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
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
Monocróm

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)

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

Creative Commons Attribution 4.0 International (CC BY 4.0)
https://creativecommons.org/licenses/by/4.0/
Some rights reserved.

Global Information Society Watch 2019 web and e-book
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.
Decolonising AI: A transfeminist approach to data and social justice

Paz Peña and Joana Varon
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,” 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. The province’s Ministry of Early Childhood worked for years with the anti-abortion NGO Fundación CONIN to prepare this system. 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. The idea that algorithms can predict teenage pregnancy before it happens is the perfect excuse for anti-women 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.

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.
was published by the Laboratorio de Inteligencia Artificial Aplicada (LIAA) at the University of Buenos Aires. According to LIAA, which analysed the methodology posted on GitHub by a Microsoft engineer, 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, and has been exported to Colombia and implemented in the municipality of La Guajira.

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 humanity. Instead, we propose to build a transfeminist 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. 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.”

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.

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:

[Models, 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.]

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
risk constructing machine intelligence that mirrors a narrow and privileged vision of society, with its old, familiar biases and stereotypes.\textsuperscript{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,”\textsuperscript{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.\textsuperscript{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 data by private companies will only benefit the population.\textsuperscript{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.”\textsuperscript{23} Along the same lines, Coudry and Mejias\textsuperscript{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”.\textsuperscript{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, shows 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.\textsuperscript{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.\textsuperscript{27}

---


\textsuperscript{27} Van’t Wout, E., et al. (2018). Capítulo II. Big data para la identificación de comportamiento criminal. In L. Irarrázaval et al. (Eds.), Propuestas para Chile. Pontificia Universidad Católica de Chile.
Many critics point to the negative impacts of predictive policing on poorer neighbourhoods and other affected communities, including police abuse, stigmatisation, racism and discrimination. 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.

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, 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:

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.

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.

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”.

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.”

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.

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

---

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

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

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

As Silvia Rivera Cusicanqui says:

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

Beyond even a human rights framework, decolonial and transfeminist approaches to technologies are great tools to envision alternative futures and overturn the prevailing logic in which AI systems are being deployed. Transfeminist values need to be embedded in these systems, so advances in the development of technology help us understand and break what black feminist scholar Patricia Hill Collins calls the “matrix of domination”38 (recognising different layers of oppression caused by race, class, gender, religion and other aspects of intersectionality). This will lead us towards a future that promotes and protects not only human rights, but also social and environmental justice, because both are at the core of decolonial feminist theories.

Re-imagining the future

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

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

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

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


39 https://codesign.mit.edu

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


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.