Artificial intelligence in public management: risks, global measures, and the brazilian context[1]

Inteligência artificial na gestão pública: riscos, medidas globais e o contexto brasileiro

Inteligencia artificial en la gestión pública: riesgos, medidas globales y el contexto brasileño

Pedro Luiz Costa Cavalcante

https://doi.org/10.36428/revistadacgu.v18i33.730

Abstract: This paper analyzes the risks and governance challenges associated with the adoption of artificial intelligence in the Brazilian public sector, drawing on international experiences for comparative purposes. It examines the potential impacts of AI on service quality, equity, and public integrity in government decision-making processes and policy outputs. Meanwhile, it explorer emerging global initiatives aimed at strengthening algorithmic accountability. The paper discusses both the opportunities and risks of AI use, presents the current Brazilian AI governance framework—including the national strategy and legislative developments—and compares these efforts with selected international approaches. Based on this analysis, it offers context-sensitive insights to advance accountability mechanisms in Brazil. The findings suggest that, although Brazil has engaged in the early stages of AI adoption, important gaps remain in the institutionalization of accountability frameworks. Addressing these gaps is essential for fostering responsible, trustworthy, and human-centered AI in the public sector.

Keywords: artificial intelligence, public integrity, digital transformation, innovation

Resumo: Este artigo analisa os riscos e os desafios de governança associados à adoção da inteligência artificial no setor público brasileiro, com base em experiências internacionais para fins comparativos. O trabalha examina os impactos potenciais da IA na qualidade dos serviços, na equidade e na integridade pública nos processos decisórios governamentais e nos resultados das políticas públicas. Ao mesmo tempo, explora iniciativas globais emergentes voltadas ao fortalecimento da accountability algorítmica. O trabalho discute oportunidades e riscos do uso da IA, apresenta o arcabouço brasileiro de governança — incluindo a estratégia nacional e as propostas legislativas, e compara esses esforços com abordagens internacionais selecionadas. A partir dessa análise, o artigo oferece reflexões contextualizadas para o progresso de mecanismos de accountability no Brasil. Os achados indicam que, embora o país tenha avançado na adoção inicial da IA, persistem lacunas importantes na institucionalização de mecanismos de responsabilização, cuja superação é essencial para promover uma IA responsável, confiável e centrada no ser humano no setor público.

Palavras-chave: inteligência artificial, integridade pública, transformação digital, inovação

1. Artigo submetido em 14/03/2024 e aceito em 14/04/2026.

CC BY

Resumen: Este artículo analiza los riesgos y los desafíos de gobernanza asociados con la adopción de la inteligencia artificial en el sector público brasileño, tomando como referencia experiencias internacionales con fines comparativos. El trabajo examina los impactos potenciales de la IA en la calidad de los servicios, la equidad y la integridad pública en los procesos de toma de decisiones gubernamentales y en los resultados de las políticas públicas. Al mismo tiempo, explora iniciativas globales emergentes orientadas al fortalecimiento de la rendición de cuentas algorítmica. El estudio discute las oportunidades y los riesgos del uso de la IA, presenta el marco brasileño de gobernanza —incluyendo la estrategia nacional y las propuestas legislativas— y compara estos esfuerzos con enfoques internacionales seleccionados. A partir de este análisis, el artículo ofrece reflexiones contextualizadas para el avance de mecanismos de rendición de cuentas en Brasil. Los hallazgos indican que, aunque el país ha avanzado en la adopción inicial de la IA, persisten importantes vacíos en la institucionalización de mecanismos de rendición de cuentas, cuya superación es esencial para promover una IA responsable, confiable y centrada en el ser humano en el sector público.

Palabras clave: inteligencia artificial, integridad pública, transformación digital, innovación

1. Introduction

To address the challenges imposed by social, economic, environmental, and technological dynamics, 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. Consequently, the concept of digital government has permeated various spheres and government sectors, affirming the perception that this trend is no longer just a novelty but has indeed become the new normal in public management.

Digital transformation refers to the process of enhancing organizational performance through the utilization of information and computer-based technology resources (Vial, 2019). 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. Its primary goal is to add value regarding transparency, accountability, efficiency, effectiveness, customer experience, and service delivery for both businesses and public organizations.

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 interaction with citizens in policymaking. The transition to digital government comprises a series of strategies to modernize government operations to benefit society. As highlighted by the OECD (2014, p. 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 emerging countries. 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 guided by networking principles.As a result, Brazil has been recognized by the World Bank as one of the leading countries in digital government maturity, ranking second globally in the 2022 GovTech Maturity Index. The country also stood out for achieving the most significant progress among the 98 countries evaluated. This advancement is largely associated with the expansion of digital public services through the Gov.br platform, which currently serves over 140 million users—approximately 80% of the adult population. The platform plays a central role in facilitating access to public services and improving interactions between citizens and the government.

However, it is important to note that different international indexes provide varying assessments of digital government maturity. Indicators such as the United Nations E-Government Development Index

CC BY

(eGDI) and the OECD Digital Government Index (DGI) position Brazil at comparatively lower levels, reflecting differences in methodology and evaluation criteria. Moreover, high performance in digital government indexes primarily reflects the availability and integration of digital services and infrastructure, rather than the effective deployment of artificial intelligence in public decision-making processes.

In brief, digital transformation is a process that empowers governments to harness technology's potential to reinforce fiscal transparency and accountability, boost the effectiveness of public expenditure, and enhance outcomes in education, healthcare service delivery, and social welfare. This represents a foundational shift in how governments function, utilizing technology to streamline operations, facilitate decision-making, manage data, engage with society, and provide a better citizen experience. 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.

The case of AI is particularly noteworthy as it has developed into a global trend that significantly influences governments' policy decisions and implementation, producing significant changes in policymaking since 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).

Using algorithms in policy decision-making is reshaping public services and economies, offering the potential for increased productivity, enhanced efficiency, and reduced costs. For instance, governments in Latin America and the Caribbean actively explore AI applications within public administration for various purposes. These include responding to the challenges posed by COVID-19, optimizing 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).

In this paper, key concepts such as accountability, transparency, integrity, and equity are understood within the framework of public sector governance. Accountability refers to the ability to assign responsibility for algorithmic decisions and ensure mechanisms for oversight and redress. Transparency concerns the availability of information regarding how algorithmic systems operate and influence decisions. Integrity relates to the alignment of such systems with ethical standards and the protection of public values, while equity refers to the fair and non-discriminatory treatment of individuals affected by automated decision-making processes. These concepts are mobilized as analytical lenses to examine the governance challenges associated with AI, rather than as objects of independent theoretical elaboration.

Indeed, despite the good purposes and benefits, not everything is rosy since AI has also ushered in a host of intricate challenges, risks, and setbacks for governments, society, and businesses alike, such as safeguarding private data and privacy. Furthermore, two other issues threaten the government's ability to deliver public services effectively, transparently, and equitably: digital divide or exclusion and algorithmic discrimination.

The digital divide represents a phenomenon that entails disparities in access, utilization, 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 becoming increasingly worrisome because a substantial portion of the population has limited or no access to the internet. For instance, the impact became increasingly apparent during the Covid-19 pandemic, particularly concerning low-income students attending public schools who encountered greater difficulties in accessing classes due to the precariousness of their internet connections.

On the other hand, algorithmic bias refers to systematic and repeatable errors in a computer system that produce unfair outcomes, such as favoring one category 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 artificial intelligence and machine learning, in which decisions rely on a dataset of inputs and other learning and decision-making techniques. Consequently, this issue has raised significant concerns regarding the fairness, justice, and transparency of criteria and automated decision-making, as well as the dehu-

CC BY

manization of services. These concerns have negatively impacted the quality of service provision and the promotion of diversity. Depending on the specific circumstances, such algorithmic discrimination may breach legal protections and perpetuate unfairness and inequalities in society.

Therefore, this policy paper aims to analyze the effects of the widespread adoption of AI, examining the potential risks it presents to service quality, equity, and transparency. Furthermore, it aims to shed light on ongoing global initiatives that foster accountability in artificial intelligence. 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 process and service effectiveness, the protection of citizens' rights, trust in government and public integrity.. For the purpose of this paper, integrity is understood primarily as public integrity, referring to the alignment of public decision-making processes with ethical standards, accountability mechanisms, and the protection of citizens' rights in the context of AI-enabled governance.

In this sense, the contributes to the debate on algorithmic accountability by examining the Brazilian public sector as its primary empirical focus, supported by selected international experiences. These case are used in an illustrative manner, aiming to contextualize the Brazilian case rather than to provide a normative or evaluative benchmark. To achieve this, it first:

1. discusses AI use's positive and negative effects, providing global examples of algorithmic discrimination. Then, the article

2. presents how the Brazilian public sector has structured and implemented its AI governance strategy

    a. and highlights the AI bill's critical points currently under discussion in the nation's parliament.

3. Next, the study scrutinizes how governments worldwide are designing transparency and accountability mechanisms to address the r

    a. repercussions of AI while aligning them with ethical and integrity standards, as well as participatory and inclusive principles valued by society.

    b. Based on these insights, some insights are offered for advancing the design and implementation of practical initiatives within the Brazilian public sector.

2. AI proliferation and consequences

Governments worldwide are increasingly employing algorithms and Artificial Intelligence (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 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 minimizing opportunities for corruption. As stated by OECD (2020), this continually evolving technology tends to make the public sector more intelligent, manifesting in increased agility, efficiency, 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." (OECD, 2020, p. 13)*

Nonetheless, using AI and algorithmic systems in public service delivery comes with inherent risks, as demonstrated by evidence indicating that they 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.

CC BY

The range of issues and setbacks associated with employing algorithms for decision-making 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 occurring in economically disadvantaged areas, perpetuating the bias that crime is higher in less affluent regions within major cities. 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 USA (O'Neil, 2017);
- In the context of child benefits, a scandal in the Netherlands, known as the "Toeslagenaffaire," emerged due to the use of an algorithm. It resulted in tens of thousands of often vulnerable families being falsely accused of fraud, separating hundreds of children 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. Consequently, 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; social values, norms, and objectives also influence it. This interplay has given rise to tensions, 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.

> In Brazil, Evidence suggests that the use of AI-related and data-driven tools in the Brazilian public sector, although not yet widespread, has expanded across key institutional domains. In tax administration, the Federal Revenue Service employs advanced data analytics and algorithmic risk assessment to support compliance and target audits (OECD, 2020; OECD, 2023). In the judiciary, courts have adopted machine learning systems—such as Victor—to assist in case classification and prioritization (Almeida, Filgueiras & Mendonça, 2022; CNJ, 2021). Similarly, the Federal Court of Accounts has incorporated data analytics and automated monitoring to strengthen auditing and detect irregularities (OECD, 2020; TCU, 2022). At the same time, the Gov.br platform provides the digital infrastructure that enables the expansion of data-driven public services at scale (World Bank, 2022; OECD, 2020), while the Office of the Comptroller General has increasingly relied on data integration to enhance fraud detection and integrity mechanisms (OECD, 2020; CGU, 2022). Taken together, these initiatives indicate a pattern of incremental and institutionally concentrated adoption, rather than a fully diffused or consolidated use of AI across government.

Despite these advances, evidence also points to unintended consequences associated with the use of automated decision-making systems, particularly when governance and accountability mechanisms are insufficiently developed. 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) resulted in a combination of algorithmic bias and a digital divide.

the situation is similar. 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) resulted in a combination of algorithmic bias and a digital divide. A recent Federal Audit Court (TCU) audit identified several issues with this approach. Notably, using algorithms significantly increased the rate of denials without providing adequate explanations to policyholders. Furthermore, the automation strategy was not accompanied by staff replacement needed to analyze the benefit demands, which led to longer waiting times and extensions of processing deadlines, sometimes four times longer than stipulated by legislation. The TCU report also indicated that the INSS initiative needs to have basic transparency standards and prioritize the interests of citizens. By diverting requests to the appeals court, it discourages the recognition of le-

CC BY

gitimate rights, fails to protect citizens, and exacerbates the already prominent digital exclusion in this policy area.

In response to these challenges, a recent trend called algorithmic accountability has emerged, as emphasized in the latest OECD report, "Global Trends in Government Innovation 2023." This emerging approach, guided by democratic and integrity principles, involves actions to hold accountable those who create, procure, and employ algorithms for their outcomes. Consequently, these managers and organizations should be obligated to improve the transparency of the values and criteria embedded in their algorithms, mitigate associated risks, and take responsibility for the results they generate.

While governments are increasingly integrating AI into policymaking, they are also actively working to ensure that the algorithms, which may appear promising initially, are free from bias and discrimination. They are also focusing on ensuring that public servants have a strong understanding of data ethics. Moreover, official entities and external stakeholders promote algorithmic accountability, emphasizing transparency and explainability to build trust with citizens and prevent injustices in public services (OECD, 2023).

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

Hence, the 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 their public organizations and servants in using algorithms in public service delivery according to ethical principles. So, it raises the question: How are Brazilian data governance stakeholders addressing these challenges?

3. Algorithmic accountability in the brazilian context

In this section, the paper will delineate how the issue is currently being addressed in Brazil and the global initiatives that can serve as a reference for advancing this debate within the country's context. To begin with, it is worth noting thatRecent evidence suggests that AI-related applications have been gradually incorporated into the Brazilian public sector, particularly in areas such as tax administration, judicial systems, and social policy management. For instance, public institutions have adopted data-driven tools for risk analysis, fraud detection, and case prioritization, while digital government platforms have expanded the use of data analytics to improve service delivery. However, these initiatives remain fragmented and unevenly distributed across institutions, and comprehensive data on their scale and impact is still limited[2].

According to Berryhill et al. (2019), several countries have formally addressed this issue by implementing strategies in the public sector AI domain. These strategies demonstrate its integration into policymaking and innovative service design processes. However, it is essential to note that these countries are at various stages of development. These strategies also 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 for citizens. They also entail cross-government, cross-sector, and international collaboration through councils, networks, communities, and partnerships. Furthermore, strategic management and the utilization of government data, including open data, play a pivotal role in fueling AI adoption across all sectors.

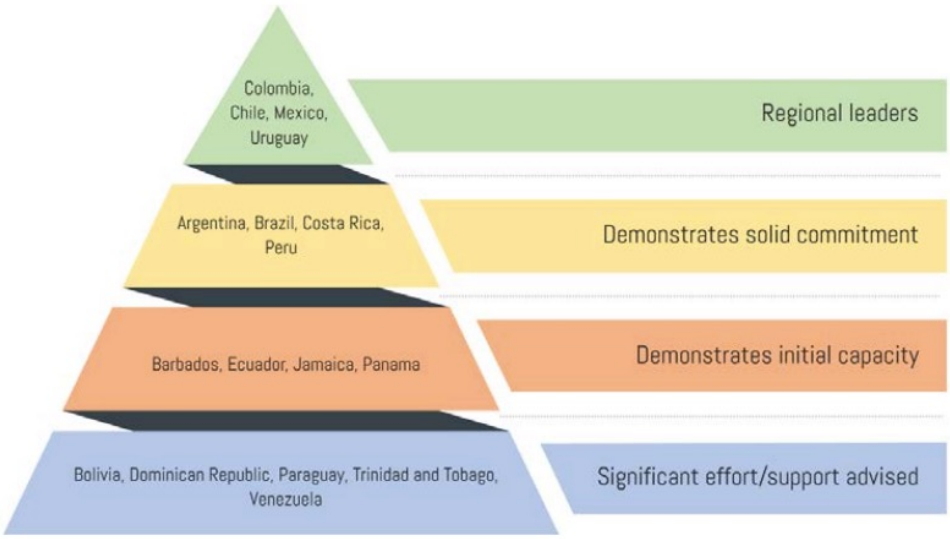
The heterogeneity is also observed inside regions, such as Latin America and the Caribbean

2. See more at https://oecd.ai/en/dashboards/countries/Brazil.

CC BY

(LAC), in which AI proliferation is not only fast-speed but also conducted with different levels of governance maturity regarding priorities, public investments, and regulations. Given the increasing importance of AI in shaping policies and its potential impact on the digital economy, more than 60 countries in the region are actively developing national AI strategies (OECD & CAF, 2022). The figure depicted below (Figure 1) illustrates these diverse levels of development:

FIGURE 1 • LAC REGION CAPACITIES FOR AI LEGAL AND ETHICAL FRAMEWORKS



Source: OECD & CAF, 2022.

Brazil occupies the second tier of the capacity pyramid, indicating its commitment to the OECD AI Principles and a willingness to implement them, although with a lower degree of maturity than regional leaders. Three noteworthy instruments in this context are the Data Governance Committee, the Brazilian Artificial Intelligence Strategy (Estratégia Brasileira de Inteligência Artificial - EBIA) and the new Bill of AI, now in the Federal Senate.

The Data Governance Committee, established by Decree in 2019, possesses the authority to make decisions, including those about the principles and guidelines for classifying broad, restricted, and specific data sharing, 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.

Turning to the EBIA, a formal document, it recognizes the enormous potential of AI to benefit people worldwide. However, as AI's impact on society grows, it becomes crucial to ensure responsible AI use and development, safeguarding fairness, safety, and privacy. Led by the Ministry of Science, Technology, and Innovation, the strategy seeks to advance technology's development and use, contributing to scientific progress and problem-solving in the country's priority areas. The expected benefits of AI encompass enhanced competitiveness, increased productivity, improved public services, enhanced quality of life, and reduced social inequalities, among others.

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 Artificial Intelligence.

CC BY

Despite these measures, the committee resolutions and the EBIA lack clear and operational goals for addressing the adverse effects of AI applications, such as bias, discrimination, and issues related to race, gender inequalities, and digital exclusion in both the public and private sectors. More recently, the Brazilian Artificial Intelligence Plan (PBIA 2024–2028) has sought to address part of this gap by introducing a more implementation-oriented approach, including defined priorities, governance mechanisms, and planned investments. The plan explicitly incorporates the use of AI to improve public services and signals a shift toward a more coordinated policy agenda. However, it remains uncertain to what extent these initiatives will effectively incorporate safeguards related to inequality, bias, and accountability, particularly in contexts marked by structural disparities.

In this regard, evidence on digital exclusion remains a critical concern. Although the latest 2023 TIC Households report from the Brazilian Internet Steering Committee (CGI.br) shows that Internet connectivity has increased, particularly among lower-income groups, approximately 29 million people are still not Internet users[3]. The report also highlights persistent inequalities in access and digital skills, disproportionately affecting individuals with lower levels of education, older populations, and those living in urban peripheries. While the expansion of connectivity represents an important advance, disparities in the quality of access continue to limit the development of digital capabilities and constrain the full realization of the benefits associated with digital transformation and AI-enabled public services.

In this context, another facet of the Legislative branch appears to be more advanced in addressing the adverse consequences and risks associated with AI. Since 2019, the National Congress has actively pursued the responsible development of AI systems by introducing and deliberating upon several bills. Given Brazil's bicameral system, proposals originating in one house undergo scrutiny by the other. The most proemiment bill, nº 2.338/2023, was approved by the Federal Senate in December of 2024 and is currently under review in the Chamber of Deputies. As of early 2026, the proposal is being analyzed by a Special Committee established to examine its content, alongside additional legislative initiatives that seek to complement and refine the regulatory framework[4].The Bill nº 2.338/2023[5], initiated by the President of the Federal Senate, aligns with assumptions and guidelines derived from legislative initiatives in the European Union and the United States and the aligned with the principles defined by the OECD AI recommendations[6].

The project's objective is two-fold: i) to establish rights that safeguard the most vulnerable party involved—the individual consistently impacted by artificial intelligence systems, spanning from content recommendations and targeted online advertising to assessments of eligibility for credit and specific public policies; ii) by implementing governance tools and an institutional framework for oversight and supervision, the initiative promotes conditions of predictability regarding its interpretation. Additionally, the bill aims to provide legal certainty for innovation and technological development. In a nutshell, the Bill nº 2.338/2023 has the following key provisions worth highlighting in this debate on algorithmic accountability:

- ***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:*** The project enumerates measures to ensure transparency and mitigate bias and standardizes the procedure for algorithmic impact assessment. The bill 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;
- ***Risk-based Regulation:*** The bill introduces a tiered risk classification system by delineating three risk levels: (i) excessive risk, warranting prohibition; (ii) high risk; and (iii) non-high risk. Prior to it, an AI provider is obligated to conduct a comprehensive self-assessment analysis for risk classification and, in the case of high-risk AI systems, additional actions such as reliability tests, measures to mitigate discriminatory biases, and technical explainability measures will be required.

3. See https://cetic.br/en/.
4. See https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2487262.
5. See https://www25.senado.leg.br/web/atividade/materias/-/materia/157233.
6. See https://oecd.ai/en/ai-principles.

CC BY

The option to regulate according to risk aims to only regulate what is essential, thereby avoiding unnecessary restrictions on systems that do not pose a high risk;

- ***Supervisory Authority:*** It 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.

Clearly, this law proposal, if approved in in the Chamber of Deputies, Brazil will contribute to a human-centric, inclusive, non-discriminatory, responsible, and ethical AI approach that not only augments the quality of life for individuals and mitigates the digital divide but also stands as a benchmark for other developing nations struggling with similar challenges.

Although these policy instruments are relatively recent or in debate, it is reasonable to assume that these efforts aim to address algorithmic accountability's foundations. So, the question that may help the initiatives to become effective is: What lessons or policy features from international best practices could enhance AI governance in Brazil?

4. 'Good' AI governance: functions and good practices

Indeed, the governance of Artificial Intelligence applications is a complex endeavor. As Almeida, Filgueiras and Mendonça (2022) point out, algorithm governance comprises a set of practices aimed at controlling, shaping, and regulating algorithms and their impacts. This field is unique due to its intricacies, complexity, and a certain degree of unpredictability. A fundamental starting point for establishing effective governance is defining the guiding principles and values upon which the framework is built, which typically encompass fairness, transparency, integrity, accountability, and explainability. However, translating these concepts into practical actions and implementation within a real-world public administration context presents significant challenges. While technology is a critical component, most of the variables in this governance framework stem from social, political, and economic dimensions within a democratic context (Almeida, Filgueiras & Mendonça, 2022).

A framework that upholds and enforces these ethical values to ensure AI is trustworthy, fair, inclusive, and accountable must always prioritize a human-centered approach. It should focus on structuring governance capacities for decision coherence, enforcement, and monitoring, as emphasized 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: by 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, which is the ultimate goal for many nations. However, this process is still in its early stages. Therefore, to comprehend the evolution of this policy agenda, it is prudent to focus on an instrument-based approach. 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." According to Vedung (1998: 21), it is "a set of techniques by which governmental authorities wield their power in attempting to ensure support and affect or prevent social change," while Howlett (2011: 415) defines it as "to deliberately affect the nature, types, quantities, and distribution of the goods and services provided in a society."

These policy instruments can be categorized 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. Alternatively, it can also cover their goals and particular features. Table 1 below provides a comprehensive overview of policy instruments employed to address issues related to algorithmic accountability, covering all these components.

CC BY

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	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
United States (District of Columbia - DC)	Stop Discrimination by Algorithms Act	Substantive	Convener and regulator	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; 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 obligating public agencies to be accompanied by making existing and future algorithms compliant with the new obligations and citizens can have access to new rights, such as an extended 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

CC BY

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 supreme agencies' audit 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 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	ATRS provides 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

It is important to emphasize that the international experiences presented in this paper are not intended to provide a comprehensive or evaluative assessment of AI governance frameworks. Rather, they illustrate the diversity of approaches adopted by different countries in addressing issues such as transparency, accountability, and risk management. These initiatives are at different stages of development—some already implemented, others still under discussion or lacking robust empirical validation.

Moreover, no country has yet fully addressed all the challenges associated with artificial intelligence, including issues related to transparency, accountability, algorithmic bias, and digital exclusion. In the Brazilian context, these challenges are further shaped by structural factors such as socioeconomic inequality and persistent digital divides.

Therefore, the purpose of this comparison is to map the variety of policy instruments and governance strategies currently emerging, rather than to

CC BY

assess their implementation or effectiveness or to establish prescriptive benchmarks. From this analytical perspective, it is possible to identify relevant patterns and emerging features that may inform the Brazilian debate. In particular, while most algorithmic accountability initiatives emphasize transparency, many are also incorporating risk-based mitigation approaches. However, fewer demonstrate the capacity for practical algorithm auditing, which would help complete the accountability cycle by ensuring the responsible and trustworthy use of AI in real-world applications.

6. Final remarks

The primary objective of this policy paper is to contribute to the debate on algorithmic accountability, with a specific focus on the Brazilian context. Although Brazil has made progress in digital government and has initiated the use of AI-related tools in specific domains, the adoption of artificial intelligence remains partial and heterogeneous, with varying levels of institutionalization. While the country is part of the broader trend of AI adoption—particularly in public services—the development of accountability mechanisms within its AI governance framework is still evolving. In this context, state initiatives remain in an early and uneven stage, reflecting ongoing efforts to structure and consolidate this policy agenda.

Compared with other countries in Latin America and the Caribbean (LAC), 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 the country's AI governance framework has evolved in recent years, incorporating instruments such as the EBIA, the PBIA, and ongoing legislative initiatives, although important gaps remain in the institutionalization of accountability mechanisms. . Nevertheless, these instruments do not adequately cover AI consequences as the required priority. To facilitate this discussion, the paper presented examples of successful global initiatives that, at various stages of implementation, have been designed to address different facets of AI consequences, both in the public and private sectors.

Table 1 also reveals that 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. However, it is worth noting that some of these legal changes have yet to be approved, which highlights the lack of consensus on this matter within political systems. Concerning governmental roles, they span from a convener to a regulator. However, this does not mean these countries are not acting as financiers, direct users, or co-developers, as OECD (2020) outlined. These initiatives have different areas of attention.

The goals of these initiatives are quite diverse, encompassing non-discrimination in various forms, improving transparency, explicability, and accountability for managers and policymakers, as well as addressing data protection and environmental concerns, among others. Notably, these experiences should have specifically mentioned digital exclusion as an aspect to be addressed. Although it may not be considered a direct consequence of AI applications, it can be exacerbated in the context of a high digital divide, which is the case in Brazil. It is essential to highlight that these experiences, whether already implemented or in the formulation phase, share a commonality: they lack concrete evidence of real-world outcomes. While well-intentioned efforts, they require ex-ante or ex-post evaluations to be deemed truly effective in achieving their intended purposes.

In the Brazilian case, the good news comes from the Legislative branch, which has advanced the debate on AI regulation through a bill in the Federal Senate to establish a framework for responsible artificial intelligence governance. The proposal not only aligns with the OECD recommendations but also closely adheres to the principles and features outlined in international best practices discussed in this paper. The bill takes a proactive stance by 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) emphasizing 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 adopt good practice policies and governance measures voluntarily. By embracing such policies, the agent may mitigate any potential administrative penalties.

However, it is essential to note 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 accountability mechanisms with

CC BY

the ongoing promotion of innovation, which is a difficult balance. Finally, regardless of the sophistication of the chosen policy instruments, they will only be effective if the government actively invests in building state capacity to implement them. It entails leadership, coordination, and gathering support for AI.

In summary, for AI governance to accomplish its comprehensive goals, it must recognize the intricate interplay between technology and society when setting priorities. This approach is essential for fostering responsible, trustworthy, democratic, inclusive, and human-centered AI implementation in both civil service and business domains. In essence, 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.

6. References

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, https://ieeexplore.ieee.org/document/9775553.

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).

Controladoria-Geral da União (CGU). (2022). *Relatórios de integridade e uso de dados*.

Conselho Nacional de Justiça (CNJ). (2021). *Relatório Justiça em Números / Inovação e tecnologia*.

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.

OECD. (2020). *Going Digital in Brazil, OECD Reviews of Digital Transformation*, Organisation for Economic Co-operation and Development. (2020). *Going digital in Brazil (OECD Reviews of Digital Transformation)*. OECD Publishing. https://doi.org/10.1787/e9bf7f8a-en.

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. (2023). *Global Trends in Government Innovation 2023*. OECD Publishing.

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.

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.

Tribunal de Contas da União (TCU). (2022). *Relatórios de auditoria e transformação digital*.

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems, 28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003.

World Bank. (2022). *GovTech Maturity Index 2022*.

CC BY

Pedro Luiz Costa Cavalcante
cavalcante.pedro@gmail.com
ORCID: http://orcid.org/0000-0001-7635-695X
Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP)

Doutor em Ciência Política pela UnB, com pós-doutorado na Universidade da Califórnia e na Columbia University. Atuou como Professor Visitante na School of Global Policy and Strategy da Universidade da Califórnia e como Visiting Fellow na Oxford University. É professor nos programas de Mestrado e Doutorado em Administração Pública da Enap e do IDP. Organizou livros e publicou dezenas de artigos científicos sobre governança, gestão pública, inovação, desenvolvimento e coordenação de políticas públicas. Gestor Governamental desde 2004, exerceu funções de direção e assessoramento em diversos órgãos do Governo Federal. Atualmente, atua como Secretário Adjunto de Coordenação e Governança de Estatais no Ministério da Gestão e da Inovação em Serviços Públicos.

CC BY