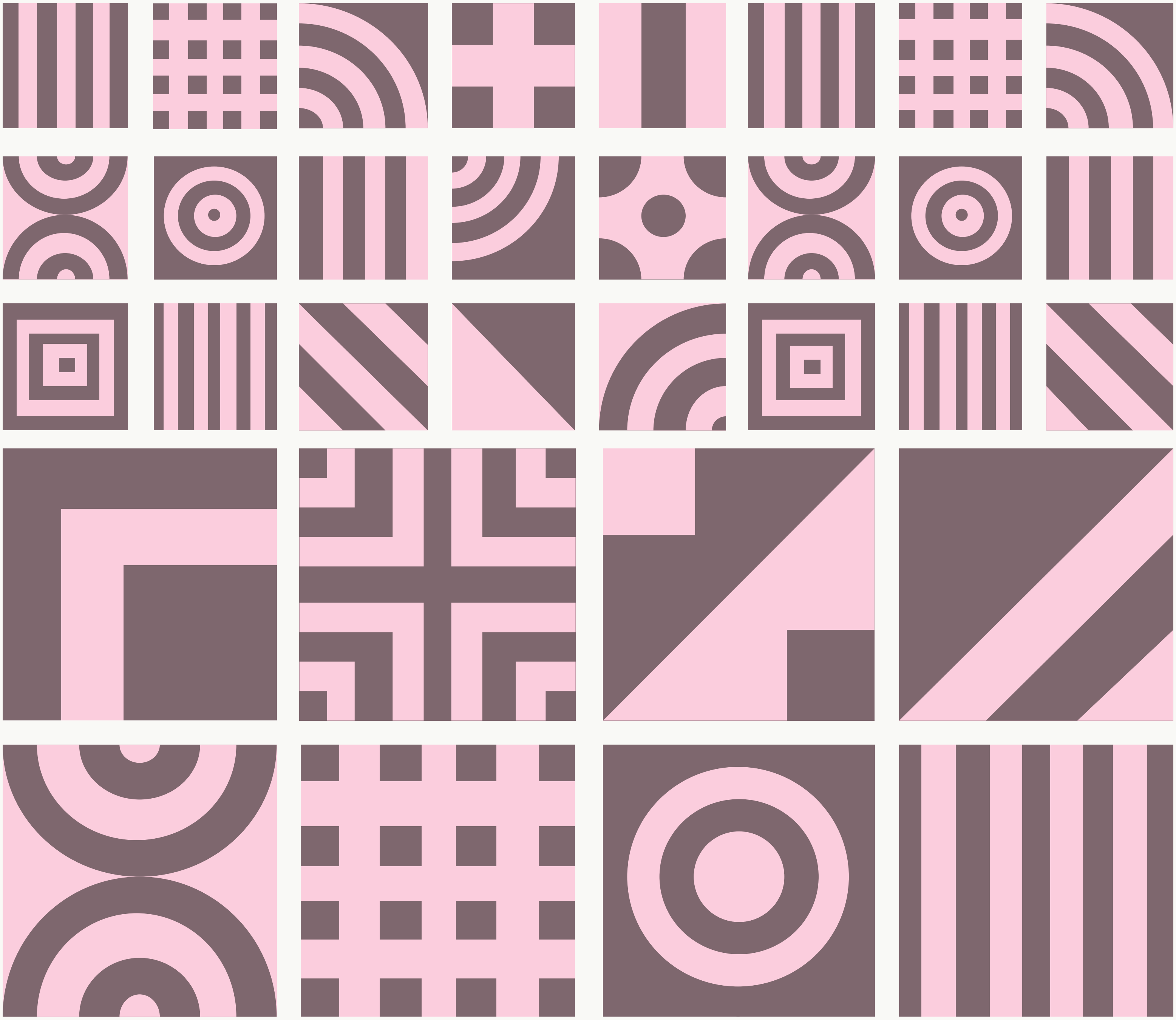

Artificial Intelligence in Government: Collective study on experiences and risks to human rights

DERECHOS
DIGITALES
América Latina

ARTIFICIAL INTELLIGENCE IN THE STATE: A COLLECTIVE STUDY ON EXPERIENCES AND RISKS TO HUMAN RIGHTS

This report was produced by Derechos Digitales, an independent, non-profit organization founded in 2005. Its mission is to defend, promote, and advance fundamental rights in digital environments across Latin America. For more information about this project, visit https://ia.derechosdigitales.org/.

Overall Supervision: Jamila Venturini and Juan Carlos Lara

Author: Juan Manuel García

Review: Paloma Lara Castro and Juan Carlos Lara

Translation into English and Portuguese: Urgas Traductoras

Design: Alter Studio

Suggested citation: García, J.M. (2024) Artificial Intelligence in the State: A Collective Study on Experiences and Risks to Human Rights. Derechos Digitales.

November, 2024

CONTENTS

ARTIFICIAL INTELLIGENCE IN GOVERNMENT: COLLECTIVE STUDY ON EXPERIENCES AND RISKS TO HUMAN RIGHTS

INTRODUCTION

I. ARTIFICIAL INTELLIGENCE AS A SOCIO-TECHNICAL ARTIFACT

Artificial Intelligence (AI) is already incorporated into the range of tools available to the public sector for planning and implementing public policies. The commonly held perception of these technologies, focused mainly on their usefulness for making processes more efficient, has a decisive influence when determining whether or not to apply them in the public sector. In that sense, these tools are implemented for making decisions about, for example, which segment of the population to prioritize over others for a given policy, or which people may, potentially, need more government assistance than others in a particular situation. In this regard, concerns arise about the use of these technologies and the implications for the exercise of fundamental rights. This situation is frequent, regardless of the specific techniques that come under the conceptual umbrella of AI, such as natural language processing (NDP), machine learning (ML), risk prediction systems and automated decision-making (ADM).

With the aim of deepening understanding of this issue and contributing evidence to these processes, since 2019, Derechos Digitales has been analyzing the problem under the Artificial Intelligence and Inclusion programmatic area. Collaborating with researchers from various Latin American countries, we have examined the areas where States are employing these technologies, the specific characteristics of their use, and the potential risks they pose to human rights. This research has taken the form of case studies, developed using a shared methodology that incorporates multiple dimensions of analysis, with a particular focus on their potential impact on individual rights. These are exploratory studies designed to provide evidence on an activity that remains emerging but is rapidly accelerating.

Data serves as the core component for the technical development of AI-based technologies. However, Latin American governments have difficulty maintaining robust practices for data use, management and storage, due in part to problems such as database fragmentation, heterogeneity in perspectives on data among different government agencies, diversity in computing systems, and the lack of a shared language (Fundar, 2024). This poses an additional challenge when considering that behind the data used by AI in the context of government administration are individuals seeking employment, those in need who may qualify for state assistance, students attending public institutions, or users of essential government services, among other scenarios.

The rights of these individuals, and of all people, are safeguarded by international human rights frameworks and local legislation, which impose obligations on and establish limits for the government in the processing of data. These obligations and limits differ depending on whether the data are classified as personal or public, as do the conditions for ensuring equitable use in the provision of resources and services. This article, drawing on evidence gathered through the Artificial Intelligence and Inclusion program, seeks to analyze how the State employs this type of technology in fulfilling its responsibilities toward citizens.

As a general analytical framework, it is essential to acknowledge that the production of data processed and analyzed by AI-based technologies is not neutral. As Buschmann (2021, p. 41) highlights in their analysis of the Urban Crime Prediction System implemented in Chile, every piece of data is part of a social production chain that can embed contextual biases and perspectives. This means that databases may include data reflecting irregular or inaccurate situations, potentially resulting in biased systems that perpetuate discriminatory practices.

Therefore, understanding the potential impact on fundamental rights requires more than merely analyzing the technology itself — such as examining the algorithms or automation processes implemented — since these do not operate in isolation or a vacuum. Policies leveraging AI as a tool exist within specific social and political contexts, shaped by diverse demographic compositions, unique legal frameworks, democratic characteristics linked to historical processes, and government actions responding to the specific circumstances of each territory. Consequently, a central focus of this document is to explore how responses to public issues are developed within a socio-technical network, involving both human and non-human agents, within a defined historical and institutional context (Velasco & Venturini, 2021, p. 11).

Based on these premises, the article is divided into three parts. First, a summary of the applicable legal frameworks for guaranteeing individuals' fundamental rights, particularly in the context of AI-based technology usage by the State. Next, an overview of the ten specific applications of AI examined within the framework of the Artificial Intelligence and Inclusion project. Finally, an analysis of how these applications pose risks to human rights, considering both legal frameworks and the intrinsic characteristics of their use.

II. METHODOLOGY FOR ANALYZING IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE

As mentioned above, given the socio-technical nature of the development and deployment of these technologies, understanding their potential impact on rights requires knowing the context of their implementation, the population they intend to affect, the legal framework applicable in that jurisdiction, and the government's data use practices, among other contextual considerations. Along these lines, the ten case studies have been analyzed using a multi-dimensional methodology that seeks to document the range of factors involved in the potential impact on individual rights.

This methodology, developed by Derechos Digitales for this project, consists of five dimensions:

1. National implementation context: attempting to understand the sociodemographic and technological characteristics of the country where the AI systems are implemented. Factors evaluated include population distribution, access to technology, and the socioeconomic conditions influencing the effective use of AI by the State.

2. Regulatory and institutional context: examining the legal and institutional framework regulating government use of AI. This includes the existence of laws, standards, specialized institutions and oversight mechanisms designed to guarantee that technological development and use respect human rights.

3. Data infrastructure: analyzing the technological and data resources that underpin government AI systems, such as the quality of the datasets, their interoperability and the existence of mechanisms for protecting the information's privacy and security. One of the key points focuses on the data's characteristics, mainly in terms of any personal data contained in the databases being analyzed and processed.

4. Decision-making process: exploring how AI systems are integrated into government decision-making processes, considering the participation of human actors, the transparency of the criteria used and the routes of accountability.

5. Technological design: focusing on the technical and operational features of the AI systems implemented by States, including their objectives and capabilities, as well as possible biases in their development. This dimension seeks to analyze whether the systems implemented are designed in alignment with the public policy's needs and objectives.

The studies were developed by researchers from each country where the cases occur. Some worked as individuals and others as part of a team, and both those representing an organization and those with no institutional affiliation were included. The researchers come from academic, civil society and even activist spaces. This diversity illustrates the broad range of people interested and involved in the agenda promoting a more democratic use of technology that is respectful of human rights.

While the approaches reflect diverse perspectives, the difficulties in implementing this methodology were common to all the studies. First, there was a lack of information available for identifying the cases. A preliminary step in selecting the studies to be conducted was a mapping exercise developed by the Derechos Digitales team, searching public available information for use cases in Latin America. This process involved search engines on the open web; a search of specific publications from different government entities; and a review of specialized reports from international organizations and programs financing use of AI in government. The information found was, in most cases, incomplete and fragmentary. This presented the first challenge, given the dependence on information provided by official sources in the different governments.

Along these lines, the second issue stems from difficulties in accessing information from official sources. The researchers found that access to information requests were essential resources for analyzing the third, fourth and fifth dimensions. However, responses to the requests were often incomplete or delayed, hampering development of the studies. Further, some government representatives were reluctant to participate in the research as official sources.

The last issue to mention stems from the short-lived nature of some of the policies found. During the mapping stage, multiple policies were detected that were using or said they used AI as a central tool. However, the search for more information revealed that many of them were discontinued or interrupted, for two main reasons. The first is the lack of financing after a pilot stage, while the second is the result of certain changes in government administration that took place in recent years, leading to changes in management plans and government priorities.

1. APPLICABLE LEGAL FRAMEWORKS

Government implementation of AI-based technologies poses significant challenges in terms of protecting fundamental rights, so referring to applicable legal frameworks is essential. This section first covers the obligations arising from international human rights treaties, such as the American Convention on Human Rights and the Protocol of San Salvador, and their relationship to the principles of legality, necessity and proportionality in AI use. In addition, specific legislation on personal data protection and access to information in different countries around the region was examined, highlighting advances, shortcomings and their intersection with bills to regulate AI. The goal is to offer a comprehensive panorama of the legal and regulatory obligations that States must comply with to guarantee the protection of rights in light of the growing use of these technologies.

International human rights treaties

As mentioned at the start, one of the questions that the Artificial Intelligence and Inclusion project sought to answer involves the extent to which governments take into account the criteria of legality, necessity and proportionality when implementing this kind of policy. This question attempts to understand how governments safeguard the provisions established by the standards of the Inter-American human rights system, specifically the obligations arising from the American Convention on Human Rights (ACHR), the Additional Protocol to the American Convention on Human Rights in the Area of Economic, Social and Cultural Rights (Protocol of San Salvador) and rulings of the Inter-American Court of Human Rights.

The impact of AI use on human rights has already been recognized internationally in several resolutions. The recent United Nations (UN) Human Rights Council Resolution A/HRC/RES/48/4 on the right to privacy in the digital age has identified some of the risks to the exercise of human rights posed by the adoption of artificial intelligence, which occur “in particular when [AI is] employed for identification, tracking, profiling, facial recognition, behavioral prediction or the scoring of individuals.” The resolution establishes that States must respect human rights in the implementation of these systems and adopt both preventive measures and effective resources to confront violations and abuses of the right to privacy, particularly in the case of women, children and persons in vulnerable situations.

Thus, the obligations established by international treaties are specific in terms of States' responsibilities when they adopt AI-based technologies. To sum up these obligations, we will use the analysis by Alimonti and de Alcântara (2024), which offers a clear picture in a report recently published by the Electronic Frontier Foundation, where the authors define States' duties in incorporating AI-based technologies. Following are a few of the implications they mention:

- **Protection of human rights:** States must ensure that the use of AI systems does not violate human rights, in compliance with the ACHR. In addition, policymaking that incorporates AI must follow a human rights approach, guided by the principles of the ACHR and Protocol of San Salvador (Article 2).

- **Social participation and transparency:** It is essential for decision-making processes involving AI to be transparent and enable meaningful participation, in alignment with the right to information and participation enshrined in the ACHR.

- **Prior assessment of impact on human rights:** Before adopting an AI system, States must conduct an exhaustive evaluation that considers the system's suitability, in compliance with the obligations set forth in the ACHR and the Protocol of San Salvador. Likewise, the UN High Commissioner for Human Rights report emphasizes the need to guarantee that these tools fully comply with international human rights law, recommending a moratorium on, and even a ban of, technologies that cannot be used compatibly with these standards (UNHCR, 2021). In addition, States must establish appropriate mechanisms for overseeing the use of AI in the public sector, in harmony with the accountability established in the ACHR.

- **Protection of groups in vulnerable situations:** Special attention must be paid to the differentiated impact that AI use may have on groups that have historically been discriminated against, in compliance with the ACHR principles of equality and non-discrimination.

- **Guarantees to ensure compliance with the principle of non-discrimination:** Policies that use AI must be designed to prevent discrimination, in alignment with Article 1.1 of the ACHR, which establishes the obligation to respect and protect rights.

Finally, given the place of private-sector initiatives in the AI agenda, it is important to emphasize that, although AI tools may be developed by private entities, States are still responsible for ensuring that their use respects the above-mentioned obligations, in accordance with the ACHR and Protocol of San Salvador. This responsibility applies not only to direct use of these tools, but also to the content of procurements, licenses or any kind of formal agreement with providers. Likewise, States must ensure that the data gathered, stored and processed in these systems meets appropriate protection and security standards.

For their part, private companies, as key actors in the development and deployment of these technologies, have specific responsibilities in safeguarding human rights. According to the UN Guiding Principles on Business and Human Rights, companies must identify, prevent, mitigate and as needed, remedy the negative impact that their operations may have on human rights, acting responsibly at all stages of their AI tools' lifecycles (UNHCR, 2011).

This summary aims to highlight the existence of an international legal framework that protects citizens from the potentially harmful use of technology. These international frameworks, which are legally binding for States, are particularly relevant in contrast to other international instruments that take on special significance in the context of AI. Here, we refer to recommendations and ethical principles on AI, which, while they may provide a framework for action, do not, in themselves, constitute binding obligations for States.

Personal data protection and access to information

To comply with international human rights frameworks, each country must develop specific legislation that protects its citizens' rights. This includes clearly and effectively incorporating into domestic legislation the rights and obligations recognized in international treaties. Likewise, it is essential for national regulations, such as Personal Data Protection (PDP) and Access to Public Information laws, to be consistent with international provisions, guaranteeing a protection framework aligned with human rights standards. In this sense, although the situation in the region is diverse and evolving, these frameworks represent a starting point, rather than a final objective, for ensuring the protection of fundamental rights.

Until recently, PDP legislation in Chile was based on the Protection of Privacy Act, enacted in 1999 as Law 19.628, which regulated the processing of data of a personal nature in registries or databases. Despite being one of the first data protection laws in Latin America, as noted by Valderrama (2021), it was criticized for how quickly it became outdated and for its inability to adequately protect people from mishandling of their data by third parties. However, the Chilean Congress, in August 2024, approved a new set of articles for the Personal Data Protection Law, providing a specific framework for data processing and creating the Personal Data Protection Agency. This is an advance in terms of the regulatory situation analyzed in the cases years ago.

Brazil also has new regulations. The General Data Protection Law (GDPL), approved in 2018 and in force since 2020, establishes a legal framework for personal data protection. It shares important similarities with the European Union's General Data Protection Regulation (GDPR), as does the new Chilean legislation. As analyzed by Cardoso et al. (2021), it applies to any personal data processing done in Brazil and defines personal data as information related to identified or identifiable individuals, including sensitive data. The law created the National Data Protection Authority (ANPD), charged with overseeing enforcement of the law and applying penalties for noncompliance.

Like Chile and Brazil with their recent legislation,[1] Argentina has a bill[2] drafted by the Access to Public Information Agency with participation from businesses, public institutions, civil society organizations and academia. The motivation behind its creation was to update the existing regulations, Law 25.326, approved in 2000, nearly a quarter century ago. However, the bill has been stalled in committee since August 2023, with no plenary session debate scheduled.

At the other end of the spectrum, as noted by Sequera and Cuevas (2024), Paraguay has no personal data protection law. The main applicable framework is Law 6534/2020, which refers specifically to personal credit data protection but does not comprehensively address personal data protection in general. It is worth highlighting the role of the Coalition for Personal Data Protection, in Paraguay, which is a group that has been actively working since 2016 to promote the creation of a comprehensive legal framework to regulate personal data processing in the country. In 2021, this coalition drafted a bill that has yet to be taken up in parliament, although in 2023 the bill was included on the Congressional agenda in four different sessions (TEDIC, 2024).

[1] As of this report's concluding date, the new Chilean law is about to be passed and, with that, the two-year vacancy clock from publication to being fully in force started.

[2] For more information (in Spanish): https://www.argentina.gob.ar/aaip/datospersonales/proyecto-ley-datos-personales

Thus, as we can see, the situation of personal data protection in the region is far from homogeneous. This is significant when we consider the complementary nature of this legislation with some bills specific to regulating AI currently under discussion in countries around the region. This is the case, for example, with the bills currently under debate in Chile and Brazil, which incorporate a risk framework similar to that of the AI Act passed by the European Parliament. It is worth noting, as a reference, that this legislation, which complements the GDPR, would classify some of the analyzed cases as **"high-risk systems."** This category includes, for instance, automated systems for accessing government-provided benefits, as well as those used for employment policy management, security policy management, and the administration of justice.[3]

In addition, as mentioned in the methodology section, the study highlights the need for governments in the region to improve their access to information practices regarding AI use and its role in the performance of public functions. This applies whether the information is disclosed proactively or made available passively, i.e., in response to a formally submitted request.

Regulations on access to public information in countries such as Chile, Argentina, Uruguay, and Mexico provide a framework for both forms of access to public information. While no specific regulations govern the provision of information specifically related to artificial intelligence, in principle, a separate framework should not be necessary, as States are already obligated to ensure the dissemination of this information under general transparency and public access to information laws. Instead, it is the implementation of these frameworks that requires review to enhance public transparency regarding AI implementation.

2. USES OF ARTIFICIAL INTELLIGENCE IN THE PUBLIC SECTOR

In this section, we analyze how different States in the region use AI-based technologies as a tool for implementing public policies, based on the evidence generated by the Artificial Intelligence and Inclusion program, as previously mentioned. Specifically, it seeks to answer the following questions: How are governments in Latin America implementing artificial intelligence, and what are its impacts on development, inclusion, and human rights? Additionally, how do they consider the principles of legality, necessity, and proportionality?

[3] European Parliament AI Law, Annex 3. Available at https://artificialintelligenceact.eu/annex/3/

The analyzed cases illustrate the use of AI in sensitive areas of public administration, including employment, social protection, public safety, education, and citizen services, as well as in the administration of justice. We will provide details on each of these cases. Each case description will be accompanied by explanatory text boxes outlining the technical characteristics of the technologies used, the databases involved, or the tech companies associated, whenever relevant and useful for comprehension.

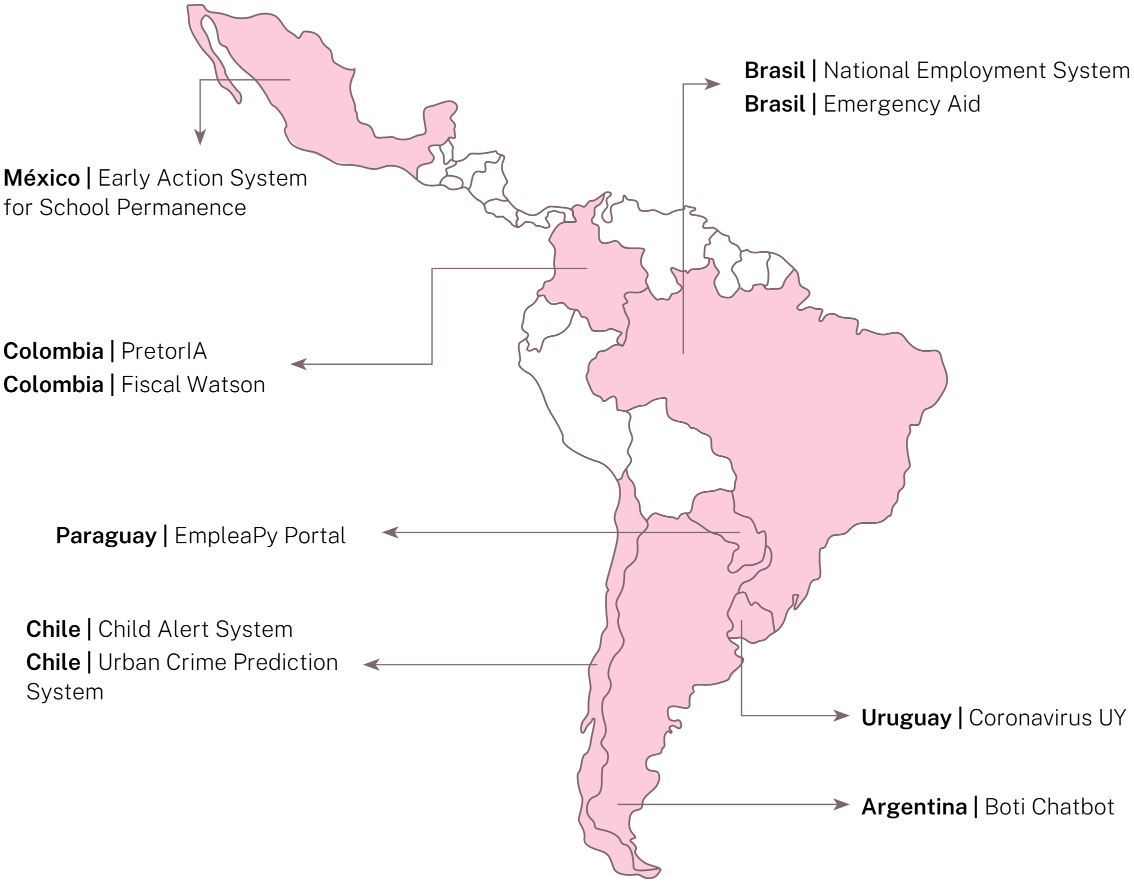
Brasil | National Employment System
Brasil | Emergency Aid
México | Early Action System for School Permanence
Colombia | PretorIA
Colombia | Fiscal Watson
Paraguay | EmpleaPy Portal
Chile | Child Alert System
Chile | Urban Crime Prediction System
Uruguay | Coronavirus UY
Argentina | Boti Chatbot

Image 1. Regional distribution of cases analyzed

EMPLOYMENT

In the area of employment, two cases were analyzed: the incorporation of AI in the framework of the National Employment System (SINE) in Brazil, a labor policy active since 1975; and the automation of processes in the context of the EmpleaPy program, managed by Paraguay's Ministry of Labor, Employment and Social Security (MTESS). Both cases involve job intermediation policies, where the goal is to facilitate connections between job seekers and companies or organizations that need to hire workers.

Dataprev, Brazil's Technology and Social Security Information Company, was created by Law 6.125 in 1974. It is a public enterprise linked to the Ministry of Economy, responsible for management of the Brazilian Social Database. It provides technological tools for the implementation of the Brazilian government's social policies.

In 2019, the **National Employment System** experienced an important transformation with the introduction of the "New SINE". This restructuring process attempted to modernize the system by implementing technological tools, such as using AI for match analysis between those seeking and those offering employment, the use and importance of digital data with job market information, and ties to the private sector (Bruno et al., 2021, p. 10).

Specifically, the implementation of AI tools falls under the SINE Digital Transformation Plan, launched in 2019 by the Secretariat for Public Employment Policies (SPPE). This plan stems from the New SINE initiative and, above all, the broader partnership between the Brazilian Federal Government and Microsoft (Bruno et al., 2021, p. 14). The project includes the use of Microsoft-provided AI tools by the Brazilian government for both employment and sustainability sectors.

In the employment sector, Microsoft's proposal involves using AI to facilitate workforce intermediation on SINE's Job Openings Portal (Portal de Vagas) and the Emplea Brasil portal. It also includes worker training through the "Workers' School 4.0," a distance-learning platform developed by the Ministry of the Economy's Special Secretariat for Productivity, Employment, and Competitiveness (SEPEC/ME), in partnership with the Brazilian Agency for Industrial Development (ABDI). This platform features Microsoft courses offered via the Microsoft Community Training tool (Bruno et al., 2021, p. 14).

In addition, the state-owned company Dataprev leverages its technological infrastructure to grant private companies access to anonymized data on workers registered in SINE. This is carried out under the Open SINE project, part of the aforementioned modernization process, which aims to open the SINE worker database to private companies and other institutions operating in the labor intermediation sector (Bruno et al., 2021, p. 12).

Microsoft's collaboration with the Brazilian government stems from a Technical Cooperation Agreement. According to the researchers, this system will rely on data, digital technology, and artificial intelligence, supported by Microsoft Dynamics, PowerBI Premium tools and licenses, and Azure cloud AI solutions. The work plan indicates that it will use the Microsoft Dynamics 365 Customer Insights module for data unification and processing, with the expected outcome being a better understanding of both workers and job openings (Bruno et al., 2021, p. 49).

Both SINE and EmpleaPy use semantic matching techniques to align those offering jobs with those seeking them. This is a computational method that identifies semantically related information. Given two graph-based structures — for example, classifications, taxonomic databases, XML schemas, or ontologies — matching is an operation that locates nodes in each structure that are semantically equivalent.

Likewise, the **EmpleaPy** case in Paraguay seeks to automate processes for connecting employers and job candidates. The initial version, known as ParaEmpleo, was created by the Swiss company Janzz Technology. Later, the Ministry of Labor, Employment and Social Security created a new version, EmpleaPy, which intends to incorporate automated decision-making. This is the version in force as of the study date (Sequera & Cuevas, 2024).

With regard to the updates and improvements made by the Ministry of Labor, Employment, and Social Security (MTESS) to Janzz Technology's software, the study authors found — during an interview with the Ministry's technical team — that the new version was developed entirely in-house from scratch, using government funding (Sequera & Cuevas, 2024, p. 16). This version reused Janzz Technology's generic data-processing approach but was adapted to meet specific needs and incorporate new functionalities.

In turn, technical sources interviewed by the researchers stated that the EmpleaPy platform does not yet use artificial intelligence. Instead, it employs a complex, automated approach for certain algorithmic processes, which helps identify users and manage both individual and company profiles. This information is then used to group them and make suggestions based on shared keywords. They also noted that the grouping process itself is done manually, and that the only fully automated part is querying and validating user profiles (Sequera & Cuevas, 2024, p. 16).

SOCIAL PROTECTION

In this area of public service, three cases were analyzed: the “Emergency Aid” program in Brazil; the “Coronavirus UY” app in Uruguay; and finally, the “Child Alert System” in Chile.

CadÚnico, established in 2001, is a database used to identify and classify the socioeconomic status of low-income Brazilian families. It is employed by more than 30 public policies in Brazil and serves as the primary tool for selecting low-income families for programs that form part of Federal Social Assistance.

Emergency Aid (EA) is an income transfer policy aimed at easing the economic and social impacts of the COVID-19 pandemic, enabling vulnerable populations to maintain access to consumer goods, especially food (Tavares et al., 2022).

Emergency Aid was automatically granted to both beneficiaries of the Bolsa Família Program[4] and people registered in the CadÚnico database who met the program’s eligibility criteria. In terms of public policies already established in the country, EA leveraged the existing structure of income transfer programs, like the Bolsa Família Program, to reach a new population that was not benefiting from any social policy (known as the ExtraCad population). In addition, new measures and technologies were implemented (Tavares et al., 2022, p. 14).

Emergency Aid involves an intense flow of data through all stages of the program. Beneficiary selection is automated and handled by the Dataprev enterprise mentioned above, which cross-references multiple databases, from different Government agencies, with CadÚnico data and with the requirements for granting the benefit to the ExtraCad population, all done via the Emergency Aid app (Tavares et al., 2022, p. 19).

The **Coronavirus UY program**, managed by Uruguay’s Ministry of Public Health (MSP), served as an information management mechanism to address the COVID-19 pandemic. Primarily developed by the Agency for E-Government, Information, and Knowledge Society (AGESIC) in collaboration with both public and private actors, the program is an information system whose key components include a mobile phone application (Yael, 2021, p. 5).

[4] The Bolsa Familia Program, implemented by the Federal Government, is the largest income transfer program in Brazil. In addition to providing income to families living in poverty, it integrates public policies to expand access to fundamental rights like health, education and social aid, coordinating complementary actions in areas such as sports, science and employment to overcome poverty. For more information (in Portuguese): https://www.gov.br/mds/pt-br/acoes-e-programas/bolsa-familia

A Predictive Risk Model (PRM) is a tool that uses established patterns in databases to automatically generate a probability (or a risk score) that a specific event will happen to a person in the future. Since PRMs tend to use data gathered for other purposes (e.g., administrative government databases) and can be automated, they can efficiently examine large populations to identify a reduced number of people who are at higher risk (AUT & UAI, 2019, p. 107; cited in Valderrama, 2021).

The Coronavirus UY app was designed to provide the public with information on infection statistics for the then-novel coronavirus, as well as details on the health measures in effect at the time. It also aimed to monitor possible infections by collecting self-diagnosis data, offer remote medical care during periods of isolation, and, starting in mid-2020, alert users to potential contact with infected individuals. By centralizing information, the system helped facilitate government response planning — both broadly and at the individual level — ranging from providing care recommendations to delivering healthcare via telemedicine (Yael, 2021, p. 5).

The database used by the Coronavirus UY app, which centralized epidemiological forms, was under the Ministry of Public Health management. Each health care provider independently managed its own database. Patient electronic clinical documents, updated by treating physicians, were stored at the medical institution where the patient had been treated. All information generated by health personnel on a specific patient had to be kept in the custody of the corresponding institution. In addition, users could access their medical history on the web portals of the providers who offered this option, or they could request it directly using their national ID (Yael, 2021, p. 23).

Finally, the **Child Alert System** attempts to estimate and predict the level of risk for children of suffering a violation of their rights, using data analysis with different algorithmic models to anticipate and intervene early for prevention in each case. In practice, the system generates a score or "risk index" for each child or adolescent, making it possible to rank the cases by priority for Municipal Offices for Children (OLN). In addition, the system has become a platform for registering, managing and monitoring the cases of children and adolescents identified as being at greater risk (Valderrama, 2021, p. 23). The predictive models were trained with data from various sources, including SENAME, Chile Crece Contigo, the Ministry of Education (enrollment and academic performance at public and private schools), the Social Registry of Households, census data on neighborhood vulnerability and crime statistics on neighborhoods provided by the Undersecretariat for Crime Prevention, considering radii of 300 and 1,000 meters around the home of each child or adolescent (Valderrama, 2021, p. 31).

This is a system developed and maintained by the Undersecretariat for Social Assessment and implemented for the Municipal Offices for Children of the Undersecretariat for Children, both undersecretariats coming under Chile's Ministry of Social Development and Family. The role of predictive models is limited to conducting an initial classification using prioritization criteria to establish an order in which cases will be handled. This risk score is shown in a column along with other columns such as territorial alerts and Chile Crece Contigo. While calculation of the lists is limited to the universe of possible children and adolescents to be served, it is up to the coordinator and case managers at each OLN to decide whether or not to use the priority order calculated by the predictive tool (Valderrama, 2021, p. 23).

PUBLIC SAFETY

In terms of public safety, the **Urban Crime Prediction System** was analyzed, a development of the Security Analysis and Mathematic Modeling Center (CEAMOS) at the Universidad de Chile, together with the Criminal Analysis Department (DAC) of the Chilean Police Force (Carabineros). This development, implemented in 58 police precincts throughout the country, sought to predict areas at higher risk of criminal events to guide preventive police patrolling in cities, defining areas for greater surveillance and control. According to the report, the Government understands police surveillance as actions intended to prevent undesirable situations from arising or to detect them for neutralization, with the following operational features: preventive surveillance, police procedures, selective searches, extraordinary services and execution of warrants (MDS, 2013, p. 10; cited in Buschmann, 2021, p. 22).

AUPOL *(Automatización de Unidades Policiales)* is the primary platform used by the Carabineros to record and store data on reports, arrests, records, and infractions. This system facilitates the generation of police reports that are then submitted to courts and prosecutors.

The technology used in the context of the Urban Crime Prediction System is based on crime prediction, which is defined as any system that analyzes existing data to predict criminal events (Buschmann, 2021, p. 9). According to the author, in this context AI systems use machine learning and data analysis techniques to identify patterns in criminal activity. These patterns are based on criminological theories suggesting that crime is not randomly distributed but rather follows environmental, situational and social patterns that can be analyzed and understood (Buschmann, 2021, p. 10).

The crime prediction system uses two kinds of data. The first are police cases, including arrests and complaints related mostly to crimes of major social significance (CMSS) grouped in Robbery with Force and Robbery with Violence. The cases are recorded by the carabineros on the AUPOL platform, including data on the official entering the complaint or arrest in the system, and personal identification data on the people affected, witnesses, complainants and/or arrestees, such as full name, unique national ID number, profession, education, sex, age, physical characteristics, height and address. The second kind of data considered is the location of urban services and attractions identified as relevant contextual factors that could motivate or facilitate the occurrence of crime. This latter point, according to the system's developers, includes the location of banks, bus stops, restaurants and ATMs (Buschmann, 2021, p. 28). These data are obtained from information recorded by the carabineros in their geographic information system and from collaborative open-source platforms such as OpenStreetMap (Baloian et al., 2017; Carabineros de Chile, 2018; cited in Buschmann, 2021).

JUSTICE

In terms of the administration of justice sector, two cases were analyzed, both in Colombia. The first involves the PretorIA system, implemented at the Constitutional Court, and the second is Fiscal Watson, used under the aegis of the National Office of the Attorney General.

PretorIA uses Natural Language Processing (NLP), which is a field of artificial intelligence that focuses on the interaction between computers and human language. Through NLP, the system can analyze, categorize and extract relevant information from tutela texts. This includes the automatic labeling of documents and production of case statistics.

The **PretorIA** system uses natural language processing to support the process of selecting cases for judicial protection of fundamental rights (tutela) at Colombia's Constitutional Court. Its main function is to classify and label tutela sentences according to categories previously set up by experts. The system works with legal texts in Spanish and provides information on the content of the rulings, as well as general statistical data. In terms of its autonomy, PretorIA does not have the ability to make legal decisions. It works as a tool supporting the process of selecting tutelas, while the final decisions are made by Constitutional Court magistrates. The system does not work autonomously, and its expected role is to simplify case review work (Saavedra & Upegui, 2021, p. 5).

Tutelas are constitutional actions established in Article 85 of the Colombian Constitution. Their aim is the immediate protection of individual fundamental rights in response to situations of violation or threat of violation of those rights. This mechanism enables a citizen to request the intervention of the Constitutional Court or of judges to protect their fundamental rights quickly and effectively (Saavedra & Upegui, 2021, p. 18). PretorIA obtains data from the tutela case files that are sent to the Constitutional Court by judges and courts of first and second instance. The system processes sentencing texts and uses categories defined by Court personnel to classify and label the information (Saavedra & Upegui, 2021, p. 46).

In Colombia, the Office of the Attorney General has implemented **Fiscal Watson**, an AI-based tool developed by IBM, to support information management for the Oral Accusatory Criminal System (SPOA). This system centralizes information related to criminal investigations, judicial actions and management of evidence, among other aspects. Fiscal Watson uses advanced algorithms to analyze structured and unstructured data, identifying patterns, tendencies and possible connections between legal cases, with the objective of facilitating decision-making by judicial officials (Palacios et al., 2024, p. 11).

Fiscal Watson operates during the investigative phase of legal proceedings, when information from various investigations is gathered and correlated based on user-defined criteria. These criteria may be geographic (e.g., event location) or qualitative (e.g., specific details from the incident report). For instance, Watson can detect links among homicide cases involving the same perpetrator, identify similar criminal patterns within a region, or find parallels in the modus operandi across different cases. This analysis helps investigators gain a broader perspective and uncover potential connections that might not be evident when manually reviewing large volumes of data (Palacios et al., 2024, p. 13).

SPOA, the main source of information for Fiscal Watson, is a vast criminal information system that consolidates data from multiple legal, police, and administrative databases. This system includes modules for entering crime reports, managing legal actions, distributing cases among officials and consulting files. One of the most critical modules for Watson's operation is the one containing the factual accounts — that is, the initial descriptions of events recorded in legal cases.

The role of the official responsible for these reports is critical, since any omission or error in the details could lead to inaccurate, discriminatory or incorrect results in Watson's analysis (Palacios et al., 2024, p. 14).

To ensure the security and integrity of the data, Fiscal Watson does not directly access the original SPOA database. Instead, it uses a mirror copy of the system, which ensures that the original information is protected while Watson conducts its analyses and queries (Palacios et al., 2024, p. 16).

EDUCATION

The Guanajuato State Secretariat for Education, in Mexico, developed the **Early Action System for School Permanence (SATPE)**, which seeks to lower the dropout rate for upper intermediate education. The system comes under the Social Contract for Education, a comprehensive strategy by the government of Guanajuato to improve education quality and ensure students stay in the school system (Ricaurte & Nájera, 2024, p. 10). This contract is structured into four main components: ensuring school attendance, guaranteeing that nobody is left behind in their learning, recognizing the role of teachers, and fostering family participation in the educational process (Ricaurte & Nájera, 2024, p. 13).

Business intelligence (BI) tools make it possible to create analysis and data visualization scorecards using intuitive user interfaces, facilitating their use by people without advanced data processing technical knowledge.

The data used by the SATPE come from several sources, including the Scholastic Records System, which gathers information on student enrollment, attendance and academic performance in public schools. Data from the Guanajuato State Official Catalogue of Schools (CEO), which provides information on schools in the state, and from the Data Collection for Improving Learning Outcomes (RIMA) survey, which focuses on learning indicators, are also used. Information related to teaching staff is also considered (Ricaurte & Nájera, 2024, p. 18). Data on education indicators and scholastic records implemented for the SATPE are processed using PowerBI software, a business intelligence system developed by Microsoft (Ricaurte & Nájera, 2024, p. 21).

The data were gathered using the Office of School Services' Simplified Privacy Notification, which informs users (parents and tutors) of the purposes for which information on children and adolescents is collected, obtaining their tacit and — in some cases — express consent, according to statements by the Guanajuato State Executive Branch Transparency Unit (Ricaurte & Nájera, 2024, p. 18).

CITIZEN SERVICES AND ADMINISTRATIVE PROCESSES

A chatbot is a computer program designed to interact with users simulating a human conversation using voice or text commands, usually over the Internet. Over the years, chatbots have evolved from their early version in the 1960s to become algorithm-driven systems that are able to learn from interactions with users, thus optimizing their future answers (Adamopoulou et al., 2020; cited in Ferreyra, 2024).

The final case study is on the "Boti" Chatbot, implemented by the Government of the City of Buenos Aires (GCBA). This is a virtual assistant that allows citizens to interact and obtain information via WhatsApp. This chatbot uses Natural Language Processing and has an open domain focus, which enables it to offer answers on a wide range of topics, including citizen services, health and urban mobility. Since it was launched, Boti has evolved to include features such as assistance during the COVID-19 pandemic, providing public health information and enabling access to public services, becoming a centralized channel of communication between city government and citizens (Ferreyra, 2024, pp. 10–12).

As highlighted by GCBA, Boti has experienced significant growth since it was first implemented. During the COVID-19 pandemic, it became the main source of official information on symptoms, prevention and managing vaccine scheduling and certificates, among other features. During the first quarter of 2022, Boti reached its historic zenith of 26 million monthly interactions, becoming the main channel of communication between GCBA and citizens. However, according to the author based on official information, the numbers later dropped to between 2 and 5 million conversations per month (Ferreyra, 2024, p. 6).

This chatbot draws data from various City government sources and management systems. Information for specific citizen services is linked to other government systems, such as the Digital Procedures System (STD) and the Remote Administrative Procedures (TAD) platform. These systems enable the relevant government office to respond to user requests. Furthermore, the GCBA states that user-provided information is safeguarded by confidentiality agreements and is used solely to run the functionalities offered by the chatbot. In exceptional cases, the data may be retained for an extended period at the administration's discretion, which maintains that this is always done in compliance with data protection regulations (Ferreyra, 2024, p. 16).

3. ANALYSIS AND DISCUSSION: CHALLENGES ASSOCIATED WITH THE STATE'S AUTOMATED USE AND PROCESSING OF DATA AS A PUBLIC POLICY INSTRUMENT

In this section, we will examine several implications arising from these use cases, focusing on the different risks to the exercise of rights based on the data used in training, the data they employ, and the algorithms that process them. Likewise, we will assess the implications for personal data protection and access to public information in light of the legislation in each country studied.

Before proceeding, however, we need to delve into the impact of one of the most influential events that occurred during the course of this research: the COVID-19 pandemic.

The pandemic's role: using technology to handle massive amounts of data

As part of a collective research project in coordination with **Consorcio Al Sur**, **Derechos Digitales** carried out a study on the implications of technology use during the pandemic (Consorcio Al Sur, 2021). This study analyzed the main features of the mobile applications implemented by the governments of 14 Latin American countries and expressed deep concern over the lack of a comprehensive approach by States to ensure respect for human rights, in line with internationally established standards. Such noncompliance with their protection obligations has, in many cases, resulted in violations of those rights (Consorcio Al Sur, 2021, p. 65).

The lack of comprehensive government measures to safeguard rights is concerning not only in relation to mobile apps, but also with respect to the technologies examined in this study. Since the Artificial Intelligence and Inclusion program began in 2019, it was possible to observe the emergence of new applications aimed at managing pandemic-related issues in the region, as well as the expanding influence of preexisting technologies or applications. In four of the ten cases analyzed, the pandemic was a central factor in the implementation and growth of these technologies.

The Coronavirus UY app, analyzed by Yael (2021), is the clearest and most representative example of this. As mentioned, the app centralizes information to guide government actions both broadly and on a case-by-case basis, providing services ranging from precautionary guidance to telemedicine care (Yael, 2021, p. 5). It was among the most noteworthy initiatives within a broader technological implementation strategy that, in addition to the app, involved various citizen services offered through government websites and popular platforms like Facebook and WhatsApp — presented in the form of a virtual assistant (Yael, 2021, p. 7).

The second example is found in the use of AI in the Emergency Aid program in Brazil. In this case, technology was used as an administrative management tool to implement a policy targeted to mitigating the economic and social damage caused by the pandemic. Unlike Coronavirus UY, the AI use was internal and limited to administrative management. Although no big tech companies participated, the state enterprise Dataprev played a central role, as mentioned above.

In addition, the pandemic boosted the development of some pre-existing implementations of AI. The first case to mention, also in Brazil, is the National Employment System (SINE). The technical cooperation agreement that enabled the implementation of AI tools was established in November 2020 between the Brazilian government and Microsoft. It arose as a response from the company to a public invitation to bid targeted to mitigating the negative impact of the COVID-19 pandemic in Brazil's productive sector (Bruno et al., 2022, p. 6).

The last example to mention, which was broadly driven by the pandemic, is the Boti chatbot, implemented by the Government of the City of Buenos Aires. This chatbot played a leading role, adopting features ranging from citizen services and vaccination appointments to the creation of a self-diagnosis tool that worked through a neural network that could classify voice sounds, breathing and coughing, analyzing audio clips of coughing sent via WhatsApp to detect possible COVID-19 cases (Ferreyra, 2024, p. 12). During its operations in the context of the pandemic, monthly user interactions with the chatbot increased exponentially, leaping from a few hundred thousand to millions, with a spike of 26 million during early 2022 (Ferreyra, 2024, p. 6).

This underscores two key points deserving attention. First is the central role of artificial intelligence in the implementation of policies targeted to a large number of users, as in the four cases mentioned. Second, the relevance of agreements with big tech firms for the provision of these services in three of the cases, where Meta (formerly Facebook) and Microsoft especially stand out. Of the ten cases analyzed, the technologies implemented during the pandemic are the ones that show the greatest dependency on these big firms, including Fiscal Watson, which fully depends on IBM services.

Access to public information: challenges ranging from opacity to inappropriate data disclosure

As noted in the introduction, one of the greatest challenges in gathering information for the case studies was securing access to data and sources necessary to analyze how the State uses these technologies.

In the report on the Urban Crime Prediction System, Buschmann (2021, p. 9) points out that although the Carabineros have taken steps to enhance administrative transparency and integrity — such as establishing the Department of Public Information and Lobby, the Department of Complaints and Suggestions, and the STOP criminal statistics platform — most published data are not disaggregated. Additionally, the distribution of police personnel is kept confidential, and there is little information about misconduct by Carabineros personnel.

Buschmann also notes that summary investigations within the Carabineros are confidential, which has been challenged by the Inter-American Commission on Human Rights and conflicts with the principle of probity set forth in Chile’s Constitution. Furthermore, the author highlights an issue related to the Transparency Council (CPLT): its oversight role, like that of any public body, is restricted to enforcing regulations rather than resolving issues concerning requests for information. This underscores, in turn, the need to recognize access to public information as a fundamental right enshrined in the Constitution (Castillo, 2009; CIDH, 2016; cited in Buschmann, 2021, p. 9).

In the study on the Boti chatbot, Ferreyra (2024, p. 4) notes that gathering information about the chatbot’s operational features and its internal management mechanisms relied on publicly available sources as well as a request for access to public information submitted to the GCBA. Due to the vagueness of many of the answers received, a second request was filed which, as of the end of the study, remained unanswered (Ferreyra, 2024, p. 4).

It is worth mentioning that the City of Buenos Aires has a recently reformed access to information law (Law 104, 2017), the result of an open public consultation process. However, this law has not guaranteed effective access to information about the GCBA’s use of technology. Another example involves the Argentinian Observatory of Computer Rights (O.D.I.A.), which filed two requests for access to public information regarding the use of AI in facial recognition cameras in the City of Buenos Aires. According to Observatory sources, neither request was satisfactorily answered, prompting them to file an *amparo* (protective action) to halt the City’s Fugitives Facial Recognition System. The action was initially rejected, but after appealing — and expanding the complaint — they requested that the use of this technology be declared unconstitutional.

In contrast, in the case of Emergency Aid, the researchers do not analyze a problem of lack of information but rather the disclosure of personal data, including sensitive data, in a manner that contradicts data protection principles and individual rights. The authors (Tavares et al., 2022, p. 33) argue that any act of public administration in Brazil must adhere to the principle of transparency, originally established by the constitutional mandate to comply with the principle of publicity, as set forth in Article 37. The effects of this principle on public administration were reinforced in 2011 with the enactment of the Access to Information Act (Law 12.527/11), which establishes that publicity is the general rule and confidentiality the exception. Within this framework, actions and procedures related to Emergency Aid are subject to these legal obligations, and the responsible authorities — particularly the Ministry of Citizenship, as well as the Caixa Econômica and Dataprev public enterprises — must make all relevant information available without the need for a prior request, provided that the public interest is confirmed and the fundamental rights involved are upheld (Tavares et al., 2022, p. 33).

However, the authors emphasize that, while there is a lack of active transparency in automated decision-making on data management, the principle of transparency is applied excessively and indiscriminately in the political realm. This is evident in the publication, on the Transparency Portal, of a list containing beneficiaries' personal data, including full names, amounts received, and payment details, among other sensitive information. Justified on the grounds of accountability and fraud prevention, this approach directly conflicts with both the Access to Information Act, which mandates the protection of personal data, and fundamental principles of data protection (Tavares et al., 2022, p. 33).

This situation highlights the poor alignment of social protection policies with data protection standards, as well as a disregard for informational self-determination, which the Federal Supreme Court recognizes as a fundamental right. By disclosing personal data in this manner, the program prevents individuals from exercising control over their own information (Tavares et al., 2022, p. 33).

The last relevant case to analyze is Chile's Child Alert System (SAN). According to Valderrama (2021, p. 8), the report was based on information obtained through an access to information request under the Transparency Act, supplemented by a thorough review of secondary sources, including news articles, presentations, procurement notices, technical proposals, technical guidelines, reports, public accounts, and purchase orders from the responsible ministry and other entities. However, the author highlights several challenges encountered during the research process, including the explicit refusal of key officials from the Ministry of Social Development and Family to participate in interviews, as well as the lack of up-to-date public documentation on the status of the SAN predictive tool, which had already been implemented in Local Child Offices (Valderrama, 2021, p. 8).

Considerations on personal data protection

Since the AI technologies studied rely on processing large volumes of data, it is essential to address the personal data protection concerns identified by researchers. In some cases, non-compliance with personal data protection standards is evident. In others, however, while the State may provide certain safeguards, studies emphasize the need for rigorous oversight to effectively enforce the protections established.

In Chile, the Child Alert System (SAN) raises specific concerns about the processing of personal data of children and adolescents, which the Council for Transparency (CPLT) considers particularly sensitive. According to the CPLT, data on children and adolescents require enhanced protection due to the lack of clear informed consent and the fact that minors may be less aware of the risks associated with data processing. As a result, the Council has restricted the disclosure of information on minors, limiting it to cases where the requester is verified as the legal guardian. In other cases, access has been denied to prevent specific and foreseeable harm to minors’ privacy (Valderrama, 2021, pp. 14–15).

The CPLT has also questioned agreements such as the collaboration between the National Minors Service and the National Intelligence Agency, arguing that they fail to meet child and adolescent data protection standards. This stance aligns with the Convention on the Rights of the Child, which prohibits arbitrary interferences with children’s privacy, and with the principle of the best interests of the child. Based on this principle, a specialist consulted by the author affirms that minors’ data processed within the educational system should not be considered public information (Valderrama, 2021, pp. 14–15).

In Colombia, the PretorIA system, developed by the Constitutional Court, automates the selection of tutela cases for review without processing sensitive personal data, according to Saavedra & Upegui (2021, p. 47). The system operates on the text of judicial rulings and does not alter access to or the processing of personal data, as the traditional process already required the submission of legal documents. The Court has stated that the system does not rely on names or specific personal identifiers and can function with anonymized data. Additionally, Law 1581 of 2012 establishes exceptions to consent for data processing necessary for judicial functions.

However, although the system does not modify access to or the processing of personal data, nor does it directly affect individual rights, its social impact is significant. This impact stems from its role within a judicial process that requires legitimacy and public trust. For this reason, the authors stress the need for rigorous technical and operational oversight to prevent malfunctions and ensure the transparency and legitimacy of the process (Saavedra & Upegui, 2021, p. 47).

In Uruguay, the Coronavirus UY system centralized patients' personal information, including age, phone number, ID card, symptoms, and pre-existing conditions. This information was organized into a single database, accessible to the Ministry of Public Health (MSP) and healthcare providers, who used it for patient follow-up. The centralized database was owned by the MSP, which was responsible for setting guidelines and protocols for its use (Yael, 2021, p. 17).

The personal data used by the app are protected under Uruguay's personal data protection legislation (Law No. 18.331), which classifies health data as particularly sensitive (Art. 4). According to Articles 17 and 19, such data can only be processed by health institutions for purposes directly related to the legitimate interests of the issuer and recipient, with the prior consent of the data subject. However, on the same day the app was launched, March 20, 2020, the Personal Data Regulation and Control Unit (URCDP) issued Ruling No. 2/020, stating that, due to the public health emergency and under legal authorization, the processing of health data — such as that collected by the Coronavirus UY app — could be carried out without the prior consent of the data subjects.

Furthermore, the MSP serves as the auditing and oversight authority for the data ecosystem. This case illustrates how governments' excessive use of sensitive data during the pandemic posed a serious risk to fundamental rights.

Problems with the automated use of data and algorithms

It is relevant to analyze the potential violation of rights that could result from how technologies are built, particularly in terms of AI-based algorithms or systems, as well as the data used for their training and processing. A pertinent example for this analysis is the case of the Child Alert System employed in Chile.

According to Valderrama (2021, p. 31), the modeling of the Child Alert System algorithm relied primarily on data from individuals who had interacted with government education and social aid services, a population that tends to have lower income or educational levels. This could create socioeconomic disparities, meaning that the model may have reduced accuracy in identifying high-risk children from higher socioeconomic backgrounds, while at the same time overestimating the risk for families from lower socioeconomic levels (Valderrama, 2021, p. 36).

It is therefore crucial to consider the representativeness of the data and explore ways to incorporate broader and more diverse information to enhance the predictive system's accuracy and fairness. This, in turn, would strengthen the public policy framework within which the system operates.

Another issue can be seen in the aforementioned Emergency Aid program, implemented by Brazil's Federal Government. As previously noted, this case relies on the extensive use of data for the automated selection of beneficiaries. Its data infrastructure is based on the cross-referencing of 34 different databases, including records from Cadastro Único, the National Registry of Social Information (CNIS), and data from various government agencies, such as the Ministry of Economy and Caixa Econômica Federal. However, this complexity faces significant challenges due to the obsolescence of the records (Tavares et al., 2022, p. 31).

In particular, key databases such as the Annual Social Information Report (RAIS) have not been updated with recent data, which negatively impacts individuals who have lost their jobs or experienced changes in their employment status since the 2018 base year. This lack of up-to-date records has led to the exclusion of individuals who, despite meeting the eligibility criteria, are unable to access the benefit. Consequently, rather than promoting social inclusion, the program's digital architecture reinforces barriers to access and weakens the effectiveness of social protection policies. As will be analyzed further, this situation led to legal action, allowing for a more detailed evaluation of the program's functioning.

Regarding the Urban Crime Prediction System in Chile, it is important to examine how it reinforces biases, not only by increasing over-surveillance in certain areas but also by perpetuating discriminatory practices through the targeted deployment of surveillance technologies in locations previously classified as high-risk. According to Buschmann

(2021, p. 41), the Urban Crime Prediction System relies primarily on two data sources: police records, which include arrests and reports of crimes deemed socially significant (CCSS), and data collected through the AUPOL platform, used by Carabineros. These datasets contain personal information about individuals involved, including names, ID numbers (RUN), professions, gender, and home addresses.

However, the production of this data is not a neutral process. As Buschmann (2021, p. 41) explains, every data point is part of a social production chain, meaning it may incorporate biases and contextual perspectives. As a result, these databases may include inaccurate or misleading information, such as arbitrary detentions or unverified complaints, leading to biased predictive systems that reinforce discriminatory policing, particularly in the enforcement of preventive identity checks. Furthermore, the research highlights the absence of evaluation protocols or external audits to ensure the integrity of the data collection process (Buschmann, 2021, p. 7).

Democratic counterweights: the role of oversight and legal bodies

This raises questions about the role of oversight bodies in balancing the protection of fundamental rights, particularly public defender offices and external auditing units, which serve as official agencies responsible for monitoring government administration. Additionally, we will examine the role of the judicial system in some of these cases, where citizens filed complaints seeking to rectify policy failures — some of which stemmed from improper data processing.

The case of Boti, in the City of Buenos Aires, serves as a strong example in this regard. According to Ferreyra in their study on the government chatbot (2024, p. 20), in 2022, the General Audit Office of the City of Buenos Aires (AGCBA) — an independent public oversight agency — conducted a comprehensive analysis of the systems, processes, and technologies ensuring the operability of the Boti chatbot, covering the year 2021. The audit report, published in March 2023, recognized the modernization efforts carried out by the City Government to improve access to information for handling citizen services and appointment scheduling. However, it also identified critical areas for improvement, particularly regarding the formalization of administrative procedures related to the chatbot. The report emphasized the need to establish robust IT policies for the management and protection of personal data, as well as the importance of effective governance in information and communication technologies. Additionally, the audit underscored that managing and storing large volumes of data necessitates a constant review of policies to ensure the security and protection of processed data.

We can cite another example from this same case. In 2022, the Ombudsman of the City of Buenos Aires, the City's personal data protection authority, conducted an investigation in response to a complaint regarding the Boti chatbot's operations (Ferreyra, 2024, p. 21). The complaint, filed by a citizen, was based on her concern for the lack of available legal notice on entering the Boti virtual assistant on the Buenos Aires government's official webpage. It also noted that, via this chatbot, anyone with knowledge of a third party's national ID number and phone number could obtain sensitive information, such as COVID-19 test results. On confirmation of these facts, the Ombudsman's office issued recommendations to improve the clarity and completeness of the legal notice, as well as to ensure the proper enrollment of databases in the corresponding register. The need for safe, ethical handling of personal data was emphasized, underscoring the importance of compliance with data protection legislation.

The judicial system has played a critical role as intermediary between citizens and the government in protecting the exercise of rights. In light of the problems noted in the context of the Emergency Aid program in Brazil, and due to the lack of administrative mechanisms for reviewing automated decisions, the judicial pathway became the main resource for challenging these decisions and requesting human analysis of how the benefit was granted. Because it involved a federal program, the legal response took place through the Federal Justice system (Tavares et al., 2022, p. 37). The program's judicialization was intensified because of how out-of-date the Federal Government system registries were. There are cases of unemployed people who, despite lacking formal work, appeared with an active employment status in the databases, which prevented them from accessing the benefit. This situation led to a significant increase in the number of legal appeal actions, which reached nearly 76,000 in September 2020 (Tavares et al., 2022, p. 22). To address this problem, an agreement was set up between the Federal Public Defender's Office and the Ministry of Citizenship, which enabled the Public Defender's Office to access a specific Dataprev system for consulting and presenting administrative objections (Tavares et al., 2022, p. 22). According to the authors, the lack of human review in automated decisions can thus be considered to have furthered judicialization, showing the system's limitations for guaranteeing inclusive access to Emergency Aid (Tavares et al., 2022, p. 31).

On this point, Buschmann (2021, p. 9) analyzes the role of the General Comptroller of the Republic, a key institution in Chile for transparency and public oversight. This autonomous agency oversees the investment of funds by different state institutions, including the police, which was the object of the case study. During accountability reviews, the Comptroller can formulate objections and observations, verifying the legality of acts via audits that assess activities, results and procedures to determine whether they comply with established regulations and principles. According to the author, these audits include financial, legality, management, results, and accountability controls, as well as evaluation of internal controls. This approach has made it possible to identify problems related to the digital platforms used by the Carabineros, which were examined in the course of the author's research (Buschmann, 2021, p. 9).

The role of oversight and supervisory bodies is an important example of how States, within their democratic functioning, can mitigate the shortcomings identified in the elements analyzed in this section. These agencies have intervened in cases of violations related to personal data protection, access to information and problems stemming from databases that are out-of-date and, ultimately, inadequate for automating processes. However, it is essential to highlight the importance of conducting prior assessments and ensuring that the implementation of these technologies respects human rights frameworks, thus preventing situations where rights have been violated before oversight agencies can intervene.

CONCLUSIONS

In the framework of the Artificial Intelligence and Inclusion project, over a period spanning nearly six years, ten cases of technology use in seven Latin American countries were analyzed. During this time, from 2019 to 2024, the development of AI-based technologies increased notably, as did the dissemination of their use, mainly starting with the mass use of generative AI at the end of 2022. Furthermore, this period saw the beginning and end of the COVID-19 pandemic, which offered an exceptional situation that intensified the use of technology as an instrument for managing not only citizen services and administrative processes, but also critical social protection policies. Likewise, the period witnessed the launch of some of the most influential ethical and regulatory frameworks in the region, such as the OECD Principles for Trustworthy AI (2019),[5] UNESCO's Recommendations on the Ethics of Artificial Intelligence (2021),[6] or the European Union's Artificial Intelligence Act (2024).[7] Below we will analyze which factors remain and what changed regarding AI use in government during this period.

[5] Available at: https://oecd.ai/en/ai-principles

[6] Available at: https://www.unesco.org/es/legal-affairs/recommendation-ethics-artificial-intelligence

[7] More information at https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

The first observation has to do with the persistence of a lack of appropriate regulatory frameworks for implementation of these technologies. On this point, information from the first Global Index on Responsible AI report, published in 2024, is illustrative. For the dimensions of Government Actions and Frameworks, based on analysis of the existence of national AI strategies, assessments of impact on rights, actions for human review, considerations on proportionality, clear guidelines for transparency and explainability, among other criteria, only two countries in the region scored over 50%. For the Human Rights and AI dimension, none of the countries scored over 50%. This dimension of analysis considers indicators on gender equality, data protection and privacy, public engagement, children's and adolescents' rights, and worker protections.[8] This is one of the conclusions that is reinforced in terms of the first comparative report published under this project (Velasco & Venturini, 2021), where the lack of specific frameworks for AI use in government was noted.

In the area of protection and use of personal data, oversight agencies played a critical role in preventing automated data processing from violating fundamental rights, such as access to social benefits. A particularly informative example is the case of Emergency Aid in Brazil, where the failure to keep databases updated caused significant harm to people attempting to access the program. This case demonstrates the risks associated with policies based on automated data processing and underscores the importance of guaranteeing the quality of the data that will be processed by the algorithms.

In terms of access to public information, a persistent problem is observed that runs through cases at all stages of publication. The incomplete response to an access to information request, as found with the Boti chatbot in the City of Buenos Aires, or the explicit refusal to provide interviews, as in the analysis of the Child Alert System in Chile, are examples that show the difficulties in investigating State use of artificial intelligence. However, lack of transparency is not the only challenge; another is publishing information without respecting personal data protection, as seen in the case of Emergency Aid in Brazil. In this case, the Transparency Portal was used to publish a list with personal data from the program's beneficiaries. Although this action was justified by reason of accountability; as noted by the authors, it comes into conflict with the Access to Information Act, which requires public authorities to protect personal data, and with data protection principles (Tavares et al., 2022, p. 33). Therefore, it is essential to create and interpret regulatory frameworks in a complementary fashion to guarantee that the protection of individual fundamental rights is prioritized.

[8] The full report is available at: https://www.global-index.ai/Region-South-and-Central-America

Another point of concern is the persistent lack of spaces for meaningful participation that ensure the diversity and inclusion of multiple stakeholders, not only in the context of policy implementation, driven by different government authorities, but also in emerging regulatory spaces. In a prior investigation, Derechos Digitales analyzed the participatory processes created in the context of so-called artificial intelligence plans and strategies promoted by different governments in the region. This study shows that, while efforts exist to create participatory spaces, they are still insufficient to empower citizens in decision-making around public policies that could directly affect the enjoyment of their rights (Hernandez et al., 2022).

In this way, the persistence of problems related to personal data protection, access to public information and the lack of spaces for meaningful engagement is seen in the implementation of artificial intelligence technology by States. While adherence to ethical frameworks is appropriate, it is not sufficient; clear rules and governance frameworks that ensure the participation of multiple stakeholders must be established. At Derechos Digitales, we consider it a priority to incorporate a human rights perspective in all processes related to the regulation of artificial intelligence, whether through executive branch or parliamentary initiatives in the region, with the goal of promoting responsible and inclusive use of these technologies in public administration, including clear limitations on their use.

BIBLIOGRAPHY

Alimonti, V., & Cavalcanti de Alcântara, R. (2024). Estándares interamericanos y uso estatal de la IA en decisiones que afecten derechos humanos: Implicaciones para los DDHH y marco operativo. Electronic Frontier Foundation. https://www.eff.org/document/estandares-de-derechos-humanos-para-el-uso-estatal-de-la-ia-en-america-latina

Bruno, F., Cardoso, P. & Faltay, P. (2021). *Sistema Nacional de Empleo y la gestión automatizada de la desocupación laboral*. Derechos Digitales.

Buschmann, J. (2021). *Sistema predictivo del delito urbano: Producción algorítmica de zonas de vigilancia y control en la ciudad.* Derechos Digitales.

Consorcio Al Sur (2021) Informe Observatorio Covid-19: Un análisis crítico de las tecnologías desplegadas en América Latina contra la pandemia. Consorcio Al Sur. https://www.alsur.lat/reporte/informe-observatorio-covid-19-consorcio-al-sur-un-analisis-critico-tecnologias-desplegadas

Ferreyra, E. (2024). Boti: *estudio sobre el chatbot con procesamiento del lenguaje natural del Gobierno de la Ciudad de Buenos Aires*. Derechos Digitales.

Hernández, L., Canales, M.P. & Souza, M. (2022) I*nteligencia Artificial y participación en América Latina: Las estrategias nacionales de IA*. Derechos Digitales.

Nájera, J., Ricaurte, P. (2021). Tecnologías de interés público: el caso de las coronapps en América Latina (Policy Report No. 1, Serie 1: TIC en tiempos de Covid-19). Centro Latam Digital. https://centrolatam.digital/publicacion/coronapps/

Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos. (2011). *Principios Rectores sobre las Empresas y los Derechos Humanos: Puesta en práctica del marco de las Naciones Unidas para "proteger, respetar y remediar".* Naciones Unidas. https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf

(2021). *Artificial intelligence risks to privacy demand urgent action –Bachelet*. https://www.ohchr.org/en/2021/09artificial-intelligence-risks-privacy-demand-urgen t-action-bachelet

Palacios, L., Forero, V., & Labarthe, S. (2024). *Fiscal Watson: estudio del uso de Inteligencia Artificial en la Fiscalía General de la Nación en Colombia.* Derechos Digitales.

Ricaurte, P., & Nájera, J. (2024). *SATPE*: *análisis del sistema predictivo para la prevención del abandono escolar del estado de Guanajuato*. Derechos Digitales

Ruiz Nicolini, J. P., Kunst, M., & Dias, J. M. (2024). *Usos inteligentes de datos en el Estado.* Fundar. https://fund.ar/wp-content/uploads/2024/09/Fundar_Usos_inteligentes_de_datos_en_el_Estado_CC-BY-NC-ND-4.0-2.pdf

Saavedra, V., & Upegui, J. C. (2021). *PretorIA y la automatización del procesamiento de causas de derechos humanos*. Derechos Digitales.

Sequera, M., & Cuevas, M. (2024). *EmpleaPY: Investigación sobre la automatización de procesos para las políticas de empleo en Paraguay*. Derechos Digitales.

Tavares, C., Fonteles, J., Simão, B., & Valente, M. (2022). *El Auxilio de Emergencia en Brasil: Desafíos en la implementación de una política de protección social datificada.* Derechos Digitales.

Tedic (2024) "Última versión del proyecto de ley de datos personales en Paraguay: Un trabajo colectivo y participativo". Tedic. Disponible en: https://www.tedic.org/ultima-version-del-proyecto-de-ley-de-datos-personales-en-paraguay/

Yael, D. (2021) *Coronavirus UY y la tecnología como solución a la pandemia*. Derechos Digitales.

Valderrama, M. (2021). *Sistema Alerta Niñez: IA e inclusión en Chile.* Derechos Digitales.

Velasco Fuentes, P. & Venturini, J. (2021) *Decisiones automatizadas en la función pública en América Latina: Una aproximación comparada a su aplicación en Brasil, Chile, Colombia y Uruguay.* Derechos Digitales.

derechosdigitales.org