WORLD WIDE WEB FOUNDATION

ALGORITHMS AND ARTIFICIAL INTELLIGENCE IN LATIN AMERICA

A Study of Implementation by Governments in Argentina and Uruguay

September 2018

www.webfoundation.org

CONTENTS

WORLD WIDE WEB FOUNDATION

This report was written by Juan Ortiz Freuler and Carlos Iglesias, Web Foundation. The case studies section benefitted from a background paper by the Latin American Open Data Initiative (ILDA).

With thanks to Yasodara Córdova, Miranda Bogen, Matías Aranguiz, Vidushi Marda, Joe Westby, Marcelo Soria, Lauran Potter and Craig Fagan for their comments and suggestions.

Suggested citation: Ortiz Freuler, J. and Iglesias, C. (2018). Algorithms and Artificial Intelligence in Latin America: A Study of Implementation by Governments in Argentina and Uruguay, *World Wide Web Foundation*.

Supported by a grant from the Omidyar Network.

The Web Foundation was established in 2009 by Sir Tim Berners-Lee, inventor of the World Wide Web. Our mission is to establish the open web as a public good and a basic right.

FOREWORD

In March 2017, our founder and inventor of the World Wide Web, Sir Tim Berners-Lee, expressed his concern about what he believed to be a growing loss of control over personal data and the role that algorithms and artificial intelligence play in our society.

In response to these concerns, the Web Foundation published a series of white papers. They provide theoretical frameworks and examples to think about how the use of personal data, algorithms and artificial intelligence creates risks and opportunities for low and middle income countries. In 2018, we went into more detail on AI and algorithms, publishing a report which documents the AI initiatives across Africa. We also published a report that explores a methodology for the analysis of the algorithms Facebook uses to define what content is served to different users, and tested it in the context of Argentina.

This report complements this series, focusing on the use of algorithms and artificial intelligence by the public sector. We are concerned about the power dynamics between those who build these systems and those who are affected by them. In particular, about how automating processes through technology may create a high degree of opacity around *who* accesses public services and resources, and why they do so.

To the extent that techniques that leverage artificial intelligence and algorithms rely on large volumes of data, we believe that the open data agenda and collaborative spirit — which we promote together with allies such as ILDA — is essential to ensure fair implementations.

We believe that whereas an adversarial approach to advocacy might be effective in other contexts, in low and middle-income countries the approach should be one of collaboration. It is time to engage government officials in conversations about good practices, and provide technical support and coordination that helps public officials to progress this agenda.

In line with this collaborative spirit, we share this report hoping you will enjoy reading it and will send us your comments and suggestions. Let us work together to ensure that new technologies lead to a fairer world.

EXECUTIVE SUMMARY

Although there is growing interest, the use of artificial intelligence systems by governments in Latin America is still very limited. This presents an opportunity to develop a cross-sector network that offers officials interested in incorporating these tools a place to discuss how different approaches maximise benefits and minimise risks to the population.

This report focuses on the two aspects officials interested in implementing artificial intelligence must address: effectiveness and legitimacy. Based on four case studies from Argentina and Uruguay we observe that although a specific official may have the power to implement an AI system, he or she does not control all the components that will determine its impact. Thus, to understand the effectiveness and legitimacy of a system's implementation, we suggest that public officials consider four key areas:

1. The process of dataset creation;
2. The design of tools;
3. The administrative protocols that surround the tool's output;
4. The legal and social norms that define the broader context in which the policy is executed.

The case studies

The Government of the Province of Salta, in Argentina, implemented a system to predict teenage pregnancy and school dropout, with the support of Microsoft. The case illustrates how a government with limited resources seeks to use technology to solve urgent social problems. The government implemented a mechanism to coordinate data collection from 200,000 people living in vulnerable areas through NGOs and government officials, and a machine learning model to generate predictions about school dropouts and teenage pregnancy among members of this population. The implementation triggered great interest on the part of other governments and criticism from activists who believe that it violates privacy and does not address the causes of the problem. The implementation had transparent stages and opaque ones.

The Government of the City of Buenos Aires, for its part, developed an algorithm to build dashboards outlining commercial opportunities in the city's neighbourhoods and thus direct investments to where it is presumed that they will generate more value. The algorithm is simple and implemented in a way that facilitates transparency.

In Uruguay we explain how the government acquired Predpol policing software to predict crimes. It is a problematic case due to its degree of opacity and the dynamics of discrimination and exclusion it may reinforce. We explain that in less than three years the Ministry of the Interior discontinued the program and replaced it with retrospective statistical tools, developed by the Ministry's own teams, which it found more useful.

In its recommendations, the report highlights the need to:

- **Develop** local infrastructure, expertise, and regulation.
- **Define** and adopt practices of transparency, public participation and accountability in the development and implementation stages.
- **Establish** criteria for evaluating the risks of different models and their implementation.

01

INTRODUCTION

In recent decades, the emergence of new information technologies has led governments around the world to produce and keep digital records.

Throughout the last decade, as part of a movement, the Web Foundation actively promoted the process of opening government data, because we understand that, in order for a robust democratic debate to take place, people must understand how the public sector works. The opening of government data allows citizens to analyse the data in order to understand, support and push for public policies on the basis of evidence.

In more recent years, advances in methodologies and technologies have allowed new and more effective uses for the mountains of data produced by governments. With the rise of the so-called *big data* progress was made in combining large databases to identify previously invisible or overlooked patterns. Some eye-catching findings boosted the discipline. They also casted doubts. Are these patterns mere coincidences or do they reflect causal links that can be fueled or broken by public policies in order to improve people's lives?

With a greater understanding of the power of data, governments around the world have begun to rely on increasingly automated procedures for processing and analysing data. This process is triggered by a belief that this technology can assist governments to become more effective in delivering services.

The effectiveness of a policy or an algorithm, however, is not the only relevant dimension. Legitimacy offers a parallel challenge, which requires a range of tools and processes, which go beyond the field of data science.

The growing adoption of new technologies has altered (and will continue to alter) the way in which governments engage with the people. Ensuring transparency, public participation, and accountability is a matter of urgency. Only this will guarantee that the implementation of new technologies by government institutions will be done in accordance with democratic values and the interests of the people.[1]

A minimalist approach to the concept of legitimacy would focus solely on whether the action or policy in question is in accordance with the law and human rights. However, in a field as novel and unregulated as artificial intelligence, this definition is not very helpful as practical guide to action for public officials. The key concepts and approaches to the problem vary between those who focus on procedure and those who focus on outcomes.

Among the group that focuses on procedural aspects are those who demand that the affected populations participate in the design.[2] For others the priority is ensuring algorithms are interpretable, explainable,[3] and accountable (ensuring it is possible to trace the chain of responsibilities once a flaw is detected).[4]

Among those who focus on results are those who demand non-discrimination, and those who are concerned with more abstract and comprehensive concepts such as *justice* and *fairness*.

The FATML[5], ASILOMAR[6], NESTA[7], AI NOW[8], Partnership on AI[9], Access Now and Amnesty International[10], Berkman Klein Center[11] and the Government of Canada[12] are among the groups and institutions that are trying to establish how legitimacy should be defined in this space.

In order to assess whether the implementation of an AI or algorithmic system is legitimate, we suggest both the process and the results are examined.

1 There is a lot of interest in what new technologies represent for the future of work. Interestingly, there seems to be less debate regarding how ICTs will impact on the relationship between state and citizen. Encouraging STEM studies, which is often a key recommendation when discussing automation and the future of work, should also be considered in relation to the future of civic debate. Citizens increasingly need basic notions of STEM to participate effectively in the public arena.

2 See for example, Costanza-Chock, S. (Ed.). (2017). Design Justice In Action. Design Justice, (3), 3-12. Available at https://static1.squarespace.com/static/56a3ad187086d771d66d920d/t/597b2d20cf81e057708841eb/1501244715717/DJ_2017_Issue3_SM.pdf

3 It should be possible to explain to the person affected *why* the algorithm settled her particular case in the way it did. More in subsequent paragraphs.

4 So as to ensure that algorithms are not used as scapegoats, and provide incentives for responsible humans to act i) diligently in order to avoid problems, and ii) swiftly to solve them when they occur. More information at Web Foundation. (Ed.). (2017). Algorithmic Accountability: Applying the concept to different country contexts. Retrieved August 5, 2018, from http://webfoundation.org/docs/2017/07/WF_Algorithms.pdf

5 Suggest four principles to ensure accountable algorithms FATML. (2017). Principles for Accountable Algorithms [Web log post]. Retrieved August 28, 2018, from https://www.fatml.org/resources/principles-for-accountable-algorithms

6 ASILOMAR. (2017). Asilomar AI Principles [Web log post]. Retrieved August 28, 2018, from https://futureoflife.org/ai-principles/?cn-reloaded=1

7 Suggests 10 principles the public sector should take into account when implementing algorithms Copeland, E. (2018, February 20). 10 principles for public sector use of algorithmic decision making [Web log post]. Retrieved August 28, 2018, from https://www.nesta.org.uk/blog/10-principles-for-public-sector-use-of-algorithmic-decision-making/

8 Suggests four objectives that should guide the implementation of tools based on artificial intelligence AI Now. (2018, February 21). Algorithmic Impact Assessments: Toward Accountable Automation in Public Agencies [Web log post]. Retrieved August 28, 2018, from https://medium.com/@AINowInstitute/algorithmic-impact-assessments-toward-accountable-automation-in-public-agencies-bd9856e6fdde

9 An initiative, launched by leading US technology companies and focusing on seven key pillars About - The Partnership on AI. (n.d.). Retrieved August 28, 2018, from https://www.partnershiponai.org/about/

10 Toronto Declaration: Protection of rights, equality and non-discrimination by machine learning tools. See paragrapgs 30-37 Access Now., & Amnesty International. (Eds.). (2018). The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems. *Access Now*. Retrieved August 28, 2018, from https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf

11 Discusses the contexts in which algorithms should be explainable. Doshi-Velez, Finale and Kortz, Mason and Budish, Ryan and Bavitz, Christopher and Gershman, Samuel J. and O'Brien, David and Shieber, Stuart and Waldo, Jim and Weinberger, David and Wood, Alexandra, Accountability of AI Under the Law: The Role of Explanation (November 3, 2017). Berkman Center Research Publication Forthcoming; Harvard Public Law Working Paper No. 18-07. Available at SSRN: https://ssrn.com/abstract=3064761

12 Is developing a questionnaire to assist government units in measuring risks arising from employing algorithmic tools, based on two aspects: the actor and sector upon which the tool impacts; and the level of autonomy granted to the tool. Karlin, M. (2018, August 7). The Government of Canada's Algorithmic Impact Assessment: Take Two [Web log post]. Retrieved August 28, 2018, from https://medium.com/@supergovernance/the-government-of-canadas-algorithmic-impact-assessment-take-two-8a22a87acf6f

These conceptual distinctions are moving centre stage. As state modernization processes advance across the continent, officials from different levels of government often face the question of whether or not to implement algorithms or artificial intelligence systems to solve policy challenges. Accordingly, this question could be broken down as follows:

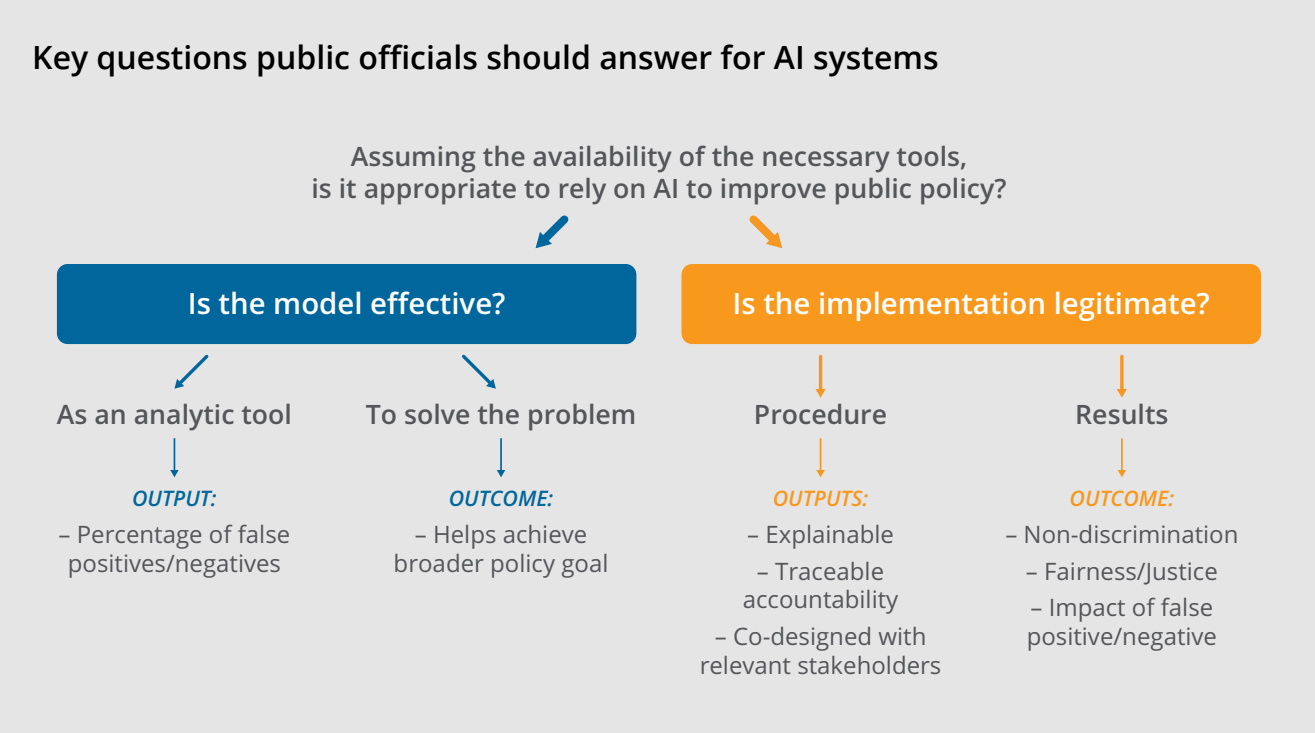


Fig. 1 — Diagram of the ecosystem in which the algorithms operate

Providing a definitive answer to these questions is not the purpose of this document. Instead, it seeks to provide tools that can help public servants answer them.

One of the issues that we want to highlight is that, although an individual official may have the power to decide whether or not to implement these systems to tackle a specific problem, the official often controls only one of the several components that determine the results. That is, the key decision-maker has limited information. This makes the task of answering questions about the expected effectiveness and legitimacy of an implementation particularly difficult. The framework we are developing seeks to highlight this situation, and the subsequent need for fluid communication between the various actors whose actions play a role in defining the impact.

The proposal focuses on four components:

1. The process of data collection/creation
2. The design of the tools
3. The administrative protocols that surround the tool's output
4. The legal and social norms that define the broader context in which the policy is executed.

As a way to showcase this proposed framework, we apply it to four case studies from Argentina and Uruguay.

DEFINITIONS AND METHODOLOGICAL APPROACH

In this report we will rely on a broad definition of *algorithm*. We employ the term to refer to a logical series of steps for organising and acting on a set of data with the aim of achieving an outcome quickly.[13] When we use the term *artificial intelligence* (AI) we refer to techniques which use algorithms as part of a broader and more dynamic analytical process. AI is an umbrella term that groups together a broad range of techniques that allow us to identify which course of action maximises the probabilities of achieving a specific objective within a changing environment. This includes popular techniques such as *machine learning*.[14] To simplify matters for the readers, and except when referring to a specific implementation, we will use *algorithm* as a generic term that encompasses both basic algorithms and artificial intelligence systems.

The effectiveness and legitimacy of an algorithm's implementation depends on many factors usually beyond the control of a single actor. Its implementation therefore requires close coordination among many actors. Yet these actors often work in silos.[15] The framework shown below attempts to highlight the ecosystem of people, data, code and norms, which surround the implementation of these tools and determine their impact.

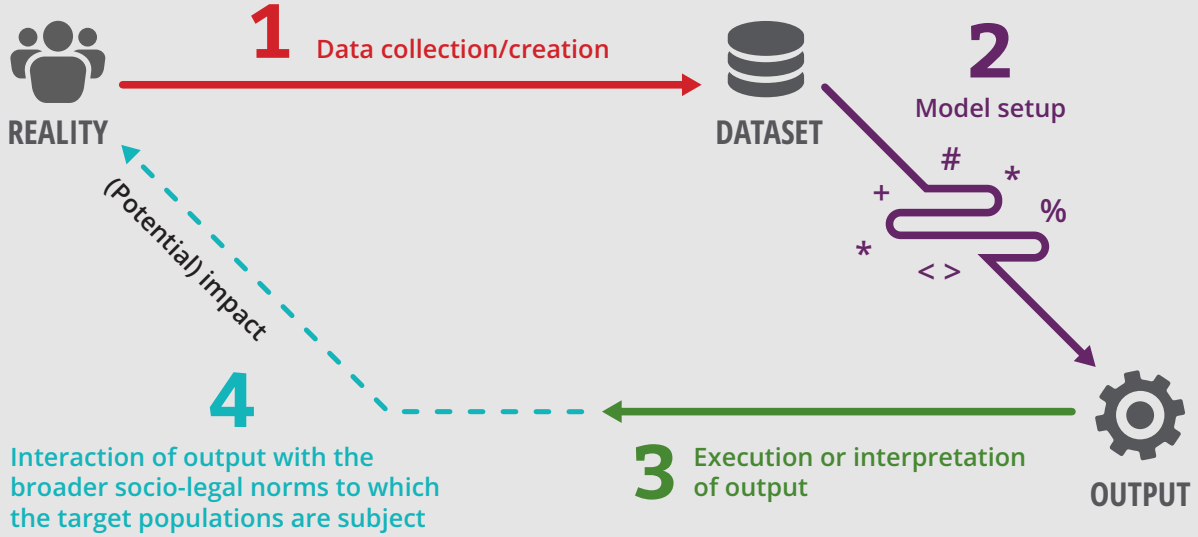


Fig. 2 — Diagram of the ecosystem in which the algorithms operate

13 For a general introduction to algorithms, see Web Foundation (Op. Cit) http://webfoundation.org/docs/2017/07/WF_Algorithms.pdf

14 For a general introduction to the subject of artificial intelligence, see Web Foundation. (Ed.). (2017). Artificial Intelligence: The Road Ahead in Low and Middle income countries. Retrieved August 5, 2018, from http://webfoundation.org/docs/2017/07/AI_Report_WF.pdf

15 For example, if algorithm design teams are alerted of a bias in the database or training set, the rules governing the algorithms can be adjusted to correct the bias before it causes harm. If this were not a viable solution and/or the algorithm has a high error rate, the designers must alert political authorities about these weaknesses, who can design administrative protocols so as to exclude the utilization of the algorithm in cases where the bias can manifest itself . Lawmakers, who define the legal frameworks that define how institutions work and social relations evolve, should be apprised of the implementation of algorithms with disruptive capacities, such as increasing demand for public services, or redistributing wealth among citizens. Implementation of some algorithms may require or benefit from legal reforms in order to minimize the risks and increase the benefits.

1. Data collection/creation

This process involves, among other steps, defining the scope of data to be collected and the appropriate method to collect it; the teams who will execute the method, and the labels that will be assigned to each variable.

Each of these decisions is subject to the same power relations that govern society,[16] which will inevitably impact the meaning of the data and how it can be used. The process of building databases must therefore be observed critically, particularly as the resulting databases play an increasingly central role in people's lives.[17]

The link with the open data agenda is straightforward: Algorithms are supported by large databases which, as a rule, should be accessible and verifiable to ensure the development of reliable solutions.[18] Openness increases the chances of identifying data that are incorrect, incomplete or unrepresentative, which can help those in charge of designing the model to neutralise subsequent problems such as bias before they impact a population. Inasmuch as algorithms are used to comply with rights, establish obligations or monitor people's behaviour, a high degree of participation is required in the process of defining what data are to be collected, as well as transparency around the methodologies used for their collection.

There are also cases where systems that are making decisions about identifiable individuals might need to balance these needs with privacy rights of those individuals. Opening up aggregates or relying on anonymisation are often times the way to balance these concerns. Releasing no data whatsoever should be considered an extreme solution, that should be accompanied by an explanation of the process and the arguments that led towards this decision.

2. Algorithm design / AI model setup

Algorithms and AI systems behave according to the rules under which they were programmed.

There are three key actors in this stage: 1) the political authorities, 2) the programmers and 3) the affected population. The political authorities decide whether an algorithm will be used to aid public management. If they decide it is useful, legal and in the interest of the population, they then establish the objectives and outline the values that must be optimised.[19] The programmers' task is to put those values into practice by deciding what variables to include, what weights to assign to them and how they should interact. Involving the target communities in the definition of the values and assumptions that will guide the design of the algorithm, as well as in the process of translating these values into code, is key to minimising the risks of harm and ensuring the legitimacy of their implementation.[20]

One of the central issues is the explainability of the algorithm or AI model. Can those affected by it understand or be informed about why a particular decision was made about them? Explainability is particularly relevant when it concerns systems that determine or affect the provision of a public service or which substantially affects an individual's interests, rights or freedoms.[21] The first step is to ensure the population is aware that automated systems are in place. The following steps imply reducing the degrees of opacity. Extreme opacity is usually referred to as a "black box". This refers to systems in which even if we are able to monitor the *inputs* and *outputs* we are never able to learn what happens in between.[22]

16 Foucault, M. (2002). *The order of things: An Archeology of the Human Sciences*. Psychology Press.

17 At the same time, steps must be taken to simplify this debate on their construction. Here you will find some suggestions to standardize metadata, and provide more critical information on how the databases were built Gebru, T. (2018). Datasheets for Datasets. Retrieved August 28, 2018, from http://jamiemorgenstern.com/papers/datasheet.pdf [Working Draft] More information can be found in Nesta (Op. Cit) https://www.nesta.org.uk/blog/10-principles-for-public-sector-use-of-algorithmic-decision-making

18 The ACM, for example, suggests that "a description of the way in which the training data were collected should be kept by the algorithm builders. This should be accompanied by an examination of possible biases induced by the human or algorithmic data collection process." Association for Computing Machinery (ACM), US Public Policy Council. (2017, January 12). *Statement on Algorithmic Transparency and Accountability* [Press release]. Retrieved August 28, 2018, from https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

19 To illustrate the relevance and potential risks of this stage, see Crockford, K., & Ito, J. (2017, December 22). Don't blame the algorithm for doing what Boston school officials asked. *Boston Globe*. Retrieved August 28, 2018, from https://www.bostonglobe.com/opinion/2017/12/22/don-blame-algorithm-for-doing-what-boston-school-officials-asked/lAsWv1Rfwqmq6Jfm5ypLmJ/story.html

20 See for example objective 4 raised by AI NOW (Artificial Intelligence Now): "Strengthening due process by giving the public the opportunity to participate in the process before, during and after the impact assessment." (Op. Cit) https://medium.com/@AINowInstitute/algorithmic-impact-assessments-toward-accountable-automation-in-public-agencies-bd9856e6fdde and paragraph 21 of the Toronto Declaration (Op. Cit) https://www.amnesty.org/download/Documents/POL3084472018ENGLISH.PDF

21 "If a system is so complex that it is impossible to make it intelligible to the people it affects, it is not doing its job." Lecher, C., What Happens When an Algorithm Cuts Your Health Care, March 2018. Available at https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy [Date of Enquiry: 06/05/2018]

22 The double meaning of the term *black box* makes it a useful metaphor: in addition to being used to refer to opaque systems, it is also used to refer to the systems that record information, such as the black boxes of aeroplanes. As Frank Pasquale explains, in the information age, we face this duality of meaning on a daily basis. On the one hand, we are constantly monitored by public and private information systems, and on the other, we do not know what is done with the information collected or what consequences they may lead to. See Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information.* Harvard University Press. Pg 3

Systems which operate as a black box usually preclude any form of decentralised control over the quality of *outputs* and can undermine affected populations' trust in government institutions.[23]

From the perspective of the affected population, a system may behave as a black box either as a result of a conscious decision or due to the intrinsic complexity of the system. Among the cases of deliberate opacity are those in which key information is withheld with the justification that such secrecy is required to protect intellectual property or prevent malicious attacks or gaming.[24] Among the cases of intrinsic complexity are some advanced models where neither the model's designers nor people with unrestricted access to it are capable of explaining the relationship between inputs and outputs. It is frequently argued that the complexity inherent to certain artificial intelligence techniques means that they are black boxes by nature. Consequently, non-intelligibility is an inherent characteristic or condition of using some systems. This line of reasoning suggests an inescapable trade-off between effectiveness and explainability. However, significant progress has been made on intelligibility over the past year. This progress suggests that, in the near future, even complex machine learning models could be implemented in such a way that their outputs are intelligible and can be explained to the populations affected[25]. For the time being, it is important for policy-makers to consider these tradeoffs, and rely on explainable models for systems that can substantially affect people's rights, and leave more opaque models for projects that have less tangible effects of people.

3. Defining the administrative protocols that surround the tool's output (decision/ interpretation of output)

In the process of designing and implementing an algorithm or AI model, political authorities must establish whether the technology will operate autonomously, allowing the system to execute decisions (e.g. reject a pension application) or be informative, in which case the final decision will be made by a public servant.

Since many decisions public officials are required to take are ethically complex, political authorities may be tempted to use these systems to disengage from this responsibility or to provide cover for decisions which can not be justified on political grounds: "*It was the algorithm!*". These technologies are still in their infancy, and often provide *outputs* which are inaccurate or incomplete. Therefore, it is important that decisions which substantially affect people's lives are under human supervision.[26] Most AI still struggles with understanding context, and this can often lead to very unexpected and harmful mistakes, if allowed to execute decisions automatically.

It should be noted that the distinction between algorithmic *decision* and *recommendation* becomes blurred when laws or administrative protocols create incentives against any departure from the criteria suggested by the algorithm. For example, when those who deviate from what the algorithm suggests and make a mistake are penalised. It is fundamental that those who establish the administrative procedures to be followed from the output understand the full meaning of the *output,* including its weaknesses. This will require close communication between the

23 The ACM, for example, states that "Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts." (Op. Cit) https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

24 It should be noted key information can be made available without putting intellectual property at risk. Public procurement processes should include an assessment of the transparency and accountability afforded by different tools. It must be noted that open systems that enable decentralized control by members of the public, increase security and reduce the probability that undetected bugs can be exploited for malicious ends. Encouraging open source should be part of the the government's strategy in these areas. See Simonite, T. (2017, October 18). AI Experts want to end black box algorithms in government. *Wired*. Retrieved from https://www.wired.com/story/ai-experts-want-to-end-black-box-algorithms-in-government/

25 See Snow, J. (2017, November 7). New Research Aims to Solve the Problem of AI Bias in "Black Box" Algorithms. *MIT Technology Review*. doi:https://www.technologyreview.com/s/609338/new-research-aims-to-solve-the-problem-of-ai-bias-in-black-box-algorithms/ And, Rieke, A., Bogen, M., & Robinson, D. (2018). Public Scrutiny of Automated Decisions: Early Lessons and Emerging Methods. Retrieved from http://omidyar.com/sites/default/files/file_archive/Public Scrutiny of Automated Decisions.pdf

26 The European Commission, for example, believes: "The data protection law establishes that you have the right not to be subject to a decision based solely on automated means, if the decision produces legal effects concerning you or significantly affects you in a similar way. A decision produces legal effects when your legal rights are impacted (such as your right to vote). In addition, processing can significantly affect you if it influences your circumstances, behaviour or choices. For example automatic processing may lead to the refusal of your online credit application". See European Commission (Ed.). (2017). Can I be subject to automated individual decision-making, including profiling? Retrieved from https://ec.europa.eu/info/law/law-topic/data-protection/reform/rights-citizens/my-rights/can-i-be-subject-automated-individual-decision-making-including-profiling_en See Art. 22 and related provisions EU. (2017). Article 22 EU GDPR "Automated individual decision-making, including profiling". Retrieved August 28, 2018, from http://www.privacy-regulation.eu/en/22.htm . The impact assessment tool proposed by members of the Canadian government aims to create incentives to reduce the number of decisions that are delegated to automated systems, by multiplying the score awarded to these designs, thus increasing the likelihood that they be categorised as "high risk". See Karlin (Op. Cit) https://medium.com/@supergovernance/the-government-of-canadas-algorithmic-impact-assessment-take-two-8a22a87acf6f

technical actors who designed the algorithm and the political staff designing the procedures that state how it will be leveraged. The specifics of each implementation will establish where each algorithm lies on the scale between an *automated* or *self-executable decision* and a *recommendation.* These are not absolutes, but two extremes of a continuum. Most implementations will lie somewhere in the middle.

4. Interaction of output with the broader socio-legal norms to which the target populations are subject

An algorithm's impact is mediated by formal rules, such as those established by the State based on laws and regulations, and also by other more informal social rules, which arise from decentralised processes.

For example, the impact of depriving a person of access to a public service will depend both on their financial situation and on the alternative safety nets which the state provides, including mechanisms and institutions through which redress can be achieved[27]. At the same time, identifying somebody as pregnant, ill or member of some other class may have varying effects in different social surroundings. This component is key when establishing, for example, the suitability of importing models developed in other contexts. At the same time, it highlights the importance of ensuring quality channels of communication between those who develop the algorithms, state institutions, and affected populations.

Given the exploratory nature of the present study, the Web Foundation used a mixed methodology to identify cases: documentary research through Boolean and Advanced Google searches to query specific websites, semi-structured interviews and public information access requests.[28]

27 See paragraph 11 of the Toronto Declaration (Op. Cit.) https://www.amnesty.org/download/Documents/POL3084472018ENGLISH.PDF

28 Further information on the methodology for identifying cases in the annex.

02

CASE STUDIES

THE LATIN AMERICAN CONTEXT

Latin America is one of the most unequal regions in the world.[29] This inequality should be understood as a warning cry to the region's governments, who should allow more substantive citizen participation in public decisions as a way to ensure the government works *for* the people. In this sense, the internet provides an opportunity for the most marginalised groups to act in concert in order to defend their rights and promote their interests more effectively. The internet is also being used by governments for service provision and in some cases as the channel through which data for automated decisions is collected.

INTERNET ACCESS AND AFFORDABILITY

According to ECLAC, in 2015, 54% of Latin Americans had connected to the internet at least once in the last 12 months — an increase of 20 percentage points since 2010.[30] Insofar as internet access is associated to income,[31] this gap implies that high-income sectors have more tools available to them to advance their interests than those with lower incomes.[32] This also limits the ability of the government to leverage the internet as a means for service delivery, since it can further inequality. Yet, it is important to note that, according to the internet affordability index from the Alliance for an Affordable Internet (A4AI), the region is taking steps in the right direction to achieve the goal of 1GB for less than 2% of the average monthly income.[33]

29 Roser, M., & Ortiz-Ospina, E. (2016, October). Income Inequality [Web log post]. Retrieved from https://ourworldindata.org/income-inequality#global-income-inequality

30 CEPAL. (2016). *Estado de la banda ancha en América Latina y el Caribe 2016* (Rep.). Santiago de Chile: CEPAL. https://www.cepal.org/es/publicaciones/40528-estado-la-banda-ancha-america-latina-caribe-2016

31 Poushter, J. (2016). *The strong relationship between per capita income and internet access, smartphone ownership* (Rep.). PEW. doi:http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ and Jansen, J. (2010). *Use of the internet in higher-income households* (Rep.). PEW.http://www.pewinternet.org/2010/11/24/use-of-the-internet-in-higher-income-households/

32 Stanley, W., & Weare, C. (2015). *The Effects of Internet Use on Political Participation: Evidence From an Agency Online Discussion Forum* (Working paper). Cambridge, MA: HKS :https://sites.hks.harvard.edu/m-rcbg/Conferences/rpp_rulemaking/Weare_Political_Participation.pdf

33 Web Foundation. (2017). *2017 Affordability Report* (Issue brief). Washington DC: Web Foundation. https://a4ai.org/affordability-report/report/2017-resumen-ejecutivo/

KEY DATA ARGENTINA

Internet Penetration (2016):
70.97%

Open Data Barometer Score (2016):
37.51 / 100

Secure Servers per 1M people (2016):
61.6

KEY DATA URUGUAY

Internet Penetration (2016):
66.4%

Open Data Barometer Score (2016):
60.85 / 100

Secure Servers per 1M people (2016):
111.2

OPEN DATA

A model of deliberative and participatory democracy, in which public decisions acquire legitimacy from a process, in which each social group has the opportunity to explain and advance their interests, requires widespread access to information about the management of public affairs. In the 21st century this means a policy of open government data, in which access to public data is facilitated in reusable formats. According to the Open Data Barometer produced by the Web Foundation[34], Latin America has an average score of 35/100, which suggests that there is still a lot of work to be done. However, the Barometer highlights the progress achieved in recent years. Latin America has a robust, open data and civic technology community, as evidenced by the popularity of the annual Abrelatam-Condatos events[35]. In 2016, Mexico, Colombia and Uruguay were among the top 20 countries worldwide of the 114 analysed. Aside from being fundamental to public debate, open data is a key tool to ensure a steady stream of quality data for the development of AI systems, and in allowing decentralised control over them.[36]

COMPLEMENTARY INFRASTRUCTURE

The development of artificial intelligence requires infrastructure to allow the secure, large-scale processing of data[37]. At the same time, this infrastructure requires other complementary infrastructure, such as stable energy sources, which are often absent in Latin America. A proxy to estimate the availability of this infrastructure might be the number of secure servers.[38] According to the World Bank (2016), while in the United States there are more than 1623 secure servers for every 1 million people, in Latin America there are only 59.[39] This lack of infrastructure often leads governments to store and process data beyond its jurisdiction. This leads to risks in sovereignty and security. Servers in foreign countries are subject to legal and institutional frameworks, which differ from those approved and controlled by the people whose data is being hosted.

34 Web Foundation. (2017). Open Data Barometer: Snapshot of Latin America (Issue brief). Washington DC: Web Foundation. https://opendatabarometer.org/4thedition/regional-snapshot/latin-america/

35 See The Emergence of Open Latin America in *From Open Government to the Open State:* https://www.cepal.org/en/publicaciones/41353-gobierno-abierto-al-estado-abierto-america-latina-caribe

36 Ortiz Freuler, J., & Fagan, C. (2018, January 11). How open data can save AI [Web log post]. Retrieved from https://webfoundation.org/2018/01/how-open-data-can-save-ai/

37 Amongst other issues, there is the risk of hostile attacks. See Huang, S., Papernot, N., Goodfellow, I., Duan, Y., & Abbeel, P. (2017). Adversarial attacks on neural network policies. https://arxiv.org/pdf/1702.02284.pdf

38 Secure Server (or SSL Server) refers to a web server that supports any of the main security protocols, which encrypts and decrypts messages to protect them from third parties. Online businesses, for example, use secure servers to ensure the security of online purchases.

39 In absolute terms, while in the US there are more than 500,000 secure servers, in Latin America there are 37,000 (figures for 2016). See World Bank. (2017). *Secure Internet servers (per 1 million people)*. Retrieved June 02, 2018, from https://data.worldbank.org/indicator/IT.NET.SECR.P6?end=2016&locations=ZJ-1W-US&start=2001&view=chart

In this context, as one would expect, the development of artificial intelligence systems is at an early stage. Yet, figures 3 and 7 suggest that there is a growing interest in these systems. International organisations are not unaware of this trend. ECLAC, for example, has examined how these technologies can be used to promote economic and social development, within the framework of the United Nations' Sustainable Development Goals (SDGs).[40] The World Bank, for its part, is developing methodologies to use these tools in the context of impact studies on development policies.[41] The Web Foundation itself has highlighted how these new technologies are being used to improve living conditions in low and middle-income countries.[42]

Although it is a relatively new item on the region's agenda, resource constraints at the local level do not necessarily translate into barriers for the adoption of these technologies. Even in the absence of local infrastructure, officials have access to online platforms, which allow systems to be developed on servers based abroad.[43] At the same time, companies that develop products for the governments of countries with more advanced ecosystems are beginning to offer their services in the region.[44] These factors, combined with an explosion in expectations regarding the potential of algorithms and artificial intelligence systems (some more reasonable than others[45]), increase the risk that unfeasible, inefficient or even harmful solutions are adopted. The establishment of policies to mitigate these risks and maximise potential benefits should be a priority for the region.

40 ECLAC, Data, algorithms and policies: the redefinition of the digital world (LC / CMSI.6 / 4), Santiago, 2018. Available at: https://repositorio.cepal.org/bitstream/handle/11362/43477/7/S1800053_en.pdf [Date of consultation: 05/08/2018]

41 McKenzie, D. (2018, May 03). How can machine learning and artificial intelligence be used in development interventions and impact evaluations? [Web log post]. Retrieved from http://blogs.worldbank.org/impactevaluations/how-can-machine-learning-and-artificial-intelligence-be-used-development-interventions-and-impact

42 See (op. Cit) https://webfoundation.org/research/white-paper-series-opportunities-and-risks-in-emerging-technologies/ and Brandusescu, A., Ortiz Freuler, J., & Thakur, D. (2017). *Artificial Intelligence: Starting the policy dialogue in Africa* (Working paper). Washington DC: Web Foundation. http://webfoundation.org/docs/2017/12/Artificial-Intelligence-starting-the-policy-dialogue-in-Africa.pdf

43 For example TensorFlow, from Google (https://www.tensorflow.org/); CNTK, from Microsoft (https://www.microsoft.com/en-us/cognitive-toolkit/); and Torch from Facebook (https://research.fb.com/downloads/torch/) among others.

44 For a broader sketch of the variables that may affect a country's readiness for the development and roll out of artificial intelligence in a country, see Sterling, R., Miller, H., & Martino-Trustwell, E. (2018). Government AI Readiness Index [Web log post]. Retrieved from https://www.oxfordinsights.com/government-ai-readiness-index

45 Piekniewski, F. (2018, May 05). AI winter is well on its way [Web log post]. Retrieved from https://blog.piekniewski.info/2018/05/28/ai-winter-is-well-on-its-way/

ARGENTINA

Argentina has made significant progress in recent years in the use of technology in public policies, particularly in open data. Among the first actions of the current national administration was the establishment of the Ministry of Modernisation (Decree No. 13/2015), which regulates open data in the national public administration.[46] Moreover, the work of the National Data and Public Information team meant that Argentina showed the one of the greatest increases worldwide in the 2017 edition of the Open Data Barometer.

The current administration, elected in December 2015, has shown particular interest in the opportunities artificial intelligence presents. One measure of this this interest is Figure 3, which shows the number of annual publications about AI found on ministries' websites doubled from 100 to 200 in just two years.

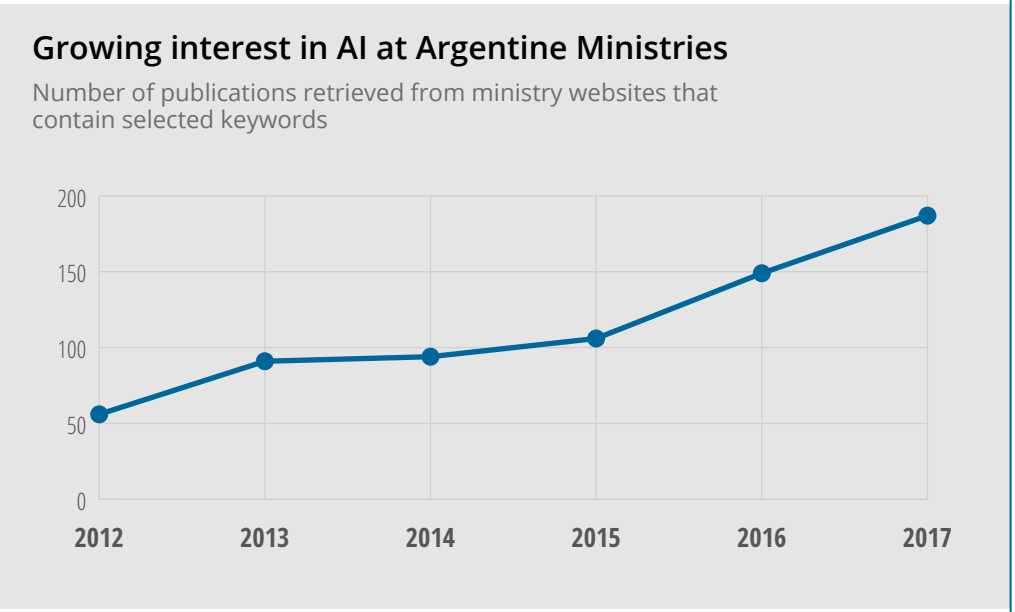


Fig. 3 — Number of publications retrieved through Google from Argentinian ministries' websites, which include the words "Artificial Intelligence" / "Machine Learning" / "Deep Learning"[47]

Among other methods used to promote these technologies are exchanges between academia and public servants. In October 2017, for example, the Ministry of Modernisation organised, together with the Sadosky Foundation, a summit for personnel from the government's computer departments to interact with academics.[48] The objective was to allow AI experts to share their experience in the use of these technologies and discuss their potential within the public administration. Some months later, in May 2018, the Ministry of Modernisation partnered with the ITU for Argentina to host the "1st Artificial Intelligence Forum and Internet of Things in Smart, Sustainable Cities in Latin America".[49]

46 Executive Decree No. 117/2016 available at http://servicios.infoleg.gob.ar/infolegInternet/anexos/255000-259999/257755/norma.htm

47 Searches were restricted to each particular year. Consequently, the number of publications is not cumulative.

48 Fundación Sadosky. (2017). Ciclo de charlas de Innovación: "Inteligencia Artificial, Ciencia de Datos y Gobierno" [Web log post]. Retrieved from ttp://www.fundacionsadosky.org.ar/ciclo-de-charlas-de-innovacion-inteligencia-artificial-ciencia-de-datos-y-gobierno/

49 ITU. (2018). 1st Forum on Artificial Intelligence and the Internet of Things in Smart Sustainable Cities in Latin America [Web log post]. Retrieved from https://www.itu.int/en/ITU-T/Workshops-and-Seminars/20180530/Pages/default.aspx

Cases

Through searches and interviews with experts on the subject matter, three particularly interesting cases of AI being used in public policy were identified: **(1) Prediction of school dropouts**, **(2) prediction of teenage pregnancies**, and **(3) a map to identify business opportunities.**

SALTA, ARGENTINA

After signing an agreement with Microsoft in June 2017, the Government of the Province of Salta developed tools to predict school dropout and teenage pregnancy. In both cases the datasets used were provided by the Province's Ministry of Early Childhood, which is in charge of the system. The government is currently in negotiations to export the model to the provinces of Tierra del Fuego, La Rioja, Chaco and Tucumán, as well as the Department of Guajira, in Colombia.

As both cases are very similar in their development, we have examined them together.

Objective:

Prediction of School Dropouts[50]

Unit Responsible: Ministry of Early Childhood of the Government of Salta, Argentina

Development: Microsoft / Ministry of Early Childhood[51]

Origin of the Data: Surveys by the Early Childhood and Education Ministries.

Public Availability of Data: No (Personal Data Protection Law 25,326)

Selection of Variables: up to 80 variables depending on the person (Personal (9), Education (7), Health (11), Employment (10), Housing (20), Family (23))[52]

Variables capable of triggering undue discrimination:[53]: Yes

Model intelligibility: Black Box

Output: Assigns a drop out probability to each member of the sample, and highlights the subset of those with the highest probability.

Reported Error Ratio: 20% false positives[54]

Takes or Assist decisions: Assists

Consequences: Ministry of Early Childhood coordinates actions with other ministries, according to the profile of the person identified

Impact (according to those responsible): The model assigned a school drop out probability of more than 70% to 418 children and adolescents. The impact of subsequent interventions is still unknown.

50 Complete file available in digital annex

51 An agreement was subscribed with Microsoft, who provided training and helped develop the first version of the model, developed in January 2017. According to sources at the Ministry of Early Childhood, the model continues to be improved and modified.

52 Complete list in the digital Annex

53 We rely on the doctrine of suspect classification: In the legal domain it is understood that certain categorizations may lead to discrimination. Consequently, in several jurisdictions, government practices or statutes that make distinctions between individuals based on certain characteristics (below) considered to be capable of fuelling undue discrimination are subject to special scrutiny by the judiciary. This study seeks to ensure algorithm designers are made aware of the risks triggered by the inclusion of certain categorizations. To such end we encourage designers to actively flag whenever they use any of these variables. Reporting error ratios for each category would also help reduce the odds that tools that discriminate are rolled out. The United Nations High Commissioner for Human Rights refers to the following classes "race, colour, sex, language, religion, political or other opinion, national or social origin, property, birth or other status, and which has the purpose or effect of nullifying or impairing the recognition, enjoyment or exercise by all persons, on an equal footing, of all rights and freedoms"

54 According to data provided by the Ministry of Early Childhood (see annex) the trained model was able to successful in identifying 80% of the adolescents who had dropped out. One of the experts consulted considers this is not the appropriate way to define the model's error ratio. As the model is prospective, it should be assessed based on its capacity to predict students who *will* drop out, not based on its capacity to identify those *who have already* dropped out. If the model has been trained to identify those who have already dropped out, it might be taking into consideration characteristics that only surface after a student has dropped out, weakening its prospective abilities and increasing its ratio of error.

Objective:

Prediction of Teenage Pregnancy[55]

Unit Responsible: Ministry of Early Childhood of the Government of Salta, Argentina

Development: Private (Microsoft) / Public (Ministry of Early Childhood)[56]

Origin of the Data: Ministry of Early Childhood surveys

Public Availability of Data: No (Personal Data Protection Law 25,326)

Selection of Variables: up to 78 variables depending on the person (Personal (8), Education (5), Health (12), Employment (9), Housing (20), Family (23))[57]

Variables capable of triggering undue discrimination:[58] : Yes

Model Intelligibility: Black box

Output: Assigns a probability of pregnancy to each member of the sample (women between 15 and 19 years old) and identifies the subset of those with the highest probability of pregnancy.

Reported Error Ratio: 15% false positives[59]

Takes or Assists Decisions: Assists

Consequences: The Ministry of Early Childhood coordinates actions with other ministries, according to the profile of the person identified

Impact (according to those responsible): Model identified 250 adolescent women with a +70% probability of pregnancy. Impact of subsequent interventions remains unknown.

Data

Pablo Abeleira, technology coordinator at the Ministry of Early Childhood of Salta, informed us[60] that the tools mainly rely on data collected in low-income areas of the south-eastern, eastern and western districts of the city of Salta in the years 2016 and 2017. This set of data is collected based on a system which allows for coordination between ministries, civil society organizations, and Microsoft, which supplies the basic technology and hosting.[61] In total, 296,612 people were surveyed, of whom 12,692 were women between the ages of 10 and 19. In the case of early school leaving, data from a Ministry of Education survey, which collected data on 10,244 children and adolescents between 6 and 20 years old, was also used.

The details of the methodology underlying these data collection processes are well documented according to the government, but not publicly available. Furthermore, in the absence of an open data law, government departments do not allow the data to be accessed by the public. Though privacy considerations might trump the release of the full dataset, partial and anonymised subsets could be useful.

As Boyd and Crawford point out,[62] it is important that those who build databases explain publicly how the data's characteristics limit their potential uses. The absence of clarifications limits the opportunity to design the model in such a way that it neutralises or corrects the biases which a database may have[63].

55 Complete file available in digital annex

56 An agreement was signed with Microsoft, which provided training and help in the first version of the model, which was developed in January 2017. According to sources from the Ministry of Early Childhood, the model continues to be improved and modified.

57 Complete list in the digital Annex

58 See footnote 53 or Annex for definition

59 According to data provided by the Ministry of Early Childhood (see annex) the trained model was able to successfully identify 85% of the adolescents who were pregnant. One of the experts consulted considers this is not the appropriate way to assess the model's error ratio. As the model is prospective, it should be assessed based on its capacity to predict which adolescents will be pregnant, not based on its capacity to identify those who already are pregnant or mothers. If the model has been trained to identify those who have already are pregnant, it might be taking into consideration characteristics that only surface after an adolescent is pregnant (eg. dropping out of school), weakening its prospective abilities, and increasing its ratio of error.

60 Internet interview conducted on 15 May, 2018

61 The system is supported by an app that standardizes the information collected by several organizations operating in the territory. While each organization is given exclusive access to the databases each has compiled, the Government has access to all the data. CONIN and TECHO are the most noteworthy of the participating organisations. For more information see Ministerio de Primera Infancia (Director). (2016, November 8). *CONIN Microsoft MPI* [Video file]. Retrieved from https://www.youtube.com/watch?v=vG6zNFLyrrg and Gracias a la Nube, crean primer mapa de desnutrición en el país. (2017, June 2). *DocSalud*. Retrieved from http://www.docsalud.com/articulo/8069/gracias-a-la-nube-crean-primer-mapa-de-desnutrición-en-el-país The data is hosted on a server outside Argentina but, according to ministerial authorities, the data is hosted in accordance with the relevant laws. In an interview published by UNO (https://uno.com.ar/tecnologia/como-funciona-el-sistema-para-predecir-embarazos-adolescentes-de-salta-04122018_rJxfqbraiM) Abeleira stresses that the consent of the people whose data is collected.

62 See point 3. boyd, D. & Crawford, K., "Six Provocations for Big Data. A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society", September 2011. Available at: http://softwarestudies.com/cultural_analytics/Six_Provocations_for_Big_Data.pdf [Date of consultation: 05/06/2018].

63 These could be corrected by suppressing sensitive attributes, changing labels, reweighting attributes, or resampling underlying data, amongst other approaches. See Rieke (Op. Cit) http://omidyar.com/sites/default/files/file_archive/Public%20Scrutiny%20of%20Automated%20Decisions.pdf

Before the implementation of any AI or algorithmic tool, the relevant authorities should release a technical report outlining the characteristics of tools. Such report should include — at a minimum — the relevant metadata for the training set and lists of all the sources of data upon which the system acts on to produce an output. In the specific case of Salta this report should clarify that, because the data used to train the model came from people who live in certain low-income areas,[64] it is, by definition, not capable of predicting cases of teen pregnancy or early school leaving among members of other groups. The model is constrained by the databases used in its development. Yet, in public presentations the tool was shown as capable of offering predictions about the general population. This led to much criticism from the technical community. In an interview, Pablo Abeleira argued this was miscommunication, and that his technical team is aware of this limitation.

Design of the Model

From the available information, it appears that the models were built through an iterative process, in which variables were added and removed in order to achieve a model with greater predictive capacity. What Abeleira defines as the first iterations are available at the online portal GitHub[65]. The developers Marcelo Felman and Facundo Davancens have made public a range of explanations about the methodology used to make the decisions, which are summarised in four steps: knowing the domain, preparing the database, defining the model and incorporating it.

The fact that the methodology is available on a website like GitHub makes the process more transparent and allows the public to provide input. Proof of this value is that two separate experts offered critical feedback regarding the teenage pregnancies model.[66] Subsequent iterations, however, were conducted behind closed doors by government technicians.

The selected method (*two class boosted decision tree*[67] — a type of *machine learning algorithm*) has a relatively high degree of intelligibility, compared to other alternative approaches. As shown in the second section of Fig. 5, the interface includes a dashboard that lists the variables that were particularly relevant in defining the classification (high probability of pregnancy / early school leaving). If well grounded, this could be considered an explanation of the *output*.

Yet, although the method allows a degree of explainability, implementation is performed inside a black box. Although we may be able to access the inputs (in this case,a private dataset) and the outputs (which are confidential for good reason), it is not possible for the people affected to learn how or why the system arrived at a particular output. There is no theoretical framework explaining the relationship between the model and the underlying problem. There is no justification for the variables or the selected methodology.[68] Affected populations are asked to trust that the opaque processes for data collection and algorithm design are effective and are designed to achieve fair results. These procedural weaknesses undermine the system's legitimacy.

The iterative process, which is described on GitHub (and which according to Abeleira continued within the Ministry) has the advantage of helping to increase the effectiveness of the predictive model. However, if mishandled, it runs the risk of creating a less sustainable model, for example, through *overfitting*[69].

64 See *Sobre la predicción automática de embarazos adolescentes* (Rep.). (2018). Buenos Aires: Laboratorio de Inteligencia Artificial Aplicada. doi:https://www.dropbox.com/s/r7w4hln3p5xum3v/[LIAA] Sobre la predicción automática de embarazos adolescentes.pdf?dl=0

65 Davancens, F., Teen Pregnancy Prediction with Machine Learning, September 14, 2017. https://github.com/facundod/casestudies/blob/master/Prediccion%20de%20Embarazo%20Adolescente%20con%20Machine%20Learning.md and Felman, M., "Predicting School Dropouts with Machine Learning", January 4, 2018. https://github.com/marcelofelman/case-studies/blob/master/Desercion%20escolar%20con%20Machine%20Learning.md

66 In the case of adolescent pregnancy detection, for example, three GitHub users left comments concerning methodological issues they argue require revisions. See comments to Davancens, F., Teen Pregnancy Prediction with Machine Learning, September 14, 2017.https://github.com/facundod/case-studies/issues/2

67 In a decision tree the leaves represent class labels and the branches represent conjunctions of characteristics that lead to those class labels. An enhanced decision tree is an ensemble learning method in which the second tree corrects the errors of the first tree, the third tree corrects the errors of the first and second tree, and so on. See more at Microsoft Azure. (2018, January 9). Two-Class Boosted Decision Tree [Web log post]. Retrieved from https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-boosted-decision-tree

68 It should be noted that although the documents provided by the Ministry of Early Childhood show how the software highlights certain variables of particular relevance (see annex), the weight of each one is not clear and neither is the role played by the rest of the variables

69 *Overfitting* occurs when a model is designed taking excessive consideration for the particular characteristics of the training set. It fits the training set very well, but overfits the data its expected to perform on afterwards. That is, when the trained model is then utilised to classify elements of a new dataset, the excessive consideration for characteristics that were particular to the training set backfire, leading to a considerable drop in accuracy. Think of a glove that is designed based on the hand of a person that has a wart, and the designers decide to include space for the protuberance and then market it to the masses.

In terms of the tool's effectiveness, according to data provided by the Ministry of Early Childhood (see annex) the trained model was able to successfully identify 85% of the adolescents who were pregnant. One of the experts consulted considers this is not the appropriate way to assess the model's error ratio. As the model is prospective, it should be assessed based on its capacity to predict which adolescents *will be* pregnant, not based on its capacity to identify those who already are pregnant or have given birth. If the model has been trained to identify those who have already are pregnant, it might be taking into consideration characteristics that only surface after an adolescent is pregnant (eg. dropping out of school), weakening its prospective abilities, and increasing its ratio of error.

The Applied Artificial Intelligence Laboratory of the Computing Department of the University of Buenos Aires also expressed its concerns. It produced a report based on the information from the initial model, publicly available through GitHub, for adolescent pregnancy. The lab highlighted three problems[70]:

1. Overestimation of the effectiveness of the tool due to the reuse of the training set as evaluation data.
2. Biased data, since it is limited to the most vulnerable sectors of the population.
3. Inadequacy of the data to answer the question that seeks to be addressed. The conditions which led to a pregnancy in the past will not necessarily lead to pregnancies in the future inasmuch as the other variables affecting the outcome do not remain fixed.

The Ministry of Early Childhood sent us their rebuttals to these criticisms, which could be paraphrased as follows:

1. The model published in GitHub re-used the training data as evaluation data. Yet the current model's reported accuracy is based on tests using independent databases.
2. There is no bias in the database because the model seeks to make predictions *only* for these vulnerable populations (which is the Ministry's scope of action).
3. The model is updated and re-trained on a regular basis based on data that is continuously updated. The models do rest on a series of assumptions about the possibility of extrapolating conclusions from patterns detected within a dataset, but these assumptions are not particular to the model; they are the bedrock of statistical analysis as a discipline.

In view of the criticisms received, it would be advisable for the Ministry of Early Childhood to publish a technical report with information about the databases used and the variables selected, as well as the hypothesis, which guides the design of the model, and the design process that lead to the definitive model.

This explanatory document should also include an analysis of similar models. For example, a study on early school leaving in the United States found that variables such as ethnicity, special education, birth in the United States, English as a mother tongue, school behaviour, and the school's geographical region did not significantly predict early school leaving.[71] However, equivalent variables were included in the Salta model. Those in charge of design should be expected to offer a theoretical explanation for their inclusion.

70 See *obre la predicción automática de embarazos adolescentes* (Op. Cit) https://www.dropbox.com/s/r7w4hln3p5xum3v/[LIAA] Sobre la predicción automática de embarazos adolescentes.pdf?dl=0

71 Wood, L., Kiperman, S., Esch, R., Leroux, A., Truscott, S., & Gilman, Richard C. (2017). Predicting Dropout Using Student- and School-Level Factors: An Ecological Perspective. School Psychology Quarterly, 32 (1), 35-49. https://hollis.harvard.edu/primo-explore/fulldisplay?docid=TN_apa_articles10.1037/spq0000152&context=PC&vid=HVD2&search_scope=everything&tab=everything&lang=en_US

Decision / Interpretation of the Output:

The output is a list of minors who are assigned a probability of dropping out of school, or pregnancy. The algorithms do not execute any policy as a consequence of a particular classification. The system is designed to help the Ministry of Early Childhood to identify individuals who require particular attention from the State[72].

Pablo Abeleira, who leads the project at the ministry, mentioned[73] that, for example, they aim to check whether there are members of the family who should receive some type of social assistance and are not receiving it. This is in line with what Abeleira said in an interview with UNO, where he stated: "It is a comprehensive approach. We send the results to the appropriate areas and work is performed among all sectors of the government. "[74]

As these are very sensitive areas, it is important that a protocol defining the follow-up actions be co-designed with the communities that the ministry is targeting, topic specialists, and political authorities in areas of government that might be impacted by a shift in government policy. Quarterly reports on the impact of the protocol would also help the community understand the ways in which their data is being used.[75]

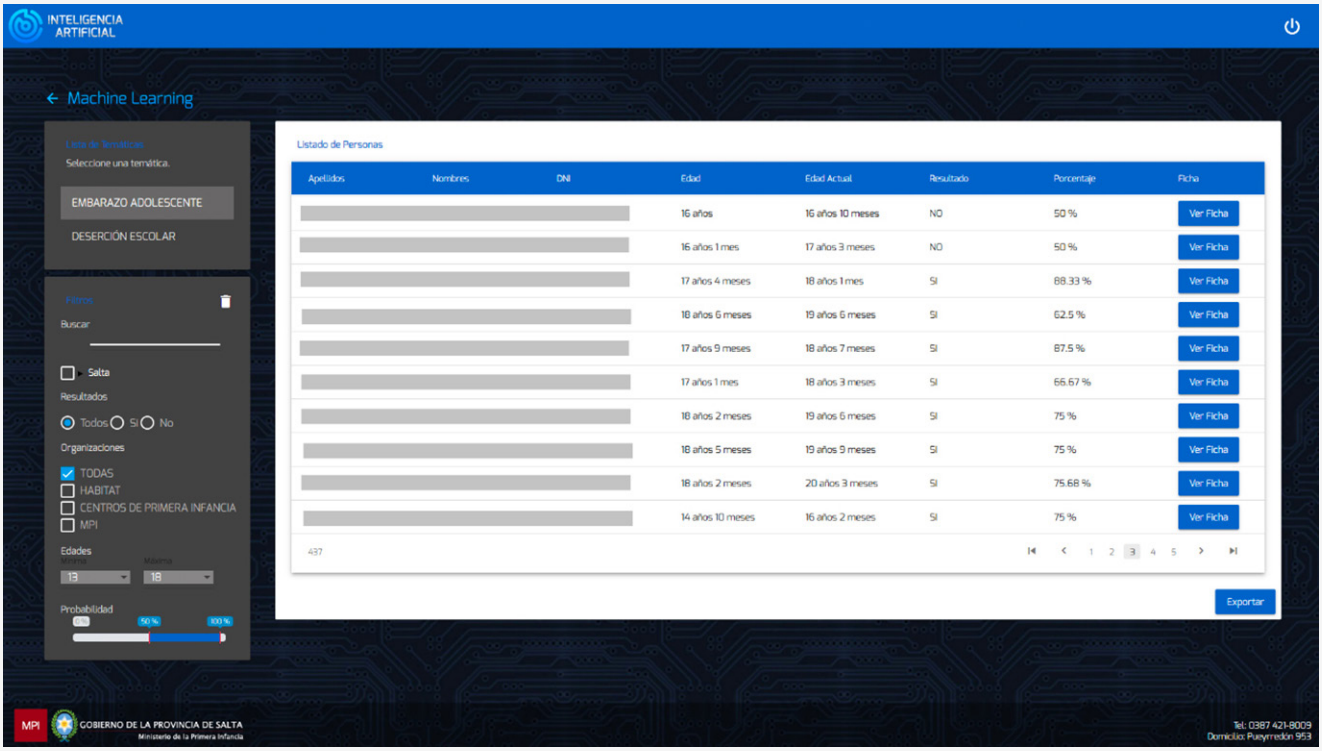


Fig. 4 — Section of the control panel developed by the Ministry of Early Childhood of Salta showing the results sheet for a group of women[76]

72 See Ministerio de Primera Infancia (Producer). (2016, January 21). *Ministerio de la Primera Infancia - Visión* [Video file]. Retrieved from https://www.youtube.com/watch?v=F8Lg9ytO9JY

73 Interview with Pablo Abeleira, technology coordinator of the Ministry of Early Childhood of Salta (5/15/2018)

74 UNO (Op. Cit) https://uno.com.ar/tecnologia/como-funciona-el-sistema-para-predecir-embarazos-adolescentes-de-salta-04122018_rJxfqbraiM

75 For example, in the interview referred to above, Pablo Abeleira was asked whether sex education and access to condoms and contraceptives were part of the governmental response to the identification of an adolescent that predicted to become pregnant. However, no specific response to this point was made. Iit would be useful to clarify the role the Church and civil society play in the program's implementation (See min. 1.48 of the institutional presentation: Ministerio de Primera Infancia (Op. Cit) https://youtu.be/F8Lg9ytO9JY?t=108)

76 Information provided by Pablo Abeleira via email. Complete document available at the digital annex.

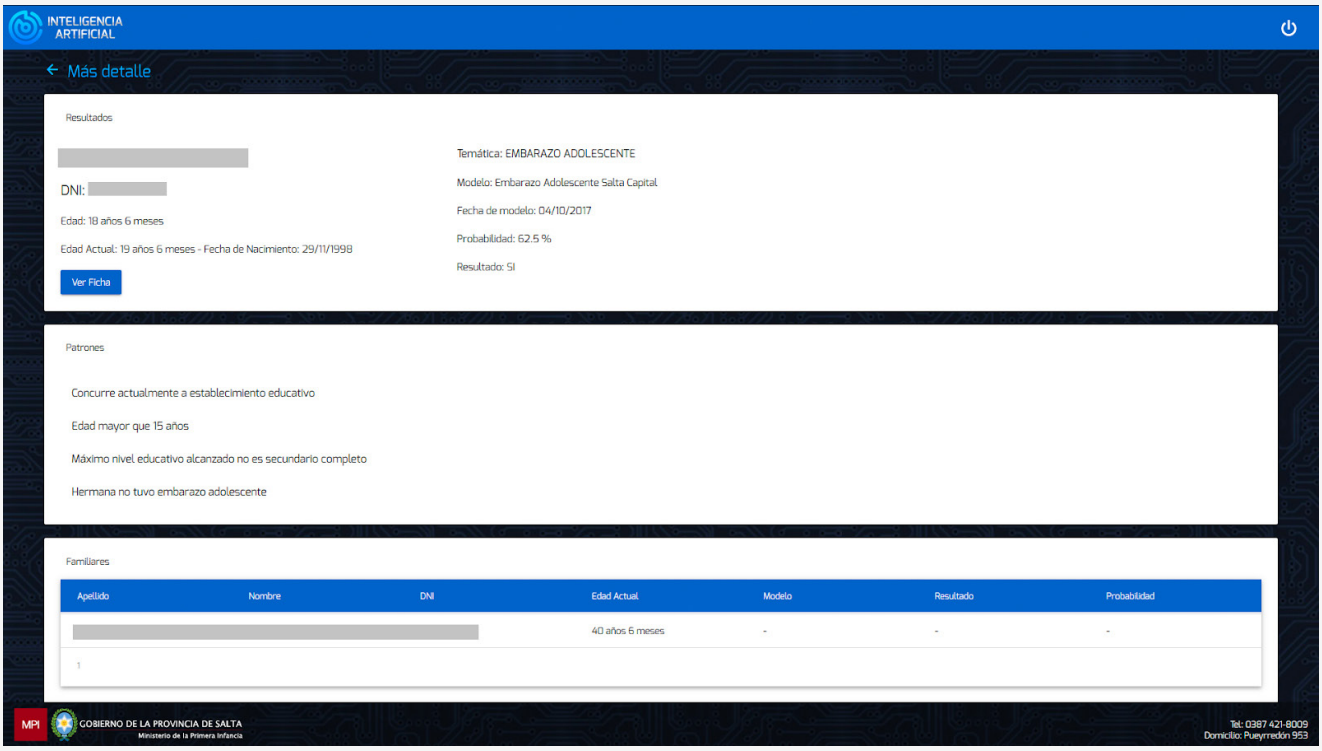


Fig. 5 — Section of the control panel developed by the Ministry of Early Childhood of Salta, showing the results sheet for a woman identified as having a high probability of pregnancy[77]

Interaction with Social and Legal Frameworks:

It is interesting to start by reviewing what the developers mention in their GitHub repositories: "Although the ideal thing would be to have an expert in the field, there will not always be one available. For this reason, it is important to consider various techniques as well as to use our common sense ". The sentence reveals a position that it is permissible to use the technology without consulting topic experts or the affected populations.[78]

This approach has received a lot of criticism. In an open letter entitled "Artificial intelligence or patriarchal artifices?",[79] the Observatory on Violence Against Women, a local NGO highlighted a range of technical[80] and political issues. Of particular concern is the infringement of minors' right to privacy[81], and the apparent ignorance of the context of structural social inequality, which prevents women from exercising their sexual and reproductive rights.

The signatories state that 20% of the births in Salta are to adolescents and that the levels of gender violence in the province are too alarming to be ignored. They highlight a contradiction between developing a complex technical system aimed at identifying particularly vulnerable individuals while the province lacks any form of comprehensive public policies aimed at eliminating the structural inequality that makes them vulnerable, or providing women with the information and resources to allow them exercise their rights effectively[82].

77 Information provided by Pablo Abeleira via email. Complete document available at the digital annex.

78 This approach contrasts, for example, with the one taken by the Latin American Open Data Initiative (ILDA) when designing an open data standard to identify cases of femicide, where female experts helped to establish how data is opened, which data will be made public. Experts are also called in to define the way in which the data are collected, and how it will be utilised. Fumega, S., "Some News of Our Work on Data, Gender and Security", 09 February 2018. https://idatosabiertos.org/algunas-novedades-de-nuestro-trabajo-sobre-datos-genero-y-seguridad/

79 Cuarto Poder Salta, "Artificial intelligence or patriarchal artifices?", 16 April, 2018 Available at: https://www.cuartopodersalta.com.ar/inteligencia-artificial-o-artificios-patriarcales/

80 The lack of transparency in the data collection and the sensitive nature of the variables utilised by the model, among others.

81 According to Art. 10 of Law 26.061 "Girls and boys and adolescents have the right to private life and privacy of and within family life. These rights may not be subject to arbitrary or illegal interference." https://www.oas.org/dil/esp/Ley_of_Protection_Integral_of_the_Rights_of_the_Nina_Ninos_and_Adolescents_Argentina.pdf

82 A few weeks after the publication of this letter, an 11-year-old girl from Salta who had been raped by her stepfather did not get a legal abortion. The authorities stated that the 11-year old girl had decided to voluntarily continue with the pregnancy. This illustrates what adolescent women endure in Salta. Luego de difusión de caso de niña embarazada tras violación, Salta adhirió a protocolo de aborto no punible de Nación. (2018, May 24). *Ámbito Financiero*. Retrieved from http://www.ambito.com/922368-luego-de-difusion-de-caso-de-nina-embarazada-tras-violacion-salta-adhirio-a-protocolo-de-aborto-no-punible-de-nacion

Summary of the Findings:

- **Effectiveness of the model:** Although the reported false positive rates are relatively low, both a group of academic specialists from the public university and one of the independent specialists consulted have questioned the methodology with which they are calculated. Disclosing further information to the public regarding the design and implementation of the model would help resolve these tensions, and possibly improve the predictive accuracy of the model.
- **Effectiveness of implementation:** The Ministry of Early Childhood coordinates actions with other ministries, but has not managed to consolidate information on the impact. Publishing the administrative protocols that establish the follow-up actions to be performed based on the outputs as well as quarterly reports on the impact achieved through these actions would strengthen the ministry's position.
- **Legitimacy (process):** The model tells the system operator what the main variables driving the probability assigned to a specific case are. This suggests the model has some degree of explainability. However, these explanations are available only to public officials. Furthermore, databases are closed and the model's design is not available to the public, which limits traceability of errors and restricts the possibility of performing independent audits.
- **Legitimacy (result):** As the ministry does not collect and consolidate information on the impact of these tools, it is not possible to determine whether their implementation tends to result in fair or unfair outcomes. Women's rights activists have questioned the decision to implement these types of tools without a framework that embeds them within a policy that targets the structural inequality suffered by the populations it claims to support.

It is difficult to design effective public policies without data. The Ministry of Early Childhood is systematically surveying excluded populations, leading to datasets that could help offer better quality of service to those who need them the most. Due to the urgency and sensitivity of the issues the ministry seeks to address, coordination with actors from different sectors (academia, women's groups, human rights defenders) and direct participation mechanisms for the affected populations are essential to provide guarantees that a comprehensive, respectful approach to the rights and interests of affected populations has been taken.

CITY OF BUENOS AIRES, ARGENTINA

Objective:

To facilitate the identification of business opportunities[83]

Unit Responsible: Min. of Modernisation of the City of Buenos Aires

Development: State

Origin of the Data: Local and Federal Government, Private (*Properati*)

Public Availability of Data: Yes

Selection of Variables: Resident population by area; Working population by area, disaggregated by gender; Prices for purchase and rental of premises; quarterly openings and closings of businesses, walkability index.

Variables capable of triggering undue discrimination:[84] No

Algorithm intelligibility: High (basic decision tree)

Output: Statistics on businesses and populations of interest

Reported Error Ratio: NA

Takes or Assists Decisions: Assists

Consequence: No actions by the government ensue. It assists potential investors

Impact (according to people in charge): Unknown

The Commercial Opportunities Map (MOC[85]) is a web-based tool developed by the Government of the City of Buenos Aires to provide information to entrepreneurs interested in starting or expanding businesses in the city. It relies on a very basic algorithm that illustrates a case of low risk and transparent use of this technology.

Once the user selects a type of business and a geographical area, the tool projects the subset of the database that is of interest to the user. In this way it provides specific, detailed information about the characteristics of the market in each area. The tool also assists government agencies in assessing the commercial impact of their policies by observing the variations of each indicator over time.

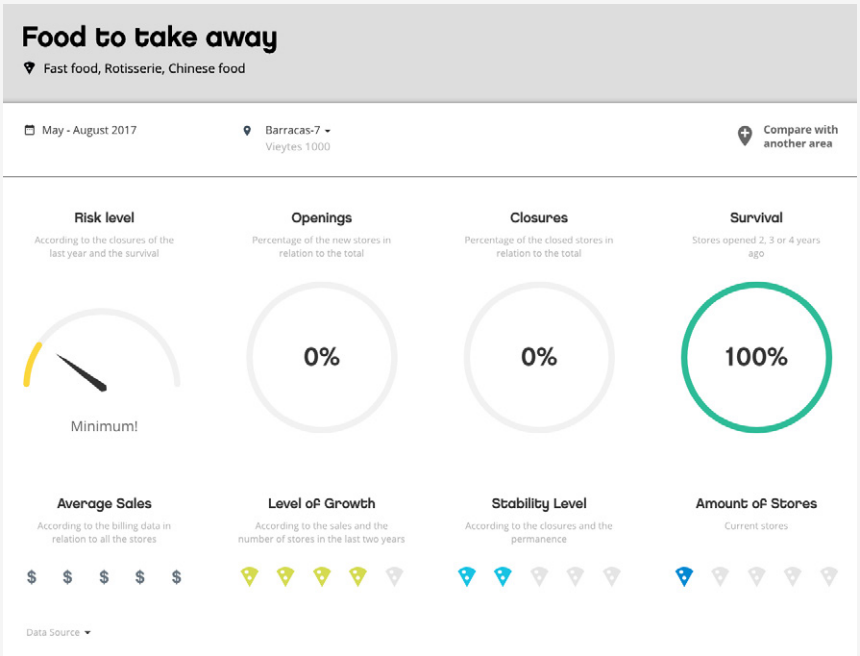


Fig. 6 — Screen print of the preliminary indicators section, which the MOC provides as output

83 Complete file available in digital annex

84 See footnote 54 or Annex for definition In some jurisdictions, what originally emerged as judicial doctrines establishing strict scrutiny of suspect classifications has been enshrined acknowledged by lawmakers. The Constitution of the City of Buenos Aires, for example, in Article 11 defines these suspect classes as "race, ethnicity, gender, sexual orientation, age, religion, ideology, opinion, nationality, physical characteristics, mental or physical disability and socioeconomic status."

85 Mapa de Oportunidades Comerciales. (2018). Retrieved from https://moc.buenosaires.gob.ar/ See also Map Of Commercial Opportunities, 2018 WSIS Prizes, Category E-Business, https://www.itu.int/net4/wsis/stocktaking/projects/Project/Details?projectId=1514904381

Data

It is worth highlighting that under each of the displays generated by the algorithm, the user is informed about the original source of the data informing the statistic (see fig. 6). This not only increases transparency, it facilitates the identification of databases with the potential to be reused for other purposes. Of the six displays presented to the user, four are based on data which are available through the government's open data portal. The maximum and minimum prices for commercial premises, on the other hand, are built using data from Properati, a private online platform that connects buyers and sellers of real estate[86]. One of the Properati's characteristic features is that it releases a lot of the data that drives its proprietary search engine in open formats (including those which are used by the government in this case). As best practice, the report made available for download should include links to each database and sources of information as well as any documentation available about them.

We were informed by the Ministry of Modernisation that, in the next and subsequent editions, the movement of people in each area will be estimated based on anonymised databases provided by the private telecommunications company *Telefonica*.[87] Leaving aside the debate about the risks involved in the government having access to this data,[88] it is important that the contract between the government and the company be available to the public together with a file including the corresponding metadata associated with the data received.

Algorithm Design

The website requires users to select i) a geographical area, and ii) a business category. It then offers users information based on their self-reported interests. This process is simple, transparent and intelligible. It can be classified as a basic decision tree, and is useful to illustrate an algorithm in its simplest form. When we consider the output, the exceptions are *level of risk*, *level of growth* and *level of stability*. These outputs seem to have been subject to additional processing. The details of this processing should be explained to users.[89]

Although the platform's focus is on the disaggregation of statistical information about the past, the "risk level" component contains a probability about the occurrence of future events. An estimate of reliability or error of this assessment would be helpful as data gets updated. As reported by María Dorado,[90] Head of the Directorate of Open Government of the Ministry of Modernisation, the second version will include a predictive model. It is important that this second version provides specific information about margins of error, and reports the confidence bands over time.

Decision / Interpretation of the Output

No action is taken by the government as a result of the output. It does not allocate a benefit or access to a service. The platform provides information that enables individual investors to make better informed decisions. In doing so, the tool seeks to reduce the information gap between investors who have access to specialists, consultants and private databases, and those who do not.[91] It would be desirable to offer access to statistics about the use of the tool, including what selections are most popular.

86 Properati Data. (n.d.). Retrieved August 29, 2018, from https://www.properati.com.ar/data

87 Dorado, M. (Personal Communication via Google Hangout) March 14, 2018.

88 Although it is said to be an anonymised database, there is a wide range of techniques that can be leveraged to deanonymize it. See, for example Hardesty, L. (2013, March 27). How hard is it to 'de-anonymize' cellphone data? [Web log post]. Retrieved from https://news.mit.edu/2013/how-hard-it-de-anonymize-cellphone-data

89 Illustrative example: Mapa de Oportunidades Comerciales (Op. Cit.) https://moc.buenosaires.gob.ar/results_redesign.html?barrio=Boedo&zona=111&address=&rubro=Comida%20al%20Paso&active_tab=commerce-pill)

90 Dorado, M. (Personal Communication via Google Hangout) March 14, 2018.

91 Practically all the data come from publicly available databases, meaning they can be reused for other purposes. Although, as previously stated, it would be useful to reveal the exact calculation performed to generate the outputs for "Risk Level", "Stability" and "Growth".

Interaction with Socio-cultural and Legal Frameworks

Improving the quality of information available in a market is clearly a government function. This reduces uncertainty and allows new players to enter, thus making the market more dynamic. Improving markets, however, does not necessarily imply benefits for all actors. The information made available to the public may lead to more resources being invested in more affluent areas, or to areas experiencing gentrification or growth. In other words, though the model may not seem, on its face, to be prone to triggering discrimination, it can still reinforce structural inequities. Another key government function is to improve people's quality of life. It would therefore be helpful if users could access explanations about:

1. The relationship between government investments in infrastructure and outputs offered by the algorithm.
2. Descriptions of the policies that are in place to help those in areas with poor or declining indicators.[92]

Summary of the Findings:

- **Effectiveness of the algorithm:** It is a simple design, which offers statistical information based on public databases. Subsequent versions will include predictive models.
- **Effectiveness of implementation:** It is difficult to make a causal link between use of the tool and changes in investment patterns. In any case, there is no data available about its impact.[93]
- **Legitimacy (process):** It is a highly intelligible algorithm. It would, however, be useful to provide more information about how some outputs are estimated, such as "risk level".
- **Legitimacy (results):** As the tool provides information to an undefined group, it is difficult to assess the impact. It would be useful for the MOC to offer information regarding how government policy will improve indicators in those areas classified as high risk for investment, in order to ensure residents of these areas observe improvements in their quality of life.

92 See, for example, Berkowitz, M., & Muggah, R. (2017, March 13). Opinion: The smartest cities are resilient ones [Web log post]. Retrieved from https://www.devex.com/news/opinion-the-smartest-cities-are-resilient-ones-89476

93 Neither is there information about general website traffic, including through third party websites Similar Web. (n.d.). Retrieved August 29, 2018, from https://www.similarweb.com/website/moc.buenosaires.gob.ar#overview

URUGUAY

Over the past ten years Uruguay has established itself as one of the regional leaders in the adoption of ICTs both for the provision of public services and for the internal administrative procedures of public bodies.

The Uruguay 2020 Digital Agenda included the objective that all ministries with large volumes of data should develop models for the descriptive and predictive analysis of phenomena that affect the community.[94] Although the figures are low in absolute terms, in relative terms Fig. 7 suggests a great increase in interest. It shows the increase in the number of publications on AI, which were found on ministries' websites. Although the number had been growing steadily since 2013, it rose dramatically in 2017, the year of publication of the Uruguay 2020 Digital Agenda.

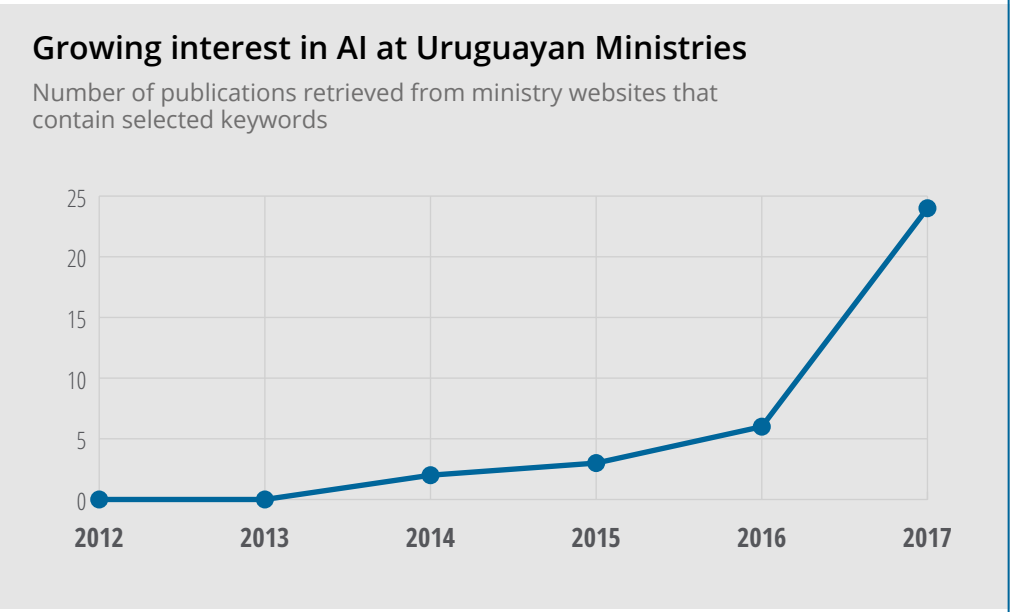


Fig. 7 — Number of publications retrieved through Google from Uruguayan Ministries' websites, which include the words "Artificial Intelligence" / "Machine Learning" / "Deep Learning"

The most striking use case we identified was a machine learning model for the prediction of crimes. The model defines and updates "hotspots" on a map, which was used to target police resources.

94 Presidencia de la República. (2017, March 17). Agenda 2020 [Web log post]. Retrieved from https://uruguaydigital.gub.uy/agenda-digital/agenda-2020

Objective:

To predict where a criminal offence will be committed[95]

Unit Responsible: Ministry of the Interior of Uruguay

Development: Private (PredPol)

Origin of the Data: Public Security Management System

Public Availability of Data: No

Selection of Variables: Type of crime, location, date and time[96]

Variables capable of triggering undue discrimination:[97] Indirectly

Model intelligibility: Black Box

Output: it identifies 150 square metre sections, in which there are high probabilities of crimes being committed

Reported Error Ratio: Unknown

Takes or Assists Decisions: Assists

Consequence: Deployment of officers to the area

Impact (according to people in charge): Crime was not reduced in absolute terms. There were [small] reductions in areas where it was implemented[98]

At the end of 2013, following a competitive tender process in which eight companies participated,[99] the Ministry of the Interior acquired the licence for *PredPol* software. Predol is the product of a joint research project between the Los Angeles Police Department (LAPD) and the University of California Los Angeles (UCLA).[100] As stated in resolutions of the Court of Auditors, the PredPol licence was initially acquired in November 2013 for a period of 12 months, for which the government paid US $123,400.[101] The licence was subsequently renewed twice: in May 2015, for a new term of 12 months at the same price,[102] and again in June 2016, for the same price but a 6 months period.[103]

Data

PredPol offered predictions based on data collected by the Ministry of the Interior, which it accessed through the Public Security Management System (SGSP), a system that consolidates all the public safety information generated by public institutions of Uruguay.

When evaluating possible biases in the datasets that feed the model, it is important to bear in mind that:

- Not all crimes are reported. Police databases are developed using *reported* crimes and some crimes are reported more than others.[104]
- The system was used to detect a particular type of crime: *rapine* (robbery).
- Police presence in an area may increase the likelihood of a crime being detected or reported.

95 Complete file available in digital annex

96 This information comes from the supplier's website (http://www.predpol.com/about/ and https://www.predpol.com/how-predictive-policing-works/). Local authorities refused a request for access to information (See Annex). Consequently, we were unable to ascertain whether the software included variables other than those mentioned on the supplier's website.

97 See footnote 54 or Annex for definition

98 Ricardo Fraiman and Pablo Martínez' introduction to the Ministry's report concludes that the programme did not reduce robberies in absolute terms. However, in those areas where it was implemented, it did cause "significant reductions" (pg 18). Alejandro Cid's chapter, supported by the methodology of differences in differences, and relying on some questionable assumptions, suggests that the reduction was between 0.4 and 3%, depending on the area used as a baseline (pg 152). See Ministry of the Interior and Inter-American Development Bank (IDB), How to avoid urban crime? The High Operational Dedication Program in the new Uruguayan police, Montevideo, November 2017. ISBN: 978-9974-8615-1-0. Available at: https://www.minterior.gub.uy/images/2017/November/Cmo-evitar-el-delito-urbano.pdf

99 PredPol achieved 100 points. The other participating companies were Quanan; Palantir; Softron; Arnaldo C. Castro; CSI Engineers; Stiler and Ciesu

100 PredPol (Op. Cit) http://www.predpol.com/

101 Court of Accounts of the Oriental Republic of Uruguay www.tcr.gub.uy/archivos/resoluciones_5486_r2013-17-1-0007104.doc or https://docs.google.com/document/d/1BSfIVjiEBVh-6gHPNjcqLey4OE1Ij7RUSveHRZ4QKlo/edit

102 Court of Accounts of the Oriental Republic of Uruguay www.tcr.gub.uy/archivos/resoluciones_12568_r2013.17.1.0007104.doc https://docs.google.com/document/d/1Gp75-SfLvNGAufp88WIT8OSXQsER0bWbJu_1snkORjA/edit?usp=sharing

103 Court of Accounts of the Oriental Republic of Uruguay, Delivered 2231/15 Folder 191, Montevideo, May 6, 2015. Available at: https://docs.google.com/document/d/1BSfIVjiEBVh-6gHPNjcqLey4OE1Ij7RUSveHRZ4QKlo/edit or www.tcr.gub.uy/archivos/resoluciones_17606_r2013-17-1-0007104.doc [Date of consultation: 04/08/2018]

104 For example financial or white collar crimes. See for example Smith IV, J. (2017, April 25). This app exposes the white collar criminals around you. *Mic*. Retrieved from https://mic.com/articles/175182/white-collar-crime-app-the-new-inquiry#.2whHEBYuP

Consequently, there are reasons to believe that the databases could be biased against marginalised groups.[105]

Far from being a public database, in 2012 the Ministry of the Interior classified all information and documentation on police activity. Although making this information freely available could affect both privacy and security, reliance on databases whose design is not subject to any type of public participation or transparency mechanism may affect the legitimacy of their use in algorithms.

Design of the model

In the context of this research, a request for access to public information was submitted. The intention was to learn about the technical aspects of the software programs used by the Ministry of the Interior in the last five years. However, the statutory deadline has expired and the ministry has yet to reply to the request.[106]

Accordingly, in addition to the use of secret databases described in the previous section, we must mention the model's design is also secret. In this way, PredPol becomes a black box, impossible for affected populations to understand why there is a more frequent police presence on their block.

Based on information publicly available through PredPol's website we could confirm the *machine learning* model relies on three variables: type of crime, location, date and time.[107] It should be noted that although *location* is not a suspect class in itself, given the problems of gentrification, particularly in Latin America, this variable may act as a proxy for the socio-economic or ethnic background of a group of people. These are suspect categories[108] prone to trigger undue discrimination.

In addition to the lack of transparency in the implementation of PredPol (legitimacy of process) and the use of variables that can trigger undue discrimination (legitimacy of result), there was a lack of public participation in the design of the tool and its implementation. Consequently, PredPol's legitimacy as a tool to manage a matter as sensitive as police resources is low. And the extended use of these tools can progressively erode the trust in police forces in general.

Decision/Interpretation of the Output

Based on the data provided, PredPol develops and updates tailored maps highlighting 150 square metre sections where its assessment suggests there are high probabilities that crimes will be committed. This information is meant to support decision-making processes of high ranking officers in charge of dispatching patrol teams.

PredPol suppliers specifically suggest that officers spend 10% of their time patrolling areas identified as high risk. According to a document published by the Ministry of Interior, once the quadrants are defined "[the officers] should remain there unless they have to attend an emergency".[109] These types of instructions place this implementation closer to the group of self-implementing algorithms and distances it from the ones which merely offer suggestions or provide information for decision-making purposes. In this case, the police force becomes the organic arm of the computer. It would be different, for example, if each commissioner were able to define where to allocate resources, and regarded the output as one of many variables to be

105 A New York Times investigation, for example, revealed that while studies show that marijuana use is equal across all racial groups, over the last three years the New York City police arrested low-income and black people on marijuana-related charges at a rate eight times higher than that of whites . Individuals from the Latino community were arrested at a rate five times higher than whites. Editorial Board. (2018, May 14). The Legacy of Stop-and-Frisk in New York's Marijuana Arrests. *New York Times*. Retrieved from https://www.nytimes.com/2018/05/14/opinion/stop-frisk-marijuana-nyc.html and Angwin, J., & Larson, J. (2016). *Machine Bias* (Rep.). Pro Publica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

106 Data on requests for access to information available in the digital annex.

107 This information comes from the supplier's website (https://www.predpol.com/how-predictive-policing-works/). Local authorities refused to comply with our request for access to information (See Annex). Consequently, we are unaware whether the model was modified at the time of implementation.

108 See, for example: Israni, E. T. (2017, October 26). When an Algorithm Helps Send You to Prison. *New York Times*. Retrieved from https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html

109 See Ministry of the Interior and Inter-American Development Bank (IDB), How to avoid urban crime? The High Operational Dedication Program in the new Uruguayan police, Montevideo, November 2017. ISBN: 978-9974-8615-1-0. Available at: https://www.minterior.gub.uy/images/2017/Noviembre/Cmo-evitar-el-delito-urbano.pdf pg. 18

considered. This example illustrates very clearly that often it is not the designers of the model, but the political authorities who define how much power is granted to an AI system.

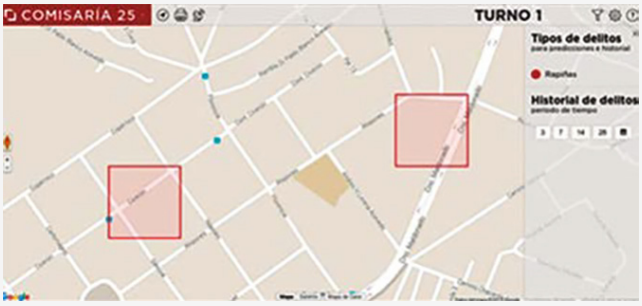


Fig. 8 — Map of Grids Created Using PredPol (Ministry of the Interior, BID, 2017)

PredPol was used between 2014 and 2017. From 2015 onwards, members of the Tactical Information Directorate (DIT) were trained in data-processing and analysis. Once trained, the Ministry of Interior carried out an evaluation: half of the police stations in Montevideo employed Predpol's forecasting and dynamic analysis tools, while the other half used the annual retrospective reporting system created by the DIT.[110]

Unlike PredPol, in the DIT reports the areas of special attention are not shown as quadrants but as street segments. According to the ministry, this allowed greater accuracy and a more efficient use of scarce human resources. In addition, the ministry considered that the street segments display was a more realistic representation of police work.[111]

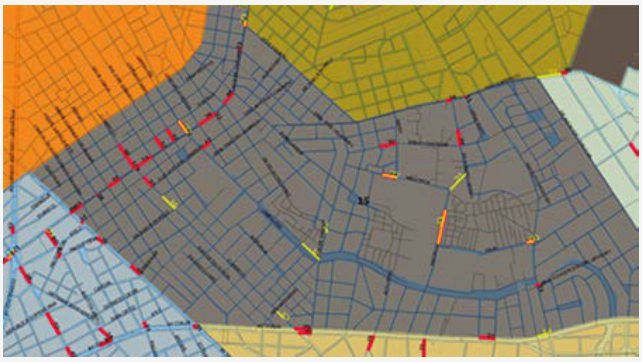


Fig. 9 — Segment Map (Ministry of the Interior, BID 2017)

The evaluation process showed that areas where PredPol was in operation did not fare better than those relying on the static system. Consequently, they decided to discontinue the use of PredPol. Currently the police personnel rely on DIT reports which they access through the Public Security Management System (SGSP).[112]

Interaction with Social and Legal Frameworks

The increasing adoption of these types of predictive models by police forces around the world has been criticised.[113] Among the critics are those who accept the theoretical bases of the model but believe that predictive police systems do not achieve their objective. At best, they displace crime hotspots from one area to another, without achieving any reduction in crime overall.[114]

There are also criticisms of the theoretical frameworks which inform the development and implementation of these tools.[115] Some argue these tools are built under the assumption that criminal behaviour occurs in a vacuum, forgetting the social and legal frameworks in which it occurs. In this way, a narrative about crime is developed, which likens it to a catastrophe that simply

110 Ministry of the Interior and Inter-American Development Bank (IDB), How to avoid urban crime? The High Operational Dedication Program in the new Uruguayan police, Montevideo, November 2017. ISBN: 978-9974-8615-1-0. Available at: https://www.minterior.gub.uy/images/2017/Noviembre/Cmo-evitar-el-delito-urbano.pdf See pg 20

111 As described in the section on the design of the algorithm, it was not possible to access any type of information on the type of variables used to design the algorithm.

112 Sosa, A., No Toquen Nada, Interview "*La comisario y la estadística policial*", Radio Del Sol, Montevideo, 06 June, 2017. Available at: http://delsol.uy/notoquennada/entrevistas/la-comisario-y-la-estadistica-policial and IDB (Op. Cit): https://www.minterior.gub.uy/images/2017/November/Cmo-evitar-el-delito-urbano.pdf . The SGSP is a broader platform through which all the information related to the police action is managed. The platform interoperates with several information sources of National Institutions and with Security and Police Agencies of MERCOSUR as well as with databases from the Ministry of the Interior. Note that by Resolution No. 5903 of July 20, 2012, the Ministry of the Interior classified as reserved all information and documentation, which involves official police activity.

113 This type of system usually focuses on "street crime" and does not include other criminal activities. In order to make this differentiation explicit, in April 2017 the Non-Governmental Organization New Inquiry presented a geo-referenced system focused exclusively on white collar crimes in New York City. Available at: https://whitecollar.thenewinquiry.com/

114 "more focused patrolling in certain places probably causes displacement, which may be temporary or functional (...) since the implementation of this new patrol strategy in 2013, the places where crime occurs have become more scattered, especially in the case of robbery. In 2006-2012 50% of the crimes were concentrated in 4% of blocks, and currently in 5.3% "See Galiani, S. & Jaitman, L. (2015): "Applied crime economics: the case of Montevideo, Uruguay" in: http://focoeconomico.org/2015/12/18/economia-del-crimen-aplicada-el-caso-de-montevideo-uruguay/ The Max Planck Institute found no conclusive evidence of a causal relationship between the implementation of these tools and a substantial reduction in thefts. See Disappointing results for predictive policing [Web log post]. (2017, September 25). Retrieved from https://digit.site36.net/2017/09/25/disappointing-results-for-predictive-policing/

115 See also Robinson, D., & Koepke, L. (2016). *Stuck in a Pattern* (Rep.). Upturn. https://www.teamupturn.org/reports/2016/stuck-in-a-pattern/

happens. Consequently, discussion about the social conditions which lead to certain crimes occurring in certain areas does not happen,[116] and the policies of inclusion required to give legitimacy to the institutions whose laws are being violated are not discussed.

A report published in 2016 by researchers from the Human Rights Data Analysis Group (HRDAG) took these criticisms further, suggesting that these models generate a feedback loop: a greater police presence in a given area makes it more likely for crime to be detected. The more crimes recorded in an area, the more likely that police forces will be deployed. According to HRDAG, this means officers will be repeatedly sent to the same areas of the city, usually those where minorities are concentrated, regardless of the true crime rate in that area.[117]

At the local level, the Human Rights organization SERPAJ expressed similar concerns about PredPol in 2017. They underlined that these programs "detect patterns in the data fed in and then repeat them in future predictions." [...] If, applied without due care, the logic is that of a vicious circle. Technology is used to legitimise the policeman's mythical *sixth sense*."[118] Basically legitimizing arbitrary police behaviour, which translates into police abuse, and is often the product of racism or classism.

Summary of the Findings:

- **Effectiveness of the model:** No information is available about PredPol's error rate. Experts from different fields have questioned the theoretical basis for these types of models. The replacement of PredPol by a basic statistical tool suggests that its effectiveness was lower than expected.
- **Implementation effectiveness:** a reduction in crimes was detected in areas in which it was implemented but not in overall terms. It has been argued that the tool simply leads to the displacement of crime hotspots.
- **Legitimacy (process):** The data is not public, and PredPol is a proprietary black box, whose outputs are not explainable.
- **Legitimacy (results):** There is a risk of discrimination. Local and international organisations have argued that tools like PredPol tend to replicate the biases of training data and the historical power dynamics between law enforcement and minority or underprivileged populations, and that they are used to justify police presence in marginalised areas.

In a democracy, police forces must be at the service of the people. Sending police forces to areas under the expectation that they will find criminals creates an environment in which abuses of power are likely to occur. The fact that these decisions are based on non-transparent systems undermines the legitimacy of the police force. Government departments in charge of financing infrastructure and protecting rights should engage the communities affected in the design of any tool concerned with the exercise of force.

116 Eubanks, V. (n.d.). We created poverty. Algorithms won't make that go away. *The Guardian*. Retrieved May 13, 2018, from https://www.theguardian.com/commentisfree/2018/may/13/we-created-poverty-algorithms-wont-make-that-go-away

117 Human Rights Data Analysis Group (HRDAG), Predictive Policing Reinforces Police Bias, 10 October 2016. Available at: https://hrdag.org/2016/10/10/predictive-policing-reinforces-police-bias/ [Date consulted: 04/07/2018]. See also, paragraph 16 of the Toronto Declaration (Op. Cit) https://www.amnesty.org/download/Documents/POL3084472018ENGLISH.PDF

118 Perz, L. & Samudio, T., "Same actions in modern disguises: Program of High Operative Dedication", in Human Rights in Uruguay. 2017 Report, Peace and Justice Service - SERPAJ Uruguay, Montevideo, December 2017. Available at: http://www.serpaj.org.uy/serpaj/index.php/documentos-de-interes/file/55-info2017

03

RECOMMENDATIONS

The reliance on algorithms and AI systems by governments in Latin America is beginning to gain traction.

Officials interested in leveraging its potential face a scenario where there are many vendors but few mechanisms to evaluate the effectiveness of the tools or the legitimacy of the proposed implementation. Consequently, in a context of growing enthusiasm for the adoption of these tools, there is a risk that systems that are useless, inefficient or capable of causing harm to the public will be purchased or built. Thus, the development of mechanisms to assess the effectiveness of different tools, the establishment of guidelines to increase the legitimacy of their implementation, minimise risks and ensure that they are used for the benefit of the target populations, become matters of urgency.

We have three core recommendations for government officials:

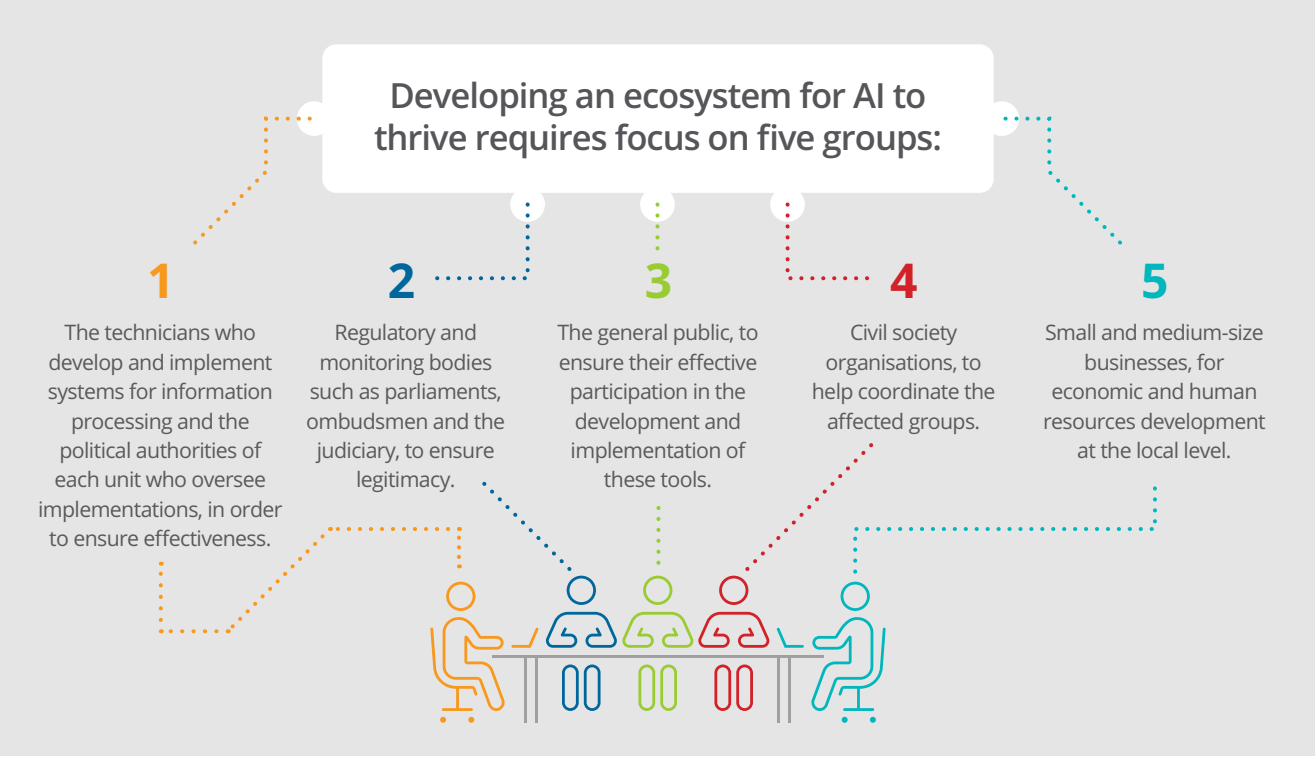
1. Develop infrastructure and government expertise for the use of AI techniques

The infrastructure required for governments to use algorithms effectively is currently absent or limited in Latin America. In the course of this research, we found that the Province of Salta in Argentina, lacking necessary infrastructure, stores sensitive data in private servers located abroad, beyond its territorial jurisdiction. At the same time, the Uruguayan government relied on a private black box system to allocate police resources for approximately 30 months.

These examples not only illustrate risks to the privacy of the population and problems regarding sovereignty, but also missed opportunities to promote strategic investments in the sector. The process of modernising the state should be leveraged to create an ecosystem that favours the development of new technologies.

Developing tools that are sensitive to the local context[119] and capable of maximising their benefits and minimising risks of AI will require greater interaction between specialists from a variety of disciplines, including data science, public policy, administrative management, and human rights.[120] Particular effort should be made towards the effective inclusion of women, a systematically excluded group that is particularly at risk.[121]

Fig. 10 — Representation of key stakeholders



119 See paragraph 20 of the Toronto Declaration (Op. Cit) https://www.amnesty.org/download/Documents/POL3084472018ENGLISH.PDF

120 The iterative model commonly known as "policy sandbox", which facilitates experimentation without generating binding results, may be useful in the current context. See, FCA regulatory sandbox: Lessons learned from the first year in operation [Web log post]. (2017, November 1). Retrieved from https://www.lexology.com/library/detail.aspx?g=22212737-374e-43bc-9cba-76b8777eb017

121 Avila, R., Brandusescu, A., Ortiz Freuler, J., & Thakur, D. (2018). *Artificial Intelligence: Open questions about gender inclusion* (Rep.). W20 / Web Foundation. doi:http://webfoundation.org/docs/2018/06/AI-Gender.pdf

2. Ensure transparency, public participation and accountability in the development and implementation of AI techniques

Public officials need a set of guiding principles as they work towards the adoption of tools that have the potential to perpetuate and reinforce inequalities and existing social biases. The Toronto Declaration on AI can serve as a useful compass.[122]

The case studies show that officials often focus most of their energy on developing the predictive capacity of statistical models, but do not take into account the aspects necessary to legitimise its implementation.

Ensuring both the legitimate implementation of these tools will require the development of mechanisms of transparency, public participation, and accountability at each stage of the process.

The following examples illustrate the types of practices public officials should consider exploring as part of any implementation of AI or algorithmic systems.

	Transparency	Public engagement[123]	Accountability
DATA COLLECTION	Publish metadata for each database, including: a. Purpose b. Potential use cases c. Known reuse cases d. Point of contact e. Methodology f. Ethical and legal considerations	Public engagement to determine: a. the data to be collected b. the collection process c. the definition of the relevant tags and labels.	Conduct annual public hearings to: a. Present the impact of the participative processes b. Discuss use cases of datasets and their possible shortcomings
MODEL SETUP	Publish a file, including: a. Contact of developers and political authority b. Purpose c. Sources of data d. Method e. Percentage of errors (eg. False Positives) f. Impact assessment g. Ethical and regulatory considerations h. Contracts	Public engagement to determine: a. The definition of the objectives pursued b. the inputs to be included c. The level of acceptable complexity and risk	Facilitate and promote third party audits a. Publish results of preliminary and periodic internal audits b. Conduct annual public hearings to discuss results and risk mitigation actions c. Disclose variation in error rates and problematic feedback loops
EXECUTION / INTERPRETATION OF OUTPUT	Publish the administrative protocols that define how the output of an AI tool is leveraged, including the level of deference to the AI tools.	Public engagement to define the protocols to be followed from an output and the degree of deference to the AI tools.	Conduct yearly assessments regarding how administrative protocols that surround the output have been complied with, and present them in town hall meetings in areas where particularly vulnerable populations affected by its implementation are based.
SOCIO-LEGAL FRAMEWORKS	Publish a document that outlines a. information regarding recourse mechanisms b. alternative safety-nets or services (if algorithm affects access to service	Public engagement to: a. design systems of remedy and justice for individuals and communities who may be adversely affected by these systems b. document public perceptions regarding the problem, and how it should be tackled	Publish: a. surveys documenting the pre/post implementation shifts in perception of the problem, and how it's being tackled b. Statistics regarding use of alternative safety-nets.

122 Toronto Declaration (Op. Cit) Linked above

123 Particular attention should be placed on engaging with typically underrepresented groups such as women, indigenous communities, and people with disabilities.

3. Establish criteria for the development of impact assessments to be considered before implementing automated systems in public decision-making

At the moment there is reason to expect that prefabricated models will be adopted for rapid deployment. As this may entail serious risks, the development of criteria to allow government officials to differentiate potentially problematic models from ones which are less so, is a matter of urgency.

To improve and develop the effective implementation of these mechanisms, the field would benefit from a multi-stakeholder forum that includes technical and political actors as well as representatives of academia and civil society.

NEXT STEPS

In order to improve the use of these technologies by governments, we recommend as immediate steps:

1. **Creating and promoting formal and informal spaces** — nationally and regionally — which facilitate fluid dialogue on the implementation of new technologies for governance, including governments, the private sector, academia and civil society.
2. **Documenting and developing good practices** in terms of transparency, public participation and accountability in the implementation of AI systems.
3. **Establishing standardised checklists** to assist developers and implementers in risk mitigation.

ANNEX

Digital Annex

- Glossary and Methodology for the Creation of Records
- Google Search Results
- Information Access Requests
- Interviewees

Available at the following link: http://bit.ly/ai-annex

Annex II

Methodology For The Identification of Cases

ILDA performed the following tasks based on the Research Manual developed by the Web Foundation[124].

Documentary research through Boolean searches using the Google search engine advanced function.

- **Objective:** Identify possible case studies, specialists and existing research.
- **Implementation:** The URLs of 77 government agency websites were identified: 28 from Argentina, 27 from Uruguay and 22 from Colombia[125]. At the same time, searches were performed on civil society organizations', academic institutions' and news websites in each country (40 URLs). In total 45 searches were performed using differing parameters on each subset of URLs.[126]

Semi-structured interviews.

- **Objective:** Find examples, trends and recommendations on the implementation of algorithm systems. Identify other actors through a snowballing approach. Investigate in-depth the technical specifications of particular examples.
- **Implementation:** Six interviews were conducted: five of them with officials of the central and local governments in Argentina and one with the Under-Secretary of the Ministry of Industry, Energy and Mining of Uruguay.[127] Attempts to arrange interviews with Colombian officials were unsuccessful.

Requests for access to public information.

- **Objective:** verify availability of data and access complementary information.
- **Implementation:** There were 18 requests for access to information. Eight in Argentina, four in Colombia, and six in Uruguay. A response was obtained in 15 cases but only a fraction of them included the information requested. In particular, the security agencies were extremely reluctant to release any type of information.[128] In the case of the Province of Salta, as it does not have a digital portal for information requests (submissions must be made on paper), applications were sent by mail and through the enquiry section of the Ministry of Early Childhood's website. These were not answered. Information was nevertheless provided after informal channels of communications were set up with government officials, who have been very open to cooperation.

124 See the Investigation Manual in https://docs.google.com/document/d/10f75NIyRMSXePr8VvTxpXTzvgRJzhbA7jvjK4S2Cdso/edit?ts=5aae376c#
125 It should be noted that the number of Ministries varies from country to country, affecting the number of URLs. In Colombia there are 16, in Argentina 21, and in Uruguay 13.
126 Search terms, URLs and results available in the Annex.
127 Names and positions available in the Annex
128 The answers obtained are documented in the Annex

WORLD WIDE WEB
FOUNDATION

World Wide Web Foundation
1110 Vermont Ave NW, Suite 500
Washington, DC 20005, USA

www.webfoundation.org

Twitter: @webfoundation