CARR CENTER FOR HUMAN RIGHTS POLICY
HARVARD KENNEDY SCHOOL

Not My A.I.
Towards Critical Feminist Frameworks to Resist Oppressive A.I. Systems

Joana Varon
and Paz Peña

Carr Center
Discussion Paper

Not My A.I.

Towards Critical Feminist Frameworks To Resist Oppressive A.I. Systems[1]

Carr Center for Human Rights Policy
Harvard Kennedy School, Harvard University
October 12, 2022

Joana Varon
Technology and Human Rights Fellow
Carr Center for Human Rights Policy

Paz Peña
Independent specialist on technology,
feminism and social justice

[1] This article is a summary of the main findings presented at notmy.ai platform, co-developed by the authors.

ABSTRACT

In the hype of A.I., we are observing a world where States are increasingly adopting algorithmic decision-making systems altogether with narratives that portray them as a magic wand to "solve" social, economic, environmental, and political problems. But in practice, instead of addressing such promise, the so-called Digital Welfare States are likely to be deploying oppressive algorithms that expand practices of surveillance of the poor and vulnerable; automate inequalities; are racist and patriarchal by design; further practices of digital colonialism, where data and mineral extractivism feed Big Tech businesses from the Global North; and reinforce neoliberal practices to progressively drain out social security perspectives. While much has been discussed about "ethical", "fair," or "human-Centered" A.I., particularly focused on transparency, accountability, and data protection, these approaches fail to address the overall picture. To deepen critical thinking and question such trends, led by case-based analysis focused on A.I. projects from Latin America that are likely to pose harm to gender equality and its intersectionalities of race, class, sexuality, territoriality, etc, this article summarizes some findings of the notmy.ai project, seeking to contribute to the development of feminist frameworks to question algorithmic decision-making systems that are being deployed by the public sector. The universalistic approach of human rights frameworks provide important goals for humanity to seek, but when we look into the present, we cannot ignore existing power relations that maintain historical relations of oppression and domination. Rights are not universally accessed. Feminist theories and practices are important tools to acknowledge the existence of the political structures behind the deployment of technologies and, therefore, are an important framework to question them. For this reason, they can serve as a powerful instrument to imagine other tech and other worlds based on collective and more democratic responses to core societal challenges, focused on equity and social-environmental justice.

Introduction

Progressively, 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. Some have named this trend "Digital Welfare States," where, to quote the U.N. Special Rapporteur for Extreme Poverty, "systems of social protection and assistance are increasingly driven by digital data and technologies that are used to automate, predict, identify, surveil, detect, target and punish."
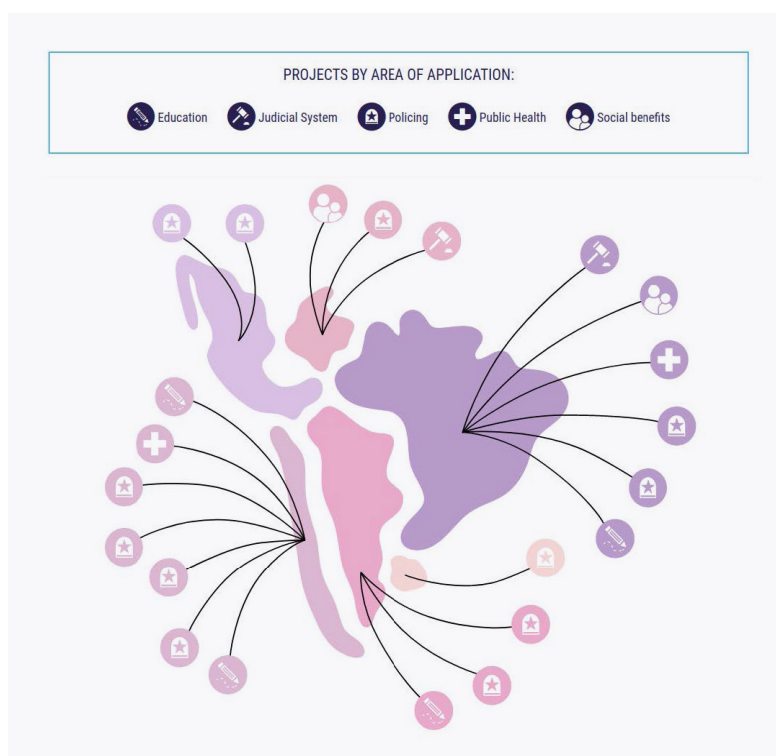
As one should expect, governments in Latin America are also following the hype to deploy A.I. systems in public services. In an initial non-conclusive mapping exercise, through desk research and a questionnaire distributed across digital rights networks in the region, we have mapped projects where algorithmic decision-making systems are being deployed by governments with likely harmful implications on gender equality and all its intersectionalities. As of April 2021, we have mapped 24 cases in Chile, Brazil, Argentina, Colombia, and Uruguay, which we were able to classify into 5 categories: judicial system, education, policing, social benefits, and public health. Several of them are in an early stage of deployment or developed as pilots. We could also see that, while the first projects were related to the area of policing, using A.I. for social protection is the most recent trend. Most of them are in the pilot stage and U.S. companies, particularly Microsoft and I.B.M. are involved in several of these initiatives. Some companies are even exporting projects from one country to another, using the region as a laboratory of ideas, normally bad ideas, which, not surprisingly, are not even tested first in the U.S., their home country.

Unlike many A.I. projects and policies that tend to depart from the start-up motto "move fast and break things," our collection of cases depart from the assumption that, unless you prove you are not causing harm, you are very likely to be, if your system targets marginalized communities. As Tendayi Achiume, the U.N. Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia, and Related Intolerance, poses in the report "Racial Discrimination and Emerging Digital Technologies," databases used in these systems are "the product of human design and can be biased in various ways, potentially leading to—intentional or unintentional—discrimination or exclusion of certain populations, in particular, minorities as based on racial, ethnic, religious and gender identity." (Tendayi, 2020).

As Cathy O'Neil observes in her book *Weapons of Math Destruction*, A.I. systems are based on models that are abstract representations, universalizations, and simplifications of complex realities where much information is being left out according to the judgment of their creators. "[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." (O'Neil, 2016). Have these opinions been challenged and tested before the deployment of these systems?

Given these problems, it should be recognized that part of the technical community has made various attempts to mathematically define "fairness," and thus provide a demonstrable standard on the matter. Likewise, several organizations, both private and public, have undertaken efforts to define ethical standards for A.I. The data visualization "Principled Artificial Intelligence" (Berkman Klein, 2020) shows the diversity of ethical and human rights-based frameworks that emerged from different sectors from 2016 onwards with the goal of guiding the development and use of A.I. systems. The study shows "a growing consensus around eight key thematic trends: privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility and promotion of human values." Nevertheless, as we can see from that list, none of this consensus is driven by social justice principles. Instead of asking how to de-

PROJECTS BY AREA OF APPLICATION:

Education   Judicial System   Policing   Public Health   Social benefits
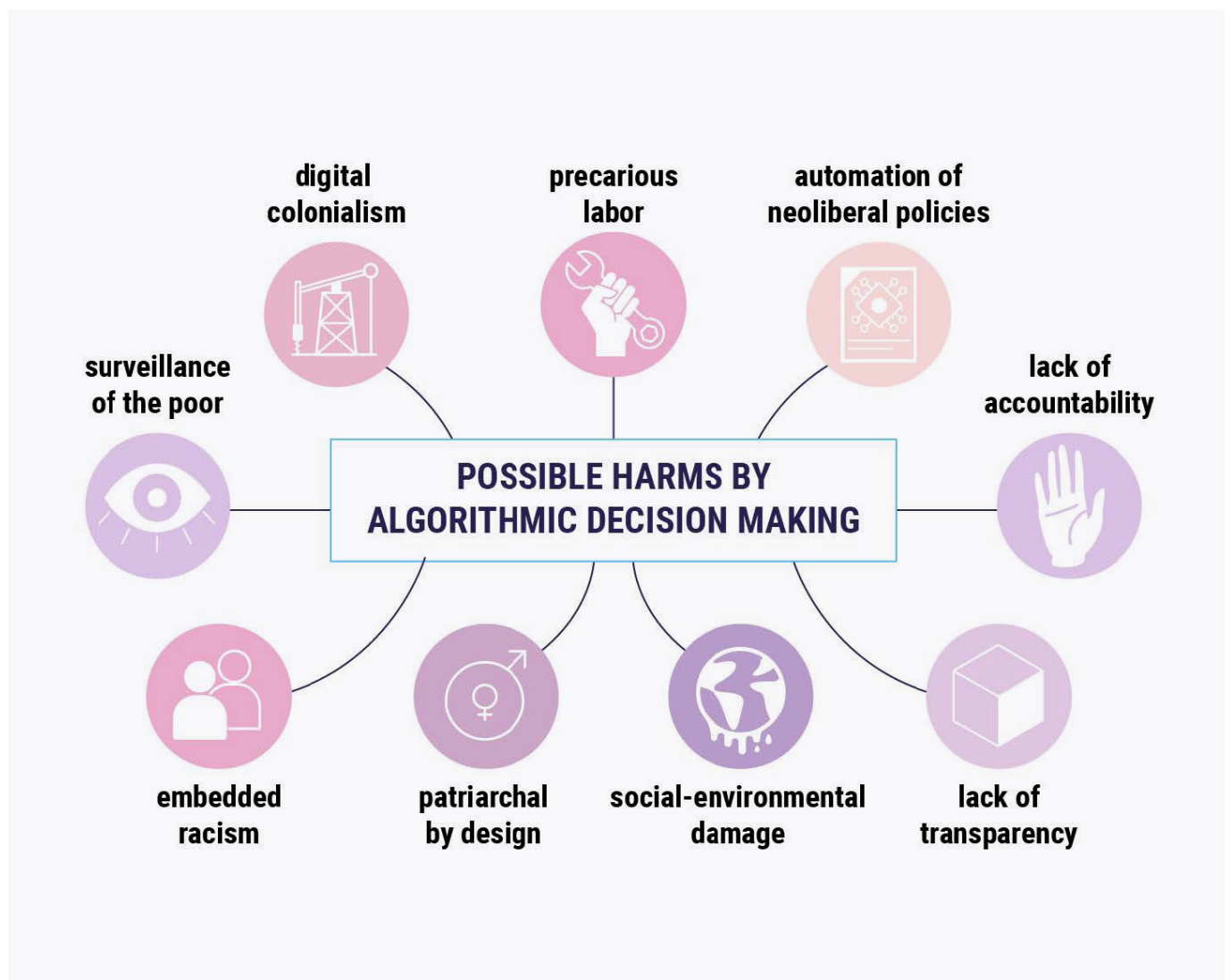
Mapping available at www.notmy.ai

velop and deploy an A.I. system, shouldn't we be asking first: "Why to build it?"; "Is it really needed?"; "On whose request?"; "Who profits?"; and "Who loses?" from the deployment of a particular A.I. system? Should it even be developed and deployed?

Based on both our bibliographic review and our case-based analysis of possible harms by A.I. programs deployed in the areas of education and social benefits, in Chile, Argentina, and Brazil, at notmy.ai, we are gradually expanding an empirically tested case-based anti-colonial feminist framework to question these systems from perspectives that go beyond criticism from the Global North. Hopefully, it will be a scheme that can help us to pose structural questions about whether a given governmental A.I. system may incur possible harm to several feminist agendas.

II. Feminist Categories to Question A.I. Systems

Artificial Intelligence programs have faced criticism on several fronts. Based on an overall bibliographical review as well as findings from the case-based analysis, we have created the following framework of analysis which intends to go beyond the discourses of fairness, ethical, or human-centric A.I. and seeks a holistic structure that considers power relations to question the idea of deploying A.I. systems in several realms of the public sector.

Oppressive A.I. Framework
by Joana Varon and Paz Peña

digital colonialism

precarious labor

automation of neoliberal policies

surveillance of the poor

POSSIBLE HARMS BY ALGORITHMIC DECISION MAKING

lack of accountability

embedded racism

patriarchal by design

social-environmental damage

lack of transparency

Design by Clarote for notmy.ai

All these categories will be briefly explained based on a bibliographical review, in what follows.

A. SURVEILLANCE OF THE POOR: *Turning Poverty and Vulnerability into Machine-Readable Data*

The former U.N. Special Rapporteur on Extreme Poverty and Human Rights, Philip Alston, has criticized the phenomenon in which "systems of social protection and assistance are increasingly driven by digital data and technologies that are used to automate, predict, identify, surveil, detect, target and punish." (A/74/48037 2019). These granular data sources enable authorities to infer people's movements, activities, and behavior, not without having ethical, political, and practical implications for how the public and private sector view and treat people. According to Linnet Tylor, in her article "What Is Data Justice?" (TYLOR, 2017), this is even more challenging in cases of low-income portions of the population, since the ability of authorities to collect accurate statistical data about them has been previously limited, but now is targeted by regressive classifications systems that profile, judge, punish, and surveil.

"[Facial recognition] systems exhibit higher error rates for darker-skinned women than for any other group, with the lowest error rates for light-skinned men."

Most of these programs take advantage of the tradition of State surveillance on vulnerable populations (Eubanks, 2018), turn their existence into data, and now use algorithms to determine the provision of social benefits by the States. Analyzing the case of the U.S., Eubanks shows how the usage of A.I. systems is subjected to a long tradition of institutions that manage poverty and that seek, through these innovations, to adapt and continue their urge to contain, monitor, and punish the poor. In doing so, these institutions turn poverty and vulnerability into machine-readable data, with real consequences on the lives and livelihoods of the citizens involved. (Masiero & Das 2019). Likewise, O'Neil (2016), analyzing the usages of A.I. in the U.S., asserts that many A.I. systems "tend to punish the poor," meaning it is increasingly common for wealthy people to benefit from personal interactions, while data from the poor are processed by machines making decisions about their rights.

70%

of poor people in the world in 1995 were women.

This becomes even more relevant when we consider that social class has a powerful gender component. It is common for public policies to speak of the "feminization of poverty." In fact, the United Nations Conference on Women, held in Beijing in 1995, concluded that 70% of poor people in the world were women. The reasons why poverty affects women have to do, not with biological reasons, but with structures of social inequality that make it more difficult for women to overcome poverty, such as access to education and employment. (Aguilar, 2011).

B. EMBEDDED RACISM

For U.N. Special Rapporteur E. Tendayi (2020), emerging digital technologies should also be understood as capable of creating and maintaining racial and ethnic exclusion in systemic or structural terms. This is also what tech researchers on race and A.I. in the U.S., such as Ruha Benjamin, Joy Buolamwini, Timnit Gebru, and Safiya Noble highlight in their case studies, which vary from facial recognition technologies to search engine algorithms. Ruha Benjamin (2019) particularly discusses how the use of new technologies reflects and reproduces the existing racial injustices in U.S. society, even though they are promoted and perceived as more objective or progressive than the discriminatory systems of an earlier era. In this sense, for this author, when A.I. seeks to determine how much people of all classes deserve opportunities, the designers of these technologies build a digital caste system structured on existing racial discrimination.

From technology development itself, Noble's research (2018) demonstrates how commercial search engines such as Google not only mediate but are mediated by a series of commercial imperatives. These imperatives, in turn, are supported by both economic and information policies that end up endorsing the commodification of women's identities. In this case, she exposes this by analyzing a series of Google searches where Black women end up being sexualized by the contextual information the search engine throws up (e.g., linking them to wild and sexual women).

Another notable study is by Buolamwini & Gebru (2018), who analyzed three commercial facial recognition systems that include the ability to classify faces by gender. They found that the systems exhibit higher error rates for darker-skinned women than for any other group, with the lowest error rates for light-skinned men. In Latin America, Tarcisio Silva, has coined the term algorithmic racism for this kind of tech deployment, while Pablo Nunes and Nina da Hora, all three from Brazil, have pointed out the dangers of automating police brutality against the Black population through facial recognition technologies, particularly in Brazilian favelas.

C. PATRIARCHAL BY DESIGN: *Sexism, Compulsory Heteronormativity, and Gender Binarism*

Many A.I. systems work by sorting people into a binary view of gender, as well as by reinforcing outdated stereotypes of gender and sexual orientation. A study co-authored by DeepMind senior staff scientist Shakir Mohamed exposes how the discussion about algorithmic fairness has omitted sexual orientation and gender identity, with concrete impacts on "censorship, language, online safety, health, and employment" leading to discrimination and exclusion of L.G.B.T.+ people.

Gender has been analyzed in a variety of ways in A.I. West, Whittaker, & Crawford (2019) argue that the diversity crisis in industry and bias issues in A.I. systems (particularly race and gender) are interrelated aspects of the same problem. Researchers commonly examined these issues in isolation in the past, but mounting evidence shows that they are closely intertwined. However, they caution, that despite all the evidence on the need for diversity in technology fields, both in academia and industry, these indicators have stagnated.

Inspired by the work of Buolamwini & Gebru (2018), Silva & Varon (2021) researched how facial recognition technologies affect transgender people and concluded that, although the main public agencies in Brazil already use these types of technology to verify identities for accessing public services, there is little transparency on their accuracy (tracking false positives or false negatives), as well as on privacy and data protection in the face of data sharing practices between public administration agencies and even between private entities.

In the case of Venezuela, amid a sustained humanitarian crisis, the State has implemented biometric systems to control the acquisition of basic necessities, resulting in several complaints of discrimination against foreigners and transgender people. According to Díaz Hernández (2021), legislation to protect transgender people is practically nonexistent. They are not allowed recognition of their identity, which makes this technology re-signify the value of their bodies "and turns them into invalid bodies, which therefore remain on the margins of the system and the margins of society."

In the case of poverty management programs through big data and A.I. systems, it is crucial to look at how poor women are particularly subject to surveillance by States and how this leads to the reproduction of economic and gender inequalities. (Castro & López, 2021).

D. DIGITAL COLONIALISM

Authors like Couldry and Mejias (2018) and Shoshana Zuboff (2019) review this current state of capitalism where the production and extraction of personal data naturalize the colonial appropriation of life in general. To achieve this, a series of ideological processes operate where, on the one hand, personal data is treated as raw material, naturally disposable for the expropriation of capital, and, on the other, where corporations are considered the only ones capable of processing and, therefore, appropriating the data.

Regarding colonialism and A.I., Mohamed et al. (2020) examine how coloniality presents itself in algorithmic systems through **institutionalized algorithmic oppression** (the unjust subordination of one social group in order to privilege another), **algorithmic exploitation** (the ways in which institutional actors and corporations take advantage of often already marginalized people for the asymmetric benefit of these industries), and **algorithmic dispossession** (the centralization of power in the few and the dispossession of many), in an analysis that seeks to highlight the historical continuities of power relations.

E. SOCIO-ENVIRONMENTAL DAMAGE

Crawford (2021) calls for a more comprehensive view of A.I. as a critical way to understand that these systems depend on the exploitation of energy and mineral resources, and cheap labor, on the one hand, and, in addition, our data at scale. In other words, A.I. is an extractive industry. All these systems are energy-intensive and heavily dependent on minerals, sometimes, extracted from areas where there are. In Latin America alone, we have the lithium triangle within Argentina, Bolivia, and Chile, as well as several deposits of 3TG minerals (tin, tungsten, tantalum, and gold) in the Amazon region, all minerals used in cutting-edge electronic devices. As Danae Tapia and Paz Pena pose, digital communications are built upon exploitation, even though "sociotechnical analyses of the ecological impact of digital technologies are almost non-existent in the hegemonic human rights community working in the digital context." (Tapia & Pena, 2021). And, even beyond ecological impact, Camila Nobrega and Joana Varon also expose that green economy narratives together with technosolutionisms are "threatening multiple forms of existence, of historical uses and collective management of territories." Not by chance the authors found out that Alphabet Inc., Google's parent company is exploiting 3TG minerals in regions of the Amazon where there is a land conflict with indigenous people. (Nobrega & Varon, 2021).

"Labeling images, and cleaning databases are all manual work very often performed in unsavory working conditions "to make the internet seem smart." Communalities of these jobs are very precarious working conditions, normally marked by overwork, low-pay, no social benefits or stability.

This is very different from the work conditions of the creators of such systems. (Crawford, 2021). Who takes care of your database? As always, care work is not recognized as valuable work."

F. AUTOMATION OF NEOLIBERAL POLICIES

As Payal Arora (2016) frames it, discourses around big data have an overwhelmingly positive connotation thanks to the neoliberal idea that private companies' exploitation of poor communities' data for profit will only benefit the population. From an economic point of view, Digital Welfare States are deeply intertwined with capitalist market logic and, particularly, with neoliberal doctrines that seek deep reductions in the general welfare budget, including the number of beneficiaries, the elimination of some services, the introduction of demanding and intrusive forms of conditionality of benefits, to the point that—as Alston has stated (2019)—individuals do not see themselves as subjects of rights but as service applicants. (Alston, 2019, Masiero and Das, 2019). In this sense, it is interesting to see that A.I. systems, in their neoliberal efforts to target public resources, also classify who the poor subject is through automated mechanisms of exclusion and inclusion. (López, 2020).

G. PRECARIOUS LABOR

Particularly focused on artificial intelligence and the algorithms of Big Tech companies, anthropologist Mary Gray and computer scientist Siddharth Suri point out the "ghost work" or invisible labor that powers digital technologies. Labeling images, and cleaning databases are all manual work very often performed in unsavory working conditions "to make the internet seem smart." Communalities of these jobs are very precarious working conditions, normally marked by overwork, low-pay, no social benefits or stability. This is very different from the work conditions of the creators of such systems. (Crawford, 2021). Who takes care of your database? As always, care work is not recognized as valuable work.

H. LACK OF TRANSPARENCY

According to A.I.N.O.W. (2018), when government agencies adopt algorithmic tools without adequate transparency, accountability, and external oversight, their use can threaten civil liberties and exacerbate existing problems within government agencies. Along the same lines, O.E.C.D. (Berryhill et al., 2019) postulates that transparency [on the part of] is strategic to foster public trust in the tool. More critical views note the neoliberal approach when transparency depends on the responsibility of individuals, as they do not have the time or the desire to commit to more significant forms of transparency and consent online. (Annany & Crawford, 2018). Thus, government intermediaries with special understanding and independence should play a role here. (Brevini & Pasquale, 2020). Furthermore, Annany & Crawford (2018) suggest that what the current vision of transparency in A.I. does is fetishize the object of technology, without understanding that technology is an assembly of human and non-human actors. Therefore, to understand the operation of A.I. it is necessary to go beyond looking at the mere object.

"Bad ideas dressed as innovation spread fast."

III. Case-Based Analysis to Deploy the Critical Feminist Framework

A. CAN AN A.I. SYSTEM ACTUALLY PREDICT TEENAGE PREGNANCY?

The Ministry of Early Childhood from the Province of Salta, Argentina is using a machine learning system as part of its program called "*Plataforma Tecnológica de Intervención Social*" ("Technological Platform of Social Intervention") with the goal of preventing school dropouts and teenage pregnancy. "Intelligent algorithms allow identifying characteristics in people that could end up in any of these problems and warn the government to work in their prevention," said Microsoft Azure's representative, which is responsible for the deployment of the system. "With technology, based on name, surname and address, you can predict five or six years ahead which girl, future teenager, is 86% predestined to have a teenage pregnancy," declared Juan Manuel Urtubey, a conservative politician and governor of Salta.

Predestined, he said. But, to predict and predestine someone for unwanted pregnancy is not an exact task, neither for fortune-tellers nor for mathematicians, fortune-tellers would be more careful perhaps. A study by *Laboratorio de Inteligencia Artificial Aplicada* (L.I.A.A.) from the University of Buenos Aires, which analyzed the methodology posted on GitHub by Microsoft engineers, heavily criticized the attempt. The study stated that the results were falsely oversized because there were statistical errors in the methodology, the database is biased due to the sensitivities of reporting unwanted pregnancy, and the data collected is inadequate to make any future prediction.

And criticisms can go beyond statistical analysis. Several feminists pointed out that the argument 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 safe abortion laws as 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. Fur-

thermore, it is also notable to point out that the system has chosen to work on a database composed only of female data. This specific focus on particular sex also reinforces patriarchal gender roles and blames female teenagers for unwanted pregnancy, as if a child could be conceived without semen.

Even under several criticisms, the initiative continued to be deployed. And worse, bad ideas dressed as innovation spread fast: the system is now being deployed in other Argentinian provinces, such as *La Rioja*, *Tierra del Fuego*, and *Chaco*. It has also been exported to Colombia, where it is being implemented in the municipality of *La Guajira*. Another iteration of that same project has also reached the Brazilian Federal Government, through a partnership with the Brazilian Ministry of Citizenship and Microsoft. Allegedly, by September 2019, Brazil was the 5th country in Latin America to Projeto Horus, which was presented in the media as a "tech solution to monitor social programs focused on child development. The first city to test the program was Campina Grande, from the State of Paraíba, in the northeast region of Brazil. Among the authorities and institutions in the kick-start meeting was a representative from Microsoft, the Ministry of Early Childhood from the municipality of Salta, and members from the Brazilian Ministry of Citizenship.

It would not be an exaggeration to say that the attempt to transpose the system to Brazil was another expression of colonial extractivism once, allegedly, not even the Brazilian government kept records of the results of the proof of concept, a closed box, even for their partners. Did Microsoft have access to the database of the Brazilians? We tried to schedule an interview with a representative from Microsoft in Brazil who was talking about the project in the media, but, after we sent the questions, the previously scheduled interview was canceled.

In summary, we can say that the "*Plataforma Tecnológica de Intervención Social*" and Projeto Horus are just very eloquent examples of how the misconception that A.I. is neutral has been increasingly deployed in some countries in Latin America to assist potentially discriminatory public policies that could undermine the human rights of unprivileged people, as well as monitor and censor women and their sexual and reproductive rights. Analyzing our framework from Oppressive A.I. we could say it ticks all the boxes:

2. SISTEMA ALERTA NIÑEZ – CHILE

The Childhood Alert System (S.A.N.) is a software-based on the use of Predictive Risk Modeling (P.R.M.) to "identify children and adolescents (N.N.A.) who are at risk of violation of rights and the

families that require support to enhance their role of providing child protection." The system provides complimentary input to data collected in Chile by the so-called Local Childhood Offices to support decision-making by managers of these offices when offering support to children and their families. It is based on a set of individual, family, environmental, and peer conditions that tend to occur when there is a high risk of violation of rights.

To identify children and adolescents in these circumstances, the model is trained through the analysis of children and adolescents who had already been subject to violation of their rights, to then study their life course and identify the family conditions experienced by them before the violations occurred. Children and adolescents facing these conditions at present would be those who may—potentially—need support to mitigate the risk factors surrounding them.

In New Zealand, a similar system was tested. Anne Tolley, then the Social Development Minister, said: "Where it goes from there is another big ethical question. Because God knows, do we really want people with clipboards knocking on people's doors and saying: 'Hello, I'm from the Government, I'm here to help because your children are going to end up in prison?' I just can't see that happening." Tolley's position was made clear by her note on the briefing paper: "Not on my watch. Children are not lab rats."

> ## "...do we really want people with clipboards knocking on people's doors and saying: 'Hello, I'm from the Government, I'm here to help because your children are going to end up in prison?' I just can't see that happening."
> —*Anne Tolley, Social Development Minister*

As soon as the creation of this system was publicly announced in Chile, civil society groups working for children's rights stated that, in addition to surveillance, the system "implied the imposition of a certain form of sociocultural normativity," as well as "socially validating forms of stigmatization, discrimination and even criminalization of cultural diversity existing in Chile." This particularly affected indigenous peoples, migrants, and low-income people, and ignored that growing

cultural diversity "demanded greater sensitivity, visibility, and respect, as well as the inclusion of culturally relevant approaches to public policies." (*Sociedad Civil de Chile Defensora de los Derechos Humanos del Niño*, et al., 2019). In this sense, Francisca Valverde, from the group of organizations *Bloque por la Infancia* and executive director of *Asociación Chilena Pro Naciones Unidas* (A.C.H.N.U.), argues that this type of system stigmatizes poor children and does not emphasize the protection of children and adolescents in different territories, including those of higher social classes.

As it is presented as a complementary system to the Local Childhood offices, technology's errors in the risk prediction amongst children and adolescents are expected to be shielded by human intervention, giving the machine some impunity to continue working. However, there are no field studies in the case we have examined that consider how caseworkers who interact with the machine deal with "automation bias" referred to the higher valorization of automated information than our own experiences. (Bridle, 2018). Furthermore, people affected by these systems, that is, poor children, adolescents, and their families, are not even the subject of consultation since they are not recognized as interested parties. Likewise, the consent to use your data submitted to get or not social benefits opens a whole discussion about the ethics of these systems that has not been resolved.

The A.I. systems examined in the case studies show that their design and use by States respond to a continuum of neoliberal policies that have abounded in Latin America, to varying degrees, during the last 40 years. (López, 2020). On the one hand, they are instruments that automate and grant a degree of technological resolution to ideological decisions: resources' focalization. (Alston, 2019). In this case, it is mainly about using big data to produce a more detailed category of poor children and adolescents (López, 2020) and, a step further, automate their social-risk determination. The vulnerable childhood approach is a classic neoliberal take and it comes from the idea of poverty as an individual problem (not a systemic one) and caseworkers as protectors of people "at-risk." (Muñoz Arce, 2018). The hasty adoption of these neoliberal instruments is also worrisome because it is hard to dissolve them once adopted. (Eubanks, 2018).

S.A.N. and these social-risk models can be analyzed at least in two aspects: first, to question if big data and A.I. could ever reflect structural elements that influence the risk of vulnerability and social inequalities of our societies, or is are they just a way to objectivize the responsibility of individuals through their data trajectory in the State. And second, how S.A.N. and these social-risk models are a continuation of the idea of data disembodiment, where technologies artificially abstract bodies, identities, and interactions from social contexts to obscure its operation as a tool for social control, aggravating its consequences on social inequalities. (Monahan, 2019).

We could say that, so far, both S.A.N. and the systems analyzed from Argentina and Brazil check the boxes of harmful automatic decision-making proposed by our framework:
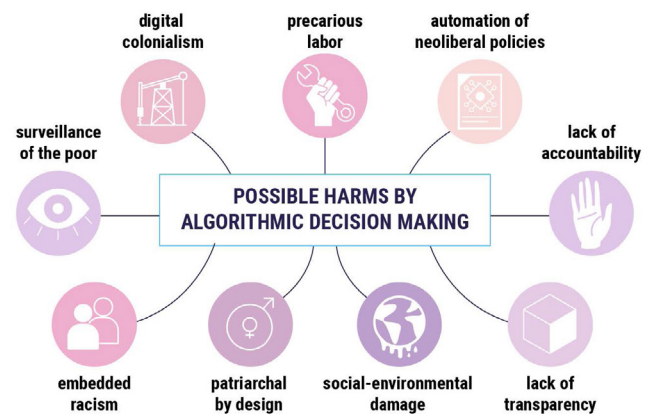
Oppressive A.I. Framework by Joana Varon and Paz Peña. Design by Clarote for notmy.ai.

The proposed Oppressive A.I. framework is not written in stone. These categories are not fixed, they can expand according to a particular context. For instance: if an A.I. system is developed to help people to access public services but solely favors ableism, it can also be considered an Oppressive A.I. as it is excluding people with disabilities from accessing public services. Therefore, this framework is just a general guide for questions, a work-in-process, and shall be re-shaped according to the particular context and its oppressions. Indeed, it has been gradually becoming more complex over a series of workshops with feminists from Brazil and other countries in Latin America.

IV. Conclusions: The Need to Decolonize Our Imaginaries Around Technologies

Many people look at our framework, which was envisioned to promote critical approaches to A.I. systems, and ask for a positive agenda, perhaps a recipe to create a "good A.I." But can an A.I. be good? Is it even possible to create a feminist A.I.? Isn't the term "artificial intelligence" just as loaded with marketing and sales purposes as "smart cities," "internet of things," or "the cloud." To what kind of intelligence are we referring to being possible to automate? Several national policies for A.I. and most start-ups and big tech corporations operate under the Silicon Valley motto of "move fast and break things," meaning, innovate first and check possible harms later. That is just a new expression of old colonial practices that have led us to a world on the verge of environmental catastrophes, far away from social justice, and perpetuating violence against whoever is different from the White male cis-hetero capitalist. Therefore, before attempting to go fur-
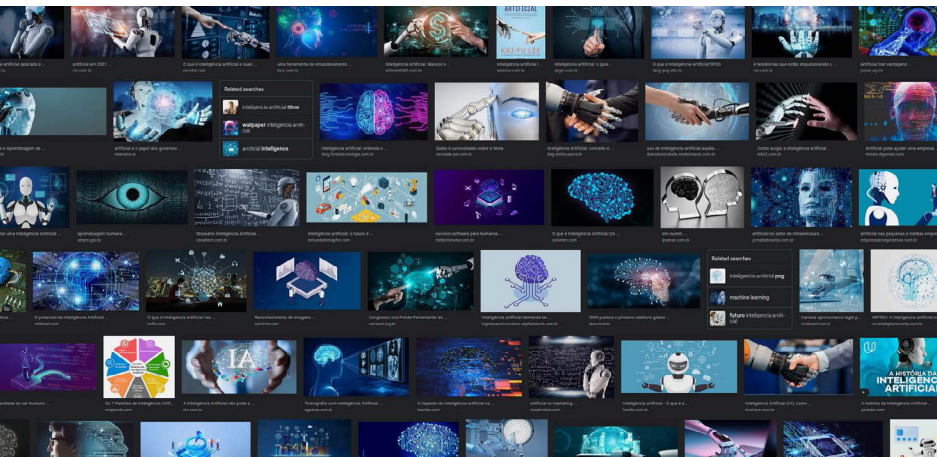
"Several national policies for A.I. and most start-ups and big tech corporations operate under the Silicon Valley motto of "move fast and break things," meaning, innovate first and check possible harms later.

That is just a new expression of old colonial practices that have led us to a world on the verge of environmental catastrophes, far away from social justice, and perpetuating violence against whoever is different from the White male cis-hetero capitalist."

ther with any propositive framework for positive A.I. systems (if they are even possible), we need to take a moment to re-set our tech imaginaries, away from the imaginary around tech coined by the so-called "tech bros" in Silicon Valley. For instance, today if we search for "Artificial Intelligence" in a search engine, this is the picture of imaginaries that we get: Robots, robots, robots, brains, bites, circuits, metallic colors, and a lot of blue. Actually not very far away from imaginaries around technologies that another powerful industry world-wide has induced us: Hollywood.

"The future of tech shall have more colors, more regions, cultures, and dreams."

Feminist Theories to Enforce Data Protection" we have shed a light on how limited the in-dividualistic notion of consent proposed in data protection frameworks is. Departing from a universalized notion of the individu-als who are consenting, it does not take into account unequal power relations. But, if we do not have the ability to say no to big tech companies when we need to access a mo-nopolistic service, we clearly cannot freely consent. If consent is something that the "Me too" movement has shown that Hollywood is not very kin to, the Big Tech compa-nies from the Silicon Valley haven't shown to think very differently. The industries of tech imaginary operate un-der very similar values.

Robots, robots, bites, screens, tech for war, female robot ser-vants. Certainly, something that also doesn't go very far away from cameras with racial biases that reinforce police brutality towards particular communities, nor from systems that are con-ceived to predestine women to unwanted futures, all these oppressive technologies that governments are deploying today. Hollywood and Silicon Valley are industries of imagi-naries and they have conquered the world or, if we could say that without sounding conspiratory, they have conquered minds and are shaping dreams about what the future should look like. But these dreams certainly do not represent the di-versity of tastes, histories, cultures, and wishes from across the world. They are more likely to represent the future busi-ness plan of companies whose C.E.O.s are tech bros (and, how odd is that: a quick search of their names also shows that they are all likely to pose in front of blue backgrounds, just like the blue imaginaries of A.I.).

The future of tech shall have more colors, more regions, cul-tures, and dreams. And this is not reachable through diversity and inclusion programs inside these companies that will keep operating under the same "move fast and break things" motto. Companies whose business model is to mainstream the regime of surveillance capitalism, where we barely have the power to say no. In the article "Consent to Our Data Bodies: Lessons from

But can we envision tech-nologies that operate under a real feminist notion of consent? What would the future look like if other people could envision technologies that help us escape, instead of reproducing what Patricia Hill Collins, in her classic book Black Feminist Thought, calls the matrix of domination* (capitalism, heteropatriarchy, white supremacy, ableism, and colonialism)? That is the main question that the project Oracle for Transfeminist Technologies is posing. De-veloped by Coding Rights, in partnership with scholar and design activist Sasha Costanza-Chock, it is a card game col-laboratively developed to help us envision what transfeminist technologies would look like. Through a series of workshops, we have been collectively brainstorming on what are the transfeminist values that shall inspire us to speculate about alternative futures. Over time, values such as agency, account-ability, autonomy, social justice, non-binarism, cooperation, decentralization, consent, diversity, decoloniality, empathy, and security, among others, have emerged in the meetings. These are values that work in the complete opposite of cur-rent values such as profit, addiction, consumerism, competi-tion, racism, hate, egocentrism, and other values that are em-bedded in several algorithms that pertain to our lives today.

Every value, turned into a card, has been fostering important conversations about how they could operate embedded into a piece of technology. While Oracle has been envisioned as a game, we believe that it can be a tool to slowly decolonize our imaginaries around technologies. As the values are used to inspire tech from a speculative future, participants are dispatched from present limitations, so imagination can run free. Science Fiction writer Ursula Le Guin once said in an interview: "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." Nevertheless, to avoid technosolutionisms, after imagination to design a speculative tech breaks loose, the last card to be open in consultation with the Oracle is the joker, which states: *Not all tech shall exist. Not everything that is new is better. The future is ancestral, what are you erasing with this idea?*

Some cards from the Oracle for Transfeminist Technologies, available at transfeministech.org.

NOT ALL TECH SHALL EXIST. NOT EVERYTHING THAT IS NEW IS BETTER. THE FUTURE IS ANCESTRAL, WHAT ARE YOU ERASING WITH THIS IDEA?
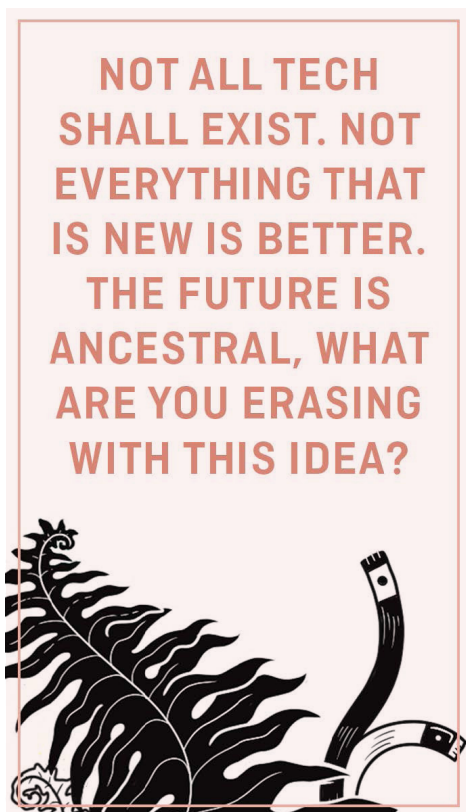
Bibliography

A.I.N.O.W. 2018. Algorithmic Accountability Policy Toolkit. *https://ainowinstitute.org/aap-toolkit.pdf*

Alston, Philip. 2019. *Report of the Special Rapporteur on Extreme Poverty and Human Rights*. Promotion and Protection of Human Rights: Human Rights Questions, Including Alternative Approaches for Improving the Effective Enjoyment of Human Rights and Fundamental Freedoms. A/74/48037. Seventy-Fourth Session. Item 72(b) of the Provisional Agenda.

Ananny, M., & Crawford, K. (2018). Seeing Without Knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability. *New Media & Society*, 20(3), 973–989. https://doi.org/10.1177/1461444816676645.

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

Ballantyne, N. (2019). The Ethics and Politics of Human Service Technology: The Case of Predictive Risk Modeling in New Zealand's Child Protection System. January 2019. *The Hong Kong Journal of Social Work* 53 (01n02):15-27 DOI: 10.1142/S0219246219000044.

Benjamin, R. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code* (Cambridge, U.K.: Polity, 2019).

Berryhill, J., Kok Heang, K., Clogher, R. & McBride, K. (2019). *Hello, World: Artificial Intelligence and its Use in the Public Sector*. O.E.C.D.

Brevini, B., & Pasquale, F. (2020). *Revisiting the Black Box Society by Rethinking the Political Economy of Big Data. Big Data & Society. https://doi.org/10.1177/2053951720935146.*

Bridle, J. (2018). New Dark Age: *Technology, Knowledge and the End of the Future* (London: Verso Books, 2018).

Bright, J., Bharath, G., Seidelin, C., & Vogl, T. M. (2019). *Data Science for Local Government* (April 11, 2019). Available at SSRN: *https://ssrn.com/abstract=3370217 or http://dx.doi.org/10.2139/ssrn.3370217.*

Clayton, V., Sanders, M., Schoenwald, E., Surkis, L. & Gibbons, D. (2020) Machine Learning in Children's Services Summary Report. What Works for Children's Social Care. U.K.

Couldry, N., & Mejias, U. (2019). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television and New Media*, 20(4), 336-349.

Cubillos Celis, P. (2019). *Neoliberalismo, Focalización e Infancia en Chile*: 1973-2010.

Revista Mexicana de Sociología, 81(3). doi: *http://dx.doi.org/10.22201/iis.01882503p.2019.3.57923.*

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. First Edition*. (New York, N.Y.: St. Martin's Press, 2018).

Fjeld, Jessica and Achten, Nele and Hilligoss, Hannah and Nagy, Adam and Srikumar, Madhulika, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for A.I. (January 15, 2020). *Berkman Klein Center Research Publication No. 2020-1*, Available at SSRN: https://ssrn.com/abstract=3518482 or http://dx.doi.org/10.2139/ssrn.3518482

López, J. (2020). *Experimentando con la Pobreza: El SISBÉN y Los Proyectos de Analítica de Datos en Colombia*. Fundación Karisma. Colombia.

Masiero, S., & Das, S. (2019). Datafying Anti-Poverty Programmes: Implications for Data Justice. Information, *Communication & Society*, 22(7), 916-933.

Mohamed, S., Png, M. & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*. https://doi.org/10.1007/s13347-020-00405-8

Muñoz Arce, G. (2018): The Neoliberal Turn in Chilean Social Work: Frontline Struggles Against Individualism and Fragmentation. *European Journal of Social Work*, 22(2), 289-300. DOI: 10.1080/13691457.2018.1529657.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York, N.Y.: New York University Press, 2018).

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

Pena, P.; Varon, J. (2019). "Consent to Our Data Bodies: Lessons from Feminist Theories to Enforce Data Protection." *Coding Rights*.

Silvia, M.; Varon, J. (2021)."Facial Recognition in the Public Sector and Trans Identities: Techno-Politics of Control, Surveillance, and Threats to Gender Diversity in Their Intersectionality of Race, Class, and Terri-tory." *Coding Rights*.

Schöngut, N. (2017). Ensamblajes Socio-Técnicos para la Producción de Intervenciones Psicosociales en un Programa del Servicio Nacional de Menores de Chile. Psicoperspectivas. *Individuo y Sociedad*, 16(3).

Taylor, L. (2017). What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally. *Big Data & Society*, 4(2), 1-14. https://journals.sagepub.com/doi/10.1177/2053951717736335

Tendayi Achiume, E. (2020). Racial Discrimination and Emerging Digital Technologies: a Human Rights Analysis. *Report of the Special Rapporteur on Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance*. A/HRC/44/57. Human Rights Council. Forty-Fourth Session. 15 June–3 July 2020.

Tijoux, M.E. (2016). Racismo en Chile. La Piel como Marca de la Inmigración. Editorial: *Universitaria*. Santiago, Chile.

Vaithianathan, R., Benavides, D., Taskova, K., Hermosilla, MP., Letelier, A., Escobar, J., Garretón, M. & Kulick, E. (2019). *Informe Final: Propuesta de Instrumento de Focalización para la Identificación de Niños, Niñas Y Adolescentes en Riesgo de Vulneración de Derechos*. "Construcción del Instrumento de Focalización Alerta de Niñez". 730566-13-LQ18. 27 de junio de 2019. Universidad Adolfo Ibáñez (UAI), Chile y Auckland Uni-versity of Technology (AUT), Nueva Zelanda.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Hu-man Future at the New Frontier of Power* (London: Profile Books, 2019).

Carr Center for Human Rights Policy
Harvard Kennedy School
79 JFK Street
Cambridge, MA 02138

carrcenter.hks.harvard.edu

79 JFK Street | Cambridge, MA 02138
617.495.5819