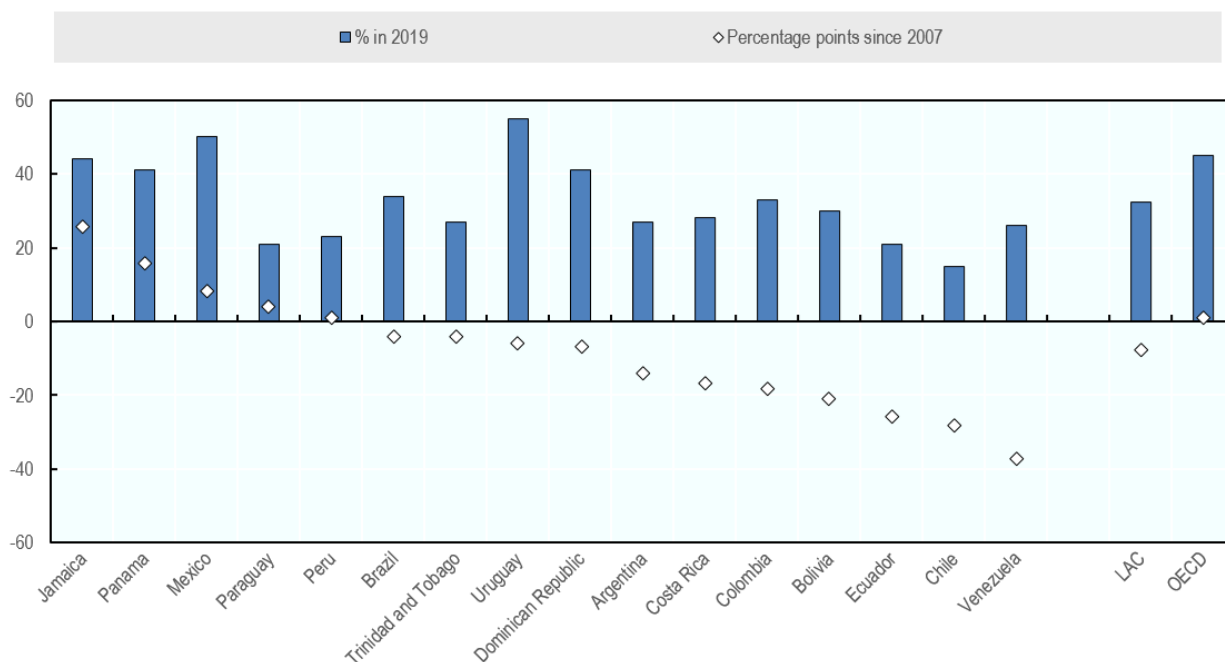


4 Efforts to develop a responsible, trustworthy and human-centric approach

As is evident from the use cases presented in Chapter 3, there is significant potential for AI applications in the public sector across the LAC region. However, the application of AI also raises many challenges and has implications that LAC government leaders and public servants need to consider when determining whether and how the technology can help them address problems and achieve their objectives. This chapter explored how LAC governments are building principles and taking actions for ensuring that they take a responsible, trustworthy and human-centric approach to AI.

Low and often declining levels of trust in LAC governments (see Figure 4.1) demonstrate that LAC governments must take a strategic and responsible approach to AI in the public sector. This approach must build confidence among the public that AI is being used in a trustworthy, ethical and fair way, and that the needs and concerns of citizens are at the heart of government decisions and actions with regard to AI.

Figure 4.1. Trust in government is declining in many LAC countries, often from a low starting point



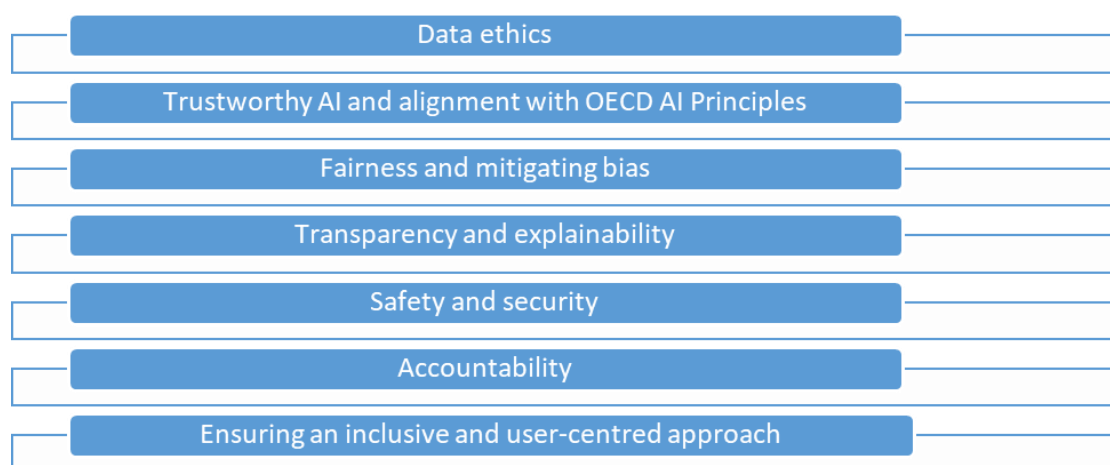
Note: Data refer to the percentage of people who answered “yes” to the question “Do you have confidence in your national government”. Data for Barbados are not available, data for Trinidad and Tobago are for 2017 rather than 2019, and data for Jamaica and Trinidad and Tobago are for 2006 rather than 2007.

Source: Gallup World Poll.

In order to achieve this, LAC governments must develop a responsible, trustworthy and human-centric approach to designing and implementing AI, one that identifies trade-offs, mitigates risk and bias, and ensures open and accountable processes and actions. Governments will also need to bring together multi-disciplinary and diverse teams to help with such determinations and to promote the development of public sector AI initiatives and projects that are both effective and ethical. Finally, a key aspect for addressing these and other considerations is for LAC countries to gain an understanding of the needs of their people and to ensure a focus on users and individuals who may be affected by AI systems throughout their life cycle.¹

This chapter explores these issues in the LAC regional context with the aim of helping government leaders and public servants to maximise the benefits of AI while mitigating and minimise potential risks. The overall topics in this chapter are present in Figure 4.2.

Figure 4.2. Issues discussed in Chapter 4



Data ethics

Most modern AI systems are built on a foundation of data. However, the availability, quality, integrity and relevance of data are not sufficient to ensure the fairness and inclusiveness of policies and decisions, or to reinforce their legitimacy and public trust. Consistent alignment and adherence to shared ethical values and principles for the management and use of data are essential to: 1) increase openness and transparency; 2) incentivise public engagement and ensure trust in policy making, public value creation, service design and delivery; and 3) balance the needs to provide timely and trustworthy data (OECD, 2020^[11]). To help countries think through the considerations around the management and use of data, the OECD has developed the *Good Practice Principles for Data Ethics in the Public Sector* (Box 4.1). Since data are foundational for AI, data ethics, by extension, are essential for the trustworthy design and implementation of AI. The forthcoming review *Going Digital: The State of Digital Government in Latin America* will provide a broader discussion of data ethics in LAC countries. Accordingly, this section focuses more specifically on aspects of trustworthy and ethical AI.

Box 4.1. OECD Good Practice Principles for Data Ethics in the Public Sector

Governments need to be prepared to handle and address issues and concerns associated with data corruption; biases in data generation, selection and use; data misuse and abuse; and unexpected negative outcomes derived from data use increase. The OECD Digital Government and Data Unit and its Working Party of Senior Digital Government Officials (E-leaders) have finalised guiding principles to support the ethical use of data in digital government projects, products and services to ensure they are worthy of citizens' trust. The principles are as follows:

- Manage data with integrity.
- Be aware of and observe relevant government-wide arrangements for trustworthy data access, sharing and use.
- Incorporate data ethical considerations into governmental, organisational and public sector decision-making processes.

- Monitor and retain control over data inputs, in particular those used to inform the development and training of AI systems, and adopt a risk-based approach to the automation of decisions.
- Be specific about the purpose of data use, especially in the case of personal data.
- Define boundaries for data access, sharing and use.
- Be clear, inclusive and open.
- Publish open data and source code.
- Broaden individuals' and collectives' control over their data.
- Be accountable and proactive in managing risks.

Source: (OECD, 2021^[2]).

Trustworthy AI and alignment with the OECD AI Principles

Ensuring trustworthy and ethical practices are in place is critical because the application of AI involves governments implementing AI systems with various degrees of autonomy. Ethical decisions regarding citizens' well-being must be at the forefront of governments' efforts to explore and adopt this technology, if they are to realise the potential opportunities and efficiencies of AI in the public sector. Trust in government institutions is contingent on their ability to be competent and effective in delivering on their mandates, while operating consistently on the basis of a set of values that reflect citizens' expectations of integrity and fairness (OECD, 2017^[3]).

The use of AI to support public administrations should be framed by strong ethical and transparency requirements, in order to complement the relevant regulations in place (e.g. in terms of data protection and privacy) and to avoid doubt regarding possible biased results and other issues arising from opaque policy procedures and AI usages. The OECD Digital Economy Policy Division² in the Science Technology and Innovation Directorate has developed the OECD AI Principles, which include the development of a reference AI system life cycle (OECD, 2019^[4]). Since 2019, the Digital Economy Policy Committee has been working to implement the OECD AI Principles in a manner consistent with its mandate from the OECD Council. The Committee has also launched the OECD.AI Policy Observatory and engaged a large OECD AI Network of Experts to analyse and develop good practices on the implementation of the OECD AI Principles.

This section of the report leverages the OECD AI Principles to assess how LAC countries are approaching trust, fairness and accountability for the development and use of AI systems. It examines the mechanisms that exist to address such concerns along the AI system life cycle. Accordingly, the analysis considers how countries respond to the ethical questions posed by the design and application of AI and associated algorithms.

Many national governments have assessed the ethical concerns raised by AI systems and applications, notably related to inclusion, human rights, privacy, fairness, transparency and explainability, accountability, and safety and security. Several countries around the world are signatories to international AI guiding principles. As touched on in the Introduction, 46 countries have adhered to the OECD AI Principles (Box 4.2), including seven LAC countries. Recently, the G20 adopted the "G20 AI Principles",³ which are drawn directly from the OECD AI Principles. Three LAC countries – Argentina, Brazil and Mexico – have committed to these principles by virtue of their participation in the G20. Some countries have also designed their own country-specific principles. Adhering to or otherwise articulating clear principles for AI represents a positive step for international co-operation, and for bringing about an environment and culture aligned

with the societal goals and values articulated in the Principles. Table 4.1 provides an overview of LAC government adherence to the OECD and G20 AI Principles and indicates where own country-specific principles have been put in place.

Box 4.2. OECD AI Principles

The OECD Principles on Artificial Intelligence support AI that is innovative and trustworthy, and which respects human rights and democratic values. OECD member countries adopted the Principles on 22 May 2019 as part of the OECD Council Recommendation on Artificial Intelligence (OECD, 2019^[5]). The Principles set standards for AI that are sufficiently practical and flexible to stand the test of time in a rapidly evolving field. They complement existing OECD standards in areas such as privacy, digital security risk management and responsible business conduct.

The Recommendation identifies **five complementary, values-based principles** for the responsible stewardship of trustworthy AI:

- AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being.
- AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.
- There should be transparency and responsible disclosure around AI systems to ensure that people understand AI-based outcomes and can challenge them.
- AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed.
- Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.

Consistent with these value-based principles, the OECD also provides **five recommendations to governments**:

- Facilitate public and private investment in research and development to spur innovation in trustworthy AI.
- Foster accessible AI ecosystems with digital infrastructure and technologies and mechanisms to share data and knowledge.
- Ensure a policy environment that will open the way to deployment of trustworthy AI systems.
- Empower people with the skills for AI and support workers for a fair transition.
- Co-operate across borders and sectors to progress on responsible stewardship of trustworthy AI.

In early 2020, the OECD convened a multi-stakeholder and multi-disciplinary OECD Network of Experts in AI (ONE AI) to develop practical guidance to implement the AI Principles (OECD, 2019^[5]). The working group on Policies for AI has developed a report on the *State of Implementation of the OECD AI Principles: Insights from National AI Policies* (OECD, 2021^[6]). The report provides good practices and lessons learned regarding implementation of the five recommendations to policy makers contained in the OECD AI Principles.

Source: <https://oecd.ai> and <https://oecd.ai/network-of-experts>.

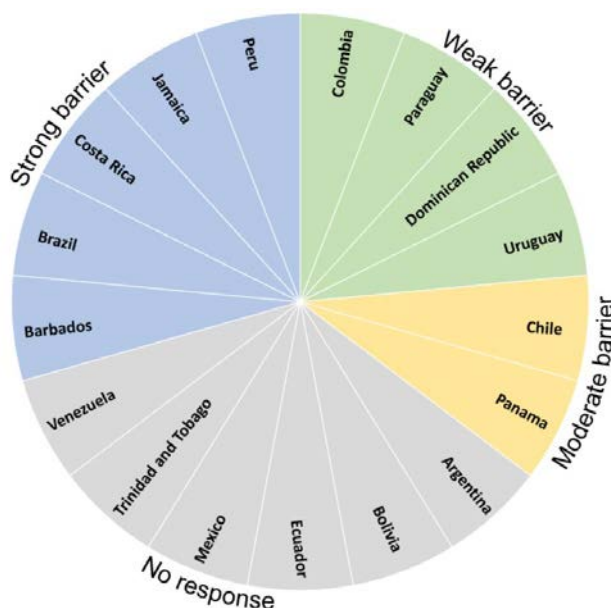
Table 4.1. LAC country establishment and adherence to the AI principles

	OECD AI Principles	G20 AI Principles	Country-specific principles
Argentina	✓	✓	
Barbados			
Brazil	✓	✓	
Bolivia			
Chile	✓		✓
Dominican Republic			
Colombia	✓		✓
Costa Rica	✓		
Ecuador			
Jamaica			
Mexico	✓	✓	✓
Panama			
Paraguay			
Peru	✓		
Trinidad and Tobago			
Uruguay			✓
Venezuela			

Source: OECD LAC Digital Government Agency Survey (2020); <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>; <https://oecd.ai>.

Committing or adhering to ethical principles is likely to be a necessary but not necessarily sufficient condition for trustworthy deployment of AI. If principles are to have maximum impact on behaviour, they must be actionable and embedded in the processes and institutions that shape decision making within governments. The OECD has found that the absence of common standards and frameworks are the obstacles most often cited by digital government officials in their pursuit of AI and other emerging technologies, largely due to growing concerns around fairness, transparency, data protection, privacy and accountability/legal liability (Ubaldi et al., 2019^[7]). Out of 11 respondents to the OECD's digital government agency survey, seven LAC countries stated that insufficient guidance on the ethical use of data represents a strong or moderate barrier for data-enabled policy making, service design and delivery, and organisational management (Figure 4.3). Among these countries were a number that have adhered to the OECD principles and/or have created their own country-specific principles. While the responses focus on the ethical use of data, they can serve as a proxy measure for AI ethics. The use cases discussed in the previous chapter also show that public data and AI developments have encountered ethical challenges that could be mitigated or clarified if ethical guidance, standards and/or frameworks were in place to help actualise high-level principles. The following sections review the main instruments and initiatives that contribute to developing responsible, trustworthy and human-centric approaches to AI in the public sector.

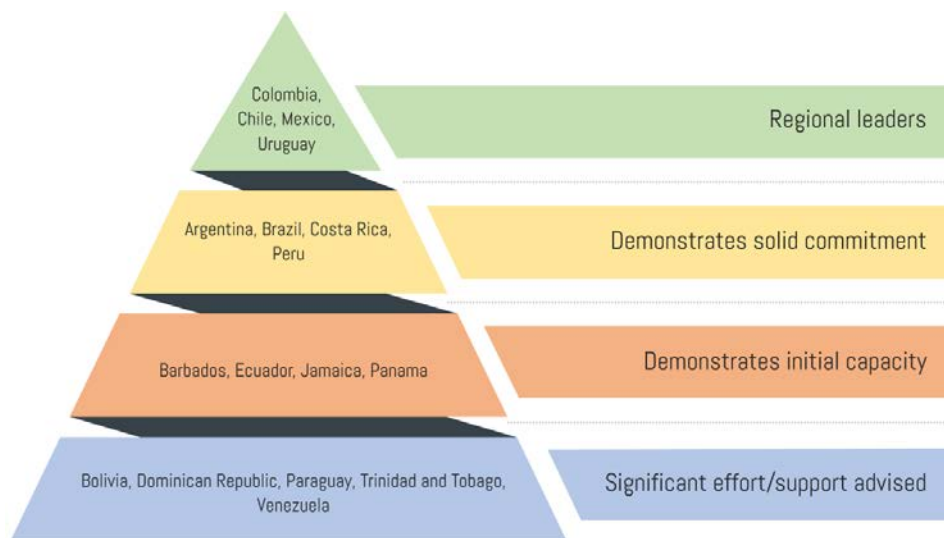
Figure 4.3. Insufficient guidance on the ethical use of data as a barrier to improved policy making, service design and delivery and organisational management



Source: OECD LAC Digital Government Agency Survey (2020).

LAC country frameworks and mechanisms for trustworthy and ethical AI in the public sector

Figure 4.4. LAC region capacities for establishing legal and ethical frameworks for AI



Note: All countries that adhere to the OECD AI Principles have been categorised as “demonstrates solid commitment” or higher.

As shown in Table 4.1, of the 17 LAC governments featured in this study, five have developed or are developing their own country-specific principles to guide their exploration and use of AI. All of these efforts have been initiated in the last few years, indicating a recent and accelerating focus specifically on ensuring trustworthy and ethical AI policies and systems. A brief overview of evolution in this area is as follows:⁴

- In 2018, Mexico published 14 principles for the development and use of AI, becoming the first country in the region to advance in setting frameworks for this technology with a focus on the public sector.
- In 2019, Uruguay included nine general principles as part of their AI strategy to guide the digital transformation of the government and provide a framework for the use of AI in the public sphere.
- In 2020, both Colombia and Chile released consultation draft principles documents to guide their AI efforts. The former published the Ethical Framework for AI, a product of commitments included in its 2019 AI strategy, and is currently organising expert roundtables to receive feedback in order to develop a final version.⁵ Chile also includes an Ethics sub-axis as part of its AI policy.
- In its 2021 national AI strategy, Brazil committed to developing ethical principles for the design and implementation of AI systems. While ethics was a strong focus in the Brazilian AI strategy, the scope and content of its country-specific ethical principles have not yet been released.

In addition to AI-specific principles, Barbados, Brazil, Jamaica, Panama and Peru have issued recent data protection legislation that better align the countries with the OECD AI Principles due to the inclusion of transparency, explainability, and fairness rights and principles with regard to data collection and processing. Brazil's data protection legislation additionally includes principles related to safety and accountability. Such rules can contribute to trustworthy and ethical design and use of AI systems, and represent a step forward in building a legal and regulatory framework to support and guide AI progress. Such updates have been cited by a number of LAC countries as essential in the light of new technologies. For instance, in Panama, there was a consensus among all public sector organisations interviewed during an OECD fact-finding mission in November 2018 that the legal and regulatory framework needed updating to reflect technologies such as AI and data analytics (OECD, 2019^[8]).

As seen in Annex B,⁶ for the most part, LAC countries developing their own principles address the same topics as the OECD Principles, although in more detail and with greater precision in order to emphasise local priorities and country-specific context. For example, when considering how countries are aligned with the first OECD Principle on “inclusive growth, sustainable development and well-being”, they generally cover inclusion, social benefit and general interest, but also stress particular issues. Uruguay states that AI technology development should bear as purpose complementing and adding value to human activities; Mexico believes that measuring impact is fundamental to ensuring AI systems fulfil the purposes for which they were conceived; Peru envisions the creation of a dedicated unit to monitor and promote the ethical use of AI in the country; Colombia incorporates a specific measure to protect the rights of children and adolescents; and Chile's approach integrates environmental sustainability (comprising sustainable growth and environmental protection), multi-disciplinarity as a default approach to AI, and the global reach and impact of AI systems.

When considering data protection legislation from countries without dedicated AI principles, there is strong alignment with OECD Principle 2 (human-centred values and fairness) and Principle 3 (transparency and explainability). In line with recent developments in other parts of the world (e.g. the General Data Protection Regulation in Europe), the latest data protection laws in LAC include safeguards against bias and unfairness, and promote the explainability of automated decision-making. This is the case for Barbados, Brazil, Ecuador, Jamaica, Panama and Peru. However, these data protection laws are not specific to AI and neglect certain aspects that more nuanced and targeted instruments such as AI ethical frameworks and principles seek to address. For instance, these laws generally do not account for the options open to individuals to contest or appeal decisions based on automated processes, nor do they consider how AI developments could support or hinder the achievement of societal goals. In addition, since they are focused on data protection, they are limited in the extent to which they consider the downstream uses of data, such as for machine learning algorithms. There may be opportunities to review current data protection laws in light of the growing number of ways that data can be used for purposes such as algorithms and automated

decision making. This implies that current legislation may need to be updated or supplemented (e.g. with AI-specific frameworks) in order to capture the new opportunities and challenges posed by AI technologies.

Colombia's Ethical Framework for AI serves as a good example in the region (Box 4.3), as it explicitly touches on all areas included in the OECD AI Principles, to which the country has adhered, while also grounding the framework in Colombia's own context and culture. Outside the LAC region, Spain's Charter on Digital Rights serves as a strong human-centred mechanism that in a manner relevant and appropriate for the country, seeks to "transfer the rights that we already have in the analogue world to the digital world and to be able to add some new ones, such as those related to the impact of artificial intelligence" (Nadal, 2020^[9]) (Box 4.4). While extending beyond Artificial Intelligence, the Charter includes important AI principles and requirements that are uniquely framed around public rights.

Box 4.3. Colombia's Ethical Framework for AI

The Ethical Framework for AI was designed as a response to the ethical implications of increased implementation of AI technologies in Colombia and to ignite a discussion around the desired social boundaries of their use. It touches upon moral problems related to data (including their generation, recording, adaptation, processing, dissemination and use), algorithms and their corresponding practices (including, responsible innovation, programming, hacking and professional codes).

The Framework consists of principles and implementation tools for public and private entities to implement AI. The table illustrates how the implementation tools interact with the principles.

	Transparency	Explanation	Privacy	Human control	Security	Responsibility	Non-discrimination	Inclusion	Youth rights	Social benefit
Algorithm assessment	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Algorithm auditing	✓	✓	✓		✓					
Data cleansing			✓		✓	✓	✓			
Smart explanation	✓	✓					✓			
Legitimacy evaluation	✓			✓			✓	✓	✓	✓
Sustainable system design					✓	✓				
Risk management			✓	✓	✓	✓				
Differential policy			✓		✓					
Codes of conduct	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
AI ethics research	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Privacy impact assessments			✓		✓		✓			
Ethical data approaches			✓	✓	✓	✓	✓			
Personal data stores			✓							
Strengthen business ethics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Several versions of the Ethical Framework were released for consultation by means of triggering multi-sectorial debate and suggestions which were incorporated into the final iteration of the document.

The final version of the Ethical Framework was launched on 12 October 2021 and can be consulted here: <https://bit.ly/3EC7wJy>.

Source: (Guío Español, 2020^[10]).

Box 4.4. Charter on Digital Rights (Spain)

Spain's Ministry of Economic Affairs and Digital Transformation and its Secretary of State for Digitalisation and Artificial Intelligence (see Box 5.1) has developed a Charter on Digital Rights to meet one of its commitments under the Spain Digital 2025 strategy.

The draft Charter included 28 sets of rights, many of which relate directly to the ethical and trustworthy development and use of AI systems and underlying data. The most relevant falls under Article 25, "Rights with regards to AI":

- Artificial intelligence shall ensure a people-centred approach safeguarding the inalienable dignity of every individual, shall pursue the common good, and ensure compliance with the principle of non-maleficence.
- In the development and life cycle of artificial intelligence systems:
 - The right to non-discrimination, irrespective of origin, cause, or nature, shall be guaranteed concerning decisions, use of data and processes based on artificial intelligence.
 - Conditions of transparency, auditability, explainability, traceability, human oversight and governance shall be established. In any case, the information provided shall be accessible and understandable.
 - Accessibility, usability and reliability shall be ensured.
- Individuals have the right to request human oversight and intervention and the right to challenge automated decisions made by artificial intelligence systems that impact their personal sphere or patrimony.

The Charter was developed in co-ordination with a variety of experts and relevant stakeholders, and was subject to an open public consultation that closed on 20 January 2021. The government reviewed and evaluated the comments received and finalised the charter in July 2021.

Source: www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf and https://portal.mineco.gob.es/es-es/ministerio/participacionpublica/audienciapublica/Paginas/SEDIA_Carta_Derechos_Digitales.aspx.

In addition to the development of AI principles, some LAC countries are seeking complementary approaches to ethical and trustworthy AI, though perhaps in a less explicit, detailed or mature manner than discussed above:

- Argentina's AI strategy includes an "Ethics and Regulation" transversal axis that pledges to "Guarantee the development and implementation of AI according to ethical and legal principles, in accordance with fundamental rights of people and compatible with rights, freedoms, values of diversity and human dignity." It also seeks to promote the development of AI for the benefit, well-being and empowerment of people, as well as the creation of transparent, unbiased, auditable, robust systems that promote social inclusion. Although the strategy does not define an ethical framework, it creates two bodies responsible for leading the design of such instruments: the AI National Observatory and the AI Ethics Committee.⁷ Argentina further pledges to "promote guidelines for the development of reliable AI that promote, whenever pertinent, human determination in some instance of the process and the robustness and explicability of the systems". It also considers the importance of a "risk management scheme that takes into account security,

protection, as well as transparency and responsibility, when appropriate, beyond the rights and regulations in force that protect the well-being of people and the public”. Finally, it recognises that it may not be appropriate to use AI systems when the following standards are not met: transparency, permeability, scalability, explicability, bias mitigation, responsibility, reliability and impact on equity and social inclusion.

- As noted above, Brazil’s AI strategy commits to developing AI principles. The strategy itself also has a strong focus on ethics, with considerations woven throughout the document. For instance, it includes a cross-cutting thematic axis on “legislation, regulation and ethical use”, and commits to “shar[ing] the benefits of AI development to the greatest extent possible and promote equal development opportunities for different regions and industries”. It also includes actions to develop ethical, transparent and accountable AI; ensure diversity in AI development teams with regard to “gender, race, sexual orientation and other socio-cultural aspects”; and commits to develop techniques to detect and eliminate bias, among other actions included in Annex B.
- Chile’s AI policy includes a section dedicated to ethical considerations and measures, with associated actions detailed in the AI Action Plan. Specific activities include conducting an ethics study, developing a risk-based system for categorizing AI systems, ensuring the agreement of national best practices for ethical AI and developing an institution to supervise AI systems, among others. Interestingly, the policy and Action Plan also call for adapting school curricula to include education on technology ethics.
- In its digital strategy, Panama envisions a co-operation agreement with IPANDETEC (the Panama Institute of Law and New Technologies) for the promotion of human rights in the digital context.⁸
- Peru’s 2021 draft national AI strategy includes a cross-cutting pillar on ethics and a strategic objective to become a regional leader in the responsible use of data and algorithms. It also commits to country-specific implementation of the OECD AI Principles, to which Peru adheres, and the creation of a unit to monitor and promote the responsible and ethical use of AI in the country. The draft further envisions the development of country-specific “ethical guidelines for sustainable, transparent and replicable use of AI with clear definitions of responsibilities and data protection”. In addition, the country’s *Digital Trust Framework* mandates the ethical use of AI and other data-intensive technologies: “Article 12.2 – Public entities and private sector organisations promote and ensure the ethical use of digital technologies, the intensive use of data, such as the Internet of Things, Artificial Intelligence, data science, analytics and the processing of large volumes of data”.⁹ However, it does not explain what is understood as ethical or a more precise set of applicable principles, although Peru adheres to the OECD AI Principles, implying that these might serve as the criteria.

In seeking to implement and operationalise high-level principles and ensure a consistent approach across the public sector, only Mexico and Uruguay have issued guidelines to assess the impact of algorithms in the public administration. Uruguay’s digital agency, AGESIC, has elaborated the *Algorithmic Impact Study Model*, a set of questions that can be used by project managers across the public sector to evaluate and discuss the risks of systems using machine learning. Mexico has published the *Impact Analysis Guide for the Development and Use of Systems Based on Artificial Intelligence in the Federal Public Administration*. As with the AI strategy and principles, this guide was developed by Mexico’s former administration and the state of top-level support for implementation is not clear. Box 4.5 presents further information about both guides.¹⁰ Such mechanisms can help to materialise many aspects of building a trustworthy approach, including items discussed later in this section.

Box 4.5. Existing LAC guidelines to assess the impact of algorithms in public administrations

Algorithmic Impact Study Model (Uruguay)

The Algorithmic Impact Study (EIA) was designed by AGESIC, the public digital agency of Uruguay, as a tool for analysing automated decision support systems that use machine learning. Aimed at project managers or teams that lead AI projects, EIA is designed to identify key aspects of systems that merit more attention or treatment. The model consists of a set of questions that evaluate different aspects of systems including the underlying algorithm, the data and their impacts. Users can then share, analyse and evaluate the results. The questionnaire is structured as follows:

- Brief description of the project.
- Project outcome or objective.
- Social impact.
- About the system.
- About the algorithm.
- About the decision.
- Impact evaluation of the automated decision system.
 - About the data.
 - Origin of the data of the automated decision system.
 - Automated decision system data types.
 - Stakeholders of the automated decision system.
 - Actions to reduce and mitigate the risks of the automated decision system.
 - Data quality.
 - Procedural fairness.

Source: (AGESIC, 2020^[11]).

Impact Analysis Guide for the development and use of systems based on AI in the Federal Public Administration (Mexico)

The Impact Analysis Guide is a tool designed to determine the societal and ethical reach of AI systems developed by the Federal Public Administration, and to define safeguards in accordance with their potential impacts. It is based on Canada's Directive on Automated Decision Making and its associated Algorithmic Impact Assessment.

The guide presents an initial questionnaire that analyses five dimensions:

- The use and management of data.
- The process.
- The level of autonomy and functionality of the system.
- Its impact on individuals, companies and society.
- Its impact on government operations.

Each question generates a score to which a multiplier is added depending on the number of areas where it has an effect (physical or mental impact, user experience, standards and regulations, objectives/goals, operation, reputation). This process produces a score for each of the five dimensions and an overall impact level.

Depending on the overall score and the resulting impact for each dimension, the guide assigns a total “level of impact” to the system on a scale of 1 to 4. According to each level, the system must fulfil a set of requirements before, during and after implementation. For example, AI systems where two or more dimensions, including socio-economic scope, have high or very high impact are assigned to level IV. Level IV systems must meet the following requirements:

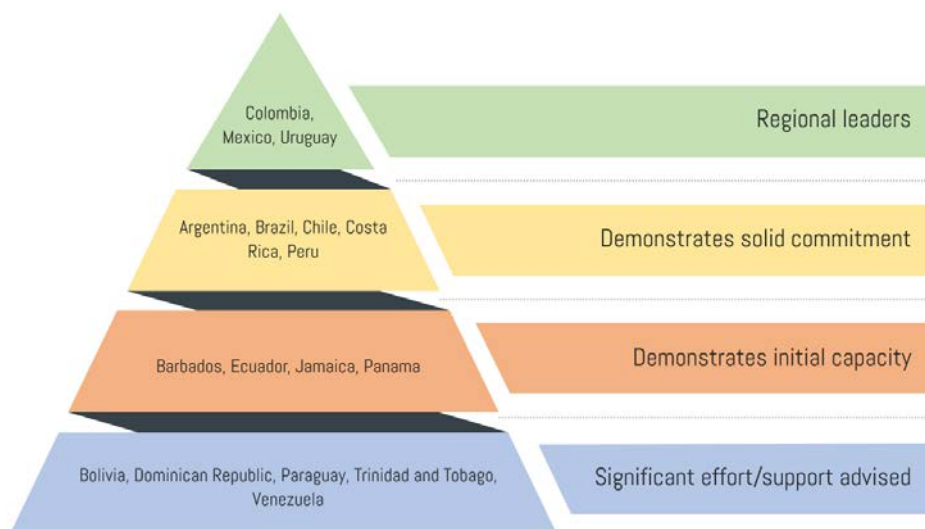
- Before implementation.
 - Register the system with the Digital Government Unit (UGD), including a clear and complete description of its function, objectives and expected impact.
 - Submit a report to the UGD detailing ethical concerns, risks and possible unplanned uses of the system.
 - Allocate resources to research the impact and implications of using the system.
- During implementation.
 - Conduct quarterly robustness, reliability and integrity testing of the system and model.
 - Publish information (variables, metadata) about the data used in the training of an algorithm and the methodology for the design of the model.
 - Share with users and the public a clear and complete description of the model and its expected impact.
- After implementation.
 - Provide a meaningful, clear and timely explanation to users about how and why the decision was made (include variables, logic and technique).
 - Publish information on the effectiveness and efficiency of the system every six months.

Source: (Coordinación de la Estrategia Digital Nacional, 2018^[12]).

Fairness and mitigating bias

While data and algorithms are the essence of modern AI systems, they can create new challenges for policy makers. Inadequate data lead to AI systems that recommend poor decisions. If data reflect societal inequalities, then applying AI algorithms can reinforce them, and may distort policy challenges and preferences (Pencheva, Esteve and Mikhaylov, 2018^[13]). If an AI system has been trained on data from a subset of the population that has different characteristics from the population as a whole, then the algorithm may yield biased or incomplete results. This could lead AI tools to reinforce existing forms of discrimination, such as racism and sexism.¹¹

Figure 4.5. LAC region capacities for establishing safeguards against bias and unfairness



Note: All countries that adhere to the OECD AI Principles have been categorised as “demonstrates solid commitment” or higher.

All the LAC countries that have adhered to the OECD AI Principles have demonstrated a strong commitment to fairness, non-discrimination and prevention of harm (Principle 2). This principle also constitutes a strong focus of LAC countries’ self-developed principles and data protection laws. Some of the most explicit aspects of these principles are as follows:

- As part its Ethical Framework for Artificial Intelligence, Colombia developed a monitoring Dashboard available for free to all citizens. The dashboard provides information about the use of AI systems across the country and implementation of the ethical principles of Artificial Intelligence in AI projects by public entities.
- Colombia, Mexico and Uruguay have established a clearer role for humans in terms of maintaining control of AI systems, resolving dilemmas and course correcting when necessary.
- Uruguay’s “General Interest” principle aligns with OECD principles 1 and 2. The first part of the principle sets a social goal, namely, protecting the general interest, and guaranteeing inclusion and equity. The second part states that “work must be carried out specifically to reduce the possibility of unwanted biases in the data and models used that may negatively impact people or favour discriminatory practices”.
- Chile’s Inclusive AI principle calls for no discrimination or detriment to any group, and emphasises consideration of children and teenagers and the need for a gender perspective, which can be compared to the gender sub-axis in the country’s AI Policy. The country’s AI strategy and action plan calls for continuous discussions across sectors about bias as well as the development of recommendations and standards regarding bias and transparency in algorithms.
- The data protection legislation of Barbados, Brazil, Jamaica, Panama and Peru includes safeguards against automated decision making and profiling that may harm the subject or infringe upon their rights. The right to not be subject to automated decision making is shared by these countries. This may apply when automated data processing leads to decisions based on or that define the individual’s performance at work, aspects of their personality, health status, creditworthiness, reliability and conduct, among others. In the case of Ecuador, although the *Guide for the Processing of Personal Data in the Central Public Administration* does not have the same

legal standing as data protection legislation, it stipulates that personal data treatment by the central public administration cannot originate discrimination of any kind (Art. 8).

Aside from aspects included in country-specific principles and data protection laws, LAC countries are establishing safeguards against bias and unfairness. Efforts that show strong potential include the following:

- Argentina’s AI Strategy recognises the risk of bias in AI systems as part of its diagnostic of the ‘Ethics and Regulation’ transversal axis, although no specific measures are explained.
- Brazil’s national AI strategy includes action items to develop techniques to identify and mitigate algorithmic bias and ensure data quality in the training of AI systems, to direct funds towards projects that support solutions that support fairness and non-discrimination, and to implement actions to support diversity in AI development teams. It also commits to developing approaches to reinforce the role of humans in a risk-based manner.
- Chile’s AI Policy proposes the creation of new institutions capable of establishing precautionary actions directed at AI. It proposes fostering research on bias and unfairness, while a unique gender element evaluates how to reduce gender-related biases, and highlights the production of biased data and development teams with little diversity. Relevant actions include:
 - Actively promoting the access, participation and equal development of women in industries and areas related to AI.
 - Working with research centres to promote research with a gender perspective in areas related to AI.
 - Establishing evaluation requirements throughout the entire cycle or life of AI systems to avoid gender discrimination.
- Colombia’s Centre for the Fourth Industrial Revolution, established by the government and the World Economic Forum (WEF), leads a project to generate comprehensive strategies and practices oriented towards gender neutrality in AI systems and the data that feed them.¹²
- Peru’s 2021 draft national AI strategy envisions the collaboration of public sector organisations to conduct an impact study on algorithmic bias and to identify ways to lessen such bias in algorithms that involve the classification of people. However, the scope of this effort appears to be limited to private sector algorithms. In addition, the strategy mandates that all public sector AI systems related to the classification of people (e.g. to provide benefits, opportunities or sanctions) must undergo a socioeconomic impact study to guarantee equity.
- Uruguay has released two relevant instruments to address bias and unfairness. The *Framework for Data Quality Management*¹³ includes a set of tools, techniques, standards, processes and good practices related to data quality. More specifically on AI, the *Algorithmic Impact Study Model* (see Box 4.5) references questions to evaluate and discuss the impacts of automated decision-making systems. The section *Measures to reduce and mitigate the risks of the automated decision system* (p. 8) includes various questions designed to mitigate bias. *The Social Impact* (p. 4) and *Impact evaluation of the automated decision system* (p. 6) sections aim to help development teams evaluate if their algorithms might lead to unfair treatment.

It should not be assumed that AI bias is an inevitable barrier. Improving data inputs, building in adjustments for bias and removing variables that cause bias may make AI applications fairer and more accurate. As it discussed earlier, codified principles and newer data protection laws are impacting how AI systems process personal data. Legislation is one option that can help address these issues and mitigate associated risks. Developing laws in this area may be a particularly useful approach in LAC countries, where the OECD has observed a strong legal focus and attention to meeting the exact letter of the law

(OECD, 2018^[14]) (OECD, 2019^[8]). While such an approach can promote trust, it can also quickly become outdated and hinder innovation or discourage public servants from exploring new approaches. Another approach involves creating agile frameworks that adopt necessary safeguards for the use of data-intensive technologies but remain adaptable and promote experimentation.

Moving ahead, LAC governments will need to couple high-level principles with specific controls and evolving frameworks and guidance mechanisms to ensure that AI implementation is consistent with principles and rules. The algorithmic impact assessments discussed earlier represent a step in the right direction (Box 4.5). Countries outside the region have also developed some examples that go beyond strategy pledges and principles. For instance, the UK government recognises that data on issues that disproportionately affect women are either never collected or of poor quality. In an attempt to reduce gender bias in data collection, it has developed a government portal devoted to gender data (OECD, 2019^[15]).¹⁴ The existence of an independent entity also facilitates progress, particularly with regard to testing ideas, setting strategies and measuring risks, as in the case of the Government of New Zealand's Data Ethics Advisory Group (Box 4.6).

Box 4.6. New Zealand: Data Ethics Advisory Group

In order to balance increased access and use of data with appropriate levels of risk mitigation and precaution, the government chief data steward in New Zealand founded the Data Ethics Advisory Group, the main purpose of which is to assist the New Zealand government in understanding, advising and commenting on topics related to new and emerging uses of data.

To ensure the advisory group delivers on its mandate, the government chief data steward has appointed seven independent experts from different areas relevant to data use and ethics as members, including experts in privacy and human rights law, technology and innovation.

The group discusses and comments solely on subjects and initiatives related to data use, not broader digital solutions provided by public bodies. Examples of topics that the Data Ethics Advisory Group might be requested to comment on include the appropriate use of data algorithms (e.g. how to avoid algorithmic bias) and the correct implementation of data governance initiatives.

Source: (OECD, 2019^[15]), www.data.govt.nz/about/government-chief-data-steward-gclds/data-ethics-advisory-group and <https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faiipo.oecd.org%2F2021-data-policyInitiatives-25665>.

A subset of AI systems that have been particularly contentious with regard to bias is facial recognition. Such systems can also have an inherent technological bias (e.g. when based on race or ethnic origins) (OECD, 2020^[16]). As discussed in Chapter 3 of this report, facial recognition represents a very small but growing use case for AI in LAC governments. For instance, Ecuador officials told the OECD that they are exploring a facial recognition identity program for access to digital services. Governments and other organisations are designing frameworks and principles to help guide others as they explore this complex field. A relevant example that may be useful for LAC countries is the Safe Face Pledge, which focuses on facial biometrics (Box 4.7).

Box 4.7. The Safe Face Pledge

The Safe Face Pledge was a joint project of the Algorithmic Justice League and the Center on Privacy & Technology at Georgetown Law in Washington, DC. It served as a means for organisations to make

public commitments towards mitigating the abuse of facial analysis technology. It included four primary commitments:

- **Show Value for Human Life, Dignity and Rights.**
 - Do not contribute to applications that risk human life.
 - Do not facilitate secret and discriminatory government surveillance.
 - Mitigate law enforcement abuse.
 - Ensure your rules are being followed.
- **Address Harmful Bias.**
 - Implement internal bias evaluation processes and support independent evaluation.
 - Submit models on the market for benchmark evaluation where available.
- **Facilitate Transparency.**
 - Increase public awareness of facial analysis technology use.
 - Enable external analysis of facial analysis technology on the market.
- **Embed Commitments into Business Practices.**
 - Modify legal documents to reflect value for human life, dignity and rights.
 - Engage with stakeholders.
 - Provide details of Safe Face Pledge implementation.

The Safe Face Pledge was sunset in February 2021; however, its general principles remain relevant.

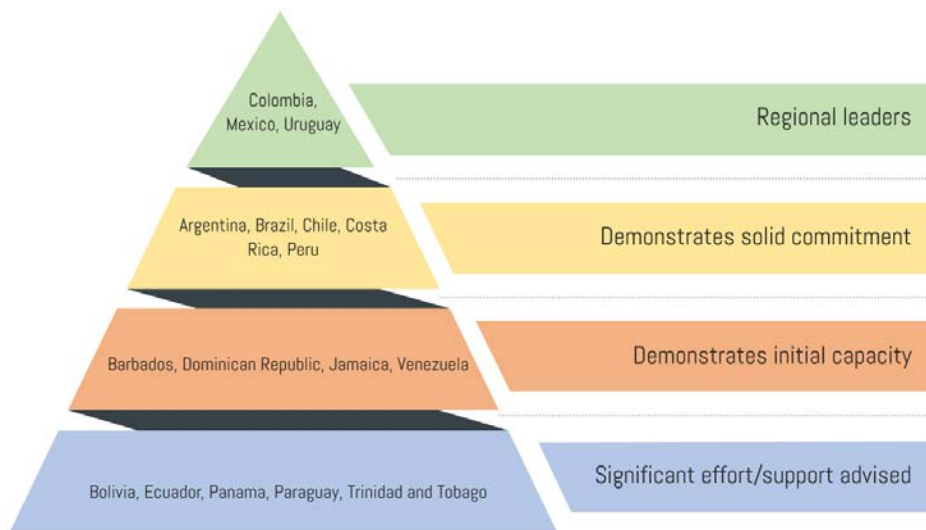
Source: www.safefacepledge.org/pledge.

Other factors are related to mitigating bias and ensuring fairness. In the field of AI, diverse and inclusive teams working on product ideation and design can help prevent or eliminate possible biases from the start (Berryhill et al., 2019^[17]), notably those related to data and algorithmic discrimination. The section *Ensuring an inclusive and user-centred approach* later in this chapter explores this issue in greater detail.

Transparency and explainability

An important component of a trustworthy AI system is its capacity to explain its decisions and its transparency for the purposes of external evaluation (Berryhill et al., 2019^[17]). In the case of PretorlA (Colombia) (Box 3.3), the Constitutional Court decided to make the explainability of this new system a top priority, on the basis that it could influence judicial outcomes through interventions in the selection process of legal complaints. Conversely, in Salta, Argentina, the algorithm designed to predict teenage pregnancy and school dropout (Bx 3.14) was opaquer, leading to uncertainty about how it was reaching its conclusions. This feature contributed to civil society scrutiny and a lack of trust in subsequent years. Overall, as part of analysis of these use cases, this study found low availability of information concerning the deployment, scope of action, status and internal operation of AI systems in the public sector.

Figure 4.6. LAC region capacities for considering the explainability of AI systems and automated decision making



Note: All countries that adhere to the OECD AI Principles have been categorised as “demonstrates solid commitment” or higher.

LAC countries are working in different ways to ensure the transparency of AI systems and decisions. Countries that have developed AI principles and ethical frameworks generally present strong alignment with OECD AI Principle 3 (Transparency and explainability). Uruguay’s principles represent a slight exception here as they consider transparency but make no mention of explainability. However, inclusion of the expression “active transparency” could open the principle up to broader interpretation. However, Uruguay’s Algorithmic Impact Study (EIA) does consider explainability. Other efforts include the following:

- Colombia’s Ethical Framework for AI includes two relevant implementation tools: an *algorithm assessment* which enables constant mapping of public sector AI systems to assess how ethical principles are being implemented, and an *intelligent explanation* model which provides citizens with understandable information about AI systems.
- Mexico’s AI Principles require that users have explained to them the decision-making process of the AI system as well as expected benefits and potential risks associated with its use. The principles also foster transparency through the publication of information allowing users to understand the training method and decision-making model of the system, as well as the results of its evaluations.

Most recent data protection legislation also extends traditional access rights by requiring greater transparency with regard to the methods and processes involved in automated decision making. For Barbados and Jamaica, right of access includes the right to know about the existence of automated decision making, as well as the algorithmic processes. Barbados further extends this right to include “the significance and the envisaged consequences”. Brazil confers access to information on the form, duration and performance of the treatment of personal data. When automated decision making is in place, subjects may access information regarding the criteria and procedures, in compliance with trade and industrial secrets.

Countries are also developing approaches to increase transparency and explainability beyond formal frameworks and laws. Such approaches include the following:

- As part of its “Ethics and Regulation” transversal axis, Argentina’s AI Strategy states that “developments that tend towards Explainable Artificial Intelligence (Explainable AI or “XAI”)

should be promoted, in which the result and the reasoning for which an automated decision is reached can be understood by human beings”. However, no specific measures are discussed.

- Brazil’s national AI strategy commits to directing funds toward projects that support transparency, and to put in place supervisory mechanisms for public scrutiny of AI activities.
- Chile’s national AI strategy and action plan provide a number of considerations for the transparency and explainability of AI systems, notably developing standards and good practices that can be adapted as the concept is better understood over time, promoting new explainability techniques and conducting research in this area. This process includes establishing standards and transparency recommendations for critical applications.
- The Dominican Republic developed a Digital Government guide¹⁵ that includes a provision on the documentation and explainability of digital government initiatives, software, services, etc. However, specific guidelines for algorithmic transparency and explainability are not provided.
- Peru’s draft 2021 national AI strategy envisions the development of a registry of AI algorithms used in the public sector and the underlying datasets used in public sector AI systems. It is unclear whether the registry would be open to the public.
- Uruguay’s AI strategy promotes the transparency of algorithms through two interrelated actions: the definition of “standards, guidelines and recommendations for the impact analysis, monitoring and auditing of the decision-making algorithms used in the [public administration]”; and the establishment of “standards and procedures for the dissemination of the processes used for the development, training and implementation of algorithms and AI systems, as well as the results obtained, promoting the use of open code and data”.
- Venezuela’s Info-government Law defines a principle of technological sovereignty, which mandates that all software adopted by the state should be open and auditable. For instance, Article 35 states that “Licenses for computer programs used in the public administration must allow access to the source code and the transfer of associated knowledge for its compression, its freedom of modification, freedom of use in any area, application or purpose, and freedom of publication and distribution of the source code and its modifications”.¹⁶

While countries have made a number of commitments, most have not yet been implemented in a manner that makes them actionable. Box 4.8 provides an example from outside the LAC region showing how a government has approached this challenge.

Box 4.8. Guidance transparency and explainability of public AI algorithms (France)

Etalab, a Task Force under the French Prime Minister’s Office, has produced a guide for public administrations on the responsible use of algorithms in the public sector. The guide sets out how organisations should report on their use to promote transparency and accountability. The guidance covers three elements:

- **Contextual elements.** These focus on the nature of algorithms, how they can be used in the public sector, and the distinction between automated decisions and cases where algorithms function as decision-supporting tools.
- **Ethics and responsibility of using algorithms to enhance transparency.** This includes public reporting on the use of algorithms, how to ensure fair and unbiased decision making, and the importance of transparency, explainability and trustworthiness.
- **A legal framework for transparency in algorithms** including the European Union’s General Data Protection Regulation (GDPR) and domestic law. This includes a set of rules to be applied

to administrative decision-making processes on what specific information must be published about public algorithms.

Etalab also proposes six guiding principles for the accountability of AI in the public sector:

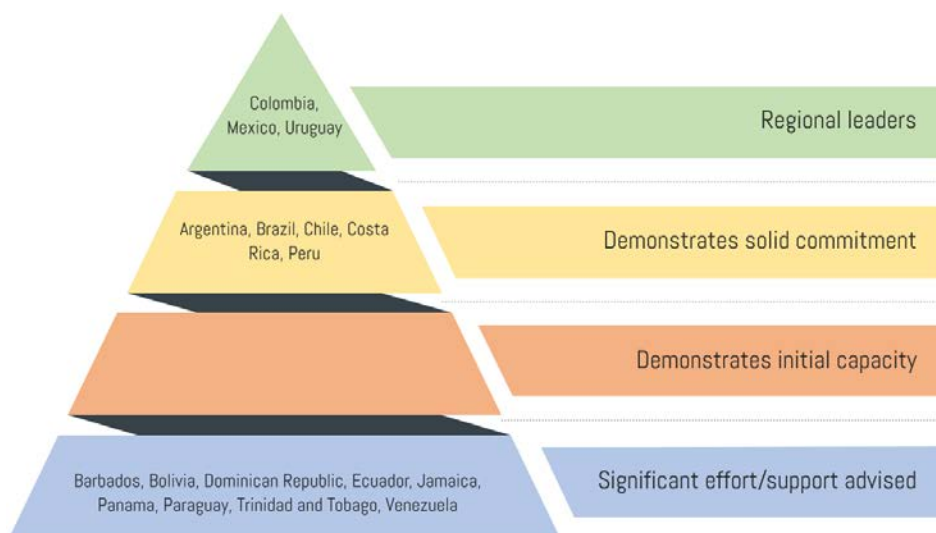
1. **Acknowledgment:** agencies are obligated to inform interested parties when an algorithm is used.
2. **General explanation:** agencies should provide a clear and understandable explanation of how an algorithm works.
3. **Individual explanation:** agencies ought to provide a personalised explanation of a specific result or decision.
4. **Justification:** agencies should justify why an algorithm is used and reasons for choosing a particular algorithm.
5. **Publication:** agencies should publish the source code and documentation, and inform interested parties whether or not the algorithm was built by a third party.
6. **Allow for contestation:** agencies should provide ways of discussing and appealing algorithmic processes.

Source: www.etalab.gouv.fr/datasciences-et-intelligence-artificielle; www.etalab.gouv.fr/how-etalab-is-working-towards-public-sector-algorithms-accountability-a-working-paper-for-rightscon-2019/, <https://etalab.github.io/algorithmes-publics> and www.europeandataportal.eu/fr/news/enhancing-transparency-through-open-data; www.etalab.gouv.fr/algorithmes-publics-etalab-public-un-guide-a-lusage-des-administrations (Berryhill et al., 2019_[17]).

Safety and security

This section examines how and to what extent LAC countries are establishing measures to develop and use safe and secure AI systems. As stated in the OECD AI Principles, “AI systems should be robust, secure and safe throughout their entire life cycle so that, in conditions of normal use, foreseeable use or misuse, or other adverse conditions, they function appropriately and do not pose unreasonable safety risk.¹⁷ Such AI systems may involve the application of a risk management approach, such as the development of an algorithmic impact assessment process, ensuring the traceability of processes and decisions, and providing clarity regarding the (appropriate) role of humans in these systems (Berryhill et al., 2019_[17]).¹⁸

Figure 4.7. LAC region capacities for promoting safety and security in public sector AI systems



Note: All countries that adhere to the OECD AI Principles have been categorised as “demonstrates solid commitment” or higher.

Adherence by LAC countries to the OECD AI principles can be interpreted as a solid commitment to safety and security. Countries in the region are also taking additional measures to ensure AI systems are safe and secure. Those that have developed national AI strategies and country-specific AI principles often emphasise the safety, security and robustness of AI systems in those principles. For instance:

- Argentina’s AI strategy commits to the creation of an ethical framework including a risk management scheme that takes into account security, protection, transparency and responsibility, with a view to protecting the well-being of people and the public.
- Chile’s AI Policy incorporates a focus on AI safety including through risk and vulnerability assessments and the enhancement of cybersecurity, with a specific goal to “position AI as a relevant component of the cybersecurity and cyber defence field, promoting secure technological systems”.
- Colombia’s Ethical Framework for AI proposes safety mechanisms such as the immutability, confidentiality and integrity of base data, and the establishment of codes of conduct and systems of risk to identify possible negative impacts. It seeks to ensure that “Artificial intelligence systems must not affect the integrity and physical and mental health of the human beings with whom they interact” (p. 34).
- Mexico’s Impact Analysis Guide for the development and use of systems based on AI¹⁹ provides a detailed set of principles on safety related to the mitigation of risks and uncertainty, design and implementation phases, and mechanisms for user data protection.
- Uruguay’s AI principles state that “AI developments must comply, from their design, with the basic principles of information security”. The country’s Algorithmic Impact Study Model helps set up a risk-based approach to AI safety and security and also includes guidelines to clarify the role of humans in algorithmic decision making.

Box 4.9. Evaluating the human role in algorithmic decision making (Uruguay)

The Algorithmic Impact Study model of Uruguay allows digital government teams to evaluate the role of humans in algorithmic decision making, prompting ethical discussions on this topic. Although the model does not clarify the appropriate role of humans in decision making, its guiding questions help public sector teams to evaluate current or proposed algorithms in the light of safety and accountability principles (see next section), and to decide which features to incorporate. The following selected questions from the Algorithmic Impact Study model refer to safety and accountability:

Impact evaluation of the automated decision system

1. Will the system only be used to help make decisions in the context of this project? (Yes or no)
2. Will the system replace a decision that would otherwise be made by a human? (Yes or no)
3. Will the system automate or replace human decisions that require judgment or discretion? (Yes or no)
4. Are the effects resulting from the decision reversible?
 - a. Reversible.
 - b. Probably reversible.
 - c. Difficult to reverse.
 - d. Irreversible.

Procedural fairness

1. Does the audit trail identify who is the authorised decision maker? (Yes or no)
2. Is there a process to grant, monitor and revoke access permissions to the system? (Yes or no)
3. Is there a planned or established appeal process for users who wish to challenge the decision? (Yes or no)
4. Does the system allow manual override of your decisions? (Yes or no)

Source: (AGESIC, 2020^[11])

Brazil is the only LAC country without country-specific AI principles that includes objectives in other laws that are aligned with the OECD AI Principles in this area. In particular, the national data protection law incorporates a “prevention principle” calling for the adoption of measures to prevent damage caused by the processing of personal data. In addition, the country’s recent national AI strategy commits to actions that ensure human review and intervention in high-risk activities and commits to directing funds towards projects that support accountability in AI systems.

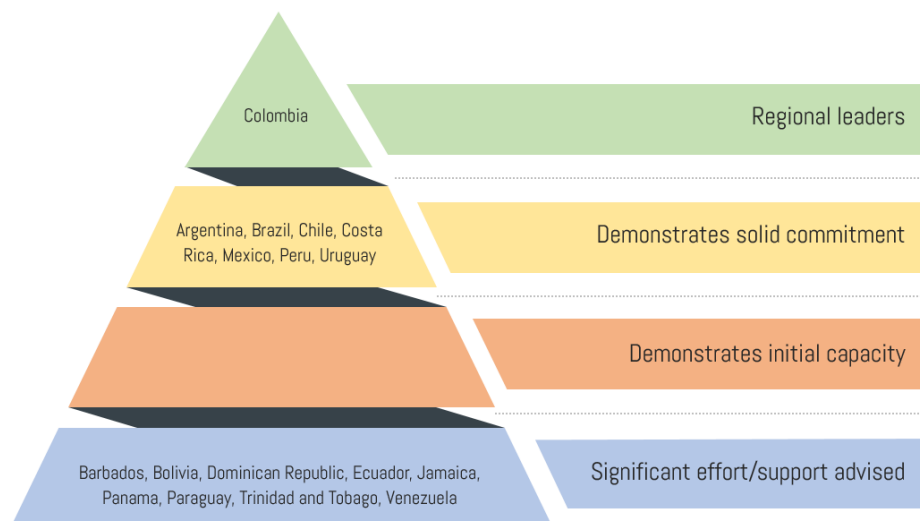
Accountability

This section examines the extent to which accountability mechanisms are present and operational in LAC countries, and ensure the proper and appropriate functioning of systems. Accountability is an important principle that cuts across the others and refers to “the expectation that organisations or individuals will ensure the proper functioning, throughout their lifecycle, of the AI systems that they design, develop, operate or deploy, in accordance with their roles and applicable regulatory frameworks, and for demonstrating this through their actions and decision-making process”.²⁰ For instance, accountability

measures can ensure that documentation is provided on key decisions throughout the AI system life cycle and that audits are conducted where justified. OECD work has found that in the public sector this involves developing open and transparent accountability structures and ensuring that those subject to AI-enabled decisions can inquire about and contest those decisions (as seen in Box 4.8) (Berryhill et al., 2019^[17]).

It is essential for LAC government pursuing AI to develop the necessary guidelines, frameworks or codes for all relevant organisations and actors to ensure accountable AI development and implementation.

Figure 4.8. LAC region capacities for promoting accountability in public sector AI systems



Note: All countries that adhere to the OECD AI Principles have been categorised as “demonstrates solid commitment” or higher.

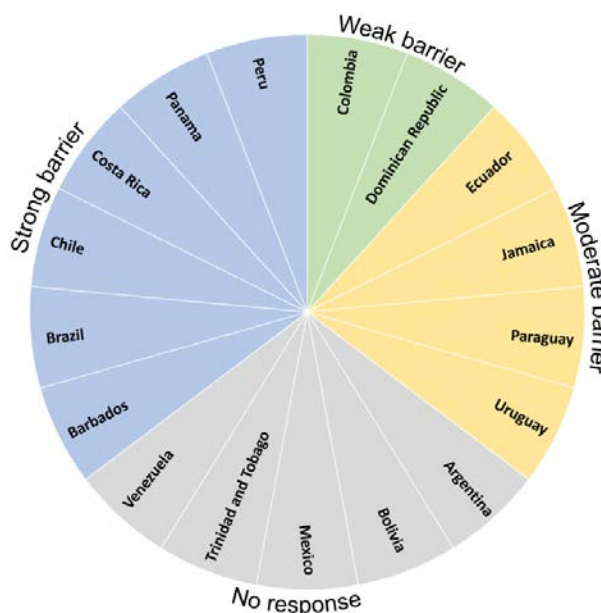
Adherence by LAC countries to the OECD AI principles can be interpreted as a solid commitment to this issue. Countries in the region are also taking additional measures to ensure AI systems are accountable, but to a somewhat lesser extent compared to certain other topics reviewed elsewhere in this chapter. Only Colombia, Mexico and Uruguay have integrated accountability into national AI strategies or principles, although clear evidence of implementation is not available in most cases. The following examples are particularly noteworthy:

- Chile’s national AI strategy includes a goal to “Develop the requirements for prudent development in an agile way and responsible use of AI”, including through the creation of an institution that can supervise AI systems at different stages of their life cycle. It also calls for organisations to have clearly defined roles and responsibilities to ensure lines of responsibility.
- Colombia’s Ethical Framework for AI states that there “is a duty to respond to the results produced by an Artificial Intelligence system and the impacts it generates”. It also establishes a duty of responsibility for entities that collect and process data and for those who design algorithms. It also recommends defining clear responsibilities for the chain of design, production and implementation of AI systems.
- Mexico’s AI Principles integrate accountability by highlighting the importance of determining responsibilities and obligations across the whole life cycle of an AI system.
- Peru’s 2021 draft national AI strategy envisions the adoption of ethical guidelines that include clear definitions of responsibilities.
- Uruguay’s AI principles include a requirement that technological solutions based on AI must have a clearly identifiable person responsible for the actions derived from the solutions.

Brazil is the only LAC country without country-specific AI principles that includes objectives in other laws that are aligned with the OECD AI Principles in this area. In particular, the national data protection law includes a responsibility and accountability objective which requests that data processors adopt measures in order to comply effectively with the data protection law, thereby ensuring accountable actors are in place. These objectives represent a solid step towards implementing the OECD’s previous recommendation that Brazil develop “transparency mechanisms and ethical frameworks to enable a responsible and accountable adoption of emerging technologies solutions by public sector organisations” (OECD, 2018^[14]). In addition, the country’s recent national AI strategy commits to directing funds towards projects that support accountability in AI systems.

The common absence of legal or methodological guidance on accountability coincides with a major perception among LAC countries that lack of clarity regarding checks and balances/accountability for data-driven decision making represents a strong or moderate barrier to the use of data in the public sector (Figure 4.9). While this is not specific to AI in the public sector, the concepts are related.

Figure 4.9. Lack of clarity regarding checks and balances/accountability for data-driven decision making as a barrier



Source: OECD LAC Digital Government Agency Survey (2020).

Finally, monitoring during the implementation stage is vital to ensure that AI systems operate as intended in accordance with the OECD AI Principles, and that organisations are accountable in this regard. Related to the topic of safety and security discussed in the previous sub-section, such monitoring should ensure that risks are mitigated and that unintended consequences are identified. A differentiated approach will be required to focus attention on AI systems where the risks are highest – for instance, where they influence the distribution of resources or have other significant implications for citizens (Mateos-Garcia, 2018^[18]). For the most part, LAC countries have not developed these types of monitoring mechanisms, with the exception of efforts being undertaken by Colombia (Box 4.10). Such mechanisms may represent the next stage of development for regional leaders once efforts to build ethical frameworks and enabling inputs have solidified.

Box 4.10. Monitoring AI in Colombia

Colombia is developing policy intelligence tools to monitor the implementation of i) national AI policies, ii) emerging good practices to implement the OECD AI recommendations to governments, and iii) AI projects in the public sector:

- SisCONPES is a tool that monitors the implementation of every action line in the AI national strategy. It reports advances and challenges in implementation to entities leading implementation of the strategy, notably the Presidency of the Republic.
- A follow-up plan to monitor implementation of the OECD AI Principles and identify good practices matches specific actions implemented by the Colombian government to the OECD recommendations.
- The GovCo Dashboard monitors the implementation of AI projects in the public sector. The dashboard includes a description of each project and highlights mechanisms through which AI is used and the progress of each project.
- A dashboard for monitoring the Ethical Framework for Artificial Intelligence, a public access tool where citizens can learn more about the use of AI systems in the State and implementation of the ethical principles of Artificial Intelligence in AI projects. The dashboard is available at <https://inteligenciaartificial.gov.co/en/dashboard-IA>.

These policy intelligence tools are also used by the Presidency and the AI Office to evaluate resource allocation and evaluate policy implementation.

Source: (OECD, 2021^[6]), Colombia officials.

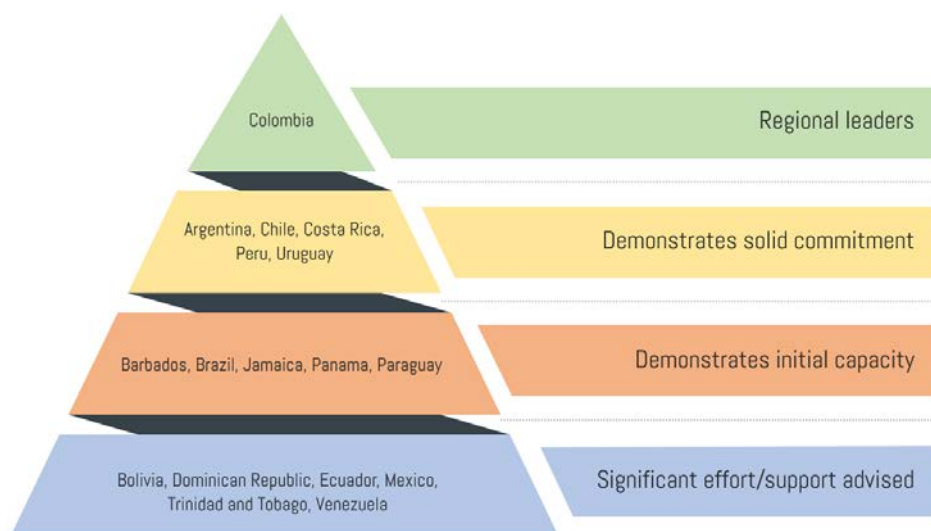
Ensuring an inclusive and user-centred approach

Inclusive

Ensuring the representation of perspectives that are **multi-disciplinary** (different educational backgrounds, professional experiences and levels, skillsets, etc.),²¹ as well as **diverse** (different genders, races, ages, socio-economic backgrounds, etc.), together in an **inclusive** environment where their opinions are valued is a critical cross-cutting factor relevant to many of the considerations discussed in this chapter and the next. This factor is fundamental to achieving AI initiatives that are effective and ethical, successful and fair. It underpins initiatives ranging from comprehensive national strategies to small individual AI projects, and everything in between. The OECD's recent *Framework for Digital Talent and Skills in the Public Sector* (OECD, 2021^[19]) affirms that the establishment of multi-disciplinary and diverse teams is a prerequisite for digital maturity and achieving a digitally enabled state.

Developing AI strategies, projects and other initiatives is an inherently multi-disciplinary process. Moreover, multi-disciplinarity is one of the most critical factors for the success of innovation projects, especially those involving tech. Pursuing such projects requires consideration of technological, legal ethical and other policy issues and constraints. Clearly, AI efforts need to be technologically feasible, but equally they need to be acceptable to a range of stakeholders (including the public) and permissible under the law.

Figure 4.10. LAC region capacities for setting guidance for building multi-disciplinary teams



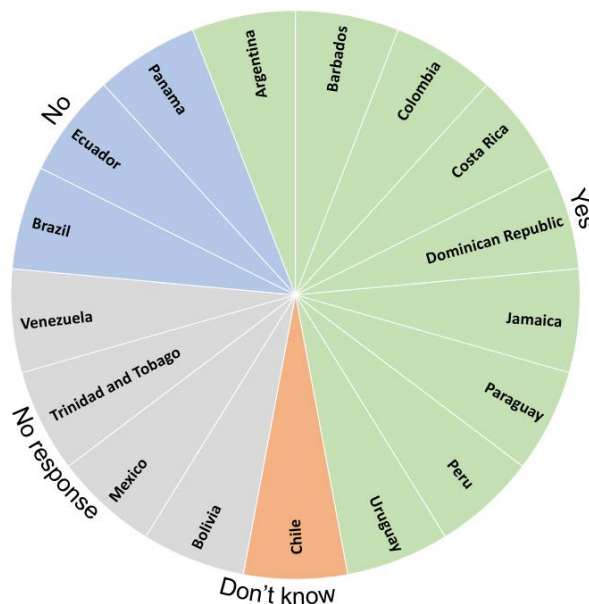
Many LAC countries have embraced multi-disciplinarity (see Table 4.2 for examples of professions involved) as a criterion for the development of digital projects, services and strategies (Figure 4.11). Nevertheless, guidance for the inclusion of multiple disciplines in the design and development of AI specifically is scarce. This trend demonstrates initial competence and commitment, but also signals that AI-specific guidance may be necessary as countries increasingly adopt and design these systems. At present, Colombia is the only country with guidance covering this topic for the development and use of AI and other emerging technologies. In their strategies, Argentina, Brazil and Uruguay recognise the importance of multi-disciplinarity for the development of AI in the public sector, but do not offer specific guidance or methods. Various other countries promote multi-disciplinarity either through innovation labs, declarations on their digital strategies and/or empirically, although not specifically for AI.

Table 4.2. Professions involved in a multi-disciplinary team

Digital professionals	Non-digital professionals
User-centred design	Law, policy and subject matter
Product and delivery	Strategy and governance
Service ownership	Commissioning and procurement
Data	Human resources
Technologists	Operations and customer service
	Psychologists and sociologists

Source: (OECD, 2021^[19]).

Figure 4.11. Use of multidisciplinary teams for delivering digital, data and technology projects in LAC countries



Source: OECD LAC Digital Government Agency Survey (2020).

In Colombia, three key guidelines for the development of digital public services emphasise the need to incorporate multiple disciplines and perspectives:

- In relation to AI, the *Emergent Technologies Handbook* proposes two measures. The first is the involvement of non-technical members in project implementation, “[working] closely with the service owners” (p. 11) and not just at the engineering level. The second is the creation of a pilot project evaluation team composed of internal and external actors (p. 9).²² Additionally, the *Task Force for the Development and Implementation of Artificial Intelligence in Colombia* states that multi-disciplinarity is an important consideration when assembling an Internal AI Working Group. The working group structure proposed by the document includes an expert in AI policy, a data scientist expert, an ethicist, an internationalist and researchers.²³
- For digital projects in general, the *Digital Government Handbook* asserts that developers should “count on everyone’s participation” (p. 33) and, more specifically, should work to: generate integration and collaboration among all responsible areas; seek collaboration with other entities; identify the project leader and assemble multi-disciplinary teams to participate in the design, construction and implementation, testing and operation of the project; and establish alliances between different actors.²⁴
- Finally, the *Digital Transformation Framework* notes that “the digital transformation of public entities requires the participation and efforts of various areas of the organization, including: Management, Planning, Technology, Processes, Human Talent and other key mission areas responsible for executing digital transformation initiatives”²⁵ (p. 21).

The Ethics and Regulation strategic axis of Argentina’s AI strategy includes an objective to “form interdisciplinary and multi-sectoral teams that manage to address the AI phenomenon with a plurality of representation of knowledge and interests” (p. 192). Additionally, this section recognises that “bias may even be unconscious to those who develop [AI] systems, insofar as they transfer their view of the world both to the selection of the training data and to the models and, potentially, to the final result. Hence the

importance of having a plural representation in the development of these technologies and the inclusion of professionals who build these methodological, anthropological and inclusion aspects” (p. 189).

One of four “transversal principles” in Chile’s national AI strategy is “Inclusive AI”. This states that all action related to AI should be addressed in an interdisciplinary way. The strategy also recommends reframing education programmes to incorporate different conceptions of AI from the perspectives of various disciplines.

Brazil’s national AI strategy explicitly discusses the multi-disciplinary nature of AI and the importance of a multi-disciplinary approach, but does not contain action items directed at supporting such an approach.

Uruguay’s strategy recognises the importance of training in multi-disciplinary contexts for public servants, in order to generate skills that enable them “to understand all the difficulties, challenges and impacts that arise when using AI in the services and processes of the Public Administration” (p. 12). Indeed, the strategy itself was developed by a multidisciplinary team representing the fields of technology, law, sociology and medicine, among others. In summary, LAC country strategies with specific references to the inclusion of multi-disciplinarity in AI development provide general, models which are applicable to every AI project. Working from the basis of the existing pool of use cases and lessons, a next step for policy makers in the region could be to provide guidance or methods for the inclusion of other disciplines to tackle key issues that have arisen in specific focus areas.

Although not specific to AI, LAC countries have also developed a considerable set of practices and guidelines for the inclusion of multi-disciplinarity in the development of digital government projects. These are relevant because guidelines and initiatives focused on broader digital government efforts should also apply to projects involving AI in the public sector. They include the following examples:

- Argentina’s public innovation lab, LABgobar, has created the “Design Academy for Public Policy”. The lab works addresses has two main purposes: 1) to identify and strengthen specific-themed communities of practice through diverse approaches that inspire action, participation and collaboration; and 2) to train interdisciplinary teams of public servants from different ministries through the Emerging Innovators executive programme, which provides real challenges for participants to solve through the application of innovation tools.²⁶
- Barbados’ *Public Sector Modernization Programme* proposes the creation of a digital team with expertise in areas such as “digital technologies, open innovation, service design, data analytics and process reengineering, among others”.²⁷
- During innovation processes, Chile’s Government Lab recommends forming “a multifunctional work team, composed of representatives of all the divisions related to the initial problem or opportunity” and provides guidance on so doing.²⁸
- The National Code on Digital Technologies of Costa Rica recommends building multi-disciplinary teams as part of its standards for digital services, including specific roles such as a product owner, project manager, implementation manager, technical architect, digital support leader, user experience designer, user researcher, content designer, back-end developer and front-end developer.²⁹
- Jamaica developed a multi-disciplinary experience as part of its COVID-19 CARE programme. Several government agencies were involved in developing an online system for the receipt of grant applications, automated validations and payment processing.³⁰
- Panama’s *Digital Agenda 2020* was designed by a multidisciplinary team (p. 2).
- The development of Paraguay’s *Rindiendo Cuentas* portal (<https://rindiendocuentas.gov.py>) for transparency and accountability involved various teams across the public administration.³¹
- Peru’s Government and Digital Transformation Laboratory includes among its objectives the “transfer of knowledge on Agile Methodologies in the public sector and [the promotion of] the creation of multidisciplinary teams” for the co-creation of digital platforms and solutions.³²

Additionally, all public administration entities are mandated to constitute a Digital Government Committee consisting of a multi-disciplinary team including, at least, the entity director, the Digital Government leader, the Information Security Officer, and representatives from IT, human resources, citizen services, and legal and planning areas.³³

- In relation to recruitment processes, Uruguay “seeks complementarity through multidisciplinary teams, complementary knowledge and different perspectives”.³⁴

These efforts show that building multi-disciplinary teams has been a recurrent practice among most LAC governments when delivering digital solutions. Nevertheless, for many of the initiatives in question, the OECD was unable to determine the process whereby the teams were built and how the different participating disciplines contributed to the end objective. It also proved difficult to ascertain the composition of the development teams of existing AI use cases. As part of transparency measures to increase trust and safety (see the chapter *Develop a responsible, trustworthy and human-centric approach*), providing further information on team composition could be a good practice for LAC countries to adopt when delivering AI solutions. As an example of this, Box 4.11 presents two non-AI cases where multiple disciplines contribute to the deliverance and governance of digital services.

Box 4.11. Multi-disciplinary teams to improve digital public services

The following two cases are not related to AI initiatives, but provide nevertheless excellent examples of multi-disciplinary approaches that could also be applied to AI at a strategic level.

Redesigning the disability certification process (Argentina)

In Argentina, an estimated 3 million people have some form of disability. In order to certify these disabilities, the Medical Evaluation Boards (MEB) issues a *Certificado Único de Discapacidad* (disability certificate, CUD) which allows people to access rights and benefits provided by the government.

The process to obtain a CUD was long and difficult. In the absence of digital support the process could take up to seven months and involved many steps, many of which required the user to visit a public office in person. Not only was turnaround slow but the process itself added extra burdens and complexity to the lives of people who needed increased levels of support.

Having identified the need to transform this service, the National Agency for Disability paired with the team at Mi Argentina, Argentina’s platform for providing citizen-centred services. This multi-disciplinary team consisted not only of software engineers and designers but also psychologists, political scientists, anthropologists and sociologists. Together, the team set out to simplify and speed up the procedure while accompanying people throughout the process.

To this end, the team interviewed people with disabilities, their families and health workers. As they built up a picture of the challenges people faced, they identified opportunities for simplifying the process and designing a one-step online approach. The new CUD application process now guides citizens through the requirements of their application rather than requiring them to attend a physical meeting to establish the necessary documentation.

Source: (OECD, 2020^[20]).

The Open Data Working Group (Uruguay)

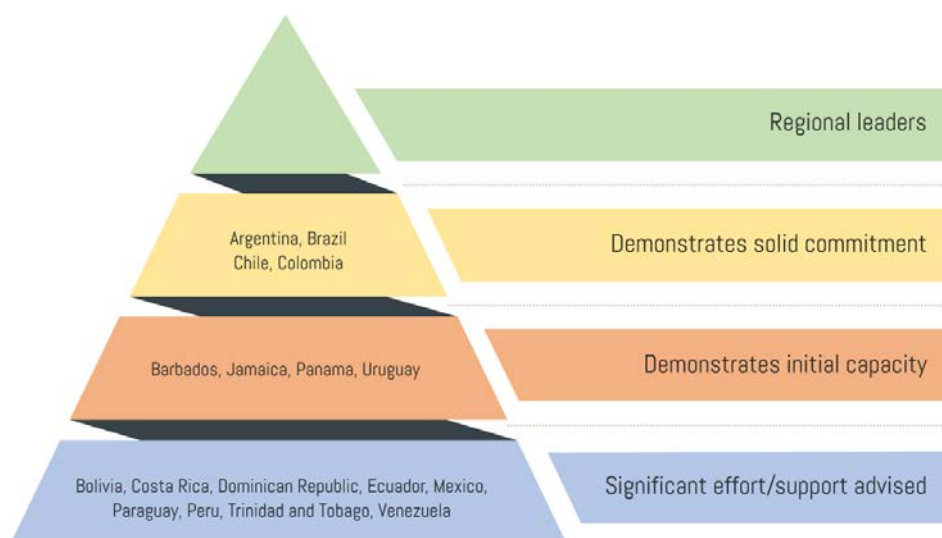
In order to implement an open data policy, co-ordinate action among different actors, and generate a collaborative work environment for public entities, civil society and academia in the field of active transparency, Uruguay created the Open Data Working Group, a multi-disciplinary entity that comprises representatives from various non-tech public sector institutions. Core members include the Ministry of

Social Development, Management and Evaluation, the Faculty of Engineering (University of the Republic of Uruguay), the National Institute of Statistics, the Open Government Network (Civil Society), the Municipality of Montevideo and the digital agency, AGESIC.

Source: www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/comunicacion/publicaciones/grupo-trabajo-datos-abiertos

Alongside multi-disciplinarity, another critical concept is **diversity**. This umbrella concept recognises that people, whilst similar in many ways, have different life experiences and characteristics, such as gender, age, race, ethnicity, physical abilities, culture, religion and beliefs (Balestra and Fleischer, 2018^[21]). These elements produce unique and important values, preferences, characteristics and beliefs in each individual that have been shaped by the norms and behaviours they have experienced over time. In the field of AI, diverse teams can better consider the needs of different users and help prevent or eliminate possible biases from the outset (OECD, 2019^[22]), because diverse representation in product ideation and design helps to minimise the possibilities of data bias and algorithmic discrimination. As touched on earlier, this benefit can only be fully realised in an environment that is *inclusive*, where the opinions of individuals are valued and they feel safe to express them.

Figure 4.12. LAC region capacities for setting guidance for building diverse AI teams

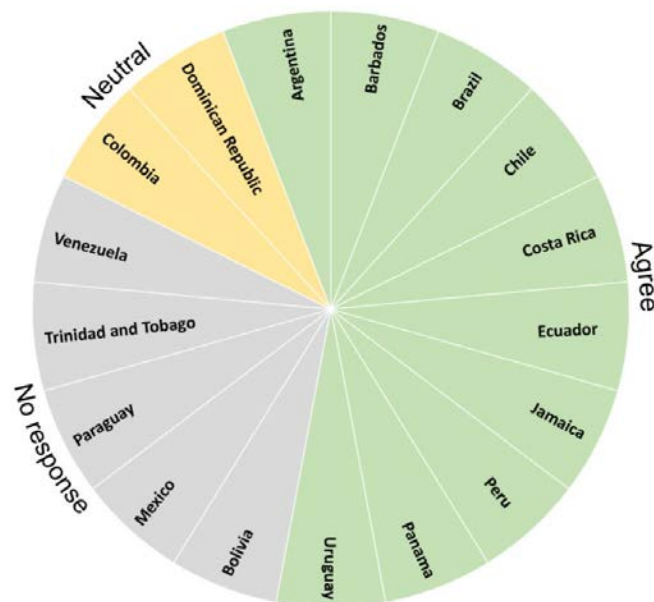


Source: OECD analysis based on research and LAC country surveys and interviews.

At the global level, lack of gender and racial diversity persists in AI research and the AI workforce, in spite of its acknowledged importance (NSTC, 2016^[23]). However, many countries in the LAC region retain the perception that digital teams in the public sector are diverse and reflect broader society (Figure 4.13). Given the scope of this study, it was not possible to assess the actual diversity of these teams; however, guidance or methods for ensuring such diversity is mostly absent in LAC countries. Although the AI strategies of Argentina, Brazil, Chile and Colombia highlight the importance of diversity for AI development, there are very few examples of specific initiatives and guidance being developed to make diversity a key factor for the composition of AI teams. One such example is the proposed design for Colombia's AI Task Force, which considers the inclusion of diverse backgrounds in the construction of its teams.³⁵ Among the considerations evaluated in this and the next chapter, diversity was the least addressed by LAC countries.

LAC countries perceptions that their digital teams are diverse coupled with scant guidance in this area creates a somewhat contradictory scenario and may indicate blind spots to potential problems. Granted, it may indicate that teams are indeed diverse, although without more solid guidance such diversity may be fleeting and subject to change. Countries should consider adopting general guidance by assessing the state of diversity in their digital teams and recognising its importance in strategies or guidelines. As previously pointed out, existing experience in the LAC region could lead to guidance tailored to focus area and contexts where team diversity has proven to be an important element of AI development.

Figure 4.13. Digital teams in the public sector are often diverse and reflect broader society in my country



Source: OECD LAC Digital Government Agency Survey (2020).

In its AI strategy, Argentina recognises the importance of “plural representation in the development of [AI] technologies and the inclusion of professionals who design (...) methodological, anthropological and inclusion aspects” (p. 189). Its chief concern is to tackle bias throughout the development process, including the selection of training data, the design of algorithms and final outcomes. More specific instructions on diversity exist for the creation of the AI Ethics Committee, “an independent, multidisciplinary and multisectoral entity consisting of professionals from different areas of knowledge and members of the community, balanced in age, sex and ethnic and cultural origin”. The Committee also emphasises the need to ensure “that its members have a constant link with civil society organisations oriented to these issues and access to external consultants with specific knowledge, if necessary, for particular cases”.

Brazil’s AI strategy commits to “stimulate the diverse composition of AI development teams with regard to gender, race, sexual orientation and other socio-cultural aspects”.

Chile’s AI Policy underlines the importance of diverse and inclusive teams, particularly from a perspective of gender and sexual diversity. In order to foster equity in the implementation of AI systems, the policy also highlights the importance of developing AI in an inclusive manner incorporating perspectives from Indigenous groups, people with special needs and the most vulnerable.

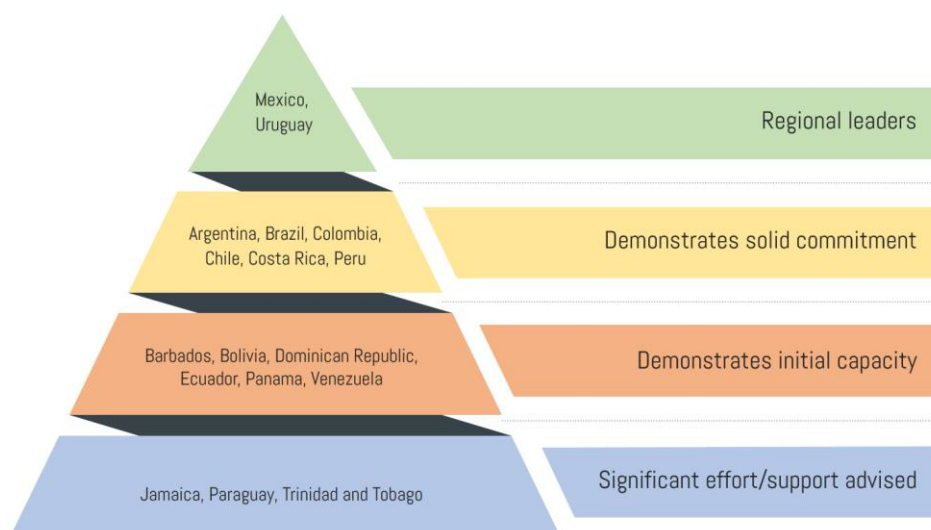
Finally, Colombia’s AI Ethical Framework affirms, as part of its non-discrimination principle, that “a diverse group of the population should participate in the design generating impact matrices that make it possible to detect any type of discrimination at an early stage and correct accordingly in a timely manner”.

User-centred

Each national approach must operate within its own unique context as well as its own culture and norms. Governments should engage with citizens, residents, businesses, public servants and anyone else who may interact with, or be impacted by, an AI-based solution, through deliberative dialogue to more clearly understand their perspectives, values and needs (Balaram, Greenham and Leonard, 2018^[24]). Users of public services may want meaningful engagement and assurances to clarify how the use of AI will impact the services on which they depend. In some instances, citizens can also become co-creators of public services that use AI, a process that involves significant user engagement (Lember, Brandsen and Tönurist, 2019^[25]). Finally, AI has the potential to help governments move towards proactive public services. Such services anticipate and handle user needs before action is required (e.g. completing a form) (Scholta et al., 2019^[26]) and would not be possible without greater understanding of these needs.

Unless they engage with potential users (both inside and outside government, as appropriate), public servants will not be able to determine accurately which problems exist and whether a potential AI application or alternative will satisfy core needs. Such engagement will become increasingly important and should be included as an integral part of national strategies and overall direction. Civil servants must also be empowered to interact with users.

Figure 4.14. LAC region capacities for setting guidance and methods for understanding user needs



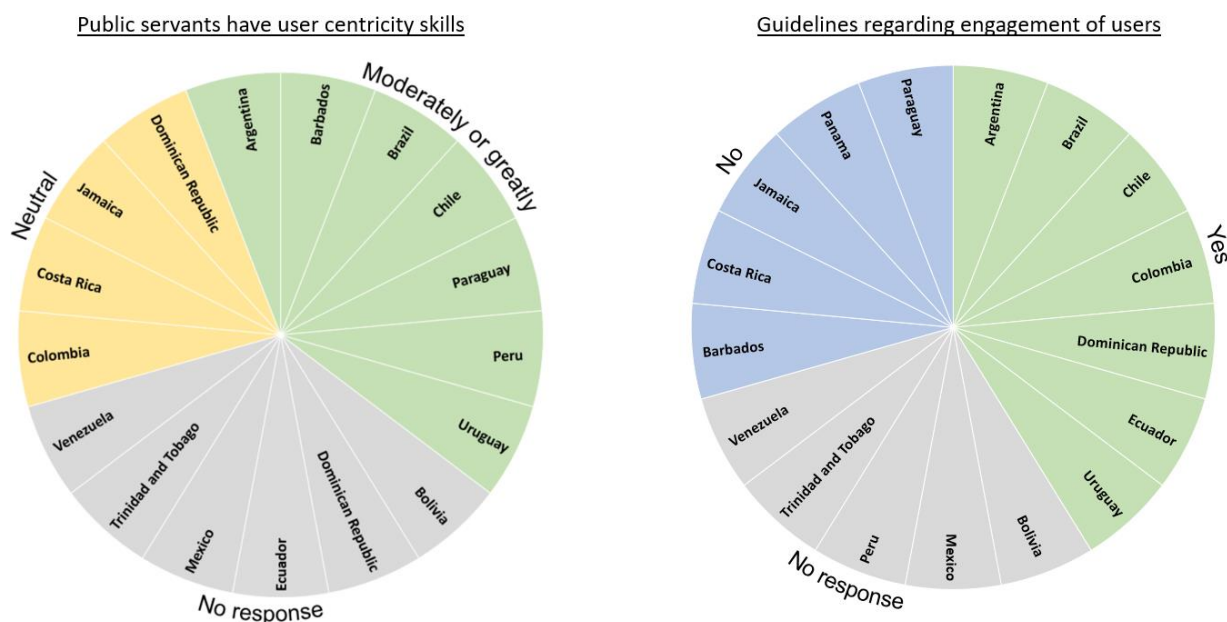
In the LAC region, countries have developed two complementary approaches to designing digital public services according to user needs. The first is a *user-driven* approach that centres on understanding users and co-designing public services. The second is a *user-informed* approach focused on adapting and designing services according to requests, response rates, usability and measured satisfaction. The OECD Digital Government Policy Framework recommends that policy processes, outputs and outcomes not just be informed, but also shaped by the decisions, preferences and needs of citizens through mechanisms for engagement and collaboration (OECD, 2020^[27]). Such an approach is designed to allow people's voices to be heard in public policy making. To this end, governments can establish new forms of partnerships with the private and third sectors, crowdsource ideas from within the public administration and society at large, and make use of methodologies such as user research, usability (UX) design and human-centred design to create and improve public services (OECD, 2020^[28]).

The difference between both approaches is illustrated by the case of Panama, where the OECD found a focus on digitising existing processes and procedures, and less attention to understanding user needs and re-orienting services accordingly.

The dominant themes of delivery in Panama are centred on digitisation and/or automation of existing processes rather than on users and their needs. Consequently, there is greater focus on the technologies that can be deployed rather than the transformation of the underlying services. This leads to the proliferation of apps and different technologies responding to particular problems from a technology point of view rather than considering critical policy actions (...) to reflect the diversity of the country's population and better serve their needs (OECD, 2019^[8]).

Perception among LAC countries of user centricity skills among public servants remains generally positive. Additionally, half of the surveyed countries that responded confirmed the existence of guidelines to encourage user engagement in the service and policy design process. Figure 4.15 illustrates the increasing inclusion of users' perceptions and needs in the design of digital services in the region. Although evidence related specifically to user-centred AI development is scarce, current work shows a solid ground to extend guidance and professional expertise in order to better understand users and take their needs into account when designing AI systems.

Figure 4.15. Perception that public servants have user centricity skills and countries have guidelines in place for user engagement



Note: For user centricity skills, no countries indicated "to some extent" or "not at all".

Source: LAC Digital Government Agency Survey (2020).

Mexico and Uruguay are the only two countries in the LAC region to explicitly consider user-centred indications for AI development within their AI impact assessment guides (Box 4.5). Mexico's *Impact Analysis Guide for the development and use of systems based on artificial intelligence in the Federal Public Administration* asks whether a system " [was] consulted or tested with interest groups and/or vulnerable groups" (Coordinación de la Estrategia Digital Nacional, 2018, p. 8^[12]) in order to assess if a system meets users' needs. In another approach to user needs, Uruguay's Algorithmic Impact Study model seeks to ascertain the existence, or not, of "a mechanism to collect feedback from system users" (AGESIC, 2020, p. 11^[11]).

Various LAC countries have developed user-driven capacities, principally focused on human-centred design methodologies, albeit not exclusively in the field of AI:

- One of the objectives of LABgobar, Argentina’s public innovation lab, is to design user-centred policies and services. To this end, it carries out ethnographic research focused on studying the habits and behaviours of citizens as they interact with the state, and delivers methodologies to incorporate people’s views, feelings and voices into decision making to bring them to the attention of the institutional actors responsible for implementing public policies.³⁶
- Argentina’s National Direction of Digital Services within the Government Secretariat of Modernisation (Secretaría de Gobierno de Modernización) established a set of principles to carry out research on user needs, advise public sector organisations and design solutions. The first principle states: “Prioritise user needs: we constantly talk with citizens, we observe their contexts, we understand what they need beyond what they say”.³⁷ In addition, this entity created the *Code of good practices for the development of public software*, which compiles various methodologies and prerequisites to understand user needs (Box 4.12).
- Brazil’s national digital government strategy also includes a principle focused on citizens’ needs.³⁸ This objective is supported by the Design Thinking Toolkit for Government, developed by the Innovation Laboratory of the Federal Court of Audit, which provides guidance on the engagement of end users in the early stages of service design, with a view to the dissemination and use of relevant techniques to public institutions. The Design Thinking Toolkit consists of five phases: empathy, (re)definition, ideation, prototyping and testing. Each phase is explained and accompanied by a set of tools.³⁹ Additionally, the federal government has created a dedicated team to collect information about the quality and adequacy of digital public services using simple and agile methodologies. As of February 2021, the team had reached 31 660 people through 2 373 interviews, 29 287 online forms and 58 research projects.⁴⁰
- The principles of Chile’s Government Laboratory (Laboratorio de Gobierno, LabGob) help guide different types of government projects, and notably include a principle to “Focus on people” in order to understand their needs, assets, motivations and capacities as agents of the innovation process (see Box 5.11). LabGob has also produced a set of guidelines entitled “How Can We Facilitate Face-to-Face Spaces for Public Innovation?” to help public sector organisations obtain external views, including from users (see Box 6.7). The OECD has previously elaborated a number of recommendations for Chile, which the country is evaluating, on how to advance in becoming more citizen-driven by uncovering user needs, among other approaches (OECD, 2020_[20]).
- Colombia has three relevant instruments for understanding user needs. The *Emergent Technologies Handbook* does not define specific guidelines but instead emphasises the need to consider “User Experience” as part of the solution’s architecture (p. 10). It also suggests the inclusion of users in pilot project evaluation teams (p. 9).⁹² Another document, the *Digital Government Handbook*, recommends to “Identify the problem or need and the stakeholders related to the project” (p. 31).⁹³ Finally, the *Guide for Characterization of Citizens, Users and Stakeholders*, which is not limited to digital government services, provides a general guideline for the characterisation of users in all government projects involving external actors: “To characterise is to identify the particularities (characteristics, needs, interests, expectations and preferences) of the citizens, users or stakeholders with whom an entity interacts, in order to group them according to similar attributes or variables” (p. 10).⁹⁴
- In Bogotá, Colombia, the city’s innovation lab, LAB Capital, developed an online course on public sector innovation for public officials to help public servants gain insights into innovating on policies and services from a user-centred perspective, as well as to foster an ecosystem of innovators among public offices.⁴¹
- Costa Rica’s National Code on Digital Technologies lists a set of applicable principles, policies and standards (see the chapter, “Digital accessibility, usability and user experience”).⁴² Among the standards for digital services, the code defines a user-centred procedure to be considered when designing and procuring digital services. This procedure includes understanding users’ needs,

performing constant research about users, building a multi-disciplinary team, using agile methodologies, iterating to achieve permanent improvement, running tests with users, and collecting performance data and indicators, among others.

- Peru's *Government and Digital Transformation Laboratory* also uses user-centred methodologies to design public services, according to the "Digital Agenda towards the Bicentennial".⁴³ Additionally, the *Guidelines for the Formulation of the Digital Government Plan* includes within its principles the importance of focusing design on the needs and demands of the citizen. It states that "public entities [must] make use of innovation, agile or other frameworks focused on the citizen's experience, and investigate and analyse their behaviours, needs and preferences" (p. 35).⁴⁴ The country has also developed a digital volunteering programme to engage academia, the private sector, civil society and citizens in various projects to design, redesign and digitise public services and policies.⁴⁵ In an interview with the OECD, Peru officials stated that they are working to shift public sector mind-sets and cultures through such guidance to ensure a continuous focus on core user needs, drawing on user research, interviews and user testing with rapidly developed prototypes and minimum viable products.
- In Uruguay, the *Social Innovation in Digital Government Lab* (Laboratorio en Innovación social en Gobierno Digital) provides co-creation and participation methodologies to find better ways to deliver public services (Box 4.12). Their process involves four stages: understand, empathise, co-create and experience.

Box 4.12. Methodologies to understand user needs

Code of Good Practices for the development of public software (Argentina)

The National Office of Information Technology within the Secretariat of Public Innovation has developed a Code of Good Practices for the development of public software. The first section of the Code requires public officers to understand the contexts and real needs of users, in order to realize the functions of the digital systems they are designing and be able to better assess development priorities. The Code provides a set of ideal and minimum good practices, as well as guidance on taking the first steps:

Ideal

- Create user story maps to understand the experience of all different types of users, including people with disabilities.
- Develop prototypes to test the service with real users.
- Make decisions based on the collection and analysis of objective data, including A/B tests, service usage metrics and usability tests.
- Design the software taking into account the entire user experience, including actions performed outside of digital systems.
- Evaluate the service using W3C tools.

Minimum

- Comply with accessibility regulations to guarantee real equality of opportunities and treatment of all people.
- Ensure familiarity with the standards of the National Directorate of Digital Services.
- Determine the type of devices preferred by users and calibrate the interface design and development and user experience accordingly.

Where to start

- Consult the “Know your project” Technological Ten Commandments.
- Consult the “Make sure your solutions are accessible” Technological Ten Commandments.
- Create a user history map in order to understand the experience of frequent users.

Source: www.argentina.gob.ar/ontfi/codigo-de-buenas-practicas-para-el-desarrollo-de-software-publico/.

Co-creation and participation methodologies of the Social Innovation in Digital Government Lab (Uruguay)

The Social Innovation in Digital Government Lab provides public institutions in Uruguay with co-creation and participation methodologies to identify better ways to deliver public services. Among the different methodologies offered, the following two templates can be employed to better understand user needs:

“Actions with Actors” template

This template seeks to understand expectations, opportunities and risks associated with a specific actor. Development teams are asked to complete a template sheet with the following information and discuss among the group. The key output is defining actions to take with regard to:

- The **expectations/motivations** of a specific actor.
- The **risks/fears** of a specific actor.
- **Opportunities** – once expectations and fears have been discussed, teams must identify an opportunity to which they can provide a response.
- **Actions to take.**

“Typology of User” template

This template seeks to identify the different types of users of a particular service and thus generate an inclusive solution. Teams must respond to the following questions/tasks:

- What is the **actor’s name, age and location**?
- What are the **needs and motivations** of the concerned actor?
- Represent the actor with a **drawing/image**.
- What is the concerned actor’s **relationship with technology**?

Development teams are asked to complete the template for each type of person identified and then analyse and systematise the information, identifying patterns and generating groups, if applicable. The type of data required to assemble the profiles may vary based on the project. For example, data on educational level and declared gender may be relevant.

Source: www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/tematica/metodologias-lab.

LAC countries are also adapting and designing services according to user requests, response rates and measured satisfaction. Although the following examples relate to user-centred methodologies, they concentrate mostly on measuring, and thus fall into the category of “user-informed” approaches, rather than prioritising a more comprehensive understanding of user needs.

- Barbados has implemented a usability testing programme for its Electronic Document and Records Management System.⁴⁶
- In Brazil, under the Digital Government Strategy, agencies are required to use public satisfaction tools. In this regard, the Strategy details three main courses of action. First, as part of its “Digital services satisfaction assessment” objective, the country aims to standardise satisfaction assessment, increase user satisfaction of public services and improve useful perception of public information. Second, strategy states that agencies will “conduct at least one hundred experience surveys with real users of public services by 2022”. And third, the strategy commits to “implement a mechanism to personalise the offer of digital public services, based on the user’s profile”.⁴⁷ This approach aligns with the digital services monitoring dashboard offered as part of the country’s one-stop-shop portal,⁴⁸ which makes available general satisfaction indicators including users’ evaluation of information and services, and average waiting time.
- Ecuador has published the *Open Data Guidelines* (under consultation), a document that provides guidance on selecting and prioritising demands for open data, creating an inventory of the most requested information, fostering citizen participation in order to better define the public’s open data needs, and evaluating perception and the re-use rate of published datasets.⁴⁹
- Uruguay assesses citizens’ response to digital services through focus groups and monitoring strategies. Research projects based on focus groups are carried out annually to evaluate aspects such as image, satisfaction and access barriers. Different focus groups consist of prioritised segments of the population, previously identified in quantitative studies. Monitoring strategies and indicators include a satisfaction survey, studies of the general population that measure completion and satisfaction with online procedures, and interoperability platform indicators.⁵⁰
- Venezuela’s *Info-government Law* includes a general guideline on the design of ICT initiatives based on accessibility and usability conditions. Article 15 states that “in the design and development of systems, programs, equipment and services based on information technologies, the necessary accessibility and usability considerations must be foreseen so that they can be used universally by those people who, for reasons of disability, age, or any other condition of vulnerability, require different types of information media or channels”.⁵¹

To assist governments in further developing their user-centred design skills, the Government of Australia’s BizLab kindly provided OPSI with its full Human Centred Design curriculum, including editable source files, which OPSI has made available on its Toolkit Navigator.⁵²

References

- AGESIC (2020), *Preguntas para la evaluación del Estudio de Impacto Algorítmico (EIA): Proyectos de sistemas automatizados para la toma de decisiones.*, <https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/sites/agencia-gobierno-electronico-sociedad-informacion-conocimiento/files/documentos/publicaciones/Gu%C3%ADa%20para%20el%20estudio%20de%20Impacto%20Algor%C3%ADtmico%20%28EI.> [11]
- Balaram, B., T. Greenham and J. Leonard (2018), *Artificial Intelligence: Real Public Engagement*, RSA, https://www.thersa.org/globalassets/pdfs/reports/rsa_artificial-intelligence--real-public-engagement.pdf. [24]

- Balestra, C. and L. Fleischer (2018), “Diversity statistics in the OECD: How do OECD countries collect data on ethnic, racial and indigenous identity?”, *OECD Statistics Working Papers*, No. 2018/09, OECD Publishing, Paris, <https://dx.doi.org/10.1787/89bae654-en>. [21]
- Berryhill, J. et al. (2019), “Hello, World: Artificial intelligence and its use in the public sector”, *OECD Working Papers on Public Governance*, No. 36, OECD Publishing, Paris, <https://dx.doi.org/10.1787/726fd39d-en>. [17]
- Coordinación de la Estrategia Digital Nacional (2018), *Guía de análisis de impacto para el desarrollo y uso de sistemas basadas en inteligencia artificial en la APF*, https://www.gob.mx/cms/uploads/attachment/file/415644/Consolidado_Comentarios_Consulta_IA_1_.pdf. [12]
- Guío Español, A. (2020), *Ethical Framework for Artificial Intelligence in Colombia*, <https://iaeticacolombia.gov.co/static/media/Marco-Etico-para-la-IA-en-Colombia.pdf> (accessed on 1 December 2020). [10]
- Lember, V., T. Brandsen and P. Tönurist (2019), *The potential impacts of digital technologies on co-production and co-creation*, pp. 1665-1686, <https://www.tandfonline.com/doi/full/10.1080/14719037.2019.1619807>. [25]
- Mateos-Garcia, J. (2018), *The complex economics of artificial intelligence*, <https://www.nesta.org.uk/blog/complex-economics-artificial-intelligence/>. [18]
- Nadal, V. (2020), *Inteligencia artificial y ‘seudonimato’: el Gobierno presenta la primera versión de la Carta de Derechos Digitales*, <https://elpais.com/tecnologia/2020-11-17/inteligencia-artificial-y-pseudoanonimato-el-gobierno-presenta-la-primera-version-de-la-carta-de-derechos-digitales.html>. [9]
- NSTC (2016), *Preparing for the Future of Artificial Intelligence*, Executive Office of the President National Science and Technology Council Committee on Technology, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NS-TC/preparing_for_the_future_of_ai.pdf. [23]
- OECD (2021), *OECD Good Practice Principles for Data Ethics in the Public Sector*, OECD Publishing, <https://www.oecd.org/gov/digital-government/good-practice-principles-for-data-ethics-in-the-public-sector.htm>. [2]
- OECD (2021), “State of implementation of the OECD AI Principles: Insights from national AI policies”, *OECD Digital Economy Papers*, No. 311, OECD Publishing, Paris, <https://dx.doi.org/10.1787/1cd40c44-en>. [6]
- OECD (2021), “The OECD Framework for digital talent and skills in the public sector”, *OECD Working Papers on Public Governance*, No. 45, OECD Publishing, Paris, <https://dx.doi.org/10.1787/4e7c3f58-en>. [19]
- OECD (2020), *Digital Government in Chile – Improving Public Service Design and Delivery*, OECD Digital Government Studies, OECD Publishing, Paris, <https://dx.doi.org/10.1787/b94582e8-en>. [20]
- OECD (2020), “Digital Government Index: 2019 results”, *OECD Public Governance Policy Papers*, No. 03, OECD Publishing, Paris, <https://dx.doi.org/10.1787/4de9f5bb-en>. [1]

- OECD (2020), *OECD Open, Useful and Re-usable data (OURdata) Index*, OECD Publishing, [28]
<http://www.oecd.org/gov/digital-government/ourdata-index-policy-paper-2020.pdf>.
- OECD (2020), “The OECD Digital Government Policy Framework: Six dimensions of a Digital Government”, *OECD Public Governance Policy Papers*, No. 02, OECD Publishing, Paris, [27]
<https://dx.doi.org/10.1787/f64fed2a-en>.
- OECD (2020), *Tracking and tracing COVID: Protecting privacy and data while using apps and biometrics*, OECD Publishing, [16]
<http://www.oecd.org/coronavirus/policy-responses/tracking-and-tracing-covid-protect-ing-privacy-and-data-while-using-apps-and-biometrics-8f394636>.
- OECD (2019), *Artificial Intelligence in Society*, OECD Publishing, Paris, [22]
<https://dx.doi.org/10.1787/eedfee77-en>.
- OECD (2019), *Digital Government Review of Panama: Enhancing the Digital Transformation of the Public Sector*, OECD Digital Government Studies, OECD Publishing, Paris, [8]
<https://dx.doi.org/10.1787/615a4180-en>.
- OECD (2019), *OECD Recommendation of the Council on Artificial Intelligence*, [5]
<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.
- OECD (2019), “Scoping the OECD AI principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO)”, *OECD Digital Economy Papers*, No. 291, OECD Publishing, Paris, [4]
<https://dx.doi.org/10.1787/d62f618a-en>.
- OECD (2019), *The Path to Becoming a Data-Driven Public Sector*, OECD Digital Government Studies, OECD Publishing, Paris, [15]
<https://dx.doi.org/10.1787/059814a7-en>.
- OECD (2018), *Digital Government Review of Brazil: Towards the Digital Transformation of the Public Sector*, OECD Digital Government Studies, OECD Publishing, Paris, [14]
<https://dx.doi.org/10.1787/9789264307636-en>.
- OECD (2017), *OECD Guidelines on Measuring Trust*, OECD Publishing, Paris, [3]
<https://dx.doi.org/10.1787/9789264278219-en>.
- Pencheva, I., M. Esteve and S. Mikhaylov (2018), *Big Data and AI – A transformational shift for government: So, what next for research?*, <https://doi.org/10.1177%2F0952076718780537>. [13]
- Scholta, H. et al. (2019), *From one-stop shop to no-stop shop: An e-government stage model*, pp. 11-26, <https://www.sciencedirect.com/science/article/pii/S0740624X17304239>. [26]
- Ubaldi, B. et al. (2019), “State of the art in the use of emerging technologies in the public sector”, *OECD Working Papers on Public Governance*, No. 31, OECD Publishing, Paris, [7]
<https://dx.doi.org/10.1787/932780bc-en>.
- Whittaker, M. et al. (2018), *AI Now Report 2018*, [29]
https://ainowinstitute.org/AI_Now_2018_Report.pdf.

Notes

¹ (OECD, 2019^[4]) presents a common understanding of what constitutes an AI system as well as a framework detailing the stages of the AI system life cycle.

² www.oecd.org/digital/ieconomy.

³ www.mofa.go.jp/files/000486596.pdf.

⁴ See Annex B for sources and details.

⁵ For instance, the OECD participated in the Expert Roundtable on International Best Practices and the Expert Roundtable on Youth Issues, hosted by the Berkman Klein Center for Internet and Society at Harvard University. A summary report of their discussions is available at:

<https://cyber.harvard.edu/story/2021-01/summary-report-expert-roundtable-colombias-draft-ai-ethical-framework>.

⁶ Annex B presents an overview of some of the mechanisms aligned with the OECD AI Principles that have been put in place by LAC governments. It should be noted that the seven LAC countries that officially adhere to the OECD AI Principles are considered to be in full alignment. Thus, for these countries, the Annex shows areas where they *further* strengthen their commitment through the elaboration of country-specific principles.

⁷ <https://oecd.ai/dashboards/policy-initiatives/2019-data-policyInitiatives-15065>.

⁸ See initiative 5.7 of Agenda Digital 2020.

⁹ https://cdn.www.gob.pe/uploads/document/file/473582/du_007_2020.pdf.

¹⁰ Outside of the LAC region, Canada's Directive on Automated Decision Making and its associated Algorithmic Impact Assessment represent the leading example of such an approach. Further details can be found at www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592 and an in-depth case study is available in the OECD report *Hello, World: Artificial Intelligence and its Use in the Public Sector* (Berryhill et al., 2019^[17]).

¹¹ See www.digital.nsw.gov.au/digital-transformation/policy-lab/artificial-intelligence for examples of risks associated with AI bias and other challenges.

¹² See https://issuu.com/c4irco/docs/brochure_c4ir_english_issuu.

¹³ www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/comunicacion/publicaciones/marco-referencia-para-gestion-calidad-datos

¹⁴ For more information see: www.gov.uk/government/publications/gender-database/gender-data.

¹⁵ <https://optic.gob.do/wp-content/uploads/2019/07/NORTIC-A1-2014.pdf>.

¹⁶ <http://conatel.gob.ve/files/leyinfog.pdf>.

¹⁷ <https://oecd.ai/dashboards/ai-principles/P8>.

¹⁸ This section does not consider broader cybersecurity and information security efforts that are not directly related to AI in the public sector.

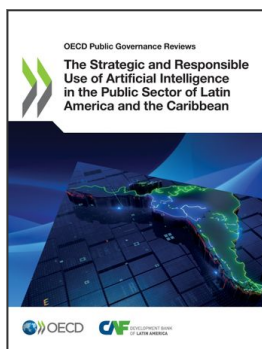
¹⁹ www.gob.mx/innovamx/articulos/guia-de-analisis-de-impacto-para-el-desarrollo-y-uso-de-sistemas-basadas-en-inteligencia-artificial-en-la-apf.

²⁰ <https://oecd.ai/dashboards/ai-principles/P9>.

²¹ Such individuals could include policy analysts and advisors, field experts, user experience designers, software developers and attorneys. Depending on the AI system and relevant applications, this may also include professions like sociologists, psychologists, medical doctors or others that have subject matter expertise in fields with which an AI initiative may interact (Whittaker et al., 2018^[29]).

²² https://gobiernodigital.mintic.gov.co/692/articles-160829_Guia_Tecnologias_Emergentes.pdf.

- ²³ <https://dapre.presidencia.gov.co/TD/TASK-FORCE-DEVELOPMENT-IMPLEMENTATION-ARTIFICIAL-INTELLIGENCE-COLOMBIA.pdf>.
- ²⁴ Section 3.2 (ICT Guidelines for the State and ICT for Society).
- ²⁵ https://mintic.gov.co/portal/715/articles-149186_recurso_1.pdf.
- ²⁶ www.argentina.gob.ar/jefatura/innovacion-publica/laboratoriodegobierno and <https://oecd-opsi.org/innovations/design-academy-for-public-policy-labgobar>.
- ²⁷ www.gtai.de/resource/blob/214860/d0599cb76af4c3f5c85df44bfff72149/pro202001315003-data.pdf.
- ²⁸ How can we solve public problems through innovation projects?
<https://innovadorespublicos.cl/documentation/publication/32/#>
- ²⁹ www.micit.go.cr/sites/default/files/cntd_v2020-1.0_-_firmado_digitalmente.pdf
- ³⁰ OECD LAC Digital Government Agency Survey (2020).
- ³¹ OECD LAC Digital Government Agency Survey (2020).
- ³² www.gob.pe/8256.
- ³³ Ministerial Resolution No. 119-2018-PCM, and its amendment Ministerial Resolution No. 087-2019-PCM.
- ³⁴ OECD LAC Digital Government Agency Survey (2020).
- ³⁵ See <https://dapre.presidencia.gov.co/TD/TASK-FORCE-DEVELOPMENT-IMPLEMENTATION-ARTIFICIAL-INTELLIGENCE-COLOMBIA.pdf> (p. 50).
- ³⁶ www.argentina.gob.ar/jefatura/innovacion-publica/laboratoriodegobierno and <https://oecd-opsi.org/innovations/design-academy-for-public-policy-labgobar>
- ³⁷ <https://github.com/argob/estandares/blob/master/principios.md>.
- ³⁸ www.gov.br/governodigital/pt-br/EGD2020.
- ³⁹ <https://portal.tcu.gov.br/inovaTCU/toolkitTellus/index.html>.
- ⁴⁰ www.gov.br/governodigital/pt-br/transformacao-digital/ferramentas/pesquisa-com-usuarios.
- ⁴¹ <https://oecd-opsi.org/innovations/online-public-innovation-course-for-public-officials-labcapital>.
- ⁴² www.micit.go.cr/sites/default/files/cntd_v2020-1.0_-_firmado_digitalmente.pdf.
- ⁴³ See pp. 43-50,
https://cdn.www.gob.pe/uploads/document/file/748265/PERU_AgendaDigitalBicentenario_2021.pdf.
- ⁴⁴ See <https://guias.servicios.gob.p> and www.peru.gob.pe/normas/docs/Anex_I_Lineamientos_PGD.pdf
- ⁴⁵ www.gob.pe/8257.
- ⁴⁶ OECD LAC Digital Government Agency Survey (2020).
- ⁴⁷ Decree 10.332/2020, www.planalto.gov.br/CCIVIL_03/_Ato2019-2022/2020/Decreto/D10332.htm.
- ⁴⁸ <http://painelservicos.servicos.gov.br>.
- ⁴⁹ https://aportecivico.gobiernoelectronico.gob.ec/legislation/processes/14/draft_versions/33.
- ⁵⁰ OECD LAC Digital Government Agency Survey (2020).
- ⁵¹ www.conatel.gob.ve/ley-de-infogobierno.
- ⁵² <https://oecd-opsi.org/toolkits/australias-bizlab-human-centered-design-curriculum>.



From:

The Strategic and Responsible Use of Artificial Intelligence in the Public Sector of Latin America and the Caribbean

Access the complete publication at:

<https://doi.org/10.1787/1f334543-en>

Please cite this chapter as:

OECD/CAF Development Bank of Latin America (2022), "Efforts to develop a responsible, trustworthy and human-centric approach", in *The Strategic and Responsible Use of Artificial Intelligence in the Public Sector of Latin America and the Caribbean*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/96d49021-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area. Extracts from publications may be subject to additional disclaimers, which are set out in the complete version of the publication, available at the link provided.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at <http://www.oecd.org/termsandconditions>.