF THE GOVERNED": PITFALLS IN DEVELOPING JAMAICA'S DIGITAL ID SYSTEM

Mona ICT Policy Centre (MICT), The University of the West Indies
Hopeton S. Dunn
https://conf.carimac.com/index.php/cybersecurity/2019

Introduction

In a recent landmark ruling, Jamaica's Chief Justice Brian Sykes observed that the government's harsh decision to impose criminal sanctions to enforce compulsory registration by all citizens in a new digital ID system was a remarkable choice "in a democracy where the exercise of executive power rests upon the consent of the governed."[1]

This report discusses the government's planned use of artificial intelligence (AI) to create a new national identification system (NIDS) as a unique verifier of every Jamaican citizen. The enabling legislation, called the National Identification and Registration Act (NIRA), was approved by Parliament in 2018 under the leadership of the Andrew Holness-led government. It was presented as a means of modernising and integrating a clutch of existing national ID data sources, including census data, tax registration metrics and electoral roll data. Loan funding to the tune of USD 68 million was being provided by the Inter-American Development Bank (IDB) to acquire the supporting AI technology, and to roll out information campaigns and other implementation services related to the nationwide capture of biometric data for machine classification, analysis and storage.

While the overall plan for a national ID system was widely deemed as an important advance for the country's development, there were some elements of the plan that generated deep public concern and became the basis of a legal challenge by the parliamentary opposition. These included the highly intrusive level of biometric data being demanded, the compulsory nature of the plan, criminal sanctions for non-compliance, and the absence of adequate safeguards for data protection. The scenario that emerged by 2019 was one in which AI was being used to undermine the privacy rights, personal choices and constitutional freedoms of an entire population.

In April 2019, the Jamaican Supreme Court largely agreed with the expressed concerns and handed down a historic ruling, designating the new NIRA law "null, void and of no effect."[2]

Legislative context

Jamaica is a parliamentary democracy that gained political independence from Britain some 57 years ago, in 1962. Since then, two major political parties, the currently ruling Jamaica Labour Party (JLP) and the opposition People's National Party (PNP), have alternated power, both claiming strong affinity to democracy and the rule of law. The apex of the judicial system is still the United Kingdom Privy Council, but the country's Supreme Court, headed by an independent chief justice, serves as the top court of original jurisdiction. The Supreme Court includes a constitutional division.

Jamaica's independence constitution was amended with bi-partisan support in April 2011 to include a Charter of Fundamental Rights and Freedoms, which established or strengthened a range of key citizen protections. Among other provisions, the Charter specifies that "Parliament shall pass no law and no organ of the state shall take any action which abrogates, abridges or infringes the guaranteed rights." These rights include the "right to equitable and humane treatment by any public authority in the exercise of any function"; the "right to protection from search of the person and property" without a warrant; "respect for the protection of private and family life, and privacy of the home"; and "protection of privacy of other property and of communication".[3]

There are as yet no legal provisions in Jamaican law conferring specific protections against the ill effects or misuse of AI systems. However, a March 2010 Cybercrimes Act addresses computer-specific offences, such as unauthorised access to computers

[1] Chief Justice Sykes, Para 23 of the Supreme Court Ruling on the National Identification and Registration Act. supremecourt.gov.jm/content/robinson-julian-v-attorney-general-jamaica

[2] Supreme Court of Jamaica. (2019). Ruling of Full Court in Claim Number 2018HCV01788 between J. Robinson, Claimant and the Attorney General of Jamaica, Defendant. supremecourt.gov.jm/content/robinson-julian-v-attorney-general-jamaica

[3] https://japarliament.gov.jm/attachments/341_The%20Charter%20of%20Fundamental%20Rights%20and%20Freedoms%20(Constitutional%20Amendment)%20Act,%202011.pdf

and illegal data alteration. Amendments to the Act in 2015 offered additional protections as well as stiffer penalties for cybercrime offences.[4] However, a key companion piece of legislation, the Data Protection Act,[5] though in draft form since 2017, has not yet been debated and approved by Parliament. In this regard, the start of the collection of people's biometric data under provisions of the NIRA in 2018 was deemed by some observers to be premature and troubling.

NIRA, biometrics and the court

The NIRA was approved in the Jamaican Parliament in December 2018 over the strong objection of sections of civil society and the parliamentary opposition. The law made formal registration by all citizens compulsory and required them to provide specific biometric information on pain of criminal sanctions. A central registering authority, created by the Act, was mandated to collect identity verifiers, including iris scans and fingerprints and using facial recognition technologies. However, in a departure which many citizens and the parliamentary opposition deemed unwarranted and extreme, the Act also required the capture of vein patterns, and if needed, footprints, toe prints, palm prints and the blood type of citizens and residents. The compilation and analysis of these biometrics were to be executed using big data analytics and pattern recognition technologies.

The new law specified that "(e)very person who refuses or fails, without reasonable excuse, to apply to the Authority for enrolment in the database... commits an offence and shall be liable on conviction to the penalty specified." The government refused to accede to demands made by civil society and the parliamentary opposition for changes, or even to extend the public and parliamentary debate time before final approval. As a result, the NIRA law was referred by the opposition to the Supreme Court for a ruling on the constitutional validity of certain key sections.

The court's ruling was delivered on 12 April 2019 in a 309-page judgment, from a panel of three judges, led by Chief Justice Sykes. In his written judgement, the chief justice paid particular attention to the compulsory nature of the law and its recourse to criminal sanctions for non-compliance:

> Here we see the ultimate coercive power of the state being enlisted to ensure compliance – the risk of imprisonment even if the risk is reduced. The learned Attorney General contended that when you have a system of compulsory registration then there has to be a means of enforcement that may be an effective method of ensuring compliance. The policy choice, it was said, was to use the criminal law. This response by the learned Attorney General suggests that persuasion was not thought to be a reasonable option, a seemingly remarkable conclusion in a democracy where the exercise of executive power rests upon the consent of the governed.[6]

At the end of the detailed written ruling, the judges of the Supreme Court announced that the legislation violated numerous sections of the Charter of Fundamental Rights and Freedoms of the Jamaican constitution. It found that data collection methods and the protocols of intended data use did not sufficiently guarantee respect for and protection of privacy, and that there were insufficient safeguards against the misuse and abuse of the data to be collected.

So riddled was the legislation with what the court deemed unconstitutional clauses, that the judges said they were obliged to disallow the entire NIRA law. Accordingly, the court ruled unanimously that the entire NIRA was "null, void and of no effect."[7]

In the event, Jamaica's first attempt to use AI on an extensive basis for public data gathering and analysis was deemed unconstitutional on the grounds of inadequate attention to the civic, personal, legal and social implications of the Act. According to one commentator, the government had "promoted the transactional value of the technology, rather than the fundamental value of the principles for which it was adopted."[8] In an editorial, the Gleaner newspaper also remarked:

> The Supreme Court's comprehensive slap-down of the government's national identification law has implications beyond the need of the Holness administration to reflect deeply on its future approach to the formulation of laws. For it raises questions, too, about our commitment to the Constitution.

The newspaper reminded readers that part of the haste in passing this legislation was related to "the need to meet the Inter-American Development Bank's funding cycle for a US$68 million loan for the project."[9]

[4] https://www.japarliament.gov.jm/attachments/339_The%20Cybercrimes%20Acts,%202015.pdf

[5] https://www.japarliament.gov.jm/attachments/article/339/The%20Data%20Protection%20Act,%202017----.pdf

[6] Supreme Court of Jamaica. (2019). Op. cit.

[7] Ibid.

[8] Morris, G. (2019, 14 April). Jamaica's NIDS setback of its own making. *Jamaica Observer*. www.jamaicaobserver.com/the-agenda/jamaica-s-nids-setback-of-its-own-making_162161?profile=1096

[9] The Gleaner. (2019, 16 April). Editorial – NIDS Ruling Breaks New Ground. *The Gleaner*. jamaica-gleaner.com/article/commentary/20190416/editorial-nids-ruling-breaks-new-ground

A large part of those loan funds would have been used to acquire the AI technology that would have been embedded in what the law called the National Civil and Identification Database (NCID).

Identity, AI and cyber risks

The proposed new national database in Jamaica was to be a prime site for big data analytics in an emerging global technology environment. According to the Harvard Business Review, the technologies that enable AI, like development platforms and vast processing power and data storage, are advancing rapidly and becoming increasingly affordable. Yet it warns that the successful deployment of AI requires a deliberate policy of "rewiring" the organisation involved in its utilisation. AI initiatives, it argues, face formidable cultural and organisational barriers if not carefully deployed.[10]

In a similar vein, digital security analyst Sarah Vonnegut says special measures are often needed to safeguard AI-generated and other digital databases. She argues that databases that are important to companies and government organisations are very attractive to hackers and can be vulnerable to numerous forms of attack. One vulnerability is the so-called "buffer overflow" vulnerability, when a programme "tries to copy too much data in a memory buffer, causing the buffer to 'overflow' and overwriting the data currently in memory." Vonnegut says buffer overflow vulnerabilities "pose an especially dangerous threat to databases holding particularly sensitive info, as it could allow an attacker exploiting the vulnerability to set unknown values to known values or mess with the program's logic."[11]

Dana Neustadter, from the internet content design company Synopsys, says secure algorithms are a large part of the value of any AI technology:

> In many cases, the large training data sets that come from public surveillance, face recognition and fingerprint biometrics, financial, and medical applications, are private and often contain personally identifiable information. Attackers, whether organized crime groups or business competitors, can take advantage of this information for economic reasons or other rewards. In addition, the AI systems face the risk of rogue data injection maliciously sent to disrupt neural network's functionality (e.g., misclassification of face recognition images to allow attackers to escape detection). Companies that protect training algorithms and user data will be differentiated in their fields from companies that suffer from the negative PR and financial risks of being exploited.[12]

In the absence of adequate data security safeguards, it is clear that the extremely sensitive biometric data that were to be collected to create the NIDS in Jamaica would have been especially vulnerable to these and other forms of malicious attack, without legal recourse to a modern data protection act.

While recognising the importance of applying AI and other data-related technologies to create a reliable national database, the merit of the Jamaican Supreme Court ruling is its insistence that the process not threaten citizen rights, freedoms and privacy and that better safeguards be introduced to mitigate the risks to citizen data.

Conclusion

Jamaica now faces the challenge of how to reform and re-establish a national identification system that is within the bounds of the constitution. The new successor legislation has to ensure respect for citizens' rights to make informed choices about the data being collected and held. While AI will doubtless aid any renewed data gathering effort, care will be needed to ensure robust data protection, secure and reliable data storage and overall data integrity on the principles laid down, for example, by the 2018 General Data Protection Regulation (GDPR) of the European Union.[13] Jamaica's own long-pending Data Protection Act will need to be debated and enacted as a matter of priority, and to precede any data collection under any revised ID law.

Against the background of the Supreme Court ruling, the compulsory provisions and criminal sanctions will have to be removed in favour of greater stakeholder consultation, public education and what the Chief Justice calls citizen "persuasion". The intrusive nature and unwarranted details required in the biometric data being sought from citizens will also have to be reviewed to provide for citizen consent and the provision of advance justification on the part of government for the collection of each type of sensitive biometric data.

[10] Fountaine, T., McCarthy, B., & Saleh, T. (2019). Building the AI-Powered Organization. *Harvard Business Review, July-August*. https://hbr.org/2019/07/building-the-ai-powered-organization

[11] Vonnegut, S. (2016, 24 June). The Importance of Database Security and Integrity. *Checkmarx*. https://www.checkmarx.com/2016/06/24/20160624the-importance-of-database-security-and-integrity

[12] Neustadter, D. (n/d). Why AI Needs Security. *Synopsys*. https://www.synopsys.com/designware-ip/technical-bulletin/why-ai-needs-security-dwtb-q318.html

[13] https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en

Action steps

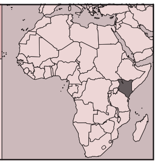
The following steps are necessary in Jamaica:

- It must be recognised that establishing a national ID system is not the sole responsibility of the government; it is necessary that civic, academic, human rights and corporate stakeholders become more involved in hosting public forums on AI, human rights and national ID systems.
- The government itself should re-commit to a thorough legal and policy review in line with the requirements imposed by the Supreme Court ruling. Any new national ID legislation must benefit from extensive public consultations and wider parliamentary deliberations.
- A significant proportion of the loan funds committed to this project by the IDB should be devoted to public education, citizen consultations and to ensure data protection and integrity.
- International case studies on the establishment of successful national ID systems should be produced by the relevant government agencies and used to inform a process of public education towards a new AI-assisted national ID system for Jamaica.
- Finally, the outcomes and recommendations of the 7th National Cyber Security Conference, which was hosted by the Mona ICT Policy Centre at the University of the West Indies in June 2019, need to be made more widely available to government and all other stakeholders.[14]

[14] https://conf.carimac.com/index.php/cybersecurity/2019

KENYA

MOBILE MONEY LENDING AND AI: THE RIGHT NOT TO BE SUBJECT TO AUTOMATED DECISION MAKING



NBO Legal Hackers
Francis Monyango
www.monyango.com, francmonyango@gmail.com

Introduction

The success of mobile money in Kenya changed lives and has created a good base for many mobile technology-based solutions. Among these solutions are mobile loan services which ride on the mobile system but also involve the use of artificial intelligence (AI).

For the purposes of this report, AI is defined broadly as computer systems designed to perform tasks in a way that is considered to be intelligent, including those that "learn" through the application of algorithms to large amounts of data.[1] Because of this, algorithmic decision making will be considered use of AI. Automated decision making is a decision made by automated means without any human involvement. Examples of this include an online decision to award a loan and a recruitment aptitude test which uses pre-programmed algorithms and criteria.[2]

In the mobile lending process, there are two instances where algorithms sift through data before making decisions. This is in the credit scoring stage and credit referencing stage. The decisions these algorithms make affect the lives of millions of Kenyans every day. In this report, I will highlight the rights enshrined in the Kenyan Bill of Rights that are being affected by the use of AI in mobile lending apps.

The rise of mobile loans in Kenya

Several factors led to the rise of mobile loans in Kenya. The first one is the success of M-PESA,[3] a mobile money service by telecommunications company Safaricom. M-PESA was launched on 7 March 2007 and by November that year it had one million active users. The first purely mobile loan product in Kenya was M-Shwari, which was launched in January 2013 in collaboration with the Commercial Bank of Africa.[4] Since then other banks and Silicon Valley-sponsored companies like Branch[5] have joined the market as service providers, all trying to replicate Safaricom's M-Shwari success in mobile lending.[6]

The second factor is a change in banking policy. In 2016, the Kenya Banking Act was amended to cap interest rates at four percentage points above the central bank rate. This policy change led many banks to reconsider their lending business model as it was now less profitable, and they became more cautious about lending to individuals and small businesses. Banks then got into financial technology by launching mobile money lending apps while others formed partnerships with existing telecommunications companies in the mobile money business.[7]

These initiatives gave the banks new customers and exposed more Kenyans to accessible credit facilities. While this bridged the financial inclusion gap, it also exposed many Kenyans to systems where the information in their phones was used to credit score them.

The mobile lending process

A borrower first downloads a mobile lending application onto his or her mobile phone. For those who want to get loans from institutions allied to their network service providers, the service is accessible

[1] https://www.giswatch.org/giswatch-2019-call-proposals-artificial-intelligence-human-rights-social-justice-and-development

[2] https://www.ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/rights-related-to-automated-decision-making-including-profiling

[3] https://www.safaricom.co.ke/personal/m-pesa

[4] FSD Africa. (2016). *The Growth of M-Shwari in Kenya – A Market Development Story*. https://s3-eu-central-1.amazonaws.com/fsd-circle/wp-content/uploads/2016/11/26122759/M-Shwari_Briefing-final_digital.pdf-9.pdf

[5] https://www.branch.co.ke

[6] Sunday, F., & Kamau, M. (2019, 25 June). Mobile loans: The new gold rush minting billions from the poor. *Standard Digital*. https://www.standardmedia.co.ke/article/2001331308/mobile-loans-the-new-gold-rush-minting-billions-from-the-poor

[7] The general result was that many banks closed their branches and created an agency model similar to the M-PESA model where you could deposit and withdraw money from your account from an agent who doubled up as a grocery seller (M-PESA agents are usually shopkeepers who sell other things as well). Banks started teaming up with telcos which had the reach they wanted and it was cheaper to use their mobile money infrastructure compared to running branches. With this system in place, banks could introduce new credit products with low operating costs.

from their SIM card menu. These USSD-based[8] services are usually available to users who do not own smartphones. For those who download the mobile applications onto their smartphones, the application will install and ask for permission to access their contacts, call logs, messages and in some cases, their social media accounts.

All this is done to enable the loan service providers to get data to enable them to create a profile of the borrower. These mobile lenders end up having access to a large volume of consumer data whose deployment is unregulated, and consumers are becoming concerned.[9]

After downloading the application, one usually feeds in personal identifiers such as your name, phone number and national identification number before the system "calculates" how you much can get as a loan. The money will then be deposited in your mobile money account, from which you can withdraw or use it anywhere, any time.

When one fails to repay a mobile loan, the service providers usually try to reach out to you to remind you to pay up. But not all do that. Some have automated systems that send out warning messages before sending the customer's name to a credit reference bureau (CRB) as a loan defaulter, hence affecting their credit score.

The General Data Protection Regulation (GDPR)[10] came into force in May 2018 in the European Union. While it is foreign to Africa, it does have an impact on the continent as it applies to entities that process and hold the personal data of data subjects residing in the European Union, regardless of the company's location. This includes any entity on the African continent that conducts business with European companies or deals with EU data subjects.

Those who do not comply with the regulations face legal fees or fines and these consequences do not just apply to businesses within the EU. Many countries like Kenya are trying to comply by enacting data protection laws that are modelled after the GDPR, which is deemed to be the global regulatory "gold standard" for the protection of personal data of consumers.

Article 22 of the GDPR restricts service providers from making solely automated decisions that have a legal or similarly significant effect on individuals. Applied to the Kenyan context of mobile loans, the algorithmic decision-making processes in the lending cycle, from the moment of credit scoring to after the possible default of payment by a borrower, including the listing of defaulters at CRBs, should have human intervention. This is especially the case in the latter part of the process, which affects a borrower's ability to borrow again.

Relevant constitutional rights in Kenya

Consumer protection rights

Article 46 of the Constitution of Kenya[11] provides for consumer protection rights. These rights require businesses to provide consumers with enough information that will enable them to protect their economic interests. These rights back up all the rights listed below and are fleshed out by the Consumer Protection Act, 2012.[12]

Right to access to information

This right is enshrined under Article 33 of the Constitution. This right may be used to defend the right to informed consent. The approach in Kenya since the enactment of the access to information law is that citizens have a right to information that affects them. This can be manifested via an entity providing the information and publishing it on its website or via individuals requesting specific information that affects them and their rights. In the case of mobile loan business entities, they are doing a poor job in consumer education and this has led to many complaints, which shows how uninformed their customers are.

Right to privacy

This right is found in Article 31 of the Constitution. Privacy and data protection issues arise in the mobile lending process because of the automated decision making which affects the borrowers. Many Kenyan borrowers have been listed as defaulters at CRBs for failing to pay KES 200 (USD 2) or less on time.

According to an industry insider who did not wish to be named, different companies have different processes. Some companies have purely automated processes, while others have instances

[8] Unstructured supplementary service data (USSD) is a global system for mobile (GSM) communication technology that is used to send text between a mobile phone and an application program in the network. Applications may include prepaid roaming or mobile chatting. https://searchnetworking.techtarget.com/definition/USSD

[9] Mwaniki, C. (2019, 24 June). Mobile loan borrowers shun betting. *Business Daily*. https://www.businessdailyafrica.com/datahub/Mobile-loan-borrowers-shun-betting/3815418-5169438-qcivc3/index.html

[10] https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

[11] www.kenyalaw.org:8181/exist/kenyalex/actview.xql?actid=Const2010

[12] www.kenyalaw.org:8181/exist/kenyalex/actview.xql?actid=No.%2046%20of%202012

where human beings intervene before a decision is made. These listings drastically affect the credit score of those listed.

The other issue that arises is the lack of informed consent. Many users of these applications do not really know what they are getting into.[13] Some mobile loan service providers inform third parties in their user contacts lists about cash owed especially when they are late with payment. This is after charging fees on the digital loans that range from 6% to 10% for a one-month loan (6% to 10% times 12 months – do the math). Despite the hue and cry, many app owners say that users should read the terms and conditions, which implies that many users of these apps do not know what they are getting into.

Right to dignity

This right is found in Article 28 of the Constitution of Kenya. When you download a mobile lending application onto your phone, the applications usually get access to your saved contacts. It is these contacts that some Kenyan mobile lenders have developed a habit of calling when they need to pressurise their customers to pay back their loans.

This debt shaming is embarrassing and against the right to dignity. Many whose close relatives and friends have been contacted by mobile lending firms to compel them to pay have found the experience to be embarrassing, while the company doing this seems to be unapologetic.

Conclusion

One can see that when it comes to mobile lending in Kenya, there is an algorithmic decision-making process that takes place without any human intervention. These decisions affect the lives of many people and households, and can be said to contravene several rights enshrined in the country's constitution.

People agree to these automated processes without full disclosure from the mobile lending companies on what they will do with their data and how it will be processed. The decisions on the amount they are eligible for and whether their names should be listed by CRBs as loan defaulters have been left to algorithms.

While the Central Bank of Kenya has expressed concerns about how digital lenders are operating and has promised to crack the whip, policy reform is needed to ensure that the credit reference listing of defaulters is not automated and that there is human intervention. Their attempts at cracking the whip at digital lenders have not been successful because most digital lenders are not regulated by the Central Bank, hence they are unaffected by legislative reforms pushed by the regulator. Only an amendment to the Banking Act to include digital lenders will force them to comply with the set laws.

The Kenyan National Assembly has tabled a Data Protection Bill.[14] The draft law contains provisions that dictate that automated decision making cannot happen without human intervention. The exemptions to this rule include when it is consensual, necessary for performance of a contract,[15] and authorised by law.

An interesting clause under this provision requires a data controller to notify a person within a reasonable period of time that a decision has been made that may produce legal effects. The person may then request the data processor to reconsider the decision or to make a new decision not based on the automated processing. This provision is a life buoy to those who have had their credit score messed up by over-eager digital lenders. If passed, the digital lenders and CRBs will be forced to contact the person exhaustively before listing them for defaulting on their loan repayments. It will also deal with algorithmic decision making in the financial sector to complement existing financial sector laws.

Kenya's neighbour, Uganda, recently passed a Data Protection and Privacy Act[16] which contains provisions on rights in relation to automated decision making. While the Act is similar to the Kenyan draft, Uganda has gone a step further to state that a data processor or controller should notify the person within 21 days, a move that removes ambiguity from the law.

The Constitution of Kenya allows for public participation during the development of legislation. These points on automated decision making need to be brought to the attention of legislators during this phase of consultation.

[13] Ondieki, E. (2019, 19 May). Outcry as mobile lenders use 'cruel' tactics to recover loans. *Daily Nation*. https://www.nation.co.ke/news/Outcry-as-mobile-lenders-use--cruel--tactics-/1056-5121620-s2dh87z/index.html

[14] www.parliament.go.ke/sites/default/files/2019-07/The%20Data%20Protection%20Bill%2C%202019.pdf

[15] "Necessary for the performance of a contract" in this context is where automated decision making is used for a process like credit scoring which is key to determining how much a borrower should get. This is usually stated in the contracts.

[16] www.ulii.org/ug/legislation/act/2019/1

Action steps

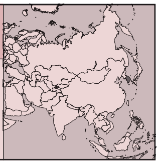
Civil society organisations need to:

- Engage legislators on the human rights implications of mobile lending apps in Kenya. Push for sector-specific laws that govern and limit the application of automated decision making in mobile lending processes.
- Create awareness-raising material to build financial literacy and to support the consumer's right to information. This could be done in collaboration with service providers and the regulators. The regulators, led by the Central Bank of Kenya Governor, have shown interest in enforcing consumer protection in the sector.[17] The service providers are aware of the chaos in the industry and some have expressed a willingness to engage in consumer protection efforts, which gives civil society a good chance to collaborate in ensuring consumers are informed.
- Call on the relevant government bodies and relevant regulators to hold players in this industry to account in processes that deploy automated decision making so that they do not violate consumer rights. Simply creating legislation is not enough – the implementation of laws needs to be proactive.

[17] Leting, T, (2019, 1 April). CBK raises alert on Mobile lending. *East African Business Times*. https://www.eabusinesstimes.com/cbk-raises-concern-on-mobile-lending

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 Data[3]

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/FileDown.do?atchFileId=FILE_000000000827254&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 forum[7] – 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 Korea[10] 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&pagesize=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 pseudnonymised 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.

LATIN AMERICA

VICTOR FRANKENSTEIN'S RESPONSIBILITY? DETERMINING AI LEGAL LIABILITY IN LATIN AMERICA

Franco Giandana (with David Morar)
Creative Commons Argentina; Universidad Nacional de Córdoba
www.unc.edu.ar

Introduction

The story of Victor Frankenstein, the scientist who lost control of his creation, is a great starting point to ask: how do we make artificial intelligence (AI) developers responsible for the software they create, and for any subsequent potential harm it causes? Was Frankenstein guilty for the harm his creation caused? What happens when it is impossible to control the AI systems that are being implemented? Do we need to revise our frameworks? Is civic responsibility in AI objective or subjective? How do we determine legal causation?

These questions have become particularly relevant in Latin America, where a new AI system developed by the Artificial Intelligence Laboratory of the University of Buenos Aires has been implemented in the judicial system in Buenos Aires, the Constitutional Court in Colombia, and at the Inter-American Court of Human Rights in Costa Rica.

The system, called Prometea,[1] is helping authorities resolve "simple" cases in different fields, for example, cases related to the right to housing, to individuals in vulnerable conditions or with disabilities, involving labour rights or children and adolescents, to do with road safety, or price control for public contracts. Even criminal cases are dealt with by the system.

While civil society has raised concerns over how this project will guarantee due process, many other cities in the region have already expressed interest in deploying similar narrow AI systems.[2]

In the context of the Prometea system, this report broadens the discussion to outline some of the legal considerations that the courts face when dealing with AI-related harm.

The rule of law or the rule of AI?

Equality before the law is not a reality in Latin America. There are still strong, concentrated elites that hold the power, both in public office and in the private sector. AI could, potentially, help in closing the space between those in power and the populace who seek, through public institutions, the fulfilment of their rights.

Prometea is operated using voice recognition, assisting with different tasks, from providing possible solutions to cases to managing procedures and processes. The operator speaks to the system to give it instructions, which the software processes at a very high speed, reducing, as the Inter-American Court of Human Rights has shown, a workload of three days to two minutes.

Prometea uses a machine-learning prediction system. It first analyses thousands of documents that have been organised into categories in order to help the system "learn" from them. From there any new file that is entered into the system can be evaluated by Prometea, which determines where strong case history exists, offering a "ruling" on the new case, which is then approved by a judge. The system has a 96% "success rate" in that its rulings are accepted nearly all the time. As stated by the Prometea developers in an article published on the Bloomberg Businessweek website: "Prometea is being used for stuff like taxi license disputes, not murder trials, but it's a significant automation of the city's justice system."[3] From 151 judgments signed using Prometea at the Attorney General's Office in Buenos Aires, 97 found the solution solely by using this system, while 54 cases only used it as a virtual assistant to automate tasks related to the legal procedures involved.

But equality before the law is not a mathematical expression. It is a more complex, emotional and holistic perspective for judging everyone according to the same behavioural standards (i.e. the law), allowing them to be able to access the justice system, and to seek recourse when the law is infringed.

[1] See also the country report from Colombia in this edition of GISWatch.

[2] "Narrow AI refers to AI which is able to handle just one particular task. A spam filtering tool, or a recommended playlist from Spotify, or even a self-driving car – all of which are sophisticated uses of technology – can only be defined via the term 'narrow AI'." Trask. (2018, 2 June). General vs Narrow AI. *Hackernoon*. https://hackernoon.com/general-vs-narrow-ai-3d0d02ef3e28

[3] Gillespie, P. (2018, 26 October). This AI Startup Generates Legal Papers Without Lawyers, and Suggests a Ruling. *Bloomberg Businessweek*. https://www.bloomberg.com/news/articles/2018-10-26/this-ai-startup-generates-legal-papers-without-lawyers-and-suggests-a-ruling

Law – that social instrument we humans have built to ensure that people can co-exist in harmony – must always contain visible, human-embedded constructions of the world. This is a core element of the rule of law; it makes it possible, and it allows legal liability and responsibility to be enforced by the state. In this way, the institutions of justice have the legitimacy they need to be trusted and valued by our societies.

As is the case, though, with anything laced with humanity, flaws abound. And they do so in legislative processes as well, at every step of the way during due process. Because of this, one can easily argue that justice, inherently flawed, is an evolving concept that needs to be strictly monitored and improved. This comes with the important caveat that these improvements must not be blind and simply for the sake of making "improvements".

However, a deeper look at AI would show that not only are we removing the visible human perspective from the equation, we are further embedding an invisible, unattributed and unspoken human element. The AI creations that would compute cases and carry out analysis are not simply ones and zeros. They are subject to the thoughts, whims and biases of their creators. The use of AI therefore doubly weakens the position of the judge, and as a consequence, the rule of law.

How sufficient is the current legal framework to deal with AI?

Justice has always been a value decisive to any human organisation, and there have been different ways to deliver justice throughout history. The concept of justice is somewhat flexible and evolving, so bringing in modifications to deliver a better quality of justice is not something unimaginable. However, these innovations deserve a detailed analysis in order to protect basic human rights. But before we analyse how implementing AI to assist judicial institutions might impact our concept of the rule of law, let us briefly analyse how prepared our legal national frameworks are for facing the challenge of AI technologies in general.

Legal liability for damages suffered due to AI systems is clearly a challenge to traditional law, especially in regions where judges and policy makers are lacking the sufficient knowledge to comprehend the potential and relevance of AI. Nevertheless, before we ask for new regulations and better legislation, we must first review the actual state of our legal frameworks and how resilient they are to the implementation of new technologies.

Weaknesses in data protection laws

Data protection laws already frame what rights must be protected when AI systems collect and process data. In countries like Brazil and Argentina, general data protection laws state a clear difference between personal data and sensitive personal data, determining boundaries and limitations for the collection and usage for every kind of data. But even when there are clear standards to determine whether certain data is to be considered sensitive or not, AI is capable of inferring and generating sensitive information from non-sensitive data. At the same time, the law specifies the importance of the data to be collected complying with the principle of consent. However, there are broad exceptions to the need for consent for certain objectives, for example, for national security purposes (e.g. article 4.III.a of the Brazilian General Data Protection Law), leaving a broad scope for interpretation that often escapes public debate and can be potentially used for implementing risky technologies.

A problem with the right to explanation

Only a specific, well-constructed right of explanation standard can allow AI to properly assist judges in the process of ruling on legal matters. So far, however, we also are not aware of how a potential right of explanation would work, since a legal explanation is enough to legitimise judicial resolutions. Citizens have the right to understand a judge's decision, and legal explanation is no longer enough when AI systems are involved. The transparency of the systems implemented in the public sphere is of the utmost importance to make understanding and legitimacy possible. It is important that these systems are permanently monitored and tested by independent committees that understand both the technical and legal features that legitimise their use. If we let algorithms determine what kind of consequences our actions legally have, we will be moving away from the standard of the rule of law and entering a judicial terrain that is mostly assisted, and decided on, by algorithms.

The problem of proportionate information

One problem with using AI technology in the judicial system is that there is no standard on how much information is adequate and necessary considering the purpose it is being collected for, even for simple repetitive cases such as those Prometea is used for, simply because human behaviour is complex and can express itself in multiple different scenarios. Any limitation of the information needed to justify or defend the subjects involved will result in arbitrary judgment (of course, collecting too much data can cause worse scenarios). It also creates the illusion that every document that Prometea uses for training is rightful without a competent analysis. There is a high possibility that out of 2,400 judgements used as a training set at the Attorney General's Office in Buenos Aires, a portion of them received some level of unfair treatment in their ruling and therefore there is a logical concern for a detailed analysis on every single document provided to train Prometea.

The legal responsibility of the AI developer

Given that AI constantly changes the algorithms in ways the developer cannot always foresee, how can they be responsible if their AI system causes damages?

The responsibility of an AI developer can be determined by two factors. One is the creation of a risk which is irrespective of any intention of those who designed, developed and implement the system. This is what we know as objective responsibility. Here it is always important to determine if the system is capable of conducting illegal actions *ab initio*, that is, from the outset of the development. For example, a developer would need to run regular risk assessments to identify, eliminate or reduce the possibilities of a system operating through bias or using information in a discriminatory way. Creating a creature like Dr. Frankenstein's could easily represent a risk, and if the "monster" suddenly turns uncontrollable, the scientist would be liable solely because of creating the risk.

The second factor focuses on the positive obligation that everyone has to not harm or violate anyone's rights. So if an AI system causes damages, the developers will be liable only if they have acted unlawfully because they have been negligent at some point of the AI life cycle, or because they had the intention to cause harm. In this case, the scientist behind Frankenstein's monster would not be responsible if he operated diligently.

Determining how responsibility will be attributed to those liable depends on what kind of system has produced the harm, the nature of the activity it was assisting, and how the harm was produced. With regard to objective responsibility, we understand it could be applied only to those systems that execute an activity which is already potentially risky. For the AI system that does not execute these kinds of activities, a subjective attribution should be considered, introducing the need to establish the nexus between the harmful outcome and the developer's actions, even when the oversight power is reduced as the system runs and the algorithms change.

Finally, since most people in Latin America do not have the opportunity to sue companies based in the European Union or in North America under their jurisdictions, as these are expensive legal procedures, many will have to choose their own jurisdictions, even with the delays, the random inefficiencies, and the lack of knowledge on how AI systems truly work.

Conclusion

New technologies such as AI, while offering an opportunity for innovation in our societies, including our legal systems, are raising a number of critical questions with respect to their application. Flawed digital technologies are increasingly at the core of our daily activities, and they interact with us. These technologies act like a substitute intelligence to which we can delegate tasks, ask for directions or answers to complex problems, unfolding the nature of reality by analysing data in ways we, as humans, cannot achieve alone.

Now, as we have seen, AI is already being used in our judicial system. Prometea is a first step towards the implementation of AI in the judicial system, and even if its only scope is "simple" cases, we are aware that this scope might widen if it keeps functioning "efficiently".

This success might lead to broader use of the technology in more complex stages of the judicial process, something the developers of the system already say they are looking forward to happening in the near future.[4]

Latin American countries face the challenge of fully integrating the everyday activities of their citizens with these new digital paradigms, while also protecting the unique characteristics and needs of a vast, diverse group of people in the region. After two centuries of somewhat stable rule of law, we have created a strong notion of the importance of our institutions and laws to guarantee the exercise of human rights. Even if AI offers a way of bringing the judiciary closer to the people, it needs to be implemented in a way that safeguards our shared sense of the importance of institutions and the law – this sense of importance is fundamental to our trust in these institutions.

Even when our current legal frameworks need revision and possible modification, we already have the legal means to protect our rights, with institutions and processes that offer legal recourse. Personal data, our privacy and freedom of expression find protection in our legal frameworks. The institutions of *habeas data,*[5] constitutional control

[4] Murgo, E. (2019, 17 May). Prometea, Inteligencia Artificial para agilizar la justicia. *Unidiversidad*. www.unidiversidad.com.ar/prometea-inteligencia-artificial-para-agilizar-la-justicia

[5] Habeas data is a constitutional remedy to rectify, protect, actualise or erase the data and information of an individual, collected by public or private subjects using manual or automated methods.

or the ordinary mechanisms to seek recourse for the harm caused by AI are already here.

But simply having the right laws at the right time is not necessarily always a winning formula. A strong, competent judicial system that understands what is at stake and how to respond to a highly technological age is needed. Procedural law and constitutional law must insure that no changes are made to due process without transparency. As Jason Tashea wrote for *Wired* magazine:

> How does a judge weigh the validity of a risk-assessment tool if she cannot understand its decision-making process? How could an appeals court know if the tool decided that socioeconomic factors, a constitutionally dubious input, determined a defendant's risk to society?[6]

Transparency is needed to protect due process and ultimately, the rule of law. If it is not provided – for instance, if the decision-making codes behind algorithms are protected as industrial secrets or intellectual property – due process and the rule of law will be in danger. Because of this, a broad understanding of the legal, ethical and rights repercussions of such a deployment should be sought.

Action steps

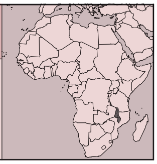
The following action steps are suggested for Latin America, and elsewhere in the world:

- Build creative narratives to communicate the risks and the repercussions of implementing AI systems within the public sphere.
- Design a capacity-building agenda for citizens to strengthen their right to due process in courts using Prometea or any similar system.
- Seek collaboration with other organisations and specialists in the region to build general consensus on the ethical use of Prometea and provide a broader understanding of the challenges it represents.
- Advocate for policy reform in order to include specific regulations on how AI should protect rights and how transparency should be realised in AI systems.
- Foster strategic litigation against AI violating human rights. Present a detailed set of evidence to support your claim.

[6] Tashea, J. (2017, 17 April). Courts are using AI to sentence criminals. That must stop now. *Wired*. https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now

MALAWI

FRAMING THE IMPACT OF ARTIFICIAL INTELLIGENCE ON THE PROTECTION OF WOMEN'S RIGHTS IN MALAWI



University of Livingstonia
Donald Flywell Malanga
donaldflywel@gmail.com www.unilia.ac.mw

Introduction

Artificial intelligence (AI) is touted to have the potential to address some of the human rights issues facing women in the world today.[1] However, literature on the current debate is mostly dominated by the developed world,[2] and few similar studies have been done in developing countries like Malawi. Consequently, there is lack of empirical evidence to substantiate how women conceptualise AI in the context of protecting their rights.

This report uses framing theory to understand how Malawian women speak about the impact of AI on human rights. Studies suggest that the way people frame (or conceptualise) an object affects the way they can engage and use it.[3] Therefore, if we know how women frame AI in their contexts, we can understand the factors that shape how they engage with AI technologies, and deploy AI in a manner that safeguards their rights.

Key objectives of this report are to analyse how women potentially frame AI and human rights in Malawi, their awareness of AI initiatives in the country, and how the rights of women are understood to be impacted positively or negatively by AI. Given this, a set of recommendations are made for improving the level of participation of women in AI initiatives in the country.

Context

Malawi gained its independence from Great Britain in 1964. It borders Tanzania, Zambia and Mozambique. The country has an estimated population of 18.1 million people, of which 85% live in rural areas.[4] The GDP per capita is roughly USD 390.[5] Most women are working in the agricultural sector, which is a backbone of Malawi's economy. Of those in non-agricultural waged employment, 21% are women and 79% are men, and the numbers have remained the same over the years. The overall mobile penetration is estimated at 45.5% while internet penetration is 6.5%.[6] About 34.5% of women own a mobile phone, 0.6% own a desktop computer, and 1.8% own a laptop, while just 4.7 % of them have access to the internet.[7] The low rates of information and communication technology (ICT) penetration in Malawi are attributed the country's weak economy, the high value-added tax (VAT) imposed on the importation of ICT gadgets, and other contextual factors.

State of women's rights in Malawi

Section 24 of Malawi's constitution stipulates that "women have the right to full and equal protection by the law, [and] have the right not to be discriminated against on the basis of their gender or marital status."[8] These rights are also operationalised in Malawi's National Gender Policy (2015),[9] National ICT Policy (2013),[10] and the Malawi Growth and Development Strategy III (2017).[11]

Despite such policy interventions, women's rights in Malawi are largely curtailed. Patriarchal beliefs and attitudes still prevail and many of the traditional cultural practices are harmful to women's rights. The unequal status of women is further exacerbated by poverty and discriminatory treatment in the family and public life. Malawi is ranked 173 out of 188 countries on the UN's Gender Inequality Index.[12]

[1] Access Now. (2018). *Human Rights in the Age of Artificial Intelligence*. https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf

[2] Cullen, D. (2018, 31 January). Why Artificial Intelligence is Already a Human Rights Issue. *Oxford Human Rights Hub*. https://ohrh.law.ox.ac.uk/why-artificial-intelligence-is-already-a-human-rights-issue

[3] Chigona, W., Mudavanhu, S. L., & Lwoga, T. (2016). Framing telecentres: Accounts of women in rural communities in South Africa and Tanzania. *CONF-IRM 2016 Proceedings*, 43. https://aisel.aisnet.org/confirm2016/43

[4] https://data.worldbank.org/indicator/SP.POP.TOTL?locations=MW

[5] https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=MW

[6] National Statistical Office. (2015). *Survey on access and usage of ICT services in Malawi – 2014: Report*. https://www.macra.org.mw/wp-content/uploads/2014/09/Survey_on-_Access_and_Usage_of_ICT_Services_2014_Report.pdf

[7] Ibid.

[8] www.malawi.gov.mw/images/Publications/act/Constitution%20of%20Malawi.pdf

[9] https://cepa.rmportal.net/Library/government-publications/National%20Gender%20Policy%202015.pdf/at_download/file

[10] https://www.macra.org.mw/wp-content/uploads/2014/07/Malawi-ICT-Policy-2013.pdf

[11] https://www.mw.undp.org/content/dam/malawi/docs/UNDP_Malawi_MGDS)%20III.pdf

[12] https://www.usaid.gov/malawi/fact-sheets/malawi-gender-equality-fact-sheet

Methodology

Ten women participants from the academic, government, civil society and private sectors participated in the interviews for this report. The author contacted them by phone and consent was given for face-to-face interviews. Some of the key questions that were posed to participants were as follows:

- What is your general understanding of AI?
- Can you describe some of the AI initiatives that you know or have participated in?
- How do you think AI is likely to impact on women's rights in the country?
- What policy/regulatory initiatives should be put in place to ensure women's rights are protected through the emerging AI regime?

Framing theory was adopted and textual analysis was used to analyse their views, opinions and insights. Framing refers to the process by which people develop a particular conceptualisation of an issue or re-orient their thinking to an issue.[13] Different people view the world in different ways and place emphasis and importance on different issues. This then results in people sometimes framing the same issues differently. This implies that frames help us to interpret the world around us and represent that world to others.[14] The frames suggested by the interviews helped to understand meanings that were attached to AI in relation to their rights. Table 1 lists the sector and type of work of the participants.

Defining AI and human rights

The majority of participants conceptualised AI in different contexts. For instance, to a lawyer, AI was understood as an automatic machine that can work independently of a human being, while other participants defined AI as a robotic, machine-learning, sensor or biometric device: "It's where biometrics or sensors are automatically able to capture data in a computer" (Resp. 9); "I think is when a computer is able to think like a human being" (Resp. 3); "It can be defined as a machine or computer that reasons like a human being" (Resp. 5).

From the definitions, it was clear that while none of the participants was able to capture the comprehensive definition of AI, their responses showed a basic awareness of AI that is in line with lay people's definitions of the technology. What was evident was that the majority of the informants attempted to define AI according to their own understanding, suggesting some measure of appropriation of the technology at a conceptual level. Likewise, on human rights, the majority of women confidently defined human rights as entitlements that any person including women are supposed to enjoy from birth: "Human rights are entitlements that women and any other person should enjoy from birth such as right to life, opinion, freedom of expression, just to name a few" (Resp. 2). It was also encouraging that some participants went further to categorise human rights as economic rights, social rights, cultural rights and solidarity rights.

TABLE 1.

Sector and type of work of interview respondents

Pseudonym	Position	Sector
Resp. 1	Lawyer	Government
Resp. 2	Human rights activist	Civil society
Resp. 3	Gender specialist	Civil society
Resp. 4	ICT expert	Government
Resp. 5	ICT expert	Private sector
Resp. 6	ICT lecturer	Academia
Resp. 7	AI lecturer	Academia
Resp. 8	Computer engineering student (Level 4)	Academia
Resp. 9	ICT student (Level 3)	Academia
Resp. 10	ICT education student (Level 2)	Academia

[13] Chong, D., & Druckman, J. N. (2007). A Theory of Framing and Opinion Formation in Competitive Elite Environments. *Journal of Communication*, *57*(1), 99-118.

[14] Chigona, W., Mudavanhu, S. L. and Lwoga, T. (2016). Op. cit.

Awareness of AI initiatives in Malawi

Most of the participants expressed ignorance of AI initiatives being implemented in the country – only a few women (from academia and the private sector) were able to mention any. This suggests that while a general awareness of AI can be found among the public, specific examples that demonstrate a practical knowledge of the implementation of AI are less readily available.

A review of AI initiatives in Malawi showed that they are at different maturity. Table 2 presents key AI initiatives in the country.

Table 2 indicates that AI projects in Malawi include machine-learning algorithms, drones,

TABLE 2.

Summary of major AI initiatives in Malawi

AI initiative	Description	Implementer/funder
Data intelligence	- Also known as planning, prediction, prompting and prodding for change (4P2C). - Uses geo-spatial and data analytic tools. - Provides accurate and real-time information for women and children and their families. - Used for monitoring and decision making in development and humanitarian processes.[a]	UNICEF Malawi in partnership with the Malawi government
Drones lab	- Provides a controlled platform for universities, the private sector and other expert organisations to explore how unmanned aerial vehicles (UAVs) can be used to deliver services that will benefit women and children in disadvantaged communities. - To make the project sustainable, UNICEF teaches local Malawians how to make drones and trains local pilots rather than relying on expatriates.[b]	Launched in June 2017 by UNICEF Malawi and the Malawi government
IBM Digital-Nation Africa (D-NA) Project	- The project offers training to students on emerging technologies such as AI, cloud computing, data science and analytics, blockchain and security, and the internet of things (IoT). - It has three platforms: (i) Watson Artificial Intelligence: Communicates with the user to build a profile, gives an overview of current job markets and suggests multiple learning paths. (ii) New Collar Advisor: A tool that performs skills gap analyses for users and helps in job matching. (iii) Innovators: Helps users explore, design and leverage cloud computing and AI services to build applications like chatbots.[c]	In December, 2018, International Business Machines (IBM) partnered with Malawi University of Science and Technology (MUST)
AI for patient diagnosis project	- An AI-powered digital pathology microscope slide scanner.[d] - The AI-powered system turns a microscope into a manual slide scanner and magnifies images of blood so that they can be linked to a computer or mobile phone and beamed across the world. - It is anticipated that over 28,000 children admitted with blood cancer are going to benefit from the initiative. - The AI microscope has improved paediatric services at the Queens Central Hospital.	A group of computing scientists and medics from Newcastle University in the United Kingdom in partnership with Queens Central Hospital in Malawi
Missing Maps Project	- Using AI-powered maps, helped to educate over 100,000 houses in three days about measles and rubella vaccines.[e]	Facebook, Red Cross and Malawi government
National Registration and Identification System	- A civil registration of national identities that uses biometric data and sensors.[f]	Malawi National Registration Bureau (NRB)

(a) https://www.unicef.org/malawi/innovation-0. (b) Ibid. (c) Mphande, J. (2018, 18 December). MUST partners IBM on ICT project. *Malawi University of Science and Technology*. https://www.must.ac.mw/2018/12/18/must-partners-ibm-on-ict-project. (d) Newcastle University. (2019, 22 February). Newcastle experts help African children with cancer. https://www.ncl.ac.uk/press/articles/latest/2019/02/malawimicroscope. (e) APO Group. (2019, 9 April). Facebook artificial intelligence (AI) researchers create the world's most detailed population density maps of Africa. *Africanews*. https://www.africanews.com/2019/04/09/facebook-artificial-intelligence-ai-researchers-create-the-worlds-most-detailed-population-density-maps-of-africa. (f) www.nrb.gov.mw/index.php/about-us/strategic-plan/item/3-the-national-registration-identification-system-nris-for-malawi imagery, biometrics and sensors, among many others.

None of the interview respondents had participated in an initiative using AI. This could be an indication of how women are neglected not only in the ICT sector in the country, but particularly from innovative future-thinking projects.

Framing the positive impact of AI on women's rights

The participants felt that AI had the potential to improve women's social and economic well-being if deployed using ethical standards and based on human rights principles: "I believe that AI is very broad and, as women, we can benefit both directly and indirectly as they affect our rights" (Resp. 6). For example, they stated that AI is likely to benefit women in the areas of education, health care, environmental management and good governance. As one participant put it, there are few areas of the economy where the use of AI will not impact on women, highlighting the need for the rights of women to be foregrounded in AI initiatives: "You know, women are represented in all sectors of the economy. Therefore, any area that AI is going to touch, women are always there. For example, if a company uses AI in banks, women are there, in agriculture, women are there, you name it... in all these areas our rights must be protected and our voices be heard" (Resp. 7).

The following were some of the specific potential positive impacts of AI mentioned by the participants:

- *AI can save women's lives in times of disaster*: Women stated that the use of AI that predicts extreme weather could help to respond to natural disasters in which the loss of lives of women and children could be prevented. "If we have AI that can tell us about bad weather to come or any disaster, it may help to save the lives of women. We had a disaster of Cyclone Idai in Malawi; you will find that the majority of people who lost their lives through floods were women" (Resp. 3).
- *Improve women's agricultural livelihoods*: In Malawi, 79% of agricultural activities are done by women. The participants stressed that the use of satellite images to collect weather data together with AI applications could provide information for women farmers to improve crop yields, and diagnose and treat crop and animal diseases. This may better the social and economic rights of women through improved livelihood outcomes: "In our country, we depend on natural rain fed for our crops to grow, and women are majority smallholder farmers. If we could have AI which predicts in time the occurring of crop pests and diseases, it means women farmers can use this information to control and prevent such pests and diseases. In the long run, they will improve their livelihoods such as increase income, improve crop and livestock yields, and be able to manage and deal with vulnerabilities" (Resp. 8).
- *Women's social and economic inclusivity*: Despite research suggesting that the use of AI in financial services can increase institutional bias, the participants felt that the use of AI applications such as credit scores for disbursement of loans can help to reduce gender bias: "The majority of women do not have access to financial credits in Malawi because sometimes they are biased due to gender stereotypes. If we have AI credit scores that do not bias on selection, it means many women may qualify for credits and loans" (Resp. 10). From the comment, it was clear that AI with the appropriate ethical standards can be used to challenge the discrimination experienced by women when trying to access credit and loans, enabling the social and economic rights of women.
- *Promote the health rights of women*: Participants observed that if they had AI technologies that could help to predict, diagnose and prevent the outbreak of diseases, the access to health care for marginalised and vulnerable women located in remote areas would be significantly improved. AI would enable health workers to intervene and contain any disease before spreading sporadically. AI could also be used to screen patients for diseases, and help improve medical services during pregnancy: "I expect in Malawi to have AI that could screen cervical cancer, help in surgery of pregnant women during delivery... because so many pregnant women die during delivery time due to inadequate midwifery professionals in our health facilities. We may reduce maternal and neonatal death by 2030 by using such technologies" (Resp. 2).
- *Equality in education and employment opportunities*: Participants expressed optimism that the use of AI for hiring workers could help more women get high-paying jobs without bias. In addition, some participants believed that AI could allow more young girls to enrol in top universities in the country such as the University of Malawi and Malawi University of Science and Technology: "AI may help young girls to enrol

at University of Malawi which is currently very competitive and many girls are left out" (Resp. 7); "When you go to an interview, maybe it is a technical job, and they see you as a woman, the likelihood that you will be picked becomes very low, even if you have the right qualification. I hope that if we can have an AI technology that interviews job applicants without asking gender and other affiliations, it may promote gender equality" (Resp. 9).

- *Fair criminal justice*: Participants also felt that AI could be used to promote fair trials in courts. In particular, it may help to provide women access to the justice system without bias and discrimination against gender: "Sometimes we have our court judges who are very corrupt, so when you have an AI technology that can provide a ruling or determination fairly, it means people may have trust in our justice systems. I do not know if this technology may work in Malawi..." (Resp. 1).

Framing the negative impact of AI on women's rights

Although women felt that the use of AI technologies could help them exercise and enjoy their social, political, cultural and economic rights both online and offline,[15] they also observed that all aspects of human interaction with ICTs in general are gendered. While AI can contribute to women's rights, a predominant concern was that a violation of rights such as privacy can occur, and discrimination can reinforce existing inequalities. They suggested that proper legislation and regulations were necessary to counteract this potential. The following were some of the concerns expressed by the participants:

- *Discrimination against women*: In general, the use of facial or speech recognition software tools can surveil and identify women and discriminate against them. AI can be used to create and disseminate disinformation, and political and socioeconomic propaganda. The targets of these campaigns are often likely to be women and other vulnerable individuals. One participant expressed a concern that AI would exacerbate the objectification of women that was already felt online: "Already the current ICTs have exacerbated privacy concerns for women. We see more nudes for women in social media platforms like YouTube, etc. than males, so how AI will protect us from such embarrassments?" (Resp. 3).
- *Censoring dissent:* Law enforcement agencies can use AI such as facial recognition to monitor the women activists who provide dissenting views to government. This may also violate the rights of women such as freedom of expression, opinion, movement, assembly, privacy and data protection.[16]
- *Right to equality in the work place and in education:* Despite expressing the view that AI could neutralise bias in work opportunities or enrolment in universities, some participants acknowledged that the opposite could also be true. AI used for hiring workers could promote bias and discriminate against women based on gender, race and ethnic backgrounds, among others. In addition, using AI to create credit scores for enrolling in a top college or university may discriminate against more deserving young girls who perhaps did not have the appropriate credit scores due to poor access to basic quality education.
- *Right to life and to health*: Participants expressed some fear of AI technology. For instance, they feared that the use of drones in conflict situations may result in more death and injury to women and children. Hospital patients undergoing surgery using AI such as robotics might die of a systems failure, something that could be avoided if a human was performing the surgery.
- *Family and marriage rights*: AI technologies that are used for reproductive screening were seen as good because they allowed women to know their reproductive status, but they may affect women's rights to get married if it was revealed that they were unable to conceive.

Conclusion

Although not comprehensive, this report suggests how women frame AI in the context of human rights in Malawi. The women interviewed showed a generalised notion of some characteristics of AI, such as automation, involving robotics, and the use of imagery, biometrics, machine learning and algorithms. However, they suggested a low level of awareness of AI projects being implemented in Malawi.

The participants believed that AI could promote their political, social, cultural and economic rights. Nevertheless, they feared that many AI initiatives being piloted in the country could cause more harm than good to their rights if legislative and regulatory frameworks are not revised to reflect the current AI

[15] Sida. (2015). *Gender and ICT*. (Gender Toolbox Brief). https://www.sida.se/contentassets/3a820dbd152f4fca98bacde8a8101e15/gender-and-ict.pdf

[16] Access Now. (2018). Op. cit.

regime. The participants suggested that AI is likely to violate women's rights such as privacy, equality, health rights and education rights, among others, all of which are rights contained in the country's constitution. Eventually, AI could perpetuate the already existing digital divide and digital inequalities that women are experiencing in the country.

A review of policies such as Malawi's National Gender Policy (2015),[17] National ICT for Development Policy (2006)[18] and National ICT Policy (2013)[19] show that gender equality and other women's rights are policy priorities for the government. However, there are no specific targets in place to monitor how such rights can be measured and achieved. Therefore, the findings in this report have implications for the role of civil society, the private sector, academia and the government for understanding the impact of AI on human rights.

Civil society should lobby for transparency and accountability in the deployment of AI so that these initiatives are more inclusive of women. This will also ensure that women are more aware of different AI deployments and be in a position to assess the risks those AI technologies are likely to pose to their rights. The government should review current policy frameworks such as the ICT Policy so that it is aligned with the new wave of AI technologies being implemented in the country. The government should work with the private sector and academia through public-private partnerships to deploy AI technologies that are beneficial to all citizens, including women. The government should also develop an affirmative action plan for women so that more young women are enrolled in AI technology and other science, technology, engineering and mathematics (STEM) graduate courses in our universities and colleges.

Action steps

The following steps are necessary in Malawi:

- The government should develop an AI strategy, with a strong focus on the role of women.
- Data protection legislation, regulations and standards should be developed by the government, the private sector and civil society organisations.
- The government should use the public-private partnership approach in the deployment of AI projects to ensure transparency and accountability.
- Human rights risk assessments should be conducted before AI technologies are deployed in the country.
- Regulatory or policy frameworks that fund and incentivise local women AI innovators need to be established.
- There should be investment in training and research into AI that includes women.
- Affirmative action is necessary to ensure more women participate and enrol in AI and other related ICT courses.
- Civil society should build knowledge and capacity among women's groups and organisations on AI.

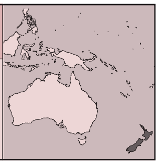
[17] https://cepa.rmportal.net/Library/government-publications/National%20Gender%20Policy%202015.pdf/at_download/file

[18] http://unpan1.un.org/intradoc/groups/public/documents/unpan/unpan033688.pdf

[19] https://www.macra.org.mw/wp-content/uploads/2014/07/Malawi-ICT-Policy-2013.pdf

NEW ZEALAND

ALGORITHMS AND SOCIAL MEDIA: A NEED FOR REGULATIONS TO CONTROL HARMFUL CONTENT?



New Zealand Law Foundation Artificial Intelligence and Law Project, Centre for Law and Policy in Emerging Technologies, University of Otago
Joy Liddicoat
https://www.cs.otago.ac.nz/research/ai/AI-Law/index.html

Introduction

On 15 March 2019, a white supremacist committed a terrorist attack on two mosques in Christchurch, murdering 51 people as they were peacefully worshipping, injuring many others and live streaming the attack on Facebook. The attack was the worst of its kind in New Zealand's history and prompted an emotional nationwide outpouring of solidarity with Muslim communities. Our prime minister, Jacinda Ardern, moved quickly, travelling immediately to the Muslim communities affected, framing the attack as one on all New Zealanders, vowing compassion, refusing to ever say the name of the attacker, issuing a pledge to ban semi-automatic weapons of the kind used in the attack, and steering her people through a difficult emotional time of grief, anger and shock. The global response led Ardern and French President Emmanuel Macron to issue the #Christchurch Call,[1] calling for, among other things, an examination of the use of algorithms by social media platforms to identify and interfere with terrorist extremist online content. This country report critically examines the events, including discussion of technical measures to find and moderate the objectionable content. In doing so, it asks whether the multistakeholder model is failing in the sphere of social media internet-related policy.

Country context

New Zealand is a small-islands country with a well-functioning democracy and stable economy. However, significant social inequalities exist among indigenous and some migrant populations, with children, young people and the elderly in particular economically, culturally and socially vulnerable in areas such as housing, health, education and income. Levels of internet access are generally high, there is good telecommunications infrastructure and an active information and communications technology sector.

The public discourse on artificial intelligence (AI) in New Zealand largely reflects that elsewhere, with an either utopian or dystopian view of how AI will affect humans, especially their employment. The positive benefits of using AI for service improvements is helping to change this, but most people still associate AI with robots and automated processing, rather than the machine learning or predictive algorithmic tools which are increasingly used in everyday life. More nuanced voices are emerging, but remain largely confined to a small range of actors mainly in the business, academic and government sectors.

Unlike other countries, there is no national AI strategy or research and development plan. However, while development, implementation and uptake of AI is patchy, it is growing. In government, for example, many departments are developing and implementing their own algorithms for a variety of service improvement purposes, rather than buying off-the-shelf products from third party foreign service providers.[2] In the private sector, a diverse range of businesses are developing AI-related products and services. There is one non-government AI organisation, the AI Forum, which was established in 2017 to bring together researchers, entrepreneurs, business and others to promote discussion and uptake of AI technologies.[3] In 2018, an inter-disciplinary initiative, the Centre for AI and Public Policy, was also established at the University of Otago.[4]

Few civil society voices offer critical analyses of AI issues, although there are pockets of activity such as in relation to harmful content, civic participation and social media use. Most of these voices appear to reflect the views of the small group of civil society actors commenting on internet policy and internet governance more generally. A small but growing number of community groups are developing their own machine-learning tools to deliver public information services, such as CitizenAI, which has developed the tools and is making them available on

[1] https://www.christchurchcall.com/christchurch-call.pdf

[2] Gavaghan, C., Knott, A., Maclaurin, J., Zerilli, J., & Liddicoat, J. (2019). *Government Use of Artificial Intelligence in New Zealand.* University of Otago and New Zealand Law Foundation. https://www.cs.otago.ac.nz/research/ai/AI-Law/NZLF%20report.pdf

[3] The AI Forum is also a member of the Partnership on AI. https://aiforum.org.nz

[4] https://www.otago.ac.nz/caipp/index.html

third-party platforms such as Facebook Messenger.[5] It was within this context that the #Christchurch Call was made, including the call for consideration of the use of algorithms by social media platforms.

What can we learn from the #Christchurch Call?

Defining harmful content

The prime minister was shocked that the attacker broadcasted the attack via Facebook's live-streaming service. The original footage was viewed 4,000 times before being removed by Facebook (the video was removed within 27 minutes of being uploaded). However, even within that short space of time the video had already been uploaded to 4chan, 8chan and other platforms. Within 24 hours the video had spread widely and 1.5 million copies of it had already been removed. There was one upload per second to YouTube alone within the first 24 hours. Like me, many people saw all or part of the video inadvertently, for example, as we followed news of the events on Twitter and had copies of the video posted in our feeds. This live streaming and sharing caused widespread public disgust and distress and there were demands that online platforms find and remove all copies of the recording.

A key question was the legal status of the video. The Films, Videos and Publications Classification Act 1993[6] regulates the distribution and sale of films, videos and other publications. The Act has a legal test relating to the harm that content might cause and established a chief censor with powers to give legal ratings to material (such as an age restriction, a parental guidance recommendation or a content warning).

The chief censor, David Shanks, viewed the attacker's video and ruled it was objectionable, within the legal meaning of the Act, considering its content was harmful if viewed by the public. The result was that distribution and possession of the video was a criminal offence. It is important to know that the chief censor did not ban the video. Instead, he ruled it could be made available to certain groups of people including experts, reporters and academics. Shanks considered that restricting distribution was a justifiable limitation pursuant to law under the New Zealand Bill of Rights Act 1990.[7] In doing so, he emphasised the video was not only a depiction of the attacks, but went further and promoted terrorism and murder:

> There is an important distinction to be made between "hate speech", which may be rejected by many right-thinking people but which is legal to express, and this type of publication, which is deliberately constructed to inspire further murder and terrorism. It crosses the line.

Shanks considered the material was of a similar nature to ISIS terrorist promotion material that had also previously been ruled objectionable. However, he said it was in the public interest for some people to have access to the video for legitimate purposes, including education, analysis and in-depth reporting, and advised that reporters, researchers and academics could apply for an exemption to access and hold a copy. For similar reasons, he also ruled the attacker's manifesto was objectionable, finding that it was likely to be persuasive to its intended audience and promoted terrorism, including mass murder.[8]

The decision was upheld on review.[9] The effect of the ruling was retrospective (the publications were deemed objectionable from the date created) with the result that it is a criminal offence to possess them; people who had already downloaded the video, for example, were advised they should "destroy it."[10]

Dancing to the algorithms

Having lawfully deemed the video to be objectionable, public debate turned to how the spread of the video could be halted. Some New Zealand internet service providers (ISPs) took their own initiative to block sites that were attempting to distribute the video on the day of the attacks.[11] These actions were criticised on the grounds that the blocking was not authorised according to law – the censor's ruling on the video was made three days after the attack – and that ISPs were wrong to take down the content before it had been declared objectionable.[12] The ISPs were certainly taking a risk in blocking content that had not been ruled unlawful. However, the ef-

[5] https://citizenai.nz

[6] legislation.govt.nz/act/public/1993/0094/latest/DLM312895.html?src=qs

[7] Office of Film and Literature Classification. (2019, 23 March). Christchurch attacks classification information. https://www.classificationoffice.govt.nz/news/latest-news/christchurch-attacks-press-releases

[8] The Censor did not impose tailored restrictions to allow journalists or researchers to access the manifesto. https://www.classificationoffice.govt.nz/news/featured-classification-decisions/the-great-replacement

[9] *Johnson v Office of Film and Literature Classification*, Film and Literature Review Board, Wellington, 14 April 2019.

[10] See Office of Film and Literature Classification. (2019, 23 March). Op. cit. Several people have already been convicted of possessing the objectionable material.

[11] https://twitter.com/simonmoutter/status/1106418640167952385

[12] See, for example, Free Speech Coalition. (2019, 25 March). Christchurch and Free Speech. https://www.freespeechcoalition.nz/christchurch_and_free_speech and Chen, C. (2019, 18 March). ISPs in AU and NZ censoring content without legal precedent. *Privacy News Online*. https://www.privateinternetaccess.com/blog/2019/03/isps-in-au-and-nz-start-censoring-the-internet-without-legal-precedent

fect of the censor's ruling was that the video was deemed objectionable *at the time it was made* and having retrospective effect, so that ISPs were not acting unlawfully in blocking access to the content on the day of the attacks. Had the censor ruled the video was not objectionable, the criticisms might have been more valid.

At the domain name service (DNS) level, the Domain Name Commissioner issued a statement saying that if necessary to protect the security of the .nz ccTLD and the DNS, the commissioner may suspend a registered domain name on the request of the government computer emergency response team or the Department of Internal Affairs.[13]

The prime minister sought to steer a balanced path between affirming internet openness on the one hand and human rights – including religious freedom – on the other. In her address to parliament four days after the attack, Ardern said:

> There is no question that [the] ideas and language of division and hate have existed for decades, but their form of distribution, the tools of organisation, they are new. We cannot simply sit back and accept that these platforms just exist and that what is said on them is not the responsibility of the place where they are published. They are the publisher. Not just the postman. There cannot be a case of all profit, no responsibility.

In preparing for a subsequent Paris summit on the #Christchurch Call in May 2019, Ardern made clear that her focus was on the harm of online terrorist extremist content, saying the "task here is to find ways to protect the freedom of the internet and its power to do good, while working together to find ways to end its use for terrorism."[14] In particular, she said: "We ask that you assess how your algorithms funnel people to extremist content and make transparent that work."

The leader of the opposition political party also weighed in, saying: "It's smart algorithms on the internet traffic into New Zealand that allow you to lawfully target, whether it's white supremacists or whatever those extremists are. I think we were overly cautious, I think we need to revisit that."[15]

Some commentators agreed with the prime minister that social media platforms can no longer say they are not publishers when their own algorithms enable and drive content sharing, while others said this was simply a new take on an old debate.[16] The risks of relying on algorithms to filter content are already known to be fraught. In New Zealand, for example, Good Bitches Baking is a not-for-profit network which shares home baking with people going through a difficult time.[17] However, whenever they attempt to post on Facebook they are blocked because of their name.

Many pointed to the futility of trying to chase copies of the terrorist attack video, likening this to playing "whack-a-mole" because the content would be constantly appearing elsewhere on the internet. The technical difficulty in identifying "copies" was highlighted, as small changes could be made to the video (such as editing to add material) making it difficult to clearly identify the relevant content. Concerns were raised about the collateral damage to legitimate and legal content on platforms. Some members of the Muslim community wanted to see the video (for example, to see if family members had survived) and some had in fact already watched it. There were concerns that prohibiting the video would drive it to the so-called dark web, thereby embedding the harm it was sought to avoid. Finally, some considered that there will "always be harmful content on the internet, outside of anyone's control" and there was little point in trying to contain this particular video.[18]

However, others pointed out that these arguments were fallacious, since algorithms were already being used to curate and feed content and could therefore be redesigned: there may be ways to interfere with recommendation algorithms to prevent the development of filter bubbles that channelled users to extremist content. Others decried Facebook's failure to implement its own community standards and its initial silence after the attack.[19] Another view was that algorithms feeding the objectionable video content to those who did not want to see it was an interference with their right to privacy (their right to be let alone and to decide for themselves what they wished to view).

[13] Carey, B. (2019, 29 March). Emergency Response to the Christchurch Terrorist Attacks. *Domain Name Commission*. https://www.dnc.org.nz/christchurchterroristattackresponse

[14] Ardern, J. (2019, 16 May). Christchurch Call opening statement. https://www.beehive.govt.nz/speech/jacinda-ardern%E2%80%99s-christchurch-call-opening-statement

[15] Simon Bridges of the National Party. See: https://www.tvnz.co.nz/one-news/new-zealand/simon-bridges-calls-tougher-cyber-security-laws-in-wake-christchurch-terror-attacks?variant=tb_v_1

[16] Brown, R. (2019, 12 April). This is not the internet you promised us. *The Spinoff*. https://thespinoff.co.nz/partner/actionstation/12-04-2019/this-is-not-the-internet-you-promised-us

[17] https://www.gbb.org.nz

[18] Moskovitz, D. (2019, 8 April). Publisher or postman? https://dave.moskovitz.co.nz/tag/freedom-of-speech

[19] Manhire, T. (2019, 19 May). Mark Zuckerberg, four days on, your silence on Christchurch is deafening. *The Spinoff*. https://thespinoff.co.nz/society/19-03-2019/mark-zuckerberg-four-days-on-your-silence-on-christchurch-is-deafening

The outcome of the #Christchurch Call included a set of voluntary commitments by governments and online service providers "intended to address the issue of terrorist and violent extremist content online."[20] Among the range of measures that online service providers agreed on was to:

> Review the operation of algorithms and other processes that may drive users towards and/or amplify terrorist and violent extremist content to better understand possible intervention points and to implement changes where this occurs. This may include using algorithms and other processes to redirect users from such content or the promotion of credible, positive alternatives or counter-narratives. This may include building appropriate mechanisms for reporting, designed in a multi-stakeholder process and without compromising trade secrets or the effectiveness of service providers' practices through unnecessary disclosure.

The topic of algorithms to promote and share content, as well as to find and limit its spread, was now squarely in the public domain – a considerable step forward in the discourse on AI. A small number of New Zealand groups have already responded to the outcome of the #Christchurch Call, urging more research with an interdisciplinary approach. The AI Forum's Ethics, Law and Society Working Group, for example, noted the three ways content is disseminated on the internet (user upload, internet searches and social media feeds), pointing out that implementing the #Christchurch Call could involve filtering at some point in each of these processes. Because some of these processes would have to be automatic, the challenge would be to identify items to be filtered in an AI classification system which can determine whether each item would be allowed or blocked.[21]

The group identified technical challenges to accurately identifying the relevant content, such as choosing the best classifiers, getting classification consistently correct, how to deal with errors, and whether classifiers should err on the side of accepting or blocking content. This in turn, the group said, raised ethical questions about freedom of expression and censorship and economic questions about the cost of running different types of filtering systems.[22] Despite these difficulties, the working group considered that "small changes to feed recommendation algorithms could potentially have large effects – not only in curbing the transmission of extremist material, but also in reducing the 'filter bubbles' that can channel users towards extremist political positions."

Other community responses

Communities responded to the attacks in a variety of ways. Many took up the prime minister's call to deny the attacker the infamy he sought by refusing to use his name and by not sharing any photos of him. For example, telecommunications company Spark called on people to support "a #ShareNoEvil movement that could help deprive terrorists of the fame and oxygen their evil needs to survive" and with the explicit ambition of "making the act of sharing terrorist content culturally unacceptable in Aotearoa." The campaign enabled supporters to download a Google Chrome extension that lets users block the attacker's name and replace it with the words "Share no evil".[23]

Muslim women leaders called out the community on the racism that they face, speaking of the efforts they made to alert the government, including police, to the harassment they were experiencing from white supremacists and other right wing groups. Islamic Women's Council spokesperson Anjam Rahman cited numerous examples of anti-Muslim and racist incidents both before and after the terrorist attacks and said "this is New Zealand."[24] Anti-racism activities sprang up throughout the country, even including a local TV series called "That's a Bit Racist".[25]

At the same time, the internet was used to spread misinformation about the attack, including to confuse or misinform the public about the proposed gun law reform, and a host of expressly racist and Islamophobic groups were set up on Facebook and other platforms.

The internet also enabled support for and sharing of the massive public outpourings of grief. Tens of thousands of people attended rallies throughout the country to decry the attack, holding public Muslim prayer vigils to show support for Muslim communities, to grieve, and to come together in

[20] Office of the Prime Minister. (2019, 16 May). Christchurch Call to eliminate terrorist and violent extremist online content adopted (press release). https://www.beehive.govt.nz/release/christchurch-call-eliminate-terrorist-and-violent-extremist-online-content-adopted

[21] AI Forum. (2019, 23 May). Reaction to the Christchurch Call from the AI Forum's Ethics, Law and Society Working Group. https://aiforum.org.nz/2019/05/23/reaction-to-the-christchurch-call-from-the-ai-forums-ethics-law-and-society-working-group

[22] Ibid.

[23] https://sharenoevil.co.nz

[24] Fitzgerald, K. (2019, 18 March). Christchurch terror attack: 'This is New Zealand' - Muslim woman reflects on past racist attacks. *Newshub*. https://www.newshub.co.nz/home/new-zealand/2019/03/christchurch-terror-attack-this-is-new-zealand-muslim-woman-reflects-on-past-racist-attacks.html

[25] https://www.tvnz.co.nz/shows/thats-a-bit-racist

acts of democratic solidarity and call for deeper examination and honesty about racism, religious intolerance and hate speech.

The New Zealand Law Society joined the debate about hate speech, to inform and educate about this form of speech and what it means to different groups of people.[26] The Free Speech Coalition expressed horror at the attack, saying that the "principle of freedom of expression should be inseparable from non-violence," but condemned the legal ruling on the status of the manifesto, saying New Zealanders "need to be able to understand the nature of evil and how it expresses itself."[27] The Coalition has so far been silent on the ruling of the video.

Algorithms, social media platforms and internet regulation

In this environment, the local internet community had to work hard to determine how best to engage with government and also create shared spaces for the community to discuss the issues. Much of this also involved educating about the nature of the internet, the various infrastructural layers and where content regulation fits within other areas of internet policy making. InternetNZ launched a forum for civil society and the technical community to participate in the lead-up to the meeting in Paris to support these discussions.[28]

Regulation of online content is fraught with problems including how to ensure lawful content (such as evidence of war crimes) is not affected by definitions of "terrorist" content. However, the international human rights standards provide a framework for balancing these different rights. At the same time, hard questions must be asked as to whether the multistakeholder cooperation processes which work to create agreed norms at the technical DNS layers of the internet are really working in the social media environment and whether, in the absence of an alternative, regulation has now arrived as the only realistic option. Jay Daley, for example, called for regulation of social media platforms primarily because these "are not the internet" and are not developed and coordinated in the same ways as other cooperative multistakeholder processes, such as IEEE, the IETF and W3C. Daley argues this is totally unlike the processes used by social media platforms, which have proven incapable of enforcing even their own moral code of conduct standards.[29]

Jordan Carter, the chief executive of InternetNZ, echoed this view, saying that the principles of an open and free internet and content regulation have been "elided, sometimes by organisations that are parts of our constituency, into that sort of cyber-libertarian ethos of government is always bad, freedom of speech is always good, any moves to regulate content or services are always bad." Given the market dominance of the social media platforms and their impact on public opinion, Carter argues, "just as the public square and the media were always regulated, it isn't obvious to me that these platforms should[30] be exempt just because they're on the internet."

The debate about regulation is continuing and it remains to be seen whether the outcome of the #Christchurch Call will have any significant, long-term impact. The role of social media in the terrorist attack has been expressly excluded from the terms of reference of the national inquiry into the attack.[31]

Conclusion

The effects of the terrorist attacks are still being felt in Christchurch and throughout New Zealand. Public support for gun control laws and wider discussion about racism shows that most New Zealanders abhor the actions of the attacker and want to take some level of personal responsibility for addressing racism in their daily life. Four months on from the attacks, our experience was that prompt legal classification of the video enabled take-down of online content according to the rule of law, thereby upholding and affirming the centrality of human rights in the midst of a horrific terrorist attack. The use of algorithms by social media platforms received considerable attention, helping to inform the public and give more nuance and depth to discussions about AI. This bodes well for future discussion of AI. However, more research is needed to understand

[26] Cormack, T. (2019, 4 April). Freedom of speech vs Hate speech. *New Zealand Law Society*. https://www.lawsociety.org.nz/practice-resources/practice-areas/human-rights/freedom-of-speech-vs-hate-speech

[27] Freedom of Speech Coalition. (2019, 23 March). Banning of manifesto is a step too far. https://www.freespeechcoalition.nz/banning_of_manifesto_a_step_too_far

[28] The forum was open to all interested civil society and technical community members in New Zealand and globally. https://christchurchcallcoord.internetnz.nz

[29] Daley, J. (2019, 16 April). A case for regulating social media platforms. *LinkedIn*. https://www.linkedin.com/pulse/case-regulating-social-media-platforms-jay-daley

[30] Brown, R. (2019, 12 April). Op. cit.

[31] Ardern, J. (2019, 8 April). Supreme Court judge to lead terror attack Royal Commission (press release). https://www.beehive.govt.nz/release/supreme-court-judge-lead-terror-attack-royal-commission. Para 6(3)(b) of the terms of reference of the inquiry limits the matters that the Inquiry has power to consider, namely "activities by entities or organisations outside the State Sector, such as media platforms." See: https://christchurchattack.royalcommission.nz/about-the-inquiry/terms-of-reference

the human rights and ethical implications of diverse algorithmic classifiers and the rules that might be created to identify and curate online content.

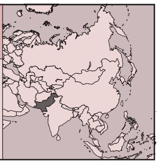
Action steps

The following action steps can be suggested for civil society:

- Foster increased public discussion about AI and provide case studies to improve and build understanding of the human rights issues involved.
- Strengthen and develop an interdisciplinary approach to AI to ensure technical, philosophical, legal and other approaches are brought together to develop responses.
- Ensure civil society, academic and technical perspectives play an equal role with government and business perspectives in developing responses to the human rights implications of AI.
- Continue to deal with specific types of harmful content according to law, such as extremist terrorist online content, rather than rejecting regulation of content *per se*.
- Develop research strategies to support technical considerations of the human rights and ethical issues that arise in the development of AI tools (for example, identifying appropriate classifiers and the use of diverse data sets for machine-learning tools).

PAKISTAN

ARTIFICIAL INTELLIGENCE AND ITS IMPACT ON FREEDOM OF OPINION AND EXPRESSION IN PAKISTAN



Internet Policy Observatory Pakistan
Arzak Khan
www.ipop.org.pk

Introduction

The field of artificial intelligence (AI) in Pakistan is evolving rapidly and poised to grow significantly over the coming decade. AI is a technology that is creating new opportunities in education, promoting equality and freedom of expression and access to information. The advancements in data collection, processing, increased computing power and low costs of storage have resulted in numerous AI start-ups offering insights into the control of new diseases, identifying patterns of human interactions,[1] setting up smart electrical systems[2] and intelligent irrigation systems,[3] powering smart cities,[4] and analysing shopping data,[5] apart from the not-so-popular uses of AI such as identifying tax evaders[6] and policing.[7]

AI-powered applications are set to play a very critical role in the cyber ecosystem in Pakistan, which will also introduce significant risks and challenges by amplifying existing bias, discrimination and ethical issues in its governance. AI-driven algorithms and applications are growing at a rapid pace, but most of the development is happening in the private sector, where little consideration or thought is given to human rights principles when designing computer code carrying instructions to translate data into conclusions, information or outputs. The potential impact on human rights is worsened given that no framework is available in Pakistan that regulates the application of AI from a human rights perspective.

The aim of this maiden scoping exercise is to look at the human rights legal framework for AI in Pakistan and to propose a theoretical framing for thinking about the obligations of government and responsibilities of private companies in the country, intersecting human rights with the expanding technological capabilities of AI systems.

AI and human rights in Pakistan

The growing digitalisation in our everyday lives has given way to the rise of a robotic age with machine-driven AI reshaping the way we connect and perceive the digital world. The benefits and advantages of the internet since its very inception have not been evenly distributed and there is a perceived risk that AI systems will further exacerbate the inequalities by facilitating discrimination and impacting marginalised populations and nations like ours.

Despite the plethora of projects, AI and machine-based learning technologies are still in their infancy in Pakistan. This means that it is very important that in employing new, emerging technologies both public and private sector organisations in the country find ways to protect human rights, primarily due to the fact that these technologies can exacerbate discrimination, inequality and exclusions already prevalent in Pakistani society.

Pakistan has a poor human rights record. The international community and human rights organisations have long been concerned about the persistent patterns of human rights violations occurring in the country, including forced disappearances, torture and extrajudicial execution.[8] The suppression of freedom of expression online has intensified and the introduction of the Prevention of Electronic Crimes Act, 2016[9] is being used to threaten, harass and detain human rights activists, silencing them online. Forced disappearance has become widespread and paid trolls harass in-

[1] McGee-Smith, S. (2019, 30 January). AT&T: Intelligent Pairing Ups Agent Success. *No Jitter*. https://www.nojitter.com/contact-center-customer-experience/att-intelligent-pairing-ups-agent-success

[2] Jazz. (2019, 22 January). Jazz reducing electricity line losses with GSMA, CISNR, and PESCO. *The Nation*. https://nation.com.pk/22-Jan-2019/jazz-reducing-electricity-line-losses-with-gsma-cisnr-and-pesco

[3] www.rcai.pk/ResearchCenterAI/project/pp1.html

[4] Sethi, M. (2019, 6 June). Opinion: Smart cities will heal the world. *SilverKris*. https://www.silverkris.com/opinion-smart-cities-ayesha-khanna

[5] Ibid.

[6] Bhutta, Z. (2019, 13 March). Govt to use NADRA's database to detect tax evaders. *The Express Tribune*. https://tribune.com.pk/story/1928627/2-govt-use-nadras-database-detect-tax-evaders

[7] Hussain, S. (2019, 27 March). Capital's cops to get facial recognition devices soon. *The Express Tribune*. https://tribune.com.pk/story/1937963/8-capitals-cops-to-get-facial-recognition-devices-soon

[8] https://www.amnestyusa.org/countries/pakistan

[9] www.na.gov.pk/uploads/documents/1470910659_707.pdf

dividuals online with impunity. Violence triggered by alleged "blasphemy" recently claimed the life of a young university student, leading to rare condemnation from the government after pressure from social media users.[10] All these issues highlight the importance of a human rights framework for AI.

Automation helps remove human interventions from decision-making processes, which can have both positive and negative impacts on human rights,[11] whether accessing business or government services, or simply accessing content online. AI also poses new risks for human rights, as diverse as non-discrimination, privacy, security, freedom of expression, freedom of association, the right to work and access to public services.[12] Pakistan has the opportunity to shape the direction of AI in the country from a socio-ethical perspective by attending to important issues such as privacy, data protection, human dignity and non-discrimination. As the AI ethics guidelines presented by the European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG) rightly suggest, "AI should be developed, deployed and used with an ethical purpose, grounded in, and reflective of, fundamental rights, societal values and the ethical principles of beneficence, non-maleficence, autonomy of humans, and justice."[13] Pakistan should also follow these guidelines in the deploying of AI systems to ensure a human-centric approach to AI with the aim of maximising the benefits of AI and minimising its risks.

The use of AI in the public sector

Because AI is driving towards greater personalisation in an era of information abundance,[14] it can also help government agencies in developing countries like Pakistan solve complex public sector issues by giving citizens a more personalised and efficient experience. There is huge opportunity for the government to capitalise on this new technology to improve people's lives, end poverty and introduce transparency in decision-making processes by reducing fraud, waste and abuse in all departments. In a bid to increase tax collection, the government has embarked on using AI and big data to identify tax evaders.[15] It is hoped that this will help widen the tax net in Pakistan and make tax auditing faster and more efficient. The National Database and Registration Authority (NADRA),[16] custodian of citizen records, in collaboration with the Federal Board of Revenue, have identified 53 million individuals as tax evaders using data mining and AI technologies.[17]

The widespread collection of citizens' data by the government, and the mining of multiple data sets, increases the risk of manipulation, bias and discrimination based on ethnicity, gender, geography and holding pluralistic or dissenting political views. One such data set involves the national identity smartcard issued by NADRA, which is required for banking transactions,[18] purchasing vehicles,[19] real estate[20] and airline tickets, and travelling abroad,[21] among other transactions. The use of facial recognition technology is another area in the public sector where the government has emphasised the use of AI algorithms. Facial features will become part of a system used for facial identification, matched against existing data sets with citizens' information. Facial recognition technology can easily identify "discriminating" features on a person and help anyone from researchers to law-enforcement authorities find who they are looking for quicker than ever.[22]

The most pressing challenge for adopting AI technologies is that they should be consistent, accurate and consciously checked for unintended bias. The algorithms used in powering AI platforms must be stable and transparent in design to ensure that small changes in inputs do not change the

[10] https://www.amnesty.org/en/countries/asia-and-the-pacific/pakistan

[11] Council of Europe. (2018, 22 March). Algorithms and Human Rights: a new study has been published. https://www.coe.int/en/web/freedom-expression/-/algorithms-and-human-rights-a-new-study-has-been-published

[12] Scientific Foresight (STOA). (2019, 19 March). Is artificial intelligence a human rights issue? *European Parliamentary Research Service Blog*. https://epthinktank.eu/2019/03/19/is-artificial-intelligence-a-human-rights-issue

[13] European Commission. (2018, 18 December). Draft Ethics guidelines for trustworthy AI. https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai

[14] Nyst, C., & Monaco, N. (2018). *State Sponsored Trolling: How Governments are Deploying Disinformation as Part of Broader Digital Harassment Campaigns*. Palo Alto: Institute for the Future. http://www.iftf.org/fileadmin/user_upload/images/DigIntel/IFTF_State_sponsored_trolling_report.pdf

[15] Sajid, H. (2018, 8 October). Pakistan government to use Big Data and AI to find tax evaders. *Samaa*. https://www.samaa.tv/news/2018/10/pakistan-government-to-use-big-data-and-ai-to-find-tax-evaders

[16] https://www.nadra.gov.pk

[17] Khan, M. (2019, 22 June). Online tax profiling system for 53m people unveiled. *Dawn*. https://www.dawn.com/news/1489672

[18] Dawn. (2010, 22 December). SBP's directive to banks on CNIC. *Dawn*. https://www.dawn.com/news/592915

[19] INCPak. (2019, 9 February). Biometric verification mandatory for registration and transfer of vehicle. *INCPak*. https://www.incpak.com/info/biometric-verification-mandatory-for-registration-and-transfer-of-vehicle

[20] Dawn. (2011, 23 June). Ban imposed on real estate deals without police, Nadra checks. *Dawn*. https://www.dawn.com/news/638786

[21] Passports are linked to the smartcard database.

[22] Khan, D. (2018, 30 March). Pakistan is Developing Its Own Facial Recognition System for Security Purposes. *ProPakistani*. https://propakistani.pk/2018/03/30/pakistan-is-developing-its-own-facial-recognition-system-for-security-purposes

output results. It is important that during design the principles of algorithmic fairness and bias-free algorithms are considered. The United Kingdom recently published its guidelines on the use of AI in the public sector, highlighting the need for a balanced, ethical, fair and safe approach to the application of algorithms.[23] These kinds of initiatives can serve as useful templates in developing a similar framework for the application of AI in the public sector in Pakistan.

The use of AI in the private sector

AI has disrupted the way businesses interact with consumers. It has already become integrated into our daily lives and AI is found in all of our digital devices, including search engines, social media platforms and chat applications. Google, Facebook, Amazon, Apple and Netflix all use AI as means of enhancing (and controlling) the user's experience.

The private sector in Pakistan has been quick to realise the immense potential of AI in improving operations and sales strategies in businesses.[24] Daraz, one of the leading e-commerce platforms in Pakistan, and recently acquired by the China-based Ali Baba group,[25] is empowering tens of thousands of sellers to connect with millions of customers using AI to power its smart search app. The app is powered by deep learning and natural language processing algorithms. It recommends products to shoppers, and informs retailers when they need to increase their inventory to keep up with the demand. The leading telecom service provider, Pakistan Telecommunication Company, also uses AI for interacting with its customers.[26]

The market for AI-powered assistants is growing in Pakistan. Similar to Apple's Siri, Google's Assistant and Microsoft's Cortana personal assistants, RUBA, the world's first Urdu-speaking AI bot has been launched.[27] The AI bot can speak and understand Urdu while working as your personal assistant in performing tasks such as messaging someone, finding a contact or calling them.

Meanwhile, using big data, machine learning and cognitive intelligence, Mindshare Pakistan has built AI-rich solutions that have transformed the way media planning is done, for instance, through customising advertising using AI.[28]

Finally, ADDO AI, a company started in Pakistan, is one of the emerging AI companies in Asia that is using historical and real-time data gathered through satellites and hyperspectral imaging for planting, harvesting and crop management.[29] The goal is to improve the crop yields for farmers and to protect seed lenders[30] against risks and climate challenges.

These are just some of the AI initiatives in the country's technological sector, which all show how seriously the potential of AI is being taken for business planning and operations at all levels.

A human rights framework for AI in Pakistan

AI is driving towards greater personalisation of the online interactions of users. This, on one end, allows easy and timely access to required information based on personal preferences as learned by an AI system using a set of behaviour patterns. On the other end, it minimises the exposure to diversified views. Such personalisation may reinforce biases and incentivise the promotion and recommendation of inflammatory content or disinformation in order to sustain users' online engagement.[31] AI-based personalisation risks undermining an individual's ability to find certain kinds of content online by deprioritising content with lower index or search rankings, especially when the system's technical design occludes any kind of transparency in its algorithms.

AI has further encouraged the use of big data-driven systems in both the private and public sector, with the widespread collection and exploitation of individuals' data, increasing risk of manipulation, discrimination and micro-targeting.

[23] Government Digital Service & Office for Artificial Intelligence. (2019). *A guide to using artificial intelligence in the public sector*. https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector

[24] Minetti, S. (2018, 21 April). Ensuring the private sector takes full advantage of AI solutions. *Global Banking & Finance*. https://www.globalbankingandfinance.com/ensuring-the-private-sector-takes-full-advantage-of-ai-solutions

[25] The Express Tribune. (2018, 8 May). China's Alibaba Group acquires Daraz. *The Express Tribune*. https://tribune.com.pk/story/1705095/2-chinas-alibaba-group-acquires-daraz

[26] Pakistan Telecommunication Company Limited (PTCL). (2017, 24 April). PTCL first to use Afiniti's Artificial Intelligence Solution (AIS) in Asia. https://ptcl.com.pk/Home/PressReleaseDetail?ItemId=540&linkId=0

[27] Rehman, S. (2019, 3 April). World's First Urdu Speaking AI Assistant Launched. *ProPakistani*. https://propakistani.pk/2019/05/03/worlds-first-urdu-speaking-ai-assistant-launched

[28] The News. (2019, 22 February). Mindshare Pakistan has built AI-rich Media Planning solutions. *The News*. https://www.thenews.com.pk/print/435128-mindshare-pakistan-has-built-ai-rich-media-planning-solutions

[29] Hynes, C. (2017, 31 August). Four Companies Using AI To Transform The World. *Forbes*. https://www.forbes.com/sites/chynes/2017/08/31/four-companies-using-ai-to-transform-the-world/#3b5284744038 #40eb31a14038

[30] In Pakistan, most farmers do not have the financial means for buying seeds for themselves, so seed lenders are people who give away seeds to be repaid later once the crops yield.

[31] Tufekci, Z. (2018, 10 March). YouTube, the Great Radicalizer. *The New York Times*. https://www.nytimes.com/2018/03/10/opinion/sunday/youtube-politics-radical.html

Both public sector organisations and private sector companies should provide meaningful information and insights about how they develop and implement standards for personalising the information environment on their platforms. The opaqueness of the algorithms powering AI systems emphasises the importance of obligations of government and the private sector towards the human rights framework that imposes responsibility on both to desist from implementing measures that violate individual human rights. At a policy level, the government and private sector should make policy commitments to respect the human rights of their users in all AI applications, relevant to the collection and use of personal data needed to feed these systems.

Given the methods of profiling individuals using third-party data in Pakistan, it is imperative that the government devise data protection laws and standards with modern notions of consent, reason, use, transparency and accountability. This is particularly important if the personal data is being used without being de-identified before developing new data sets.

As in the government's tax profiling initiative, a further complication is introduced when multiple data sets are used for raising a new data set, in this case of suspected tax evaders. Such data sets have a tendency to lose the original context of where and when the data was acquired, and risk decisions against individuals with mostly out-of-date or inaccurate data.[32]

Conclusion and recommendations

Many countries are involved in devising global, regional and national AI policies to maximise the potential and minimise the risks of this technology. It is imperative that Pakistan should develop its AI strategy and guidelines with a human rights framework at the core of its policy. The government should not focus only on public sector regulation of the use of AI, but also regulate its use in the private sector. This is because the private sector is more likely to dominate the AI evolution due to its ability to invest long-term in AI development and initiatives.

Furthermore, there is the risk that governments might hand over sensitive data to companies[33] for use in AI platforms, without the necessary safeguards in place. Pakistan should look at the framework provided by the European Union's General Data Protection Regulation (GDPR)[34] and other similar policy frameworks[35] to prevent the unaccountable use of AI and the use of personal data for training these systems. Efforts should be made to provide guidelines on ethical coding, design and application, and attention paid to the need for public oversight of AI machine-learning systems. Transparency is crucial in AI systems given that they are destined to have a significant impact on the lives of communities and individuals in Pakistan.

[32] Schwartz, B. (2019, 2 June). Google pre-announces June 2019 core search algorithm update. *Search Engine Land*. https://searchengineland.com/google-pre-announces-june-2019-core-search-algorithm-update-317698

[33] Due to political pressure, corruption and the importance of smartcard data to businesses.

[34] https://gdpr-info.eu

[35] Dutton, T. (2018, 28 June). An Overview of National AI Strategies. *Politics + AI*. https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

PANAMA

ARTIFICIAL INTELLIGENCE IN PANAMA: THE PROTOTYPE PHASE

IPANDETEC (Panama & Central America)
Lia Hernández and Kayc James
www.ipandetec.org

Introduction

It is important to recognise the need for artificial intelligence (AI) applications, not only in Panama, but in the world. AI has been shown to have great potential in its application in a variety of fields, including health, agriculture, finance and education. Panama is, however, right at the beginning of the AI learning curve. This is not surprising: there are 2.8 million internet users in Panama, representing 70% of the country's population.[1] Among the first rudimentary steps in using AI in the country was the use of fingerprint technology to clock in workers. But while the country has made advances in AI, it has yet to be used on public service platforms, making the benefits of AI accessible to citizens. This report discusses Panama as being in the "prototype phase" of implementing AI.

Ready or not?

The main interest in AI in Panama can be found in the private sector, which has been primarily engaged in providing AI business solutions. While the government is out of step with these developments, it is nevertheless focused on issues to do with data protection and open government which can serve as a springboard for the future incorporation of AI in public services.

One example of the government's interest in the application of ICTs in the public service sector is Panama Online,[2] which is one of its flagship projects. This platform aims to streamline the way in which citizens transact with state services. For example, citizens can access their police clearance records using an automated system that searches the police database and provides them with the necessary document, without any visit to a police station being necessary.

Despite e-government initiatives like these, one just has to look at other countries in Latin America, with AI projects in the banking, insurance and justice sectors, among others, to see how far Panama is lagging behind.

Below we briefly discuss the potential use of AI in the finance, health and education sectors in Panama.

Finance

E-commerce in Panama is regulated by Law No. 51 of 2008 and by Executive Decree No. 40 of 2009, with subsequent laws dealing with e-commerce also being passed.

In September 2018, the General Directorate of Income initiated a pilot electronic billing programme with 43 companies. The purpose of the programme was to facilitate e-commerce transactions in the country, including electronic invoicing, and digital archiving of transactions. Banks also offer online banking, allowing clients to access customer care using chat, open accounts, and transact. However, it is only private sector banks that offer online services. The state bank has yet to get online.

It is clear that while there is potential for the use of AI in the finance sector – for instance, in predictive modelling of government budgets, in economic forecasts, or in customer services – the context does not necessarily exist for this to become a widespread reality. It was only in 2018 that the government launched its e-billing programme, showing the extent to which the public sector on the whole is behind the e-commerce curve.

Health

There is perhaps more potential for AI pilot projects in the health sector. In April 2019 an initiative was launched by the Ministry of Health (MINSA) which involved using electronic bracelets for newborn babies and their mothers to avoid mix-ups at hospitals and clinics, and to curb baby snatching, a common crime in the country.

The Bracelet System for Neonatal Control works using an ankle bracelet with a transmitter that is placed on the newborn baby, and a receiver bracelet which the mother wears. It includes a hand-held programming device, control antennas on clinic doors (which light up red or green), and an alarm.[3]

[1] https://www.elcapitalfinanciero.com/el-70-de-la-poblacion-en-panama-usa-internet-pero-que-hace

[2] www.panamaenlinea.gob.pa

[3] https://www.prensa.com/sociedad/Presentan-sistema-seguridad-control-nacidos_0_5290720902.html

The bracelets also have a built-in coloured light – green for compatible and red for incompatible – that turns on when mother and baby are together.

This project will be piloted in the province of Chiriqui. However, it has been met with mixed reactions. These reactions show that public understanding and sentiment with any new initiative are important to ensuring its success – and the same will be the case with respect to AI initiatives. A key challenge in the use of AI in Panama will be the public's willingness to participate in these kinds of initiatives.

Education

Panama has implemented e-learning platforms to help with its teaching and learning. Some platforms have been developed by schools and universities themselves, and others have been through partnerships with technology service providers.

Some of these platforms include alerts sent to students to remind them of assignment deadlines, the time that has passed since they logged onto a platform, and messages relating to absenteeism. These are first-level e-education platforms, and there is some way to go to explore other potential areas of AI use in schools, such as the potential use of AI to reduce school drop-out rates as seen in India.[4]

Prototype phase

As suggested by the above initiatives, the first-step platforms have been created to lay the ground for more sophisticated use of AI in these sectors. Because of this, we consider Panama as being in the prototype phase of AI implementation. But there remains much room for improvement. Most initiatives that are looking to the use of AI in the country are isolated, including in the private sector.

Two of the most notable initiatives in the private sector involve Microsoft. In the first, Microsoft has partnered with Copa Airlines. The two have reached an agreement on the use of Microsoft's cloud services, productivity solutions and applications related to AI. This partnership is seen as critical in a sector where efficiency and punctuality are essential.[5]

Another company with which Microsoft is working is Atento, a leader in contact centres and customer relations in Latin America, which is also using the cloud platform and cognitive services of the tech giant.

Operating in Guatemala, Panama and El Salvador, Atento created a new "customer service intelligence" area. According to César Cernuda, the president of Microsoft Latin America:

> The implemented solution applies modern voice transcription services, understands the reason for contact (interpretation of intention), and analyses the client's sentiment (nervous, anxious, aggressive, happy). Artificial intelligence allows it to develop the profile more deeply, the inclinations and preferences of the consumer, establishing more effective contact and greater satisfaction.[6]

These examples suggest a clear benefit of AI in the business sector. According to reports, businesses experience a 15% to 30% improvement in efficiency when applying AI to a process or service – and it is this potential that could also prove useful in the public sector.

Conclusion

For IPANDETEC, it is very important to take the necessary steps to lay the ground for the wide-scale implementation of AI in Panama. This needs to be done in a way that AI enables the human rights of citizens, rather than diminishes them. From this perspective we can learn a lot from the implementation of AI in other countries where the use of technologies is more advanced, whether in the region, or further afield, as in Asia. It is a matter of growing as a country and developing the most promising initiatives that balance the need for technological development, but do not continue to disadvantage groups who are already marginalised.

We see great potential for the use of AI in numerous ways, including the treatment of diseases like cancer (MIT, for example, recently developed an IT-based algorithm that can predict breast cancer up to five years before it appears);[7] in the use of natural language processing for government services; in the support offered to farmers, including in pest identification; and in education. Many of these solutions require a detailed case analysis of the experiences of their use in other countries, especially those that are, like Panama, in the prototype phase of development.

[4] See the India country report in this edition of GISWatch.

[5] https://www.estrategiaynegocios.net/lasclavesdeldia/1180122-330/inteligencia-artificial-c%C3%B3mo-las-empresas-en-centroam%C3%A9rica-ya-la-utilizan

[6] Ibid.

[7] Conner-Simons, A., & Gordon, R. (2019, 7 May). Using AI to predict breast cancer and personalize care. *MIT News*. https://news.mit.edu/2019/using-ai-predict-breast-cancer-and-personalize-care-0507

Action steps

The following action steps in Panama will further help to lay the ground for the widespread application of AI:

- Renew the debate on the Law on Protection of Personal Data with greater participation by citizens and civil society, and include aspects such as territorial jurisdiction, among others.
- Push for Panama to implement the international conventions to which it is signatory in terms of digital rights. In particular, implement the Budapest Convention on Cybercrime[8] in its entirety in Panama.
- Panama should improve regulation of areas such as cryptocurrencies, e-commerce and e-invoicing, among others, to encourage foreign investment and protect the national financial system.
- The different institutions of the state must be encouraged to cooperate with ANTAI – the national authority on access to information and transparency – in terms of the country's open government and open data programme. This is important in order to increase transparency and trust in the government.
- The government must promote human and digital rights throughout the country to have an informed and aware population.

[8] https://www.coe.int/en/web/cybercrime/the-budapest-convention

PERU

ALGORITHMIC BIAS: A FIRST LOOK AT DISCRIMINATION AND THE FINTECH SECTOR IN PERU

Hiperderecho
Carlos Guerrero Argote
www.hiperderecho.org

Introduction

This report considers the use of artificial intelligence (AI) in the so-called "fintech" sector in Peru. Fintech refers to emerging industries that use technology to improve activities in the financial market, for example, to offer loans or mobile bank services. The report frames the discussion in terms of the potential of AI to unfairly discriminate against potential customers, laying the basis for more detailed studies in the sector.

A set of recommendations aimed at reducing potential discrimination in the fintech sector is also proposed, in order to exploit the potential of fintech to ensure the inclusion of vulnerable groups in the financial system.

Background

In 2018, the Inter-American Development Bank (IDB) and Finnovista, an organisation that promotes fintech ventures in the Latin American region, published the *Report on Fintech in Latin America 2018: Growth and Consolidation*, a follow-up study on the development of fintech in the region. It was reported that Peru was ranked sixth out of a total of 18 countries in the region, with 5% of the total fintech market and a total of 57 fintech ventures in the country. Compared to the measurement taken in 2017, the Peruvian fintech market experienced growth of 256%, the second largest growth experienced in Latin America.[1]

In May 2019 the Lima Fintech Forum 2019 was held in the capital, the third such event that, according to its own description, "brings together the main representatives in fintech, banking, insurance, regulation, cybersecurity, digital transformation and innovation in Peru."[2] The event had the participation of a high authority in the Peruvian government, the vice minister of economy, who expressed his satisfaction with the growth of the fintech companies and his commitment to support a regulatory framework favourable to their interests.

However, despite what seems like universal enthusiasm with which this new business segment has been received, the understanding of the impact of fintech in Peru seems to be limited, even within the Peruvian financial industry. There are several elements that confirm this. Despite its exponential growth, the information available on fintech is scarce in the media, and when it is available, it is usually self-promotional, and basically limited to the websites where they offer products and services. In addition, many of the traditional financial actors still do not recognise fintech ventures as important agents of their ecosystem, which translates into a low level of strategic cooperation and "coopetition".[3] Finally, studies conducted by universities on the subject are scarce and are focused almost exclusively on describing the construction and operation of the business models of these companies.[4]

In Peru, discussions on fintech tend to revolve around economics and financial regulation, but do not touch on other topics that are already widely discussed in more mature ecosystems, such as the ethics of the use of technology or its implications for human rights. In this scenario, trying to diagnose the use of AI in the Peruvian fintech sector to establish the existence of discrimination seems premature, and can be compared to talking about rocket science to a pre-industrial audience. However, given the rapid growth of the sector, it is at least necessary to frame the discussion for further debate and investigation.

Algorithmic bias and discrimination

Nowadays, it is an indisputable fact that technologies such as AI can generate or reproduce situations of discrimination against people as a product of a design contaminated by the prejudices and unfair

[1] Inter-American Development Bank (IDB), IDB Invest, & Finnovista. (2018). *Report on Fintech in Latin America 2018: Growth and Consolidation*. https://publications.iadb.org/en/fintech-latin-america-2018-growth-and-consolidation

[2] https://limafintechforum.com

[3] ASBANC. (2017). *Una mirada al fenómeno fintech en el Perú y el mundo*. https://www.asbanc.com.pe/publicaciones/asbanc-semanal-242.pdf

[4] When the term "fintech" is used to search the National Repository of Research Papers (RENATI), there are only eight matches: renati.sunedu.gob.pe/simple-search?query=fintech

biases of its creators. This "contamination" can be conscious or unconscious and occur in different stages. It includes definitions of concepts such as "beauty" or "creditworthiness" under which an algorithm will look for patterns that allow it to assign values to certain attributes such as age, sex, address, etc., that help the AI to make decisions.[5]

What is meant by discrimination in Peru?

Discrimination, in its broadest sense, is an act of differentiating between individuals or groups of people based on attributes of subjective evaluation whose result is the benefit of one individual or group (better valued) at the expense of another (less valued). As a social phenomenon, discrimination has been widely studied and its traces can be found throughout Peruvian history. In recent years, discrimination has been understood as a negative action that undermines people's rights to equality and that must therefore be discouraged through laws.

In the Peruvian constitution of 1993, currently in force, discrimination of any kind is prohibited.[6] Over the last 26 years, different laws have been developed in order to reinforce this prohibition, covering almost all areas of social life. For example, performing an act considered discriminatory constitutes a crime that can currently be punished by up to four years in prison. Likewise, this act may be subject to an administrative sanction in some jurisdictions, which may involve the imposition of economic penalties or even the closure of the establishment where the event occurred. Discrimination can also be subject to sanctions in the field of work, education and even at the level of consumer relations. In all these scenarios, judges and administrative entities have over the years made different rulings and decisions that have helped define the scope of the laws and codes of conduct. Some cases are very relevant to this report and will be presented later.

Why will the fintech sector make a good basis for further research?

In terms of technological ventures that have started up in Peru in the last five years, fintech companies are positioned in a particularly interesting place for three reasons. The first is their intensive use of all types of technologies, including AI tools. The second is that more than half of the local fintech ventures offer B2C (business-to-consumer) services, which means that they have a direct impact on the users of financial services. The third is the dominant narrative that affirms that these companies will contribute to the greater financial inclusion of vulnerable and historically excluded groups. These make fintech ventures an obvious choice if we want to research whether the use of AI generates or reproduces situations of discrimination against Peruvians.

Although they have in common that they all use technology, fintech ventures can be sub-categorised according to the type of services they offer. The most popular at the regional level are those that provide payment and remittance services, those that offer loans and those offering personal or corporate finance management. Fintech ventures that grant loans can serve as a useful reference for study because it is widely recognised that they use AI tools in their business models, compared to others focused on activities such as credit scores or online payments that do not necessarily involve the use of AI.[7] In these "lending fintechs", AI usually interacts directly with customers through application forms, which allows it to collect, contrast and analyse the data received to make decisions. Depending on how the algorithms have been designed or how the AI has been trained, these decisions can exhibit acts of discrimination and, in certain cases, could be sanctioned.

Is it illegal (or even possible) for an algorithm to discriminate?

While discrimination of any kind is prohibited in Peru, there are cases in which a person can be "discriminated" against legally. For example, a restaurant that has a policy of not serving people who are in shorts and flip-flops may refuse to serve a customer that dresses in that way without discriminating against him or her. Another case, closer to the topic of this report, is a person who asks for bank loans and never bothers to pay them, which generates a negative credit record. If a new bank receives this person's loan application, they can reject it based on the negative record and not commit an act of discrimination even though the person has never been a client of the bank.

[5] 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-happensand-why-its-so-hard-to-fix

[6] Political Constitution of Peru, Article 2: Every person has the right: (...) To equality before the law. No person shall be discriminated against on the basis of origin, race, sex, language, religion, opinion, economic status, or any other distinguishing feature.

[7] Schueffel, P. (2016). Taming the Beast: A Scientific Definition of Fintech. *Journal of Innovation Management, 4*(4), 32-54. https://journals.fe.up.pt/index.php/IJMAI/article/view/2183-0606_004.004_0004/221

These two examples are almost common sense. The freedom of contract is a concept that determines that two parties can agree their own rules when contracting. In Peru, freedom of contract is a right recognised in the constitution. If a bank can establish a policy to not give loans to clients with a history of defaulting, it seems logical that an algorithm can perform the same discriminatory behaviour and not break any law.

But what is the limit? Below are two cases that, with 15 years of separation between each, seem to answer this question.

In 1999, a woman requested a credit card from Ripley (a popular department store), for which she delivered the required documentation consisting of her personal data, a copy of her identification document, and the credit card of another bank. Two weeks later, the woman received the news that the credit card request had been denied and when she asked why, she was told that the district in which she lived (La Victoria) was "not verifiable", which meant that it was difficult to carry out actions of verification and debt collection in La Victoria, which was labelled as "dangerous". The woman decided to report Ripley for discrimination before the Peruvian consumer protection authority (INDECOPI), arguing that the store had only considered where she lived and not her credit record or income. In the first decision by INDECOPI, the complaint was dismissed. However, on appeal, the complaint was accepted based on an interpretation of discrimination currently in force in law.

The rationale of the case was that there are two forms of treatment in consumer relations: differentiated treatment and discriminatory treatment. Differentiated treatment is legal and is based on the existence of objective and justified causes that allow restricting and even denying the provision of services or products to a consumer. On the contrary, discriminatory treatment is based on purely subjective and arbitrary reasons, hence it is illegal. In this case, when it comes to credit, having a restrictive policy for customers that live in certain places can be an objective cause for differentiated treatment, but it should not be the only one. If it was the only one, it can become discriminatory, as has happened in this case.[8]

Fourteen years later, in 2013, a retired woman filed a lawsuit before the Constitutional Court against Banco de la Nacion (the state bank) for a similar reason. The bank had denied her a loan because she was 85 years old and its policy was to only grant loans to clients up to 83 years of age. The woman claimed that this was a violation of her right to equality and non-discrimination, while the bank argued that it was a case of differentiated treatment based on objective causes. To resolve this case, the court used an argument similar to that of INDECOPI. Based on international treaties and local legislation, it determined that age could not be the only requirement to deny access to credit and that doing so constituted discriminatory treatment. In addition, they pointed out that the vulnerable quality of the applicant, who was a senior citizen, had to be considered.[9]

So if an AI decides not to grant a loan to a client solely and exclusively because of his or her age, address, gender, etc. without considering other factors, is it committing an act of discrimination? Using the doctrine of the type of treatment (differentiated or discriminatory), the answer seems to be affirmative. Although in the cases cited the discriminatory behaviour was carried out by people, there is no reason why this doctrine could not be applied to machine-learning systems that have algorithms for decision making. However, while it is possible to recognise discrimination when interacting with a person, it is not so simple when it is the result of AI, because the way these tools work is not usually transparent or self-explanatory to those who interact with it.

Given that these systems collect and process huge amounts of data, identifying how and when a discriminatory act occurs seems an impossible task. For instance, a fintech service provider could have configured its algorithms to deny loans to all women, young people in rural areas or anyone with a surname of indigenous origin and it would be very difficult to prove that this has happened. Perhaps an indicative test could be the terms and conditions that appear on the websites of these companies and contain discriminatory clauses, which would help predict the discriminatory behaviour of the AI. At least two fintech companies that we looked at showed discrimination in terms of age, but it is likely that others also have built-in discriminations, without making these public.[10]

[8] See Final Resolution Nº 747-2000-CPC and Final Resolution Nº 0517-2001/TDC-INDECOPI for the case file Nº 307-1999-CPC. https://www.scribd.com/document/344636065/0517-2001-1-Cecilia-Reynosa-Contra-Ripley-y-Recaudadora-Discriminacion

[9] See the final ruling of the Constitutional Court for case file N° 05157 2014-PA/TC. http://tc.gob.pe/jurisprudencia/2017/05157-2014-AA.pdf

[10] The Fintech companies who carry out discrimination based on age are Fio (https://www.fio.pe) and Tappoyo (https://www.tappoyo.com). In the case of Fio, in the frequently asked questions section of its website, it is indicated that applicants must be between the ages of 21 and 65. In the case of Tappoyo, the restriction is between 20 and 70 years of age.

An important aspect that should also be part of this reasoning is how fintech ventures are different from other actors in the financial system in terms of their obligations to not discriminate. Could they argue that their need to increase restrictions on consumers is greater because they are more vulnerable and face greater risks? There is still plenty of room to debate.

Conclusion

It is clear that the fintech sector offers useful opportunities to understand the application of AI in Peru, and its potential impact on citizen rights. In particular, we suggest above that there is scope to explore potential bias and discrimination in the algorithms used by companies operating in the sector, including those that offer loans to vulnerable and disadvantaged groups. We have outlined the difficulties that are likely to be faced by such an analysis, including a lack of transparency in the algorithms used, and that there might be cause to allow a greater degree of differentiation in the sector given its vulnerability. At times this may be considered discriminatory. We have also highlighted, however, how legislation governing discrimination in the country, which is well defined, should be brought to bear on the business practices of fintech companies. Below we suggest some ways in which civil society can engage with the sector in order to strengthen its positive benefits for providing financial services to the vulnerable and marginalised in Peru.

Action steps

The following recommendations can be made for civil society organisations in Peru:

- Peruvian civil society organisations should begin to participate in the spaces related to the use and development of AI in order to acquire skills and capacities that allow them to form an opinion from the perspective of human rights.

- They must demand that ethical practices in the use of AI are incorporated into the development of business models in the local fintech ecosystem, especially in terms of transparency in the use of AI tools.

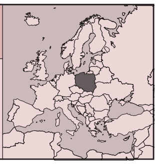
- Civil society organisations must demand that the Peruvian government adhere to international guidelines for the creation of public policies on AI that promote the development of the human being.

- They should encourage universities to conduct in-depth studies on the use of AI in the economic, social and regulatory fields, among others.

- Civil society organisations should support the financial inclusion of vulnerable populations through fintech and other technological ventures, but with respect for human rights, especially in terms of preventing discrimination.

POLAND

ALGORITHMS DECIDING THE FUTURE OF SOCIAL RIGHTS: SOME EARLY LESSONS FROM POLAND



Jedrzej Niklas
jedrzej.niklas@gmail.com

Introduction

Governments all around the world are looking toward artificial intelligence (AI) and other automated systems as an attractive solution for many complicated social problems. One of the significant promises of these new technological developments is the increase of efficiency and cost-effectiveness of public administration. Automated systems and algorithms are already being used to determine health insurance, check eligibility for welfare benefits or detect potential fraud.[1] These technologies become a vital element in procedures that have a significant effect on the enjoyment of social rights. At the same time, they also create many problems for transparency and accountability and can amplify existing inequalities.

This report will try to illustrate some of these problems by analysing already existing automated systems and algorithms used by the Polish welfare administration. While these systems are not very sophisticated, they can provide some early lessons for data-intensive practices and their impact on people's rights and needs. Because of their technological and mathematical nature, the systems are very often portrayed as objective and apolitical, while they in fact have significant social justice consequences. What we learn from them can be crucial for the discussion about more advanced technologies, including AI, and their human rights implications.

Background: E-government imperatives, complex regulation and austerity

In 2018 the Ministry of Digital Affairs published a document that serves as a base for a future national AI strategy in Poland.[2] The report focuses on investments in research, cooperation between business and universities and potentially AI companies, and, most notably, public administration. While the Ministry also included some ethical and human rights concerns (mostly concerning privacy and non-discrimination), the primary narrative stressed AI's impact on organisational efficiency, innovation and economic growth. In comparison to other countries, the Polish approach is rather typical and sees AI as a strategic technology for state and business operations.[3]

However, the hype around AI is not only driven by government plans and policies. Big international corporations – like Microsoft, IBM or SAS[4] – are already offering sophisticated machine learning and analytical tools that can be used by public agencies, including welfare administration.[5] Among those innovations are technologies designed to facilitate client services (e.g. verification of eligibility for a service or services tailored to the individual needs of citizens) and for planning and policy-related purposes (such as doing cost-effective analyses).[6] This is, of course, not a new phenomenon. As part of the e-government agenda, many similar systems have been around for decades.[7] Over the years, Poland has been investing a lot in the digitalisation of public services, mostly thanks to funds from the European Union (EU).[8]

These technologies fall under a complicated and fragmented regulatory regime. From the perspective of citizens' rights, the most significant laws

[1] See, for example: Spielkamp, M. (Ed.) (2018). *Automating Society: Taking Stock of Automated Decision Making in the EU*. www.algorithmwatch.org/wp-content/uploads/2019/01/Automating_Society_Report_2019.pdf

[2] Ministerstwo Cyfryzacji. (2018). *Założenia do strategii AI w Polsce*. Warszawa. https://www.gov.pl/documents/31305/436699/Za%C5%82o%C5%BCenia_do_strategii_AI_w_Polsce_-_raport.pdf

[3] Dutton, T. (2018, 28 June). An Overview of National AI Strategies. *Medium*. https://www.medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

[4] https://www.sas.com. Considers itself a leader in analytics.

[5] For example, in 2013, IBM was lobbying the city of Lodz (in central Poland) to optimise the governance of social assistance programmes by using data analytics. In: www.archiwum.uml.lodz.pl/get.php?id=12361

[6] See, for example: IBM. (n.d.). *Government Health & Human Services Solutions*. www.ibm.com/downloads/cas/BAWQZPL2; SAS. (n.d.). *SAS provides the DWP with powerful predictive insights in highly complex policy areas*. https://www.sas.com/en_gb/customers/dwp.html

[7] In the 1980s, many countries developed so-called expert systems that were an early version of AI. See: Weintraub, J. (1989). Expert Systems in Government Administration. *AI Magazine*, *10*(1). https://pdfs.semanticscholar.org/4c36/058a0d5bde0c1c13db139e9d30246ee5c0dc.pdf

[8] European Commission. (2016). *E-Government in Poland*. https://joinup.ec.europa.eu/sites/default/files/inline-files/eGovernment_Poland_June_2016_v4_01.pdf

are the administrative law, the data protection law (the EU's General Data Protection Regulation and its national implementation) and numerous laws and regulations related to social benefits, health care and other public services. The mix of rules creates an uncertain and complex framework that sets how public administration makes decisions about services, its use of digital systems in this process, how citizens may apply for benefits, and what kind of safeguards and oversight systems are in place.

Given these uncertainties and complexities, the discussion around AI shows that at least at the EU level, there is a growing consensus for the separate regulation of automated systems that will address at least the problem of liability and the opacity of such systems.[9]

Beside these legal problems, there is also a growing international discussion on how such computerised systems can amplify existing inequalities and create concerns for social justice. Some researchers prove that the deployment of such systems in welfare can create harms for people experiencing poverty, and other vulnerable populations.[10] Very often these technologies are justified by austerity policies and cost-reduction strategies, and therefore directly affect the enjoyment of certain social rights.

Early lessons from the datafication and algorithmisation of welfare

Healthcare insurance verification: Errors and limitations to safeguards

In 2018 the Polish press reported on the case of a migrant woman of Romani origin who was denied medical service for her ill daughter.[11] The incident, which quickly escalated into conflict when the police intervened, was caused by anti-Romani sentiment and the discriminatory attitude of medical personnel. However, the direct reason for denying the service was an error in an automated system which checks eligibility for health care insurance. The Polish health care system is based on a public insurance scheme; however, a significant segment of the population (around two million people) is excluded from it. This group includes mostly undeclared workers, freelancers, migrants and the homeless.[12]

Introduced in 2013, e-WUŚ (*elektroniczny system weryfikacji uprawnień świadczeniobiorców* in Polish) is used before each visit to the doctor or hospital.[13] The system uses data shared between the National Health Fund and the Polish Social Insurance Institution to verify insurance status automatically. Due to inaccurate, erroneous or outdated data, e-WUŚ has in the past made thousands of mistakes, creating chaos in Polish health care.[14] To reduce the scale of this problem, the government introduced some special safeguards and procedures.[15] If the system indicates that a person does not have insurance, she can still visit a doctor but under certain conditions. She has to write a statement expressing disagreement with the system's verification result, and in two weeks provide the necessary evidence indicating that she should be covered (usually pay slips).

While this procedure works well in most cases, it can create some problems for vulnerable populations like migrants or the homeless. To verify insurance statuses, e-WUŚ uses a national ID number. Some migrants do not have this, even if their status is fully regulated. Similar problems face homeless people who lack either official addresses or documentation.[16] These situations create bureaucratic restrictions to the enjoyment of the universal right to health. Some also argue that developing and maintaining the system itself costs more than just expanding health insurance to the whole population.[17]

[9] Stolton, S. (2019, 27 June). 'Adverse impacts' of Artificial Intelligence could pave way for regulation, EU report says. *Euractiv.com.* https://www.euractiv.com/section/digital/news/adverse-impacts-of-artificial-intelligence-could-pave-way-for-regulation-eu-report-says

[10] Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor.* New York: St. Martin's Press; Gilliom, J. (2001). *Overseers of the Poor: Surveillance, Resistance, and the Limits of Privacy.* Chicago: University of Chicago Press.

[11] Lehmann, A. (2018, 2 March). Lekarka wyrzuciła za drzwi romską dziewczynkę i jej matkę, bo nie miały dokumentów. *Gazeta Wyborcza.* www.poznan.wyborcza.pl/poznan/7,36001,23092992,lekarka-wyrzucila-za-drzwi-romska-dziewczynke-i-jej-matke-bo.html

[12] Komuda, L. (2016, 5 December). Fakturka za uratowanie życia. 2,5 miliona Polek i Polaków nie ma ubezpieczenia NFZ. *Oko press.* www.oko.press/fakturka-uratowanie-zycia-25-miliona-polek-polakow-ubezpieczenia-nfz

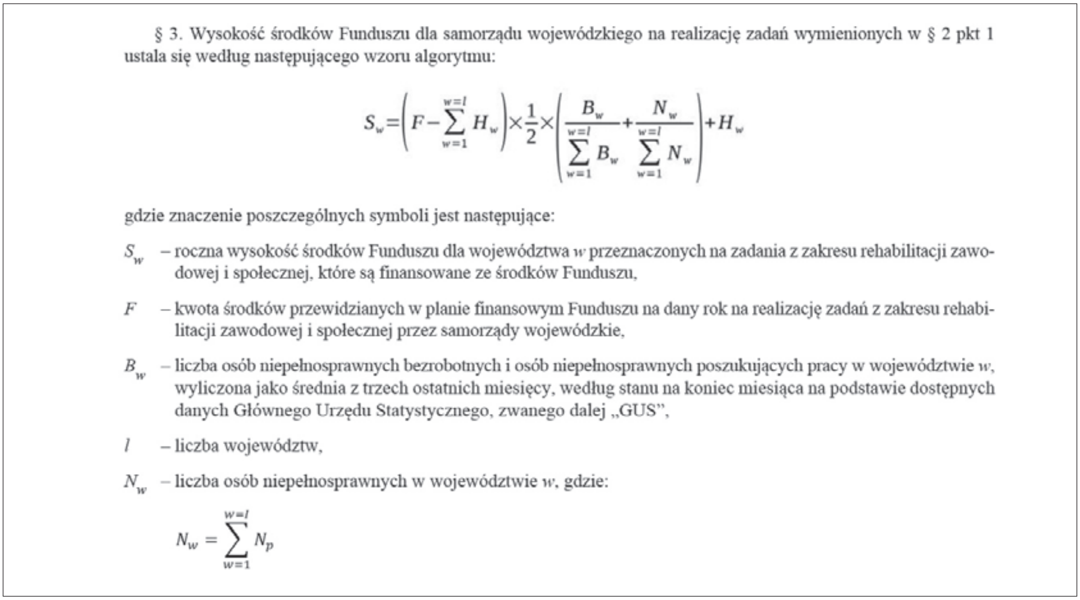
[13] Narodowy Fundusz Zdrowia. (n.d). *eWUŚ - Elektroniczna Weryfikacja Uprawnień Świadczeniobiorców.* www.nfz-warszawa.pl/dla-pacjenta/ewus-elektroniczna-weryfikacja-uprawnien-swiadczeniobiorcow

[14] Cichocka, E. (2013, 23 September). System eWUŚ pozbawia Polaków bezpłatnego leczenia. *Gazeta Wyborcza.* www.wyborcza.pl/1,76842,14651136,System_eWUS_pozbawia_Polakow_bezplatnego_leczenia.html

[15] Narodowy Fundusz Zdrowia. (n.d.). *Co zrobić, gdy eWUŚ wyświetli nas "na czerwono".* www.nfz.gov.pl/dla-pacjenta/zalatw-sprawe-krok-po-kroku/co-zrobic-gdy-ewus-wyswietli-nas-na-czerwono

[16] Gangadharan, S., & Niklas, J. (2018). *Between Antidiscrimination and Data: Understanding human rights discourse on automated discrimination in Europe.* London: London School of Economics. https://eprints.lse.ac.uk/88053/13/Gangadharan_Between-antidiscrimination_Published.pdf

[17] Nyczaj, K. (2016, 2 March). Co dalej z eWUŚ? *Medexpress.pl.* https://www.medexpress.pl/co-dalej-z-ewus/63134



§ 3. Wysokość środków Funduszu dla samorządu województwa na realizację zadań wymienionych w § 2 pkt 1 ustala się według następującego wzoru algorytmu:

$$S_w = \left(F - \sum_{w=1}^{w=l} H_w \right) \times \frac{1}{2} \times \left(\frac{B_w}{\sum_{w=1}^{w=l} B_w} + \frac{N_w}{\sum_{w=1}^{w=l} N_w} \right) + H_w$$

gdzie znaczenie poszczególnych symboli jest następujące:

S_w – roczna wysokość środków Funduszu dla województwa w przeznaczonych na zadania z zakresu rehabilitacji zawodowej i społecznej, które są finansowane ze środków Funduszu,

F – kwota środków przewidzianych w planie finansowym Funduszu na dany rok na realizację zadań z zakresu rehabilitacji zawodowej i społecznej przez samorządy wojewódzkie,

B_w – liczba osób niepełnosprawnych bezrobotnych i osób niepełnosprawnych poszukujących pracy w województwie w, wyliczona jako średnia z trzech ostatnich miesięcy, według stanu na koniec miesiąca na podstawie dostępnych danych Głównego Urzędu Statystycznego, zwanego dalej „GUS”,

l – liczba województw,

N_w – liczba osób niepełnosprawnych w województwie w, gdzie:

$$N_w = \sum_{w=1}^{w=l} N_p$$

New challenges for civil society: Example of mathematical formula for algorithm used for allocating resources in Poland, which is included in legislation.

Profiling the unemployed: Assumptions of practice and a successful human rights intervention

Between 2014 and 2019, Polish job centres were using an automated decision-making system to categorise unemployed people and allocate them different types of assistance automatically.[18] The so-called profiling mechanism used demographic information and data collected during a computer-based interview conducted by frontline staff. Each data entry was assigned a score from 0 to 8. Based on the final calculation, the algorithm decided on the profile of the individual, and, as a result, determined the scope of assistance a person can apply for. The system divided all unemployed into three profiles, which differed from each other in terms of demographics, distance from the labour market and the chance for re-employment. The main reasons for this technology were to reduce costs, increase efficiency and offer greater individualisation of services.

However, the profiling mechanism caused numerous controversies. For example, citizens applying for assistance did not know how the system worked: what the scope of input data was (e.g. what kind of demographic data was taken into account), what answers to the interview questions were chosen by the frontline staff, how information was processed and how scores were calculated. At the same time, the outcomes of the algorithmic calculations could be severe. In many cases, the profile assigned to the person limited her access to specific types of assistance. This was mostly visible for the so-called “third-profiled” who could only apply for a very limited set of assistance. There were also accusations of discrimination against some groups (single mothers, people with disabilities or rural residents). Many unemployed persons articulated their dissatisfaction with the profiling mechanism and even tried to game the system or submit formal complaints.[19] In addition, according to an official evaluation of the system, staff at the job centres were also unhappy with the categorising tool: 44% of them said that it was useless in helping them with their everyday work.[20] In practice, they used the system in different ways. For many, the computer was an ultimate decision maker. For others,

[18] Niklas, J. (2019, 16 April). Poland: Government to scrap controversial unemployment scoring system. *Algorithm Watch*. www.algorithmwatch.org/en/story/poland-government-to-scrap-controversial-unemployment-scoring-system

[19] Niklas, J., Sztandar-Sztanderska, K., & Szymielewicz, K. (2015). *Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making*. Warszawa: Fundacja Panoptykon. www.panoptykon.org/sites/default/files/leadimage-biblioteka/panoptykon_profiling_report_final.pdf

[20] Ministerstwo Rodziny, Pracy i Polityki Społecznej. (2019). *Analiza Rozwiązań Wprowadzonych Ustawą z Dnia 14 Marca 2014 r. o Zmianie Ustawy o Promocji Zatrudnienia i Instytucjach Rynku Pracy oraz Niektórych Innych Ustaw.*

the profiling was just a part of the broader individual assessment process of an unemployed person. Sometimes they even adjusted the profile to meet the expectations of the unemployed person.[21] These examples show that in practice, the use of automated technologies depends on organisational culture, competencies and individual preferences. Designers' intentions may be significantly different from the actual use of technology. The level of automation results not only from initial assumptions but, above all, from the practice of users.

The profiling mechanism was also heavily criticised by civil society and human rights institutions (the Personal Data Protection Office and Human Rights Commissioner). The Panoptykon Foundation, the country's leading digital rights organisation, launched a long and successful campaign against the system.[22] Raising concerns about transparency, anti-discrimination and privacy, activists convinced the Polish Human Rights Commissioner to refer the profiling case to Poland's Constitutional Court. In 2018, the Court ruled that the bill which introduced a profiling mechanism violated the Polish constitution, and as a consequence, the government decided to stop using the system a year later.[23]

Detecting welfare fraud: The depoliticisation and opacity of the digital system

Another example of a system used by the Polish welfare administration is a complicated mix of different databases with automated functions that help to detect fraud among welfare recipients. At the core of this mechanism is a system called the Central Base of Beneficiaries (CBB), which contains millions of data records of people receiving welfare benefits.[24] During the application process, the system allows the official to check if a person is receiving the same or a similar benefit in another commune. Social workers may also run a query within other databases to verify the applicant's employment history, taxation, etc. This automated data analysis can indicate that a person is not eligible for certain benefits or be a sign of a potential fraud.[25]

The CBB was introduced as a part of a larger project of digitalisation of the welfare sector called "Empatia" (or Empathy in English).[26]

While the introduction of automated systems in welfare is necessary, especially when big social programmes are in place, the initiative created significant problems from the perspective of transparency. The mechanism of detecting fraud was designed and expanded without any social discussion and sometimes without clear legal bases. Data-sharing agreements between public agencies were seen as a technical issue, and not a political problem that can determine individuals' social rights. Additionally, while there was no information about errors or harms caused by the system, many social workers said they were frustrated and explained that the system created problems in their daily operations.[27] The use of it is time consuming and the procedures are not always easy to follow. This raises another question: To what extent does the introduction of such a system create greater efficiency in the work of social workers, allowing them to focus on their actual job, which is helping people in need?

Algorithm for allocating resources: Low-quality data and the need for expert advocacy

The Polish welfare administration is also using non-automated models and algorithms for the allocation of crucial resources. One such example is the mathematical formula used by the State Fund for Rehabilitation of Disabled People (SFRDP).[28] It allows the distribution of vital financial resources for rehabilitation and assistance for people with disabilities. The algorithm regulates the allocation of resources to local governments, and it is described in detail in law. It uses such data as the number of people with disabilities, the number of children with disabilities and unemployment statistics. While the mechanism is supposed to be objective and technical, it creates a lot of controversies and has been contested.

[21] Sejm. (2019). *Projekt zmiany ustawy o promocji zatrudnienia I instytucjach rynku pracy*. https://orka.sejm.gov.pl/Druki8ka.nsf/0/07BB9C4DDB659D71C12583D10069F1B1/%24File/3363.pdf

[22] Niklas, J., Sztandar-Sztanderska, K., & Szymielewicz, K. (2015). Op. cit.

[23] Trybunal Konstytucyjny. (2018). *Zarządzanie pomocą kierowaną do osób bezrobotnych*. www.trybunal.gov.pl/postepowanie-i-orzeczenia/komunikaty-prasowe/komunikaty-po/art/10168-zarzadzanie-pomoca-kierowana-do-osob-bezrobotnych

[24] Najwyższa Izba Kontroli. (2016). *Realizacja i Wdrażanie Projektu Emp@tia Niklas*. https://www.nik.gov.pl/plik/id,11506,vp,13856.pdf; Niklas, J. (2014, 18 April). Ubodzy w prywatność. *Fundacja Panoptykon*. www.panoptykon.org/wiadomosc/ubodzy-w-prywatnosc

[25] Ministerstwo Rodziny, Pracy i Polityki Spolecznej. (2018). Pojedyncze Usługi Wymiany Informacji udostępnione lub planowane w ramach Centralnego Systemu Informatycznego Zabezpieczenia Społecznego (CSIZS). https://empatia.mpips.gov.pl/documents/10180/1185925/2019-02-01+us%C5%82ugi+pojedyncze+CSIZS.pptx/73dd7dee-fff7-41a8-8145-44eb9f942a55;jsessionid=ba791fc14c89eda480e921ffeb46?version=1.0

[26] https://empatia.mpips.gov.pl

[27] Web forum of social workers. www.public.sygnity.pl/forums/viewforum.php?f=73&sid=058ae5fe15fc8a89c34c3f93a58b0c5d

[28] Malinowska-Misiąg, E., et al. (2016). *Algorytmy Podziału Środków Publicznych*. www.ibaf.edu.pl/plik.php?id=598

One of the biggest problems is the quality and accuracy of data used in the allocation procedure. The sources of that information are the national census from 2011 and administrative databases. According to organisations that fight for the rights of people with disabilities and representatives of local governments, this data is misrepresenting (due to methodological problems) the population with disabilities.[29] Because of this misrepresentation, some local governments received inadequate funds and as a consequence, many people with disabilities were left without necessary assistance. This problem primarily affected the education and support for children with disabilities. In 2018, civil society organisations were able to successfully advocate for some changes to the algorithm, resulting in more significant resources being allocated for specific types of support.[30] However, some of the most pressing problems (e.g. the methodology of the census) remain unresolved. While this case is not an example of automation, it shows that low-quality data and use of models can cause problems for crucial decision making about social rights. It also demonstrates that for civil society organisations to successfully advocate for their interests, they must engage in the technical language of algorithms and mathematical formulas.

Conclusion

Existing examples of automated systems and mathematical models used in welfare can provide some valuable lessons for implementing and problematising more advanced technologies like AI. One of them is related to the political nature of digital technologies and algorithms. While very often portrayed as objective and technocratic and as a result the sole realm of technical experts, the systems (their design, architecture and targets) are in fact deeply political in that they create social constructs and play a crucial role in a decision-making process that affects thousands of individuals in the allocation of resources. The use and design of automated systems is a result of individual choices about policies, priorities and cultural norms. Therefore their deployment and implementation should be subject to democratic control. However, as was shown in some of the Polish examples, this is not always an easy task.

Understanding the impact of these systems is difficult due to the complexity of the technological layers that demands some specific expertise. In such situations, activists and human rights institutions need to learn new skills and engage with a different language, concepts and communities. For example, in the case of the algorithm for allocating resources, activists had to propose specific changes to the law using the complicated mathematical formulas in the law. Therefore there is a need to look for new ways of articulating social justice *vis-à-vis* automated systems and algorithms. Privacy and data protection remain central frames in this context. However, they have limitations – they focus on quite narrowly understood informational harms, and very often ignore collective injustices created by computer systems. The application of a social rights lens can extend the discussion around automated systems and create some necessary connections between social status, discrimination, inequity and the use of technology and its outcome. A social rights framework involves procedural elements (participation in creating policies or transparency in the individual decision-making process) and substantive considerations (access to some specific set of social services). Thanks to this framework, it is easier to position AI as a political and social justice issue.

There is also a space for more radical political advocacy that would not only engage in changes or improvements to algorithms, but also call for the abolition of specific systems that cause harm. The campaign against the mechanism for profiling the unemployed was a great example of when human rights, social justice and the rule of law helped to determine which processes could be automated and which should not, and under what conditions.[31]

It is also important to acknowledge that many of those technologies function in very complex organisational and institutional environments. The use of them depends on different organisational cultures, individual motivation, conflicts between institutions and more. As indicated in the profiling case, frontline staff can, for example, use systems in a different way to that intended. Understanding this environment can be very helpful in any campaign related to technologies used in the welfare administration, or any government service-orientated institution.

[29] Związek Powiatów Polskich. (2016). *Stanowisko w sprawie koniecznych zmian w zakresie dysponowania przez samorządy środkami PFRON*. www.zpp.pl/storage/library/2017-06/adedc562e4734b57e47ed71d7235693c.pdf

[30] Sejm. (2018). Interpelacja nr 22410 w sprawie apelu skierowanego przez pracowników warsztatów terapii zajęciowej z województwa małopolskiego. www.sejm.gov.pl/sejm8.nsf/InterpelacjaTresc.xsp?key=5304F8FD

[31] Fundacja Panoptykon. (2019, 14 April). Nieudany eksperyment z profilowaniem bezrobotnych właśnie przechodzi do historii. *Fundacja Panoptykon.* https://panoptykon.org/wiadomosc/nieudany-eksperyment-z-profilowaniem-bezrobotnych-wlasnie-przechodzi-do-historii

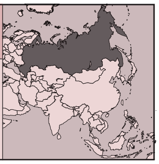
Action points

The following advocacy steps are suggested for civil society in Poland:

- It is crucial that AI and other new technological innovations are examined from the perspective of social justice and inequalities and address the needs and struggles of marginalised communities. The debate around these systems should focus on their consequences and not efficiencies; people instead of technical details about automation.
- Learn from what is already being implemented. There is a range of existing technologies and analytical models used by the welfare administration. Organisations that try to engage in the debate about the use of AI can learn from the successes and pitfalls of these models.
- There is an emerging need to connect different advocacy strategies. In the case of welfare technologies, digital rights activists should join forces with anti-poverty and anti-discrimination organisations and groups that have a greater connection with affected communities.
- Human rights advocacy should engage a plurality of claims that combine, for example, privacy, anti-discrimination and social rights. Activists should conceptualise and advocate for new ways of political and democratic control and supervision over technologies used in sensitive areas like welfare.

RUSSIA

THE FUTURE IS NOW: RUSSIA AND THE MOVEMENT TO END KILLER ROBOTS



J. Chua

Introduction

Of all the contentious issues regarding misuses of artificial intelligence (AI), the most frightening might be the use of AI in weapons. In films such as the *Terminator* series or *2001: A Space Odyssey*, malevolent machines in the form of Skynet or HAL the computer can think for themselves, and aim to control or eliminate humans. But what might seem like something from a science fiction movie is already a reality: weapons now can function using AI to kill without active human control.

Countries including the United States (US), South Korea and Russia are investing in AI technologies for use in lethal autonomous weapons systems (LAWS), also dubbed "killer robots". LAWS are distinguished from other forms of AI-enabled warfare, such as cyberwars, which are not directly lethal. The concept of autonomous weapons is also not new. The landmine is an early example of an autonomous weapon, a device that is triggered autonomously and kills without active human intervention. But the use of AI brings such weapon systems to an entire new level, for it allows machines to independently search, target and/or eliminate perceived enemies.

The US has already developed weapons such as AI-enabled drones and tanks, as well as the Sea Hunter, an autonomous war ship. So far, it has placed humans in the loop for these machines, meaning there is no firing of weapons for lethal purposes without direct human intervention. However, these weapons systems have the capacity to be operated by AI alone.[1] The US spends far more than any other country annually on weapons research and development, to the tune of USD 100 billion on "everything from hypersonic weapons to robotic vehicles to quantum computing."[2]

The Kremlin has announced AI as a priority in Russia.[3] Although at a rate far lower than the US, Russia is investing heavily in AI with the view of developing AI-driven defence capabilities. This investment has included AI-enabled missiles with the capability to change their target mid-flight, and AI-assisted tanks, although they are currently not fully autonomous yet.[4] Russia has even invited foreign investors to fund research and development in AI, raising USD 2 billion to support domestic tech companies.[5]

Russian activism against LAWS

This misuse of AI technology represents a clear danger to humanity on many fronts. Algorithms cannot as yet make perfect decisions, especially in varying warfare conditions. For example, conventional soldiers (i.e. humans) might not kill a child holding a toy weapon, but a robotic device might not be able to distinguish between a child and an adult soldier. The chance of a software glitch accidentally firing upon friendly soldiers or civilians is not just a remote possibility. This has happened before, such as a robotic cannon mistakenly killing nine soldiers and wounding 14 others during a military exercise in South Africa over a decade ago.[6]

There is currently a global campaign against weapons controlled by AI. Ban Killer Robots started as a project of Human Rights Watch, but now has chapters and affiliates in countries around the world.[7] The Russian affiliate is the NGO Ethics and Technology, established and managed by Alena Povova, Russia's primary campaigner against LAWS.[8] She recently attended a Convention on Cer-

[1] Piper, K. (2019, 21 June). Death by algorithm: the age of killer robots is closer than you think. *Vox*. https://www.vox.com/2019/6/21/18691459/killer-robots-lethal-autonomous-weapons-ai-war

[2] Thompson, L. (2019, 8 March). Pentagon May Come To Regret Prioritizing R&D Spending Over Weapons It Needs Now. *Forbes*. https://www.forbes.com/sites/lorenthompson/2019/03/08/weapons-budgets-are-way-up-so-why-isnt-the-pentagon-buying-weapons-faster/#29b7c2a84263

[3] Daws, R. (2019, 31 May). Putin outlines Russia's national AI strategy priorities. *AI News*. https://artificialintelligence-news.com/2019/05/31/putin-russia-national-ai-strategy-priorities

[4] RT. (2018, 5 May). Race of the war machines: Russian battlefield robots rise to the challenge. *RT*. https://www.rt.com/news/425902-war-machines-russian-robots

[5] bne IntelliNews. (2019, 31 May). Russia Raises $2Bln for Investment in Artificial Intelligence. *The Moscow Times*. https://www.themoscowtimes.com/2019/05/31/russia-raises-2bln-for-investment-in-artificial-intelligence-a65824

[6] Shachtman, N. (2007, 18 October). Robot Cannon Kills 9, Wounds 14. *Wired*. https://www.wired.com/2007/10/robot-cannon-ki

[7] https://www.stopkillerrobots.org/members

[8] www.ethicsandtech.ru/en

tain Conventional Weapons (CCW) meeting[9] at the UN in Geneva (20-21 August 2019) for formal and informal consultations as an expert. One of her main missions is to raise awareness of the situation among ordinary Russians. To this end she promotes the anti-LAWS cause on both Russian and English social media and by blogging on the issue. Povova has spent many years working with the Russian parliament, serving as an advisor and expert. Her NGO takes a five-stage approach in the campaign against killer robots. Firstly, it begins by collecting information, finding available information and creating a relevant database. Secondly, it processes the information, preparing data for its working group and experts. Thirdly, it develops solutions via open discussions, with the aim to eventually turn them into formalised legal acts. Fourthly, it informs the public via social media, news releases, press conferences, etc. Fifthly, it persuades and exerts public pressure on the legal system to adopt laws and measures to implement the solutions, and keep track of the results.[10]

At the CCW meeting in Geneva, she appeared optimistic about the eventual success of the campaign. Issuing a video statement during the CCW meeting, she told her Russian audience that in fact the sci-fi scenario of algorithmic weapons is already a reality. She appealed strongly to the scientists in the Russian Academy of Sciences – indeed scientists everywhere – to join the fight against killer robots. She asked all technology workers to join the movement in declaring they will not support the development of LAWS. She also urged Russians to be a part of this movement before it is too late to stop this nightmare.[11]

To this end, she also connects with other Russian-speaking campaigners from Commonwealth of Independent States (CIS) countries, including the prominent Kazakh activist Alimzhan Akhmetov, the director of the Center for International Security and Policy in Nur Sultan.[12] While Kazakhstan itself is not developing AI-enabled weapons, activists understand there must be regional cooperation to stop Russia from implementing such weapons. Akhmetov takes a moral and legal approach in trying to ban LAWS. He cites the Martens Clause,[13] first introduced in the 1899 Hague Convention II – Law and Customs of War on Land and added to the Geneva Convention in the additional protocols of 1977:

> In cases not included in the present Protocol or other international treaties, civilians and combatants remain under the protection and the rule of the law of nations, as they result from the usages established among civilized peoples, from the laws of humanity and the dictates of public conscience.[14]

Applying the Martens Clause to the ban on LAWS, Akhmetov argues that public conscience demands "principles of morality and ethics, the exclusive sphere of human responsibility" be applied and that "robots are not able to appreciate the value of human life" to make life or death decisions.[15] It is arguable that countries should already stop developing LAWS based on the Martens Clause. According to the Arms Control Association:

> The Martens clause requires in particular that emerging technology comply with the principles of humanity and dictates of public conscience. Fully autonomous weapons would fail this test on both counts.[16]

Meanwhile, Ethics and Technology has managed to hold meetings to work on a range of issues, including meaningful human control; current and new legal acts; responsibility for system operations; elements of system operation; defence and attack; and expansion beyond the military sphere.[17] So far, the NGO is not banned and its campaign against LAWS has not faced major overt opposition from the Russian government. In fact, the Russia government simply moves forward in developing its AI-enabled weapon systems. Most notably, on 22 August 2019, Russia sent a Skybot F-850 robot called FEDOR (Final Experimental Demonstration Object Research) to the International Space Station as the sole passenger on a Soyuz rocket. FEDOR, a humanoid-looking robot, was sent to practise routine maintenance and repair tasks on the space station.[18] But FEDOR is a dual-use robot, also capable

[9] https://www.giplatform.org/events/group-governmental-experts-lethal-autonomous-weapons-systems-gge-laws-2nd-meeting-2019

[10] www.ethicsandtech.ru/en

[11] https://publish.twitter.com/?query=https%3A%2F%2Ftwitter.com%2Falenapopova%2Fstatus%2F1164166965361074181&widget=Tweet

[12] www.cisp-astana.kz

[13] https://en.wikipedia.org/wiki/Martens_Clause

[14] Email exchange with author, 29 July 2019.

[15] Ibid.

[16] Docherty, B. (2018, October). REMARKS: Banning 'Killer Robots': The Legal Obligations of the Martens Clause. *Arms Control Association*. https://www.armscontrol.org/act/2018-10/features/remarks-banning-%E2%80%98killer-robots%E2%80%99-legal-obligations-martens-clause

[17] www.ethicsandtech.ru/en

[18] Cuthbertson, A. (2019, 22 August). Russia sends gun-toting humanoid robot into space. *The Independent*. https://www.independent.co.uk/life-style/gadgets-and-tech/news/russia-robot-space-fedor-humanoid-iss-a9074716.html

of very lethal actions. In a video shared on Twitter by Russia's Deputy Prime Minister Dmitry Rogozin, FEDOR can be seen firing pistols with high accuracy on a shooting range.[19]

While dual-use machines arguably cannot be prevented, Russia has developed other LAWS that are less likely to be used outside of war. They are not fully autonomous, performing mostly via remote control, but could soon potentially be fitted with AI technologies to work independently. They include tanks, mine-clearing bulldozers and surveillance and utility seacrafts.[20] The most advanced of these might be the Soratnik, an "unmanned vehicle capable of autonomously picking, tracking and shooting targets" without direct active human intervention.[21]

The rising tide against LAWS

Many countries have already expressed a desire for a global ban on AI-enabled weapons, with 29 agreeing to such a declaration, Jordan being the latest addition.[22] Significantly, the People's Republic of China is also on this list, arguing for a ban on the use of fully autonomous weapons, but not their development or production.[23] Arguably, China might have a point in wanting to continue developing such machines, because they can have dual-use capabilities. Although AI-enabled machines that have peaceful functions can be converted into killing machines, an outright ban would, for example, stop the development of a useful machine such as the above-mentioned Skybot F-850.

At the recent CCW meeting in Geneva, the US and Russia continued to argue for the legitimisation of LAWS, but it appears it is a "losing fight against the inevitable treaty that's coming for killer robots."[24] The only countries developing LAWS that we know of are China, Israel, Russia, South Korea, the United Kingdom and the US. None of them are so far along the AI weapons development path that the course cannot be reversed. Just like the ban on landmines was a success, a ban on AI weapons is within grasp.

Supporting the Ban Killer Robots campaign is a worldwide movement and cannot be just country-specific. With enough global support, Russia can be persuaded to come around to support this ban as well, since the huge costs involved in AI weapons development puts them at a disadvantage compared to the spending in the US, which is further along in the development of and investment in AI weapons. Russia reportedly spends an estimated USD 12.5 million on developing AI use in weapons, although the true figure is unknown.[25] A ban on LAWS would disarm the US and neutralise its economic advantage, and actually benefit Russia's military position as a result.

Action steps

Activists working in this field need to step up their work by connecting to more countries around the world than the ones currently in the movement.[26] Establishing organisations in more countries could eventually persuade additional nations to turn against LAWS at the United Nations, isolating those few countries that do develop such weapons. Another approach is convincing tech workers to pressure companies not to develop such weapons. This has worked at Google, whose workers refused to work on an AI project that has implications for weapons.[27] Finally, NGOs should work directly with their country's government using productive and frank discussions. This can happen in Russia and elsewhere. The budget for LAWS in Russia can be better spent elsewhere. While the global campaign might seem optimistic, the ban on such lethal weapons appears inevitable and governments should be persuaded to consider the benefits of coming on board.

[19] https://twitter.com/Rogozin/status/852869162493935617?ref_src=twsrc%5Etfw%7Ctwcamp%5Etweetembed%7Ctwterm%5E852869162493935617&ref_url=https%3A%2F%2Fnews.yahoo.com%2Frussia-sends-gun-toting-humanoid-133245592.html

[20] RT. (2018, 5 May). Op. cit.

[21] Ibid.

[22] Campaign to Stop Killer Robots. (2019, 22 August). Russia, United States attempt to legitimize killer robots. https://www.stopkillerrobots.org/2019/08/russia-united-states-attempt-to-legitimize-killer-robots

[23] Campaign to Stop Killer Robots. (2019, 21 August). Country Views on Killer Robots. https://www.stopkillerrobots.org/wp-content/uploads/2019/08/KRC_CountryViews21Aug2019.pdf

[24] Campaign to Stop Killer Robots. (2019, 22 August). Op. cit.

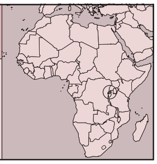
[25] Bendett, S. (2018, 4 April). In AI, Russia Is Hustling to Catch Up. *Defense One*. https://www.defenseone.com/ideas/2018/04/russia-races-forward-ai-development/147178

[26] https://www.stopkillerrobots.org/members

[27] Campaign to Stop Killer Robots. (2019, 14 January). Rise of the tech workers. https://www.stopkillerrobots.org/2019/01/rise-of-the-tech-workers

RWANDA

AI EYED TO TRANSFORM HEALTH CARE IN RWANDA



Emmanuel Habumuremyi
https://www.linkedin.com/in/emmanuel-habumuremyi-b808861a

Introduction

Although artificial intelligence (AI) has started to raise human rights concerns at the global level, this is not yet the case in Rwanda, which still sees it mainly as an opportunity for meeting the country's development needs. On one side, the country is witnessing young innovators thriving to put in place AI solutions while waiting curiously for the time when they will benefit from its socioeconomic promise. On the other side, this has awakened policy makers' minds, who are now thinking about the necessary policy and regulation framework to be established in Rwanda to cater for the use of AI in various domains.

This report identifies existing policies and regulations in place governing the use of AI in service delivery, before considering how AI is being piloted in the health sector to bring the country closer to achieving the United Nations Sustainable Development Goals (SDGs) on issues to do with ending health care inequality. It ends by offering a set of first-base action steps to strengthen the use of AI in Rwanda.

Country context

The presence of Sophia, a humanoid robot,[1] at the Transform Africa Summit held in Kigali on 9-11 May 2019 raised awareness in Rwanda of the existence of AI and its potential. The summit sparked a plethora of ideas in the area: "We were inspired by the technology the robotic company used to create Sofia."[2] However, a large number of students with dreams to become national and global leaders in software development face challenges ranging from a lack of facilities and exposure to competition from expatriate IT experts.[3]

Rwanda, as a small land-locked country located in eastern Africa's Great Lakes region, thrives despite the fundamental development constraints it faces following the 1994 genocide committed against Tutsis that left the country on the list of "failed states". Information and communications technology (ICT), at the centre of its economic transformation agenda, made the country one of the fastest growing economies – with a GDP growth of around 8% per year between 2001 and 2014.[4] In the 2019 Index of Economic Freedom published by the Heritage Foundation, Rwanda's economic freedom score is 71.1, making its economy the 32nd freest in the 2019 index.[5] In terms of connectivity, the highest 2G technologies cover 99.13% of the country and reach 99.92% of the population. Deployed 3G technologies cover 77.40% of the country and reach 93.37% of the population. 4G LTE technology has a 94.2% geographic coverage and reaches 96.6% of the population, as highlighted by the Rwanda Utilities and Regulatory Authority (RURA).[6]

The country's leadership believes that emerging technology can help it leapfrog the digital divide. With the world fast embracing the "Fourth Industrial Revolution", the president of Rwanda, Paul Kagame, said that Africa's challenge is to catch up in providing universal broadband:

> To succeed in making African homes, offices, schools and cities smart, we have to harness opportunities in exponential technologies. These technologies include Artificial Intelligence, Robotics, Drones, Big Data, Block chain, and 3D printing.[7]

For Rwanda, AI is an emerging technology to be embraced, not to be avoided. According to Antoine Sebera,[8] the government's chief innovation officer, most of the current computer applications being

[1] Xinhua. (2019, 15 May). Humanoid robot Sophia addresses Africa technology summit in Rwanda. *Xinhua*. www.xinhuanet.com/english/africa/2019-05/15/c_138061183.htm

[2] Buningwire, W. (2019, 1 July). Kigali High School Students Create A Robot, Calls It 'Sofia's Brother'. *KT Press*. https://ktpress.rw/2019/07/kigali-high-school-students-create-a-robot-calls-it-sofias-brother

[3] Rwirahira, R. (2018, 20 December). AI experts seek government help. *Rwanda Today*. rwandatoday.africa/news/AI-experts-seek-government-help/4383214-4904052-4dpp28/index.html

[4] Hutt, R. (2016, 7 April). 5 things to know about Rwanda's economy. *World Economic Forum*. https://www.weforum.org/agenda/2016/04/5-things-to-know-about-rwanda-s-economy

[5] https://www.heritage.org/index/country/rwanda

[6] https://rura.rw/fileadmin/Documents/ICT/statistics/Quarterly_Telecom_Statistics_report_as_of_March_2019_.pdf

[7] paulkagame.com/?p=5257

[8] Interview with the government's Chief Innovation Officer Antoine Sebera, May 2019.

manufactured have components that are AI-enabled. It is a global trend which raises ethical, social and cultural issues that necessitate the establishment of policies to grow the economic sector, but at the same time to protect people and promote ethics. The ongoing policy development to cater for AI is being supervised by the Ministry of ICT and Innovation, in partnership with RURA, the Rwanda Information Society Authority (RISA), and all relevant stakeholders from the public and private sectors and civil society.

In September 2018, AI officially entered the university curriculum, thanks to a master's degree launched by the Senegalese expert Moustapha Cissé, head of Google's AI research centre in Ghana, and by the African Institute of Mathematical Sciences (AIMS) in Kigali.[9]

Citizen awareness programmes specifically focusing on AI are not yet in place in the country. But community awareness on digital literacy is covered under the Rwanda Digital Ambassador Programme (Rwanda DAP). Started in 2017, the DAP initiative aims to transform the lives of five million citizens by bringing them online through training held in their respective communities.[10] The programme mobilises 5,000 young leaders who help Rwandans acquire digital skills and adopt e-services, driving inclusion and growth. In alignment with the Rwanda National Digital Talent Policy[11] and the World Economic Forum's Internet for All Initiative,[12] Rwanda's DAP is being implemented countrywide in partnership with Digital Opportunity Trust (DOT),[13] a civil society organisation. This initiative drives digital adoption, and helps bridge the ICT skills gap – exposed to digital devices, citizens are educated on opportunities, rights and security online.

Securing the right to health using AI in Rwanda

Since 2010, Rwanda has been involved in various Broadband Commission for Sustainable Development[14] discussions on how digital tools can increase access to health, empower patients, and provide better health information, and how real-time data can ensure that monitoring systems are more action-oriented and prioritise limited resources.[15] Below is the current status of the use of IT, including AI, in the health system.

The health system in Rwanda[16]

Rwanda believes in health care for all where health care coverage is ensured through a health insurance scheme called "Mutuelles de Santé", started in 1999. Citizens pay premiums into a local health fund via an online platform called Irembo,[17] and can draw from it when in need of medical services.

Rwanda's health care system is one of the most advanced in Africa.[18] In 1995, the country established a community health workers' framework (CHW), which was aimed at increasing the uptake of essential maternal and child clinical services through the education of pregnant women, the promotion of healthy behaviours, and encouraging follow-ups at health centres. Today CHW health workers treat malaria, diarrhea and other health challenges by visiting homes in the community, while also providing family planning, nutrition and hygiene advice, and consultations for expectant mothers. If a patient suffers a serious condition, they are transferred to a nearby health centre. When asked by *The Medical Futurist*,[19] Zuberi Muvunyi, director general of the Clinical and Public Health Services Department at the Rwanda Biomedical Centre, said that the CHW programme is one of the reasons Rwanda performed well in achieving the Millenium Development Goals (MDGs).[20]

However, as far as human resources are concerned in the health sector, Muvunyi said that Rwandans only have "one doctor for more than 10,000 [and] that is way below WHO recommendations. For nurses, maybe we have one nurse for 5,000, while the WHO recommendation is one for 3,000."

The most challenging obstacles to high quality health care in Rwanda are summarised in the Fourth Health Sector Strategic Plan 2018-2024:

- Critical shortage of skilled health workers
- Poor quality of health worker education

[9] Nkusi, A. (2019, February). The Rwandan miracle. *UNESCO Courier*. https://en.unesco.org/courier/2019-2/rwandan-miracle

[10] ITU News. (2017, 20 February). Digital Ambassadors Program kicks off in Kigali. *ITU News*. https://news.itu.int/digital-ambassadors-program-kicks-off-in-kigali

[11] https://minict.gov.rw/fileadmin/Documents/Policies2019/National_Digital_Talent_Policy.pdf

[12] http://www3.weforum.org/docs/White_Paper_Internet_for_All_Investment_Framework_Digital_Adoption_2017.pdf

[13] https://www.dotrust.org

[14] https://www.broadbandcommission.org/Pages/default.aspx

[15] Broadband Commission for Sustainable Development. (2017). *Digital Health: A Call for Government Leadership and Cooperation between ICT and Health*. https://www.broadbandcommission.org/Documents/publications/WorkingGroupHealthReport-2017.pdf

[16] https://dhsprogram.com/pubs/pdf/SPA3/02Chapter2.pdf

[17] https://irembo.gov.rw/rolportal/en/web/rssb/cbhi

[18] https://youtu.be/Iym_AnMzSrk

[19] The Medical Futurist. (2018, 29 August). Rwanda and the Dreamers of Digital Health in Africa: Wakanda Is Real. *The Medical Futurist*. https://medicalfuturist.com/digital-health-in-rwanda

[20] Ibid.

- Inadequate infrastructure and equipment in health facilities
- Inadequate management of health facilities.[21]

Digital health transformation

Given the above context, there is a belief that telehealth solutions, phone-based platforms and medical drones are possible ways to continue to improve health care in Rwanda. Through the Smart Rwanda 2020 Master Plan,[22] the digitisation of the health sector plays a pivotal role in national development, along with a number of initiatives, including to do with policy, legislation and investment, that have been put in place to enable and promote the government's overall digital transformation agenda. Rwanda, together with its partners in the private sector and funders, are looking for ways to align the uptake of digital services in the health sector with the emerging ICT ecosystem. This includes looking at issues such as point-of-care, data security, cost and financing in the health sector.

Two start-ups, Zipline[23] and Babyl,[24] are among various innovation companies attracted by Rwanda's policy of 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 counties.[25] These start-ups are expanding and innovating with the full support of the Rwandan government. Zipline is deploying its second drone launching site, while Babyl is working on its AI solution for tablets used by community health workers.

Zipline, a California-based start-up, began talking to the Rwandan government in early 2015, when the company approached a number of African governments with the idea for the delivery of medicines and blood to those who need it most using drones. As for Babyl, it has started to use AI in a call centre as a method for triage, while its final goal is to use trained AI to support community-based health workers, essentially offering them access to the expertise of doctors.[26]

Another AI initiative is spearheaded by the Rwandan government in partnership with the World Economic Forum. It aims to increase the country's diagnostic capacity for detecting cancer. As examples such as the use of AI for screening breast cancer at the Massachusetts General Hospital and Harvard Medical School prove, AI can be used to accurately assess scans, make recommendations for treatment, and reduce unnecessary surgeries caused by false positives. With the right kind of policy and infrastructure in place, the potential benefits of AI-driven medicine would be enormous for Rwanda.[27]

The Babyl pilot in health service provision

Babyl Rwanda is a digital health care provider, registered in Rwanda with its headquarters in London. Its main objective is "to put an accessible and affordable health service into the hands of everyone, by combining the computing power of machines with the best medical expertise of humans to create a comprehensive, immediate and personalised health service and make it universally available."

Babyl was launched in 2016. Since then, it has two million registered users in Rwanda and has conducted tens of thousands of consultations. For Babyl patients, this has meant minimising waiting time for consultations and also offers medical insurance to help patients pay for the service. Babyl has the doctors and nurses on its payroll, and meets its costs through the insurance coverage and the registration fees. Sixteen percent of the patients who belong to the poorest of the poor are exempted from paying the fees. In Rwanda, since 90% of Rwandans are covered by government health insurance, Babyl also entered into an agreement with Rwanda's Ministry of Health and Mutuelles de Santé, the government-subsidised community insurance scheme. This has allowed public health facilities to be used for lab tests and advanced consultations, and the costs are then covered by Mutuelles de Santé.

Babyl has taken further steps to revolutionalise the way health services are provided in the country through using AI. Its first steps in this area have aimed to offer a way of easing the burden on crowded hospitals.[28] An interview with a staff member from Babyl Rwanda confirmed that the company has already concluded its pilot phase that tested the use of an AI chatbot that uses machine learning to interact with patients using

[21] https://moh.gov.rw/fileadmin/templates/Docs/FINALH_2-1.pdf

[22] https://minict.gov.rw/fileadmin/Documents/Mitec2018/Policies___Publication/Strategy/SMART_RWANDA_MASTER_PLAN_FINAL.pdf

[23] https://flyzipline.com

[24] www.babyl.rw

[25] Norbrook, N. (2019, 8 March). Clare Akamanzi: 'We have made it easier to do business'. *The Africa Report*. https://www.theafricareport.com/9722/clare-akamanzi-we-have-made-it-easier-to-do-business

[26] Ito, A. (2018, 16 August). This 27-Year-Old Launches Drones That Deliver Blood to Rwanda's Hospitals. *Bloomberg*. https://www.bloomberg.com/news/articles/2018-08-16/this-27-year-old-launches-drones-that-deliver-blood-to-rwanda-s-hospitals

[27] Schwartz, P. (2019, 14 March). Why AI will make healthcare personal. *World Economic Forum*. https://www.weforum.org/agenda/2019/03/why-ai-will-make-healthcare-personal

[28] https://www.youtube.com/watch?v=O50XldW05I8

voice or free text. This is a computer programme specialised in medical triage that can provide medical diagnoses.[29] The idea is to provide easy, affordable, AI-enabled primary public health care. Babyl's system is an integrated AI platform for patients, including an AI triage symptom checker, a health assessment, and virtual consultations with a doctor when referral is needed.[30]

In the future, Babyl hopes that with increased smartphone penetration and a better internet network, more processes will be supported by AI. With its growing database of patient history, it hopes to be able to predict outbreaks, detect epidemics and even enable better diagnosis based on regional health trends.

Issue of health data management

In using AI, the data management issue is raised. A report called *GDPR – Not Just an EU Concern: The implications for Africa* states that outside of constitutional references to data protection, Rwanda and other countries appear to have very limited focus on data protection in their legal systems.[31] In an interview with Antoine Sebera on data collection, he said that Rwanda has considered the European Union's General Data Protection Regulation (GDPR) to protect the privacy rights of Rwandans. As AI relies on data, a data protection policy in the pipeline will be in place by 2020. At the moment, permission to collect the data of citizens to do with their health status is provided by the Ministry of Health and collected data is hosted in data centres within the country's borders.

Conclusion and action steps

The potential of AI to transform the health sector in Rwanda is at its early stage in the country. It is expected to dramatically strengthen all aspects of the health system and expedite the achievement of universal health coverage and better disease surveillance and response. However, the use of digital health solutions and technologies is not yet at the required level that fully satisfies the seeker of health services.

To mitigate challenges and deal with existing gaps, the following steps are needed:

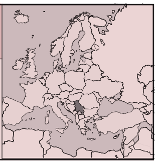
- A specific department coordinating innovations and emerging technology in the health sector should be established under the Ministry of Health.
- Policy frameworks need to be in place that determine AI governance structures and set roles for stakeholders for better monitoring and evaluation of the benefits, and to mitigate the challenges that may emerge through the use of AI.
- Financial and capacity-building support is needed for young innovators to establish and grow a start-up, including health tech start-ups.
- Human capacity needs to be strengthened: most African countries, among them Rwanda, rely on foreign human resources. This is a call for more effort to be put into developing local capacity.
- A large number of citizens are not at all familiar with AI. Public awareness is needed in order to enable them to understand the opportunities and challenges associated with AI.
- In the same way that there is a code of conduct for the use of citizen data (despite the absence of strong legislation), custodians of data collected through AI should have a known code of ethics on what is allowed or not while using collected data.

[29] Bizimungu, J. (2018, 11 January). Babyl's chatbot to enhance digital healthcare platform. *The New Times*. https://www.newtimes.co.rw/section/read/227369

[30] USAID Center for Innovation and Impact. (2019). *Artificial Intelligence in Global Health: Defining a Collective Path Forward*. https://www.usaid.gov/sites/default/files/documents/1864/AI-in-Global-Health_webFinal_508.pdf

[31] Miller, R, et. al. (2018). *General Data Protection Regulation (GDPR) – Not Just an EU Concern: The implications for Africa*. Dalberg Advisors. https://www.dalberg.com/system/files/2018-05/GDPR_Implications%20for%20Africa_EMAIL%20PDF-vFinal%20March2018.pdf

SERBIA



LIVING UNDER THE WATCHFUL EYE: IMPLICATIONS OF FACIAL RECOGNITION SURVEILLANCE IN SERBIA

SHARE Foundation
Bojan Perkov and Petar Kalezić
https://www.sharefoundation.info

Introduction

In early 2019, the minister of interior and the police director of Serbia announced[1] that 1,000 cutting-edge security cameras with facial recognition capabilities will be installed in 800 locations in Belgrade, the Serbian capital, in partnership with Chinese tech giant Huawei. However, despite the flaws of facial recognition and the intrusiveness for citizens' privacy when it is used for surveillance,[2] there is no transparency about the cameras and the partnership between the Ministry of Interior and Huawei, which is part of a broader cooperation between the Serbian and Chinese governments.

In November 2018, Serbia adopted a new Law on Personal Data Protection[3] based on the European Union's (EU) General Data Protection Regulation (GDPR).[4] The application of the law starts on 21 August 2019, after a nine-month adaptation period provided for compliance with the new rules. SHARE Foundation,[5] a Serbian non-profit organisation established in 2012 to advance human rights and freedoms online, submitted freedom of information requests to the Ministry of Interior asking for information about the cameras and supporting documents (e.g. memorandums, contracts, letters of intent). The Ministry withheld this information, meaning that the public in Serbia was left in the dark about a very problematic technology which can greatly impact the privacy of all citizens.

Legislative context

Similar to other countries with a history of repressive regimes and a broad state surveillance apparatus, there is little of a culture of privacy in Serbia. For example, the first data protection law in Serbia was adopted in 2008. After nearly 10 years of application, it turned out that the law was not good enough to provide an adequate level of protection, especially in a world of expanding technologies such as targeted advertising and a whole new digital economy based on personal data. Also, the law did not regulate video surveillance, which opened space for numerous abuses when it comes to data processing through CCTV systems, both state and privately owned. Introducing a new law was an opportunity to regulate this area of personal data processing, but the provisions on video surveillance were not included in the final text of the Law on Personal Data Protection.

In its annual report for 2018, the Commissioner for Information of Public Importance and Personal Data Protection, Serbia's independent authority for both freedom of information and protection of citizens' personal data, highlighted the fact that the Ministry of Justice argued to keep the regulation of video surveillance out of the new Law on Personal Data Protection. The Ministry's view was that this area of personal data processing should be regulated by a special law and that the GDPR does not contain provisions on video surveillance.[6] However, more than six months after the new Law on Personal Data Protection had been adopted in the National Parliament of Serbia and just two months before it is scheduled to start being applied, no specific law regulating video surveillance has been drafted or even proposed.

[1] SHARE Foundation. (2019, 29 March). New surveillance cameras in Belgrade: location and human rights impact analysis – "withheld". https://www.sharefoundation.info/en/new-surveillance-cameras-in-belgrade-location-and-human-rights-impact-analysis-withheld

[2] Big Brother Watch. (2018). *Face Off: The lawless growth of facial recognition in UK policing*. https://bigbrotherwatch.org.uk/wp-content/uploads/2018/05/Face-Off-final-digital-1.pdf

[3] Republic of Serbia. (2018). Law on Personal Data Protection. Available in Serbian at: www.pravno-informacioni-sistem.rs/SlGlasnikPortal/eli/rep/sgrs/skupstina/zakon/2018/87/13/reg

[4] European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679

[5] https://www.sharefoundation.info/en

[6] Commissioner for Information of Public Importance and Personal Data Protection of the Republic of Serbia. (2019). *Summary report on the implementation of the Law on Free Access to Information of Public Importance and the Law on Personal Data Protection for 2018*. https://www.poverenik.rs/images/stories/dokumentacija-nova/izvestajiPoverenika/2018/ENGRezime2018.pdf

Over the past six months, at the time of writing, Serbia has been in a state of political turmoil. There have been anti-government protests across the country after an opposition politician was assaulted by a group of men in December 2018.[7] It is in contexts such as these where facial recognition surveillance systems, which store large amounts of biometric data, could potentially be used for pressuring citizens who are protesting, as well as their families, because of their political views. Beyond political protest, the everyday use of the cameras comes with the risk of data breaches, which includes records of the daily routines and movements of citizens, and which could potentially result in harm to those unwittingly surveilled.

An opaque Panopticon: Citizens in the dark

As soon as Huawei's facial recognition cameras were announced in the media by the highly ranked officials in Serbia's internal affairs, SHARE Foundation decided to find out more about Huawei's cameras in terms of their location, public procurement and other relevant procedures by submitting freedom of information requests to the Ministry of Interior.

In their responses to our requests, the Ministry stated that all information about the procurement of Huawei cameras is "confidential" and therefore not for public access. Also, in an interview for Radio Television of Serbia, Police Director Vladimir Rebić said that the locations of stationary cameras were already determined based on "a broad examination and analysis of events, referring primarily to the criminal offences in Belgrade." We also requested a copy of this analysis, but the Ministry responded that the information, as well as the location of the cameras, were not contained in any document or other medium, meaning they cannot be provided upon a freedom of information request.[8]

According to Article 54 of the new Law on Personal Data Protection, if it is likely that certain data processing will present a high risk to the rights and freedoms of natural persons, the data controller is obligated to conduct a data protection impact assessment before the beginning of data processing.[9] When SHARE Foundation requested a copy of this impact analysis, the Ministry simply stated that the provisions of the new law are not yet being applied. Having read the official responses of the Ministry of Interior, which suggest that this information does exist, it seemed strange that information provided by a freedom of information officer about such a sensitive topic for citizens' privacy was contradictory to the statements made by the minister and the police director in the media.

While the Ministry was reluctant to provide any official information about the cutting-edge cameras and their procurement, Huawei on the other hand was more transparent about its cooperation with the Serbian authorities. A case study titled "Huawei Safe City Solution: Safeguards Serbia" was available on Huawei's official website and it provided detailed information about the cameras and related video surveillance solutions, claiming the cameras were already installed in Belgrade. SHARE Foundation published an article about Huawei's case study,[10] which strangely disappeared from the company's website shortly after the publication of our article. Having in mind the sensitivity of this content, we saved an archived copy[11] of the page so the case study can still be accessed online.

Huawei stated that for the test phase, nine cameras in five locations were deployed, with the locations being the Ministry of Interior headquarters, a sports arena, a commercial centre and a police station. After this test deployment, it is stated in the case study that Huawei and the Ministry achieved a Strategic Partnership Agreement in 2017 and that in the first phase of the project 100 high-definition video cameras were installed in more than 60 key locations, with the command and data centre in Belgrade being remodelled.[12]

It is very worrying that such advanced technology, which has great implications for privacy, is being deployed without citizens knowing about this digital "watchful eye" collecting and storing large amounts of their biometric data, even if they have done nothing wrong. Saša Đorđević, a researcher at the Belgrade Centre for Security Policy,[13] a Serbian think tank dedicated to advancing the security of citizens and society, said that

[7] Vasovic, A. (2018, 8 December). Thousands protest in Serbia over attack on opposition politician. *Reuters*. https://www.reuters.com/article/us-serbia-protests/thousands-protest-in-serbia-over-attack-on-opposition-politician-idUSKBN1O70S7

[8] SHARE Foundation. (2019, 29 March). New surveillance cameras in Belgrade: location and human rights impact analysis – "withheld". Op. cit.

[9] Republic of Serbia. (2018). Op. cit.

[10] SHARE Foundation. (2019, 29 March). Huawei knows everything about cameras in Belgrade – and they are glad to share! https://www.sharefoundation.info/en/huawei-knows-everything-about-cameras-in-belgrade-and-they-are-glad-to-share

[11] https://archive.li/pZ9HO

[12] SHARE Foundation. (2019, 29 March). Huawei knows everything about cameras in Belgrade – and they are glad to share! Op. cit.

[13] www.bezbednost.org

although video surveillance can improve security and safety, primarily in road traffic safety, the list of unknown things about Huawei's video surveillance in Belgrade is long. "The situation can still be corrected if and when it is determined which video surveillance equipment is being purchased, how much it costs the citizens of Serbia, where it is placed and how the personal data will be processed and protected," he added.[14]

Another problematic aspect of facial recognition technology when used for video surveillance is that it is prone to mistakes, which is especially important for law enforcement and legal proceedings. Research by Big Brother Watch has shown that in the United Kingdom the overwhelming majority of the police's "matches" using automated facial recognition have been inaccurate and that on average, 95% of "matches" made by facial recognition technology wrongly identified innocent people as crime suspects.[15]

In addition, Big Brother Watch found that the police stored photos of all people incorrectly matched by automated facial recognition systems, meaning that biometric photos of thousands of innocent people have been stored.[16] Storing such sensitive biometric data of citizens is also a privacy and security risk, which is even greater taking into account information leaks during police investigations, which are common in Serbia. "The media in Serbia frequently publish information relating to investigations conducted by the police and the prosecution, quoting mostly unknown sources allegedly 'close to the investigation' and sometimes with photos," explained Đorđević. He mentioned an example when information from the investigation of the murder of a Serbian singer was constantly published on the front pages of tabloid newspapers. Another case Đorđević highlighted occurred in February 2017, when one daily newspaper covered the arrest of a member of a football club supporters' group on its front page the evening before the police informed the public about his arrest.[17]

In other parts of the world there are similar concerns. With all the recent controversy surrounding facial recognition, two cities in the United States have so far banned the use of facial recognition by the city administration – the first was San Francisco, California, followed by Sommerville, Massachusetts.[18] It is highly likely that more cities will join them, particularly since there is more and more awareness of the negative impacts of facial recognition surveillance.

Conclusion

SHARE Foundation will again approach the Ministry of Interior for information, especially relating to the data protection impact assessment of data processing using Huawei's cameras, after the application of the new Law on Personal Data Protection starts. Of course, data processing through video surveillance systems should be regulated without delay, either though amendments to the Law on Personal Data Protection or through a separate law. It is also important to introduce citizens to the risks of such invasive technologies and call them to action, as it will provide momentum to further pressurise the authorities and demand more transparency. People feel more secure when they see a camera, as Đorđević noted,[19] but there is also a general lack of understanding of who may collect their personal data, the purposes for which it will be collected, and their rights as data subjects.

Moreover, it is also necessary that the new Commissioner for Information of Public Importance and Personal Data Protection is appointed as soon as possible and in a transparent manner,[20] as the second and final term of Rodoljub Šabić, the previous Commissioner, expired in December 2018. As head of an independent institution, the Commissioner plays a key role in protecting citizens' personal data and freedom of information and it is of utmost importance that the position is given to a person who has personal integrity, expertise and political independence. Otherwise, affairs such as the one with Huawei's facial recognition cameras may never be resolved, which would leave citizens exposed to huge risks to their privacy and without appropriate safeguards in cases of data breaches and abuse of personal data. There will also be many doubts around how to apply the new Law on Personal Data

[14] Email correspondence with Belgrade Centre for Security Policy researcher Saša Đorđević, 29 June 2019.

[15] Big Brother Watch. (2018). Op. cit.

[16] Ibid.

[17] Email correspondence with Belgrade Centre for Security Policy researcher Saša Đorđević, 29 June 2019.

[18] Haskins, C. (2019, 28 June). A Second U.S. City Has Banned Facial Recognition. *VICE*. https://www.vice.com/en_us/article/paj4ek/somerville-becomes-the-second-us-city-to-ban-facial-recognition

[19] Email correspondence with Belgrade Centre for Security Policy Researcher Saša Đorđević, 29 June 2019.

[20] Delegation of the European Union to the Republic of Serbia. (2019, 28 January). Fabrizi: Appointment of new Commissioner for Information of Public Importance should be kicked off as soon as possible. https://europa.rs/fabrizi-appointment-of-new-commissioner-for-information-of-public-importance-should-be-kicked-off-as-soon-as-possible/?lang=en

Protection, which could prove to be quite a challenge if the Commissioner is not up to the task or is easily influenced by other state institutions and political actors.

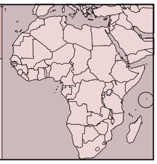
Action steps

Having taken into account the lack of transparency surrounding Huawei's surveillance cameras in Belgrade, we propose the following action steps:

- Insisting on the proper application of the new Law on Personal Data Protection and conducting the necessary data protection impact assessment.
- Advocating for the regulation of video surveillance by law in order to provide legal certainty.
- Engaging the wider community (e.g. civil society organisations, human rights defenders, tech experts, journalists) to help raise awareness among citizens about the impact of video surveillance.
- Pressuring the Ministry of Interior and other relevant state institutions to provide information about video surveillance and facial recognition in a transparent way.

SEYCHELLES

WHAT PEOPLE THINK OF ARTIFICIAL INTELLIGENCE IN A SMALL AND REMOTE ARCHIPELAGO



Janick Bru
janickbru@hotmail.com

Introduction

The Seychelles is increasingly finding more innovative ways to sustain its development – for example, building on the potential of the "Blue Economy" by launching the world's first sovereign blue bond, described by the World Bank as "a pioneering financial instrument designed to support sustainable marine and fisheries projects."[1] Yet another example is the Fishguard project,[2] which is meant to be integrated into the fisheries patrol routines of the Seychelles air force and the Seychelles coast guard. The project, piloted in October 2018, monitors large marine areas using a combination of short- and long-range drones equipped with artificial intelligence (AI). It was said that the drones "are programmed to be self-reliant and capable of making independent decisions based on data collected."[3] The aim of the project is to reduce the effects of "illegal, unreported and unregulated fishing on Seychelles' marine resources."

The Seychelles exclusive economic zone (EEZ), which was 1,374,000 km^2 (about six times the size of the UK), was extended by 14,840 km^2 in September 2018. In addition, the country jointly manages a marine space of 397,000 km^2 with Mauritius as part of the Extended Continental Shelf[4] – this being the first such collaboration at global level. Considering that the country only has a total population of 96,762 people, there are, as a result, many areas where AI could be used to boost its capacity, support development and protect its resources.

During a visit to the Seychelles in early July 2019,[5] the African head for the World Economic Forum (WEF) referred to the country's role in the Blue Economy as being "not just theory". She said Seychelles is a "working model" and that "everyone needs to wake up and realize that the ocean is the beating heart of the world, as people we are not acting with the sense of urgency that it requires." She also mentioned the so-called Fourth Industrial Revolution and its impact on the world, saying, "It is important to hold discussions on this issue as emerging technologies, artificial intelligence, blockchain are fast transforming how we live and work." The WEF website elaborates in the following words:

> These advances are merging the physical, digital and biological worlds in ways that create both huge promise and potential peril. The speed, breadth and depth of this revolution is forcing us to rethink how countries develop. [...] The Fourth Industrial Revolution is about more than just technology-driven change; it is an opportunity to help everyone, including leaders, policy-makers and people from all income groups and nations, to harness converging technologies in order to create an inclusive, human-centred future.[6]

But before any of this can happen in the Seychelles, the people of the country need to have a basic understanding of AI. This report will look at general perceptions that exist regarding AI in the Seychelles and whether or not the average person understands what AI is and what it does.

Perceptions of AI in Seychelles

To test how familiar people in the Seychelles are with the concept of AI, a small number of individuals (14) were asked about it. Three categories of people were targeted: young people, working/

[1] World Bank. (2018, 29 October). Seychelles launches World's First Sovereign Blue Bond. www.worldbank.org/en/news/press-release/2018/10/29/seychelles-launches-worlds-first-sovereign-blue-bond

[2] Bonnelame, B. (2018, 18 September). Project FishGuard: Seychelles to monitor illegal fishing with unmanned drones. *Seychelles News Agency*. www.seychellesnewsagency.com/articles/9754/Project+FishGuard+Seychelles+to+monitor+illegal+fishing+with+unmanned+drones

[3] Ibid.

[4] www.seyccat.org/about-us/seychelles

[5] Laurence, D. (2019, 5 July). Technology, Blue Economy on Seychelles' agenda at World Economic Forum meeting in South Africa in September. *Seychelles News Agency*. www.seychellesnewsagency.com/articles/11277/Technology%2C+Blue+Economy+on+Seychelles%27+agenda+at+World+Economic+Forum+meeting+in+South+Africa+in+September

[6] www.weforum.org/focus/fourth-industrial-revolution

retired members of the general public, and some individuals from formal institutions.

Most of the members of the general public who were adults and employed, or who were retired, tended not to know what "artificial intelligence" was, despite the fact that one of them was a front-desk worker for a telecommunications company and another had recently been working for a popular internet café. Those who said that they knew what it was, referred to robots generally as examples of AI, although one person referred specifically to a "Honda robot prototype that feeds disabled people." All of the people in this group said that they used Google regularly, while a couple also used Facebook. Only two of them said that they were aware that Google and/or Facebook used AI to control content and to learn about users, with one of them stating that both companies "are very tricky and subtle" in how they operate. When prompted, a third individual who had said he was "not sure" about big technology companies using AI added, "When I think about it, I always get adverts about things connected to cars and car parts, because this is what I usually look for online, I never get to see adverts about gluten-free foods."

A 15-year-old student at a state school did not know what AI was and was unable to provide an example. When asked about whether she used Google or Facebook, she said that she did but admitted that she had no idea that these companies used AI to control content or to learn about users. Nonetheless, she did feel that there was a need to regulate social media because "a lot of people do things on there that don't make sense." Another student, also from a state school, said that she knew what AI was and suggested that an example would be a robot. She said that she was not aware that Google and Facebook used AI but believed that there was a need for regulations because "the advanced technology used in AI could be used for the wrong reasons."

A third student, from a private international school, confidently stated that he knew what AI was and gave the following three examples: Google Assistant,[7] Siri[8] and Alexa.[9] He said that he used Google but not Facebook, and that he was very aware of how big technology companies control content and harvest information about users. He believed that there needed to be regulations and that technology companies should have the equivalent of internal audit departments to do this. He felt that such regulations should be applied only when necessary, for example, "if AI starts doing things it should not do," but that otherwise it would be better not to tamper with it. He added, "In case one day AI gains control of itself, it will be necessary for humans to find ways to keep overall control, just in case learning machines go rogue and act against human interests."

Representatives of institutions who agreed to respond to emailed questions were from the media, the public sector, a semi-autonomous agency and an educational institution. All of the individuals from this group knew what AI was. They used sentences such as "when computers think for themselves" or "machines/gadgets that are engineered to automate and 'think' like humans" or "machine intelligence – AI platforms having capacity to solve complex problems". Four of them thought that AI was relevant for the Seychelles, in that the world was developing "in that direction" and that technology was everywhere and constantly advancing. Only one person did not think that AI was relevant for the Seychelles, since the country has a small population and it also has some unemployment. She felt that people would be displaced by AI and that "having a computer think for you, makes you lazy!"

This group also felt unanimously that there should be regulations to control the use of AI. One thought that regulations should apply to national security and military issues, while another felt that there should be forms of control to ensure that "we reap the benefit of such technologies in an *ethical* and *legal*[10] manner (since) we need to ensure that this lucrative industry does not grow unregulated." Another point of view was that most people in the Seychelles tended to be very accepting of what is on the internet and that the only way to protect the country in this area was through "more awareness and much education" because AI needed to be well understood by the general public. One person commented that the Seychelles seemed to be quite vulnerable to cybercrime at both personal and institutional levels. This group strongly felt that AI had to be regulated and all of them thought this should be done by the government through appropriate agencies, with one person adding that the people involved in regulating AI needed to be well trained and thoroughly knowledgeable in this area.

Regarding the type of control and regulations that should exist, the following ideas were put forward: regulations which ensure that legal norms

[7] https://assistant.google.com

[8] https://www.apple.com/siri

[9] https://en.wikipedia.org/wiki/Amazon_Alexa

[10] Words emphasised by respondent.

and ethical standards prevail, while keeping a balance between encouraging innovations and stifling growth, restrictions on where and when AI can be used, and protocols that are designed "to prevent anything getting out of control." One person thought that as long as it was possible to regulate and control the physical and real effects of technology, then the virtual world itself could remain free and unregulated.

The Seychelles Institute of Technology (SIT), which provides technical and vocational education and training after secondary school in the engineering, built environment and ICT fields, was recently recognised by UNESCO-UNEVOC as one of the world's 10 innovative technical and vocational education and training (TVET) institutions.[11] Among other things, it trains future technicians in the installation of photovoltaic panels and solar water heaters, and trains students to retrofit containers as energy-efficient premises that can be used as offices. Additionally, SIT students were trained to build the first gabion rock barrage that provided sufficient water supply for a farming community in the south of the main island of Mahe that had suffered from water shortages for years – this was done through a project that emphasises an approach to advanced engineering that works with and for nature. A brief conversation with a representative of the institute indicated that it had not yet incorporated aspects of robotics or AI in its curriculum but that it was hoped the SIT would be able to do this eventually.

The level of access of the general population to the internet and other forms of telecommunication is high: figures indicate that in December 2018 there were 97,783 internet subscriptions as well as 178,946 mobile phone subscriptions (many having access to the internet) in the country.[12] In mid-2018 there was the launch of digital TV by the Seychelles Broadcasting Corporation, which offers free access to eight television channels (three local channels and five international news channels). To encourage use of this facility, every household in the country was offered a free decoder. Moreover, there are other telecommunications service providers that offer paying alternatives, which are also very popular.

However, while there is clearly a general understanding of AI within institutions, it is also obvious that overall knowledge and understanding of AI in the Seychelles is currently very limited. Despite occasional programmes on television about AI or competitions organised by the National Institute for Science, Technology and Innovation encouraging groups of students to create basic robots,[13] AI is not a topic that appears much in the local news nor is it one that is a subject of discussion professionally or socially, despite the considerable attention it is currently attracting in other parts of the world.

One of the students who answered questions summed up the situation by saying, "People use things but they are not told what it is they are using." On the other hand, the country has been at the forefront of numerous innovative schemes and ventures for sustainable development on land and in the marine environment. It is reasonable to believe that forms of AI, if used appropriately, could enhance the impact of these projects, broaden their reach and make it possible for relatively small teams to effectively manage the large areas of the Seychelles marine territory. A first step would be to educate people in the Seychelles about AI, so that they learn about the potential benefits of this form of technology, but also understand how such advances can be used in ways that may negatively affect them and their way of life.

Action steps

The following steps are needed in the Seychelles:

- Initiate a national conversation around AI that is inclusive.
- Include classes in state schools that prepare students from a relatively early age to understand what new technologies are, how they function and what they do, and progressively move to more advanced forms of the new technologies in secondary schools.
- Increase investment in training in new technology development (including AI) in relevant professional centres, in particular the SIT.
- Develop a practical and conceptual understanding of AI and its potential positive and negative impacts within both the National Institute for Science, Technology and Innovation and the national Department of ICT.

[11] Seychelles News Agency. (2019, 22 July). Seychelles Institute of Technology recognized as one of the world's 10 innovative TVET centers. *SBC*. https://sbc.sc/news/seychelles-institute-of-technology-recognized-as-one-of-the-worlds-10-innovative-tvet-centers

[12] www.ict.gov.sc/ReportsStatistics/Reports.aspx

[13] In July 2017, "Team Seychelles, which was the youngest team and came from the smallest country, came 105th out of 163 countries overall in the FIRST Global Challenge held from 16 to 18 July in Washington. Seychelles was ranked 25th out of 40 countries from Africa." Laurence, D. (2017, 28 July). Science organisation in Seychelles wins support award at int'l robotics competition. *Seychelles News Agency*. www.seychellesnewsagency.com/articles/7658/Science+organisation+in+Seychelles+wins+support+award+at+int%27l+robotics+competition

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-6b0fd073e2c2

[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 schools[11]
- 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 information[18] 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 placement[24] 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 recognition[31] 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 MIIA Meetup at Rise Africa, Cape Town, 27 October. https://drive.google.com/file/d/0BxzNs-HspAzYSDJoMWpVcDdfYnc/view

[25] https://dsideweb.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., McAlister, 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/VKA3Wwqd69r7rydZ.

[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/cmore-south-africas-new-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/201409/3706726-11act4of2013protectionofpersonalinforcorrect.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" perspective[45] 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 cooperatives[46] and data commons[47] 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 data[49] 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-happensand-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

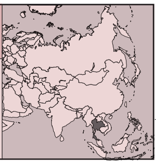
[50] https://ossafrica.com/esst/index.php?title=Summary_of_the_Municipal_Systems_Act%2C_no._32_of_2000

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.

THAILAND



APPRISE: USING AI TO UNMASK SITUATIONS OF FORCED LABOUR AND HUMAN TRAFFICKING[1]

United Nations University Institute on Computing and Society (UNU-CS), Macau and Thailand
Hannah Thinyane and Monticha Puthawong
https://cs.unu.edu/research/migrant-tech-apprise

Introduction

Forced labour and human trafficking affect more than 24.9 million men, women and children globally who are exploited for their labour or forced into prostitution.[2] Figures released by the US State Department indicate that in 2018, only 85,613 victims were identified worldwide.[3] These figures illustrate that there are a large number of people, often migrant workers, who are exploited in slavery-like conditions, yet only a small fraction are being successfully identified and subsequently helped.

The terms "human trafficking" and "forced labour" are often used interchangeably by popular media and practitioners, so it would pay to define each now. In our work, we draw on Skřivánková's continuum of exploitation that defines "decent work" and "forced labour" as two ends of a continuum, with any situation between the two end points representing different forms of labour exploitation.[4] These work situations can range from "cooperative, consensual, mutually beneficial relationships between migrants and their facilitators" to "highly coercive and exploitative".[5] Using this continuum, we can see human trafficking as a process, consisting of a series of exploitative acts that move a worker towards a situation of forced labour. In this report, we use the term "frontline responder" (FLR) to collectively refer to the broad range of stakeholders whose role it is to assess working conditions and to help potential victims access help or remediation channels – including police, labour inspectors, auditors and non-governmental organisations (NGOs).

In this report we draw from the findings of a two-and-a-half-year project aimed at understanding how digital technology can be used to support exploited workers in vulnerable situations. It starts by describing the development process that we have undertaken in Thailand to create Apprise, a system to support proactive and consistent screening of workers in vulnerable situations. The report then frames the potential of using artificial intelligence (AI) to support an understanding of changing practices of exploitation.

Apprise

Our work takes a value sensitive design (VSD) approach which is based on the understanding that technology is shaped by the biases and assumptions of its designers and creators. VSD proactively integrates ethical reflection in the design of solutions, using an integrative and iterative tripartite methodology comprised of conceptual, empirical and technical investigations.[6] With its value focus, this self-reflexive approach seeks to be "proactive [in order] to influence the design of technology early in and through the design process."[7] In doing so, VSD shows a commitment to progress, not perfection.[8]

We began our work in Thailand early in 2017 with a series of focus groups with a broad range of

[1] We would like to acknowledge the funding provided by Humanity United and Freedom Fund to develop the initial Apprise system. Apprise was developed in collaboration with The Mekong Club, an anti-trafficking NGO based in Hong Kong.

[2] International Labor Organization, & Walk Free Foundation. (2017). *Global Estimates of Modern Slavery: Forced Labour and Forced Marriage*. https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/documents/publication/wcms_575479.pdf

[3] US Department of State. (2019). *Trafficking in Persons Report*. https://www.state.gov/wp-content/uploads/2019/06/2019-TIP-Introduction-Section-FINAL.pdf

[4] Skřivánková's, K. (2010). *Between decent work and forced labour: Examining the continuum of exploitation*. Joseph Rowntree Foundation. https://www.jrf.org.uk/report/between-decent-work-and-forced-labour-examining-continuum-exploitation

[5] Weitzer, R. (2014). New Directions in Research on Human Trafficking. *The ANNALS of the American Academy of Political and Social Science*, *653*(1), 6-24. https://doi.org/10.1177/0002716214521562

[6] Friedman, B., Kahn, P., & Borning, A. (2002). *Value Sensitive Design: Theory and Methods*. University of Washington. https://faculty.washington.edu/pkahn/articles/vsd-theory-methods-tr.pdf

[7] Friedman, B., Kahn, P. H., & Borning, A. (2008). Value Sensitive Design and Information Systems. In K. E. Himma, & H. T. Tavani (Eds.), *The Handbook of Information and Computer Ethics*. Wiley & Sons.

[8] We refer the interested reader to the following papers for a full discussion of the motivation for and subsequent design of Apprise: Thinyane, H. (2019). Supporting the Identification of Victims of Human Trafficking and Forced Labor in Thailand. In K. Krauss, M. Turpin, & F. Naude (Eds.), *Locally Relevant ICT Research*. Springer International Publishing; Thinyane, H., & Bhat, K. (2019). Supporting the Critical-Agency of Victims of Human Trafficking in Thailand. Paper presented at the ACM CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, 4 May.

stakeholders, including survivors of exploitation, NGOs, Thai government officials and intergovernmental organisations. These focus groups aimed to understand current practices and problems in identifying victims of human trafficking; their access to and use of technology; and their perception on the ways technology could support them to overcome the problems that they face. To summarise the findings of this initial consultation, focus groups suggested that support was needed during the initial screening phase of victim identification. The core problems that were identified at this stage were:

- *Communication:* Due to a lack of resources (and knowledge of languages that would be required), FLRs commonly faced problems of being unable to speak the same language as workers and were therefore unable to interview them.[9] Translators were also not always available.
- *Privacy:* Initial screening occurs in the field and sometimes in front of potential exploiters. Workers fear retribution if they answer questions honestly.
- *Training:* There is a lack of understanding of the common indicators of labour exploitation and forced labour, with some FLRs focusing on physical indications of abuse, rather than the more subtle forms of coercion such as debt bondage and the withholding of wages and important documents.

Based on these findings, we developed Apprise, a mobile-based expert system to support FLRs to proactively and consistently screen vulnerable populations for indications of labour exploitation. The tool is installed on the FLR's phone, but ultimately it serves to allow workers to privately disclose their working conditions. Questions are translated and recorded in languages that are common among workers in each sector, and when combined with a set of headphones, this provides workers with a private way to answer while in the field. By analysing their responses to a series of yes/no questions, Apprise provides advice on next steps that the FLR should take to support the worker. Responses to questions are stored on the FLR's phone and uploaded to a server when they next log in with network reception, to support *post-hoc* analysis. As well as co-designing the system itself, our consultations with participants uncovered the current indications of exploitation to inform the lists of questions asked. From April 2017 to June 2019, over 1,000 stakeholders in the anti-trafficking field in Thailand contributed to the design or evaluation of the system.

Since March 2018, NGOs have been using Apprise in the field to support proactive and consistent screening in their outreach activities in the following sectors: fishing, seafood processing, manufacturing and sexual exploitation. In May 2018 we started to work closely with the Ministry of Labour (specifically the Department of Labour Protection and Welfare) and Royal Thai Navy in Thailand to understand how Apprise could support proactive and consistent worker screening at government inspection centres at ports (Port-in/Port-out or "PIPO" Inspection Centres) and at sea.

Through this process of working on the ground with FLRs, we have noticed that exploiters continually tweak and refine their own practices of exploitation, in response to changing policies and practices of inspections. When exploiters change their practices, it takes time for these changes to be recognised as a new "pattern" of exploitation. Information is often siloed by different stakeholders, and not shared for a wide variety of different reasons. After some time, stakeholders do begin sharing these changing patterns, often through their informal networks. At this point, the new practice is identified as a pattern and a new policy or practice of inspection is developed.

This game of cat and mouse continues over time, with exploiters again tweaking their behaviour to avoid detection. In response to this, we developed Apprise to allow new questions to be added to question lists, as well as new languages to be supported on the fly. When an FLR logs in to their phone, Apprise checks for any updates to lists and downloads new audio translations of questions. This adaptive support allows FLRs to question on current patterns of exploitation, obtaining further information on exploitative practices once a new pattern has been identified.

Based on this observation, we began to ask ourselves if there was a role for machine learning to support a more timely and more accurate identification of these changing practices of exploitation. While this work is still in its nascent stages, our aim is to determine sector-specific practices of exploitation in order to create targeted education and awareness-raising campaigns; support FLRs to proactively screen against current practices of exploitation; and inform evidence-based policy to support the prosecution of exploiters.

[9] Most migrant workers in Thailand migrate internally from northern Thailand, or from the neighbouring countries of Myanmar, Laos and Cambodia. Across these regions there are hundreds of languages and dialects that are frequently spoken.

Machine learning to detect patterns of exploitation

At its broadest, machine learning works by identifying patterns in existing data. Its main goal is to be able to generalise, so that the patterns identified in training data can be accurately applied to unseen data. Machine learning has been applied in a wide number of criminal justice contexts, including predicting crimes, predicting offenders, predicting perpetrator identities and predicting crime victims.[10] It has also been used in the anti-trafficking field for predictive vulnerability assessments and crime mapping in order to improve government resource allocation.[11] In our work, we aim to understand if there is a role for machine learning to predict changing patterns of exploitation, an area that currently has received little focus.

While there are obvious benefits that accurate forecasting tools could bring,[12] governments, civil society and academics have not always spoken so favourably about these tools, citing cases where they "can reproduce existing patterns of discrimination, inherit the prejudice of prior decision makers or simply reflect the widespread biases that persist in society. [They] can even have the perverse result of exacerbating existing inequalities by suggesting that historically disadvantaged groups actually deserve less favourable treatment."[13]

While recognising different notions of human rights (moral, ethical and philosophical), our work takes a legal approach, based on the Universal Declaration of Human Rights (UDHR),[14] the United Nations Guiding Principles on Business and Human Rights[15] and the International Labour Organization (ILO) Declaration on Fundamental Principles and Rights at Work.[16] These international legal instruments provide an established framework for "considering, evaluating and ultimately redressing the impacts of artificial intelligence on individuals and society."[17]

In order to analyse the human rights impact of machine learning on identifying changing practices of exploitation, we note that an important first point of consideration is the quality of data that is provided in initial screening interviews using Apprise, an issue closely linked to privacy.

Significant attention was paid in the design phase of Apprise to include strict limitations on how much data is collected from individual workers, and also who can access screening responses (and what access they have). As an example, Apprise aims to support accountability and transparency by automatically sharing a summarised version of screening responses with the FLR's immediate supervisor. However, this process limits the accuracy of GPS locations[18] of screening sessions, and only shares responses to the yes/no questions.

To support the privacy of workers, we do not collect any personally identifiable information, as we believe that the risks associated with this would unfairly disadvantage those who chose to answer questions. However, there is no way to delete a particular individual's responses later (should they be able to request this).

Over the past year and a half, we have evaluated and refined Apprise based on feedback from workers in vulnerable sectors as well as survivors of trafficking. The aim of this has been to increase the privacy that workers feel in these initial screening sessions. We note that while no screening system can guarantee truthful responses from workers, Apprise provides more privacy than current methods of interviewing workers, which often occur in groups and in front of potential exploiters (and in the worst cases, using supervisors as translators when language barriers occur).

Within a machine learning system, interview responses would obviously need to be shared further, which requires special consideration. The new patterns of exploitation themselves are intended to be shared with other FLRs, to inform initial screening of workers in vulnerable situations. However, care must be paid as to *who else* has access to them. As soon as exploiters realise that their patterns of exploitation have been identified, they are likely to adapt them more quickly.

[10] Perry, W. L., McInnis, B., Price, C. C., Smith, S., & Hollywood, J. S. (2013). *Predictive Policing: Forecasting Crime for Law Enforcement*. RAND Corporation. https://www.rand.org/pubs/research_briefs/RB9735.html

[11] https://delta87.org/2019/03/code-8-7-introduction

[12] Berk, R., & Hyatt, J. (2014). Machine Learning Forecasts of Risk to Inform Sentencing Decisions. *Federal Sentencing Reporter*, *27*(4), 222-228.

[13] Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review, 671*. https://doi.org/10.2139/ssrn.2477899

[14] https://www.un.org/en/universal-declaration-human-rights

[15] https://www.ohchr.org/documents/publications/GuidingprinciplesBusinesshr_eN.pdf

[16] The Declaration, among other things, commits states to take action to eliminate all forms of forced labour. https://www.ilo.org/declaration/lang--en/index.htm

[17] Raso, F., Hilligoss, H., Kirshnamurthy, V., Bavitz, C., & Kim, L. (2018). *Artificial Intelligence & Human Rights: Opportunities and Risks*. Berkman Klein Center for Internet & Society. https://cyber.harvard.edu/sites/default/files/2018-09/2018-09_AIHumanRightsSmall.pdf

[18] Some of the decimal points in the position are dropped.

In the cases where responses are accurate, and the tool is able to identify new practices of exploitation, there are obvious implications on the rights of exploited workers: the right to freedom from slavery (UDHR Article 4); the right to freedom from torture and degrading treatment (UDHR Article 5); the right to desirable work (UDHR Article 23); the right to rest and leisure (UDHR Article 24); the right to an adequate standard of living (UDHR Article 25); and freedom from state or personal interference in the above rights (UDHR Article 30). An important note is that while the system takes input from a subset of workers (those who have been interviewed), there is potential to impact the working conditions of many more.

Like any system, Apprise may misidentify patterns, resulting in attention being paid in the wrong direction. While this represents an inefficient use of resources (FLR and worker time), it does not have any significant implications on the rights of workers. This input would be used to inform investigations, which themselves would disprove the prediction.

Conclusion

Machine learning has been applied in a wide number of criminal justice contexts.[19] In our work, we aim to understand if there is a role for machine learning to predict changing patterns of exploitation, an area that currently has received little focus.

In this report we describe work that we are undertaking to proactively and consistently screen workers in vulnerable situations for signs of labour exploitation and forced labour. The report introduces Apprise, an expert system that we have developed and that FLRs are currently using in Thailand to support the initial screening stage of victim identification. The report also discusses the potential use of machine learning to draw on the responses to the screening interviews and to predict changing patterns of exploitation. We reflect on this proposed system, to understand the human rights implications that this new technology would include. While there is an obvious implication on workers' right to privacy, we describe steps taken to minimise this imposition. We also advocate the use of the system to support the fundamental human rights of workers who are currently trapped in exploitative work situations.

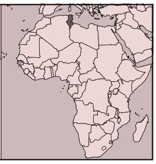
Action steps

We suggest the following steps for civil society organisations who are considering (or are using) AI systems:

- Consider AI as a tool to complement existing efforts and capacity, rather than as a solution in itself.
- Adopt a human rights-based approach to evaluating AI systems, which considers the positive and negative impacts of an innovation prior to rollout.
- Share your stories that reflect on the impact of AI on human rights in order to broaden the types of voices that are included in the global discourse.
- Ensure data privacy and protection are given adequate consideration in the design and development of an AI system: both the raw data itself, but also the predictions that the system generates.

[19] Perry, W. L., McInnis, B., Price, C. C., Smith, S., & Hollywood, J. S. (2013). Op. cit.

TUNISIA



AN ARTIFICIAL INTELLIGENCE REVOLUTION IN PUBLIC FINANCE MANAGEMENT IN TUNISIA

iGmena
Hamza Ben Mehrez
www.igmena.org

Introduction

Public sector institutions in Tunisia are open to harnessing the potential of artificial intelligence (AI) to promote policies for sustainable and equitable development, without forgetting the challenges posed by this emerging technology. In 2015, the government embarked on a series of public sector reforms to improve government operations and to meet the needs of citizens. It also recognises the need for further public sector reforms through the implementation of an accountable public finance management information system (PFMIS), and is proposing the introduction of AI in the current financial system, which has been identified as one of the areas most vulnerable to corruption.[1]

Public sector reforms and AI

In 2016, the government set out its vision for AI, among other imperatives, in a strategy document detailing a five-year development plan for Tunisia (2016-2020). This strategy document was later supplemented by the government's economic and social roadmap for 2018-2020.[2]

The economic and social roadmap for 2018-2020 seeks to accelerate the reforms started under the five-year development plan two years earlier. The aim of the development plan is to guarantee human rights, as well as social and economic growth in Tunisia.[3] Furthermore, the UNESCO Chair on Science, Technology and Innovation Policy, in partnership with the National Agency for Scientific Research Promotion, set up a task force to unlock Tunisia's AI potential.[4] The primary goal of the task force is to facilitate the emergence of an AI ecosystem that acts as a strong lever for equitable, sustainable development and job creation.

The Ministry of Finance is in the process of undertaking fundamental reforms in public finance management. In tandem, it is revising and improving its legal framework governing financial controls and improving transparency policies in managing public finances.[5] To complement these reforms, the government financial system also seeks to integrate its disparate financial and accounting systems to improve administrative and budgetary transparency, as well as to adapt its control structures to a more systematic approach.

However, the pace of AI use has been slower than anticipated. Areas such as multi-year fiscal planning, managing investments, public financial accounting and reporting require a complex AI infrastructure that takes time to set up.

AI application in the public finance management information system (PFMIS)

The PFMIS consists of core sub-systems that provide the government with the necessary information to plan, execute and monitor public finances.[6] The scope and functionality of the PFMIS include fraud detection, budget efficiency and financial analytics.[7]

By leveraging a mix of machine learning, big data and natural language processing techniques, AI is helping the auditors and finance officials at the Ministry of Finance deal with the massive amounts of data that need to be processed in line with the transparency and accountability requirements of their fiduciary responsibilities to the Tunisian

[1] National Agency for the Promotion of Scientific Research. (2018). *National Artificial Intelligence Strategy: Unlocking Tunisia's capacity potential*. www.anpr.tn/national-ai-strategy-unlocking-tunisias-capabilities-potential

[2] Bennani, F. (2018). *Tunisia – Innovative Startups and SMEs Project (P167380): Concept Project Information Document*. The World Bank. documents.worldbank.org/curated/en/133791541834427775/Concept-Project-Information-Document-PID-Tunisia-Innovative-Startups-and-SMEs-Project-P167380

[3] World Food Programme. (2017). *Draft Tunisia Country Strategic Plan*. docs.wfp.org/api/documents/40c35cf3-4055-4588-9c89-e3e59ac6a483/download

[4] Tim, D. (2018, 28 June). An Overview of National AI Strategies. *Politics+AI*. https://www.medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

[5] OECD. (2016). *Open Government in Tunisia*. https://www.oecd-ilibrary.org/docserver/9789264227118-en.pdf?expires=1559928898&id=id&accname=guest&checksum=AE439C572C8737A9C6F792AC3177604C

[6] Kanzari, R. (2018, 11 June). Programme-based budgeting reform in Tunisia – Lessons learned. *CABRI*. https://www.cabri-sbo.org/en/blog/2018/programme-based-budgeting-reform-in-tunisia-lessons-learned

[7] Underwood, C. (2019, 2 February). Machine Learning for Fraud Detection – Modern Applications and Risks. *Emerj*. https://www.emerj.com/ai-podcast-interviews/machine-learning-fraud-detection-modern-applications-risks

taxpayers, which in turn will help restore the public's trust in the government.

Below I consider the application of AI in the detection of fraud, budget efficiency, and financial analytics, and how AI applications can restore citizen trust in the Ministry of Finance through creating fiscal accountability that guarantees development, social justice and human rights.

Detection of fraud

Tunisia is the only country in the Middle East and North Africa (MENA) region to publicly publish its annual budget and audited financial reports.[8] This approach has been attributed to the open government movement and has been seen by many as a citizen-inclusive approach to solving the problem of fraud and corruption in the government, and of promoting transparency and accountability.[9] Releasing financial data to the public should make it possible for Tunisian citizens to dig into data and find errors and mistakes, or instances of fraud, and share the burden of analysis with each other.

However, this approach was not able to put an end to the public's mistrust of government officials. Missteps in areas such as the standardisation of data formats, the application programming interface (API) and the frequency of updates to the data sets have limited the potential for analysis by citizens.[10] In addition, when larger data sets are released, Tunisian citizens do not have the capacity to perform a full analysis of the data in a single pass. For example, Microsoft Excel, the world's most widely available financial analysis tool, has a million-row limit for data processing, while other available data science and accounting tools and resources are out of reach of the majority of Tunisian citizens. The data might be readily available, but the professional tools and skills are not.[11]

The Ministry of Finance is meanwhile planning to set up a fraud hotline to uncover corruption in public finance. AI will help the ministry to review the tip-offs received through the fraud hotline, given the requirement that citizens who report on corruption anonymously have to be convinced to surrender their anonymity to prove the claim being made. Upon receiving a tip, an AI tool can be directed to review any claim made that includes financial data. This would allow the financial data to speak for itself, relieving Tunisian citizens of having to reveal their identity early on in the investigative process, and not resulting in them being unnecessarily exposed.

Budget efficiencies

AI has the ability to dramatically improve the efficiency of developing the national budget, and its usability,[12] by ingesting large amounts of financial data from different sources. The Ministry of Finance is also testing the use of AI in risk assessments of all transactions against current and past data. In addition, the ministry is producing reports that allow auditors and financial officers to gain better insight into the use of financial data and take corrective actions accordingly.[13]

With the application of AI, the ministry could load and analyse financial data for public review, thereby applying the open government standard of transparency. The algorithms they are using could also be made available for public review.[14] In addition, the application of AI in the finance department will help to direct the limited resources to the departments that need them most, instead of burning resources using round-robin audit approaches and a random sampling of transactions for review.[15]

Financial analytics

The Ministry of Finance is testing the use of both supervised and unsupervised algorithms to categorise all the financial transactions loaded onto its financial platform.[16]

Supervised algorithms are based on trained data using known patterns of fraud that are provided by forensic accountants.[17] Unsupervised algorithms are special, because they are developed to allow the data to speak for itself, meaning that transactions are clustered into neighbourhoods of

[8] Trabelsi, K. (2014, 5 February). Tunisia's Citizens Budget: One More Step Toward the "Open Budget". *International Budget Partnership*. https://www.internationalbudget.org/2014/02/tunisias-citizens-budget-one-more-step-toward-the-open-budget/

[9] https://www.huffpostmaghreb.com/2014/01/16/tunisie-open-government_n_4608045.html

[10] https://www.webopedia.com/TERM/A/API.html

[11] https://www.educba.com/financial-analytics

[12] How spending occurs, and how this can be evaluated, reviewed, etc.

[13] Bisias, D., Flood, M., & Valavanis, S. (2012). A Survey of Systemic Risk Analytics. *Annual Review of Financial Economics, 4,* 255-296. https://www.annualreviews.org/doi/abs/10.1146/annurev-financial-110311-101754

[14] Open Government Initiative. (2018). *Driving Democracy Forward: Year in Review 2018*. https://www.opengovpartnership.org/ogp_annual-report-2018_20190227

[15] Macfarlane, A. G . (2016, 14 November). Using the Round Robin Method for Efficient Board Meetings. *Jacobson Jarvis*. https://www.jjco.com/2016/11/14/using-round-robin-method-efficient-board-meetings

[16] Brownlee, J. (2016, 16 March). Supervised and Unsupervised Machine Learning Algorithms. *Machine Learning Mastery*. https://www.machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms

[17] Craig, J. (2019, 29 April). How AI restores the public's trust in the fiscal accountability of governments. *MindBridge*. https://www.mindbridge.ai/ai-restores-public-trust-in-fiscal-accountability-of-governments

numbers that are interesting to accountants, for example, when they make rare connections between two accounts.[18]

These algorithms can also help the ministry to identify transactions that fall outside neighbourhoods of numbers, called outliers.[19] Both supervised and unsupervised algorithms run with standard accounting rules and statistical techniques, such as Benford's Law,[20] which allows every transaction in a financial ledger to be scored.[21] The data is also indexed for rapid search capabilities.[22]

Training in AI

The United States Agency for International Development (USAID) has funded a project called Fiscal Reforms for a Strong Tunisia (FIRST).[23] The FIRST project supported the Ministry of Finance in enhancing its capacity to develop and deliver tax policy and other fiscal reforms using AI and general algebraic modelling system (GAMS) software.[24] Ministry of Finance officials were trained in using the computable general equilibrium (CGE) model.[25]

Restoring public trust

The inability of the Ministry of Finance's audit departments to analyse 100% of its financial data has been a major factor in its inability to spot financial anomalies. While fraud hotlines and open data techniques are a step in the right direction, AI offers an opportunity for the ministry to actively pursue the detection of financial anomalies before even whistleblowers need to act.

Private sector audit firms in Tunisia are already turning to platforms developed by start-ups specialising in AI technology development, such as InstaDeep. InstaDeep raised USD 7 million in funding from AfricInvest and Endeavor Catalyst to expand the use of AI in the public sector by delivering AI products and solutions.[26]

The ministry has also used an AI auditor in their accounting system. This is a promising measure to help Tunisian taxpayers advocate for their human rights for an accountable and transparent tax collection system. Citizens are already becoming aware of the benefits of AI in Tunisia, and are demanding the use of AI at all levels of government to help in the detection of fraud, errors and omissions in financial data and tax collection procedures.

While the adoption and implementation of reforms using AI in public finance has been slow, there is a hope that the government of Tunisia will conduct public consultations with respect to the modification of policies, laws and regulations concerning the use of AI. This is also an opportunity for civil society and the private sector to have their say on the development of new AI policies and approaches in the country.

Action steps

The following steps are necessary in Tunisia to support the deployment of AI technologies:

- The government should help AI start-ups raise funds to boost their role in using different advanced machine-learning techniques, including deep learning. Funding will also help them to scale their AI developments and take their products to market.
- The private sector can offer a set of different AI products and solutions including optimised pattern-recognition, GPU-accelerated analytics, and self-learning decision-making systems. AI solutions are currently being used in different industries including logistics, automation, manufacturing and energy.
- Civil society organisations should participate in cutting-edge research in AI in order to encourage a human rights focus to the development of AI. Civil society can also raise awareness of the benefits and challenges in using AI in both the public and private sectors.

[18] Marr, B. (2017, 7 July). Machine Learning, Artificial Intelligence – And The Future of Accounting. *Forbes*. https://www.forbes.com/sites/bernardmarr/2017/07/07/machine-learning-artificial-intelligence-and-the-future-of-accounting/#7185fcf92dd1

[19] BBVA. (2018, 3 July). Five contributions of artificial intelligence in the financial sector. *BBVA*. https://www.bbva.com/en/five-contributions-artificial-intelligence-financial-sector

[20] Benford's Law can often be used as an indicator of fraudulent data.

[21] https://www.marutitech.com/ways-ai-transforming-finance/

[22] Statistical Consultants Ltd. (2011, 14 May). Benford's Law and Accounting Fraud Detection. https://www.statisticalconsultants.co.nz/blog/benfords-law-and-accounting-fraud-detection.html

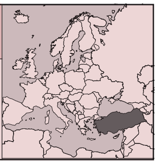
[23] https://tn.usembassy.gov/embassy/tunis/usaid-tunisia/economic-growth/fiscal-reform-for-a-strong-tunisia-first

[24] Williams, R. N. (2019, 1 February). Making faster decisions with AI. *Financial Management*. https://www.fm-magazine.com/issues/2019/feb/make-faster-decisions-with-ai.html

[25] https://tn.usembassy.gov/embassy/tunis/usaid-tunisia/economic-growth/fiscal-reform-for-a-strong-tunisia-first

[26] Jackson, T. (2019, 10 May). Tunisia's InstaDeep raises $7m funding to expand AI in Africa. *Disrupt Africa*. www.disrupt-africa.com/2019/05/tunisias-instadeep-raises-7m-funding-to-expand-ai-in-africa

TURKEY



AI AND TAKEDOWNS: A CHALLENGE TO MEDIA FREEDOMS IN TURKEY

Hun Consultancy
Gürkan Özturan
gurkhan@gmail.com

Introduction

In an atmosphere where 200 media organisations have been shut down, 170 journalists imprisoned, and where no nationally operating news agency is left with 96% of mass media under direct or indirect government control, Turkey continues its downward spiral on press freedom indexes.[1] With restrictions targeting freedom of expression and the right to access accurate, reliable and verified information, many have turned to the internet as a way to circumvent censorship. However, the increasing use of artificial intelligence (AI) for the purposes of detecting copyright infringement has had a negative impact on the right to access information. This report explores this issue in a country that, alongside Russia, has one of the highest number of internet content takedown requests in the world.

Background

While journalism in Turkey has a controversial history and independent media has always been targeted by the authorities, in the last three decades, through the introduction of the internet, a multitude of new ways to overcome restrictive measures have emerged. In 1995, when the pro-left *Evrensel Daily* started publishing its news on a website, the full potential of a critical media in Turkey had never been tested; it had always relied on distribution channels approved by the government.

Similarly, in the early 2000s when alternative media and progressive journalists in Turkey started opening accounts on social media platforms, they had discovered a new medium of reaching far more people than they previously could through print or even through the web versions of newspapers. A few years later, when these platforms were still only starting to attract internet users in the country, the Turkish government had already passed Law No. 5651, the Internet Regulations Act (2007),[2] which was updated in 2017.[3] While there is no specific clause on copyright infringement in this law, the claim of copyright infringement has recently become a way to put pressure on newly emerging critical media organisations publishing online in Turkey.

This new method of censorship applies the United States (US) standard on copyright protection, the Digital Millennium Copyright Act (DMCA).[4] The DMCA has been in effect since 1998 and is divided into five chapters, including the Online Copyright Infringement Liability Limitation Act, which creates limitations on the liability of online service providers for copyright infringement. Despite it being the legislation that is used in the US, a news organisation operating outside of the US, while using social media services that originated in the US, can be subjected to the legal responsibilities outlined in this act.

Encyclopaedia Britannica states that modern copyright was not issued in order to protect the authors' and publishers' rights, but to allow government oversight on the content published within their dominion.[5] While it is claimed that contemporary legislation on copyright is aimed at protecting the intellectual property and usage rights of legal rights-holders, it is still possible to make use of copyright infringement claims to prevent publications from circulating content critical of authorities. In an offline world (until about the 1990s, although as recently as 2015 in Turkey), if a pro-government publisher was instructed not to allow a certain article to be published, they would buy the extensive copyright of this material but never publish it, thus preventing it from reaching larger audiences. The legislation that allowed this was first contested in Sweden on 2 December 1766 with the world's first Freedom of the Press Act that allowed the publication and circulation of public documents, free from copyright restriction, for news purposes.[6]

Today in the digital age, the practice of restricting the circulation of content that is in the public interest using copyright law takes a new

[1] https://rsf.org/en/turkey

[2] Law Number 5651: Internet Regulations Bill (2007). www.resmigazete.gov.tr/eskiler/2007/05/20070523-1.htm

[3] Law Number 5651: Internet Regulations Bill (2017). www.resmigazete.gov.tr/eskiler/2017/04/20170411-3.htm

[4] https://www.copyright.gov/legislation/dmca.pdf

[5] https://www.britannica.com/topic/copyright

[6] https://sweden.se/society/20-milestones-of-swedish-press-freedom

form resembling the older methods of censorship. As costs for conventional media platforms (paper, radio, television) continuously increase and news organisations either sell their businesses to pro-government investors or close down and lay off journalists, digital media gains even more significance for independent media organisations and journalists. In this context, modern copyright laws are being instrumentalised to allow a new level of censorship to prevent independent media from operating online in Turkey.

Independent media targeted through copyright violation complaints

In October 2013, activists from the video collective "Seyr-i Sokak", working in the Turkish capital Ankara, filmed multiple far-right groups assaulting university students at Hacettepe University and published the footage on their YouTube channel as a news piece, sharing it with news networks using the title "Fascist assault on Hacettepe University's Beytepe Campus".[7] Although the video was aimed to inform society of a violent assault against students, it was taken down due to copyright infringement. As one of the activists of the collective explained, YouTube had received multiple complaints from the people featured in the video who stated that they had been filmed without their consent, and would like the video to be taken down; as a result, it was taken down, with Seyr-i Sokak being given a warning against unauthorised use of video content on the social media platform. (Later, in 2015, the same video was uploaded to another social media content platform, Vimeo.)[8]

The period when the assault featured in Seyr-i Sokak's video took place was only months after Turkey's greatest social protest in its history, the popular Occupy Gezi protests, which were suppressed violently as the governing Justice and Development Party, AKP, wanted to hold its parliamentary majority intact ahead of the March 2014 municipal elections and August 2014 presidential elections. Simultaneously, waning public support for the AKP had resulted in increased pressure targeting democratic civil society organisations and independent media.

Four years and five elections later, Turkey was to hold yet another election, this time for the biggest city in Europe, Istanbul. On 31 March 2019, Turkey held municipal elections across the country which resulted in the governing AKP losing the biggest metropolises in Turkey, including cities with the highest average young population, higher average income, and higher industrial production. In Istanbul, the governing AKP declared the opposition won through election fraud, contesting the results for the top seat of the city and as a result, on 6 May the Supreme Election Council cancelled the opposition's victory, declaring that a repeat of the elections in Istanbul would be set for 23 June 2019. This instantly ignited mass protests across the country's largest city.[9] Independent media covered the protests that continued for days after the cancellation of the election results in Istanbul, and "election safety networks" were created through the mobilisation of over 100,000 people, with the aim of ensuring that the elections were conducted safely and that news coverage of the elections continued, so that claims of election fraud could not be made again.[10] The media focus of this network was taken up by citizen journalism-focused news agency dokuz8NEWS, which also ran Regional Media Coordination (the entity working with journalists from the different regions in Turkey) and brought in dozens of local journalists to cover the new round of Istanbul elections.[11]

While dozens of local journalists from across Turkey arrived in Istanbul, several warnings arrived in dokuz8NEWS' inbox with a takedown notice from the DMCA, accusing dokuz8NEWS of unauthorised use of content. The content in question was a video of two Turkish spectators during the Madrid Open tennis tournament, shooting a video of themselves shouting the slogan of the opposition candidate.[12] While the original account that shared the video still contained the content – even at the time of writing of this report – dokuz8NEWS' publication of it as a news piece with a reference to the original publication got caught up in the algorithm detecting copyright infringement and notified ATP Media, which holds exclusive rights to broadcast tennis games.[13] Despite this information being

[7] https://vimeo.com/123256268

[8] Ibid.

[9] dokuz8NEWS. (2019, 6 May). Thousands of people are marching on Istanbul's Kadıköy at Bahariye Avenue protesting Supreme Election Council's decision to repeat municipal elections in Istanbul. https://twitter.com/dokuz8_EN/status/1125493963069317120

[10] T24. (2019, 8 May). Istanbul Volunteers: Around 100,000 people want to assume responsibility for Istanbul elections. https://t24.com.tr/haber/istanbul-gonulluleri-yaklasik-100-bin-kisi-istanbul-seciminde-sandiklarda-gorev-almak-istedi,820181

[11] dokuz8NEWS. (2019, 2 January). Regional media organizations declared united front ahead of municipal elections. https://dokuz8haber.net/gundem/yerel-medya-kuruluslarindan-secimlere-yonelik-guc-birligi-secim2019-yerel-medya-koordinasyonu-kuruldu

[12] Oymak, A. (2019, 7 May). Son nefesimize kadar... Madrid'den selamlar#her eyçokgüzelolacak #MMOpen @ekrem_imamoglu. https://twitter.com/alperoymak/status/1125842537124696070

[13] dokuz8NEWS. (2019, 21 June). Dokuz8NEWS' Flagship Twitter Account @dokuz8haber Suspended & Later Reinstated. https://dokuz8haber.net/english/dokuz8news-flagship-twitter-account-dokuz8haber-suspended

newsworthy and being shared by thousands of Twitter users in Turkey, dokuz8NEWS received a takedown notice, which led to a suspension of its Twitter account – which it uses to reach most of its audience – only three days before the elections, on 20 June. Even though the account was later reinstated in less than 15 hours by Twitter, after widespread calls for restoring the account by the International Press Institute (IPI),[14] the European Centre for Press and Media Freedom (ECPMF),[15] ARTICLE 19,[16] Reporters Without Borders (RSF) Turkey,[17] PEN Norway[18] and OSCE Representative of Freedom of the Media Harlem Desir,[19] as well as many of their readers, the network nonetheless lost a whole day of preparations ahead of the critical elections for Istanbul's top position.

Apart from political content, which may receive complaints from government-linked troll accounts, another dokuz8NEWS publication triggered DMCA takedown notices. The publication in question was a news piece promoting the teaser for the internet series "The Society", distributed by the publisher Netflix. The series features the song "Bury A Friend" by Billie Eilish in its soundtrack. Universal Music Publishing Group (UMPG), which holds the rights to the song, must have signed an agreement with Netflix allowing them to use the song for the promotion of the series, including for media promotion ahead of the release of the series. As a result of the DMCA takedown notice, this content has been removed from dokuz8NEWS following a complaint from Universal.[20]

At the same time, another independent news network, Ileri News, was also notified of copyright infringement. On 13 June, Ileri News' Twitter account was suspended for publishing a news story with video content showing an excerpt from Islamist theologian Nurettin Yıldız's talk in 2013, which is published on the Social Fabric Foundation's YouTube channel.[21] In the video, Yıldız promotes marriage with little children, suggesting "there should be no problem in marrying a six-year-old girl." Yıldız filed complaints against the media organisation, stating that this was unauthorised use of his copyrighted material and should not be allowed, even though he has lost multiple court cases in Turkey against the same news organisation in his attempt to punish journalists for covering his talks.

Yıldız's complaints resulted in the removal of the videos from Ileri News' Twitter account, and its account being suspended.[22] While the account was reinstated on the same day as dokuz8NEWS' account following the outcry by international media freedom organisations, it had stayed closed for a week, causing the news organisation to also lose time, resources and audiences prior to the Istanbul elections. Today, if one is to search Ileri News' Twitter account with an intention to access the independent media outlet's archive of videos on the popular social media platform, multiple videos and content cannot be accessed as a result of complaints and content being removed.

Conclusion

Even though AI can be used in many creative and productive ways, the use of AI for copyright protection limits people's right to information, freedom of the media, and the circulation of content online. In all three examples cited earlier, independent news organisations have received complaints and requests for removal of content, and as a result they lost access to their audiences at a time when

[14] IPI. (2019, 20 June). Shocked to see @dokuz8HABER Twitter account suspended. Essential for voters to have access to independent news ahead of #Istanbul revote in #Turkey. We urge @Twitter to reinstate without delay. @dokuz8_EN @obefintlig pic. twitter.com/lE6FSRA5xj. https://twitter.com/globalfreemedia/status/1141698959469096960

[15] ECPMF. (2019, 20 June). @dokuz8HABER's @Twitter account was suspended based on copyright violation complaints. #Socialmedia remains a major channel for dissemination in #Turkey. Just three days before the renewed election in #Istanbul this is a restriction of internet freedom! https://twitter.com/ECPMF/status/1141658266965086208

[16] ARTICLE 19 ECA. (2019, 20 June). The main @twitter account of #Turkey's news agency @dokuz8_EN has been suspended today, just 3 days ahead of the re-run of the Istanbul mayor elections. #MissingVoices. https://twitter.com/article19europe/status/1141663511078232066

[17] RSF Turkey. (2019, 20 June). RSF temsilcisi Erol Öndero lu: @Twitter bir içerikten tüm bir medya hesabına engel getirmek gibi yıllardır ele tirdi imiz prati e ortak olmamalıdır. Türkiye'de bunca idari & yargı sansürü varken böylesi bir müdahaleye ihtiyaç yoktu. Yeniden @dokuz8haber 'e eri im istiyoruz! pic.twitter.com/LCFynnZOlW. https://twitter.com/RSF_tr/status/1141803260044632066

[18] PEN Norway. (2019, 20 June). The Turkish news organization @dokuz8HABER's account has just been suspended by @Twitter based on coyright violations complaints. This happens 3 days before #IstanbulElections. Coincidence? Hardly, this is a restriction of internet freedom. https://twitter.com/norsk_pen/status/1141684118402605056

[19] Desir, H. (2019, 20 June). Terrible decision to suspend news outlet @dokuz8HABER Twitter account. Population of #Turkey needs access to pluralistic news. @Twitter needs to urgently reinstate access. @dokuz8_EN @obefintlig. https://twitter.com/OSCE_RFoM/status/1141772087411040257

[20] The producers use an algorithm to track the spread of their content online, and if the list of where it is published does not match the list of publishers authorised to use the content, they receive an alert. However, when the re-use of content is news-related, this should not be a problem.

[21] https://www.youtube.com/watch?v=J3wBppWjMHo

[22] Ileri News. (2019, 14 June). Nurettin Yıldız who promotes child marriage filed copyright complaint, Ileri News suspended. https://ilerihaber.org/icerik/cocuklar-evlendirilsin-diyen-nurettin-yildiz-telif-sikayetinde-bulundu-ileri-haberin-twitter-hesabi-askiya-alindi-99231.html

citizens' rights to access accurate, reliable and verified information were violated as well. Two of these complaints were the result of the use of AI.[23] In a country where government or pro-government businesses control 96% of the media landscape, leaving a very limited space for independent media to operate, the copyright legislation may serve as a restrictive measure on media freedoms, directly opposing the foundational Press Freedom Act as adopted in Sweden in 1766.

Over the last five years, Turkey has been the top country when it comes to filing content removal requests with social media platforms. According to Twitter's transparency report on country-specific content removal requests filed in the second half of 2018, Turkey and Russia combined had filed 74% of the total global volume of all requests, despite a 44% decrease in Turkey's content removal requests on Twitter.[24] It was also announced in the report that Twitter had filed objections to multiple court orders in Turkey given that the decisions might violate media freedoms. Furthermore, the report suggests that requests from countries other than Turkey increased by approximately 90% since the first half of 2018, showing how global the problem is becoming.

The other side of the coin is that while registered news organisations are striving to produce quality journalism that offers accurate, reliable and verified information, content that is based on manipulative and false information can circulate freely on social media platforms through distribution channels using trolls; and this opens up room for echo-chambers that grow along with social polarisation.

Action steps

The following steps are suggested for Turkey:

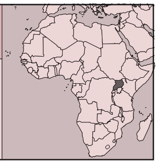
- Social media platforms which target billions of internet users, and have become some of the most popular platforms for media and news distribution, should develop their relationships with global media freedom and journalism associations such as the Committee to Protect Journalists, the International Federation of Journalists, IPI, ECPMF and RSF. This could help to verify independent news organisations that are operating digitally.
- International media freedom organisations should form legal support teams for small independent news organisations so that they have legal representation and support in case of takedown notices being issued for newsworthy material or accounts being closed.
- Transnational news networks among independent news organisations should be established in order to bypass the information asymmetry caused by authoritarian governments' control over mainstream media. This would allow voices to be heard globally that deal with emerging problems through discussion.
- Independent news organisations that are critical of the governments of countries where they are based should adopt a more neutral publication policy and avoid an accusatory tone in their articles when referring to government officials, pro-government people and voters.[25]
- An academic and philosophical debate on the value of intellectual property, its significance to human civilisation, and the impact of copyright in modern culture should be initiated.

[23] The examples of dokuz8NEWS and Ileri News. In the latter case, AI was used by Yıldız's team.

[24] https://transparency.twitter.com/en/removal-requests.html

[25] For Turkish media organisations – and more recently many in the United States – this is a growing problem. More and more journalists seem to be losing their temper in their writing, and some do not even claim to be "independent" or "unbiased" anymore, further isolating the people of "the other side".

UGANDA



“CAMERAS, MOBILES, RADIOS – ACTION!”: OLD SURVEILLANCE TOOLS IN NEW ROBES IN UGANDA

Collaboration on International ICT Policy for East and Southern Africa (CIPESA)
Daniel Mwesigwa
https://cipesa.org

Introduction

Artificial intelligence (AI) has dramatically changed the ways in which machines can “watch”, “listen”, “act” and importantly, “learn”; in other words, how machines can be used in the social, economic and political spheres. In this report, I discuss broadly how neural networks, deep learning and natural language processing (NLP)[1] – all components of AI – are deployed in Uganda by actors in the public and private sectors and international development for different purposes.

To illustrate this, I highlight three pathways of AI deployment in Uganda: the government has recently installed surveillance CCTV cameras in the capital Kampala and surrounding metropolitan areas; a large Ugandan telco is using big data to target subscribers with micro loans via mobile money; and a UN agency is mining radio content to analyse sentiment for policy and planning purposes.

While there are benign intentions and even justifications for the deployment of the aforesaid technologies in the given contexts, I highlight how their implementation might enhance the actors’ surveillance capabilities and result in the abuse of fundamental human rights such as the right to privacy.

I use the labels “big brother”, “big tech” and “big other” to represent the new surveillance tools used by the state, private sector and international development agency respectively.[2]

Background

Uganda is a landlocked country in East Africa bordering Kenya to the east, Tanzania in the south, the Democratic Republic of Congo in the west and South Sudan in the north. It has a population of 37.7 million people according to the bureau of statistics.[3] It has a GDP per capita of USD 604 according to the World Bank.[4] Over 69% of the working population is employed in the agriculture sector,[5] which contributes 25% to the GDP.[6] Most notable is the rise of telecommunications and mobile telephony that have become pervasive nationwide. Internet penetration currently stands at 13.5 million subscribers, a rate of 35%, according to the Uganda Communications Commission.[7]

The new Panopticon

The country has had a long reign of relative peace and stability in the past 33 years, thanks to the ruling President Yoweri Kaguta Museveni who ascended to power in 1986 after years of guerrilla warfare against the sitting governments of the time. The country has since then registered considerable socioeconomic growth but now faces the state’s growing assertiveness in managing security threats, perceived or real, and the fast-rising opposition to the incumbency. In this context, the war on terror and dissent are conveniently categorised under the rosy covers of “national security”.

A spate of extrajudicial killings and untold assassinations of high-profile Ugandan citizens, including Muslim clerics, military and police officers, among others, led the government to expedite the procurement of 24-hour CCTV cameras in crime-prone areas in Kampala and surrounding areas. The first phase of installation comprised 1,940 cameras

[1] In *AI Superpowers*, Kai-Fu Lee explains that neural networks work based on the amount of test data fed to them, upon which the networks themselves identify patterns within the data. Deep learning entails more efficiently trained layers in neural networks. Deep learning’s most natural application is in fields like insurance and loans, where relevant data on borrowers is abundant (credit score, income, recent credit card usage), and the goal to optimise the use of this data is clear (i.e. to minimise default rates). Both terms could be used interchangeably for the first two CCTV and micro loans examples. Lee, K. (2018). *AI Superpowers: China, Silicon Valley, and the New World Order*. Boston: Houghton Mifflin Harcourt.

[2] Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology, 30*(1), 75–89. Shoshana Zuboff frames the tension about big tech’s overtures as “surveillance capitalism”, where users and even non-users are surveilled largely for commercial purposes. In this article, I use big tech and big other independently of each other.

[3] UBOS. (2014). National Population and Housing Census 2014. https://www.ubos.org/onlinefiles/uploads/ubos/NPHC/CENSUS%20FINAL.pdf

[4] https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations=UG

[5] https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS?locations=UG

[6] https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?locations=UG

[7] https://twitter.com/UCC_Official/status/1088826901702078466

in the capital in 2018.[8] However, the planned second phase will cost the public purse over USD 104 million in a procurement deal rushed through parliament under unclear terms.[9]

The alchemists

As fable goes, alchemists of yesteryear tried to reverse engineer precious metals through a combination of different elements. However, most of their efforts were in vain. Today, mobile operators are the new alchemists. Just like the alchemists, telcos are collating precious registration data alongside other data that would have been considered useless decades ago (when machine learning was not as advanced as today). They have mastered the art of extracting value out of this mix of data.

Telcos have benefited immensely from government regulation and the loopholes in regulation. For example, the mandatory SIM card registration in Uganda entailed the collection of biometric data and personally identifiable information in spite of a lack of sufficient constitutional guarantees on data protection and privacy. The evidence shows that the main reason for this mandatory registration – crime – has not been fully addressed, if at all.[10]

This kind of registration data, supplemented through often opaque data agreements with data brokers, is the juicy stuff that telcos – and others – are leveraging to build deep-learning models for their product offerings. For example, MTN Uganda in partnership with IBM rolled out a mobile loans programme on the country's largest network. IBM uses customer and economic data to model customer behaviour and risk. They have built credit scoring models based not only on MTN's data but also on national identity card data. This sort of targeting has become so precise that telcos and tech giants like IBM are not just the new alchemists, extracting value *ex nihilo*, but also "surveillance capitalists".

The "benign" surveillance of radio

In Uganda, radio is the most pervasive form of media. There are over 117 radio stations today[11] – up from only one in the 1990s. More than 400 licences have been granted to FM radio operators. At least 98% of the population listen to radio once in a year. Contrasted against other forms of media such as newspapers (100,000 copies circulated per week), TV and the new kid on the block, the internet (13.5 million subscriptions), radio's reach remains unmatched.

However, there has been so little innovation happening in the radio space that even broadcasting radio over the web could well pass as groundbreaking. In this context, the UN Global Pulse's radio analysis tool promises much.[12] It is an NLP machine that identifies key words which are then analysed to determine "sentiment" and enable the UN to chart advocacy efforts or even policy interventions. Good intentions aside, there are pertinent questions that the radio content analysis tool raises, and which possibly might open a can of worms, especially when it comes to bias or data abuse.

Big Brother Uganda

The Constitution of Uganda emphasises that the government through its agencies must guarantee security of persons and property. The issue of national security gives the government powers among other things to surveil, monitor and intercept communications, and track movement as it deems fit, in order to secure the country's territorial integrity against internal and external aggression. Through different enforcement mechanisms, the government's mandate to facilitate state surveillance is enabled by the following select laws: the Anti-Terrorism Act, 2002; the Regulation of Interception of Communication Act, 2010; the Anti-Pornography Act, 2014; the Communications Act, 2013 (amended 2017); and the Data Protection and Privacy Act, 2019.[13]

The ubiquity of mobiles and motorcycles (also known as *boda bodas*) marked the turn of the last decade in Uganda. Both have been used to do as much good as they have been used to do harm. For example, they have enabled last-mile communication and cheap transport to hard-to-reach places such as rural areas. But a great deal of the runaway crime in the country that pre-empted mandatory SIM card registration among other interventions by the government was committed by hitmen using unregistered mobiles and *boda bodas*.

[8] The Independent. (2019, 15 April). More than 1900 CCTV cameras installed in Kampala. *The Independent*. https://www.independent.co.ug/more-than-1900-cctv-cameras-installed-in-kampala

[9] Misairi, T. K. (2019, 26 April). Uganda: MPs Okay Shs386b Loan for City Spy Cameras. *AllAfrica*. https://allafrica.com/stories/201904260210.html

[10] Wanyama, E. (2018, 18 April). The Stampede for SIM Card Registration: A Major Question for Africa. *CIPESA*. https://cipesa.org/2018/04/the-stampede-for-sim-card-registration-a-major-question-for-africa

[11] Kalyegira, T. (2013, 31 December). 20 years of FM radio stations in Uganda. *African Centre for Media Excellence*. https://acme-ug.org/2013/12/31/20-years-of-fm-radio-stations-in-uganda

[12] https://www.unglobalpulse.org/projects/radio-mining-uganda

[13] https://www.ulii.org/ug/legislation/act/2015/2002; https://ulii.org/ug/legislation/act/2015/18-2; www.ug-cert.ug/files/downloads/The-Anti-pornography-act-2014; parliamentwatch.ug/wp-content/uploads/2016/10/The-Uganda-Communications-Amendment-Bill-2016.pdf; https://ulii.org/ug/legislation/act/2019/1

The government's installation of CCTV cameras in the city and surrounding areas is an attempt to curb the spate of assassinations and urban crime, at least according to Museveni.[14] However, there is little to write home about. The high-profile killings continue episodically.

At least one expert has warned that the government intends to adopt more uncanny approaches to address infrastructural deficits including using facial recognition.[15] A government which said it was buying a "porn machine" to monitor and track pornography online (it denied this later), and which then silently deployed an Intelligent Network Monitoring System (INMS) on mobile operators' infrastructure, is surely capable of surreptitiously deploying any technology as long it justifies the end.[16] Evidence in countries that have poor human rights records such as Ethiopia, Angola and Zimbabwe shows how governments are rushing to secure facial recognition to manage "traffic" and foster "social cohesion".[17] Uganda is no exception.

Big tech

Mobile telecommunications are an essential mark of communications in modern society including the developing world. The advancements in value-added services means that value is no longer confined to traditional offerings such as voice/text and data. In fact, ancillary services such as mobile money are said to be the future of mobile operators. With the growth of targeted approaches through deep learning, MTN Uganda and IBM debuted a mobile loan application dubbed "MoKash".[18] Other telcos such as Airtel Uganda, microfinance institutions and shiny upstarts have launched related programmes; however, MTN's offering is noteworthy because it has scale, reach and first-mover advantage.[19]

MoKash utilises subscribers' usage patterns and histories to weight appropriate credit scores. As mentioned, this is then coupled with biometric data from the national ID database.

Deep learning tools employed by the telco are extractive technologies; for example, they are efficient in nudging subscribers to take micro loans which attract exorbitant annualised repayment rates. But since the learning happens within the telco's ecosystem without knowledge of extant factors such as a customer's participation in the informal economy, the loans algorithm potentially leaves out subscribers who would have otherwise qualified. Others, meanwhile, are manipulated into taking loans to satiate a craving rather than out of necessity. The massive troves of data collected by telcos are subject to abuse. The complex web of players in the black market who offer data brokerage services is growing wider by the day.

Big other

Following the privatisation of the economy in the early 1990s, radio emerged as a medium of choice for many wishing to enter the media sector. It is available in the majority of spoken languages in Uganda. But radio station ownership is dominated by political and religious groups. These circumstances mean that state actors, especially security agencies and the communications regulator, have kept radio programmes under their cross-hairs despite considerable development of other forms of mass media such as broadcast television and the internet.

It is not uncommon to find unsuspecting people intently listening to radio especially on *boda bodas* or in local public places. More than 25,000 people call in to radio stations every day. Local programming is hyper localised and often entails relevant grassroots political and educational talk shows. Meanwhile, there are known lobby groups of radio callers in Uganda that corral and attempt to influence conversation on topics *du jour*.

The public nature of radio means that people do not provide their consent when data processors such as the UN Global Pulse's radio tool are used.[20] The automated speech recognition tool listens to dozens of radio stations simultaneously, flags relevant content when specific keywords are mentioned and generates transcripts for deeper analysis.[21]

[14] New Vision. (2018, 9 October). Museveni commissions CCTV cameras. *New Vision*. https://www.newvision.co.ug/new_vision/news/1487292/museveni-commissions-cctv-cameras

[15] An ICT policy expert interviewed for this article warned that the adoption of facial recognition for the CCTV project could be in the offing.

[16] Mwesigwa, D. (2016, 26 August). Uganda's 'Pornography-Blocking Machine' Appears To Be Part Of A Darker Censorship Agenda. *iAfrikan*. https://www.iafrikan.com/2016/08/26/ugandas-pornography-blocking-machine-appears-to-be-part-of-a-darker-censorship-agenda/

[17] Gwagwa, A., & Garbe, L. (2018, 17 December). Exporting Repression? China's Artificial Intelligence Push into Africa. *Council on Foreign Relations*. https://www.cfr.org/blog/exporting-repression-chinas-artificial-intelligence-push-africa

[18] https://www.mtn.co.ug/en/mobile-money/banking/Pages/mokash.aspx

[19] Cambridge Centre for Alternative Finance, & MicroSave. (2018). *Fintech in Uganda: Implications for Regulation*. https://www.jbs.cam.ac.uk/fileadmin/user_upload/research/centres/alternative-finance/downloads/2018-ccaf-fsd-fintech-in-uganda.pdf

[20] Muhangi, K. (2019, 4 March). Overview of the data protection regime in Uganda. *New Vision*. https://www.newvision.co.ug/new_vision/news/1495211/overview-protection-regime-uganda

[21] Rosenthal, A. (2019, 18 April). When old technology meets new: How UN Global Pulse is using radio and AI to leave no voice behind. *United Nations Foundation*. https://www.unglobalpulse.org/news/when-old-technology-meets-new-how-un-global-pulse-using-radio-and-ai-leave-no-voice-behind

However, some particular audio segments are queued for human review.[22] Upon analysis, the local government or the UN can gather insights that help inform policy decisions.

The tool is tuned to the central and northern regions of the country, where the Luganda and Luo languages are, respectively, widely spoken (there are over 41 spoken languages in Uganda, according to Ethnologue, which geographically limits the tool's coverage).[23]

While the radio tool adheres to UN privacy and data protection principles, these principles might not be aligned with the laws of the land, particularly the Data Protection and Privacy Act, given that the national laws are often in flux.[24] This potentially could create a diplomatic rift if data is used for purposes that are effectively illegal according to the national data laws.

Conclusion

In this article, I have highlighted three pathways of actual and expected AI deployment in Uganda. We now have the big brother, big tech and big other – the state, private sector and an international development agency respectively.[25]

The government has taken steps to combat threats, real and imaginary, through procurement of advanced technology, among other measures. Chiefly, the procurement of CCTV cameras has raised concerns on data protection and privacy. The infrastructural deficiencies (poor or no street lighting, limited connectivity, low standards of maintenance) undermine potential benefits. However, the state is expected to use AI facial recognition technologies to arbitrarily deal with persons of interest including dissidents. The lack of trust, openness and transparency that shrouds state-led interventions impacts on society negatively. For example, the unexplainable leakage of footage from CCTV cameras in Kampala raises questions on ethical standards and requirements to manage the retrieval, sharing and erasure of public CCTV footage.[26]

"It's now imperative that government engages with citizens on these [surveillance] tools," a respondent to this report said. "We're not talking about the future we want to see. We perhaps do not want facial recognition at this stage."

African writer and political analyst Nanjala Nyabola in a *Financial Times* article said "using technology as a substitute for trust creates this black box. But most of us don't understand how these [AI] systems are built. So what comes out is just chaos." Nanjala also said that telcos are at the stage of a "mass data sweep" in which data about an expanding consumer class is being busily devoured. In our context, big tech products such as MoKash are examples of the extraction and commodification of user data.

Meanwhile, machine learning is not as advanced as we might think. Most of it is labour and sweat. For example, the UN's Global Pulse radio tool methodology is laborious and subject to false positives – it is fraught with problems such as changes in the accents of callers and so-called "serial callers" who regularly phone in to radio stations to support a party, meaning that the data is not representative of the reality on the ground.

The enactment of the Data Protection and Privacy Act in Uganda in February 2019 is a positive step, given that it coincides with an aggressive push for CCTV camera installation and other activities that engender massive data sweeps. While the law has not yet been operationalised, it is important that the state fast tracks its implementation. The responsible state actors should ensure compliance and high ethical standards for data processing, especially in the cases covered in this report.

The government needs big tech – and big other – to help it understand what legislation and policies, including oversight and enforcement mechanisms, are necessary to strengthen the protection of human rights in the rapidly changing digital world.

We have been led to believe that data is the new oil.[27] Instead, we should be challenged to think of meaningful ways we could collectively participate in the data economy where privacy, security and profit are all held in high regard.

[22] An ICT policy respondent interviewed for this article raised concerns about the opacity of the radio tool methodology. They say that transcription is manual. There is no automatic voice-speech synthesis.

[23] https://www.ethnologue.com/country/UG/languages

[24] United Nations Development Group. (2017). *Data Privacy, Ethics and Protection: Guidance Note on Big Data for Achievement of the 2030 Agenda*. https://undg.org/wp-content/uploads/2017/11/UNDG_BigData_final_web.pdf

[25] Although the private sector and development world might have closer overlaps, and therefore the "big otherness" might be used loosely and interchangeably.

[26] https://twitter.com/dispatchug/status/1121455937724788736

[27] Mwesigwa, D. (2019, 8 April). Is data really the new-found oil? *Daily Monitor.* https://www.monitor.co.ug/OpEd/Letters/-data-oil-World-Wide-Web-Google-search-digital/806314-5061646-9hfmesz/index.html

Action steps

The following are the overarching action steps for civil society:

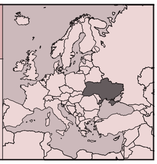
- Civil society, media and academia, among others, should unpack the conversations on AI and reframe them in digestible and locally relevant ways.
- Civil society should work hand in hand with the government to fast track the implementation of the recently passed data protection and privacy law.
- It should contribute to the regular reviews of laws to keep them in line with rapid advancements in technology, particularly AI.
- It should advocate for best and ethical design practices for AI products and services.
- It should act as a watchdog with regard to potential abuses of fundamental human rights through deployment of AI technologies by different actors.

Author's postscript

On 15 August 2019, *The Wall Street Journal* published an investigative piece called "Huawei Technicians Helped African Governments Spy on Political Opponents". The article detailed how Huawei had helped the Ugandan police infiltrate encrypted communication channels used by an opposition leader. Notably, it also mentioned Uganda's plans to open a new six-storey USD 30 million hub in November 2019, which will be linked to the over USD 104 million "Smart Cities" project implemented by Huawei. The project includes surveillance using CCTV cameras equipped with Huawei facial-recognition technology.[28] This information was not in the public domain at the time of writing of this country report. However, the report remains valid and prescient.

[28] Parkinson, J., Bariyo, N., & Chin, J. (2019, 15 August). Huawei Technicians Helped African Governments Spy on Political Opponents. *The Wall Street Journal*. https://www.wsj.com/articles/huawei-technicians-helped-african-governments-spy-on-political-opponents-11565793017

UKRAINE



ALGORITHMIC GOVERNMENTALITY: AI AND E-GOVERNMENT IN UKRAINE AFTER THE 2019 ELECTIONS

University of Greenwich
Maria Korolkova
https://www.gre.ac.uk/people/rep/fach/maria-korolkova

Introduction

This report presents a general overview of the future vision for artificial intelligence (AI) in modern Ukraine. With the 2018-2019 election campaign by the country's new president Volodymyr Zelenskyy – in which a digital agenda was placed squarely in the spotlight – as a backdrop, it considers the potential, needs and human rights implications of making the country a prominent actor in the field of AI globally.

"Artificial intelligence should replace the mentality of officials"

In April 2019, following a successful election campaign largely based on digital tools,[1] a comedian, Volodymyr Zelenskyy,[2] won the Ukrainian presidential elections with an unprecedented 78% of the vote, bringing the promise of much awaited change to the county.[3] Commenters reported that much of Zelenskyy's success is owed to his appeal to the younger generation of Ukrainians, who see digital technology as an everyday necessity, and expect the new leader to embrace the digital future.[4] This was anticipated during the campaign, with Zelenskyy's team – officially called "Ze!Team"[5] – appointing a digital campaign leader right from the start. Mikhailo Fedorov, in charge of the digital campaign, gave several interviews stating that the priority of the new president would be to develop the digital sphere in the country in order to optimise the everyday experience of Ukrainian citizens.[6]

These promises are expected to be kept by Zelenskyy, who sees AI as the best tool to reboot the mind-set of Ukraine. During his official visit to Canada on 2 July 2019, Zelenskyy outlined the country's priorities:

> With the digitalisation of processes, and the implementation of "a state in a smartphone", artificial intelligence should replace the current mentality of officials. It will be used for overcoming monopolies, and fighting smuggling; for protecting property rights, improving the country's credit rating, and attracting large-scale investment. All this should be done in order to improve the living standards of Ukrainian citizens.[7]

However, does the country have enough resources to implement these ambitious strategies, and most importantly, is the new government aware of the negative sides of AI?

[1] For detailed figures see Shishatskii, E., & Yurasov, S. (2019, 26 April). U nas est' proekt Chernaya biblioteka Poroshenko: onlain-strateg Ze [The lead online strategist of Ze Team: We have a project of Poroshenko's black library]. *tech.liga.net*. https://tech.liga.net/technology/interview/pochemu-poroshenko-proigral-intervyu-s-onlayn-strategom-zelenskogo

[2] There are several transliteration options of Volodymyr Zelenskyy's name due to the different spelling of his name in Ukrainian and in Russian, as well as different transliteration styles. In this article I am using transliteration from the official website of the president of Ukraine (https://www.president.gov.ua/en). However, if his name appears in a quote in a different transliteration, the spelling used in the original source is maintained.

[3] The elections took place in an important time in Ukrainian political life, following the legacy of the 2014 Ukrainian Revolution (also known as the Maidan Revolution), the annexation of Crimea, and an ongoing military conflict in the Luhansk and Donetsk regions of the country.

[4] The phenomenon of Zelenskyy's rapid success is of course more complex than just the reference to new technologies. Some suggest that this is a part of a "global trend of rebelling against the government systems [...] when the masses get tired of the old elites and raise populists and other 'friends of the people' to rebel against these systems" [AFRIC. (2019, 22 April). Elections in Ukraine: The Zelensky phenomenon. *AFRIC*. https://afric.online/11555-elections-in-ukraine-the-zelensky-phenomenon]; some root it in the success of the highly popular TV series "Servant of the People", which portrays a school teacher played by Zelenskyy winning the presidential elections [Fisher, J. (2019, 22 April). Zelensky win: What does a comic president mean for Ukraine. *BBC*. https://www.bbc.co.uk/news/world-europe-47769118]. There are other speculations among political analysts. For the purposes of this report, such speculations will be only mentioned marginally, and the report will largely focus on the strategic implementation of AI and its related context during the rule of the new president.

[5] Importantly, the name of the campaign mimicked the branding of the IT giant Apple. Following the branding scheme of iTunes, iPhone, iPad, iWatch, etc., the election campaign introduced Ze!Team, Ze!Academy, Ze!Elections, etc. Interestingly, "Ze" also implies the English definite article "the", thus hinting at the European/international nature of the campaign.

[6] https://www.bbc.com/ukrainian/features-russian-48014443

[7] Author's translation of the original. https://www.president.gov.ua/news/gromadyanin-kliyent-vlada-servis-prezident-u-kanadi-nazvav-p-56169

New president, new government – new (AI) life?

Fedorov, who was appointed as Advisor to the Head of State in May 2019,[8] thinks that there is a great potential for Ukraine to enter the AI world. Ukraine is among the top four countries with the largest number of IT specialists,[9] and in the last four years, according to PwC, the number of IT specialists has more than doubled, from just over 40,000 to nearly 92,000. In terms of available talent, Ukraine already outpaces its competitors in the region, including Poland and Hungary, and PwC believes the number of IT professionals will double again by 2020.[10]

The effectiveness of this potential was clearly demonstrated during the election campaign led by Fedorov and largely based on volunteers. By the time of the elections, Ze!Team had 600,000 followers on Instagram, 500,000 on Facebook and 160,000 on Telegram, with four billion website visits, and more than 18 billion campaigning emails, which is a significant achievement taking into account that they started from scratch just four months before the elections.[11]

Fedorov stresses that AI algorithms were used for detailed analysis of campaign data, resulting in 32 segments of data:

> Based on these segments, we understood who was most interested in us, who wanted to interact with us the most. We identified key segments: IT specialists, mothers, people who supported certain aspects of our programme, and worked with these segments. Plus, we identified people who supported the [then] current government and those who did not. [...] We used geolocation, targeting cities, because the CTR[12] is always higher in cities. [...] We tested a lot.[13]

When asked whether Ze! campaigners were looking into the way Trump had run his election campaign, Fedorov responded that Ukrainian AI specialists used data much better than Trump, and also in a more open manner, noting however that a detailed analysis of both campaigns was yet to be presented.[14]

The use of AI in the Ze! election campaign is somehow illustrative of the situation with AI in the country as a whole – there is a lot of testing, which produces some good results; however, there is no systematic approach to this testing.

In October 2018, Ivan Primachenko, a co-founder of the Ukrainian educational platform Prometheus that provides online courses from top world universities, published an article questioning whether Ukraine was ready to enter the era of AI. His conclusions expressed disappointment in government structures and state universities, which, in Primachenko's opinion, failed to embrace AI technologies, especially at the level of legislation and teaching.[15] The best initiatives in this sphere were still located within private companies and individual initiatives.[16]

However, with Fedorov joining the presidential office, there seem to be radical changes with respect to AI initiatives at the governmental level. The first large presidential initiative is called "State in a Smartphone", an advanced e-government project that would move all government-related services online. Some steps towards this direction have been made in the past four years; but the development strategy is now more ambitious. "The ultimate goal for most services should be full automation, when the decision is made not by the official, but by the system, based on a clear algorithm provided by a regulatory framework," says Oleksiy Viskub, first deputy head of the State Agency for e-Governance.[17]

Importantly, to realise this strategic plan, Ukrainian officials arranged multiple consultations with Estonian e-governance representatives. Over the past 20 years, the Estonian government has

[8] Before this appointment, like Zelenskyy himself, Fedorov had never been in any public or political service. He ran a small digital agency, which was hired to promote Zelenskyy's comedy club "Kvartal 95", later accepting Zelenskyy's offer to run his presidential campaign. https://strana.ua/news/202366-mikhail-fedorov-naznachen-sovetnikom-prezidenta-ukrainy-vladimira-zelenskoho.html

[9] After the United States, India and Russia. See: Ukraine Digital News & AVentures. (2016). *IT Ukraine: IT services and software R&D in Europe's rising tech nation*.

[10] Borys, C. (2018, 18 January). Ukraine's economic secret: 'Engineering is in our DNA'. *BBC*. https://www.bbc.co.uk/news/business-42403024; see also Andrienko-Bentz, O. (n.d.). *Export-oriented segment of Ukraine's IT services market: Status quo and prospects*. EBA and PwC. https://eba.com.ua/static/export_it_industryfinal_29092016.pdf

[11] Shishatskii, E., & Yurasov, S. (2019, 26 April). Op. cit.

[12] Click-through rate (CTR) is the ratio of users who click on a specific link to the number of total users who view a page, email or advertisement.

[13] Shishatskii, E., & Yurasov, S. (2019, 26 April). Op. cit.

[14] Ibid.

[15] Primachenko, I. (2018, 24 October). Voidet li Ukraina v eru iskusstvennogo intellekta? [Will Ukraine enter the era of artificial intelligence?]. *NV.ua*. https://nv.ua/opinion/vojdet-li-ukraina-v-eru-iskusstvennoho-intellekta-2502284.html

[16] To list a few: People AI start-up by Oleg Roginsky with USD 30 billion investment by Andreessen Horowitz; Augmented Pixels by Vitaliy Goncharuk, which also holds the biggest AI conference in Ukraine; Artificial Intelligence Platform within the Everest Innovation Integrator by Yuri Chubatuk; initiatives in education include master classes in machine learning in the private Ukrainian Catholic University; and free online courses in machine learning available on the Prometheus platform.

[17] https://www.president.gov.ua/news/radnik-prezidenta-ukrayini-mihajlo-fedorov-obgovoriv-iz-pred-55621

reached significant milestones in the digitalisation of its services. Having built a society where public and private digital services are woven into the fabric of everyday life, including the introduction of electronic ID cards linked to national registers, the first electronic elections, and 99% of government services online, the Estonian government now plans to build next-generation public services based on AI, according to its National Digital Advisor.[18] Ukraine is ready to join this future, as the two countries are seeking to "deepen cooperation" in the implementation of a new type of e-government.[19]

While it is still unclear which particular AI algorithms from the Estonian experience will be implemented in the Ukrainian e-government programme – a detailed action plan for 2019 and strategic goals until 2024 were presented to the European Parliament on 10 July 2019,[20] but the text has not been made public yet – it is important that the public joins in on the debates on its future.

The first survey on the public attitude towards AI in Ukraine, called "Artificial Intelligence: The Ukrainian Dimension", was conducted by Gorshenin Institute and Everest Innovation Integrator in September 2018,[21] giving promising results, but also outlining concerns. Almost 85% of respondents had heard the term "artificial intelligence", and 74.1% experienced the influence of AI on their life. Half of the respondents said they were interested when receiving information about AI. Finally, and notably, AI caused anxiety and fear in almost 23% of respondents.[22]

According to the survey, the majority of those who welcomed the development of AI in Ukraine considered it capable of replacing humans in dangerous workplaces, as well as increasing the productivity of industrial enterprises. Fewer respondents anticipated that AI would help in extending human life and preventing diseases, or would provide protection against natural disasters, catastrophes, wars and crime. On the other hand, the range of negative consequences was seen as somewhat more serious. The respondents feared that AI could result in oppression, while the most pessimistic predict the possibility of establishing an AI dictatorship and the destruction of human civilisation.

Importantly, some survey questions were related to the use of AI in national and local government. Ukrainians are convinced that technological intelligence can ensure fair elections, reduce the level of bureaucracy, overcome corruption and optimise public spending. In urban areas, respondents said AI could regulate street lighting, traffic and parking, and garbage collection and processing, and monitor public order and the health of the environment.

It would be fruitful to compare these results with the ones from the year-long public debate over the algorithmic liability law in Estonia (a.k.a. the Kratt[23] law), which initiated the "opinion shift toward avoiding sector-based regulation, opting for general algorithmic liability instead."[24] The Kratt law debates generated the important idea of providing algorithms with a separate legal status, similar to companies (a draft bill should enter the Estonian parliament for debate in summer 2019).

As Ukraine starts to have these ethical, moral and philosophical debates on AI, it is important they are discussed in all their complexity, and infused with human rights concerns, with respect to using AI in both the public and private sectors. To quote Yuri Chubatyuk, president of the Everest group of companies, which just launched a large Ukrainian AI platform:

> We are at the stage when it is necessary to discuss an effective public-private partnership, which, with a holistic, deliberate concept of innovation and development, can create the expected technological leap for the country. We need a national strategy in AI development to provide a phased transformation of each industry, especially the educational sector, which directly affects the ability of Ukrainians to compete in the technological market in the near future. We must understand that the process of consolidating efforts in this direction should start now, involving business and research communities, members of the government, politicians, and the public.[25]

Conclusion

With the election of the new president, who included a focus on new technologies as one of the priorities of his programme, Ukraine seems to be

[18] Kaevats, M. (2018, September). AI and the Kratt momentum. *Estonian Investment Agency.* https://investinestonia.com/ai-and-the-kratt-momentum

[19] https://www.president.gov.ua/news/ukrayina-ta-estoniya-pogliblyat-spivpracyu-dlya-realizaciyi-55861

[20] https://www.president.gov.ua/news/yevropejskij-soyuz-pidtrimaye-realizaciyu-koncepciyi-derzhav-56337

[21] https://www.youtube.com/watch?v=AXOxQ5GgglA

[22] Trapeznikova, D. (2018, 19 December). Iskusstvennyi intellect nam pomozhet. [Artificial intelligence will help us]. *day.kyiv.ua*. https://day.kyiv.ua/ru/article/obshchestvo/iskusstvennyy-intellekt-nam-pomozhet

[23] Kratt is a magical creature in Estonian mythology. Essentially, Kratt was a servant built from hay or old household items. The Estonian government uses this character as a metaphor for AI and its complexities.

[24] Kaevats, M. (2018, September). Op. cit.

[25] Trapeznikova, D. (2018, 19 December). Op. cit.

looking towards accelerating the development of AI in both the public and private spheres. By some estimates, the country has good chances of becoming internationally visible in the sector. It does indeed have an impressive base of IT specialists and private initiatives in the AI field; however, this needs to be backed up with coherent governmental policies that allow public-private partnerships.

The positive move is that the new e-government initiative was launched in the first few days of the president taking office; however, it is still unclear what exactly it will entail.[26] Importantly, it will be necessary to invite the public to consider the issue, as well as to think of the human rights implications of any development. Currently, not all regions of Ukraine have sufficient telecommunications coverage. While this particular issue is being addressed,[27] it is important to continue addressing the digital divide in the country, alongside a focus on AI.

While the results of the national survey on AI showed that there was a high level of public awareness of AI in Ukraine, there is less recognition of Ukraine as an AI-progressive country on the international scene. While an annual international conference on AI held in Ukraine plays an important role in developing this visibility,[28] there are still concerns that the country is seen as an underdeveloped state with respect to AI.[29]

Most importantly, while the new president Zelenskyy sees AI almost as a mythological superpower,[30] which "should replace the mentality of officials," it is important to consider the dangers of such "algorithmic governmentality". What we face here is what Antoinette Rouvroy calls – following Foucault – a crisis of the regimes of truth: "To my mind, we are less facing the emergence of a new regime of truth than a crisis of regimes of truth. A whole range of notions are in crisis: the notions of person, authority, testimony."[31]

If Ukraine is going to implement these changes, they will be radical changes in how we see the world, and how the government works. Most importantly, if the implementations of AI and e-government allow the state to know in great detail about a citizen's day-to-day life, the state would no longer need to ask people about their lives, thus dislocating the axis of power in the citizen-state relationship necessary for democracy to function, which might lead to unpredictable consequences.

Action steps

The following are key needs in Ukraine:

- *National AI strategy:* There is a need to develop a national AI strategy to create a common framework for implementation in both the public and private sectors with a specific focus on human rights and the digital divide.
- *Algorithmic identity:* It is necessary to define – legally – what kind of algorithms are used in e-government and private sector initiatives, and who owns them.
- *An effective data protection policy:* How does one ensure the integrity of decision making with algorithms that evolve and change constantly? How can we be sure that sensor data used in algorithms has not been hacked or changed? Estonia's experience in using KSI blockchain technology to secure its citizens' medical records may help.[32]
- *A balanced debate on AI:* There is a need to have an honest, meaningful public debate on the technical and legal aspects of AI, including AI's controversial attributes and threats. Any discussion must involve the public.
- *Visibility of AI:* It is also important to ensure that all AI strategies and initiatives are reported on and critically reviewed by the press and social media, both nationally and internationally. Their purpose and use needs to be clear and publicly known.
- *Education in AI:* There is a need to create AI and machine-learning courses in schools and universities nationwide.

[26] The plan was presented to the European Parliament on 10 July 2019, but the documents have not yet been released to the public.

[27] https://www.president.gov.ua/news/radnik-prezidenta-mihajlo-fedorov-obgovoriv-z-predstavnikami-56201

[28] https://aiukraine.com

[29] The futurist and author of international bestsellers Yuval Noah Harari in a conversation with Mark Zuckerberg put Ukraine in line with Honduras and Yemen when talking about the country's AI development level. See: https://fbnewsroomus.files.wordpress.com/2019/04/transcript_-marks-personal-challenge-yuval-noah-harari.pdf; for a discussion of this mistake in Ukrainian media, see Goncharuk, V. (2019, 4 May). *Ukraina – ne Gonduras: gre nashi mesto v oblasti iskusstvennogo intellekta.* [Ukraine is no Honduras: Where is our place on the AI scene]. *Ekonomicheskaya Pravda*. https://www.epravda.com.ua/rus/columns/2019/05/4/647525

[30] Interestingly, the current major exhibition on AI at the Barbican Centre, London, showcases ancient beliefs, such as myths, magic, illusion and religion, as the predecessors of AI. See: https://www.barbican.org.uk/whats-on/2019/event/ai-more-than-human

[31] Morison, J. (2016). Algorithmic Governmentality: Techo-optimism and the Move towards the Dark Side. *Computers and Law*, *27*(3). https://pure.qub.ac.uk/portal/files/89325400/Algorithmic_Governmentality.pdf

[32] KSI is a blockchain technology designed in Estonia after cyberattacks in 2007 and used globally to make sure networks, systems and data are free of compromise, and have 100% data privacy. See https://e-estonia.com/solutions/security-and-safety/ksi-blockchain and https://guardtime.com

VENEZUELA

ARTIFICIAL INTELLIGENCE AND SOCIAL DEVELOPMENT IN VENEZUELA

Fundación Escuela Latinoamericana de Redes (EsLaRed)
Sandra L. Benítez U.
www.eslared.org.ve

Introduction

The economic, political and social crisis in Venezuela is a national tragedy that significantly affects social development. It is a humanitarian crisis that has resulted in levels of impoverishment unprecedented in the history of the country. In this context, it is necessary to evaluate different ways to overcome the crisis and promote the social development of the country, including through the use of artificial intelligence (AI) and the construction of environments that open development opportunities in critical sectors. Emerging technologies, such as blockchain, chatbots, robotics and biometrics, have been developed in recent years by different actors in the public and private sectors in the country, and show significant promise in addressing many of its socioeconomic needs.

This report considers several initiatives using AI that have been implemented in Venezuela. These include the use of a cryptocurrency, the application of AI in health care, and the use of robotics for surgery and for military needs, among others. The laws, plans and treaties that are relevant to the use of emerging AI technologies are also listed. Finally, recommendations are made to encourage the optimal use of AI for social development in Venezuela, so that the country can return to a path of sustainable prosperity.

Policy and legal framework

Venezuela has a regulatory framework that guarantees the basic human rights of citizens, such as economic, social and cultural rights (ESCRs) and internet rights, and which allows the social development of the country to be strengthened. These rights are enshrined in the Constitution of the Bolivarian Republic of Venezuela (CBRV)[1] in the following articles: 52, 57, 59, 60, 61, 67, 75, 95, 110, 118, 184, 199, 201 and 308 (internet rights), and 3, 80, 83, 84, 85, 86, 305 (ESCRs). The use and management of emerging technologies are part of public policy and are covered by the following laws and plans: a) Organic Law of Science, Technology and Innovation,[2] b) Reform of the Organic Law of Science, Technology and Innovation,[3] c) National Science, Technology and Innovation Plan,[4] d) Law of the Government,[5] e) National Plan of Information Technologies for the State,[6] f) Telecommunications Law,[7] g) Law Against Computer Crimes,[8] h) Law on Data Messages and Electronic Signatures,[9] and i) Law on the Simplification of Administrative Procedures,[10] among others. The use of technology is also included in the following development plans and programmes: the Second National Economic and Social Development Plan 2013-2019,[11] The Homeland Plan 2019-2025,[12] the Economic Recovery Programme for Growth and Prosperity,[13] and the Plan for the Country's Future, which was created by the political opposition.[14] In addition, the National Constituent Assembly proposes adding a clause in the CBRV against the military use of science (warning of the danger of using AI and robotics for military purposes, and the risk that this may imply for people globally).[15]

In recent years, the Venezuelan government launched its own cryptocurrency called Petro and created a series of decrees, regulations,

[1] www.conatel.gob.ve/constitucion-de-la-republica-bolivariana-de-venezuela-2

[2] www.conatel.gob.ve/ley-organica-de-ciencia-tecnologia-e-innovacion-2

[3] transparencia.org.ve/project/reforma-parcial-de-ley-organica-de-ciencia-tecnologia-e-innovacion

[4] www.tic.siteal.iipe.unesco.org/politicas/999/plan-nacional-de-ciencia-tecnologia-e-innovacion-2005-2030

[5] www.conatel.gob.ve/ley-de-infogobierno

[6] www.conatel.gob.ve/pueblo-y-gobierno-potencian-tecnologias-de-informacion

[7] www.conatel.gob.ve/ley-organica-de-telecomunicaciones-2

[8] www.conatel.gob.ve/ley-especial-contra-los-delitos-informaticos-2

[9] www.conatel.gob.ve/ley-sobre-mensajes-de-datos-y-firmas-electronicas-2

[10] www.conatel.gob.ve/ley-de-simplificacion-de-tramites-administrativos-2

[11] plataformacelac.org/politica/232

[12] www.psuv.org.ve/wp-content/uploads/2019/01/Plan-de-la-Patria-2019-2025.pdf

[13] www.minci.gob.ve/lineas-programa-de-recuperacion-crecimiento-y-prosperidad-economica-2

[14] www.elinformador.com.ve/wp-content/uploads/2019/01/Jueves-29nov2018-Presentacion-LVQV.pdf

[15] cienciaconciencia.org.ve/clausula-uso-militar-la-ciencia-aportes-la-asamblea-nacional-constituyente-anc

measures and plans to support the development of AI technologies, specifically: the Petro white paper,[16] Petro regulations,[17] the national cryptoasset plan involving the development of the Petro,[18] the superintendency for the cryptocurrency,[19] the treasury for cryptoassets,[20] regulations for the exchange of cryptoassets,[21] the registration for services in cryptoassets,[22] the exchange for cryptoassets,[23] and the superintendency for digital mining,[24] among others.

In the international arena, Venezuela is committed to promoting social development and the use of technologies in a series of pacts, treaties and declarations, such as: the International Convention on ESCRs,[25] the Declaration on the Right to Development,[26] the Millennium Development Goals (Goal 8: Promote a global partnership for development),[27] the American Convention on Human Rights,[28] the Universal Declaration of Human Rights,[29] and the Declaration on Social Progress and Development.[30] There are also agreements[31] between Venezuela, Palestine, China[32] and Russia[33] on tourism and mining that incorporate the Petro as a currency.

The country is active in global events such as the First International Meeting on Cryptoassets[34] and the St. Petersburg International Economic Forum,[35] where it highlighted the development of the Venezuelan cryptocurrency.

AI applications for social development in Venezuela

Venezuela's economic, political and social crisis is resulting in a significant delay in the development of the country. According to the 2018 report[36] of the National Survey of Living Conditions (ENCOVI),[37] poverty has increased in Venezuela. The report reveals that the "number of poor households in Venezuela rose by two percentage points and stood at 48% in 2018." Furthermore, it says that "vast sectors of the population across the social spectrum have been forced to migrate to seek opportunities in other countries to meet essential needs and generate income that helps sustain the survival of relatives in Venezuela." Similarly, Feliciano Reyna,[38] founder of the NGO Acción Solidaria,[39] says that in Venezuela there is a "complex humanitarian emergency; that is to say, a type of humanitarian crisis that produces a change in the political, economic and social life of a country, and whose characteristic is that it severely affects the population's capacity to survive, and to live with dignity."

The situation is not helped by the serious political conflict in the country, which includes a power struggle between two national assemblies, a presidential crisis,[40] a lack of independent powers of institutions, and an erosion of the rule of law.[41] There are two legislative governmental bodies, the National Assembly (NA)[42] (democratically elected in 2015, but controlled by the opposition) and the National Constituent Assembly (NCA)[43] (promoted by the president in 2017 according to Decree 2.830,[44] with an official majority).[45] These pass resolutions and laws and set the direction of public policies. However, both bodies follow different plans for the country's development – the NA follows its plan for

[16] www.telesurtv.net/news/libro-blanco-white-paper-petro-criptomoneda-venezuela--20180131-0064.html

[17] petro.gob.ve

[18] www.conatel.gob.ve/venezuela-presenta-plan-nacional-de-criptoactivos-en-europa

[19] sunacrip.gob.ve

[20] tcv.com.ve

[21] sunacrip.gob.ve/normativa.html

[22] risec.sunacrip.gob.ve/login

[23] sunacrip.gob.ve/casas.htm

[24] sunacrip.gob.ve/mineria.html

[25] https://en.wikipedia.org/wiki/International_Covenant_on_Economic,_Social_and_Cultural_Rights

[26] https://www.ohchr.org/EN/ProfessionalInterest/Pages/RightToDevelopment.aspx

[27] https://en.wikipedia.org/wiki/Millennium_Development_Goals

[28] https://en.wikipedia.org/wiki/American_Convention_on_Human_Rights

[29] https://en.wikipedia.org/wiki/Universal_Declaration_of_Human_Rights

[30] https://www.ohchr.org/Documents/ProfessionalInterest/progress.pdf

[31] www.efe.com/efe/america/politica/venezuela-y-palestina-llegan-a-acuerdos-sobre-el-petro-turismo-mineria/20000035-3608292

[32] mppre.gob.ve/2018/09/15/china-respalda-programa-de-recuperacion-economica-en-venezuela

[33] www.cripto247.com/altcoins-icos/venezuela-y-rusia-buscan-acuerdo-con-criptomonedas-182186

[34] www.telesurtv.net/news/venezuela-encuentro-internacional-criptoactivos-sunacrip-20181116-0046.html

[35] vtv.gob.ve/venezuela-mecanismos-escala-internacional-petro

[36] elucabista.com/wp-content/uploads/2018/11/RESULTADOS-PRELIMINARES-ENCOVI-2018-30-nov.pdf; elucabista.com/2018/11/30/se-incrementa-la-pobreza-venezuela-segun-resultados-preliminares-encovi-2018; cpalsocial.org/indicadores-de-la-situacion-social-actual-en-venezuela-2784

[37] encovi.ucab.edu.ve

[38] www.diariolasamericas.com/america-latina/estiman-que-venezuela-necesitara-ayuda-humanitaria-tres-anos-recuperar-sus-capacidades-n4176011

[39] www.accionsolidaria.info/website

[40] https://es.wikipedia.org/wiki/Crisis_presidencial_de_Venezuela_de_2019

[41] www.msn.com/es-ve/video/soccer/onu-denuncia-erosi%C3%B3n-del-estado-de-derecho-en-venezuela/vp-AADUph6

[42] https://es.wikipedia.org/wiki/Asamblea_Nacional_de_Venezuela

[43] https://es.wikipedia.org/wiki/Asamblea_Nacional_Constituyente_de_Venezuela_de_2017

[44] www.panorama.com.ve/politicayeconomia/En-Gaceta-Decreto-de-convocatoria-a-Asamblea-Nacional-Constituyente-20170503-0084.html

[45] elmercurioweb.com/noticias/2019/1/21/tsj-legtimo-cataloga-a-la-constituyente-como-rgano-de-facto-e-rrito

the country,[46] and the NCA its homeland plan[47] – which worsens the crisis in the country and limits the effectiveness and sustainability of measures aimed at its recovery.

For example, between 2018 and 2019, laws were passed that supported the creation of the Petro[48] (backed by natural resources such as Venezuelan oil), and established the Comprehensive Cryptoasset System,[49] with the aim of strengthening the Economic Recovery, Growth and Prosperity Programme. The NCA supported these initiatives; however, the NA declared them unconstitutional,[50] since the CBRV defines the Venezuelan currency as the bolivar, and there is no way to change it without reforming the constitution. It is also unconstitutional because, unlike the bolivar, it is not guaranteed by the country's oil reserves.

Faced with these challenges, it is critical to look for alternatives to overcome the crisis so that the country can be put on a path of socioeconomic recovery, and so that the living conditions of citizens, as well as the relations between citizens, groups and institutions that make up the social fabric of society, can be improved. With this in mind, a number of initiatives have emerged using AI technologies:

- *The use of blockchain technology:* In 2018 the Venezuelan government launched its own cryptocurrency called the Petro.[51] PetroApp[52] is an application for the exchange and purchase of goods and services using the cryptocurrency. It revitalises the digital economy in Venezuela, and is backed by a legal framework approved by the NCA. The platform is developed using blockchain technology, and offers the following services:[53] a) the purchase of Petros using other cryptocurrencies (Bitcoin and Litecoin); b) Patria Remesas,[54] which is a platform allowing people in Venezuela to receive remittances in cryptocurrencies[55] quickly, safely and transparently; and c) access to a cryptocurrency savings plan using the Patria platform.
- *Chatbots:* There are many examples of the use of chatbots in Venezuela. For instance, "Mia"[56] is a chatbot set up by the Mercantile Bank, and is a virtual assistant designed to answer frequently asked questions quickly and conveniently; "My Health Insurance Calendar"[57] by Liberty Mutual Caracas allows members to receive information and personalised health reminders; and 3) "Pásalo" ("Pass It"),[58] which is used for secure online payments, and which uses a chatbot.
- *Robotics:* There are also numerous examples of the use of robotics in Venezuela. To name a few, "Da Vinci"[59] is a robot that allows robotic surgery at the Hospital de Clínicas Caracas. It is used by 23 surgeons with different specialties. "Commander IEV01-02"[60] is an autonomous robot developed by students at the National Experimental University of the Bolivarian Armed Forces (UNEFA) that has the capacity to move, on a small scale, solid waste containers from a port to a floating oil platform. "Arpia"[61] is an unmanned aerial vehicle (UAV) that is used to patrol border areas and oil zones. Finally, "ANT-1X (Gavilán)",[62] also an UAV, performs environmental monitoring tasks as well as operating in areas where natural disasters have occurred.
- *Biometrics:* Examples of biometrics use in Venezuela include "BiopagoBDV",[63] a biometric payment system set up by Banco de Venezuela, which allows for the purchase of goods and services, and is used to control the sale of food in supermarket chains and pharmacies; and the Integrated Authentication System (SAI),[64] a biometric

[46] www.elinformador.com.ve/wp-content/uploads/2019/01/Jueves-29nov2018-Presentacion-LVQV.pdf

[47] www.psuv.org.ve/wp-content/uploads/2019/01/Plan-de-la-Patria-2019-2025.pdf

[48] www.mppef.gob.ve/anc-aprueba-decreto-en-respaldo-al-petro-y-demas-criptoactivos/

[49] www.mppef.gob.ve/aprobada-ley-del-sistema-integral-de-criptoactivos-en-la-anc/

[50] www.asambleanacional.gob.ve/documentos_archivos/acuerdo-sobre-la-implementacion-del-petro-191.pdf; www.asambleanacional.gob.ve/noticias/_an-declaro-nula-la-emision-del-petro-y-todas-sus-obligaciones

[51] www.telesurtv.net/news/libro-blanco-white-paper-petro-criptomoneda-venezuela--20180131-0064.html

[52] www.petro.gob.ve/petro-app.html

[53] www.petro.gob.ve/servicios.html

[54] remesas.patria.org.ve/es/login

[55] www.patria.org.ve

[56] www.mercantilbanco.com/mercprod/content/tools/principales/3788_chatbot.html?utm_source=webpage&utm_medium=slideshows&utm_term=home_1&utm_campaign=slideshow_home_3

[57] prevencion.contactamed.com; www.ciberespacio.com.ve/2017/10/software/la-tecnologia-al-servicio-de-los-pacientes

[58] elcooperante.com/asi-funcionara-pasalo-un-sistema-de-pagos-para-el-mundo-hecho-por-venezolanos

[59] www.youtube.com/watch?v=10E2KBPVemo; www.youtube.com/watch?v=CdXp7eomEJQ

[60] lacalle.com.ve/2015/11/23/jovenes-creadores-del-robot-comandante-ievo1-02-resaltan-avances-del-pais-en-soberania-tecnologica

[61] cienciaconciencia.org.ve/armas-autonomas-una-carta-abierta-de-investigado

[62] www.infodefensa.com/es/2011/11/25/noticia-la-fuerza-aerea-venezolana-exhibe-sus-vehiculos-aereos-no-tripulados-ant-1x-2.html

[63] www.ex-cle.com/venezuela-implementa-el-primer-sistema-de-pago-biometrico-del-mundo; www.presidencia.gob.ve/Site/Web/Principal/paginas/classMostrarEvento3.php?id_evento=10438

[64] www4.cne.gob.ve/web/sistema_electoral/tecnologia_electoral_descripcion.php

authentication system set up by the National Electoral Council (CNE). It uses e-election technology developed by the multinational Smartmatic.[65]

The above shows that the private sector in Venezuela has used chatbot technology for banking, insurance, medical care, and digital payment systems. No widespread use of this technology in public institutions was observed, suggesting an opportunity to explore the potential of chatbots to optimise government processes, and simplify administrative procedures.

As can be seen, the use of robotics in Venezuela is oriented towards improving medical procedures, performing military operations such as patrolling territorial borders, in monitoring the environment, and for transporting waste. Actors from the academic and business sectors[66] have stated that there is little government financial support to promote robotics, which does not allow them to collaborate as part of a national technological development strategy. However, companies such as Vehiculum[67] are exploring the feasibility of using small robots to harvest vegetables in Venezuela's Andean region. Robotics training programmes[68] have also been created at different educational levels, and entrepreneurship is promoted through the digital business acceleration programme run by the transnational company Telefónica.[69]

When it comes to blockchain and biometric technology, the government has used these technologies to promote a digital economy using cryptoassets, establishing national control mechanisms for the distribution and sale of basic necessities, and to manage electoral processes. The aim has been to encourage and develop an ecosystem of supply and consumption of goods and services based on the Petro, and to establish control mechanisms in critical sectors. Likewise, the government recognises the political, economic and social value of blockchain technology in managing big data, which is why the National Centre for Development and Research of Free Technologies (Cenditel) has developed the Blockchain Project[70] and the Automated Open Consultation System.[71] These projects allow for the automated processing of data for modelling, based on latent Dirichlet allocation (LDA),[72] and facilitating decision making in electoral processes.

The research for this report has shown that in Venezuela there are efforts to integrate AI technologies into different processes of innovation, development and management; however, there are no Venezuelan state strategies that integrate initiatives in the public and private sectors that make it possible to take advantage of the benefits of AI to guarantee sustained social development.

Conclusion

The serious frictions and differences that exist in different sectors of the country with respect to the development model and policies that must be implemented to overcome the political, economic and social crisis are significant barriers to integrating AI technologies in the country's development efforts. This amounts to a lost opportunity to promote the social development of the country, based on the use and management of technologies. This report shows that AI is being used in disconnected environments with little cooperation between the project proponents. The lack of national consensus means that the national productive sector[73] is largely unaware of the capabilities of emerging technologies already in use in the country. Those with the skills to implement the technology, are, in turn, unaware of the needs of the sector. In this environment, while the government must correct the economic distortions that are impoverishing citizens, it must also evaluate new ways to implement public policies and seek a national consensus that unites people for the good of the nation. Part of this involves allowing the strategic integration of AI in critical sectors in the country. Likewise, the private sector must contribute to developing democratic ways that promote change, that are necessary to guarantee social development in the country and the optimal use of AI.

The use of AI in the socioeconomic development of Venezuela, in some cases, has only just started and has responded to specific needs that are the result of the humanitarian social crisis in the country. This is particularly visible in the diaspora of Venezuelans, including the migration of medical specialists, which has resulted in the use of chatbots to attend to patients and manage virtual clinics, as well as using cryptocurrencies for remittances. The latter is a way to support Venezuelans who stay in the country and who face one of the worst economic

[65] https://en.wikipedia.org/wiki/Smartmatic

[66] www.ing.ula.ve/averod/?page_id=14

[67] fedecamarasradio.com/inteligencia-artificial-en-venezuela-podria-aumentar-en-2019

[68] www.conatel.gob.ve/ninos-desarrollan-robotica-creativa-con-vision-social

[69] www.ciberespacio.com.ve/2013/01/industria/la-robotica-se-abre-espacio-en-venezuela

[70] blockchain.cenditel.gob.ve

[71] analisisdatos.cenditel.gob.ve/2018/11/12/big-data-y-tecnologias-libres-fortalecen-la-democracia-participativa-desde-cenditel

[72] https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation

[73] Collectively referring to sectors where economic activity occurs.

crises in their history, where there is a galloping hyperinflation[74] impacting on the cost of goods and services. A blockchain-based cryptoasset economy can introduce opportunities for economic innovation, such as using the Petro as an everyday medium of exchange, a unit for setting prices, and a form of savings for citizens, despite the unconstitutionality of the Petro, as argued by the AN, and the distrust in its use by many citizens. Also in the financial sector, the shortage of hard currency in the economy has prompted the use of digital payments supported by chatbots, and the management of banking services with virtual assistants.

The need to control the country's borders, both for the country's sovereignty and because of illicit activities, has led the government to deploy UAVs. Their use is also seen in environmental management, and in the monitoring of disaster zones. Robotics have also been used to manage waste in high-risk areas.

Finally, while AI has helped with the analysis of big data for socio-political purposes, biometric systems have been a way to control critical processes in the economy, such as managing the scarcity of goods. Biometrics have also been useful in electoral processes, helping to control electoral irregularities.

Action steps

The government has been taking measures to boost the digital economy in Venezuela as a way to establish alternative mechanisms for the development of the country. It has also used AI technologies to analyse big data, control processes in the distribution of goods and services, for electoral processes, and, although not discussed in this report, in the management of social networks.[75] The strategic management of data allows the government to react in a timely and effective way to changes in the environment. However, this should be handled with caution, as political factors can influence decision making and subject citizens to unfair measures and restrictions, which can generate social chaos and discrimination. On the other hand, it is imperative that the use of the Petro reaches national consensus so that trust in the financial environment can be created. The government should build bridges to alleviate an atmosphere of distrust with regards to cryptocurrencies. It should promote technological innovation, not by imposition, but through debate and consensus. For their part, the private sector and civil society must consider how to optimise the use of AI, and in this way contribute to the creation of solutions that help overcome the crisis impacting so negatively on the development of Venezuela.

[74] www.finanzasdigital.com/2019/05/an-inflacion-de-abril-2019-fue-447-anual-1-304-494

[75] There is AI behind what we see on social networks. AI decides what is shown on the page every time we log into a social network. AI allows for dissemination strategies that can influence social processes through, for example, fake news or using algorithms to influence people's online preferences. Social media is also being used for espionage, and the unauthorised use of personal data for political and economic purposes by large technological monopolies, governments and computer giants, among other things. www.eluniversal.com/tecnologia/38629/las-redes-sociales-y-la-inteligencia-artificial, www.aporrea.org/tecno/n344005.html

GLOBAL INFORMATION SOCIETY WATCH 2019

Artificial intelligence:

Human rights, social justice and development

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

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

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

GLOBAL INFORMATION SOCIETY WATCH
2019 Report
www.GISWatch.org

APC ARTICLE 19 Sida