

# Addendum on AI and Digital Government

## A.1 Introduction

The integration of artificial intelligence (AI) in public administration has attracted global attention in recent years. It is widely accepted that AI technologies can improve public sector operations by replacing administrative tasks with automated processes, improving efficiency, and eliminating backlogs and redundancies. This shift is not without risks, however; the lack of careful scrutiny of AI advances during the past decade and the limited understanding of the nature and extent of their consequences have raised red flags in many countries, leading to what is sometimes referred to as the “AI regulation race”. Because the development of powerful AI technologies such as large language models (LLMs) outpaces the development of national policies and regulatory frameworks, AI has become a focal point in discussions on digital transformation in the public sector.

In March 2024, the United Nations adopted resolution *A/RES/78/265*,<sup>1</sup> underscoring the potential of AI to support or hinder the achievement of the 17 Sustainable Development Goals (SDGs). Followed by the resolution *A/RES/78/311*, which explores ways to enhance international cooperation on capacity-building for using and managing artificial intelligence. Furthermore, AI has been a key issue of discussion in multiple UN committees and working groups, including the global digital compact by the General Assembly as part of the Pact for the Future. At the national level, many leading countries, including Canada, China, the members of the European Union, the Republic of Korea, Singapore, and the United States of America, are actively pioneering the regulation and use of AI.

This addendum to the *2024 E-Government Survey* focuses on the integration of AI in the public domain and digital government. While this is a specific and narrower aspect compared to the broader discussions on AI regulation and governance in general, it remains a crucial venue for review. Furthermore, this addendum aims to explore the integration of AI in the public domain, which is impacted by the much broader discussion on the regulatory frameworks aimed at benefiting society. The addendum provides a short summary of the perceived opportunities and challenges associated with AI technology, specifically in the public sector, followed by an overview of national, regional, and global efforts to regulate AI. It highlights the benefits of using regulatory sandboxes (a framework that allows being exempted from specific regulations to test a new product/process) in the development of regulatory frameworks, showcasing a number of successful initiatives in this area.



Photo credit: [shutterstock.com](https://www.shutterstock.com).

A.1 Introduction	159
A.2 AI in the public sector: opportunities and challenges	160
A.2.1 Opportunities	160
A.2.2 Challenges	160
A.3 AI governance and regulatory frameworks	161
A.3.1 United Nations initiatives	162
A.3.2 National AI strategies	163
A.3.3 Human-centric approaches	164
A.4 AI literacy and capacity-building	164
A.4.1 Robust structure for data and digital governance	165
A.4.2 AI literacy	166
A.4.3 Regulatory sandboxes	167
A.5 Key Recommendations	168
A.5.1 Building upon existing efforts	168
A.5.2 Laying the appropriate foundations for the advancement of AI technologies	169
A.5.3 Engaging in collective action	169

The addendum offers examples of existing global and national regulatory frameworks and agendas, providing recommendations on how to maximize the benefits while minimizing the negative effects of AI implementation in the public sector. A section highlighting methodologies and successful cases of AI capacity-building, including expanding AI literacy, is included to promote the inclusive development of AI technologies.

## A.2 AI in the public sector: opportunities and challenges

### A.2.1 Opportunities

AI is recognized as an effective tool for enhancing work efficiency and productivity. It can be used for a multitude of purposes, including detecting defects, classifying data, and making recommendations.<sup>2,3,4</sup> There are numerous examples of AI use for products and services in both the public sector and the private sector, though the latter has been particularly active and innovative in AI integration.

The advantages and opportunities associated with AI extend beyond the private sector; AI integration in the public sector can also deliver enormous benefits. AI has proven instrumental in effective public service delivery during crises such as the COVID-19 pandemic. In Togo, for example, the social protection programme Novissi utilized satellite imagery and household consumption data to identify the poorest villages.<sup>5</sup> These villages were then prioritized using machine learning algorithms and mobile phone data to effectively distribute \$22 million through three monthly mobile phone payments to 600,000 citizens in urban areas.<sup>6</sup> Other examples include the “virtual doctor” self-assessment tool powered by AI in Croatia and the use of sensors and AI algorithms in London to control traffic. These are excellent examples of how AI can contribute to addressing society’s problems.<sup>7</sup>

AI technologies can also support the achievement of the SDGs. The global community has been reminded time and again that the world is not on track to meet the Goals articulated in the 2030 Agenda for Sustainable Development. A comprehensive midpoint assessment of the 135 trackable SDG targets was conducted for The Sustainable Development Goals Report 2024, and the findings revealed only moderate or marginal progress for almost half the targets and either no progress or regression for 35 per cent of the targets since 2015.

The urgent call to action to accelerate progress towards the SDGs comes at a particularly challenging time, as the global economy is still reeling from the effects of the pandemic, with only a few countries having experienced substantial recovery. For most countries, there is a pressing need for enhanced efficiency with constrained budgets – a challenge ideally suited for AI algorithms. There are numerous instances where Governments have leveraged AI to enhance efficiency across various sectors. The Singapore government, for example, implemented AI in its service moments of life (now called Life SG) to streamline government services, including birth registration and elder care. The Indian government has implemented AI technologies into the agriculture sector, improving the value chain for more than seven thousand chili producers. Specifically, the ‘Saagu Baagu’ pilot project developed in cooperation between the World Economic Forum and the Telangana state government has benefitted participating farmers with a 21 per cent increase in chili yields per acre, a 9 per cent reduction in pesticide use, a 5 per cent decrease in fertilizer usage, and an 8 per cent improvement in unit prices due to quality enhancements.<sup>8</sup>

### A.2.2 Challenges

While the potential benefits of AI technologies are substantial, so are the potential risks. The ethical, security and social implications of AI must be carefully addressed. One ethical concern is data bias. AI algorithms are intrinsically data-based, meaning they rely heavily on accumulated data to produce results. Consequently, any bias in these data can lead to the misrepresentation or underrepresentation

of certain groups. This bias is especially problematic when Governments employ AI in the development of public policies intended to serve the entire population, including marginalized groups. The persistent digital divide highlighted in previous United Nations E-Government Surveys<sup>9,10</sup> and digital governance studies poses a substantial challenge to implementing AI technologies in the public sector, especially in middle-income, low-income and least developed countries.

Chapter 4 of the *2022 E-Government Survey* and chapter 3 of the *2024 E-Government Survey* provide extensive insight into digital disparities within and between countries. According to both Surveys, significant progress has been achieved in bridging the digital divide. As noted in chapter 3 of the present Survey, the proportion of the population without digital access has declined from 45 per cent (3.5 billion) to just over 22 per cent (1.73 billion) in recent years. However, these gains derive primarily from developments in Asia; Africa, Oceania and the Americas have seen little to no narrowing of the digital divide. Data disparities are particularly problematic when AI technologies are applied in critical areas such as health care. In the 2021 publication *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*, inclusiveness and equity are identified as key ethical principles for AI use in health care. The publication further emphasizes the need for careful monitoring and evaluation of AI technologies to avoid disproportionate impacts on specific groups.<sup>11</sup>

Moreover, the growing influence of AI technologies in the labor market signals a profound transformation and the potential risk of job displacement across various sectors. This necessitates the implementation of strong universal social protection systems to support those adversely affected by these rapid changes and facilitate smoother transitions to alternative forms of employment. Such measures should include unemployment benefits coupled with active labor market policies aimed at retraining displaced workers.

### A.3 AI governance and regulatory frameworks

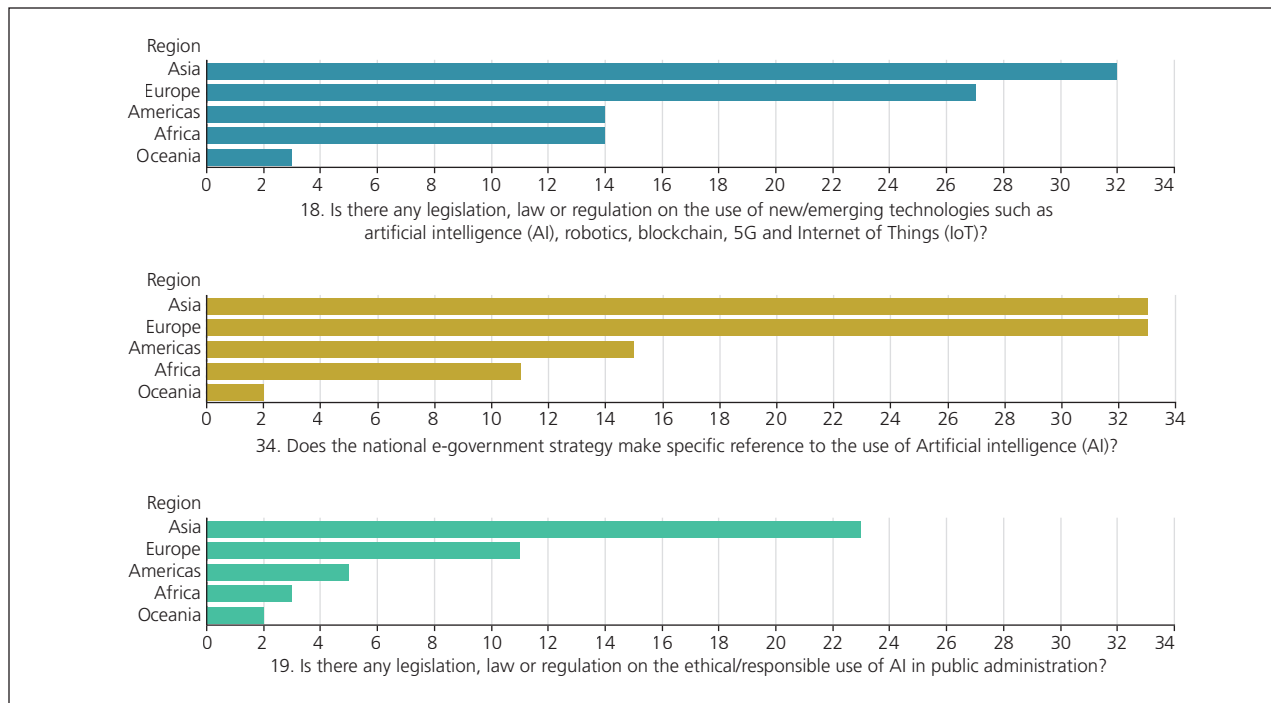
With the advent of LLMs and other powerful AI systems, the AI discourse has shifted from empowerment to regulation. The rapid evolution of AI technologies has outpaced the regulatory capacities of Governments; however, national and international authorities are aware of this dynamic and are taking steps to address the gap. Many United Nations member States have established different types of institutions to oversee the development and regulation of AI technologies and safeguard citizens against potential risks and threats.

The past several Surveys have endeavored to capture this transition and explore the evolving landscape of AI utilization and regulation by eliciting feedback on specific questions. Responses to the Member States Questionnaire (MSQ) disseminated by UN DESA indicate that 63 per cent (90 out of 142) of the countries surveyed for the 2022 to 2024 editions report having legislation or regulations on the use of new/emerging technologies such as AI, robotics, blockchain, 5G, and the Internet of Things. Almost half (44 out of 93) of the countries responding have already adopted legislation or regulations on the ethical/responsible use of AI in public administration. While these figures are notable, given that the discourse on the regulation of AI is relatively new, they remain well below the 88 per cent of countries (121 of 137) that have established e-government strategies or the equivalent. This disparity underscores the need for accelerated efforts to establish AI regulatory frameworks at the national level.

In response to the question “Does the national e-government strategy make specific reference to the use of artificial intelligence (AI)?”, a total of 94 countries from among the 2022 and 2024 respondents indicate that their national e-government strategies include explicit reference to AI, an increase from 65 countries in 2020. While this represents solid progress, it also points to the absence of AI provisions in many national e-government strategies. Countries that have not yet integrated AI into their digital agendas should carefully consider doing so to assess the potential benefits and challenges the technology may bring, as AI can greatly enhance governance and service delivery.

MSQ responses on the regulation of AI technologies reveal a regional imbalance. Asia is the leading region, ranking first in all questions relating to AI governance, closely followed by Europe. However, the Americas, Africa and Oceania lag far behind in the adoption of AI regulations and policies, with compliance rates averaging less than half of those achieved in Asia and Europe. Oceania, in particular, requires substantial support if it is to catch up with the other regions assessed, as only two to three of its countries have adopted AI-related regulations or policies.

**Figure A.1** Numbers/Percentages of countries responding affirmatively to MSQ questions on AI regulation, 2022 and 2024 cumulative total



### A.3.1 United Nations initiatives

The United Nations has been actively involved in discussions and activities surrounding AI technologies. The AI for Good Global Summit, first held in Geneva in June 2017, is an annual event organized by the International Telecommunication Union (ITU) and various partners within and outside the United Nations system. The Summit serves as a prominent platform for advocating the use of artificial intelligence in advancing the SDGs.<sup>12</sup> Other efforts undertaken by the United Nations and its agencies to enhance understanding and implement AI, are well documented in the ITU publication *United Nations Activities on Artificial Intelligence (AI) 2022* (Please also refer to the Annex A.18 for a detailed list of UN initiatives on AI). The United Nations addresses AI topics and issues internally and through its specialized agencies, and discussions on AI are incorporated into the work of existing committees and bodies, including the Commission on Science and Technology for Development and the Committee of Experts on Public Administration.

One milestone in the regulation of AI technologies by the United Nations and its agencies is the publication of the UNESCO Recommendation on the Ethics of Artificial Intelligence in 2021. This first-ever global standard on AI ethics, adopted by all 193 Member States in 2023, is intended to serve as a universally accepted normative instrument on AI technologies. It incorporates four critical values and ten principles that can be promoted through amendments to existing legislation and the elaboration of new legislation.<sup>13</sup>

The High-Level Advisory Body on Artificial Intelligence was established by the Secretary-General of the United Nations in October 2023 announcing the interim report on governing AI for humanity in December of 2023 in which they call for strengthened international governance of AI carrying out seven critical functions. Likewise, the first resolution on artificial intelligence (A/RES/78/265) was adopted by the United Nations General Assembly in March 2024 seeking methods to implement a safe, secure, and trustworthy artificial intelligence systems, followed by the resolution that calls for enhanced international cooperation (A/RES/78/311), marking another set of milestones in the international governance of AI technologies. The resolutions recognize that AI technology, when used responsibly, can contribute significantly to the achievement of all 17 SDGs by fostering economic, social and environmental progress to improve the global welfare and advance sustainable development. Although the resolutions do not impose any immediate binding obligations, it identifies the need for cooperation with and support for developing countries in narrowing the digital divide and improving digital literacy to ensure more inclusive access to AI technologies.

Through these resolutions, the United Nations affirms that the ultimate objective of AI is to enhance human welfare and achieve sustainable development, consistent with the provisions of the Charter of the United Nations, the Universal Declaration of Human Rights, and the 2030 Agenda for Sustainable Development. Towards this end, the United Nations is committed to establishing a global consensus on the development and implementation of AI systems that are safe, secure and trustworthy.

### A.3.2 National AI strategies

Many countries have established or are moving towards the establishment of AI regulations and policies. China, a leader in this area, announced its New Generation Artificial Intelligence Development Plan in 2017; this has since been supplemented by various regulations, including the 2023 Measures for the Management of Generative Artificial Intelligence Services, which address the meteoric rise of LLMs such as ChatGPT. Under these Measures, companies providing services to Chinese users are required to comply with Chinese regulations on personal information, take steps to protect the physical and psychological well-being of individuals, and uphold core values.<sup>14</sup>

The United States and the United Kingdom of Great Britain and Northern Ireland are not far behind in the regulatory race. In November 2023, the two countries released the joint Guidelines for Secure AI System Development, with contributions from various global ministries and agencies.<sup>15</sup> These Guidelines are important for two main reasons. First, they emphasize the importance of maintaining security not only during the initial development phase but also throughout the entire AI system development life cycle, adhering to secure-by-design principles. Second, they collectively represent one of the first internationally agreed-upon guidelines on AI development, setting a precedent for global cooperation in this field. Additionally, the United Kingdom of Great Britain and Northern Ireland hosted the “AI Safety Summit” in November 2023, marking the first global summit focused on AI safety. During this event, 28 countries adopted “The Bletchley Declaration on AI Safety.” This was followed by the second summit, the “AI Seoul Summit,” held in Seoul in 2024.

The European Commission has released its first-ever legal framework on AI, known as the AI Act, focusing on the risks associated with AI and positioning Europe to play a leading role globally. Announced in April 2021 and adopted in March 2024, this regulation addresses the risks of specific uses and applications of AI, including generative AI, by classifying them into four different levels: unacceptable risk, high risk, limited risk, and minimal risk. The primary objectives are to guarantee the safety and fundamental rights of people and businesses while also strengthening AI uptake, investment and innovation across the European Union. Under this Act, high-risk AI systems – those used for critical infrastructure, education, product safety, employment, essential services, law enforcement, migration, and the administration of justice – will be subject to strict obligations in the areas of risk assessment, the quality of data sets, traceability, documentation, user information, human oversight, and robustness.

The Brazilian Artificial Intelligence Strategy guides the actions of Brazil in research, innovation and the ethical use of AI. The Strategy is based on five principles developed by the Organisation for Economic Co-operation and Development for the responsible management of AI systems: inclusive growth, human-centred values, transparency, robustness, and accountability. The Strategy aims to develop ethical principles, promote AI research and development, remove barriers to innovation, educate professionals, and foster international cooperation.<sup>16</sup>

### A.3.3 Human-centric approaches

The public sector plays a vital role in the smooth functioning of society and can have a significant impact on people's lives. To best serve the public interest, governance should be based on core values that include integrity, equity, sustainability and accountability.<sup>17</sup> AI tools used within the public sector must reflect and reinforce these core values, but this may be problematic, as AI algorithms are set up to return the most likely result for a given task with no regard for ethics, social norms or societal standards. A human-centric approach to the adaptation and application of AI technologies is needed to ensure that AI-driven e-government is secure, effective and aligned with social values.

One potential strategy is to establish certification standards akin to those the United States Federal Communications Commission or the European Union has for electronic products but specifically tailored to AI. The widespread adoption of such standards would make it possible to detect and prohibit the use of AI systems that may pose hazards to humans.

Another potential human-centric approach involves incorporating a human element in the automation process. Although AI tools are powerful, they are not accountable for the results they provide, so human supervision is needed to close gaps in the chain of responsibility for AI processes and outcomes. Countries should implement a humans-in-the-loop or humans-on-the-loop approach to oversee the use and application of AI and ensure accountability. Where possible, countries should explore ways to integrate explainable AI (XAI) in digital administration and oversight to enhance transparency and allow for a thorough review of AI algorithms by human coordinators.

A great example of the human-centric approach is the aforementioned AI Act adopted in Europe, which seeks to regulate AI technologies using a four-tiered risk-based framework. This framework prohibits AI applications that present unacceptable risks to humans. Under this legislation, all suppliers introducing AI products or deploying AI systems in the European Union market must assess the risk level of their product or system and comply with the corresponding regulations.

## A.4 AI literacy and capacity-building

The technology landscape is rapidly evolving as new digital models replace older ones to achieve superior performance; as part of this evolutionary dynamic, AI advancement is inevitable. The fear of negative outcomes from a premature implementation of AI technology should not hold countries back from exploring its potential. Regulations must be accompanied by capacity-building measures to advance AI technologies and optimize the benefits the technology brings. Countries worldwide are channeling substantial financial resources into improving and expanding AI technologies. The Coordinated Plan on Artificial Intelligence, a joint commitment between the European Commission, European Union member States, Norway, and Switzerland, was published in 2018 to ensure a future-proof, human-centric, value-based digital transformation. The Plan provides for accelerated investment in AI, the implementation of comprehensive strategies and programmes, and the alignment of AI policies to prevent fragmentation within Europe.

AI development, adoption, and regulation are affected by the digital divide. Developing countries frequently encounter significant challenges in establishing robust data infrastructures. The inability to keep up with developed countries in digital development can pose a substantial threat to global stability, as the combined impact of AI and robotics is expected to transform market structures dramatically –

similar to the transitions observed during the early stages of the Industrial Revolution but at a markedly accelerated pace. As the transformation progresses, traditional job markets are expected to experience major restructuring, with many traditional jobs being replaced by AI and automated systems<sup>18,19,20,21</sup>. According to the International Labour Organization (ILO), high-income countries have the potential to automate 5.1 per cent of jobs and 13.4 per cent of augmentation potential. In contrast, low-income countries have significantly lower automation and augmentation potential, with only 1.3 per cent and 10.4 per cent of jobs, respectively. This translates to high-income countries experiencing more disruptive effects in technological transition as enjoying higher net gains from the process. Thus, the failure to address key disparities risks exacerbating global wealth polarization.

However, the lack of infrastructure in developing countries will make this transition challenging. A large segment of the population may find themselves in a position similar to that of typists or telegraph operators in the past, losing their jobs without viable opportunities for transition, further polarizing the global economy.

This underscores the urgent need to enhance the AI-related capacity of developing countries. Highlighting key elements essential for strengthening AI capabilities in these nations is crucial for bridging the gap between developed and developing countries. Building AI capacity will enable developing countries to harness the benefits of AI technology, promote innovation, and ensure they are not left behind in the digital transformation.

#### **A.4.1 Robust structure for data and digital governance**

A pivotal starting point in promoting AI capacity-building in developing countries is the establishment of robust data governance and digital governance structures. AI technologies fundamentally rely on data for their implementation, irrespective of the specific methodology employed (whether it be supervised, unsupervised, or reinforcement learning). AI is a product of machine learning algorithms that use historical data divided into three sets: training, validation and testing.

Developing AI technologies without a robust data foundation is impractical and could result in investments becoming non-performing assets, risking the sustainability of AI technologies. Even the latest generative models (including transformer-based models) require the input of accurate data to return accurate results. In other words, if the accumulated data are inconsistent or have wrong instances, they will lead to the creation of underperforming algorithms or algorithms that provide wrong results. The importance of a solid data infrastructure cannot be overstated.

AI technologies are intrinsically digital. Developing an AI framework or governance system disconnected from well-studied data or digital governance frameworks is not only inefficient but also risks overlapping with or contradicting existing efforts to govern the digital environment. While technological development does not follow a strict technological sequence or linear model, countries should be open to leapfrogging into an AI-savvy state. However, the belief that AI development can occur without substantial investment in the fundamental infrastructure (including data infrastructure) required for the development of AI is a misconception. Integrating AI governance within the broader context of digital governance is crucial. This integration ensures coherence and effectiveness by building upon existing efforts in data and digital governance and prevents countries from wasting the opportunity to benefit from prior advancements in the digital environment while preserving the potential for further AI development. It is recommended that new AI development initiatives be linked to and aligned with existing efforts to support the development of a connected digital infrastructure, including the digital public infrastructure and global digital cooperation initiatives (addressed in greater detail in chapter 3 of the present Survey).

This approach is consistent with a competency framework developed by UNESCO to support national digital capacity-building efforts. The three competency domains – digital planning and design, data use and governance, and digital management and execution – are essential for digital transformation and AI adoption. The approach also aligns with the conclusions of one of the first comprehensive

overviews of AI use and impact in public services, which asserts that Governments should view AI governance as an extension of existing regulatory tools.<sup>22</sup>

### A.4.2 AI literacy

Well-established data and digital governance structures must be complemented by strong human input for AI integration to be effective. Extensive research has shown that while knowledge in general and the understanding of technology more specifically are inherently non-rivalrous, they are not universally inclusive. Exclusion arises not only from protective measures such as patents and intellectual property rights but also from the substantial investment required to acquire the necessary background knowledge, particularly for complex concepts such as AI algorithms. Without that foundational understanding, technology-related knowledge and ultimately technology itself are largely inaccessible to the general public. The ability to understand and take advantage of widely distributed knowledge and technology is known as social or absorptive capacity.<sup>23,24</sup>

Digital capacity-building should target both producers (AI developers) and consumers (the end users or beneficiaries) of AI technologies in order to realize optimal economic and social benefits. Governments must strengthen citizens' basic awareness and understanding of AI concepts and applications by increasing their exposure to relevant concepts and providing AI education to build AI literacy. Bootcamps are an effective way to increase AI literacy within the general population. Singapore offers an integrated bootcamp programme aligned with their AI capability programme (AI Singapore) designed to identify and train AI professionals (see Box 1).

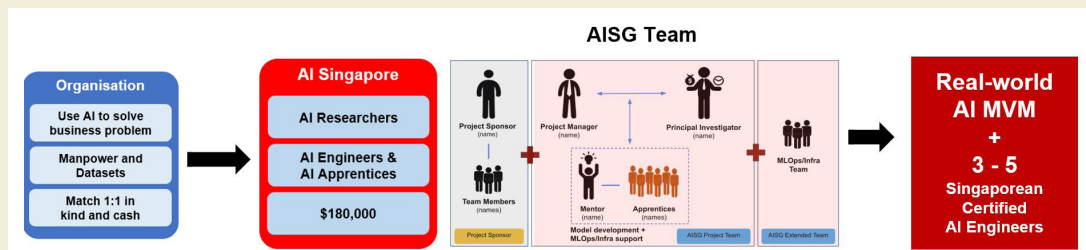
#### Box A.1 AI Singapore



AI Singapore was launched in 2017 to deepen AI capabilities in the country. This is an excellent example of an initiative that evolved as it was implemented, with programmes and subprogrammes added to support the development of a successful AI empowerment scheme. At present, the initiative incorporates six pillars: AI research, AI governance, AI technology, AI innovation, AI products, Learn AI.

The initiative started with a flagship programme called 100 Experiments. The two main objectives were to build real-world, deployable AI products and solutions for industries and to transfer knowledge and well-trained talent into the industrial sector. On the ground, the cycle was as follows: proposals were received from industries, engineering teams were formed, team members collaborated on creating AI solutions, and the solutions were transferred to the industries to enhance productivity and efficiency.

AI Singapore differs from other bootcamp programmes in that it focuses not only on creating AI solutions but also on identifying and training promising candidates to participate in the development of those solutions. In ideal circumstances, these apprentices are positioned as project coordinators for the same solutions when they complete the AI Apprenticeship programme.



100 Experiments programme structure

Source: AI Singapore, "100 Experiments programme structure", available at <https://aisingapore.org/innovation/100e/>.



The strengthening of AI capabilities in developing countries has also been undertaken through partnerships with world-renowned universities. Carnegie Mellon University Africa, for example, offers master's of science degrees in information technology, electrical and computer engineering, and engineering artificial intelligence. These programmes are designed to train innovative and technically proficient engineers within an African context. Leveraging the strong reputation of Carnegie Mellon, the University extends its influence beyond Rwanda, attracting students from across the African continent. Currently, the university has more than 300 students and more than 550 alumni representing 19 nationalities.

Improving AI capacity is an urgent priority for developing countries but is actually recommended for developed countries as well since AI literacy is required at all levels. To address this need, the AI4GOV programme – developed by Italy and Spain<sup>25</sup> and co-financed by the Connecting Europe Facility of the European Union – is administering a master's programme in artificial intelligence for public services. This ten-month graduate programme is designed to prepare future leaders in digital transformation, equipping them with the knowledge and skills they need to manage the development of AI and its adoption in the public sector.

### A.4.3 Regulatory sandboxes

As noted previously, AI development is occurring at a much faster pace than the implementation of AI regulations and policies. The increasing complexity of AI systems makes regulation particularly challenging. Discussions are starting to shift from narrow AI to general AI as development moves into uncharted territory, where outcomes are unknown but potentially transformative (see Box 2). Governments often lack the capabilities and expertise to fully understand the implications and anticipate the repercussions of emerging technologies. This can result in the overregulation or underregulation of AI technologies, depending on the strategy each country adopts. Overregulation can undermine the development of the AI industry but can be necessary in critical sectors such as health and education. Likewise, underregulation can create risks for end-users, lead to policy failures, and exacerbate digital exclusion. In this context, regulations must be carefully designed and accompanied by policies that can complement existing regulations and do not limit creativity or the potential to develop emerging technologies such as AI.

Regulatory sandboxes are innovation-friendly regulatory testing grounds that can help prevent overregulation and underregulation scenarios and give pioneering companies a temporary break from regulations while customized regulatory solutions are being developed. Sandboxes allow the testing of new products, services or processes in a controlled environment with a limited number of users, providing a safe testbed. The UN DESA policy brief on sandboxing highlights numerous successful cases of sandbox processes and frameworks, including those used for the adoption of digital currency in the Maldives and the promotion of cost-effective energy sources in Kazakhstan.<sup>26</sup> The practicality of this approach has been widely recognized, with a total of 218 regulatory sandboxes created by 2020.<sup>27</sup>

The success of regulatory sandboxes in sectors such as fintech and blockchain has paved the way for similar initiatives in artificial intelligence. In 2022, Spain started developing its first AI sandbox, anticipating the gradual enforcement of the European Union AI Act. This sandbox is designed to provide a secure environment for experimentation while ensuring compliance with the Act. With the enactment of Royal Decree 817/2023 in November 2023, the sandbox will be implemented in the coming years to test the application of AI Act regulations on high-risk AI systems, general-purpose AI systems, and foundational AI models. In 2024, the Government of Singapore set up the country's first sandbox on generative AI to allow small and medium-sized enterprises to gain experience in developing and applying relevant AI solutions. UN DESA is promoting regulatory sandboxes for AI, recognizing the increasingly important role AI will play in digital transformation in the public sector to support the effective delivery of public services and the achievement of the SDGs. During the

### Box A.2 Classifying AI technologies

There are multiple ways to classify AI technologies, with capability-based classification being the most widely used.<sup>28</sup>

*Narrow AI* is also known as weak AI. It can be trained to accomplish a specific task but cannot operate beyond that defined task. This is the only type of AI existing today.

*General AI*, also known as strong AI, remains a theoretical concept at this point. This type of AI can use previous learning and skills to accomplish new tasks without the need for humans to train the underlying model.

*Super AI*, or artificial superintelligence, is also strictly theoretical. It refers to those types of AI that surpass humans in cognitive abilities such as thinking, reasoning, learning, and making judgements. At present, these types of AI are projected to experience emotions, have needs, and possess their own beliefs and desires.

Source: IBM, "Understanding the different types of artificial intelligence", article, 12 October 2023, available at <https://www.ibm.com/think/topics/artificial-intelligence-types>.

United Nations Public Service Forum and Awards Ceremony held in Incheon in June 2024, UN DESA held a workshop on fostering innovation through digital sandboxes and AI sandboxes.

The key recommendations presented below are aligned with the aims of these initiatives, highlighting the importance of fostering innovation while ensuring the establishment of robust regulatory frameworks.

## A.5 Key Recommendations

There are three key recommendations that highlight essential steps for creating and sustaining a cohesive, innovative, and inclusive AI landscape in the public sector.

### A.5.1 Building upon existing efforts

**Key recommendation: Integrating AI governance within existing digital frameworks is crucial for creating a cohesive regulatory environment that reduces redundancy, supports innovation, enhances efficiency and security, and promotes public trust in digital technologies.**

The significance and potential of emerging AI technologies compel countries to develop new frameworks that both regulate and empower this assortment of advanced digital tools. However, AI technologies are ultimately part of a broader array of digital technologies created by scientists and engineers that are interdependent and collectively drive digital transformation. Governing these technologies separately can lead to duplicated efforts, conflicting regulations, and fragmented systems that waste resources and create operational inefficiencies. For example, data privacy and security regulations applied to AI should be consistent with those governing other digital technologies to avoid confusion and legal discrepancies.

The integration of AI governance within existing digital frameworks is essential for streamlining regulatory processes, maximizing efficiency, reducing redundancy, and ensuring cohesive regulatory practices. This approach not only supports the development and deployment of AI technologies but also strengthens overall digital governance, promoting innovation and protecting public interests. Implementing independent AI management technologies can conflict with established digital governance policies, leading to disruptions in AI advancements, increased operational costs, and potential legal disputes.

Ensuring data privacy and security becomes more challenging when AI governance is not aligned with broader digital strategies. An integrated framework helps avoid these conflicts, facilitating a smoother implementation of AI technologies. The digital and informational nature of AI, along with its heavy reliance on historical data, necessitates that AI regulation and governance be closely integrated with digital and data governance efforts. This integration ensures that AI technologies benefit from existing data management practices, enhancing their effectiveness and security. By building upon established digital governance structures, countries can better manage the complexities of AI, ensuring that these technologies are both well regulated and empowered to contribute to societal advancement.

### A.5.2 Laying the appropriate foundations for the advancement of AI technologies

**Key recommendation:** The advancement of AI requires a balanced approach that includes regulation, investment in strengthening AI capacity and literacy, and the creation of supportive infrastructures. This holistic strategy will enable countries to harness the full potential of AI technologies, ensuring that they contribute positively to sustainable growth and development and societal advancement.

The advancement of AI technologies is inevitable, given the wide range of efficiencies and other benefits they offer, raising the prospect of a fourth industrial revolution and significant disruptions to the job market. However, these advancements may also lead to considerable social unrest, such as unemployment resulting from rapid changes in the job market. Unlike in previous eras, the world has raised global awareness, underscoring the necessity of investing in universal solid social protection systems to support marginalized groups and creating appropriate institutions and mechanisms to enable a smoother transition.

Thus, under this new condition, countries should not limit themselves to regulating AI technologies but should also invest in strengthening AI capacity and AI literacy. For the most developed countries, setting up various types of regulatory sandboxes can be an effective strategy. These sandboxes allow for rapid development within a controlled environment, preventing risks from causing unforeseen threats to humanity. This approach enables innovation while maintaining safety and ethical standards. For countries lacking the appropriate infrastructure, establishing robust data frameworks and enhancing national AI literacy are crucial steps. However, these efforts must be integrated and coordinated in a holistic way rather than pursued independently by each institution. By creating a cohesive strategy, developing countries can generate both supply and demand for AI technology, fostering sustainable growth and development in this sector.

### A.5.3 Engaging in collective action

**Key recommendation:** The advancement and regulation of AI technologies require collective global action to ensure inclusive, safe and effective development. By working together, countries can establish a comprehensive and inclusive framework to ensure that AI technologies are used safely and effectively, benefiting all of humanity.

A few countries with the potential to pioneer development on this new frontier are spearheading the advancement and regulation of AI technologies. However, AI technologies are among those transformative phenomena that cannot be defined or regulated by a small group of nations since AI has the potential to alter daily human life fundamentally. Given this transformative potential, international rule making bodies such as the United Nations must adopt a proactive stance. The United Nations should promote the harmonization of national perspectives on AI and work towards establishing a general normative framework that all nations can agree upon and adhere to. This supports the rationalization behind the adoption of resolution A/RES/78/265, which focuses on achieving safe, effective AI regulation while allowing the technology to develop to its full potential.

International organizations should recognize that AI exclusion can occur at the local, national, regional and international levels, so it is crucial that an inclusive approach be pursued to prevent specific groups from being marginalized by AI technologies. As part of this effort, it is essential to ensure that all stakeholders, including those from less developed regions, have a voice in the global dialogue on AI. In a broad sense, international institutions must be persistent in their efforts to help lay the necessary foundations for a highly digitalized, rapidly changing world. This will involve supporting the development of different strategies tailored to the needs of each country based on their level of development and AI readiness. By addressing the challenges and opportunities unique to each country, international institutions can help create and sustain a balanced and equitable global AI landscape.

## Endnotes

- 1 United Nations, General Assembly, “Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development”, 1 April 2024 (A/RES/78/265), available at <https://documents.un.org/doc/undoc/gen/n24/087/83/pdf/n2408783.pdf?token=2pQCjT5fXQxtqIXeHP&fe=true>.
- 2 Yang Zhao and others, “Artificial intelligence-based fault detection and diagnosis methods for building energy systems: advantages, challenges and the future”, *Renewable and Sustainable Energy Reviews*, vol. 109(C) (July 2019), pp. 85-101, available at [DOI:10.1016/j.rser.2019.04.021](https://doi.org/10.1016/j.rser.2019.04.021).
- 3 Henry Y.T. Ngan, Grantham Kwok Hung Pang and Nelson Hon Ching Yung, “Automated fabric defect detection – a review”, *Image and Vision Computing*, vol. 29, No. 7 (June 2011), pp. 442-458, available at DOI:10.1016/j.imavis.2011.02.002.
- 4 Stoyan Mitov, “Three custom AI solutions to implement to Increase efficiency In 2024”, *Forbes*, 1 March 2024, available at <https://www.forbes.com/sites/forbestechcouncil/2024/03/01/three-custom-ai-solutions-to-implement-to-increase-efficiency-in-2024/>.
- 5 United Nations, Department of Economic and Social Affairs, *E-Government Survey 2022: The Future of Digital Government*, chap. 4 (New York, 2022), available at <https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2022>.
- 6 Malaka Gharib, “The pandemic pushed this farmer into deep poverty. Then something amazing happened”, *NPR*, 15 February 2021, available at <https://www.npr.org/sections/goatsandsoda/2021/02/15/966848542/the-pandemic-pushed-this-farmer-into-deep-poverty-then-something-amazing-happene>.
- 7 UN DESA, *E-Government Survey 2020: Digital Government in the Decade of Action for Sustainable Development* (New York, 2020), available at [https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2020-Survey/2020%20UN%20E-Government%20Survey%20\(Full%20Report\).pdf](https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2020-Survey/2020%20UN%20E-Government%20Survey%20(Full%20Report).pdf).
- 8 World Economic Forum, *AI for agriculture: How Indian farmers are harvesting innovation*, 11 January 2024, available at <https://www.weforum.org/impact/ai-for-agriculture-in-india/>.
- 9 UN DESA, *E-Government Survey 2018: Gearing E-Government to Support Transformation Towards Sustainable and Resilient Societies*, chap. 2 (New York, 2018), available at [https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2018-Survey/E-Government%20Survey%202018\\_FINAL%20for%20web.pdf](https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2018-Survey/E-Government%20Survey%202018_FINAL%20for%20web.pdf).
- 10 UN DESA, *E-Government Survey 2020: Digital Government in the Decade of Action for Sustainable Development* (New York, 2020), available at [https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2020-Survey/2020%20UN%20E-Government%20Survey%20\(Full%20Report\).pdf](https://publicadministration.un.org/egovkb/Portals/egovkb/Documents/un/2020-Survey/2020%20UN%20E-Government%20Survey%20(Full%20Report).pdf).
- 11 World Health Organization, *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance* (Geneva, 2021), available at <https://www.who.int/publications/i/item/9789240029200>.
- 12 International Telecommunication Union, *AI for Good Global Summit 2018*, available at <https://www.itu.int/en/ITU-T/AI/2018/Pages/default.aspx>.
- 13 United Nations Educational, Scientific and Cultural Organization, *Recommendation on the Ethics of Artificial Intelligence, Adopted on 23 November 2021* (Paris, 2022), available at <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.
- 14 China, Cyberspace Administration of China, Office of the Central Cyberspace Affairs Commission, Notice of the Cyberspace Administration of China on Soliciting Public Opinions on the Draft Measures for the Administration of Generative Artificial Intelligence Services (in Chinese), *Current Affairs News*, 11 April 2023 (source: China Internet Information Office), available at [https://www.cac.gov.cn/2023-04/11/c\\_1682854275475410.htm](https://www.cac.gov.cn/2023-04/11/c_1682854275475410.htm).
- 15 United Kingdom, National Cyber Security Centre and others, “Guidelines for secure AI system development” (2023), available at <https://www.ncsc.gov.uk/files/Guidelines-for-secure-AI-system-development.pdf>.

## ADDENDUM

- 16 OECD.AI Policy Observatory, Brazilian AI Strategy (Brazil, Ministry of Science, Technology and Innovation, 2021), available at <https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-27104>.
- 17 Janet V. Denhardt and Robert B. Denhardt, *The New Public Service: Serving, Not Steering*, 4th ed. (New York and London, Routledge, 2015).
- 18 Economist, “New research shows the robots are coming for jobs – but stealthily”, Finance and Economics page, 16 January 2021, available at <https://www.economist.com/finance-and-economics/2021/01/16/new-research-shows-the-robots-are-coming-for-jobs-but-stealthily>.
- 19 Serenity Gibbons, “How AI might impact the job market in 2024”, article, 7 December 2023, available at <https://www.forbes.com/sites/serenitygibbons/2023/12/07/how-ai-might-impact-the-job-market-in-2024/>.
- 20 Kristalina Georgieva, “AI will transform the global economy. Let’s make sure it benefits humanity”, IMF Blog post, 14 January 2024, available at <https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity>.
- 21 Andrew Green, “Artificial intelligence and the changing demand for skills in the labour market”, OECD Artificial Intelligence Papers, No. 14 (Paris, OECD Publishing, 10 April 2024), available at <https://doi.org/10.1787/88684e36-en>.
- 22 European Commission, Joint Research Centre, “AI watch – artificial Intelligence in public services”, report, 1 July 2020, available at [https://ai-watch.ec.europa.eu/publications/ai-watch-artificial-intelligence-public-services\\_en](https://ai-watch.ec.europa.eu/publications/ai-watch-artificial-intelligence-public-services_en).
- 23 Moses Abramovitz, “Catching up, forging ahead, and falling behind”, *The Journal of Economic History*, vol. 46, No. 2 (June 1986), pp. 385-406.
- 24 Wesley M. Cohen and David A. Levinthal, “Absorptive capacity: a new perspective on learning and innovation”, *Administrative Science Quarterly*, vol. 35, No. 1, Special Issue: Technology, Organizations and Innovation (March 1990), pp. 128-152.
- 25 AI4 Gov, Master - Artificial Intelligence for Public Services (ai4gov-hub.eu), available at <https://www.ai4gov-hub.eu/master/>.
- 26 Wai Min Kwok and others, “Sandboxing and experimenting digital technologies for sustainable development”, Future World Policy Brief, No. 123 (New York, UN DESA, December 2021), available at [https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/PB\\_123.pdf](https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/PB_123.pdf).
- 27 World Bank, “Key data from regulatory sandboxes across the globe”, Brief, 1 November 2020, available at <https://www.worldbank.org/en/topic/fintech/brief/key-data-from-regulatory-sandboxes-across-the-globe>.
- 28 IBM, “Understanding the different types of artificial intelligence”, article, 12 October 2023, available at <https://www.ibm.com/think/topics/artificial-intelligence-types>.