BLAVATNIK SCHOOL OF GOVERNMENT

WORKING PAPER

PROVIDING ACCESS TO THE LATEST POLICY-RELEVANT RESEARCH

ARTIFICIAL INTELLIGENCE AND ALGORITHM ACCOUNTABILITY

POTENTIAL RISKS AND PROPOSED GOVERNANCE MEASURES

Pedro Cavalcante
Visiting Fellow, Lemann Foundation Programme,
Blavatnik School of Government, University of Oxford

BSG-WP-2024-60
September 2024

CONTENTS

1. ABSTRACT

This working paper aims to analyse the effects of the widespread adoption of artificial intelligence (AI) in government and its potential risks to service quality, equity, and integrity, with a focus on the debate on algorithmic accountability. To achieve this, it first discusses positive and negative effects of AI use in general, and then provides examples of algorithmic discrimination from different countries.

Next, it presents the AI governance strategy of the Brazilian public sector and the critical points of an AI bill under discussion in Congress, as of August 2024. It then scrutinises mechanisms used by governments around the world to address the repercussions of AI while aligning them with ethical and integrity standards, as well as participatory and inclusive principles. Some insights are offered to advance the design and implementation of similar initiatives within the Brazilian public sector.

Although Brazil has been part of the growing adoption of AI, especially in public services, the country has yet to make progress on strengthening the accountability aspect of its AI governance strategy. This working paper highlights this dimension as essential for fostering responsible, trustworthy, democratic, inclusive, and human-centered AI implementation in both the public and private sectors. It also recognises that crafting a robust governance framework for AI is a challenging process of experimentation, learning, and adaptation, and marked by progress and setbacks.

This working paper was produced by Pedro Cavalcante, a professor at the National School of Public Administration (Enap), during his visiting fellowship of the Lemann Foundation Programme for the Public Sector at the Blavatnik School of Government, University of Oxford in October 2023. Additional editing by João Pedro Caleiro.

2. INTRODUCTION

Innovations, particularly those related to digital transformation, have become a strategic cornerstone in processes, services, and policies within the public sector over the past two decades. Their main goal is addressing challenges brought upon by shifting social, economic, environmental, and technological dynamics. Consequently, the concept of digital government is no longer a novelty, but indeed the new normal in public management.

Digital transformation refers to the process of enhancing organisational performance through the utilisation of information and computer-based technology resources (Vial, 2019). The primary goal of digital transformation is to add value regarding transparency, accountability, efficiency, effectiveness, customer experience, and service delivery for both businesses and public organisations. This concept integrates the application of technology into conventional problem-solving strategies, which is now prevalent in numerous domains such as government, finance, labor market, education, medicine, the arts, science, global communication, and more.

In the realm of the public sector, digital transformation, often called digital government or *govtech*, encompasses a wide array of innovations that have been embraced and implemented over the past few decades, marking a significant shift from analog to digital government practices (OECD, 2020). This transformative process not only introduced a digital mindset into policy design but also redefined the government's structure, functions, and interactions with citizens in policymaking.

The transition to digital government comprises a series of strategies to modernise government operations to benefit society. As highlighted by the OECD (2014:14), this transformation is pivotal as it:

> *"emphasizes the crucial contribution of technology as a strategic driver to create open, innovative, participatory and trustworthy public sectors, to improve social inclusiveness and government accountability, and to bring together government and non-government actors to contribute to national development and long-term sustainable growth."*

Adopting digital technologies as a framework for the public sector has become a global phenomenon, extending to developed and developing economies. Brazil, in particular, presents an intriguing case study as it has undergone a profound transformation in this regard since the late 1990s. This transformation was spearheaded by the federal government, which initiated numerous digital innovations based on a complex legal framework and networking orientation, i.e. a range of informal and formal interactions between public and private actors with distinct but interdependent interests.

As a result, Brazil earned recognition from the World Bank, ranking as the second country worldwide in terms of maturity in digital government (the 2022 GovTech Maturity Index[1]). Brazil stood out even more for its progress, as the country had significant progress, especially in its online public service delivery during the past two years (World Bank 2022). This remarkable achievement can be attributed, in large part, to the provision of a multitude of digital services through the sophisticated platform Gov.br. Notably, this platform boasts 140 million users, equivalent to 80% of Brazil's adult population, and plays a key role in facilitating access to information and enhancing citizens' interactions with the government.

In summary, digital transformation is a process that can empower governments to harness technology's potential to reinforce transparency and accountability, boost the effectiveness of public expenditure, and enhance outcomes in areas like education, healthcare service delivery, and social welfare. This represents a foundational shift in how governments function, utilising technology to streamline operations, facilitate decision-making, manage data, engage with society, and provide a better citizen experience in their interactions with government. This transformation is achieved by developing modern digital platforms and adopting innovative technologies, including artificial intelligence (AI), machine learning, blockchain, and the Internet of Things (IoT), among others.

Artificial intelligence (AI)

The case of AI is particularly noteworthy as it has developed into a global trend that influences governments' policy decisions and implementation, producing significant changes in policymaking. The reason is that AI assumes agents that make decisions based on data to recommend or accomplish courses of action to humans or can be gradually thought of as autonomous agents as technology advances (Almeida, Filgueiras & Mendonça, 2022).

The use of AI in policy decision-making is reshaping public services and economies, offering the potential for increased productivity, enhanced efficiency, and reduced costs. This tool can be explored for various purposes, including responding to the challenges posed by COVID-19, optimising government operations, improving interactions with and services for citizens and businesses, enhancing public safety and security, reinforcing integrity and accountability in the public sector, and bolstering educational systems (OECD/CAF, 2022).

[1] The GovTech Maturity Index (GTMI) measures the key aspects of four GovTech focus areas: enhancing service delivery, supporting core government systems, mainstreaming citizen engagement, and GovTech enablers, based on the World Bank's definition of GovTech (https://www.worldbank.org/en/programs/govtech/2022-gtmi).

However, despite its benefits, AI has also ushered in a host of intricate challenges and risks for governments, society, and businesses alike, such as safeguarding private data and privacy. Two issues in particular threaten the government's ability to deliver public services effectively, transparently, and equitably:

1. Digital divide or exclusion

2. Algorithmic bias or discrimination.

The digital divide refers to the disparities in access, use, and outcomes related to information and communications technology among various population groups. This can lead to their disproportionate participation in public processes and exclusion from most *govtech* service benefits. In Brazil, this issue is worrisome because a substantial portion of the population has limited or no access to the internet. According to the 2023 TIC Households report[2] from the Brazilian Internet Steering Committee (CGI.br), internet connectivity increased in households when compared to the previous year, and particularly in the most economically disadvantaged groups. However, approximately 29 million people were still not internet users, with 24 million of those having only primary education, 17 million of the 29 million self-identifying as black or mixed race, and 16 million of them being 60 years or older. The report highlights that while the growth of nationwide connectivity is positive news, there is a persisting and worrying disparity in the quality of access across the population, blocking the cultivation of digital competencies and the comprehensive realisation of many advantages offered by the internet. This became apparent during the COVID-19 pandemic, for example, particularly concerning low-income students from public schools who encountered greater difficulties in accessing online classes due to the precariousness of their internet connections.

Algorithmic bias, on the other hand, refers to systematic and repeatable errors in a computer system that produce unfair outcomes, such as favouring one group of people over another in ways that deviate from the algorithm's intended purpose. This bias can have wide-ranging consequences, ranging from unintentional breaches of privacy to the reinforcement of different forms of discrimination related to race, gender, sexuality, age, ethnicity, religion, national origin, disability, etc.

Biased algorithms are found in both the public and private sectors, predominantly within AI and machine learning, in which decisions rely on a dataset of inputs and other learning and decision-making techniques. This has raised significant concerns regarding the fairness, justice, and transparency of criteria and automated decision-making, as well as its effect on the dehumanisation of services, the quality

[2] See https://cetic.br/en/.

of service provision, and the promotion of diversity. Depending on the circumstances, algorithmic discrimination may breach legal protections and perpetuate unfairness and inequalities in society.

This working paper aims to analyse the effects of the widespread adoption of AI and the potential risks it presents to service quality, equity, and transparency, with a focus on global initiatives that foster algorithm accountability. Adopting AI in public administration is undoubtedly a positive and primarily irreversible paradigm. However, it is equally clear that the outcomes of these innovations, particularly the indiscriminate deployment of algorithms in the context of the digital divide, are not always predictable and can adversely affect processess and service effectiveness, the protection of citizens' rights, public integrity and trust in government.

3. AI PROLIFERATION AND CONSEQUENCES

Governments worldwide are increasingly employing algorithms and AI to either automate or enhance decision-making processes in their operations and the provision of public services. This shift is driven by a rising demand for efficiency and effectiveness, prompting governments to explore solutions that enable them to respond to citizens' needs more swiftly and efficiently. AI is well-suited to meet these demands due to its capacity to rapidly and accurately process vast amounts of data, thereby empowering governments to make more informed decisions and enhance the quality of public services.

The integration of this new general-purpose technology has left a substantial impact on nearly every area of public policy, spanning fields such as agriculture, healthcare, education, science, and technology, among others. Deploying AI tools is often seen as a means to enhance efficiency and reduce public service costs. For instance, it can lead to a reduction in front-office personnel as well as minimising opportunities for corruption. As stated by OECD (2020:13), this continually evolving technology tends to make the public sector more intelligent, manifesting in increased agility, user-friendliness, and consequently, enhanced trustworthiness, as elaborated below:

> *"For instance, AI can be used to deliver more effectively personalized services and to foster citizen engagement with public institutions through the design of human-centric interfaces; enhance operational efficiency and the quality of administrative procedures through increased automation of physical and digital tasks; and to enable greater predictive capabilities for better decision making and policy outcomes, through the use of algorithms designed to uncover trends and patterns in large volumes of data."*

Nonetheless, using AI and algorithmic systems in public service delivery comes with inherent risks, and there is evidence indicating that it can lead to harm, infringe upon human rights, and result in adverse

outcomes. As Jamie Berryhill et al. (2019) pointed out, while AI can foster innovation in government, it should not be considered a panacea. The range of issues associated with employing algorithms for decision-making in particular is extensive, spanning various countries and policy domains, as exemplified in the cases below:

- In public safety, algorithms to predict crimes are often grounded in historical data reported to the system by police officers. Unfortunately, this data predominantly relates to crimes in economically disadvantaged areas, perpetuating the bias that crime is higher in less affluent regions. More troubling consequences arise when algorithms are employed in facial recognition systems, which, fueled by prejudice, may lead to the wrongful detainment of innocent individuals or even the pressing of charges based solely on their appearance, with Black and Hispanic citizens disproportionately affected in the United States (O'Neil, 2017);
- In the context of child benefits, a scandal known as the "Toeslagenaffaire" emerged in the Netherlands due to the use of an algorithm. It resulted in tens of thousands of often vulnerable families being falsely accused of fraud and hundreds of children being separated from their families (Henley, 2021);
- Australia's "robodebt scheme" employed a data-matching algorithm to calculate overpayments to welfare recipients, issuing nearly half a million incorrect debt notices and placing many welfare recipients under undue financial burdens (OECD, 2023);
- In Serbia, the 2021 Social Card law allowed data collection on social assistance beneficiaries using an algorithm to assess their socio-economic status. As a result, over 22,000 individuals lost their benefits without proper explanation, prompting legal petitions from a network of advocacy groups (Caruso, 2022);
- The Public Employment Service Austria (AMS) employs algorithmic profiling for job seekers to enhance the efficiency of its counseling process and the effectiveness of active labor market programs. However, the design of these algorithms is not solely shaped by technical considerations, but also influenced by social values, norms, and objectives. This interplay has given rise to challenges and questions surrounding the presence of inherent biases that might undermine the objectivity and neutrality of data-based claims and evidence-driven decision-making (Allhutter *et al.*, 2020).

In Brazil, there have been issues, as well. For instance, the use of automated systems in the initial assessment of benefit requests by the National Institute of Social Security (Instituto Nacional da Seguridade Social - INSS) illustrates the combined impact of algorithmic bias and the digital divide. A Federal Audit Court (TCU) audit identified several issues with their approach. Notably, using algorithms significantly increased the denial rates without providing adequate explanations to policyholders. The automation strategy was also not accompanied by staff replacement needed to analyse the benefit demands, which

led to longer waiting times and extensions of processing deadlines, sometimes four times longer than stipulated by law. The TCU report also indicated that the INSS initiative needs to uphold basic transparency standards and prioritise the interests of citizens. By diverting requests to the appeals court, it discourages the recognition of legitimate rights, fails to protect citizens, and exacerbates the (already prominent) digital exclusion in this policy area.

In response to these challenges, a trend called algorithmic accountability has emerged, as emphasised in the OECD report, "Global Trends in Government Innovation 2023." This approach, guided by democratic and integrity principles, encompasses actions to hold accountable those who create, procure, and employ algorithms for their outcomes. The goal is to oblige managers and organisations to improve the transparency of the values and criteria embedded in their algorithms, mitigate associated risks, and take responsibility for the results they generate. National governments, external stakeholders and other entities are also joining efforts to ensure that algorithms are free from bias and discrimination, including by ensuring that public servants have a strong understanding of data ethics.

As a general rule, public administration should assess whether AI is the best solution for a given problem by analysing alternatives and considering trade-offs, all while understanding the needs of their users. The OECD has launched an AI Policy Observatory, accessible to all actors and stakeholder groups in developed and developing countries, to support this approach. Its purpose is to share knowledge on policy instruments, data, and analysis, and to stimulate discussions and initiatives of data governance, addressing various aspects, including AI risks and accountability.

Ethical principles of fair and responsible AI must be reflected in the solutions that public services acquire or develop. These principles are outlined in guidelines drawn from the OECD's AI Principles Overview and UNESCO's Recommendation on the Ethics of Artificial Intelligence, which include:

- Ensuring transparency and explainability;
- Maintaining human supervision when necessary;
- Safeguarding the right of citizens to appeal decisions made by AI;
- Pursuing non-discrimination and absence of bias;
- Implementing privacy and security measures;
- Establishing data governance and accountability mechanisms.

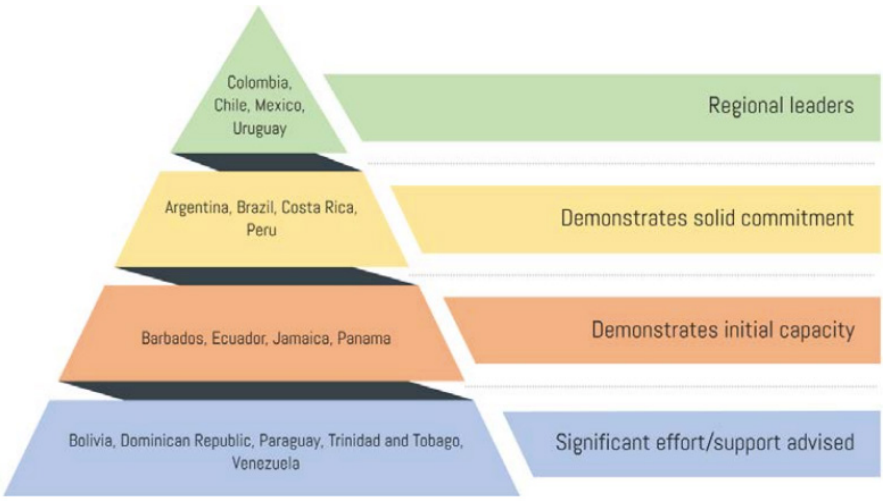
Nevertheless, most developing countries have yet to invest in initiatives to create public awareness and involve public organizations and civil servants in using algorithms in public service delivery according to ethical principles. In the next section, this paper will explore how Brazilian data governance stakeholders are addressing some of these challenges.

4. ALGORITHMIC ACCOUNTABILITY IN THE BRAZILIAN CONTEXT

AI technologies are a growing trend in the Brazilian public sector and instances of AI applications are spread across all three levels of government and the policy sector[3]. This section of the paper will explore how issues around AI and algorithm accountability are being addressed in the country, and outline global initiatives that can serve as a reference for this debate.

According to Berryhill et al. (2019), several countries have formally implemented strategies in the public sector AI domain, integrating it into policymaking and innovative service design processes. These strategies exhibit distinct configurations, albeit with common themes. For instance, they involve experimentation with and, sometimes, funding for government AI to automate processes, guide decision-making, and develop anticipatory services, meaning foreseeing the citizens demands and providing what they need before they ask. They also entail cross-government, cross-sector, and international collaboration through councils, networks, communities, and partnerships. Strategic management and the utilisation of government data, including open data, also play a pivotal role in fueling AI adoption across all sectors.

Despite these common features, different countries are at various stages of development. The heterogeneity is observed across and within regions, including Latin America and the Caribbean (LAC), where more than 60 countries are actively developing national AI strategies (OECD & CAF, 2022). In essence, AI proliferation is fast, but with different levels of governance maturity regarding priorities, public investments, and regulations. The figure below illustrates these diverse levels of development:



Source: OECD & CAF, 2022.

[3] See more at https://oecd.ai/en/dashboards/countries/Brazil.

AI data capacity relies on a well-defined data strategy that enables governments to access rich, accurate, and valuable data, maintain privacy, and adhere to societal and ethical norms. These elements are essential preconditions for the effective deployment of AI (OECD & CAF, 2022). The leaders, Uruguay and Colombia, not only have comprehensive national strategies focused on data for AI but are also transparent, dedicated, and effectively implementing them.

Brazil occupies the second tier of the capacity pyramid, signalling its commitment to the OECD AI Principles and a willingness to implement them, albeit with a lower degree of progress compared to the regional leaders.

Three policy instruments in Brazil should be highlighted in this context: (i) the Data Governance Committee; (ii) the Brazilian Artificial Intelligence Strategy (Estratégia Brasileira de Inteligência Artificial - EBIA): and (iii) the new Bill of AI.

(i) The Data Governance Committee, established by Decree in 2019, has the authority to make decisions on principles and guidelines for classifying data sharing as either broad, restricted, or specific, as well as the methods and means of publishing this classification, concerning personal data protection and the integration of entities with the Citizen Base Registry. Since its inception, the committee has been active, issuing numerous resolutions on various topics. Regrettably, none of these resolutions have focused on addressing the risks and responsibilities associated with AI or machine learning in the public and private sectors, with the majority centering on data privacy safeguards.

(ii) The EBIA, a formal document conceived under the guidance of the Ministry of Science, Technology and published in July 2021, recognises both the enormous potential benefits of AI and the need to ensure responsible AI use and development, safeguarding fairness, safety, and privacy. This strategy seeks to advance this technology's development and use, contributing to scientific progress and problem-solving in the country's priority areas. EBIA's starting point lies in defining strategic objectives encompassing the entire technological ecosystem, which can be broken down into specific actions. The strategy is anchored in the following objectives (Brazil, 2021):

- Contribute to the development of ethical principles for the responsible development and use of AI;
- Foster sustained investment in AI research and development;
- Eliminate obstacles to AI innovation;
- Train and educate professionals for the AI ecosystem;
- Encourage innovation and the development of Brazilian AI within an international context;
- Promote collaboration between public and private entities, industry, and research centers in developing AI.

Despite these measures, both the committee resolutions and the EBIA lack clear goals for addressing the adverse effects of AI applications, such as bias, discrimination, issues related to race and gender inequalities, which tend to affect more intensively those groups with restricted access to the digital transformation. The Legislative branch is more advanced in this matter, with Congress having introduced and deliberated upon several bills since 2019. Given Brazil's bicameral system, proposals originating in one house undergo scrutiny by the other. The Federal Senate has amalgamated one Chamber of Deputies proposal with other existing bills, and at the time of this paper, is deliberating on a new version of the bill, incorporating insights from a Commission of Legal Experts on Artificial Intelligence, specialising in technology law and regulation. More details below:

(iii) Bill nº 2.338/2023[4], introduced by the Senate's President, aligns with assumptions and guidelines derived from legislative initiatives in the European Union and the United States, and the principles defined by the OECD AI recommendations[5]. The bill's objective is two-fold: I) to establish rights that safeguard the most vulnerable party involved—the individual consistently impacted by AI systems, spanning from content recommendations and targeted online advertising to assessments of eligibility for credit and specific public policies; II) to implement governance tools and an institutional framework for oversight and supervision, which would ensure predictability regarding its interpretation. Additionally, the bill aims to provide legal certainty for innovation and technological development.

In a nutshell, these are the key provisions of Bill nº 2.338/2023 worth highlighting in the context of the algorithmic accountability debate:

- ***Human Rights-Centric Approach:*** The bill restates the entitlements of individuals affected by AI systems, encompassing rights such as preliminary information for individual interactions, an explanation of AI-driven decisions, non-discrimination, correction of biases, and privacy protection.
- ***Transparency and Explicability[6]:*** The bill enumerates measures to ensure transparency and mitigate bias, and standardises the procedure for algorithmic impact assessment. It also reinforces protection against discrimination through various instruments, including the right to information and understanding, the right to challenge, and a specific right to correct direct, indirect, illegal, or abusive discriminatory biases, coupled with preventive governance measures;

[4] See https://www25.senado.leg.br/web/atividade/materias/-/materia/157233.
[5] See https://oecd.ai/en/ai-principles.
[6] Explainability of AI, also referred to as interpretability, is the concept that the algorithms and their output can be explained in a way that makes sense to a human being at an acceptable level. This enables users to understand and trust the decisions made by the AI.

- ***Risk-based Regulation:*** The bill introduces a tiered risk classification system with three levels: (i) excessive risk, warranting prohibition; (ii) high risk; and (iii) non-high risk. Prior to classification, an AI provider is obligated to conduct a comprehensive self-assessment analysis of risk. In the case of high-risk AI systems, additional actions are required, such as reliability tests, measures to mitigate discriminatory biases, and technical explainability measures. The tiered regulation avoids unnecessary restrictions on systems that do not pose a high risk;
- ***Supervisory Authority:*** The bill determines that the Executive Branch designates a supervisory authority with several competencies encompassing the regulation and enforcement of legislation, the promotion of the National AI Strategy (EBIA), and the facilitation of coordination with sectoral authorities, given the cross-sectoral nature of AI systems.

If this law proposal is approved in the Senate and, further, in the Chamber of Deputies, Brazil would be well placed, depending on implementation and coordination among agencies, to contribute to a human-centric, inclusive, non-discriminatory, responsible, and ethical AI approach. This could not only improve quality of life and mitigate the digital divide, but also could set Brazil up to be as an example for other developing nations struggling with similar challenges. To ensure that these and other policy instruments are effective and enhance AI governance in Brazil, it is worth exploring next the lessons and policy features from international best practices.

5. 'GOOD' AI GOVERNANCE: THE ROLES OF GOVERMENT AND BEST PRACTICES

A fundamental starting point for establishing any effective governance framework is defining the guiding principles and values upon which it is built, typically encompassing fairness, transparency, integrity, accountability, and explainability. Translating these concepts into actions and implementation within the context of a real-world public administration, however, presents significant challenges. While technology is a critical component, most of the framework variables stem from social, political, and economic dimensions within a democratic context (Almeida, Filgueiras & Mendonça, 2022).

A framework that upholds and enforces these ethical values must always prioritise a human-centered approach. It should focus on structuring governance capacities for decision coherence, enforcement, and monitoring, as emphasised by OECD (2020). They outline various multifunctional roles the public sector can play in this process, including:

i. **Convener:** adopting a comprehensive strategy demonstrating high-level political commitment;

ii. **Financier:** providing direct or indirect funding to support research, development, and the adoption of emerging technologies;

iii. **Direct user and co-developer:** engage in innovative procurement practices or collaborate proactively through public-private partnerships to create tailored solutions;

iv. **Regulator:** reevaluating existing policy frameworks and adopting holistic approaches to ensure policy coherence and international regulatory cooperation.

The challenge lies in developing an AI governance framework that effectively combines these functions. This is the ultimate goal for many countries but the process for that is still in its early stages. To understand how this policy agenda is evolving, an instrument-based approach is advisable. The definition of policy instruments involves the idea of government accomplishing goals, as Salamon (2002:19) puts it: "an identifiable method through which collective action is structured to address a public problem." These instruments can be new government agencies with specific purposes and responsibilities towards checking the work of others, new laws or procedures, or guidance to agencies and civil servants implementing new laws and procedures.

Policy instruments can be categorised based on their purpose. Substantive instruments alter the distribution of goods and services, while procedural instruments influence policy outcomes by changing the players and rules of the policymaking process. Table 1 below provides a comprehensive overview of policy instruments employed to address issues related to the theme of algorithmic accountability. These include documents, such as white papers and directives, legal instruments, such as bills, and the creation or convening of new groups and institutions, such as agencies or councils.

Table 1 – Policy Instruments of Algorithmic Accountability

Country or Region	Initiative	Instrument Type	Govt Function	Goal	Other features
European Union	EU AI Act and AI Liability Directive	Procedural and substantive	Convener and regulator	To foster safe, transparent, traceable, non-discriminatory, and environmentally friendly systems	The bill establishes that humans must supervise algorithmic systems. Systems with unacceptable risks, such as those manipulating cognitive-behavioral or social scoring, will generally be prohibited. Systems with high risks will be evaluated before being placed on the market

Country or Region	Initiative	Instrument Type	Govt Function	Goal	Other features
United States (District of Columbia - DC)	Stop Discrimination by Algorithms Act	Substantive	Convener and regulator	To prevent the algorithm effects of discrimination on race, color, religion, national origin, sex, gender identity or expression, sexual orientation, familial status, source of income, or disability	The bill prohibits both for-profit and non-profit organizations from using algorithms that make decisions based on protected personal traits
United States	AI Bill of Rights Principles	Substantive	Convener	To establish a guide for a society to protect the American public in the age of artificial intelligence	Five principles that should guide the design, use, and deployment of automated systems: Safe and Effective Systems; Algorithmic Discrimination Protections; Data Privacy; Human Alternatives, Consideration, and Fallback
Singapore	Model AI Governance Framework	Substantive	Convener	A model that seeks to translate ethical principles into implementable practices in the AI development process. An algorithm must be "explainable" or "interpretable"	To establish mechanisms that allow for preventing and eliminating errors, which can occur both from the algorithms used and also from the databases used for their training
France	Digital Republic Bill	Substantive	Regulator	A new legal framework for algorithmic accountability and transparency will require public agencies to ensure that both existing and future algorithms comply with new obrigations. This framework will also grant citizens rights, including an expanded right to information.	The law principles are the default opening of public data, net neutrality, an obligation of loyalty for online platforms, as well as increased protection for the personal data of Internet users
Canada	Artificial Intelligence and Data Act (AIDA)	Substantive	Regulator	This code temporarily provides Canadian companies with common standards and enables them to demonstrate, voluntarily, that they are developing and using generative AI systems responsibly until formal regulation is in effect	AIDA will be the foundation for the responsible design, development, and deployment of AI systems to ensure that AI systems deployed in Canada are safe and non-discriminatory and will hold businesses accountable for how they develop and use these technologies

Country or Region	Initiative	Instrument Type	Govt Function	Goal	Other features
Finland, Germany, the Netherlands, Norway, and the UK	Auditing machine learning algorithms (white paper)	Substantive	Convener	To safeguard personal data rights; inexplicable and therefore unjustifiable decisions; or potentially institutionalized discrimination by algorithmic bias	A group of public audit agencies created a catalog with a set of guidelines based on risks and methodology to perform audit tests
Spain	Spanish Artificial Intelligence Supervision Agency (AESIA)	Substantive	Regulator	To inspect, verify and sanction AI systems focused on responsible, reliable, and sustainable use of algorithms to protect the user and avoid discrimination	Pioneering government entity with direct control, monitoring, and regulation over AI, both for the public and private sectors
Chile	Chilean Transparency Council			In a first for the Latin American region, the independent council (not governmental) is developing an open and participatory design for a binding "General Instruction on Algorithmic Transparency" for public entities	The general instruction will mandate more than a thousand public agencies to report the algorithms they use to serve the population, as a further obligation of active transparency
The Netherlands	Fundamental Rights and Algorithms Impact Assessment (FRAIA)	Substantive	Convener	To facilitate an interdisciplinary dialogue to help identify the risks to human rights from the use of algorithms and determine measures to address these risks	FRAIA aims to ensure that all relevant focus areas regarding the use of algorithms are addressed at an early stage and in a structured manner. This prevents the premature use of an algorithm that has not been adequately assessed in terms of the consequences
UK	Algorithmic Transparency Recording Standard (ATRS)	Substantive	Convener	To provide a clear and accessible format and mechanism designed to support public sector bodies providing information about the algorithmic tools they use in decision-making processes that affect members of the public	The Standard is designed to be an enabler for more effective and joined-up use of algorithmic tools to support public service delivery with transparency

Though this is not an exhaustive review, it is possible to see that developed countries are more strongly present, likely to their higher ability both to develop AI, from the private sector perspective, and to regulate it, from a state capacity point of view. It is also worth noting that while most algorithmic accountability initiatives currently emphasise transparency, many are also integrating risk-based mitigation approaches. However, fewer demonstrate the capability for practical algorithm auditing, which would effectively complete the accountability cycle by ensuring AI's responsible and trustworthy utilisation in real-world applications.

So, what lessons can Brazil draw from these experiences to enhance its AI governance framework? Regarding the key provisions of Bill 2.338/2023 discussed in the previous section, a range of fronts must be covered and complementary initiatives must be designed and implemented, as detailed below:

- Human Rights-Centric Approach: Inform individuals about AI interactions, explaining the AI-driven decisions; ensure non-discrimination and correct biases in AI systems through mandatory procedures; and protect individuals' privacy rights;
- Transparency and Explicability: Set clear guidelines requiring disclosure of AI systems' design, data sources, and decision-making processes; require robust techniques to make AI decisions comprehensible to non-technical users; develop and mandate advanced tools for detecting and mitigating bias in AI systems; and engage the public and stakeholders in feedback processes to refine AI transparency practices and address real-world concerns;
- Risk-based Regulation: Establish dynamic, continuously updated risk assessment protocols; encourage proactive risk management by AI developers through self-assessments and regular reviews; customize regulations based on specific risks, with stricter oversight for high-risk systems; and collaborate internationally to adopt best practices and harmonise AI regulation standards;
- Supervisory Authority: Empower a centralised body to oversee AI regulation and enforcement; coordinate with sectoral regulators for industry-specific AI challenges; train civil service in AI ethics, law, and technology; and continuously support and implement the National AI Strategy (EBIA) to promote innovation and responsible AI use.

In sum, Brazil must build state capacity that, through an integrated and holistic approach, promotes transparency, effectively manages risks, and establishes a robust supervisory authority to oversee AI systems. The most recent rapporteur's report on PL/2388, at the time of writing, designates the Autoridade Nacional de Proteção de Dados (ANPD) – the Brazilian National Data Protection Authority (ANPD) – as the

centralised body responsible for coordinating inter-agency efforts in a broader AI regulating system[7]. This is in line with the recommendations of a previous Lemann Foundation Programme policy brief[8], emphasising coordination of existing regulators in the specific case of digital platforms, while recognising that AI regulation involves a broader set of concerns that might require a different set of arrangements and take new institutional forms.

6. CONCLUSION

The primary objective of this working paper is to contribute to the debate surrounding AI governance in general and algorithmic accountability in specific, with a focus on the Brazilian context. Despite being part of the growing adoption of AI, especially in public services, Brazil has yet to make progress in strengthening the accountability aspect of its AI governance strategy. State initiatives have been relatively shy and lagging behind the prevailing trend of algorithmic accountability embraced by many other countries, as described in the previous section.

Compared with other countries in Latin America and the Caribbean (LAC) region, Brazil finds itself in the second tier of capacities. While there is alignment with the OECD AI Principles and a commitment to implementation, the country's AI governance framework is limited to a data committee and the Brazilian Artificial Intelligence Strategy (EBIA). Though undoubtedly relevant, these policy instruments do not adequately cover AI consequences as the required priority. To inform this discussion, the paper presented examples of initiatives from around the world, of various natures and at different stages of implementation, that have been designed to address different facets of AI consequences, both in the public and private sectors.

Most of these policy instruments fall under the substantive category, with only the EU AI Act incorporating procedural features. Some of these initiatives are guidelines, while most involve legal changes, often yet to be approved, which highlights the lack of internal political consensus on this matter. Concerning governmental roles, their focus span from convener to regulator. However, this does not mean these countries are not also acting in separate realms as financiers, direct users, or co-developers, as the OECD (2020) model outlined.

---

[7] https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-e-formalizada-como-coordenadora-do-sistema-nacional-de-inteligencia-artificial

[8] Beatriz Kira. "The Path Forward for the Brazilian National Data Protection Authority (ANPD): Strengthening Capacity and Autonomy." 12 April 2023. Blavatnik School of Government, University of Oxford https://www.bsg.ox.ac.uk/research/publications/path-forward-brazilian-national-data-protection-authority-anpd-strengthening

The goals of these initiatives are diverse, encompassing non-discrimination in various forms, improving transparency, explicability, and accountability for managers and policymakers, data protection, and environmental concerns, among others. Notably, these instruments fail to mention or address digital exclusion. Although not a direct consequence of AI, the effects of digital exclusion can be exacerbated in its wake. This is a significant concern in countries with a relevant digital divide such as Brazil. These experiences, whether in the formulation or implementation phase, share a commonality: they lack concrete evidence of real-world outcomes, and will require *ex-ante* or *ex-post* evaluations to be deemed effective or not in achieving their intended purposes.

In Brazil, there are promising developments in the Legislative, with a bill in the Federal Senate proposing to establish a framework for responsible AI governance. The bill aligns with the OECD recommendations and the principles and features outlined in international best practices discussed in this paper, encompassing four key dimensions of algorithmic accountability: i) adopting a human rights-centric approach, addressing concerns related to non-discrimination, correction of biases, and privacy protection; ii) emphasising transparency and explicability; iii) focusing on a risk-based approach and the necessity for both pre- and post- evaluation; and iv) proposing the creation of a supervisory authority. The proposal also distinguishes itself by advocating for co-regulation and encouraging AI agents to voluntarily adopt good practice policies and governance measures, and by doing so, mitigating any potential administrative penalties.

However, it should be noted that, even if approved, the bill may not be sufficient to address the challenges posed by the widespread use of this new disruptive technology. Additionally, AI governance must bridge the difficult balance between accountability mechanisms and the ongoing promotion of innovation. The chosen policy instruments, no matter how sophisticated, will only be effective if the government invests in building state capacity to implement them – which entails leadership, coordination, and gathering support for AI.

In summary, crafting a robust governance framework for AI focusing on algorithmic accountability is a challenging process of learning, adaptation, and experimentation, marked by progress and setbacks. For AI governance to accomplish its comprehensive goals, it must recognize that there is an intricate interplay between technology and society. This outlook is essential for fostering responsible, trustworthy, democratic, inclusive, and human-centered AI implementation in both the public and private sectors.

7. REFERENCES

Allhutter, D., Cech, F., Fischer, F., Grill, G. & Mager, A. (2020). Algorithmic Profiling of Job Seekers in Austria: How Austerity Politics Are Made Effective. Frontiers in Big Data, volume 3, Article 5.

Almeida, V.; Filgueiras, F. & Mendonça, R. F. (2022). Algorithms and Institutions: How Social Sciences Can Contribute to Governance of Algorithms, in IEEE Internet Computing, vol. 26, no. 2, pp. 42-46, doi: 10.1109/MIC.2022.3147923.

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://doi.org/10.1787/726fd39d-en.

Brazil (2021). Estratégia Brasileira de Inteligência Artificial – EBIA. MCTI. Avaliable at https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial (accessed on October 11th 2023).

Caruso, F. (2022). Serbia, algorithmic discrimination rehearsals. Available at https://www.balcanicaucaso.org/eng/Areas/Serbia/Serbia-algorithmic-discrimination-rehearsals-222242 (accessed on October 11th 2023).

Henley, J. (2021). Dutch government faces collapse over child benefits scandal. The Guardian. Archived from the original on January 14th 2021.

Howlett, M. (2011). Designing public policies: Principles and instruments. London: Routledge.

Lythreatis, S.; Singh, S. K. & El-Kassar, Abdul-Nasser (2022). The digital divide: A review and future research agenda. Technological Forecasting and Social Change, Volume 175.

OECD. (2020). Going Digital in Brazil, OECD Reviews of Digital Transformation, OECD Publishing, Paris.

OECD (2020).The OECD Digital Government Policy Framework: Six dimensions of a Digital Government. OECD Public Governance Policy Papers, No. 02, OECD Publishing, Paris, https://doi.org/10.1787/f64fed2a-en.

OECD & CAF (2022). The Strategic and Responsible Use of Artificial Intelligence in the Public Sector of Latin America and the Caribbean, OECD Public Governance Reviews, OECD Publishing, Paris.

OECD. Recommendation of the Council on Digital Government Strategies (2014). www.oecd.org/gov/digital-government/recommendation-on-digital-government-strategies.htm. (accessed on October 11th 2023).

O'Neil, C. (2017). Weapons of math destruction. Penguin Books.

Salamon, L. M. (2002). The tools of government: a guide to the new governance. Oxford: Oxford University Press.

Supreme Audit Institutions of Finland, Germany, the Netherlands, Norway and the UK Auditing Machine Learning Algorithms: A White Paper for Public Auditors.

Vedung, E. (1998). Policy instruments: Typologies and theories. In M.-L. Bemelmans-Videc, R. Rist, & E. Vedung (Eds.). Carrots, sticks and sermons: Policy instruments and their evaluation (pp. 21–58). New Brunswick, NJ: Transaction.

Vial, G. (2019). Understanding digital transformation: a review and a research agenda. The Journal of Strategic Information Systems. 28 (2): 118–144.

World Bank (2022). GovTech Maturity Index, 2022 Update: Trends in Public Sector Digital Transformation. Equitable Growth, Finance and Institutions Insight – Governance. Washington, DC.