# Not My AI

### A feminist framework to challenge algorithmic decisionmaking systems deployed by the public sector

#### Introduction

In the hype of artificial intelligence (AI), we are observing a world where states are increasingly adopting algorithmic decision-making systems as a magic wand that promises to "solve" social, economic, environmental and political problems. As if machines were able to erase societal biases and structural inequalities, instead of just automating them, we are gradually observing states using narratives around tech innovation for spending public resources in questionable ways, sharing sensitive citizen data with private companies and, ultimately, dismissing any attempt of a collective, democratic and transparent response to core societal challenges.

Latin America is not an exception. Throughout the region, governments are in the stage of testing and piloting a wide variety of AI systems to deploy public services. In an initial mapping exercise,<sup>1</sup> we could identify five trending areas: education; judicial system; policing; public health, and social benefits. Among such trends, we decided to focus our case-based analysis on AI projects applied in the overlap of education and distribution of social benefits. What are the feminist and human rights implications of using algorithmic decision-making to determine the provision of social benefits and other public services? As machines are designed and operated by the very same humans in power, these AI systems are mostly likely to cause or propagate harm and discrimination based on gender and all its intersectionalities of race, class, sexuality, age, territoriality, therefore, posing worrisome trends that should be of concern to feminist movements.

Taking Latin America as a point of departure, as it is where we both as researchers and feminists originate from, this investigation seeks to contribute to the development of an anti-colonial feminist framework to question AI systems that are being deployed by the public sector, particularly focused social welfare programmes. Our ultimate goal is to develop arguments that enable us to build bridges for advocacy with different human rights groups, particularly feminists and LGBTIQ groups, especially in Latin America, but not only. We hope that, in collectivity, we can foster conversations towards an overarching anti-colonial feminist critique to address governmental trends of adopting AI systems that are not only disregarding human rights but are also, once again, replicating heteropatriarchy, white supremacy and colonialism through neoliberal techno-solutionist narratives exported to the world by Silicon Valley.

<sup>&</sup>lt;sup>1</sup>More information in the Annex I

This article is the result of research conducted by the authors in close collaboration with the Feminist Research Network (FIRN)<sup>2</sup> and currently composes the core structure of the notmy.ai platform. The platform continues to be developed with the goal to increase critical thinking through a series of conversations around the development of a feminist toolkit to question algorithmic decisions-making systems that are being deployed by the public sector. Going beyond the liberal approach of human rights, feminist theories and practices, it builds political structures for us to imagine other worlds based on solidarity, equity and social-environmental justice. As AI is gradually pervading several issues that are in the core of feminist agendas, the need for supporting feminist movements to understand the development of these emerging technologies becomes key in order to fight against automatised social injustice and to imagine feminist futures. Therefore, this report seeks to bring the feminist movements closer to the social and political problems that many algorithmic decisions carry with them. To reach such end, we start by posing three research questions:

- What are the leading causes of governments implementing AI and other methods of algorithmic decision-making processes in Latin America to address issues of public services?
- What are the critical implications of such technologies in the enforcement of gender equality, cultural diversity, sexual and reproductive rights?
- How can we learn from feminist theories to provide guidelines to balance the power dynamic enforced by the usages of AI and another algorithmic decision-making systems?

To address them, this text is divided into four sections. We start by addressing the overarching question of this work: Why AI is a feminist issue? We want to address this inquiry empirically, starting from an initial mapping of AI systems being deployed by the public sector in Chile, Brazil, Argentina, Colombia, Mexico and Uruguay to determine the provision of social benefits and other public services, but actually, are more likely to be causing harm and challenging feminist agendas. Then we review critical thinking around AI used in the so-called Digital Welfare Systems towards drafting a feminist framework to grasp what would constitute an oppressive AI. Then we dig deeper into two cases in which AI is being deployed in distribution of social benefits and educational systems in the region: the Childhood Alert System in Chile and a system to predict school dropouts and teenage pregnancy developed for Microsoft Azure in partnership with governments from Argentina and Brazil. These case analyses will be based on an anti-colonial feminist approach, and not only human rights, as one of the starting points to interrogate the algorithmic decisions and will serve as a test of the oppressive AI framework, drafted as empirical feminist categories to understand power dynamics behind automated decision-making systems. This report ends with considerations about the next steps of notmy.ai

3

<sup>&</sup>lt;sup>2</sup><u>https://firn.genderit.org</u>

towards using oppressive AI framework as a first tool to expand the conversations about feminist implications in the deployment of AI systems. In addition, more positively, the report concludes with the potential of hacking oppression by envisioning transfeminist technologies through feminist values that were brainstormed in a series of workshops conducted with the Oracle for Transfeminist Technologies. In this way, we can foresee the power of conversations that playfully envision speculative transfeminist technologies as a tool to take us from imagination to action.

#### Why AI is a feminist issue?

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. Referring to the term "Digital Welfare States", the former United Nations Rapporteur on Extreme Poverty and Human Rights, Philip Alston, has criticised 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."<sup>3</sup> Particularly within the U.S.A., where some of these projects have been developed further than pilot phases, confronted with the evidence on bias and harm caused by automated decisions, AI programmes deployed in public services have faced criticism on several fronts.<sup>4</sup> More recently, governments in Latin America are also following this hype, sometimes with the support of U.S.A. companies that are using the region as a laboratory of ideas which, perhaps fearing criticism in their home countries, are not even tested in the U.S.A. first.

With the goal to build a case-based, anti-colonial feminist critique to question these systems from perspectives that go beyond well-put criticisms from the global North, through desk research and a questionnaire<sup>5</sup> 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 Tendayi Achiume, Special Rapporteur on contemporary forms of racism, racial discrimination, xenophobia and related intolerance, poses in the report "Racial discrimination and emerging digital technologies",<sup>6</sup> databases used in these systems are the product of human design and can be biased in various ways, potentially leading to – intentional or

<sup>&</sup>lt;sup>3</sup>Alston, P. (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. <sup>4</sup>Examples include: O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown.; Eubanks, V. (2018). Automating inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.; Noble, S. U. (2018). Algorithms of Oppression: How Search engines Reinforce Racism. New York University Press.; Benjamin, R. (2019). Race after technology: Abolitionist Tools for the New Jim Code. Polity.

<sup>&</sup>lt;sup>5</sup>https://notmy.ai/do-you-know-other-projects

<sup>&</sup>lt;sup>6</sup>Tendavi Achiume, E. (2020, 15 June). Racial discrimination and emerging digital technologies: a human rights analysis. A/HRC/44/57. Human Rights Council. Forty-fourth session.

https://www.ohchr.org/EN/HRBodies/HRC/RegularSessions/Session44/Documents/A HRC 44 57 AdvanceEditedVe rsion.docx 4

## unintentional – discrimination or exclusion of certain populations, in particular, minorities as based on racial, ethnic, religious and gender identity.<sup>7</sup>

As a result, as of April 2021, we have mapped 24 cases with likely harmful implications on gender equality and all its intersectionalities in Chile, Brazil, Argentina, Colombia, Mexico and Uruguay<sup>8</sup>, which we were able to classify into five categories: judicial system, education, policing, social benefits and public health. Several of them are in an early stage of deployment or developed as pilots.

It is important to highlight that this mapping was not intended to present an overall and comprehensive record of all the existing cases of AI deployed by the public sector in Latin America that might have such harmful implications. That is a particularly difficult task, mostly if we consider the lack of transparency about these projects that exists in many of our countries and very common press announcements full of shiny promises that are then difficult to follow through other channels. The reason we left an open form at notmy.ai was to continue collecting information on new projects and possible harms. Nevertheless, above anything, our mapping had a less ambitious goal which was to point to general trends about the areas of application and collect evidence that shows that AI in the public sector is already a reality in the region which demands critical opinion and awareness raising.

5

<sup>&</sup>lt;sup>7</sup>Ibid. <sup>8</sup>https://notmy.ai/mapping-of-projects/



#### A.I. PROJECTS IN THE PUBLIC SECTOR IN LATIN AMERICA

BIAS AND DISCRIMINATION ON GENDER AND ITS INTERSECTIONALITIES



The database of these cases is accessible here: https://notmy.ai/mapping-of-projects/

It should be recognised that part of the technical community has made various attempts to mathematically define **"fairness"**, and thus meet a demonstrable standard on the matter. Likewise, several organisations, both private and public, have undertaken efforts to define ethical standards for AI The very useful data visualisation "Principled Artificial Intelligence", from the Berkman Klein Center,<sup>9</sup> shows the diversity of ethical and human rights-based frameworks that emerged from different sectors from 2016 onwards with the goal to guide the development and use of AI 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."<sup>10</sup> Nevertheless, as we can see from that list, **none of this consensus is driven by a debate on** 

<sup>&</sup>lt;sup>9</sup>https://wilkins.law.harvard.edu/misc/PrincipledAI\_FinalGraphic.jpg

<sup>&</sup>lt;sup>10</sup>Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*. Berkman Klein Center Research Publication,

**power dynamics or automation of oppressions.** Instead of asking how to develop and deploy an AI system, shouldn't we be asking first: "Why build it?"; "Is it really needed?"; "On whose request?"; "Who profits?"; "Who loses from the deployment of a particular AI system?"; "Is it oppressing a particular group of the population?"; "Should it even be developed or deployed at all?"

We believe these are some of the guiding questions to be asked when considering an overarching question: How would a feminist framework to question AI systems look? Trying to depart from empirical experiences, so far, we have endured a task to analyse possible harms by AI programmes deployed in the areas of education and social benefits in Chile, Argentina and Brazil.<sup>11</sup> As a result, based on both our bibliographic review and our case-based analysis, we are gradually expanding an empirically tested case-based framework to serve as one of the instruments for our feminist anti-colonial toolkit to help us to pose structural questions about whether a given governmental AI system may incur possible harm to several feminist agendas.

8

<sup>(2020)1. &</sup>lt;u>https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3518482</u># 11\_\_\_\_\_https://notmy.ai/

### **Oppressive AI – empirical feminist categories to understand power dynamics behind automated decisionmaking systems**

Based on an overall bibliographical review of criticisms that are being posed to AI systems deployed by the public sector, and also based on our findings from the two case-based analyses, the scheme below is an attempt to create a work-in-progress framework of analysis that goes beyond the discourses of fairness, ethical or human-centric AI and seeks a holistic structure that considers power relations to question the idea of deploying AI systems in several helms of the public sector.



Oppressive AI Framework by Joana Varon and Paz Peña. Design by Clarote for notmy.ai by Coding Rights

Next, we will briefly explain and refer to the bibliographical review that substantiated that framework and, afterwards, apply it to our case-based analysis from pilot experiments in three countries from Latin America:

# A. Surveillance of the poor: turning poverty and vulnerability into machine-readable data

Analysing the case of the U.S.A., Virginia Eubanks shows how the usage of AI systems are 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.** She mentions that most of these programmes take advantage of the <u>tradition of s</u>tate surveillance on vulnerable populations,<sup>12</sup> which is to turn their existence into data and use algorithms to determine the provision of social benefits by the states.

In a similar take, Linnet Taylor, in her article "What is data justice?" says that "granular data sources enable authorities to infer people's movements, activities, and behaviour, not without having ethical, political, and practical implications of how the public and private sector view and treat people".<sup>13</sup> 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 classification systems that profile, judge, punish and surveil. In this way, poverty and vulnerability is turned into machine-readable data, with real consequences on the lives and livelihoods of the citizens involved.<sup>14</sup> Likewise, Cathy O'Neil, also analysing the usages of AI in the U.S.A., asserts that many **AI systems "tend to punish the poor"**,<sup>15</sup> 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.

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 "feminisation of poverty." In fact, the IV United Nations Conference on Women, held in Beijing in 1995, stated that 70% of poor people in the world were women. It is never enough to mention that the reasons why poverty affects women more commonly have nothing to do 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. In the case of poverty management programmes through big data and AI 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.**<sup>16</sup>

<sup>&</sup>lt;sup>12</sup>Eubanks, V. (2018). Automating inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

<sup>&</sup>lt;sup>13</sup>Taylor, L. (2017). *What is data justice? The case for connecting digital rights and freedoms globally*. Big Data & Society, 4(2). <u>https://journals.sagepub.com/doi/10.1177/2053951717736335</u>

<sup>&</sup>lt;sup>14</sup>Masiero, S., & Das, S. (2019). *Datafying anti-poverty programmes: implications for data justice*. Information,

Communication & Society, 22(7), 916-933. <u>https://www.tandfonline.com/doi/full/10.1080/1369118X.2019.1575448</u> <sup>15</sup>O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*.

Crown. <sup>16</sup>Castro, L. & López, J. (2021). Vigilando a las "buenas madres". Aportes desde una perspectiva feminista para la

investigación sobre la datificación y la vigilancia en la política social desde Familias En Acción. Fundación Karisma. Colombia.

#### **B. Embedded racism**

For the UN Special Rapporteur, E. Tandayi, emerging digital technologies should also be understood as **capable of creating and maintaining racial and ethnic exclusion in systemic or structural terms.**<sup>17</sup> This is also what tech researchers on race and AI in the U.S.A., to name a few, Ruha Benjamin,<sup>18</sup> Joy <u>Buolamwini,<sup>19</sup> Timnit Gebru<sup>20</sup> and Safiya Noble<sup>21</sup> highlight in their case studies. Likewise, focused in Latin America, researchers Nina da Hora,<sup>22</sup> Tarcisio Silva<sup>23</sup> and Pablo Nunes,<sup>24</sup> all from Brazil, have pointed to similar findings while investigating facial recognition technologies, police violence, criminal (in)justice systems and other oppressions. Ruha Benjamin points out how the use of new technologies reflects and reproduces the existing racial injustices in U.S.A. society, even though they are promoted and perceived as more objective or progressive than the discriminatory systems of an earlier era.<sup>25</sup> In this sense, for this author, when AI 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.</u>

From technology development itself, in her research, Noble demonstrates how commercial search engines such as Google not only mediate but are mediated by a series of commercial imperatives that, 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 analysing a series of Google searches where black women end up being sexualised by the contextual information the search engine throws up (e.g., linking them to wild and sexual women).<sup>26</sup>

A game change study by Buolamwini and Gebru<sup>27</sup> analysed three commercial facial recognition systems that include the ability to classify faces by gender. They found out 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. The authors attribute these race and gender biases to the composition of the data sets used to train these systems, which were overwhelmingly composed of lighter-skinned male-appearing subjects.

# C. Patriarchal by Design: sexism, compulsory heteronormativity, and gender binarism

<sup>26</sup>Noble, S. U. (2018). Algorithms of Oppression: How Search engines Reinforce Racism. New York University Press.
 <sup>27</sup>Buolamwini, J. & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:77-91, 2018.

<sup>&</sup>lt;sup>17</sup>Tendayi Achiume, E. (2020, 15 June). Op. cit.

<sup>&</sup>lt;sup>18</sup><u>https://www.ruhabenjamin.com</u>

<sup>&</sup>lt;sup>19</sup><u>https://www.media.mit.edu/people/joyab/overview/</u>

<sup>&</sup>lt;sup>20</sup>http://proceedings.mlr.press/v81/buolamwini18a.html

<sup>&</sup>lt;sup>21</sup><u>http://algorithmsofoppression.com</u>

<sup>&</sup>lt;sup>22</sup>https://www.ninadahora.dev

<sup>&</sup>lt;sup>23</sup>https://tarciziosilva.com.br/blog/

<sup>&</sup>lt;sup>24</sup>https://theintercept.com/equipe/pablo-nunes/

<sup>&</sup>lt;sup>25</sup>Benjamin, R. (2019). Race after technology: Abolitionist Tools for the New Jim Code. Polity.

Many AI systems work by sorting people into a binary view of gender,<sup>28</sup> as well as by reinforcing outdated stereotypes of gender and sexual orientation. Nevertheless, a recent 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 LGBTIQ people.<sup>29</sup>

Inspired by Buolamwini and Gebru, Silva and Varon, in 2021, researched how the deployment of facial recognition technologies affect transgender people<sup>30</sup> in its intersectionalities of race and territoriality, particularly when used by governmental authorities to authenticate identities to ensure access to public services. In an empirical analysis of Brazilian cases, the researchers could point out that there is little transparency about the accuracy rate (tracking false positives or false negatives), and that, when there is any data, there is no disaggregation considering the demographics of error rates. Meaning that, even though tech audits show that, in the current state of the art, these technologies fail on particular demographics, the government, deploying them as a means to access public services, is not keeping track of who is getting excluded and discriminated against.

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<sup>31</sup>, legislation to protect transgender people is practically nonexistent. They are not allowed recognition of their identity, which makes this technology resignify 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".

West, Whittaker and Crawford<sup>32</sup> argue that the diversity crisis in industry and bias issues in AI 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.

#### D. Colonial extractivism of data bodies and territories

<sup>&</sup>lt;sup>28</sup>Silva, M. R., & Varon, J. (2021, 31 March). Threats in the usage of facial recognition technologies for authenticating transgender identities. *Coding Rights*. <u>https://medium.com/codingrights/threats-in-the-usage-of-facial-recognition-technologies-for-authenticating-transgender-identities-e0cd602f9c60</u>

 <sup>&</sup>lt;sup>29</sup>Tomasev, N., Kay, J., McKee, K.R., & Mohamed, S. (2021). *Fairness for Unobserved Characteristics: Insights from Technological Impacts on Queer Communities.* DeepMind. <u>https://arxiv.org/pdf/2102.04257.pdf</u>
 <sup>30</sup>Silva, M.R., & Varon, J. (2021, 31 March). Op. cit.

<sup>&</sup>lt;sup>31</sup>Díaz Hernández, M. (2020) *Sistemas de protección social en Venezuela: vigilancia, género y derechos humanos.* In Sistemas de identificación y protección social en Venezuela y Bolivia. Impactos de género y otros tipos de discriminación. Derechos Digitales.

<sup>&</sup>lt;sup>32</sup>West, S.M., Whittaker, M., & Crawford, K. (2019). *Discriminating Systems: Gender, Race and Power in AI*. AI Now Institute.

Authors like Couldry and Mejias<sup>33</sup> and Shoshana Zuboff<sup>34</sup> review this current state of capitalism where the production and **extraction of personal data naturalise 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, appropriate the data.

Regarding colonialism and AI, Mohamed, Png and Issac examine how **colonialism presents itself in algorithmic systems through institutionalised "algorithmic oppression"** (the unjust subordination of one social group at the expense of the privilege of another), "**algorithmic exploitation"** (ways in which institutional actors and corporations take advantage of often already marginalised people for the asymmetric benefit of these industries) and "**algorithmic dispossession"** (centralisation of power in the few and the dispossession of many), in an analysis that seeks to highlight the **historical continuities of power relations**.<sup>35</sup>

Crawford<sup>36</sup> calls for a more comprehensive view of AI as a critical way to understand that these systems depend on exploitation: on the one hand, of energy and mineral resources, of the cheap labour, and in addition, of our data at scale. In other words, AI is an extractive industry. Even though neither Google, nor Grammarly spellcheckers recognise the word "extractivism".

All these systems are energy intensive and heavily dependent on minerals, sometimes extracted from areas where there are social-environmental conflicts. 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 Peña poses, 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."<sup>37</sup> Even beyond ecological impact, Camila Nobrega and Joana Varon also expose the green economy narratives altogether with techno-solutionisms are "threatening multiple forms of existence, of historical uses and collective management of territories", not by chance the authors found out

<sup>&</sup>lt;sup>33</sup>Couldry, N., & Mejias, U. (2019). *Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject*. Television & New Media, 20(4), 336-349. <u>https://journals.sagepub.com/doi/abs/10.1177/1527476418796632</u>

<sup>&</sup>lt;sup>34</sup>Zuboff, S. (2019). *The Age of Surveillance Capitalism: the Fight for a Human Future at the New Frontier of Power*. Profile Books.

<sup>&</sup>lt;sup>35</sup>Mohamed, S., Png, M., & Isaac., W. (2020). *Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence*. Philosophy & Technology, 33, 659-684. <u>https://doi.org/10.1007/s13347-020-00405-8</u>

<sup>&</sup>lt;sup>36</sup>Crawford, K. (2021). *Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

<sup>&</sup>lt;sup>37</sup>Tapia, D., & Peña. P. (2021). White gold, digital destruction: Research and awareness on the human rights implications of the extraction of lithium perpetrated by the tech industry in Latin American ecosystems. In A. Finlay (Ed.), *Global Information Society Watch 2020: Technology, the environment and a sustainable world: Responses from the global South.* Association for Progressive Communications (APC) & Swedish International Development Cooperation Agency (SIDA). <u>https://www.giswatch.org/node/6247</u>

that Alphabet Inc., Google parent company is exploiting 3TG minerals in regions of the Amazon where there is a land conflict with indigenous people.<sup>38</sup>

#### E. Automation of neoliberal policies

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.**<sup>39</sup> 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, individuals do not see themselves as subjects of rights but as service applicants.<sup>40</sup> In this sense, it is interesting to see that AI systems, in their neoliberal efforts to target public resources, also classify who the poor subject is through automated mechanisms of exclusion and inclusion.<sup>41</sup>

#### **F. Precarious Labour**

Particularly focused on AI and the algorithms of Big Tech companies, anthropologist Mary Gray and computer scientist Siddharth Suri point out the "ghost work"<sup>42</sup> or invisible labour that powers digital technologies. Labelling images and cleaning databases are manual work very often performed in unsavoury working conditions "to make the internet seem smart". Communalities of these jobs are very precarious working conditions, normally marked by overwork, underpaid, with no social benefits or stability – very different from the work conditions of the creators of such systems.<sup>43</sup> Who takes care of your database? As always, care work is not recognised as valuable work.

<sup>&</sup>lt;sup>38</sup>Nobrega, C., & Varon, J. (2021). Big tech goes green(washing): Feminist lenses to unveil new tools in the master's houses. In A. Finlay (Ed.), *Global Information Society Watch 2020: Technology, the environment and a sustainable world: Responses from the global South.* Association for Progressive Communications (APC) & Swedish International Development Cooperation Agency (SIDA). <u>https://www.giswatch.org/node/6254</u>

<sup>&</sup>lt;sup>39</sup>Arora, P. (2016). The Bottom of the Data Pyramid: Big Data and the Global South. *International Journal of Communication*, *10*, 1681-1699. <u>https://ijoc.org/index.php/ijoc/article/view/4297</u>

<sup>&</sup>lt;sup>40</sup>Alston, P. (2019). Op. cit.; Masiero, S., & Das, S. (2019). Datafying anti-poverty programmes: implications for data justice. *Information, Communication & Society, 22*(7), 916-933.

https://www.tandfonline.com/doi/full/10.1080/1369118X.2019.1575448

 <sup>&</sup>lt;sup>41</sup>López, J. (2020). Experimentando con la pobreza: el SISBÉN y los proyectos de analítica de datos en Colombia.
 Fundación Karisma. <u>https://web.karisma.org.co/wp-content/uploads/download-manager-</u>files/Experimentando%20con%20la%20pobreza.pdf

<sup>&</sup>lt;sup>42</sup>Gray, M.L., & Suri, S. (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass.* Harper Business.

<sup>&</sup>lt;sup>43</sup>Crawford, K. (2021). Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press.

#### G. Lack of Transparency

According to AINOW, 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.<sup>44</sup> Along the same lines, OECD\_postulates that transparency (on the part of) is strategic to foster public trust in the tool.<sup>45</sup>

More critical views comment on 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.<sup>46</sup> Thus, government intermediaries with special understanding and independence should play a role here.<sup>47</sup> Furthermore, Annany and Crawford suggest that what the **current vision of transparency in AI does it fetishise the object of technology,** without understanding that technology is an assembly of human and non-human actors.<sup>48</sup> Therefore, to understand the operation of AI, it is necessary to go beyond looking at the mere object. AI is not an object, it is an assemblage of values, peoples, places and processes – all need to be in the open.

Are these seven categories enough to think about a feminist framework to question AI systems? In the next chapter we will focus on using this proposed framework for analysing two cases from Latin America: the Childhood Alert System – SAN, in Chile and Plataforma Tecnológica de Intervención Social, from Argentina, exported to Brazil under the name Projeto Horus. While much of the analysis and framework of critique on AI is located in the U.S.A. (where many of AI projects have surpassed the pilot phase), this work, while building from important insights made especially by researchers and activists who questions discrimination of AI systems in the U.S.A., has Latin America as an empirical field, which we believe, is also an important contribution to deepen and broaden knowledge and engagement in this area as well as a decolonial endeavour.

 <sup>&</sup>lt;sup>44</sup>AINOW. (2018). Algorithmic Accountability Policy Toolkit. AINOW. <u>https://ainowinstitute.org/aap-toolkit.pdf</u>
 <sup>45</sup>Berryhill, J., Kok Heang, K., Clogher, R. & McBride, K. (2019). *Hello, World! Artificial intelligence and its use in the Public Sector*. OECD. <u>https://oecd-opsi.org/wp-content/uploads/2019/11/AI-Report-Online.pdf</u>

<sup>&</sup>lt;sup>46</sup>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

<sup>&</sup>lt;sup>47</sup>Brevini, B., & Pasquale, F. (2020). *Revisiting the Black Box Society by rethinking the political economy of big data*. Big Data & Society, 7(2). <u>https://doi.org/10.1177/2053951720935146</u>

<sup>&</sup>lt;sup>48</sup>Ananny, M., & Crawford, K. (2018). Op. cit.

### **II.** Case studies

# 1. A Childhood Alert System (SAN, in its Spanish acronym, for Sistema Alerta Niñez), Chile<sup>49</sup>

#### **General description**<sup>50</sup>

The Childhood Alert System (SAN)<sup>51</sup> is a software that provides complimentary input to data collected in Chile by the so-called Local Childhood Offices (OLN, in its Spanish acronym, for Oficinas Locales de la Niñez) in order to support decision making by managers of these OLNs when offering support to children and their families, considering the individual needs of each child and adolescent (NNA, in its Spanish acronym, for Niños, Niñas y Adolescentes), as well as family and environmental factors.

Several requests for public information submitted to the government have made it clear that SAN is one additional tool – and not the main one – available for these managers to prioritise cases identified through on-site alerts, that is, those generated and verified by people who interact directly with NNA and their families. However, as Derechos Digitales has pointed out in\_its report<sup>52</sup> on the system, the AI system was first designed, and then came the institutionalism with the creation of the OLNs.

The software is based on the use of Predictive Risk Modeling (PRM). Its purpose is to identify NNA who are at risk of violation of rights and the families that require support to enhance their role of providing child protection, **based on administrative data**, **detecting a set of individual**, **family**, **environmental and peer conditions that tend to occur when there is a high risk of violation of rights**.

The idea is that the tool is run periodically (for example, once a month) through the list of NNA in the databases of the Ministry of Social Development and Family (MDSF, in its Spanish acronym, for Ministerio de Desarrollo Social y Familia). The tool scores each individual. The score can then be used to prioritise NNA and families already known to the software, such as those identified by professionals who interact with NNA and their families in the social protection system, as in the case of the workers at Chile Grows With

<sup>&</sup>lt;sup>49</sup>The extract of the Argentinian component of this case study was developed and originally publish in the article "Decolonize AI: a feminist critique towards data and social justice", written by the same authors of this report for the publication Giswatch: Artificial Inteligence Human Rights 2019, available at: https://www.giswatch.org/node/6203 <sup>50</sup>This short questionnaire remains open, if you got to know a project that you think should be included in the mapping,

<sup>&</sup>lt;sup>50</sup>This short questionnaire remains open, if you got to know a project that you think should be included in the mapping, please, send a link or further information accessing this form: <u>https://notmy.ai/do-you-know-other-projects/</u>. You can do it anonymously or if you wish, be credited in the mapping of that project after we double check the information. <sup>51</sup><u>https://notmy.ai/project-item/sistema-alerta-ninez-san-2/</u>

<sup>&</sup>lt;sup>52</sup>Valderrama, M. (2021). *Chile: Sistema Alerta Niñez y la predicción del riesgo de vulneración de derechos de la infancia.* Derechos Digitales & IRDC / CRDI. <u>https://www.derechosdigitales.org/wp-</u>

<sup>&</sup>lt;u>content/uploads/CPC informe Chile.pdf</u>; León, P. (2019, 29 January). Alerta Infancia: el software que expone los datos personales de niños y niñas en riesgo social. *diaroUchile*. <u>https://radio.uchile.cl/2019/01/29/alerta-infancia-el-software-que-expone-los-datos-personales-de-ninos-y-ninas-en-riesgo-social/</u>

You (ChCC, in its Spanish acronym, for Chile <u>Crece Contigo</u><sup>53</sup>). In addition, NNA, who obtain a higher risk score and who were not yet in contact with other state programmes, could be proactively contacted. According to the government, prioritisation has a preventive purpose, under the responsibility of the Local Childhood Office, which will provide support and prevention programmes beneficial to the NNA and their family.

To identify NNA in these circumstances, it was necessary to train the model **through the analysis of NNA who had already been subject to violation of their rights, to then study their life course and identify the family conditions experienced by these NNA 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.

#### Who develops SAN pilot?

The pilot and the related consultancy were developed in partnership between two universities: the Centre for Social Data Analytics (CSDA) at the Auckland University of Technology (AUT Ventures Limited) and the Public Innovation Laboratory (GobLab) at the Universidad Adolfo Ibáñez.

The pilot – developed with public funds through a public tender process – was awarded to these two universities after the public announcement made in March 2018 by Chilean President Sebastián Piñera about the Great National Agreement for Children, which included among its measures the creation of an early warning system, called "Childhood Alert". The director of GobLab invited professor Rhema Vaithianathan, co-director of the Centre for Social Data Analytics, to Chile. In Santiago, they had a\_series of meetings with authorities.<sup>54</sup>

#### **SAN** audits

According to government information accessed through our transparency request, "after a few months of operation of the Childhood Alert Pilot System, the need to start an algorithmic audit was identified in order to study and evaluate the hypothetical existence of biases, both in the data used for training the model and the characteristics used for the production model".<sup>55</sup> The results would imply implementing model enhancement. This algorithmic audit was financed by the Inter-American Development Bank (IDB) and performed by the Spanish consultancy Eticas Research

<sup>&</sup>lt;sup>53</sup>https://www.crececontigo.gob.cl/acerca-de-chcc/que-es/

<sup>&</sup>lt;sup>54</sup><u>https://csda.aut.ac.nz/news-and-events/2018/csda-shares-child-welfare-predictive-risk-modelling-work-in-chile;</u> <u>https://noticias.uai.cl/modelos-predictivos-de-riesgo-para-proteger-a-la-infancia/</u>

<sup>&</sup>lt;sup>55</sup>Candia, A. (2020). Carta N°020 /3805. Available in annex 1, entitled 1Carta\_Nro\_3805\_Fol\_02359. https://nuvem.codingrights.org/index.php/s/F675xEHTHCjSPkz

Consulting. The Chilean government rejected the possibility of learning about the results of this audit.

#### International controversies over the system

In this section, we refer specifically to the controversies faced by CSDA and their projects:

• Allegheny Family Screening Tool (AFST), U.S.A.

Allegheny Family Screening Tool (AFST), is a predictive risk modelling tool designed to assist with child welfare call-screening decisions in Allegheny County, Pennsylvania, U.S.A. The Centre for Social Data Analytics developed the algorithm.

In her book, "Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor", Dr. Virginia Eubanks refers to AFST as an example of a system that oversamples households that rely on public assistance programmes and, thus tends to over-surveillance and over-punish poor and working-class families.56

A report delivered to Chile by the Centre for Social Data Analytics analyses the AFST case and refers to the controversy with Eubanks as follows: "One exception to the general trend of positive reports was a book by Virginia Eubanks (2018). She was concerned that AFST stood for poverty profiling, that is, that it would culminate in more children being removed from their families just because they were poor. The County disputed her claims and posted a point-by-point rebuttal on their website".57

Effectively, after this public controversy with the county, Eubanks declared: "My larger point is not that we must wait to entirely eradicate poverty before we try to improve Children, Youth and Family Services, but that the AFST only has access to data collected on families using public services, not on those that access private resources for parenting support. Because this will result in higher risk scores and more scrutiny of poor and working-class families, I believe the system is unjust and discriminatory".58

#### Vulnerable Children PRM, New Zealand

Initiative undertaken by the New Zealand government between 2011 and 2015 to develop an algorithm that would allow child protection services to predict future child maltreatment at the

<sup>58</sup>https://www.sostenibilidad.com/desarrollo-sostenible/licencia-social-operar-dialogo-requisito-empresarial/

<sup>&</sup>lt;sup>56</sup>Eubanks, V. (2018). Op. cit.

<sup>&</sup>lt;sup>57</sup>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". Universidad Adolfo Ibáñez (UAI), Chile y Auckland University of Technology (AUT), Nueva Zelanda

point of birth and to preempt it before it occurs.<sup>59</sup> Two models were created, but neither was implemented. Version 1 was intended for use with families who registered in the public welfare system. Version 2 was for use when a child was born. The project was led by Dr Rhema Vaithianathan, then of the University of Auckland, now at CSDA, in order to build a Predictive Risk Model to be tested retrospectively for children born between 2003 and 2006. The goal was to check the model's predictions against what had actually happened to the children. It was an observational study purely. No policies changed and no frontline worker would have accessed the prediction scores.<sup>60</sup>

There is no clear information on why system implementation was halted, but apparently, the Minister responsible had ethical reservations that the observational study might identify children at risk but not really address any case.<sup>61</sup> The discomfort with the project also related to its substance. Anne Tolley, then the Social Development Minister, told Stuff in 2015:<sup>62</sup> "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."<sup>63</sup>

#### **Controversies over SAN within Chile**

As soon as the creation of this system was publicly announced, 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", also "socially validating forms of stigmatization, discrimination and even criminalization of cultural diversity existing in Chile".<sup>64</sup> 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." In this sense, Francisca Valverde, from the group of organisations Bloque por la Infancia and executive director of Asociación Chilena Pro

<sup>&</sup>lt;sup>59</sup>Ballantyne, N. (2019). *The Ethics and Politics of Human Service Technology: the Case of Predictive Risk Modeling in New Zealand's Child Protection System.* The Hong Kong Journal of Social Work, 53(01n02),15-27. https://doi.org/10.1142/S0219246219000044

<sup>&</sup>lt;sup>60</sup>https://notmy.ai/mapping-of-projects/

<sup>&</sup>lt;sup>61</sup>Ballantyne, N. (2019). Op. cit.

<sup>&</sup>lt;sup>62</sup>Including a version of this writing in Portuguese and Spanish to be launched soon to engage more people in these conversations.

<sup>&</sup>lt;sup>63</sup> Kirk, S. (2015, 30 July). Children 'not lab-rats' – Anne Trolley intervenes in child abuse experiment. *Stuff*. <u>https://www.stuff.co.nz/national/health/70647353/children-not-lab-rats---anne-tolley-intervenes-in-child-abuse-experiment</u>

<sup>&</sup>lt;sup>64</sup>Sociedad Civil de Chile Defensora de los Derechos Humanos del Niño et al. (2019, 28 enero). Día Internacional de la protección de datos. Carta abierta de la Sociedad Civil de Chile Defensora de los Derechos Humanos del Niño. ONG Emprender con Alas.

Naciones Unidas (ACHNU), argues<sup>65</sup> that this type of system stigmatises poor children and does not emphasise the protection of children and adolescents in different territories, including those of higher social classes.

#### Analysis

In its public addresses, the Chilean government highlights "Childhood Alert" as a neoliberal targeting mechanism of social programmes rather than an AI software or system. The reason for that may be strategic, given the controversy that systems similar to SAN have faced both in the U.S.A. and in New Zealand. This makes a lot of sense when government explanatory documents, provided pursuant to transparency legislation, repeatedly emphasise that "Childhood Alert" is one more source of information for Local Childhood Offices, but that the final decision is made by humans. In fact, it should be noted that this last remark is one of the recommendations made by CSDA (New Zealand) and GobLab (Chile), which analysed the AI systems from the US and New Zealand referred to above (some of which were developed by part of the CSDA team).

However, it seems difficult to prove with evidence – at this point where it is still a recent model – that the results produced by an AI system specially designed for the occasion are not considered as a neutral fact by social workers. On the one hand, there may be pressures for them to do so, due to the investment made in the system and because the Local Childhood Offices were designed considering this technological tool. Also, it has not yet been studied how the system interacts with professionals and whether the alleged neutrality of digital technologies ends up influencing final decision-making. How do humans work alongside AI in social welfare related schemes? That is perhaps an interesting research question to explore in the near future.

In any case, the risk score generated by "Childhood Alert" seems to serve the logic of neoliberal public policies that Chile has implemented, especially with regard to children, from the dictatorship of Augusto Pinochet (1973-1990) and onwards.

### Context of Neoliberal policies in the Latin American and the dangers of amplifying them through automatisation

In order to bring some context, it is important to refer to the fact that there were two waves of neoliberalism recognised in Latin America. The first was through several coups d'état in the Southern Cone in the 1970s, with a more orthodox model implementation. The second big wave came about due to the debt crisis in countries such as Peru, Argentina and Brazil. With this, economic policies were strongly constrained by the structural adjustments of international financial institutions such as the IMF and the World Bank.<sup>66</sup>

<sup>&</sup>lt;sup>65</sup>Valderrama, M. (2021). Op. cit.

<sup>&</sup>lt;sup>66</sup>Rodríguez, J.P. (2021). The politics of neoliberalism in Latin America: dynamics of resilience and contestation. *Sociology Compass*, *15*(3). <u>https://compass.onlinelibrary.wiley.com/doi/full/10.1111/soc4.12854</u>

An essential part of the neoliberal reconfiguration of the relationship between society and the state is seen with social services. The reforms to social protection systems in Latin America have been particularly strong as they are part of the economic adjustment policies demanded by international financial institutions, aimed at privatisation, decentralisation, targeting and emergency social programmes (funds).<sup>67</sup> In Chile, for example, social policy was transformed into services conditioned by relationships between individuals and private service provider companies that left behind the articulation of diverse collective social subjects that directly confronted the states.<sup>68</sup>

Some of the criticisms of these measures are that they introduce precariousness and discontinuity to social policy, tending to make it welfarist, arbitrary, stigmatising of beneficiaries and, according to Sônia M. Draibe, "the deep segmentation that can be caused in the citizenry, through the duplication of social policy: a policy for the poor (in general a poor policy) next to a policy for the rich (in general, rich, sophisticated and often also financed with public resources)".<sup>69</sup> So, we should ask, how technology automates and grants the degree of technological resolutions to ideological decisions based on neoliberal principles?<sup>70</sup>

The vulnerable childhood approach is a classic neoliberal take, enforced by entities like the World Bank in the region, and it comes from the idea of poverty as an individual problem (not a systemic one) and caseworkers as protectors of people "at-risk".<sup>71</sup> Like in the case of SAN, the state approach to childhood is permeated by discourse that conceives linear models of individual development, which consider almost exclusively the individual psychic dimension, without inserting it into governmental processes that produce consequences in other equally relevant dimensions of the children. In other words, childhood is seen not as a subject capable of affecting the social fabric but as an object that must be protected from the adult world.<sup>72</sup> This can be seen, for example, in the use of the word "symptom" (usually related to some disease) by the undersecretary of children, when referring to SAN:

We are implementing a warning system that is called "Childhood Alert", which will aim to forecast the first symptoms that, for example, may lead a child to drop out of school, have their first contact with drugs or commit a crime for the first time, and thus start understanding directly what is happening to that child, to their family. All of this will be

 <sup>&</sup>lt;sup>67</sup>Draibe, S. M. (1994). Neoliberalismo y políticas sociales: Reflexiones a partir de las experiencias latinoamericanas. *Desarrollo Económico, 34*(134), 181-196. <u>https://www.jstor.org/stable/3467315?origin=crossref</u>
 <sup>68</sup>Taylor, M. (2003). The Reformulation of Social Policy in Chile, 1973—2001. *Global Social Policy: An Interdisciplinary Journal of Public Policy and Social Development, 3*(1), 21–44.

https://doi.org/10.1177%2F14680181030030010101

<sup>&</sup>lt;sup>69</sup>Draibe, S. M. (1994). Op. cit.

<sup>&</sup>lt;sup>70</sup>Alston, P. (2019). Op. cit.

<sup>&</sup>lt;sup>71</sup>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. <u>https://doi.org/10.1080/13691457.2018.1529657</u>

<sup>&</sup>lt;sup>72</sup>Schöngut-Grollmus, 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).* https://www.psicoperspectivas.cl/index.php/psicoperspectivas/article/view/1049

done through information crossing, which will allow us to determine which children are in the alert phase or at risk.

Once again, vulnerability is associated with the concept of "social disease" (determined by school dropout, drugs and crime). There are symptoms of this disease that focus not only on the NNA but also on their environment. AI, then, becomes an automated diagnostic tool that allows "ordering the queue" (another metaphor frequently used by the authorities to refer to AI) to make more efficient the programmes offered by the state to prevent social diseases. In this sense, instead of offering a system focused on social guarantees, the Chilean state automates targeting decisions, which is a classic mechanism of neoliberal policies, especially for children, in Chile.<sup>73</sup>

In this sense, SAN is actually an excellent opportunity to reflect on whether these systems let the state arrive earlier and better prepared to help children and adolescents on social risk, or in the end, it simply follows the neoliberal principle of controlling social expenditure, making more "efficient" individual interventions.

Thus, what SAN promises to seek, further than scoring children and adolescents, is to reduce spending through what is considered increased "effectiveness" of programmatic interventions, offering neoliberal technocratisation as the foundation for automated decisions rendered by the latest technology. But effective for whom?

This case study shows that it's design and use by the state responds to a continuum of neoliberal policies that have abounded in Latin America, to varying degrees, during the last 40 years.<sup>74</sup> On the one hand, it uses instruments that automate and grant the degree of technological resolutions to ideological decisions: resources' focalisation.<sup>75</sup> And, in this case, it is mainly about using big data to produce a more detailed category of poor children and adolescents<sup>76</sup> and, a step further, automate their social risk determination.

In continuity with neoliberal logic, SAN responds to the technocratisation of public policies. By the way, this instrument has been designed and deployed without an open conversation in the country. The degrees of participation in their design and the process's transparency are doubtful, which results in an impact on democracy itself. In particular, the same people affected by these systems – poor children, adolescents and their families – are not even the subject of consultation since they are not recognised as interested parties. This is even more concerning if we consider that the protests that took Chilean streets were specifically

<sup>&</sup>lt;sup>73</sup>Cubillos Celis, P. (2019). Neoliberalismo, focalización e infancia en Chile: 1973-2010. *Revista Mexicana de Sociología, 81*(3). <u>http://www.scielo.org.mx/pdf/rms/v81n3/2594-0651-rms-81-03-611.pdf</u>

<sup>&</sup>lt;sup>74</sup>López Solano, J. (2020). Experimentando con la pobreza: el SISBÉN y los proyectos de analítica de datos en Colombia. Fundación Karisma. <u>https://web.karisma.org.co/wp-content/uploads/download-manager-files/Experimentando%20con%20la%20pobreza.pdf</u>

<sup>&</sup>lt;sup>75</sup>Alston, P. (2019). Op. cit.

<sup>&</sup>lt;sup>76</sup>López Solano, J. (2020). Op. cit.

against neoliberal policies deployed in the region in the last years, a public outcry that has resulted in a new constitutional process for Chile.

Ironically, there is even evidence that the use of AI to predict possible vulnerabilities not only does not work well in the social care of children and adolescents,<sup>77</sup> but also ends up being quite costly for the states, at least in the firsts stages, contradictorily even to the neoliberal doctrine.<sup>78</sup> The hasty adoption of these neoliberal instruments is also worrisome because it is hard to dissolve them once adopted.<sup>79</sup>

Even so, it is interesting to watch how authorities take advantage of the innovative side of technologies to present as revolutionary an instrument that, on the contrary, perpetuates a policy that in Chile is over 40 years old. The undersecretary of Social Evaluation, Alejandra Candia, said<sup>80</sup>:

This innovative preventive targeting instrument will allow the Government to change the course of life of hundreds of children at risk of vulnerability, allowing us to reach those who require help in time to avoid complex situations that often end up being irreversible for them and also for their families.<sup>81</sup>

Developers excuse: "service acceptability" as a shield against social accountability and "social license" as an excuse for data extractivism

In contrast to the discourse of the government, there is that of the developers of the SAN pilot, carried out in the academic environment by two specialised centres working on data at universities in New Zealand and Chile. This aspect may be the reason that none of the typical language of technology, such as "innovation", "modernity" and "the future", is used to show their conclusions.

Despite this absence, the developers repeatedly state that AI is a desirable technology. Although they make reference to the possibilities for improvement and the need to always check for quality, at no time do they question it as technology nor as a valid tool to support public policies. More than responding to criticism over the system, they recommend preparing in advance for such criticism and having answers ready at hand.

In this sense, their reasoning in the conclusions and recommendations focuses on stressing the concept of "service acceptability" by society. Firstly, this suggests that developers understand that technologies are part

<sup>&</sup>lt;sup>77</sup>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>https://whatworks-csc.org.uk/wp-</u>content/uploads/WWCSC machine learning in childrens services does it work Sep 2020 Accessible pdf.

content/uploads/WWCSC machine learning in childrens services does it work Sep 2020 Accessible.pdf <sup>78</sup>Bright, J., Bharath, G., Seidelin, C., & Vogl, T. M. (2019). *Data Science for Local Government*. Oxford Internet Institute & University of Oxford. <u>https://ssrn.com/abstract=3370217</u> <sup>79</sup>Fubarles, V. (2018). On ait

<sup>&</sup>lt;sup>79</sup>Eubanks, V. (2018). Op. cit. <sup>80</sup> National IAL (2011)

Noticias UAI. (2018, 17 October). Modelos predictivos de riesgo para proteger a la infancia.
 Noticias UAI. <u>https://noticias.uai.cl/modelos-predictivos-de-riesgo-para-proteger-a-la-infancia/</u>
 <sup>81</sup>Peña, P. (2021, 16 November). Shielding Neoliberalism: "Social Acceptability" to Avoid Social Accountability of AI https://notmy.ai/news/case-study-a-childhood-alert-system-sistema-alerta-ninez-san-chile/

of socio-technical systems ("Even the best-designed services in the world might not be successful if families do not adopt them".)<sup>82</sup> Secondly, as indicated earlier, this may be due to the developers' experience with criticism over the implementation of other systems they have worked on.

The "service acceptability" is not a strictly technological concept, but a communicational one. Basically, it consists of improving communication to present the PRM to society. It is not about introducing an enhancement to the system or not implementing it at all. Rather, it is about taking advantage of the ability of humans to communicationally induce and seduce non-human systems. In other words, "service acceptability" is a human factor that, in a way, shields the machine from providing social accountability.

For example, many of the pilot developers' recommendations revolve around training humans to improve communication about the PRM. First, they suggest that presentation of the system to the affected community must be formulated in a "positive" way, based on the deliberate idea of hiding what the PRM precisely provides is a classification of social risk:

MDSF2 should also consider carefully how these families are contacted. Prepare guidelines for front-line Ministry workers that balance the need for transparency with the desire to have families contacted with a positive formulation, rather than a deficiency-based approach that emphasizes risk factors in the life of the family.<sup>83</sup>

At the same time, it is recommended that the ministry identify areas where the need for social benefits forces citizens to accept the implementation of the PRM without further questioning:

It is our recommendation that the MDSF focuses on the pilot with newborns, as this is a time of high need, but also a time when families are open to accepting additional support.<sup>84</sup>

The "service acceptability" as a shield against the social accountability of the non-human system becomes clear when hypothetical problems of "inequality" within the system are addressed, specifically with regard to "race" and socioeconomic status. On this last aspect, instead of seeing it as a circle of stigmatisation of poverty (as has been the criticism of Eubanks),<sup>85</sup> they consider its potential to discriminate against richer children and adolescents (although they do not interact with the system). In any case, human shields must

<sup>&</sup>lt;sup>82</sup>Vaithianathan, R., Benavides, D., Taskova, K., Hermosilla, MP., Letelier, A., Escobar, J., Garretón, M. & Kulick, E. (2019). Op. cit.

<sup>&</sup>lt;sup>83</sup>Ibid.

<sup>&</sup>lt;sup>84</sup>Ibid.

<sup>&</sup>lt;sup>85</sup>Eubanks, V. (2018). Op. cit.

bear the costs of software bias: "This indicates that front-line personnel should be more attentive to assessing risk factors among the wealthiest families that have been alerted."<sup>86</sup>

Moreover, it is strange how developers so easily dismiss the racial bias in SAN. Developers have stated: "In the Chilean context, frankly, we do not see similar concerns regarding race".<sup>87</sup> It is difficult to understand the reason for this statement. On the one hand, as we have already mentioned in this document, public criticism from organisations close to children's rights specifically points out racial discrimination as an issue. In addition, there is abundant literature on racism and xenophobia in Chile.<sup>88</sup> Likewise, they do not even examine the problem of bias from the point of view of the development of the system, and not only from the data collection. These omissions really cause concern, especially considering the evidence on how dangerous the racist idea of believing that algorithms do not see skin colour may be.<sup>89</sup>

Therefore, although there is the notion of technologies as a sociotechnical system, AI's conception of an objective process prevails, both in data collection and processing. This is concerning in several ways. First, because states and developers pay little attention to biases in social class and race, repeating the racial idea of colour blindness.<sup>90</sup> Second, 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 valorisation of automated information than our own experiences.<sup>91</sup>

Furthermore, as citizens have to submit their data to access state support, we can question if that consent is actually qualified. Considering that most are not even informed about how such data can be used for predictive social risk ranking, there is also a lot to discuss in terms of privacy, data protection and overall ethics of these systems. Another argument defended by the developers is the "social license" for the use of personal data, which would consist of society legitimating the use of such data. In their opinion, the criticism directed against the Chilean personal data protection law (enacted in a time when internet use was not massive) may predispose SAN to criticism. However, in addition to complying with the low standards of Chilean legislation, what would be important for the operation of AI systems would be the "social license":

Fortunately, in Chile, the problem of obtaining a social license is somewhat simpler as it is the citizens' custom to have their data processed to stratify their socioeconomic status

<sup>&</sup>lt;sup>86</sup>Vaithianathan, R., Benavides, D., Taskova, K., Hermosilla, MP., Letelier, A., Escobar, J., Garretón, M. & Kulick, E. (2019). Op. cit.

<sup>&</sup>lt;sup>87</sup>Ibid.

<sup>&</sup>lt;sup>88</sup>Tijoux, M.E. (2016). Racismo en Chile. La piel como marca de la inmigración. Universitaria. Santiago, Chile.

<sup>&</sup>lt;sup>89</sup>Benjamin, R. (2019). Race after technology: Abolitionist Tools for the New Jim Code. Polity.

<sup>&</sup>lt;sup>90</sup>Noble, S. U. (2018). *Algorithms of Oppression: How Search engines Reinforce Racism*. New York University Press.; Benjamin, R. (2019). Op. cit.

<sup>&</sup>lt;sup>91</sup>Bridle, J. (2018). New Dark Age: Technology and the End of the Future. Verso Books.

for the purpose of targeting social benefits. This practice has had a strong social license due to the benefits it brings to families in need.<sup>92</sup>

In the context of data extractivism required for AI, it is interesting that developers use the concept of "social license", which was popularised by the United Nations in 2004 to encourage companies (mainly in the extractive sectors) to involve indigenous peoples in their projects and to obtain their "consent" before implementing them. Consent has to be free, prior and informed. It is worth asking whether in SAN there is proper consent and, therefore, "social license", considering that the use of personal data from "beneficiaries" goes hand in hand with state support. There is not any explicit mention of risk rankings of vulnerability when the state looks for consent from families,<sup>93</sup> and when its developers seem to believe that the "custom" of providing personal data to the state is a good source of legitimacy.

As a result of this critical analysis, we urge feminists to examine SAN and likewise social-risk models in at least two ways: first, to question if big data and AI could ever reflect structural elements that influence the risk of vulnerability and social inequalities of our societies, or is just a way to objectivise the responsibility of individuals through their data trajectory at the state. And second, to question how SAN 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 their operation as a tool for social control, aggravating its consequences on social inequalities.<sup>94</sup> Referring back to our framework of oppressive AI, we can say – at least as far we have dug into it – SAN ticks most of the boxes.

<sup>&</sup>lt;sup>92</sup>Vaithianathan, R., Benavides, D., Taskova, K., Hermosilla, MP., Letelier, A., Escobar, J., Garretón, M. & Kulick, E. (2019). Op. cit.

<sup>&</sup>lt;sup>93</sup>Valderrama, M. (2021). Op. cit.

<sup>&</sup>lt;sup>94</sup>Monahan, T. (2009). Dreams of Control at a Distance: Gender, Surveillance, and Social Control. *Cultural Studies* ↔*Critical Methodologies*, *9*(2), 286–305. https://doi.org/10.1177/1532708608321481



Oppressive AI Framework by Joana Varon and Paz Peña. Design by Clarote for notmy ai for Coding Rights

### 2. Technological Platform for Social Intervention -Argentina / Projeto Horus – Brazil<sup>95</sup>

Let's say you have access to a database with information from 12,000 women between 10 and 19 years old, who are inhabitants of some poor province of South America. Datasets include age, neighbourhood, ethnicity, country of origin, education level of the head of household, physical and mental disabilities, number of people sharing a house and the presence or absence of hot water in their services. What conclusions would you extract from such a database? Or, maybe the question should be: is it even desirable to have any conclusion at all? Sometimes, and sadly more often than ever, just the possibility to extract sheer 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. It is used by public authorities, initially in the municipality of Salta, Argentina, piloted since 2015, under the name *Plataforma Tecnológica de Intervención Social* (Technological Platform for Social Intervention). Theoretically, the goal of the system was to prevent school dropouts and teenage pregnancy.

#### Who develops it?

The project started as a partnership with the Ministry of Early Childhood from the Province of Salta, Argentina, and Microsoft. The system is presented by a representative of both as a very accurate, almost magic, predictive tool: "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,<sup>96</sup> the machine learning system of the programme. "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",<sup>97</sup> declared Juan Manuel Urtubey, a conservative politician and governor of Salta by the time of the pilot deployment.

#### System Audits and other criticisms

But to predict and even to predestine someone for pregnancy is not that simple, not for mathematicians, neither for fortune-tellers. Criticism about the "*Plataforma Tecnológica de Intervención Social*" started to arise. <u>Some called the system a lie, an intelligence that does not think, a hallucination</u> and a risk for poor women's and children's sensitive data. A very complete technical analysis about its failures was published by the *Laboratorio de Inteligencia Artificial Aplicada*<sup>98</sup> from the University of Buenos Aires. According to

<sup>&</sup>lt;sup>95</sup>Sternik, I. (2018, 21 April). La inteligencia que no piensa. *Pagina 12*. <u>https://www.pagina12.com.ar/109080-la-inteligencia-que-no-piensa</u>

<sup>96</sup> Ibid.

<sup>&</sup>lt;sup>97</sup>Ibid.

<sup>&</sup>lt;sup>98</sup>Laboratorio de Inteligencia Artificial Aplicada – LIAA. (2018). *Sobre la predicción automática de embarazos adolescentes*. LIAA. <u>https://liaa.dc.uba.ar/es/sobre-la-prediccion-automatica-de-embarazos-adolescentes/</u>

LIAA, which analysed the methodology posted on GitHub by Microsoft engineer, the results were falsely oversized due to statistical errors in the methodology. The database is biased; it does not take into account the sensitivities of reporting unwanted pregnancy. Therefore, data collected is inadequate to make any future prediction, and it is likely to include pregnancies from one particular sector of society than others, stigmatising the poor. According to Luciana Ferrer, a researcher from LIAA:

If you are assuming that those who answered the surveys said the truth about being pregnant before or at the moment of the survey, our data is likely to be inaccurate. On such a delicate topic as teenage pregnancy, it would be cautious to assume that many teenage girls do not feel safe to tell the truth, about all if they have or want to have an abortion. This implies that using these data, we will be learning from biased information, influenced by the fact that in some privileged sectors of the population there was to access save abortion and in others the issue is a taboo, therefore, it is something that the adolescent would hide in an interview.<sup>99</sup>

In Argentina, just like in several countries in Latin America, access to <u>safe abortion was only legalized in</u> <u>cases of rape or when the mother's health was at risk. The situation in the country changed only in December</u> <u>2020, when a historical bill was</u> approved legalizing freedom of choice to interrupt pregnancy to the 14th week. It is interesting to note that Ministry of Early Childhood worked for years with the anti-abortion NGO, Fundación CONIN, to showcase this system.<sup>100</sup> Urtubey's <u>declaration mentioned was made in the middle of</u> <u>the campaign to change the law towards legalis</u>ing abortion in Argentina,<sup>101</sup> a social demand that in 2018 took over the local and international news for months. The idea that algorithms can predict teenage pregnancy before it happens was 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. Moreover, there was also a common, but mistaken, belief that "if it is recommended by an algorithm, it is mathematics, so it must be true and irrefutable."

Furthermore, 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 a particular sex also reinforces patriarchal gender roles and, ultimately, blames female teenagers for unwanted pregnancy, as if a child could be conceived without a sperm. It can also be seen as an initiative that originates from a logic of blaming the victim, particularly if we consider that the database includes girls aged 10 years old and minors a bit older, whose pregnancy would only be a result

<sup>&</sup>lt;sup>99</sup>Sternik, I. (2018, 21 April). Op. cit.

<sup>&</sup>lt;sup>100</sup>Ibid.; 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

<sup>&</sup>lt;sup>101</sup>Ibid.;Mora, B.P. (2019, 27 March). "Primera Infancia es el ministerio que defiende a los niños desde su concepción". *El Tribuno*.<u>https://www.eltribuno.com/salta/nota/2019-3-27-0-39-0--primera-infancia-es-el-ministerio-que-defiende-a-</u>los-ninos-desde-su-concepcion

of sexual violence. How can a machine say you are likely to be the victim of a sexual assault? How brutal it is to conceive such calculus?

Even in face of several criticisms, the initiative continued to be deployed. Bad ideas dressed as innovation spread fast: the system is now being <u>deployed in other Argentinian provinces</u>, such as *La Rioja*, Chaco,<sup>102</sup> and *Tierra del Fuego*.<sup>103</sup> It also has been exported to Colombia, implemented at least in the municipality of *La Guajira*,<sup>104</sup> and, as we will see, to Brazil.

#### From Argentina to Brazil

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 fifth country in Latin America to adopt Projeto Horus, presented in the media as a tech solution to monitor social programmes focused on child development. The first city to test the programme was Campina Grande, from the State of Paraíba, in the northeast region of Brazil, one of the poorest regions of the country.<sup>105</sup> Among the authorities and institutions in the kick-off meeting were representatives from Microsoft, the Ministry of Early Childhood from the municipality of Salta, and members from the Brazilian <u>Ministry of Citizenship.<sup>106</sup></u> Romero Rodrigues, the mayor of Campina Grande, is also aligned with evangelical churches.<sup>107</sup>

#### Analysis

Through access to information requests (Annex I), we have consulted the National Secretary of Early Childhood Care (SNAPI) and the Subsecretary of Information Technology (STI) from that ministry to acquire more information about the partnership. These institutions stated that:

The Ministry of Citizenship has signed with Microsoft Brasil LTDA the technical cooperation agreement n° 47/2019, for a proof of concept for an artificial intelligence tool to subsidize

PARAÍBA. <u>https://g1.globo.com/pb/paraiba/noticia/2019/09/25/campina-grande-testa-projeto-pioneiro-para-monitoramento-tecnologico-do-crianca-feliz.ghtml</u>
<sup>106</sup>Ibid.

<sup>&</sup>lt;sup>102</sup>Ibid.; Gobierno de Salta. (2016, 12 May). Tierra del Fuego aplica la plataforma tecnológica de Salta en su política social. *Gobierno de Salta*. <u>https://www.salta.gob.ar/prensa/noticias/tierra-del-fuego-aplica-la-plataforma-tecnologica-de-salta-en-su-politica-social-49653</u>

<sup>&</sup>lt;sup>103</sup> Ibid.; <u>http://www.boletinoficialsalta.gob.ar/NewDetalleDecreto.php?nro\_decreto=658/18</u>

 <sup>&</sup>lt;sup>104</sup> https://cidades.ibge.gov.br/brasil/pb/campina-grande/pesquisa/36/0?localidade1=33&localidade2=530010
 <sup>105</sup> <u>https://cidades.ibge.gov.br/brasil/pb/campina-grande/pesquisa/36/0?localidade1=33&localidade2=530010;</u> G1 PB. (2019, 25 September). Campina Grande testa projeto pioneiro para monitoramento tecnológico do 'Criança Feliz'.

<sup>&</sup>lt;sup>107</sup><u>https://paraibaonline.com.br/2020/02/prefeito-campinense-prestigia-abertura-do-22o-encontro-da-consciencia-crista/</u>

improvements in the actions of the programme Happy Child/Early Childhood (Crianç<u>a</u> <u>Feliz/Primeira Infância</u>).<sup>108</sup>

The Brazilian Minister of Citizenship Osmar Gasparini Terra, who signed the agreement, is, believe it or not, is a sympathiser of flat <u>Earthism</u>,<sup>109</sup> which is a theory that claims the Earth is not spherical. <u>Just as with</u> climate change denial and creationism, some allege <u>that the flat Earth theory has its base on Christian</u> <u>fundamentalism</u>. Terra also had a denialist discourse about COVID-19 pandemic, but in that case, he believes in math and AI as a sole tools to produce diagnostics to inform public policies, as expressed in the agreement:

The Ministry wishes to carry out an analysis for the Criança Feliz programme, using technological data processing tools based on artificial intelligence as a diagnostic mechanism aimed at detecting situations of social vulnerability as a guide for the formulation of preventive and transformative public policies.<sup>110</sup>

Just like in the Chilean<sup>111</sup> and Argentinian cases, a neoliberal vision was behind the rationale of belief in an algorithm to "automate, predict, identify, surveil, detect, target and punish the poor."<sup>112</sup> Something that was also explicit in the technical cooperation agreement, when it states that the purpose of the system was to "**optimize resources** and the construction of initiatives that can improve the offer of services aimed at early childhood, in a more customized way and with greater effectiveness."<sup>113</sup>

Therefore, the logic of **automating neoliberal policies** is stated once again. More specifically, this is about establishing a mechanism for a Digital Welfare State, heavily dependent on data collection and the conclusions that emerge from them:

The cooperation aims to build, together, a solution that collects data through electronic forms and the use of analytical and artificial intelligence tools on this data to subsidize actions of the Happy Child programme.<sup>114</sup>

<sup>109</sup>Cowie, S. (2019. 6 November). Brazil's Flat Earthers to get their day in the sun. *The Guardian*. <u>https://www.theguardian.com/world/2019/nov/06/brazil-flat-earth-conference-terra-plana</u>; Sena, M. (2020, 16 December). APÓS FICAR NA UTI POR COVID-19, OSMAR TERRA SEGUE MINIMIZANDO A DOENÇA. *Uoi*. <u>https://congressoemfoco.uol.com.br/temas/saude/apos-ficar-na-uti-por-covid-19-osmar-terra-segue-minimizando-a-doenca/</u>

<sup>&</sup>lt;sup>108</sup>Presidency of the Republic General Secretariat Deputy Head for Legal Affairs. (2018, 22 November). DECREE No. 9,579. <u>http://www.planalto.gov.br/ccivil\_03/\_ato2015-2018/2018/decreto/D9579.htm</u>

<sup>&</sup>lt;sup>110</sup>Answers to FOIA requests to the Brazilian government are available in Annex II, this particular answer is from the technical cooperation agreement: Acordo de Cooperação Técnica 47/2019 – Processo n 71000.036620/2019-43 <sup>111</sup>Peña, P. (2021, 16 November). Op. cit.

<sup>&</sup>lt;sup>112</sup>Special Rapporteur. (2019, 1 October). *Digital technology, social protection and human rights: Report.* A/74/493. Presented to General Assembly. <u>https://www.ohchr.org/en/calls-for-input/2019/digital-technology-social-protection-and-human-rights-report</u>

<sup>&</sup>lt;sup>113</sup>Acordo de Cooperação Técnica 47/2019 – Processo n 71000.036620/2019-43 <sup>114</sup>Ibid.

When inquired about databases used, they listed:

- Sistema Único de Assistência Social SUAS (Unique System of Social Assistance)
- Cadastro Único
- CADSUAS, from the Ministry of Social Development

They are all databases of social programmes in Brazil, in this case, used under the logic of **surveillance of vulnerable communities**, who are not only poor but also consist of children – all that in a partnership with a foreign company. Did Microsoft have access to all these databases? What was the counterpart for the company to enter the agreement as no transfer of financial resources were agreed:

This Agreement will not involve the transfer of financial resources between the parties. Each party will assume its own costs as a result of the resources allocated in the execution of the scope and its attributions, with no prior obligation to assume obligations based on its results.<sup>115</sup>

How many hours of human resources of public officials were deployed to a proof of concept in which results are not documented and publicly available? Did Microsoft have access to the database of the Brazilians? Is Microsoft using database from the poor in Latin America for training their machine learning systems? 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 questions, the previously scheduled interview was canceled. These were the questions sent to Microsoft:

1) Considering the technical cooperation agreement signed in September 2019 between the Ministry of Citizenship and Microsoft to carry out a proof of concept to implement artificial intelligence tools that support improvements in the actions of the Happy Child programme:

a) What is the result of the proof of concept?

b) What datasets were used by the algorithm to detect situations of social vulnerability?

c) What kind of actions would be suggested by the platform in case of vulnerability and risk detection?

d) What are the next steps in this proof of concept?

2) Does the company have other contacts with the Ministry of Citizenship or other ministries of the Brazilian government for proof of concepts or implementation of AI projects for social issues?

3) Does the company have an internal policy to promote research and development of activities of AI and the public sector focused on the country?

As we could not have a position from Microsoft, we also asked the ministry for more information about the promised "greater effectiveness", which nevertheless, was never proved. When the agreement was signed back in September 2019, there was already a lot of published criticism analysing the case of Salta, even though the agreement only recognised Microsoft "experience and intelligence gained", as per below:

Whereas Microsoft has already developed a similar project (...) with the PROVINCE OF SALTA, in the Argentine Republic, and all the experience and intelligence gained from it can be used, with this agreement, a cooperation is established for development, adaptation, and use of a platform in Brazil.<sup>116</sup>

So we requested information from the minister about error margins and data about the result of the proof of concept. They responded that "there is no information regarding the margin of error used in the technologies involved."<sup>117</sup>

Once again, the degree of participation of people affected as well as the accountability for using citizen's data was null. The proof of concept agreement had a six months work plan, which makes evident that there was no intent to conceive a broader and inclusive process consulting the targeted community, solely deploying a tech tool. When asked about the results of the pilot and the statistical data used for building the algorithms and database, the ministry has provided us with practically no information. They simply affirmed that the agreement lasted six months, from 23 September 2019 onwards. Therefore, the agreement was no longer in force by the time of the answer (18 December 2020), as such they restated:

We emphasize that such technology is not being implemented by the Happy Child Program, therefore, we are unable to meet the request for statistical data on its use and effectiveness.<sup>118</sup>

Even if the technology is not being implemented, if the government is using citizens' data to test a tool with a private company, there should be transparency and accountability obligations. Therefore, such an answer was completely unacceptable, particularly considering that in the work plan, attached to the agreement, both Microsoft and the ministry were assigned to activities towards analysis and evaluation of results. Under these circumstances and answers, we could say that the attempt to export the system from Argentina to Brazil was another expression of digital **colonialism and colonial extractivism**. Allegedly, not even the Brazilian government kept records of the results of the proof of concept, which can be seen as a **closed box**. The only

<sup>&</sup>lt;sup>116</sup>Ibid.

<sup>&</sup>lt;sup>117</sup>Answers to FOIA requests to the Brazilian government are available in Annex II.

<sup>&</sup>lt;sup>118</sup>Ibid.

thing we know is that, according to the agreement, Microsoft was exempt from any responsibility of possible harm caused by the project:

Microsoft does not guarantee or assume responsibility for losses and damages of any kind that may arise, by, for example: (i) the adequacy of the activities provided in this Agreement and the purposes of the Ministry or for the delivery of any effective solution; and (ii) the quality, legality, reliability and usefulness of services, information, data, files, products and any type of material used by the other parties or by third parties.<sup>119</sup>

In summary, we can say that The "*Plataforma Tecnológica de Intervención Social*" and Projeto Horus are just a very eloquent example of how AI's pretended neutrality has been increasingly deployed in some countries in Latin America to assist neoliberal and 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. Analysing our framework from oppressive AI, we could say it ticks all the boxes.



Oppressive AI Framework by Joana Varon and Paz Peña. Design by Clarote for notmy ai for Coding Rights

<sup>&</sup>lt;sup>119</sup>Acordo de Cooperação Técnica 47/2019 - Processo n 71000.036620/2019-43

### **Building a Feminist toolkit to question AI systems**

This article compiles what is currently the core insights of core notmy.ai platform, an ongoing effort, work in progress debate that seeks to contribute to the development of a feminist framework to question algorithmic decision-making systems that are being deployed by the public sector.



Notmy.ai platform

We have seen examples how these systems tend to be developed by privileged demographics, against the free will and without the opinion or participation from the onset of those who are likely to be targeted, or "helped", resulting in automated oppression and discrimination that use math as an excuse to skip any political responsibility. Ultimately, this trend has the power to dismiss any attempt of a collective, democratic and transparent response to core societal challenges.

To face this pervasive trend, we depart from the perspective that decolonial feminist approaches to life and technologies are great instruments to envision alternative futures and to overturn the prevailing logic in which AI systems are being deployed. As Silvia Rivera Cusicanqui poses: "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?"<sup>120</sup> If we follow Cusicanqui, it is easy to grasp that answers such as "optimisation of biased algorithms", "ethic", "inclusive", "transparent" or "human-centric" AI, "compliant with data protection

<sup>&</sup>lt;sup>120</sup>Rivera Cusicanqui, S. (2012). Ch'ixinakax utxiwa: A Reflection on the Practices and Discourses of Decolonization. *The South Atlantic Quarterly, 111*(1), 95-109. <u>http://www.adivasiresurgence.com/wp-content/uploads/2016/02/Silvia-Rivera-Cusicanqui-Chixinakax-Eng1.pdf</u>

legislation" or even solely a human rights approach to AI systems fall short in a bigger political mission to dismantle what black feminist scholar Patricia Hill Collins calls the "matrix of domination".<sup>121</sup> Simply adding a layer of automation to a failed system means magnified oppression disguised by a false sense of mathematical neutrality.

Current debates of AI principles and frameworks are mostly focused on "how to fix it?", instead of "Why we actually need it?" and "for whose benefit". Therefore, the first tool of our toolkit to question AI systems is the scheme of oppressive AI that we drafted based on both empirical analysis of cases from Latin America and bibliographic review of critical literature. Is there a particular AI system based on surveilling the poor? Is it automating neoliberal policies? Is it based on precarious labour and colonial extractivism of data bodies and resources from our territories? Are its developers part of the group it targets? Is it likely to restate structural inequalities of race, gender and sexuality? Can the wider community have enough transparency to check the accuracy in the answers to the previous questions by themselves? These might be some of the questions to be considered.

Several national policies for AI and most start-ups and big tech corporations operate under the motto of "move fast and break things" – meaning, innovate first and check for possible harm later. We propose the opposite: before developing or deploying, AI should be checked if it is likely to automate oppression. Furthermore, if that AI system is not focused on exposing the powerful nor developed by and with participation of those who will be using it, it is pretty likely that such system would fall into the categories of an oppressive AI.

In addition, those categories are not meant to be fixed; they can expand according to a particular context. So, watch out. The proposed oppressive AI framework is not written in stone. It is just a general guide for questions – a work in progress that shall be reshaped according to the particular context and its oppressions. In this sense, we recall Design Justice Network Principles<sup>122</sup> as an important guideline to assess the context of oppression, since it "centers people who are normally marginalized by design and uses collaborative, creative practices to address the challenges faces by a particular community."<sup>123</sup>

Going beyond, why not ask ourselves:

36

<sup>&</sup>lt;sup>121</sup>Collins, P.H. (2002). *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment.* Routledge.

<sup>&</sup>lt;sup>122</sup><u>https://designjustice.org/read-the-principles;</u> Flecha 1 – A serpente e a Canoa: <u>https://www.youtube.com/watch?v=Cfroy5JTcy4</u>

<sup>&</sup>lt;sup>123</sup>Costanza-Chock, S. (2018). Design Justice, AI, and Escape from the Matrix of Domination. *Journal of Design and Science*. <u>https://doi.org/10.21428/96c8d426</u>

# What is a transfeminist AI? What does it mean to develop a feminist algorithm?

We believe that transfeminist values can be embedded in AI systems, just as values such as profit, addiction, consumerism and racism are currently embedded in several algorithms that pertain to our lives today. To push this feminist approach into practice, we at Coding Rights, in partnership with our dear and brilliant scholar and design activist Sasha Costanza-Chock, have been experimenting with a card game collaboratively developed to design tools in speculative futures: the "Oracle for transfeminist Technologies".<sup>124</sup> Through a series of workshops, we have been collectively brainstorming what kind of transfeminist values shall inspire and help us envision speculative transfeminist futures.

Indeed, tangible present proposals of changes emerged once we were imagining the future in the workshops. Over time, values such as agency, accountability, autonomy, social justice, non-binarism, cooperation, decentralisation, consent, diversity, decoloniality, empathy, security, among others have emerged in the workshop brainstormings and were progressively transformed into value cards of the Oracle.

<sup>&</sup>lt;sup>124</sup><u>https://www.transfeministech.codingrights.org;</u> Varon, J. (2020, 3 July). The Future is TransFeminist: from imagination to action. *Medium*. <u>https://deepdives.in/the-future-is-transfeminist-from-imagination-to-action-6365e097eb22</u>



Value cards deck from the Oracle for Transfeminist Technologies: www.transfeministech.org

### **ORACLE FOR TRANSFEMINIST TECHNOLOGIES**



### transfeministech.org

Five card decks that compose the Oracle for Transfeministech and an example of a consultation. More information available at: www.transfeministech.org

While it has been envisioned as a card game for speculative futures, we believe that the ensemble of transfeminist values, brainstormed over a series of workshops with feminists from different regions and identitary feminist agendas, can also inspire different tech towards envisioning transfeminist AI projects, alternative tech or practices that are more coherent with the present and future we want to see. 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."<sup>125</sup>

<sup>&</sup>lt;sup>125</sup>Le Guin, U. (2019). Ursula K. Le Guin: The Last Interview: And Other Conversations. D. Streitfeld. (Ed.). Melville House.

Can we also take these values from speculation to action?<sup>126</sup> We have humbly started to do that with the article<u>"</u>Consent to our Data Bodies: Lessons from feminist theories to enforce data protection".<sup>127</sup> Addressing the questions: What is a feminist approach to consent? How can it be applied to technologies? Those simple questions were able to shed light on how limited the individualistic notion of consent proposed in data protection frameworks is. That universalised approach doesn't take into account unequal power relations. If we do not have the ability to say no to big tech companies when accessing a monopolistic service, we clearly cannot freely consent.

Maybe it is a long way to go, but perhaps, laying down an extensive analysis of these values, as we did with the notion of consent, can gradually shed more light on future tech we want to see. Technologies that consider power imbalances that are present in the context they are developed and deployed and that do not erase existences in order to take into account that, in the words of indigenous leader Ailton Krenak: the future is ancestral. What would a transfeminist AI look like for you?

----

#### Stay tuned.

We are currently brainstorming other possible steps to both amplify the reach and the tools of notmy.ai. Some activities involve continuing to test its applicability through workshops, developing local partnerships for increasing case-based analysis, continuously studying new bibliography with critical approaches to AI systems and launching translations to Spanish and Portuguese to allow for wider regional engagement in these debates. Also, if you know other AI projects being deployed in Latin America by the public sector with possible implications to advance feminist agendas, our mapping is collaborative, and you can submit them at notmy.ai.<sup>128</sup> If you have feedback about these frameworks, you can reach us in social media @CodingRights or at contact@codingrights.org

<sup>&</sup>lt;sup>126</sup>We are thankful to all the FIRN network for all the feedback received through meetings, emails and workshops, and we also wish to express enormous gratitude to Tigist Hussen, Namita Aavriti Malhotra, Jac sm Kee and Catalina Alzate, who read and commented extensively and in detail on different versions of this research. Cheers to all the feminists around the globe that have been engaging in online and offline sessions of the Oracle for Transfeminist Technologies, so far in English, Portuguese and Spanish. Let's keep the imagination flowing! And a special thanks to all Latin American colleagues who considered addressing the questionnaire and/or sent links to help us map AI systems being deployed in the region. Thanks to the Tech and Human Rights fellows and staff at the Carr Center for Human Rights Policy from Harvard Kennedy School, who also read and commented on a presentation about one version of this research. Love to all the Coding Rights team that keeps hacking patriarchy everyday and to Sasha Costanza-Chock for jamming in the invention of the Oracle for Transfeminist Tech. Lots of respect to all the multitude of other inspiring antiracist and decolonial feminists who seeded our path with fierce thoughts.

<sup>&</sup>lt;sup>127</sup>Peña, P., & Varon, J. (2019). Consent to our Data Bodies: lessons from feminist theories to enforce data protection. Coding Rights. <u>https://codingrights.org/docs/ConsentToOurDataBodies.pdf</u>

<sup>&</sup>lt;sup>128</sup> Soon available in Spanish and Portuguese: <u>https://notmy.ai/do-you-know-other-projects/</u>



Summary of our work in progress presented during Firn workshop illustrated by Sonaksha

Iyengar

#### About the authors

Joana Varon is the Founder Directress and Creative Chaos Catalyst at Coding Rights. Believing in art, creativity and coding as tools for revolutions, she is a co-creator of several creative projects operating in the interplay between activism, arts and technologies, such as notmy.ai, transfeministech.org, museamami.org, chupadados.com, #SaferSisters, Safer Nudes, From Devices to Bodies, protestos.org, Net of Rights and freenetfilm.org. Former Mozilla Media Fellow, Joana is currently a Technology and Human Rights Fellow at the Carr Center for Human Rights Policy from Harvard Kennedy School and affiliated to the Berkman Klein Center for Internet and Society at Harvard University.

**Paz Peña** is an independent consultant on Human Rights, intersectionality, and digital technologies. She is also the secretary of Al Sur, an aggrupation of 11 organizations from seven countries in Latin America working on digital rights, and the co-creator of Acoso.Online, a website with information and recommendations for victims of "non-consensual pornography" on the internet in Latin America and the Caribbean. Since 2014 she is also part of the collaborative council at Coding Rights Brazil. For eight years, she was the director of advocacy at Derechos Digitales América Latina and the public leader of Creative Commons Chile (2013-2015).

# Annex I – Access to information requests and answers (Chilean case)

- <u>Acuerdo Confidencialidad con el Ministério de Desarollo Licitación Alerta Niñez\_2018\_C</u>
- Acuse de Recibo de Pedido de Información AI001T0002359
- Análisis Nuevo proceso de asignación responsables derivaciones V1\_C
- <u>Cambio de Perfil y Roles\_C</u>
- <u>Carta\_Nro\_3805\_Fol\_02359</u>
- Documento de apoyo para la Gestión de Casos 2020
- Informe Final Alerta Niñez
- Manual Alerta Niñez final
- <u>Of 3804 Fol 02359</u>
- Orientaciones Técnicas para la implementación del Piloto de la Oficina Local de la Niñez
   2020\_como funciona

# Annex II – Access to information requests and answers (Brazilian case)

- Pedido&resposta\_LAI\_MinCidadania\_71003\_129432\_2020\_71
- <u>Acordo\_Cooperacao\_Tecnica\_47\_Microsoft</u>
- Pedido&resposta\_LAI\_MinCidadania\_71003\_129428\_2020\_11
- <u>Pedido&resposta\_LAI\_MinEcon\_00106\_030439\_2020\_31</u>
- Pedido&resposta LAI\_MMFDH\_00105\_003197\_2020\_12