Brazil
CENTRE FOR THE FOURTH INDUSTRIAL REVOLUTION

A Partner of the World Economic Forum Network
for Global Technology Governance

WORLD ECONOMIC FORUM

Unpacking AI Procurement in a Box: Insights from Implementation

WHITE PAPER

MAY 2022

Cover: Shuoshu, Getty Images – Inside: Getty Images

Contents

Foreword

Giovanni Guido Cerri
President, Innovation Council, InovaHC; Professor, Hospital das Clínicas, Brazil

Kay Firth-Butterfield
Head, Artificial Intelligence and Machine Learning, World Economic Forum

Maria Lia P. Porto Corona
State Attorney-General, São Paulo State Government, Brazil

Silvani Pereira
Chief Executive Officer, Metrô de São Paulo, Brazil

Public sector entities are increasingly aware of the benefits that artificial intelligence (AI) can bring to government activities, such as tax services, law enforcement, health and education. Nevertheless, this technology bears risks, such as bias and discrimination, which require appropriate risk mitigation strategies.

The World Economic Forum's Centre for the Fourth Industrial Revolution has created AI Procurement in a Box in collaboration with the Government of the United Kingdom, Deloitte and Splunk. This pragmatic guidebook aims to unlock public-sector adoption of AI and boost innovation through more agile procurement processes. Since its launch in June 2020, the guidelines have been successfully implemented globally, including two recent pilot projects in Brazil, with Metrô de São Paulo and Hospital das Clínicas. Both pilots have reaffirmed the importance of guidance on how to procure AI tools ethically and efficiently. They also explored in-depth additional challenges for its widespread use in the public sector.

Through this white paper, the Forum, the São Paulo State Government and the Centre for the Fourth Industrial Revolution Brazil aim to share with the international community the key insights arising from these recent use cases to expand this project framework and discuss cross-cut challenges in implementing the toolkit in countries in the Global South. We look forward to continuing to develop the public procurement work on AI with the global multistakeholder community to keep AI technologies responsible, safe and ethical.

Executive summary

Revisiting the AI Procurement in a Box toolkit with new themes to guide the widespread adoption of AI in the public sector.

In 2021, the Centre for the Fourth Industrial Revolution Brazil implemented the AI Procurement in a Box Guidelines in two unprecedented pilots, one with the Metrô de São Paulo and the other with the Hospital das Clínicas of the University of São Paulo. Both pilots reaffirmed the importance of guidance on how to procure safe and ethical artificial intelligence and machine learning (AI/ML) tools, exploring additional challenges for their widespread use in the public sector.

These pilots have brought insights into and understanding of the cross-cutting challenges in implementing the toolkit in other countries, especially in the Global South. These countries often lag in terms of AI governance frameworks, institutional and technical maturity, and skills for developing and deploying these technologies. Therefore, new topics and mechanisms are needed to make AI Procurement in a Box accessible to nations in the Global South.

- Before acquiring AI, governments must self-assess their access to information technology (IT) infrastructure and appropriate data, as well as their levels of institutional maturity and internal skillsets.

- Governments should consider a mix of policy instruments, such as algorithmic impact evaluations and certifications, to ensure the responsible use of AI within their organization and expand these policies to internally developed AI models and systems that do not go through traditional procurement processes.

- Governments should carefully consider the interactions between humans and AI models to mitigate both algorithmic bias and human-based bias.

In addition, a monitoring and evaluation framework should be adopted globally to guide the deployment of future AI Procurement in a Box pilots and allow for cross-country and regional comparisons by looking at success metrics for project implementation and the wider dissemination of best practices in the public sector.

Introduction

AI Procurement in a Box has been implemented worldwide. Nevertheless, more evidence is needed to support its market-shaping role in achieving safe and ethical artificial intelligence.

Launched in June 2020, the AI Procurement in a Box toolkit is a practical guidebook to unlock public-sector adoption of artificial intelligence and machine learning (AI/ML) tools.[1] Since its publication, various pilots and use cases have taken place in the United Kingdom, Bahrain, the United Arab Emirates and India. The toolkit remains one of the most replicated projects across the World Economic Forum's network of Centres for the Fourth Industrial Revolution. Centres such as in Israel and Japan are exploring the toolkit, and other government partners are also adapting its recommendations to local contexts and jurisdictions.

Recently, the Centre for the Fourth Industrial Revolution Brazil implemented the toolkit in two separate pilots with very different objectives. The pilot with the University of São Paulo's Hospital das Clínicas (HC) – the largest health complex in Brazil, with eight specialized institutes and two auxiliary hospitals – focused on the institution's technical and data access maturity, particularly on how to ethically use AI in an area of significant public scrutiny, such as healthcare. The other pilot, carried out by Metrô de São Paulo, one of the largest subway networks in Latin America with over 5 million passengers a day, shed light on the challenges of implementing an innovation procurement procedure for the first time, as well as the unprecedented use of an algorithm impact assessment tool to identify and mitigate possible risks in the deployment of a predictive maintenance AI tool.

TABLE 1 **AI Procurement in a Box: new pilots from Brazil**

	Metrô de São Paulo	Hospital das Clínicas
Use case description	Metrô's objective was to purchase an AI-powered predictive maintenance system for online real-time monitoring of rail tracks throughout the subway network. High-definition cameras on trains and sensors will enable the use of deep learning and automated image processing to support decisions from Metrô's engineering team.	HC envisioned integrating its more than 60 information technology (IT) systems into a cohesive apparatus that would enable it to efficiently make use of data it collects in its other fields of work, like teaching, research and innovation.
AI Procurement in a Box guidelines piloted	#1 Select a needs-driven procurement process #4 Incorporate relevant legislation and codes of practice #7 Multidisciplinary team #10 Market engagement	#5 Assess technical and administrative data accessibility #6 Assess data bias and quality at a higher level #8 Algorithmic accountability and transparency
Key challenges and accomplishments	– Adoption of a mission-oriented pre-commercial procurement procedure (*encomenda tecnológica*) to prototype and deploy a predictive maintenance algorithm. – First use of algorithmic impact assessments in Brazil, as recommended by Brazil's National AI Strategy. – Creation of an independent board of specialists to assess risks associated with the project and lower asymmetries of information between the contracting entities.	– Deep institutional and internal organizational changes, with the creation of new departments and restructuring of old ones. – Development of a new platform to integrate other systems in use and enable better standards for privacy and data anonymization. – Promotion of a culture of engagement of the people involved throughout the entire data collection, processing, knowledge-derivation and innovation development life cycle.

Source: Centre for the Fourth Industrial Revolution Brazil, 2022.

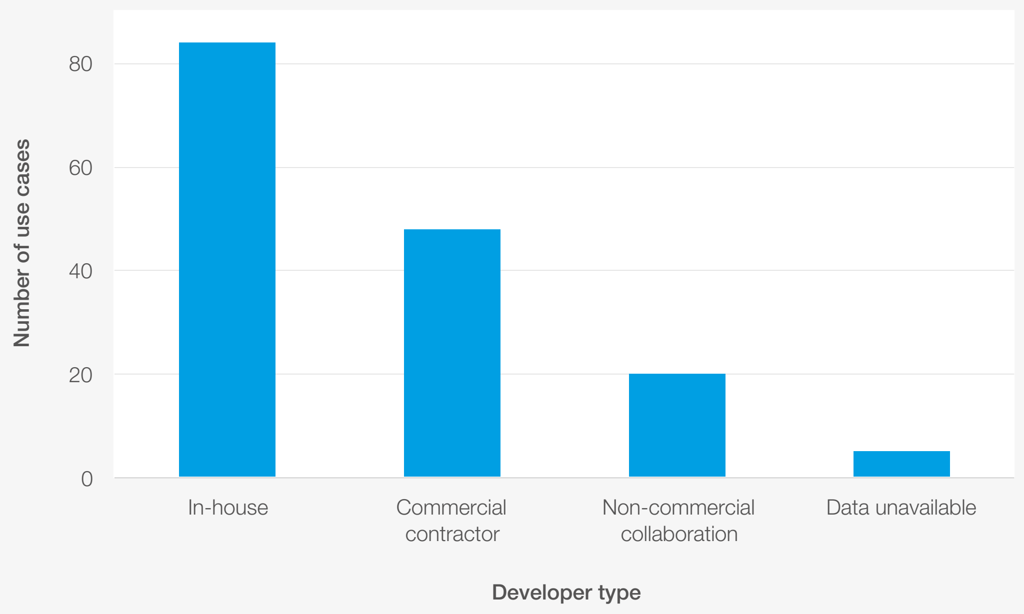
Challenges and bottlenecks faced during the implementation of these pilots in Brazil provided valuable insights into the toolkit as a whole. In addition, community meetings and surveys with local stakeholders converged on the need to provide country-specific modules to AI Procurement in a Box, adapting the Forum's guidelines to Brazilian laws and regulations. As a result, the centre in Brazil has put together a community with members from the public sector, industry, civil society and academia to co-design a pragmatic guide to procure AI solutions in Brazil,[2] along with a thorough description of the use cases at Metrô and HC.

Public procurement or in-house developments?

A report issued in 2017 shows that 81% of Organisation for Economic Co-operation and Development (OECD) countries have implemented policies and strategies to support innovative goods and services through public procurement.[3] There are countless examples of instruments like pre-commercial procurement (PCP) and public procurement of innovative solutions (PPI) being used for the development, scaling and diffusion of disruptive technologies in the market. However, and notwithstanding several PCP and PPI projects in the field of artificial intelligence (AI),[4] more evidence is needed to support the market-shaping role of public procurement in fostering ethical and human-centred AI in the IT market.

It is interesting to note that a large part of AI systems currently in use by governments worldwide cannot have their origins traced back to procurement. In the United States, for example, a report analysed the use of 142 AI tools by federal departments, agencies and sub-agencies and found that more than half (84 tools, or 53%) were actually in-house developments.[5]

FIGURE 1 AI use cases in the US federal government by type of developer

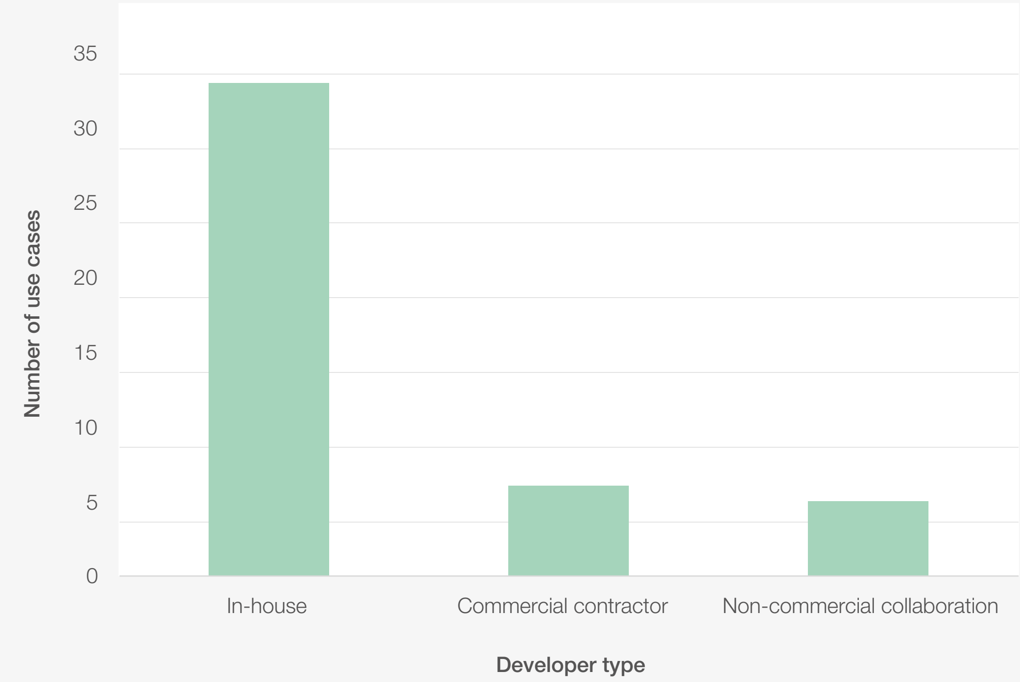


Source: Engstrom, D. F. et al., 2020.[6]

However, the technical demands imposed by this make-or-buy choice do not mean that governments have already reached adequate levels of internal capacity and staff expertise to design, develop and audit AI solutions. On the contrary, pressured by recruitment challenges and growing demand for IT professionals, public entities often struggle with the lack of personnel with proper training and adequate skill sets. To fill this gap, the literature recommends internal capacity building programmes and the formation of centres of expertise[7] to provide shared technical knowledge and a systemic approach to the use of technology in government activities.[8]

A similar scenario is found in Brazil. Information gathered by Transparência Brasil[9] reveals that among 44 AI tools in use by entities in the public sector, at least 33 were built internally; almost half of the remaining (5 out of 11) are the result of non-commercial partnerships with universities and research institutions.

FIGURE 2 | AI use cases in the Brazilian public sector by type of developer



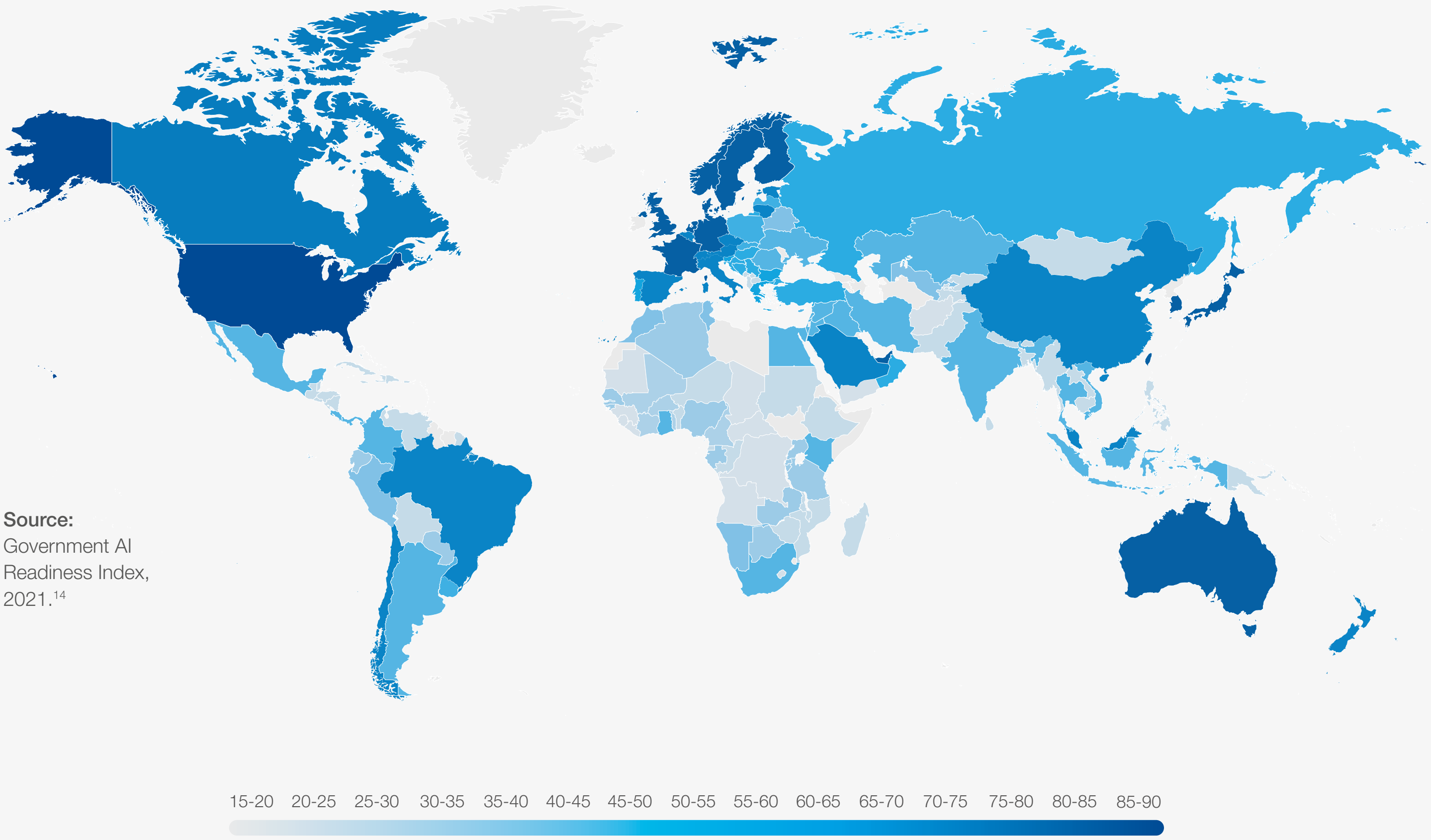
Source: Transparência Brasil, 2020,[10] adapted to match the categories adopted by Engstrom et al., 2020.[11]

The examples illustrating the widespread use of in-house developments in countries with very different levels of digital maturity and AI readiness, as shown by the United States and Brazil, point out the need for further research to investigate the causes of relatively low adoption of procurement among other sourcing mechanisms, as well as to collect more evidence to assess how government contracts might shape the AI industry. Importantly, in-house AI solutions may also be developed in contexts that may have fewer transparency and accountability structures than those acquired from procurement processes, so it is important that governance mechanisms be deployed for all uses of AI within the public sector, not just those sourced from third parties.

AI procurement in the Global South

The literature on AI governance in the Global South – a term used to highlight the uneven distribution of resources and knowledge to implement AI between high, medium and low-income countries – is growing.[12] Some AI-powered technologies, such as natural language processing, are examples of the disparate representation of the Global South in AI governance and ethics,[13] even though these countries face their own challenges in adopting AI, particularly in the public sector.

FIGURE 3 **Government readiness for AI adoption in selected countries**



Source: Government AI Readiness Index, 2021.[14]

The global picture is one of highly unequal AI readiness that negatively affects the widespread implementation of the Forum's toolkit. For example, the regional score of the three lowest-ranked regions in the Government AI Readiness Index of 2021 – sub-Saharan Africa (31.61), Central & South Asia (40.93) and Latin America and the Caribbean (41.26) – is less than half the average of the highest-ranked region, North America, at 82.94. In addition, these countries often face difficulties in developing their own AI technologies to supplement those ecosystems due to brain drain – the exodus of skilled workers to developed nations.

Key challenges include (i) **skills** to supplement the deployment of AI technologies, underscoring the diversity gap among teams usually involved in these activities; (ii) **institutional capacities and regulation** due to the lack of regulatory and governance frameworks mature enough to both support innovation and mitigate ethical risks; and (iii) **infrastructure** due to poor data governance protocols and the scarcity of datasets to train algorithms developed in the Global South, leading to a data divide in an increasingly data-driven economy.[15]

What is the purpose of this document?

This white paper aims to support the implementation of AI Procurement in a Box in a wide array of jurisdictions, incorporating new elements arising from the literature and practice that should be added to update the Forum's overall framework. As new case studies on the use of AI in the public sector are rolled out, new opportunities arise to expand the set of best practices put together in the AI Procurement in a Box project, especially use cases and challenges faced by the Global South.

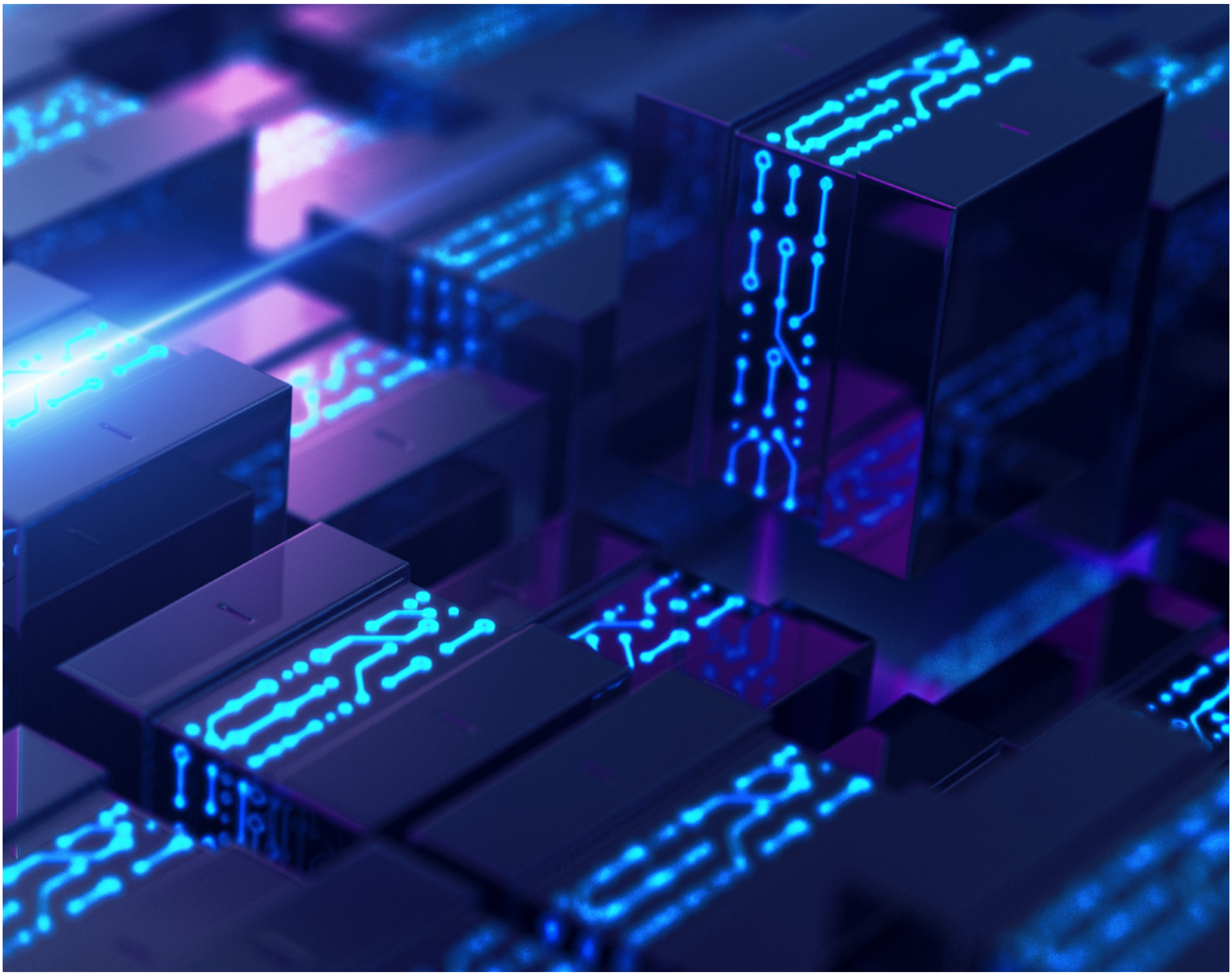
This white paper is divided into one introduction and four content sections:

– **Phase Zero** is an exploration of what maturity for AI adoption means, in terms of IT literature and insufficient or poor data quality, as well as in terms of institutional changes to prepare for the deployment of these technologies and workforce capacity building.

– **Leveraging public trust in AI through accountability structures** looks at how governments use a mix of policy mechanisms, such as algorithmic impact assessments and certifications, to ensure the responsible use of AI within their projects.

– **Human beyond-the-loop? Mitigating risks arising from human oversight** brings understanding that, although human-in-the-loop is a common best practice for AI oversight, it can also bring new risks that must be mitigated accordingly.

– **What does success look like for AI Procurement in a Box? Indicators for monitoring and evaluation** discusses how to monitor and evaluate an AI Procurement in a Box project, seeking to understand how this project can bring value to pilot institutions and impact national AI governance regimes. Evaluation is divided into six pillars based on what successful implementation should look like, each with its own set of indicators, highlighting the kinds of goals implementing partners should have in mind while simultaneously enabling pilot comparisons across different contexts and jurisdictions.



1

Phase zero: Prerequisites for the widespread adoption of AI/ML in the public sector

Understanding infrastructural, institutional and data access maturity aspects to implement successful AI projects in government

A primary concern for government adoption of AI/ML technologies is the institutional and technical maturity needed for its adoption. A recurring question is about the "phase zero" – or the minimum digital maturity needed to enable AI adoption. This topic is briefly discussed in AI Procurement in a Box (guideline #5), particularly in the lens of communicating specific technical limitations of the procuring organization to suppliers.

Most organizations – public agencies included – have a basic understanding of their data assets, such as what kind of data they have and the infrastructure used to organize it. But they are unable to harness it for the development of AI solutions due to a lack of understanding of the potential they entail.[16] In addition, many organizations do not have data governance processes in place, such as the existence of chief data officers or enterprise data champions, jeopardizing data privacy and integrity.

Guideline #5 starts a meaningful conversation but digital maturity for AI in the public sector still needs to be explored in depth. Disparate access to data and the related IT infrastructure can result in uneven adoption between different levels of government. In Brazil, for example, conversations held with local stakeholders reveal that municipal government entities are much less likely to adopt AI/ML technologies than those at the federal or state level, mainly due to technical maturity barriers and readiness.

The consequences of disparate infrastructure and data access can be severe, depending on the context. For example, AI can bring many opportunities to facilitate access to critical government services but many countries lack the institutional maturity to realize that. According to Smith and Neupane,[17] there is an existing AI gap "between those who have the ability to design and deploy AI applications and those who do not", creating large power asymmetries. Countries in the Global South often import AI technologies and therefore should retrain foreign commercial off-the-shelf algorithms to keep accuracy levels and prevent distortions. However, public entities frequently lack the data and technical capabilities needed to readjust AI models.

Low technical maturity and data access can also be detrimental to suppliers. If governments cannot get the data and other necessary inputs to develop, train and test an AI/ML model, they are more likely to procure from established market incumbents that already have access to data, creating an uneven playing field.

1.1 What constitutes AI maturity for public sector organizations?

Maturity for AI adoption encompasses many themes, from data access and IT infrastructure to institutional capacity building and governance. Because of the variety of elements, it is nearly impossible to prescribe a one-size-fits-all approach, so solutions should be designed on a case-by-case basis.

Many assessments look at digital maturity as a whole. For example, the World Bank's GovTech Maturity Index evaluates four different components to understand a country's level of digital transformation.[18] Likewise, the Broadband Commission's Working Group on Digital and AI in Health, chaired by the Novartis Foundation, published a report that addresses the maturity challenges that public healthcare sectors face globally,[19] including a self-assessment tool with recommendations on how to allocate resources in these areas.

Table 2 presents an outline of the prerequisites that public sector entities should cover before implementing artificial intelligence (AI).

TABLE 2 | Phase zero: prerequisites for widespread adoption of AI/ML in the public sector

IT infrastructure	What kind of IT infrastructure did the entity acquire in preparation for digitalization?	– Discuss the need for an IT rollout covering hardware, software, storage and architecture of datasets the public entity needs. – Open data initiatives and federated learning systems can be alternatives to enable access to the necessary data.
People	How can public entities hire (and retain) highly skilled IT employees in their workforces?	– Highly skilled workers are already scarce in the IT industry and hiring them can be challenging for most governments. Such staff is a condition for being able to take what AI can offer, including desired skills and job classifications. – Disseminate AI knowledge by investing resources in training and internal capacity building. – Foster a culture of open innovation through partnerships with private sector organizations, civil society and academia.
Institutional changes	How did the government entity restructure itself for digitalization?	– Organizational and institutional changes often go side by side with the integration of legacy IT systems, especially when specific departments of an agency have developed such tools to fit only their own needs. – Develop a data governance culture and draft an agenda for its continuous implementation. – Establish a governance process for procurement, development, implementation and monitoring of AI-enabled tools.
Insufficient or poor-quality data	How should the buyer deal with low-quality data?	– Knowledge of existing data limitations from previous efforts should incentivize public agencies to organize datasets to adopt AI in the future. – Establish clear data inputs for AI, such as creating a minimal set of key performance indicators (KPIs) that need to be collected and reported, harmonizing reporting mechanisms and establishing a data-centric institutional culture. – Discuss the trade-off between re-organizing a disorganized database (keeping relevant past information) and establishing a cutline to recollect data in an automated and organized manner. – Lack of metadata standardization and poor data quality may jeopardize the purchase of AI. Data quality should consider representativeness, provenance and bias.

Private companies can be more agile in digital maturity than public entities. This is due to the obligation to undergo formal procurement and also because public officials are risk-averse and lack incentives for digital transformation.[20] One of the main recommendations for governments is to consolidate their data sources into a data centre, as shown in figure 4.

FIGURE 4

Five categories for public data centre costs

	Share of total data centre cost, %	Description of cost categories	Action for value capture
Labour	50-75	– Performers of core infrastructure maintenance activities (e.g. provisioning of new environments, incident management, change management)	– Reducing the number of roles – Repurposing resources for other functions
Software	12-25	– Software and maintenance contracts	– Reducing the number of licenses – Shifting to open source – Leveraging enterprise license agreements
Hardware	5-12	– Server- and storage -refresh costs – Maintenance cost of hardware servicing	– Rightsizing the computer, memory and storage needs – Timing the refresh cycle with migration to cloud
Facilities	5-10	– Rent – Power – Building maintenance	– Optimizing the lease – Repurposing the space or equipment
Infrastructure	5-10	– Equipment – Network – Middleware	– Repurposing the space or equipment – Optimizing any contracts or service agreements

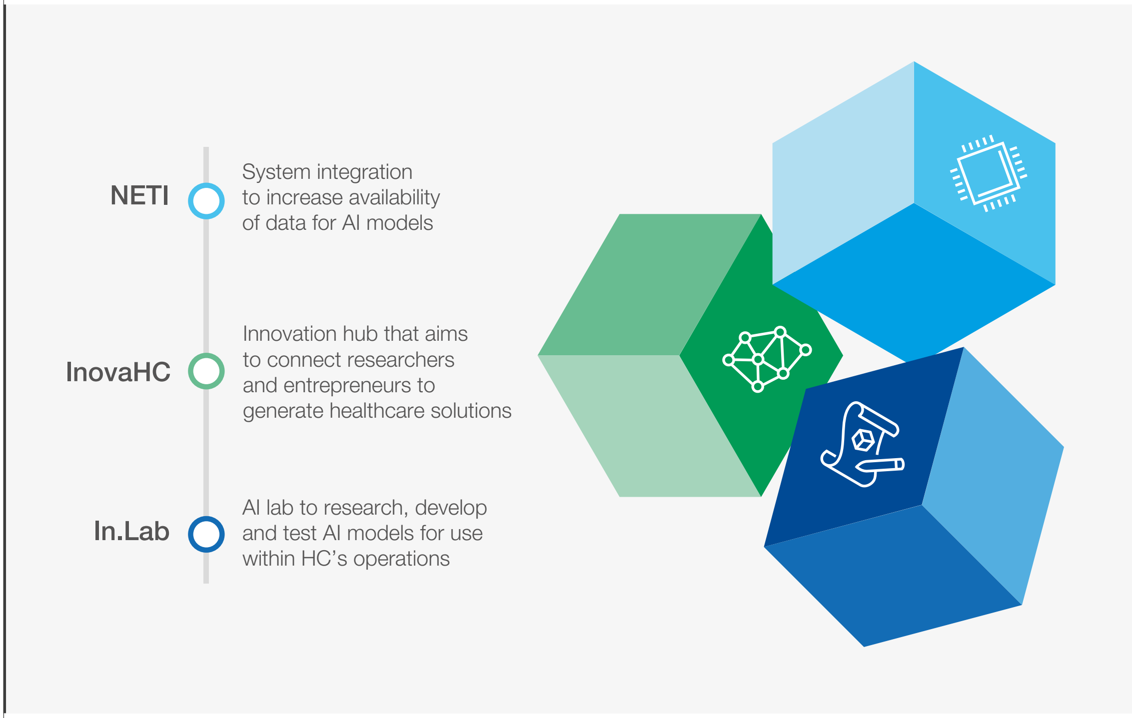
Source: McKinsey and Company, 2019.[21]

1.2 Moving towards digital maturity

Each institution has its roadmap for reaching a level of maturity where AI adoption is feasible. By sharing key lessons from experiences, other organizations can better understand the roadblocks in their unique pathway.

Documenting the journey to institutional maturity was the intention behind the Centre for the Fourth Industrial Revolution Brazil's pilot with HC. To get to a level where all the different entities that compose the hospital were able to integrate AI systems within operations, HC worked to integrate more than 60 of its IT systems and protocols into a cohesive apparatus that would enable the use of AI in healthcare and research. To achieve this goal, HC focused on integrating data sources into a data lake, providing technical training, and optimizing data management between different areas. Several internal institutional changes came in the form of the creation and integration of new departments, as shown in figure 5.

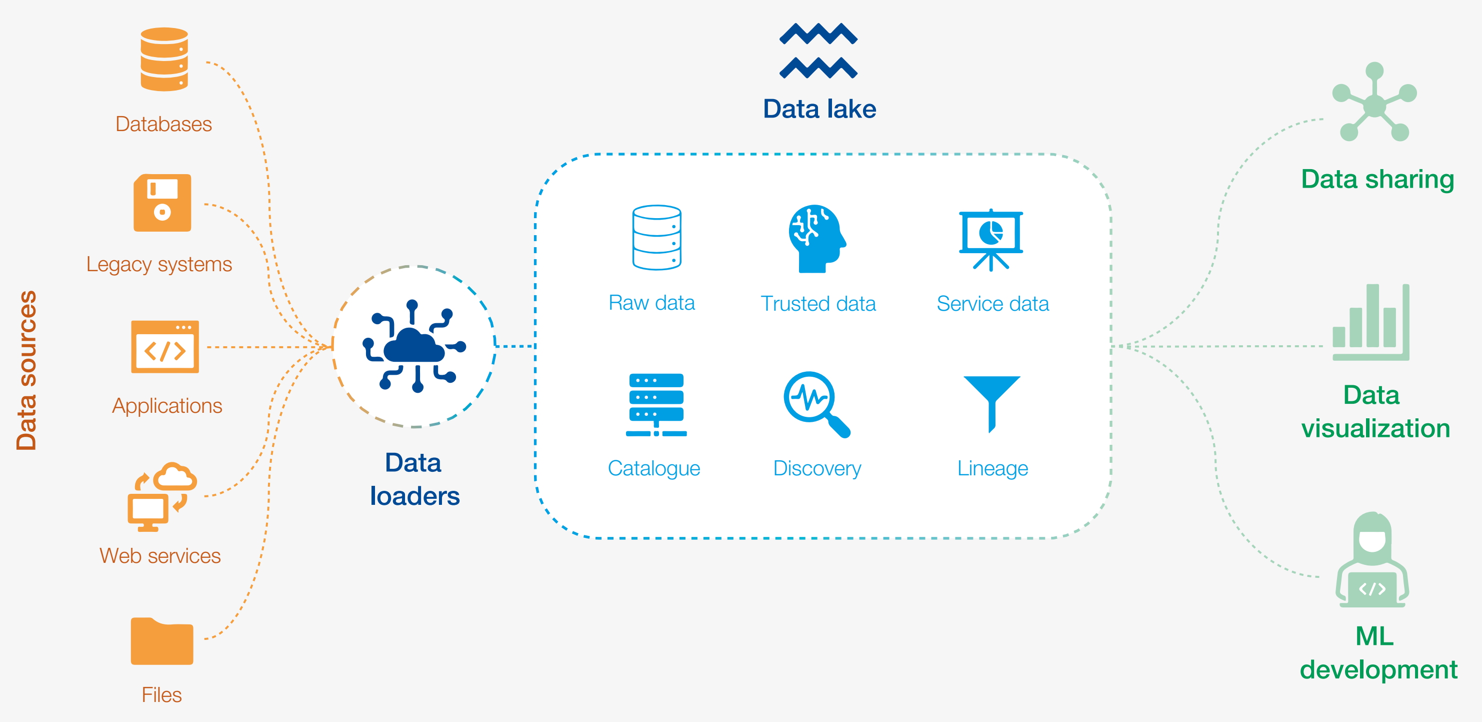
FIGURE 5 | Institutional changes in the centre's HC pilot



The development of these institutions was fundamental in achieving HC's objective. The data lake is still being built but it has already helped train successful projects, such as the Rad-vid AI model, which supports the diagnosis of COVID-19 through the analysis of lung x-rays.[22]

In addition to this reorganization of both the institution and its datasets, the hospital is also creating protocols to anonymize health data and evaluate internally developed algorithms and those procured from third-party vendors for ethical and bioethical considerations. HC is also working on improving its data culture, with a better understanding of quality documentation and correct data input, which is essential to the ability to move to the application of any complex data analytics. Although not indistinctly applicable for all entities, HC's experience shows an organization's digital maturity journey and the wide set of activities needed to get there.

FIGURE 6 | Data lake system as envisioned by HC



Source: HC.

2 Leveraging public trust in AI through accountability structures

Public trust in AI is essential to its development and use in the public sector. What steps can be taken to ensure it?

Government use of AI/ML is already important in deploying better public services. However, misused applications and documented cases of algorithmic bias may erode the public's buy-in of AI. For example, a survey carried out by the University of Oxford found that 8 in 10 Americans "agree that AI and/or robots should be carefully managed."[23] In order to instil public trust, a regulatory ecosystem that guarantees the trustworthiness of AI is needed.[24] However, how can policy-makers create AI regulations that mitigate risks while still incentivizing innovation?

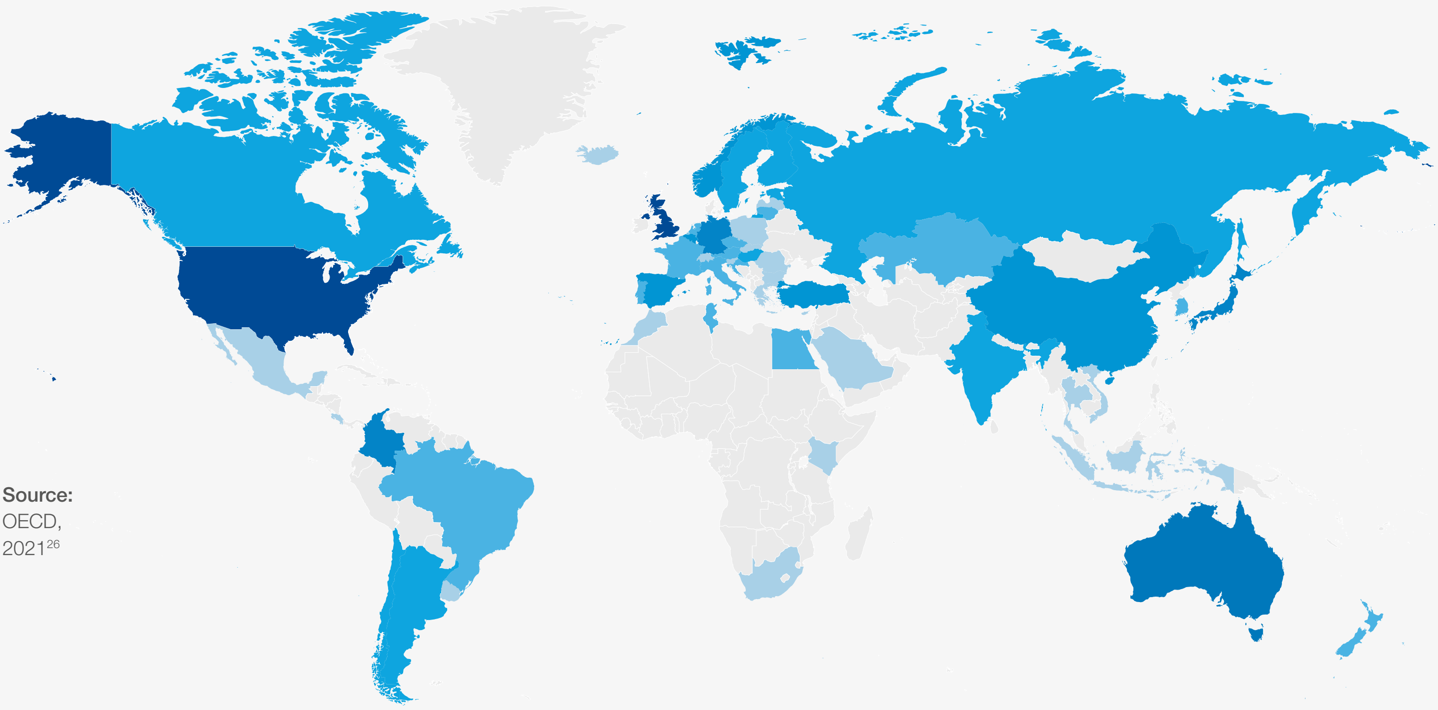
One of the main objectives of the AI Procurement in a Box project is to embed accountability and AI ethics into public procurement processes. On average, OECD member countries spend about 12% of their GDP on procuring goods and services.[25] By leveraging those investments, governments can proactively influence the development of ethical AI in the market, shaping technology governance in ways that the industry cannot.

Regarding procurement, guideline #1 advocates using innovative procurement processes to purchase AI, mainly because it focuses on outlining problems that the procurer should frame in a challenge-based, mission-oriented procedure instead of on prescribing detailed specifications for the solution. However, in most jurisdictions, PPI and PCP instruments are seen as exceptional in a system where regular procurement is the rule. Thus, public officials often must justify its use and explain why traditional procurement is unfit for the purchase of each object, aggravating the problem of risk aversion.

Self-assessment tools, such as algorithmic impact assessments, certifications and grants are examples of policy instruments with complementary approaches to incentivize supply and demand for ethical AI. Diversity and the availability of different innovation policy instruments are crucial to nations in the Global South, especially within their national AI strategies. Figure 7 shows how widespread national AI strategies have become, highlighting the marked contrast between nations:

FIGURE 7

National strategies and policies for AI in the public sector
(by number of initiatives)



Source:
OECD, 2021[26]

This section builds upon the AI Procurement in a Box guidelines to look at additional ways public sector organizations can incorporate different policy instruments and mechanisms to ensure trustworthy and ethical AI projects.

2.1 Revisiting the role of risk evaluations in public procurement

One of the most significant recommendations is using an algorithmic impact assessment (AIA), such as that used by the Canadian Directive on Automated Decision Making,[27] to analyse the potential risks and negative impacts of each AI-related project, creating strategies to mitigate these risks.

In 2022, the European Law Institute published its *Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration*.[28] As in the Canadian AIA, the suggested rules categorize automated decision-making systems (ADMS) in tiered categories, with those at the highest risk needing more oversight. In addition, the proposal aims to raise awareness of the risks of using ADMS in public administration and increase participation, transparency and accountability mechanisms in AIAs.

AIAs can be a valuable tool in leveraging participation and creating public trust because "the public has a meaningful opportunity to respond to and, if necessary, dispute the use of a given system or an agency's approach to algorithmic accountability."[29] Many countries' procurement regulations already require the publication of general, multi-purpose risk assessments in requests for proposals (RFPs). In this spirit, the Centre for the Fourth Industrial Revolution Brazil built an AIA template within the pre-existing risk assessment structures in the Metrô use case, integrating the Forum's best practices into existing bureaucracy.

BOX 1 Metrô de São Paulo and Brazil´s first Algorithmic Impact Assessment

Metrô de São Paulo, a state-owned enterprise, was the Centre for the Fourth Industrial Revolution Brazil's first pilot. Metrô's engineering team wanted to purchase an AI-powered predictive maintenance system for online, real-time monitoring of rail tracks throughout the subway network. The occurrence of failures generates a long interference time, caused mainly by the need for maintenance teams to enter the track. In the last four years, there was a total of 1,840 minutes of interference on the tracks, jeopardizing the mobility of approximately 460,000 passengers.

The proposed *Sistema de Monitoramento de Via Permanente*, an AI-powered predictive maintenance system, would allow real-time online monitoring of rail tracks. This system would support maintenance-related decisions by using high-definition cameras on trains and sensors to enable deep learning and automated image processing.

To find the ideal solution, Metrô adopted a pre-commercial procurement procedure (*encomenda tecnológica*) which allows the development of a new service, product or innovative process when technological risk is involved. Following the recommendation of the Federal Court of Auditors (Tribunal de Contas da União) and other controlling entities in Brazil, the *encomenda* contract is awarded after the publication of an RFP, which includes a risk matrix to evaluate risks that may compromise the project's success.

To fit Metrô's needs, the Centre for the Fourth Industrial Revolution Brazil developed an algorithmic impact assessment model adapted to Brazilian laws and regulations. The model was based on the Canadian AIA, given its completeness and excellent detail, and was able to incorporate into the RFP the main risks associated with the use of AI. Metrô has published both documents on its website.

Source: Centre for the Fourth Industrial Revolution Brazil, 2022.[30]

The widespread use of in-house developments in the public sector may become a limitation to AI Procurement in a Box. If in-house developments are not subject to the same accountability structures as government procurement, this sourcing choice may cause a lack of transparency on how the public sector uses AI/ML tools. The less outside stakeholders know about each project, the harder it will be to ascertain if the use case is fair and legal, impacting accountability.[31] Thus, it is vital that the same transparency and governance structures used in procurement, like AIAs, are also conducted for in-house developments. Therefore, governments should expand the use of AIAs to encompass all AI/ML projects in the public sector, not limiting them only to government acquisitions.

This point also raises questions for future research, such as if the government should create entities or governance structures to examine these risk assessments and oversee algorithmic accountability holistically and systematically, regardless of how automated decision-making systems are implemented.

2.2 Moving towards a more cohesive AI policy mix

AI governance and regulation are already complex when discussed in the public sector. When that debate reaches stakeholders like third-party vendors, the complexity grows. This is due to sophisticated algorithms and because the use of interconnected data may decrease transparency, amplifying the risk of interaction bias and other distortions. These risks are not unique to the public sector but the use of AI in government activities usually influences decisions that can highly impact human rights and communities.

There are several ways to regulate AI. Table 3 presents an outline of different accountability strategies, types of rules, enforcement and timing of proposed regulations.

TABLE 3 **Key choices and options in AI/ML regulatory design**

Accountability	– Legal accountability (judicial review or agency action) – Political accountability (notice-and-comment procedures or impact assessments)
Types of rules	– Hard rules (prohibition on types of models, data uses or requirement to prior licensing requirements or agency approvals) – Soft rules (impact assessments, bundles of notice, consent, correction and erasure rights)
Enforcement	– Public enforcers, whether public prosecutors or an administrative agency – Private enforcers (rights of action to sue in courts or whistleblowing bounty schemes)
Timing	– Ex-ante regulation, before a model runs (licensing scheme or prohibitions on uses or model types) – Ex-post regulation of results, as with lawsuits seeking damages

Source: Adapted from Engstrom and Ho, 2020.[32]

One proposal that has permeated the field is the use of AI certifications. Examples include the Institute of Electrical and Electronics Engineers (IEEE) Standards Association's Ethics Certification Program, with indicators related to transparency, accountability and algorithmic bias.[33] Another system is the Responsible AI Institute's certification programme developed in partnership with the World Economic Forum's Global AI Action Alliance.[34] It certifies AI/ML systems based on five dimensions: accountability, bias and fairness, data quality, explainability and interpretability, and robustness.

Certifications can reduce information asymmetries between buyers and vendors, addressing problems of selection (identifying desirable suppliers) and monitoring (evaluating whether the developer met specifications).[35] However, in some jurisdictions, the use of certifications in procurement processes depends on legal provisions authorizing its use by contracting entities, especially when it could affect competition by creating pre-qualified vendor lists. If the certification is not made available to many suppliers, such as start-ups and small and medium-sized enterprises (SMEs), much care should be taken to ensure market fairness. In addition, there is currently no unified certification programme for ethical AI, and those that exist differ in their variables for assessment. There is no systematized approach for when and how often developers should certify their AI tools – if it is a one-off procedure or if it should be certified periodically as the AI model changes and evolves.

Until then, public entities should keep integrating risk assessments and other governance mechanisms into their procurement processes. However, in high-risk use cases, AIAs and other self-assessments might not be enough to mitigate all the risks involved, raising the case for using certifications in the subsequent procurement stage. Most regulations do not require those impacted to be notified that an AI system was applied to them, much less offer redress and remediation when harm occurs. Thus, self-assessments and certifications should be seen as complementary instruments to be combined by government entities whenever AI/ML tools are used in high-scrutiny decision-making.

As more governments figure out how to govern the use of artificial intelligence within their agencies, it is increasingly evident that mechanisms for accountability and oversight cannot exist in a vacuum or be implemented in a disjointed way. The method of oversight, be it hard law or soft law, AIAs or certifications, must be implemented systematically and uniformly, ensuring that the subjects involved know what is being evaluated, how often and the risks involved in each tool.

3

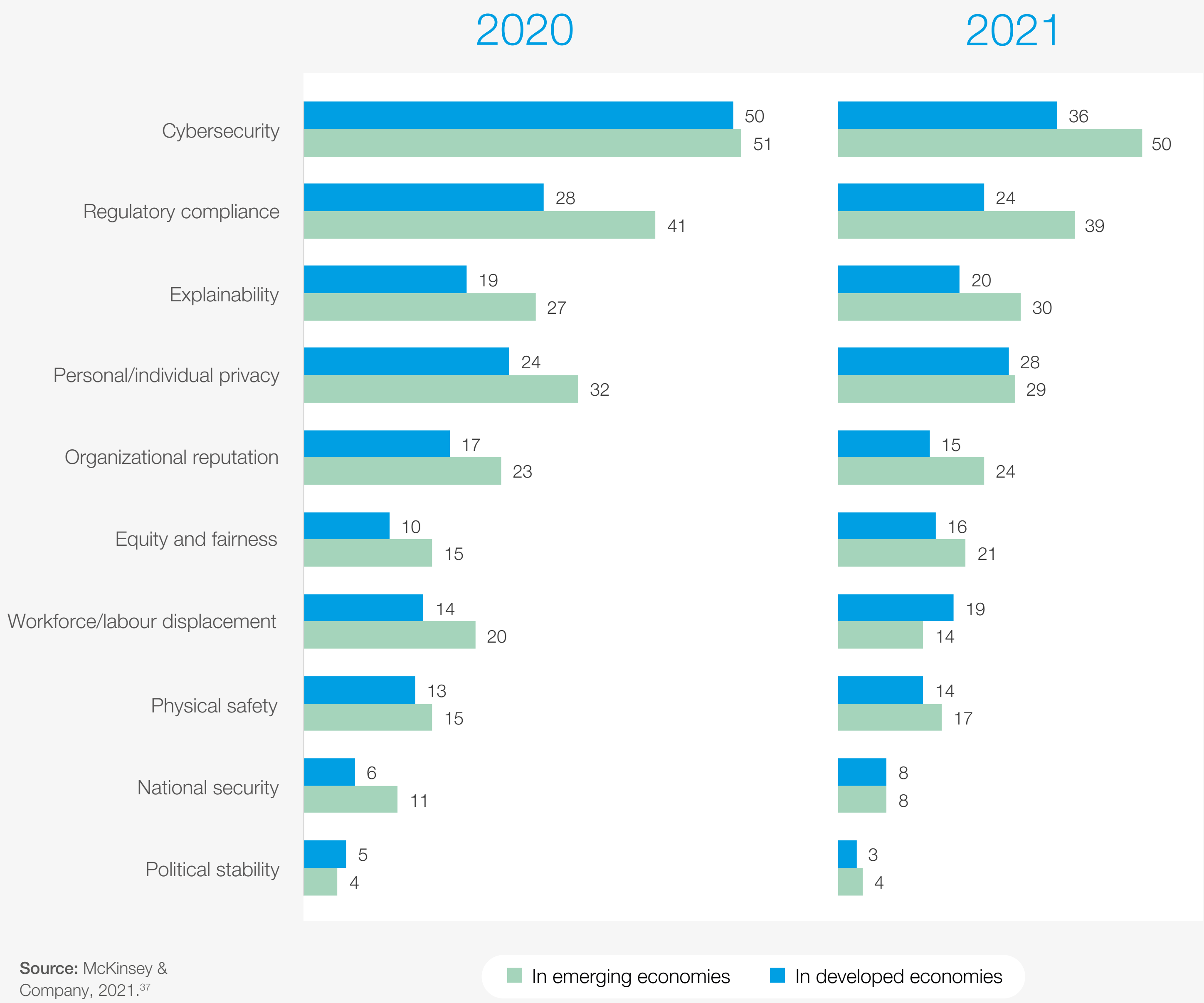
Human beyond-the-loop? Mitigating risks arising from human oversight

Human oversight is needed to reduce bias and hazards in AI/ML tools but this approach embeds risks that public sector organizations must mitigate accordingly.

One of the few near-consensus in AI policy is that algorithms – if left unchecked – can lead to decisions that may have a disparate, harmful impact on specific groups of people. However, the rapidly increasing adoption of AI/ML tools by public and private sector organizations worldwide is not being matched by similar efforts to oversee and mitigate risks like errors and unintended negative consequences. Although focused mainly on private sector companies, McKinsey's 2021 global AI survey[36] shows that respondents from developed nations and emerging economies have different perceptions of AI risks and willingness to adopt mitigation measures. The survey also highlights that respondents often cannot address the full range of risks they face and choose which ones should be prioritized, especially in the absence of regulations about AI risk mitigation.

FIGURE 8

AI risks that organizations are working to mitigate
(% of respondents by office headquarters)



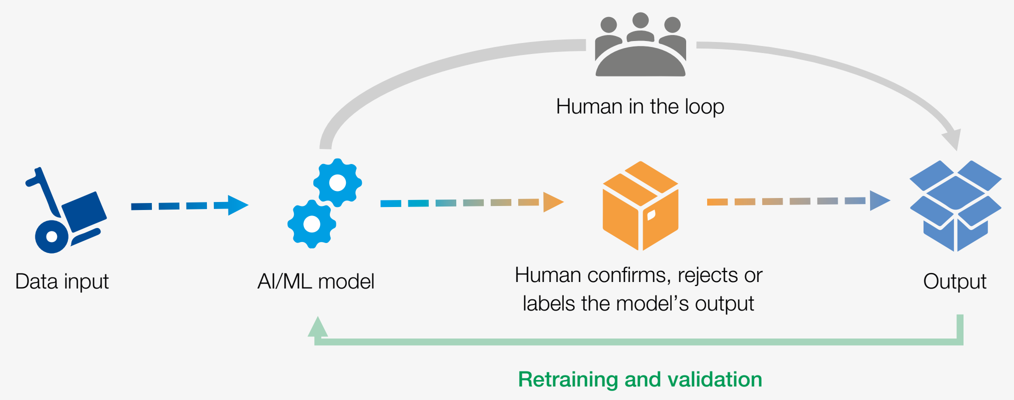
Source: McKinsey & Company, 2021.[37]

One of the most difficult aspects of risk mitigation is that it is difficult to define metrics for fairness, explainability, adversarial robustness, and distribution shift in AI systems.[38] These elements usually rely on judgement calls from humans, who may not be adequately prepared to make such decisions. In addition, there is no consensus about the various explainability requirements that adopters should follow, making any pathway highly unclear.[39]

3.1 Risks and mitigation strategies in human-machine interactions

Risk mitigation and oversight remain urgent and unresolved tasks in the public sector, where AI/ML tools are used in areas of high public scrutiny.[40] The broad expression "human-in-the-loop" here means that human agents should assess, review and remain accountable for the algorithm's outputs, ensuring compliance with legal obligations and alignment with user expectations. The European Commission also adopts oversight in its proposed AI regulations,[41] emphasizing the role of a human-machine interface to guarantee that high-risk AI systems remain responsive in the long run.

FIGURE 9 **Human-in-the-loop as a risk mitigation strategy in AI/ML tools**



However, bringing human operators into the loop raises its own set of concerns.[42] For example, individuals to whom human oversight is assigned may defer to the tendency of automatically relying on the outputs of the algorithm or willingly try to compensate its results based on perceived errors in the algorithmic scores,[43] as shown in table 4.

TABLE 4 **Risks arising from human oversight in AI/ML tools**

	Description	Mitigation measures
Automation bias	Human operators may become overconfident and excessively reliant on algorithmic outputs, whether because of insufficient attention, time or resource constraints, or lack of understanding about how the AI/ML tool works, deferring to machine outputs over time.	– Provide guidance and specific training to human-in-the-loop staff – even ML experts can misinterpret how a model is making decisions. – Conduct continuous monitoring and evaluation of human reviewer performance against patterns and repeated standards. – Implement reverse Turing tests to monitor if human-made and algorithmic decisions remain indistinguishable. – Conduct routine audits both internally and externally. – Ensure that a team of diverse, cross-functional and multidisciplinary professionals carry out supervision. – Provide adequate resources to the oversight function. Align incentives properly to keep staff from rubber-stamping results. – Document good processes, iterate and share best practices.
Compensations	Humans may overcompensate or under-compensate errors in algorithmic scores of an AI/ML tool. These interventions may not be fair or legal and, even when aiming to correct bias, can create distortions that affect the model by influencing outputs that cannot be traced back to the system's databases.	

Source: Mulligan e Bamberger, 2019[44] and Rubenstein, 2021.[45]

This scenario calls for a comprehensive set of mitigation strategies. In addition to internal capacity building and specific training, it is necessary to ensure that agency staff functioning as human-in-the-loop are given the information, time and incentives to be true safety guardrails. In addition, the literature supports the use of a reverse Turing test,[46] requiring that public sector entities set aside a random test to sample analogue, human-made decisions and then compare the results to those achieved using algorithms. Provided that analogue and automated decisions are indistinguishable from human reviewers, the whole merged set would be used to update the algorithm, leveraging human-machine collaboration. This approach can be incorporated into government acquisitions through contract terms and procurement requirements and may provide valuable exogenous data to update and train AI/ML models.

3.2 A shift towards contestable design

Inherent risks brought up by AI/ML tools of growing complexity call for an expansion of human participation throughout the system life cycle, not only review and oversight. Placing humans at the end of the decision-making process seems insufficient to ensure fairness and accountability in AI solutions, which should be trustworthy in all stages from design to execution,[47] as recently restated in the US Department of Defense Ethical Principles for AI.[48] Moreover, participant-centred development of algorithms can be key factors in mitigating risks from human oversight.[49]

Public sector AI systems must be designed to promote **contestability**. The literature has argued that contestability is a more active and dynamic notion than explainability,[50] which typically looks at static, retroactive clarification of the decision model and its effective operation. In contrast, to be contestable, a system should also trigger human engagement and foster users' understanding of models and outputs, allowing dynamic feedback for continuous and collaborative construction.

Table 5 exemplifies general human attributions within the life cycle of AI solutions.

TABLE 5 | **Beyond the loop: human participation throughout the life cycle of AI/ML tools**

Stage	World Economic Forum Guideline	Human participation
Input data	#5	Human behaviours, practices and preferences generate data that fuels AI/ML systems to generate predictions, insights and sometimes decisions.
Fitness of purpose/ risk evaluations	#2	Humans undertake a landscape review of the project's context and components to determine whether AI is the best tool to address the problem, as well as the level of risk in each project.
Design of the algorithm and ML decision model	#5, #6	Humans determine if data is sufficient, accurate, appropriate, available, adequately obtained, administered and secured. If so, the algorithm is deployed and has its readiness and proficiency checked.
AI training and supervision	#5, #6	Humans select datasets for training, clean up unsuitable data, label and categorize to improve accuracy and evaluate possible bias concerns.
Assessment of outputs	#6, #7, #8	Humans input, label, select, review, assess and interpret AI functions, creating feedback data to increase accuracy and effectiveness. They also approve the use in real-world settings, adopting strategies for mitigating foreseeable risks and unintended outcomes.
User experience, rights and expectations	#4	Humans manage and are accountable for users' rights and expectations and set the standards for compliance with relevant laws, regulations, internal policies and best practices.
Review, monitoring and oversight	#7, #8	Ongoing review is necessary to ensure AI tools adhere to their original purposes. Reviewers evaluate the appropriateness over time and for different populations and contexts, creating more feedback data.

Source: Adapted from World Economic Forum, 2020[51] and Silverman, 2021.[52]

Therefore, human-in-the-loop is not a failsafe mechanism. The US National Security Commission on Artificial Intelligence[53] has issued recommendations to leverage human-machine interactions that span the entire AI solution life cycle, such as clearly defining functions and responsibilities of individual human operators, defining feedback loops and oversight opportunities through the AI life cycle, auditing the human-AI pair, clarifying interpretability and explainability requirements, and leveraging traceability.

Internalizing these concerns into procurement processes is key because, as seen above, human oversight is not always able to neutralize the risks of using AI in public sector decision-making. Moreover, it is vital to have flexible contract terms to allow alterations and use contract-based mechanisms to create governance structures (boards, committees) to include user perspectives over time.

4 What does success look like for AI Procurement in a Box? Indicators for monitoring and evaluation

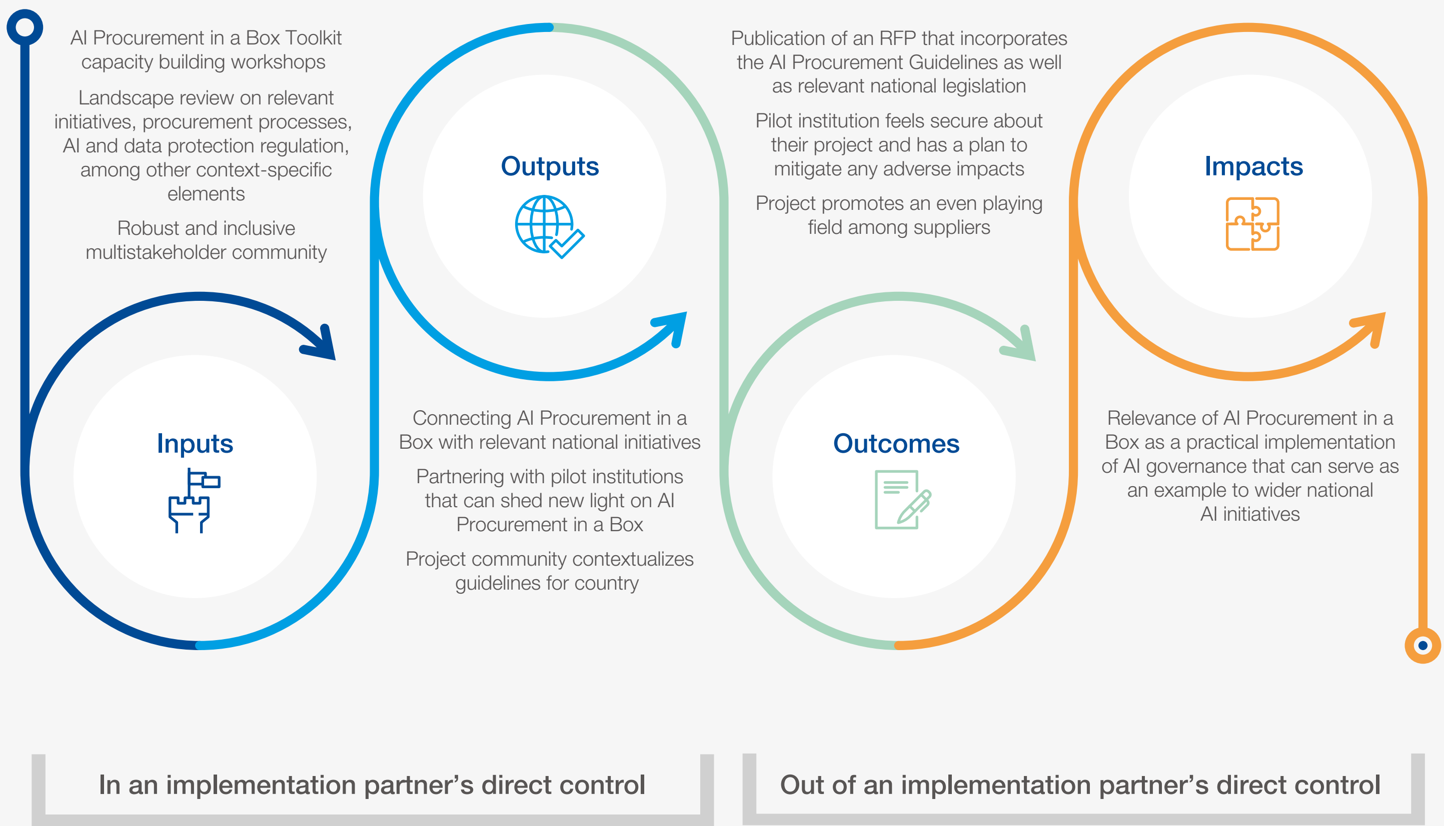
How to monitor the implementation of AI Procurement in a Box as the project scales to other continents and countries.

As with any global project, it is crucial to monitor and evaluate the performance of AI Procurement in a Box, especially when scaled to different contexts and jurisdictions. This section analyses the desired impact, highlights possible gaps and provides a roadmap for the metrics and elements needed to get it there.

Government entities should implement the AI Procurement toolkit in actual procurement processes, taking adaptation to local laws and regulations into account. As mentioned above, more evidence is needed to confirm the market-shaping role of procurement to foster the development of ethical and human-centred AI. However, the pilots hosted in Brazil, India, the United Kingdom, Bahrain, and the United Arab Emirates confirm that the project scaling phase is highly dependent on the context and that there is no one-size-fits-all solution to successful implementation.

Real-world procurement and data governance use cases rely heavily on organizations hosting such pilots. Implementation partners[54] have limited control over some actions, like workshops, landscape review, community building and the definition of pilot projects and adaptation of the guidelines to local legislation. Figure 10 shows how the implementation of the toolkit depends on activities within and outside of the implementation partner's direct control and depends on high-level commitment from pilot organizations to produce impact.

FIGURE 10 **AI Procurement in a Box: selected aspects from project implementation**



4.1 Indicators and metrics

Table 6 presents five indicators that implementing partners can use to monitor the implementation of the Forum's toolkit in public sector entities and partner organizations. Whereas AI Procurement in a Box aims to describe the journey that governments take in acquiring AI, these indicators will hopefully show how implementing partners have helped transform their AI governance within their specific contexts. Ideally, the criteria presented above will show what success can look like and reveal obstacles and bottlenecks faced during implementation. Through continuous monitoring and evaluation, the AI Procurement in a Box will be reiterable for future pilots worldwide.

TABLE 6 **AI Procurement in a Box: indicators and metrics for project monitoring and evaluation**

Indicator	Content	Metrics and evaluation criteria
Knowledge absorption	a. *Comfort level with guidelines* – How are public sector entities responding to the guidelines throughout the AI Procurement workshop? b. *Understanding of relevancy of guidelines to projects* – Are the public sector entities able to grasp how these guidelines can affect current projects? c. *Potential barriers to guideline execution* – Are there any context-based barriers that can hinder the use of the guidelines?	– Number of capacity-building workshops conducted – Number of audience members actively engaging in workshop activities – Number of public sector entities that want to accompany the results of the pilot and commit to using guidelines in future projects – Number of potential barriers to guideline adoption – Did the pilot organization enter a formal commitment to implement the guidelines?
Institutional maturity	a. *Team diversity and multidisciplinary hiring* – Does the pilot organization take guideline #7 into account when creating a team? b. *Technical and data-related limitations* – How did the pilot organization evaluate its preparedness in terms of data access and IT infrastructure? c. *Inter-agency cooperation* – At what points has the pilot organization cooperated with other government agencies?	– Number of different disciplines represented in teams – Number of people with different socio-economic backgrounds – Is there an IT department? – Is there a specific office for artificial intelligence (AI)-related projects? – Is there a data governance department or chief data officer? – Can the institution discuss its level of AI maturity?
Procurement process	a. *Legislative certainty* – Are there any current laws that could create barriers to the use of AI within the public sector? b. *Innovation procurement procedures* – What innovation procurement procedures currently exist in the country's context? c. *Incorporation of guidelines within RFP* – Is the pilot organization able to communicate to potential suppliers/vendors the technical limitations of the project, potential risks, and any other important information regarding the project during the procurement process? d. *Proposal evaluation* – Does the pilot organization feel comfortable in choosing a supplier or vendor?	– Types of different procurement procedures that can be applied to acquire innovative goods and services – Number of previous AI procurement RFPs successful in that country/context – Number of guidelines effectively applied in pilot RFP – Number of proposals received considering AI procurement evaluation criteria – Use of the Forum's evaluation and specification tool during selection procedure

Life-cycle preparedness	a. *Definition of ongoing relationship with vendor/supplier* – Will the public sector organization be able to call the supplier if there is a distortion or concept drift within the AI model? b. *Capacity building* – Will the supplier/vendor train the public sector organization's staff on how to use this technology? c. *How will the model be monitored?* Who will monitor, how and when? Will there be a defined auditing schedule?	– Was the relationship between public organization and vendor supplier specified in the contract? – Is there a clause in the contract regarding bias mitigation or concept drift? – Number of capacity-building sessions held by vendor/supplier – Number of employees that attended the above-mentioned capacity building session – Number of times the model will be audited during a year – Number of risk assessments or other mechanisms performed in a year
Privacy and ethics	a. *Data protection and privacy* – What steps were taken to ensure that data protection legislation was implemented? b. *Governance structures* – What are the current oversight structures within the pilot organization? How will the decision-making process be delegated? What are the mechanisms for conflict resolution? c. *Risk analysis execution* – Does the pilot organization undertake a risk evaluation such as an AIA? d. *Additional policy instruments* – What other mechanisms were made to ensure the ethical oversight of AI?	– Does the country have data protection legislation and does it extend to AI applications? – Was a risk assessment executed? – Number of risk assessments or other mechanisms performed in a year – Will there be a specific entity to oversee the AI project? – Did suppliers seek ethical AI certifications? – Are there pre-approved vendors for ethical AI?
Dissemination	a. *Communication to the public* – What transparency measures were incorporated? b. *Interdepartmental dissemination* – How was the project communicated to other public sector agencies? Will the department adopt the guidelines as norms? c. *Standardization* – Were entities and authorities involved in discussing the procurement guidelines? Is there a possibility of adopting the guidelines at a national level?	– Number of media publications or other communication regarding the AI project – Number of internal/external events for the AI project – Was the national or state digital transformation authority or audit agency looped into the project? – Number of presentations with the above-mentioned authorities

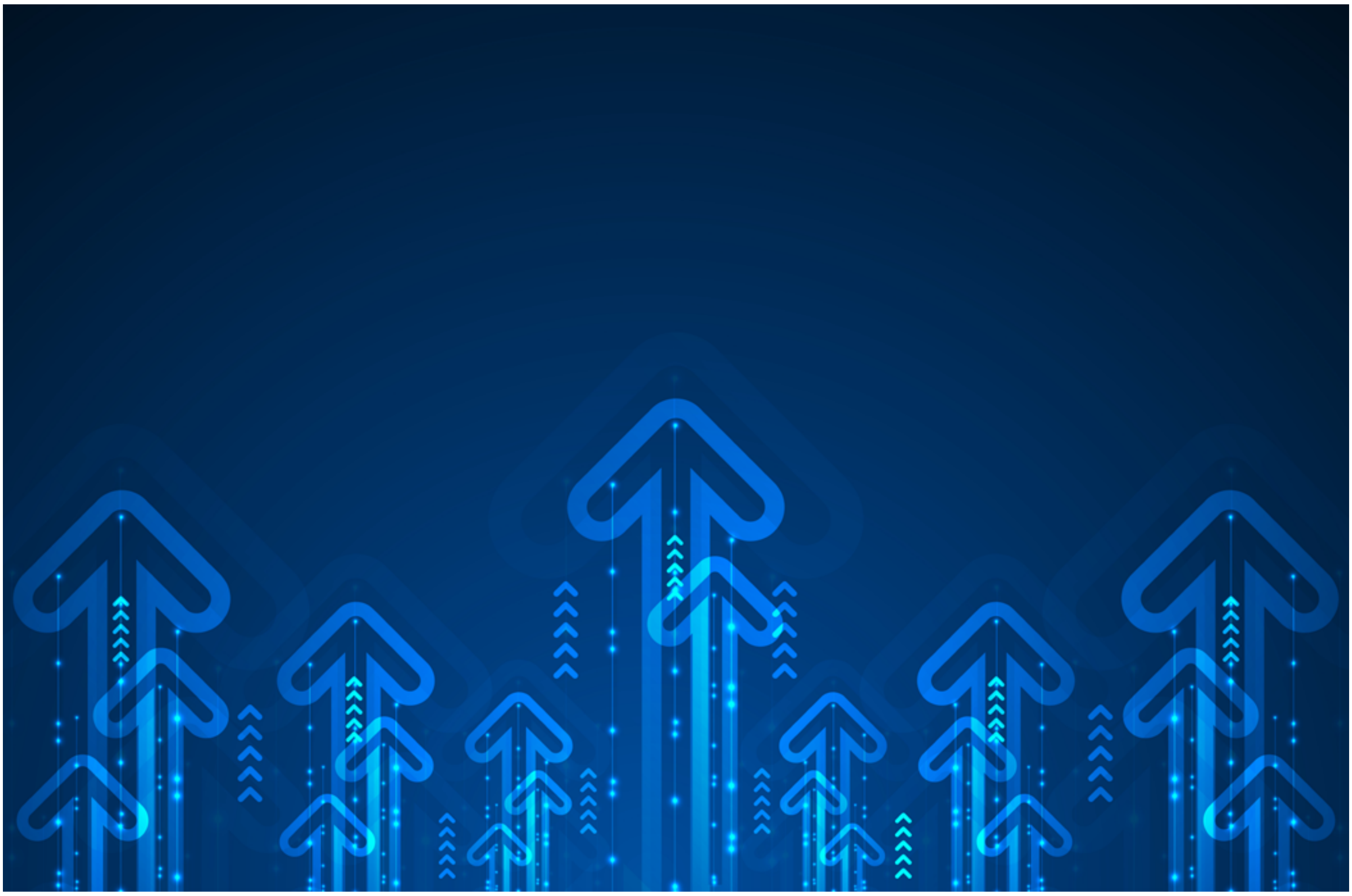
Conclusion

Moving forward with implementation of the AI Procurement in a Box project with new pilots and use cases in multiple jurisdictions.

Two years after publishing the toolkit, the World Economic Forum's guidelines on how to acquire AI ethically are still relevant across countries and contexts. Due to the loss of mobility during the COVID-19 pandemic, many countries rushed to digitalize government services. In Latin America alone, the number of citizens accessing the internet for government transactions jumped from 21% to 39% during the pandemic.[55] Government reliance on emerging technologies such as AI is set to grow exponentially, so it is imperative to prepare institutions around the globe through best practices such as those associated with AI Procurement in a Box to best use it.

With the resurgence of new literature and new pilots, the Forum and its Centre for the Fourth Industrial Revolution Network have identified new topics that need to be addressed in future pilots. As AI technologies evolve, it is also necessary for governance structures to encompass new issues, as well as ensure the inclusive implementation of AI/ML worldwide. The experiences of piloting the project in Brazil can support nations in the Global South as they address the challenges in adopting public sector AI.

The pilots and their outcomes add to the overall AI Procurement in a Box project by exploring elements to facilitate AI adoption in the Global South and incentivize debates within the public sector on how to procure AI responsibly and efficiently. As more countries consider adopting AI Procurement in a Box, more elements on how to best to do so and build upon these best practices will come to the forefront. This project will therefore continue to evolve – just like the technology it aims to demystify – to ensure that different countries with different contexts can effectively adopt it, even at different government levels.



Contributors

Lead authors

World Economic Forum

Rafael Carvalho de Fassio
Artificial Intelligence and Machine Learning Fellow, seconded from the State Attorney-General's Office in São Paulo, Brazil

Centre for the Fourth Industrial Revolution, Brazil

Clara Clemente Langevin
Project Lead, Artificial Intelligence and Machine Learning

Contributors

World Economic Forum

Hubert Halopé
Artificial Intelligence and Machine Learning Platform Curator

Emily Ratté
Artificial Intelligence and Machine Learning Project Specialist

Centre for the Fourth Industrial Revolution, Brazil

Julien Marc Hannigan Pigot
Senior Analyst

Centre for the Fourth Industrial Revolution, Brazil Executive Committee

President

Secretariat of Economic Development, State Government of São Paulo

Vice-President

Silveira, Fernando
Brazilian Association for Health Technology Industries (ABIMED)

Members

Alvim, Paulo
Ministry of Science, Technology, and Innovation of Brazil

Bernucci, Liedi
Institute of Technological Research, State Government of São Paulo

Calvet, Igor
Brazilian Agency for Industrial Development

Coelho, Milene
Astrazeneca

Conca, Jackline
Secretariat of Innovation and Digital Transformation, Ministry of Economy of Brazil

Lopes, Eduardo
Meta

Mendonça, Saul
Eletrobras

Silvestre, Julio
Bracell

Tonisi, Luiz
Qualcomm

See the full list of contributors

Endnotes

1. In this white paper, the terms artificial intelligence (AI) and machine learning (ML) are used interchangeably, since ML is the most prevalent AI technique in use today.
2. Centre for the Fourth Industrial Revolution Brazil, *Guia de Contratações Públicas de Inteligência Artificial*, March 2022, https://ideiagov.sp.gov.br/guia-de-contratacoes-publicas-de-inteligencia-artificial/.
3. Organisation for Economic Co-operation and Development (OECD), *Public Procurement for Innovation: Good Practices and Strategies*, OECD Public Governance Reviews, OECD Publishing, Paris, 2017, https://doi.org/10.1787/9789264265820-en.
4. Naudé, W. & Dimitri, N., *Public Procurement and Innovation for Human-Centered Artificial Intelligence*, IZA Discussion Paper n. 14021, 2021, http://dx.doi.org/10.2139/ssrn.3762891.
5. Engstrom, D. F. et al., *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, February 2020, https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf.
6. Ibid.
7. For example, the models adopted by the United States Digital Service (https://www.usds.gov) and 18F skunk work team for the General Services Administration (https://18f.gsa.gov).
8. Rubenstein, D. S., "Acquiring Ethical AI", *Florida Law Review*. vol. 73, November 2021, https://papers.ssrn.com/abstract=3731106.
9. Transparência Brasil, *Recomendações de Governança: Uso de Inteligência Artificial pelo Poder Público*, February 2020, https://www.transparencia.org.br/downloads/publicacoes/Recomendacoes_Governanca_Uso_IA_PoderPublico.pdf.
10. Ibid.
11. Engstrom, D. F. et al., *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, February 2020, https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf.
12. Arun, C., *AI and the Global South: Designing for Other Worlds*, Rochester, NY: Social Science Research Network, 9 June 2019, https://papers.ssrn.com/abstract=3403010.
13. Mbayo, H., *Data and Power: AI and Development in the Global South*. Oxford Insights, 9 March 2022, https://www.oxfordinsights.com/insights/2020/10/2/data-and-power-ai-and-development-in-the-global-south.
14. Nettel, P.F. et al., *Government AI Readiness Index 2021*, Oxford Insights, 2022, https://www.oxfordinsights.com/government-ai-readiness-index2021.
15. Smith, M. & Neupane, S., *Artificial intelligence and human development: toward a research agenda*, IRDC/CRDI, April 2018, http://hdl.handle.net/10625/56949.
16. Santeli, J.T. & Gerdon, S., "5 Challenges for Government Adoption of AI", *World Economic Forum*, 16 Aug. 2019, https://www.weforum.org/agenda/2019/08/artificial-intelligence-government-public-sector/.
17. Smith, M. & Neupane, S., *Artificial intelligence and human development: toward a research agenda*, IRDC/CRDI, April 2018, p. 12, http://hdl.handle.net/10625/56949.
18. Dener, Cem et al., *GovTech Maturity Index: The State of Public Sector Digital Transformation*, World Bank, 2021, https://doi.org/10.1596/978-1-4648-1765-6.
19. Broadband Commission's Working Group on Digital and AI in Health, *Reimagining Global Health through Artificial Intelligence: The Roadmap to AI Maturity*, Novartis Foundation, September, 2020, https://www.broadbandcommission.org/wp-content/uploads/2021/02/WGAIinHealth_Report2020.pdf.
20. Das, Arnab et al., "Modernizing Public Sector IT Infrastructure", *McKinsey and Company*, 10 Nov. 2019, https://www.mckinsey.com/industries/public-and-social-sector/our-insights/capturing-value-from-it-infrastructure-modernization-in-the-public-sector.
21. Ibid.
22. Novartis Foundation, *AI platform in São Paulo, Brazil helps diagnose COVID-19 patients faster*, 11 March 2022, https://www.novartisfoundation.org/transforming-population-health/healthtech-innovation/brazil-covid-19-response.
23. Zhang, B. & Dafoe, A., *Artificial Intelligence: American Attitudes and Trends*, Center for the Governance of AI, Future of Humanity Institute, University of Oxford, Jan. 2019, https://governanceai.github.io/US-Public-Opinion-Report-Jan-2019/us_public_opinion_report_jan_2019.pdf.
24. Knowles, B.& Richards, J.T., "The Sanction of Authority: Promoting Public Trust in AI", *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, ACM, 2021, pp. 262–71. *DOI.org (Crossref)*, https://doi.org/10.1145/3442188.3445890.
25. Organisation for Economic Co-operation and Development (OECD), *Reforming Public Procurement: Progress in Implementing the 2015 OECD Recommendation*, OECD Public Policy Reviews, OECD Publishing, Paris, 2019, https://doi.org/10.1787/1de41738-en.

26. Organisation for Economic Co-operation and Development (OECD), *Database of National AI Policies*, European Commission/OECD, 2021, 15 March 2022, https://oecd.ai.

27. Government of Canada, *Directive on Automated Decision-Making*, April 2019, 15 March 2022, https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592&section=html.

28. European Law Institute, *Model Rules on Impact Assessment of Algorithmic Decision-Making Systems Used by Public Administration*, European Law Institute/Universitat Wien, 2022, https://www.europeanlawinstitute.eu/projects-publications/completed-projects-old/ai-and-public-administration/.

29. Reisman, D. et al., *Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability*. AI Now Institute, Apr. 2018, p.5, https://ainowinstitute.org/aiareport2018.pdf.

30. Centre for the Fourth Industrial Revolution Brazil, *Guia de Contratações Públicas de Inteligência Artificial*, March 2022, https://ideiagov.sp.gov.br/guia-de-contratacoes-publicas-de-inteligencia-artificial/.

31. Rubenstein, D. S., *Acquiring Ethical AI*, SSRN Scholarly Paper, ID 3731106, Social Science Research Network, 1 Nov. 2021, p. 35, https://papers.ssrn.com/abstract=3731106.

32. Engstrom, D. F. & Ho, D. E., "Artificially Intelligent Government: A Review and Agenda", in: Vogl, R. (org), *Research Handbook on Big Data Law*, Northampton, MA: Edward Elgar, 2021, https://ssrn.com/abstract=3551549.

33. Cihon, P. et al., "AI Certification: Advancing Ethical Practice by Reducing Information Asymmetries", in *IEEE Transactions on Technology and Society*, vol. 2, no. 4, pp. 200-209, Dec. 2021, https://ieeexplore.ieee.org/abstract/document/9427056.

34. Responsible Artificial Intelligence Institute, *A certification for Responsible AI*, White Paper, 1 February 2022, https://www.responsible.ai/certification.

35. Cihon, P. et al., "AI Certification: Advancing Ethical Practice by Reducing Information Asymmetries," in *IEEE Transactions on Technology and Society*, vol. 2, no. 4, pp. 200-209, December 2021, https://doi.org/10.48550/arXiv.2105.10356.

36. McKinsey & Company, *The state of AI in 2021*, December 2021, https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2021.

37. Ibid.

38. Singh, M. et al., "An Empirical Study of Accuracy, Fairness, Explainability, Distributional Robustness, and Adversarial Robustness", ArXiv:2109.14653[Cs], September 2021, http://arxiv.org/abs/2109.14653.

39. Chen, J. & Storchan, V., "Seven challenges for harmonizing explainability requirements", ArXiv:2108.05390, August 2021, http://arxiv.org/abs/2109.14653. See also: Kahn. J., "What's wrong with explainable AI", *Fortune*, 22 March 2022, https://fortune.com/2022/03/22/ai-explainable-radiology-medicine-crisis-eye-on-ai/.

40. Engstrom, D. F. et al., *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, February 2020, https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf. See the table on p. 10 for an example of the wide array of AI tools currently being used by the US federal government.

41. European Commission, *Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence*, Brussels, COM(2021) 206, April 2021, https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206.

42. Green, B., *The Flaws of Policies Requiring Human Oversight of Government Algorithms*, September 2021, http://dx.doi.org/10.2139/ssrn.3921216.

43. Mulligan, D. K. & Bamberger, K. A., "Procurement as Policy: Administrative Process for Machine Learning", *Berkeley Technology Law Journal*, vol. 34, 2019, p.854-855, http://dx.doi.org/10.2139/ssrn.3464203. Rubenstein, D. S., "Acquiring Ethical AI", *Florida Law Review*, vol. 73, November 2021, p. 19, https://papers.ssrn.com/abstract=3731106.

44. Mulligan, D.K. & Bamberger, K. A., "Procurement as Policy: Administrative Process for Machine Learning". *Berkeley Technology Law Journal*, vol. 34, 2019, p.854-855, http://dx.doi.org/10.2139/ssrn.3464203.

45. Rubenstein, D. S., "Acquiring Ethical AI", *Florida Law Review*, vol. 73, November 2021, p. 19, https://papers.ssrn.com/abstract=3731106.

46. Engstrom, D. F. & Ho, D. E., "Artificially Intelligent Government: A Review and Agenda". In: Vogl, R. (org). *Research Handbook on Big Data Law*, Northampton, MA: Edward Elgar Publishing, 2021, https://ssrn.com/abstract=3551549.

47. Umbrello, S. & Yampolskiy, R.V., "Designing AI for Explainability and Verifiability: A Value Sensitive Design Approach to Avoid Artificial Stupidity in Autonomous Vehicles", *International Journal of Social Robotics*, 2021, https://doi.org/10.1007/s12369-021-00790-w.

48. US Department of Defense, "DOD Adopts Ethical Principles for Artificial Intelligence" 20 February 2020, https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/.

49. Norori, N. et al., "Addressing bias in big data and AI for health care: A call for open science", *Patterns*, volume 2, Issue 10, 2021, https://doi.org/10.1016/j.patter.2021.100347.

50. Mulligan, D. K.; & Bamberger, K. A., "Procurement as Policy: Administrative Process for Machine Learning", *Berkeley Technology Law Journal*, vol. 34, 2019, http://dx.doi.org/10.2139/ssrn.3464203. See also Kaminski, M. E. & Urban, J. M., "The right to contest AI", *Columbia Law Review*. vol. 121, no. 7, November 2021, https://columbialawreview.org/wp-content/uploads/2021/11/Kaminski-Urban-The_Right_to_Contest_AI.pdf.

51. World Economic Forum, AI Procurement in a Box: Workbook, 11 June 2020, https://www.weforum.org/reports/ai-procurement-in-a-box/workbook/.

52. Silverman, K., "We Are the Loop, Not Just in It: Success Needs a Focus on Humans at Every Step in the AI Lifecycle", *The AI Journal*, 14 May 2021, https://aijourn.com/we-are-the-loop-not-just-in-itsuccess-needs-a-focus-on-humans-at-every-step-in-the-ai-lifecycle/.

53. US National Security Commission on Artificial Intelligence, *Key Considerations for Responsible Development & Fielding of Artificial Intelligence*, Quarter 2 Report, July 22, 2020, p.30, https://www.nscai.gov/wp-content/uploads/2021/01/Key-Considerations-for-Responsible-Development-Fielding-of-AI.pdf.

54. Centre for the Fourth Industrial Revolution Affiliate Centres, government entities, academia and other partner institutions.

55. Roseth, B. et al., Servicios Públicos y Gobierno Digital Durante La Pandemia: Perspectivas de Los Ciudadanos, Los Funcionarios y Las Instituciones Públicas. Inter-American Development Bank, 2021, https://doi.org/10.18235/0003122.

COMMITTED TO IMPROVING THE STATE OF THE WORLD

The World Economic Forum, committed to improving the state of the world, is the International Organization for Public-Private Cooperation.

The Forum engages the foremost political, business and other leaders of society to shape global, regional and industry agendas.

World Economic Forum
91–93 route de la Capite
CH-1223 Cologny/Geneva
Switzerland

Tel.: +41 (0) 22 869 1212
Fax: +41 (0) 22 786 2744
contact@weforum.org
www.weforum.org