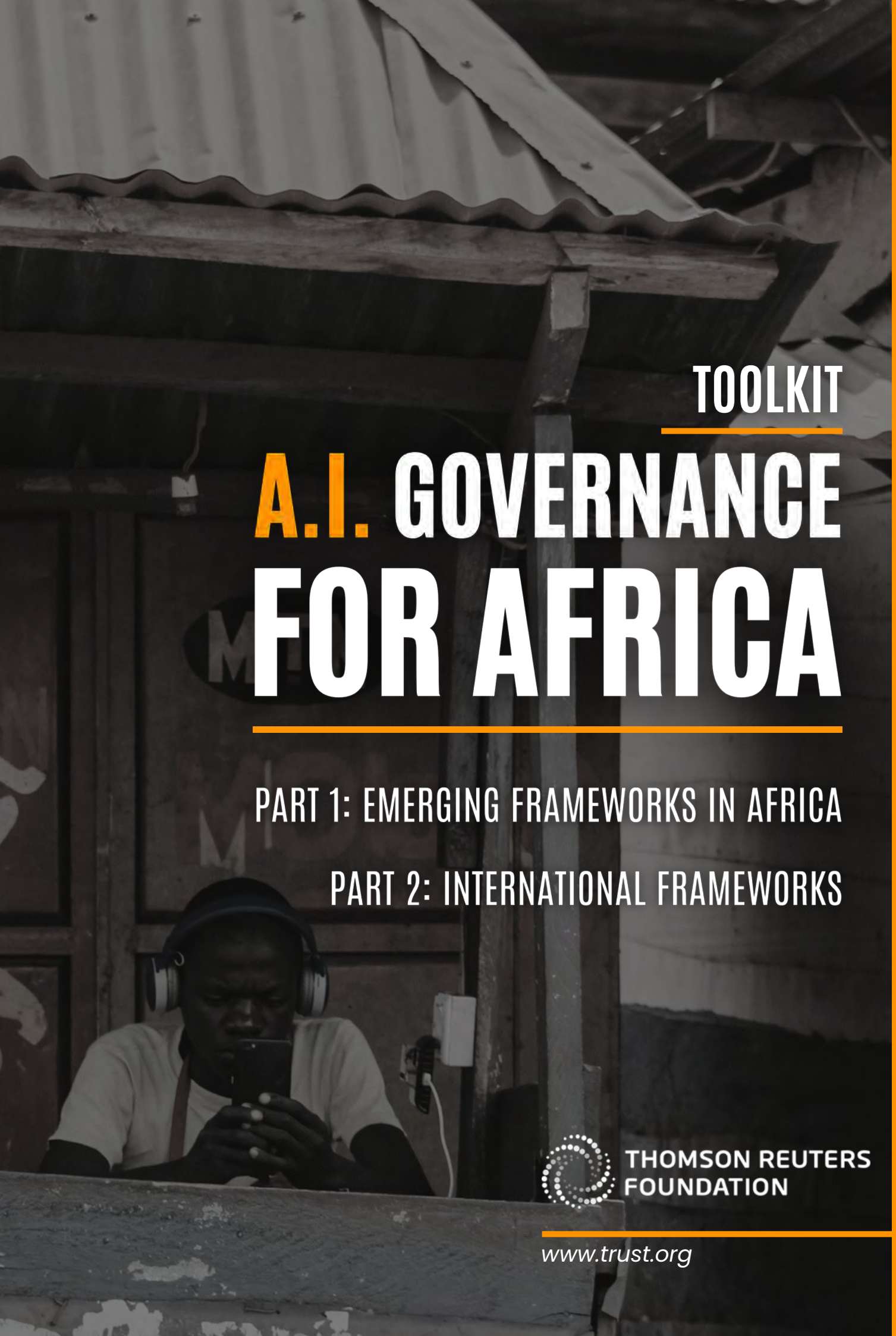


TOOLKIT

A.I. GOVERNANCE FOR AFRICA

PART 1: EMERGING FRAMEWORKS IN AFRICA

PART 2: INTERNATIONAL FRAMEWORKS

THOMSON REUTERS FOUNDATION

www.trust.org

AI Governance for Africa toolkit series

November 2023

This is **Part 1 and Part 2** of a three-part toolkit series on **AI Governance for Africa**.

Part 1 examines existing AI governance instruments on the continent, especially East Africa.
Part 2 unpacks existing international governance measures.

Part 3, published separately, explores options to build an advocacy strategy in pursuit of AI governance.

Published by the Thomson Reuters Foundation (trust.org)

Prepared by Tara Davis, Tharin Pillay, and Murray Hunter, from ALT Advisory (altadvisory.africa)

Supported by the Patrick J. McGovern Foundation (mcgovern.org)

Contents

PART 1
EMERGING AI GOVERNANCE FRAMEWORKS IN AFRICA

THOMSON REUTERS FOUNDATION

www.trust.org

INTRODUCTION

Artificial intelligence (AI) has permeated our everyday life. It is no longer confined to use by big tech or billionaires; ordinary individuals have it in their pockets – and they are using it.

Such common use has arguably been spurred by the public release of ChatGPT – OpenAI's chatbot which is capable of generating novel content in response to prompts. Within two months of its release in 2022, ChatGPT had 100 million active users, "making it the fastest-growing consumer application in history." [1] Its public release has made generative AI more accessible to the ordinary population, and people are using it in novel ways.

As AI systems become increasingly embedded in everyday life, they raise important questions about regulation, ethics, and their potential impact on human rights. These questions find form in debates about the governance of AI – how do we provide protection without stifling innovation? How can the law keep pace with the evolving nature of AI? Should AI be governed internationally or domestically?

Despite the complexity of AI governance, it is clearly a global concern. Countries are at different phases of resolving these questions and have implemented a range of governance instruments in response to concerns.

This toolkit unpacks the context of AI governance, in Africa and globally, and considers advocacy approaches for future governance. It does so in the following ways:

Part 1 examines existing AI governance instruments in Africa with a particular focus on the East African Community. This chapter outlines continental responses and details existing governing measures in Africa.

Part 2 unpacks existing international governance measures. In doing so it considers governance trends and important considerations included in governance instruments.

Part 3, which is a separate document, explores a series of key questions for the design of advocacy strategies on AI governance, particularly in African contexts.

As governance responses continue to develop, our hope is that this toolkit empowers journalists and civil society organisations to inform public discourse, drive policy and regulatory change and advocate for inclusive and responsible AI deployment.

[1] Krystal Hu "ChatGPT sets record for fastest growing user base – analyst note" *Reuters (1 February 2023)*. (Accessible here.)

Understanding Governance

Governance instruments are the tools, mechanisms and strategies that are used to guide, regulate, and manage the various aspects of AI. Some of the instruments used include:

- **Guidelines and Standards**: These are generally non-binding documents created by organisations, professional bodies and governments that provide a framework or set of principles to guide developers, users of AI or policy makers on important considerations such as fairness, transparency, accountability, and the avoidance of harm.

- **Government Strategy:** A government strategy is a high-level plan or approach that outlines the goals, priorities, and actions that a government intends to take in regard to AI. Strategy documents don't have the binding force of laws but serve as a guide for understanding a government's intended response.

- **Policy:** A government policy is a more detailed and specific set of guidelines, rules, or principles that guide decision-making and actions in a particular area. Policies provide a roadmap for implementing a government's strategy. Depending on domestic law, some policies may be enforced through measures such as monitoring or audits and could incur penalty such as disciplinary action or the revocation of a benefit or license.

- **Law:** Laws or regulations are codified rules enacted by a legislative body. The rules are enforceable, and non-compliance can result in penalty. In the context of AI, a country could implement one single law to deal with all aspects of AI (as evidenced by the EU-AI Act) or could take a fragmented approach where different laws regulate different aspects.

These instruments often interact with each other: government strategies guide the creation of policies, and policies can guide the drafting of laws to ensure that legal frameworks align with broader strategic goals. All of these instruments would likely consider the principles developed in guidelines and standards.

More about the various governance instruments, including examples, is provided in Part 2 of the Toolkit

The Continental Response

Significant impetus has been placed on the role that science and technology can play in the achievement of Africa's Sustainable Development Goals. [2] This is evident in various domestic policy documents, including those adopted by Mauritius, Egypt, and Rwanda. In light of this, there has been a steady increase in the deployment of AI in Africa, particularly in the areas of healthcare, agriculture, and mining.

[2] African Union High-Level Panel on Emerging Technologies (APET) and the African Union Development Agency (AUDA-NEPAD) *AI for Africa: Artificial intelligence for Africa's Socio-Economic Development* (2023) at page 46 ((AUDA-NEPAD report). (Accessible here).

The awareness of the benefits of AI is coupled with an acknowledgment of the myriad risks and challenges it poses. These include the common challenges around discrimination, bias and fairness, transparency, accountability, and data privacy.[3] However, there is also an acknowledgment of the unique challenges faced on the continent including digital inequalities, the lack of a structured data ecosystem[4] and concerns around access. This has prompted calls for context-specific responses. As noted by the African Union High-Level Panel on Emerging Technologies (APET) and the African Union Development Agency (AUDA-NEPAD):[5]

> "Africa's collective efforts cannot afford to continue with the habit of seeking for already-made solutions, from some other contexts attempting to counter African problems, as a matter of course. It should be noted that African problems are African context defined, and so, should be the approach to AI solution provisioning, which should be African home-grown."

AI is a policy concern for the continent and the African Union (AU) has encouraged the implementation of governance measures. In this regard, it was noted:

> "it is critical to have policies and regulatory frameworks in place that promote productive AI harnessing, by encouraging innovation and investment. AU should encourage African governments to take deliberate and proactive approach, to implement supportive regulation, policies, and initiatives."[6]

Although countries are taking domestic responses, there has also been a collective continental response, as evidenced by the development of several instruments included in Table 1. Some of these instruments are briefly detailed below.

Table 1 | Continental Instruments

Year	Instrument	Access
2013	The Smart Africa Manifesto	The Manifesto is available here
2014	The African Union Convention on Cyber Security and Personal Data Protection (The Malabo Convention)	The Convention is available here.
2019	Sharm El Sheikh Declaration adopted by the Specialised Technical Committee on Communication and Information Technologies of the African Union	The Declaration is available here.
2020	The African Union's Digital Transformation Strategy for Africa	The Strategy is available here.
2021	African Commission on Human and Peoples Rights adopts Resolution 473	Resolution 473 is available here.
2021	Artificial Intelligence for Africa Blueprint	The report is available here.
2023	Report titled 'AI for Africa: Artificial Intelligence for Africa's Socio-Economic Development' by the African Union High-Level Panel on Emerging Technologies (APET) and the African Union Development Agency (AUDA-NEPAD)	The report is available here.

[3] Ibid at 50.
[4] For more information about these challenges, see Abdessalam Jaldi *Artificial Intelligence Revolution in Africa: Economic Opportunities and Legal Challenges* (July 2023). (Accessible here.)
[5] AUDA-NEPAD Report at 49.
[6] Ibid at 45.

The African Union Convention on Cyber Security and Personal Data Protection (The Malabo Convention)

The AU Assembly adopted the Convention on Cyber Security and Personal Data Protection (the Malabo Convention) in 2014, and it finally came into force on 8 June 2023. This is a significant development as the Convention aims to establish a comprehensive legal framework for data protection, electronic commerce, and cybersecurity. Now that it is in force, all 55 AU member states are required to implement domestic laws that conform to the principles in the Convention. [7]

Although the Malabo Convention does not specifically address AI, it provides some useful standards concerning data protection. This is significant in light of the vast amounts of data required to train AI models. There are two notable provisions: first, Article 9 provides that the scope of data protection laws should include 'automated processing' within their scope of application. This means that AI systems have to comply with data protection laws when they process personal data. Second, Article 14.5 provides that a person should not be subject to a consequential decision that is based solely on the automated processing of their personal data. This means that an important decision about a person cannot be made entirely by a machine – there must be some human involvement. Most comprehensive data protection laws in African countries already include similar provisions, but the Convention is a welcomed development for countries that don't provide any protections.

Sharm El Sheikh Declaration and the AU Working Group on AI

In 2019, the African Union (AU) Ministers in charge of Communication and Information and Communication Technology and Postal Services adopted the Sharm El Sheik Declaration (the Declaration). It was adopted during the Third Ordinary Session of the AU Specialised Technical Committee on Communication and ICT. The Declaration acknowledges that digital transformation requires political commitment, the alignment of policies and regulation, and an increase in resources and investment. It further recognises that the AU requires a Digital Transformation Strategy in order to inform a coordinated response to digital technologies and the Fourth Industrial Revolution (4IR).

Importantly, the Declaration established a Working Group on Artificial Intelligence which is mandated to study the following: "the creation of a common African stance on AI; the development of an Africa wide capacity building framework; and establishment of an AI think tank to assess and recommend projects to collaborate on in line of Agenda 2063 and SDGs."

According to Egypt's Ministry of Communications and Information Technology, the purpose of the working group is as follows: [8]

[7] ALT Advisory *AU's Malabo Convention set to enter force after nine years* (May 19 2023). (Accessible here.)
[8] Egypt's Ministry of Communications and Information Technology *Egypt Hosts Second Meeting of African AI Working Group* accessed (14 December 2022). (Accessible here.)

> "Forming the AI Working Group is aimed at crafting an African AI strategy, creating a common stance in AI areas, and playing an active role in AI-related discussions on the global level. This common stance of African countries will reflect their needs and aspirations and ensure that the African voice is heard across international fora. Other objectives include addressing the various challenges that the continent is facing on that front, ensuring the governance of AI and the protection and availability of data, and developing AI regulations that might be a good starting point for such stance."

The working group is made up of experts from Egypt, Ghana, Kenya, Mali, Algeria, Cameroon, Ethiopia, and Uganda. [9] Egypt has been elected as the Chair of the Working Group, Uganda is the Vice Chair and Djibouti is the Rapporteur. The Working Group has met three times since its formation in 2019. [10] Despite the important role that this Working Group could play, there is limited public information about how it has, or intends to fulfil, its mandate.

Resolution 473

In March 2021, the African Commission on Human and Peoples' Rights (ACHPR) adopted Resolution 473 which concerns AI, robotics, and other new and emerging technologies. The resolution calls on State Parties to ensure that the development and deployment of such technologies is compatible with the rights in the African Charter. [11] Notably, it calls for State Parties to acknowledge these technologies on their agendas and to work towards a comprehensive governance framework. [12] It appeals to State Parties to maintain human control over AI, noting that the requirement should be codified as a human rights principle. [13] The Resolution commits to undertake a study to develop standards to address the challenges posed by such technology. [14] The study is not yet completed.

SMART Africa and the AI for Africa Blueprint

In 2013, seven Africa Heads of State [15] adopted the SMART Africa Manifesto which aimed to accelerate socio-economic development through the use of ICTs. Importantly, in 2014, the Manifesto was endorsed by all heads of State and Governments of the African Union, and now has 53 signatories. The SMART Africa Alliance has been formed to action and monitor compliance with the SMART Africa Manifesto.

The Smart Africa Alliance, together with several partners, developed an AI for Africa Blueprint in order to "outline the most relevant opportunities and challenges of the development and use of

---

[9] Egypt's Ministry of Communications and Information Technology *Egypt Hosts AU Working Group on AI First Session* (6 December 2019) (Accessible here.)

[10] Egypt's Ministry of Communications and Information *Technology Egypt Chairs AU Working Group on AI* (25 February 2021) (Accessible here.)

[11] African Commission on Human and Peoples Rights, Resolution 473 on the need to undertake a study on human and peoples' rights and artificial intelligence (AI), robotics and other new and emerging technologies in Africa, 10 March 2021 (Resolution 473), section 1.

[12] Section 4 and 5 of Resolution 473.

[13] Section 6 of Resolution 473.

[14] Section 7 of Resolution 473.

[15] Rwanda, Kenya, Uganda, South Sudan, Mali, Gabon, and Burkina Faso.

AI for Africa and how to address them"; and "to make concrete policy recommendations to harness the potential and mitigate the risk of AI in African countries." [16]

The Blueprint provides actionable recommendations to assist states with the implementation of national AI strategies. In doings so, it acknowledges the diversity of African states and accordingly does not propose a single AI policy solution. Instead, it provides guidelines that can be used by states to formulate their own, context specific policy. The Blueprint details 5 areas that it recommends should be considered in the formulation of a national policy. These include human capital, AI adoption (from lab to market), networking, infrastructure, and Regulation. [17]

Importantly, the Blueprint acknowledges the critical role that data plays in the development of AI and promotes the concept of open data. It acknowledges that open data can promote transparency and accountability and enable innovation by the private sector. [18] It notes that governments are in a position to make datasets publicly available while still adhering to data protection and privacy requirements. In this regard, it recommends that a common standard be developed on open data in order to ensure consistency and assist the private sector to prove access to dataset.

The Blueprint recognises the critical need for a robust governance framework to regulate AI. It notes that although some elements of Artificial Intelligence are regulated by existing laws, its unique nature, and the specific challenges it poses require legislative intervention. The Blueprint acknowledges the difficulty with the regulation of AI by stating: [19]

> "Uniformed approaches to governance can lead to systemic biases and overregulation that can and will stifle innovation, thus limiting the opportunities that can be leveraged and further creating an environment for political abuse. At the same time, underregulation will result in cultivating a culture whereby trust and confidence is absent, with consumers and citizens being left unprotected."

The Blueprint provides that an adequate legal framework should consider the following elements: [20]

1. "AI applications require copyright, patents, unfair competition laws.
2. Data requires various mechanisms such as data protection, data sharing, open data, decision on data localisation.
3. Ethics such as ethical driven design or guidelines for public procurement.
4. Legal provisions to enable the business environment such as incentives, infrastructure, cybersecurity, liability issues, licences.
5. It cuts across multiple regimes and industries such as financial markets, health and life insurance, taxation, telecommunication, etc."

---

[16] SMARTAfrica *Artificial Intelligence for Africa Blueprint* (2021) page 14 (AI Blueprint). (Accessible here).
[17] See Chapter 3 of the AI Blueprint.
[18] AI Blueprint at 38.
[19] AI Blueprint at 41.
[20] AI Blueprint at 41.

It further notes that the governance of AI will require a combination of hard and soft approaches. Hard approaches refer to the adoption of laws and regulations which it suggests are only necessary in response to a particular concern which cannot be solved through other measures. It recommends that a hard approach be taken for issues concerning copyright and patents, investment and intellectual property and unfair competition.[21] Soft law refers to substantive expectations that are not enforceable by governments, including guidelines, standards, codes of conduct and best practice. The Blueprint acknowledges that soft law will likely fill governance gaps while regulatory measures are being developed.

Regulatory Sandboxes

An example of a soft law measure which is often used in response to innovative technologies, including AI, is a regulatory sandbox.

A regulatory sandbox is a framework that allows "start-ups and other innovators to conduct live experiments in a controlled environment under a regulator's supervision."[22]
Mauritius has a Sandbox Framework for the Adoption of Innovative Technologies in the Public Service.[23] It enables the public sector to better understand the challenges, costs, and capabilities of emerging technologies before conducting a formal procurement process.

Regulatory sandboxes have also been used in Ghana, Nigeria, South Africa, Zimbabwe, and Rwanda.[24] Kenya's Communication Authority also recently held consultations on a sandbox framework for emerging technologies.[25]

Domestic Governance

As evidenced by continental initiatives, AI governance is on the agenda in Africa, and countries have taken various steps to implement domestic governance measures. Overall, progress is slow and there are stark differences in progress between states. Mauritius, for example, has taken significant strides – it was the first country to adopt an AI policy[26] and it has implemented a strategy,[27] data protection measures, and has certain rules in place that regulate service providers who use AI-enabled algorithms.[28] This is in strong contrast to several countries that have no measures in place, such as Burundi, the Democratic Republic of Congo and South Sudan.

[21] Ibid.
[22] AI Blueprint at 42.
[23] Republic of Mauritius Ministry of Public Service, Administrative and Institutional Reforms *Sandbox Framework for Adoption of Innovative Technologies in the Public Service* (March 2021). (Accessible here.)
[24] For more information on Regulatory Sandboxes and their uses by these countries, see Africa Observatory on Responsible Artificial Intelligence *Sandboxes in Mauritius* (8 June 2023). (Accessible here.)
[25] The consultation process has closed, but information about it can be accessed here.
[26] Ganiu Oloruntade "Where is Africa in the Global conversation on regulating AI?" *Techcabal* (26 May 2023). (Accessible here.)
[27] Mauritius *Artificial Intelligence Strategy* (November 2018). (Accessible here.)
[28] For more on this see: Nico Van Zyl "Mauritius issues AI-enabled advisory services rules" *Sovereign* (26 July 2021). (Accessible here.)

Table 2 below details the various governance measures that each country has implemented. According to our research we found the following:

Table 2 | Governance Instruments in Africa

Country	AI legislation	Data protection legislation addresses automated decision-making	Has a national AI strategy	Has a policy or draft policy on AI	Expert body on AI
Algeria	No	Yes	Yes	No	Yes
Angola	No	Yes	No	No	No
Benin	No	Yes	Yes	No	Yes
Botswana	No	Yes	No	No	No
Burkina Faso	No	Yes	No	No	Yes
Burundi	No	No	No	No	No
Cabo Verde	No	Yes	No	No	No
Cameroon	No	No	No	No	No
Central African Republic	No	No	No	No	No
Chad	No	No	No	No	No
Comoros	No	No	No	No	No
Congo	No	Yes	No	No	No
Cote d'Ivoire	No	Yes	No	No	No
Democratic Republic of Congo	No	No	No	No	No
Djibouti	No	No	No	No	No
Egypt	No	No	Yes	No	Yes
Equatorial Guinea	No	Unknown	No	No	No
Eritrea	No	No	No	No	No
Eswatini	No	Yes	No	No	No
Ethiopia	No	No	No	Yes (draft)	Yes
Gabon	No	Yes	No	No	No
The Gambia	No	Partial	No	No	No
Ghana	No	Yes	No	No	No
Guinea	No	Yes	No	No	No
Guinea-Bissau	No	No	No	No	No
Kenya	No	Yes	No	No	Yes
Lesotho	No	Yes	No	No	No
Liberia	No	No	No	No	No
Libya	No	No	No	No	No
Madagascar	No	Yes	No	No	No
Malawi	No	No	No	No	No
Mali	No	Yes	No	No	No
Mauritania	No	Yes	No	No	No
Mauritius	Partial	Yes	Yes	Yes	Yes

Morocco	No	Yes	Yes	No	Yes
Mozambique	No	No	No	No	No
Namibia	No	No	No	No	Yes
Niger	No	Yes	No	No	No
Nigeria	No	Yes	No	No	Yes
Rwanda	No	Yes	No	Yes	Yes
Sao Tome & Principe	No	Yes	No	No	No
Senegal	No	Yes	No	No	No
Seychelles	No	No	No	No	No
Sierra Leone	No	No	Yes	No	Yes
Somalia	No	No	No	No	No
South Africa	No	Yes	No	No	Yes
South Sudan	No	No	No	No	No
Sudan	No	No	No	No	No
Tanzania	No	Yes	No	No	No
Togolese Republic	No	Yes	No	No	No
Tunisia	No	Yes	No	Yes	Yes
Uganda	No	Yes	Yes	No	Yes
Sahrawi Arab Democratic Republic	No	No	No	No	No
Zambia	No	Yes	No	No	No
Zimbabwe	No	Yes	No	No	No
Total 55	0/55	31/55	7/55	4/55	15/55

0 of 55 countries have **dedicated AI legislation** in force.

31 of 55 countries have adopted a **data protection law** that addresses AI to some extent. Specifically, these 31 countries guard against automated decision-making.

7 of 55 countries have adopted a national **AI strategy.**

4 of 55 counties have a national **AI policy** in place.

15 of 55 countries have an **expert body or taskforce** that is engaging on AI questions.

As evidenced, Africa countries have developed limited governance responses to AI. No country has adopted dedicated AI legislation, and the most prominent form of governance is enabled through data protection laws. Although very few countries have adopted a national policy, several of them have expert bodies or developed strategies. This may suggest that their next steps will likely be the adoption of a formalised policy – following consultation and recommendations by

the body. A notable development in the policy space is Rwanda's recent adoption of an ambitious policy which includes commendable recommendations and a strong focus on ethical AI.

Despite the limited measures, it is clear from continental instruments such as the AI for Africa Blueprint that AI is gaining traction as a policy concern. However, more is required to effectively govern AI in most countries.

Understanding how data protection laws govern AI

As evidenced in Table 2, the most prominent form of the governance of AI is currently through data protection laws. This section briefly details how data protection laws regulate AI. Although there is some difference in domestic laws, data protection legislation generally does two things that concern AI: first, they include automated processing within the scope of their application and second, they provide a right against automated decision-making.

Automated processing: In the context of AI, automated processing involves the use of algorithms, rules, or instructions to perform tasks that would otherwise require human effort or decision-making. It can range from simple repetitive tasks to highly complex decision-making processes. It often involves the use of data and computational algorithms to analyse information, make predictions, optimize processes, and generate outputs. For example, automated processing can be used for data analysis where AI systems analyse large data sets to identify patterns or trends. It can be used for facial recognition where AI systems automatically process images to recognise objects or faces. These applications can have very real consequences – for example, they could determine a medical diagnosis or decide whether someone qualifies for a loan.

Data protection laws often included automated processing within the scope of application of the law. This means that when a company or government uses an AI system to collect personal data, collate it or analyse it, it has to be done in compliance with the data protection law. Generally, this means complying with rules around consent, minimality, purpose limitations and security safeguards.

The right against automated decision-making: The right against automated decision-making is a fundamental aspect of data protection laws and is provided in most legislation around the world. The right is aimed at safeguarding individuals from potentially harmful or unfair decisions made solely by automated systems without human intervention. In the examples above, AI applications could perform analytic or recognition functions without any human involvement. This means that they could analyse all available information about a specific individual and determine whether or not they qualify for a loan. Such decision could have legal or consequential implications for an individual. Such decisions carry further risk as the results may be discriminatory or biased and would lack explainability.

The right provided in data protection law attempts to guard against these risks by providing individuals with a right not to be subject to such a decision. The right generally requires that a data subject be notified when such a decision has been made and allows them to request it to be reconsidered. The purpose of the right is to ensure that there is a degree of human oversight and involvement in these decisions.

Governance in East Africa

This section looks at AI governance in East Africa in more detail. We consider the scope in terms of the AU's East African Community (EAC) which includes Democratic Republic of Congo, Burundi, Kenya, Rwanda, South Sudan, Uganda, and Tanzania. [29]

As noted in Table 3 below, no country in the EAC has dedicated AI legislation, but 4 out of the 7 countries regulate the processing of personal data by AI systems through their data protection laws. This provides a degree of protection against automated processing and decision-making as detailed above. Data protection laws are the most prominent form of AI governance in the EAC.

Notably, Uganda is the first country in the EAC to develop a national strategy on AI. The strategy forms part of its Digital Transformation Roadmap [30] which was released in August this year. The Roadmap includes several recommendations relating to the governance of AI, and we can accordingly expect the development of additional instruments from Uganda in the next few years. Rwanda has also taken notable strides in governance by being the first country in the EAC to adopt a national policy. It is an ambitious policy that includes commendable recommendations including strengthening regulatory measures and establishing a Responsible AI Office. Despite these developments, it is concerning that they are the only two countries who have formalised a national response to AI.

A summary of the findings is provided in Table 3 and the governance measures of each country are discussed in greater detail below.

Table 3 | Governance Instruments in the EAC

Country	Dedicated AI legislation	Data protection legislation addresses AI	Has a national AI strategy	Has a policy or draft policy on AI	Expert body on AI has been established
Burundi	No	No	No	No	No
Democratic Republic of Congo	No	No	No	No	No
Kenya	No	Yes [31]	No	No	Yes [32]

[29] East African Community "EAC Partner States". (Accessible here.)
[30] Uganda Ministry of ICT and National Guidance *Digital Transformation Roadmap 2023/2024 – 2027/2028* (2023). (Accessible here.)
[31] Data Protection Act 24 of 2019. (Accessible here.)
[32] The Taskforce on Distributed Ledgers and Artificial Intelligence. For more information See *The Kenya Gazette*, No.2095 of 3 August 2018. (Accessible here.)

Rwanda	No	Yes	No	Yes [33]	Yes [34]
South Sudan	No	No	No	No	No
Tanzania	No	Yes [35]	No	No	No
Uganda	No	Yes [36]	Yes [37]	No	Yes [38]

Burundi, Democratic Republic of Congo, and South Sudan

Country	Dedicated AI legislation	Data protection legislation addresses AI	Has a national AI strategy	Has a policy or draft policy on AI	Expert body on AI has been established
Burundi	No	No	No	No	No
Democratic Republic of Congo	No	No	No	No	No
South Sudan	No	No	No	No	No

These three countries currently have no governance instruments in place that regulate AI. This includes a lack of data protection laws.

Kenya

Country	Dedicated AI legislation	Data protection legislation addresses AI	Has a national AI strategy	Has a policy or draft policy on AI	Expert body on AI has been established
Kenya	No	Yes	No	No	Yes

Artificial Intelligence is squarely on Kenya's agenda; however, they only have two governance mechanisms in place – a data protection law, and a taskforce.

In 2018, The Ministry of Information, Communications and the Digital Economy of Kenya (ICDE) established a taskforce on Distributed Ledgers and Artificial Intelligence. [39] The taskforce was mandated to "develop a roadmap for emerging technologies that will define the evolving Fourth Industrial Revolution." [40] Their terms of reference included to critically review AI technologies, contextualise how they could contribute to Kenya's agenda and provide a roadmap for their

[33] Republic of Rwanda *The National AI Policy* (2023) (Accessible here.)
[34] Rwanda Centre for the Fourth Industrial Revolution. More information about it is accessible here.
[35] Personal Data Protection Act 11 of 2022. (Accessible here.)
[36] The Data Protection and Privacy Act, 2019. (Accessible here.)
[37] Uganda Ministry of ICT and National Guidance *Digital Transformation Roadmap 2023/2024 – 2027/2028* (2023). (Accessible here.)
[38] A National Expert Taskforce on the 4IR was established in 2018. See page 4 of Uganda's National 4IR Strategy. (Accessible here.)
[39] Government of Kenya, *The Kenya Gazette*, No.2095 of 3 August 2018. (Accessible here.)
[40] International Telecommunication Union Development Sector *Collaborative regulation for digital transformation in Kenya: A country review (2023),* page 25 (ITU report). (Accessible here.)

implementation. [41] The taskforce delivered a report to the Ministry of ICDE in July 2019. [42] The report considered how AI, and emerging technologies in general, could contribute to the following areas of intervention: food security, housing, healthcare, manufacturing, cyber security and land rights.

Kenya has also developed a Digital Economy Blueprint [43] which acknowledges the potential of AI to assist in Kenya's development of a digital economy. Although it provides important recommendations concerning the development of AI capacity, it is silent on the need for dedicated AI regulation.

Although Kenya does not have a dedicated AI law, its data protection Act [44] 24 of 2019 regulates artificial intelligence to some extent. The processing of personal information by AI systems falls within the scope of application of the Act. This is evident in section 4, which provides that the Act applies to the processing of personal data by automated means. Accordingly, the processing of personal data by an AI system would have to comply with the following data protection principles included in the Act: [45]

- it must be processed in accordance with the right to privacy;
- be processed in a lawful, fair, and transparent manner;
- be collected for a specific and legitimate purpose which must be specified to the data subject; and
- it cannot be further processed in ways that are incompatible with the stated purpose.

Further, section 35 provides data subjects with a right not to be subject to a decision based solely on automated processing, including profiling, which has significant consequences (legal or otherwise) for a data subject. Profiling is defined in the act to mean: "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person." Accordingly, the right ensures that consequential decisions, such as someone's eligibility for a loan, cannot be made without human intervention.

In terms of the right, if such a decision is taken, the data subject must be notified and may request that the decision be reconsidered, or that a new decision be made which is not based entirely on automated processing. [46] If this right has been infringed, a data subject may lodge a complaint with the Data Protection Commissioner, [47] which is the public body responsible for ensuring compliance with the data protection law. The Data Protection Commissioner is empowered to investigate the complaint and may issue an enforcement notice requiring an individual or company to take specified steps. Accordingly, such an enforcement notice could require that the decision be re-taken with a degree of human involvement.

---

[41] ITU report at 25.
[42] Ministry of Information, Communications and Technology "Emerging Digital Technologies for Kenya: Exploration and Analysis" (July 2019). (Accessible here.)
[43] Republic of Kenya *Digital Economy Blueprint: Powering Kenya's Transformation* (2019) (Accessible here.)
[44] Data Protection Act 24 of 2019. (Accessible here.)
[45] The Principles are provided in section 25 of the Data Protection Act 24 of 2019 and expanded upon throughout the Act.
[46] Section 35(3)(b) of the Data Protection Act 24 of 2019.
[47] Section 8(1)(f) & section 56 of the Data Protection Act 24 of 2019.

The right accordingly provides a degree of protection against consequential decision-making by AI and ensures a degree of human oversight.

Rwanda

Country	Dedicated AI legislation	Data protection legislation addresses AI	Has a national AI strategy	Has a policy or draft policy on AI	Expert body on AI has been established
Rwanda	No	Yes	No	Yes	Yes

Rwanda has taken huge strides in AI governance in the last few months, with a notable focus on responsible AI. Although it has not adopted AI specific legislation, the Cabinet of Rwanda approved a National Artificial Intelligence Policy (the Policy) in April 2023.[48] This is a notable development, making it the first country in the EAC with a national policy on AI. Further, Rwanda's data protection law regulates AI to some extent, by providing data subjects with a right against solely automated decision making. These instruments are discussed in greater detail below.

Rwanda's policy considers the use of AI in the following sectors: healthcare, banking and digital payments, e-commerce and trade, transportation, agriculture, public administration and education, manufacturing, and construction. Its purpose is to provide a roadmap for Rwanda to harness the benefits of AI for sustainable and inclusive growth and mitigate its risks. The Policy notes Rwanda's objective to position itself as "Africa's AI Lab and Responsible AI Champion."[49]

Importantly, one of the recommendations is to strengthen AI policy and regulation. In this regard, the Policy notes "trust is critical to public confidence and acceptance of AI. By strengthening the capacity of regulatory authorities to understand and regulate AI aligned with emerging global standards and best practices, we will build transparency and trust with the public."[50] Trustworthy AI is recognised as a priority area and the roadmap includes the establishment of a policy and regulatory capacity building program within 2023/24.

Significant focus is given to the development and operationalising of ethical guidelines. In this regard, the Policy notes that "ethical and safety precautions are required to ensure that AI solutions benefit citizens and do not cause harm."[51] They are in the process of developing Guidelines on the Ethical Development and Implementation of AI and intend to create AI Ethics Officers in government institutions to champion the Guidelines.

The Policy further recommends establishing a Responsible AI Office (RAI Office) which will be responsible for implementing the AI policy and participating in global AI governance fora such as

[48] Communiqué of the Republic of Rwanda, Office of the Prime Minister, 20 April 2023. (Accessible here.)
[49] Republic of Rwanda *The National AI Policy* (2023) at 1 (Accessible here.)
[50] Republic of Rwanda *The National AI Policy* (2023)at 4 (Accessible here.)
[51] Ibid at 5.

the OECD AI Policy Observatory and UNESCO. Although its scope is not yet fully defined, the establishment of this office may serve as an important accountability measure.

The policy is a commendable step and will likely springboard Rwanda's national response to AI.

Rwanda's data protection law[52] regulates the processing of personal data by automated means.[53] This means that the processing of personal data by AI has to comply with the data protection requirements prescribed by the law.

Further, it provides a degree of protection to data subjects by providing a right against automated decision-making.[54] However, unlike Kenya's law, there is no obligation to notify the data subject that such a decision has been taken, or the right to request the decision be reconsidered. This poses some practical difficulties for the operation of the right – it is impossible for a data subject to know whether their personal information has been used for automated processing, and accordingly whether such a decision has been made. Without understanding this, it is impossible for a data subject to enforce their right. This undermines the effectiveness of the right and accordingly diminishes the protection it may provide data subjects. Further, the law provides that the right does not apply if the data subject explicitly consents to it, if it is necessary for the performance of a contract, or if it authorised by law. These are common conditions that are used to narrow the scope of the right.

Tanzania

Country	Dedicated AI legislation	Data protection legislation addresses AI	Has a national AI strategy	Has a policy or draft policy on AI	Expert body on AI has been established
Tanzania	No	Yes	No	No	No

Tanzania does not regulate AI through a comprehensive law or have a national strategy or policy in place. The most recent strategic plan from the Ministry of Communication and Information Technology notes the ability of AI to assist in development but does not provide any comprehensive guidelines on its intended application or governance.[55] A degree of regulation is provided by Tanzania's recently implemented data protection law. Tanzania has accordingly been slow to implement instruments that govern AI. However, recent developments, such as the creation of the AI4D-Lab, which conducts multidisciplinary, responsible AI research, may prompt engagement in governance considerations.[56]

[52] Rwanda Law No. 058/2021 Relating to the Protection of Personal Data and Privacy ('Law no.058/2021'). (Accessible here.)
[53] Article 2 of Law No. 058/2021.
[54] Article 21 of Law No. 058/2021.
[55] Tanzania Ministry of Communication and Information Technology *Strategic Plan for the Period of 2021/22 – 2025/6*. (Accessible here.)
[56] More about Tanzania's AI4D-Lab can be found here.

Tanzania passed the Personal Data Protection Act 11 of 2022 which came into effect on 1 May 2023. The law applies to "any collection and processing of personal data performed wholly or partly by manual or automated means." [57] AI systems that process personal data will accordingly have to comply with the data protection requirements. The data protection principles are noted as follow:

> "A data controller or data processor shall ensure that personal data is-
> (a) processed lawfully, fairly, and transparently;
> (b) collected for explicit, specified, and legitimate purposes and not further processing in a manner incompatible with those purposes;
> (c) adequate, relevant, and limited to what is necessary in relation to the purposes for which it is processed;
> (d) accurate and where necessary, kept up to date, with every reasonable step taken to ensure that any inaccurate personal data is erased or rectified without delay;
> (e) stored in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data is processed;
> (f) processed in accordance with the rights of a data subject;
> (g) processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against any loss, destruction, or damage, using appropriate technical or organisational measures; and
> (h) not transferred abroad contrary to the provisions of this Act." [58]

The law provides data subjects with rights concerning automated decision making. Specifically, the right provides:

> "A data subject may, through the procedures prescribed in the regulations, require the data controller to ensure that any decision taken by or on behalf of the data controller which significantly affects data subject shall not [be] base[d[solely on the processing by automatic means." [59]

The right accordingly only applies in instances where the decision was based *solely* on automatic means. It is unclear whether it would apply in circumstances where a decision was made by an AI system, but then confirmed by a human.

Significantly, when such a decision has been made, the data controller is obligated to notify the data subject. [60] This requirement goes some way to ensure that a data subject is able to enforce their right. After receiving notification, a data subject may 'require' the data controller to reconsider the decision. The wording in this section is stronger than 'requesting' a reconsideration which is discretionary and appears to indicate that the data controller would be obligated to reconsider the decision.

[57] Section 22(1)(a) of the Personal Data Protection Act 11 of 2022.
[58] Section 5 of the Personal Data Protection Act 11 of 2022.
[59] Section 33(1)(c) of the Personal Data Protection Act 11 of 2022.
[60] Section 36(2)(a) of the Personal Data Protection Act 11 of 2022.

Similarly to other data protection laws in the region, the right does not apply if the decision is necessary for the execution or performance of a contract, it is authorised by law, or the data subject has provided consent. [61]

Tanzania's data protection law establishes the Personal Data Protection Commission, which is empowered to monitor compliance with the law and investigate complaints of non-compliance. This provides a degree of accountability by enabling data subjects to approach the Commission to ensure the automated processing of their data complies with the law.

Uganda

Country	Dedicated AI legislation	Data protection legislation addresses AI	Has a national AI strategy	Has a policy or draft policy on AI	Expert body on AI has been established
Uganda	No	Yes	Yes	No	Yes

Uganda's Ministry of ICT and National Guidance recently released the Digital Transformation Roadmap (the Roadmap). [62] In it, the Ministry acknowledged that the government has generally been slow to respond to emerging technologies [63] citing gaps in knowledge and the lack of a formalised approach as the reasons. The Roadmap includes a strong focus on AI, which was given almost no focus in Uganda's National 4IR Strategy, [64] marking a clear shift in priority.

The Roadmap notes the important role that innovative technologies can play in the achievement of Uganda's development goals and outlines certain enablers that should be implemented to do so. Some of the enablers concerning AI include to: [65]

- "**Develop a National AI Strategy** that will provide guidance on the social value, societal unity, and social impact arising from the use of artificial intelligence and other data-driven technologies."

- "**Invest in AI literacy and research** to empower people to effectively use and interact with AI systems, reduce digital divides, stimulate ethical AI development and further understanding of AI-related social, legal and ethical implications."

- "**Develop ethical framework** guides and self-assessment tools to help empower people to effectively use and interact with AI systems, reduce digital divides, stimulate ethical AI development and further understanding of AI-related social, legal and ethical implications."

[61] Section 36(3) of the Personal Data Protection Act 11 of 2022.
[62] Uganda Ministry of ICT and National Guidance *Digital Transformation Roadmap 2023/2024 – 2027/2028* (2023). (Accessible here.)
[63] Uganda Ministry of ICT and National Guidance *Digital Transformation Roadmap 2023/2024 – 2027/2028* (2023) at (Accessible here.)
[64] Uganda *Uganda's National 4IR Strategy* (Accessible here.)
[65] Uganda Ministry of ICT and National Guidance *Digital Transformation Roadmap 2023/2024 – 2027/2028* (2023) at 40. (Accessible here.)

- "**Establish a Data and AI Ethics Council** to act as a "steward" of the AI Ethics Principles and to co-ordinate independent research into best practices and standards for the ethical application of data and AI technologies to benefit society. This Data and AI Ethics Council should have representation from relevant academic and industry stakeholders and should engage with national, regional, and international expertise as needed."

The Roadmap clearly indicates Uganda's intention to formalise its approach to the governance of AI with a strong focus on ethical and responsible AI. The recommendation to establish a Data and AI Council is a welcomed move that will likely lead to increased accountability and participation.

Beyond Uganda's promising moves in the governance space, it has also had a data protection law in place since 2019. The law [66] applies to the processing of personal data by AI systems [67] and provides data subjects with a right against automated decision-making. [68] The right is similar to the one provided in most data protection laws, but includes two notable additions: first, it enables data subjects to ensure compliance by pre-emptively writing to the data controller to ensure decisions aren't taken solely by automatic means. [69] Second, it prescribes timeframes (21 days) within which the data controller must notify the data subject of how it has reconsidered the decision. [70] Legislated timeframes enable greater accountability, and can ensure data controllers actually respond to such requests. The law further ensures accountability by providing that a data subject can approach the Data Protection Authority if they are not satisfied with the decision. The Authority can order the data controller to comply with the right. [71]

Closing Commentary

Despite the significant strides African countries have taken to govern AI, more is required to ensure the ethical and safe deployment of AI across the continent. This concludes Part 1 of this toolkit. Part 2 will unpack the trends in global governance and Part 3 will explore advocacy approaches.

[66] The Data Protection and Privacy Act, 2019. (Accessible here.)
[67] Section 1(a) read with the definition of "processing" in section 2 of the Data Protection and Privacy Act, 2019.
[68] Section 27 of the Data Protection and Privacy Act, 2019.
[69] Section 27(1) of the Data Protection and Privacy Act, 2019.
[70] Section 27(3) of the Data Protection and Privacy Act, 2019.
[71] Section 27(6) of the Data Protection and Privacy Act, 2019.

PART 2
INTERNATIONAL FRAMEWORKS FOR AI GOVERNANCE

THOMSON REUTERS FOUNDATION

www.trust.org

INTRODUCTION

Countries across the world are scrambling to regulate artificial intelligence.

In Part 1 of this toolkit, we focused on the state of regulatory efforts in Africa. Now, in Part 2 we zoom out from the African continent and provide an overview of notable developments in the global sphere. We do this in three parts. First, we examine the social and political context in which AI governance is being created. We then highlight the different kinds of regulations and other instruments currently being developed, before focusing on a few globally influential regions. Finally, we discuss the key trends and themes that emerge across different regions.

Social and Political Context

There are several factors currently influencing the creation of international AI governance.

- **Asymmetric AI development:** It costs hundreds of millions of dollars to train cutting-edge AI models, due to the immense amounts of computing power, data, and other raw materials required.[1] This limit who can create these models. At present AI development is primarily being driven by private companies (like OpenAI, Meta, and Anthropic) rooted in the United States, with Chinese labs, also developing their own impressive models, in second place.[2] This gives disproportionate influence to the countries in which major AI labs reside, as these labs are bound first and foremost by national regulation. It also suggests that the AI labs themselves have significant influence as, through self-regulation and other industry measures, they may act as de-facto regulators in this space, influencing the behaviour of the rest of the world.

- **Global power dynamics**: AI systems have clear military applications, from use in autonomous weaponry to advanced surveillance and intelligence operations. They also have the potential to be incredibly useful in enhancing a state's control over its people, which may be particularly appealing to more authoritarian regimes; and to enhance economic productivity more generally. Thus, some states may aim to advance this technology as quickly as possible, and to prioritise their own national security objectives, rather than to support collaborative, globally harmonised regulatory frameworks.

- **Pre-existing regulatory cultures**: The process by which policies are made, as well as the institutional arrangements supporting different policies, vary across the world in line with the differing legal traditions and social priorities of different countries. The European Union, for example, is known for its more precautionary approach and its comprehensive regulatory frameworks, such as the General Data Protection Regulation (GDPR), while the United States' approach is typically more sector-specific, federal, and tilted in favour of

[1] Knight, "OpenAI's CEO Says the Age of Giant AI Models Is Already Over" *Wired* (17 April 2023) (Accessible here).
[2] Ding and Xiao, "Recent Trends in China's Large Language Model Landscape" *Centre for the Governance of AI* (April 2023) (Accessible here).

innovation instead of caution. As we shall see below, these pre-existing cultural differences represent different starting points from which states craft their AI regulation.[3]

- **Wave of public interest:** Since the release of ChatGPT – and its associated 'hype cycle' – there has been considerable appetite across the world for more comprehensive AI governance. Most of the CEOs of the major AI companies have been supporting this call for regulation, although the precise form they believe it should take remains ambiguous.[4]

- **Role of civil society and advocacy groups:** Efforts from international civil society and related groups can be divided into two broad camps – those primarily concerned with AI risks related to misinformation, prejudice, and copyright ("AI ethics"),[5] and those concerned with AI risks related to catastrophic harm ("AI safety").[6] These groups diverge in their beliefs about what is most important, although they overlap considerably in the steps they believe need to be taken to reduce risk from AI (for example, in their calls for AI models to be transparent, accountable, and subject to third-party safety audits). Both groups wield considerable influence, with the former being more influential among human rights and traditional civil society groups, and the latter carrying more weight amongst technical researchers and the AI labs themselves. The interests and arguments of these groups thus exert considerable influence on the regulatory environment.

Governance Efforts

As in many areas of technology regulation, the regulation of AI faces what is sometimes called the "Collingridge dilemma": if one creates regulation before the impacts of a technology are clear, that regulation may not function as intended (for example, it may stifle innovation without creating commensurate benefits); but by the time its impacts are clear, the technology may be too entrenched to regulate effectively – or at least, its regulation will become more challenging over time, as it becomes embedded in our everyday lives.[7] We have seen such dynamics at play in relation to social media over the last fifteen years. The challenge is even more acute in relation to AI, as the technology is advancing rapidly, and it is difficult to predict what capabilities will emerge from these systems over the next decade.

Types of governance

In recent years, countries across the world have been responding to this regulatory challenge in different ways. As regulation is rapidly evolving in this area, it can be useful to think about this by

[3] Engler, "The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment", *Brookings Institute* (April 2023) (Accessible here).
[4] See for example Rozen, "AI Leaders Are Calling for More Regulation of the Tech. Here's What That May Mean in the US" *The Washington Post* (27 July 2023) (Accessible here).
[5] This is typified by the November 2021 UNESCO "Recommendations on the Ethics of Artificial Intelligence", accessible here.
[6] A list of scientists, academics, policymakers, industry professionals, and other notable figures who hold this view can be found here.
[7] More information on the Collingridge Dilemma is available here.

reference to the different types of regulatory efforts currently underway – only a small portion of which are binding.

There are four dimensions we can use to categorise these efforts: area of focus, jurisdiction, stakeholder involvement, and type of document. Let's say a little more about each dimension, before turning to some of the key regulatory efforts in this area.

There are three key components in a modern AI system: the computing power necessary to train and run the system, the data used to train the system, and the model architecture and machine-learning algorithms that produce the system's intelligent behaviour.[8] Different regulatory efforts aim to regulate components, with some focusing on data processing and protection, some on algorithmic bias and transparency, and others on the regulation of the computing power itself. As we shall see, many regulations cut across these different categories, or try to regulate AI systems in general.

Another way to understand the areas of focus is through the different analytic lenses used to craft regulation. Here we can distinguish between regulation that is rooted in human rights and social equity, that focuses on data protection and privacy, that focuses on promoting innovation, or that focuses on liability and accountability. These different lenses identify different risks as being the most salient, which implicates the kinds of regulation they produce.

In terms of jurisdiction, regulation is being made at the global, regional, and national levels. And in terms of stakeholder involvement, we can distinguish government-led, industry-led, and multi-stakeholder processes.

The different types of governance in this area can be thought of on a spectrum from least to most binding. In terms of least binding, we have discussion documents or "white papers", such as the World Economic Forum's 2019 "Framework for Developing a National Artificial Intelligence Strategy".[9] Next are guidelines, declarations of principle, and related soft law instruments. These are frequently made by stakeholders who do not necessarily have regulatory enforcement powers but can be an important form of soft law, influencing future regulation. Examples here include the OECD's 2019 "AI Principles",[10] UNESCO's 2021 "Recommendations on the Ethics of Artificial Intelligence",[11] the 2021 Hyderabad Declaration on AI and Digital Wellness, with which

[8] On these components, see OpenAI, "AI and Compute" (May 2018) (Accessible here). Note also that Machine learning algorithms and model architecture are technically distinct – while algorithms are statistical techniques used for performing certain tasks, architecture in the context of neural networks refers to the arrangement of neurons and layers, and the connections between them – it is what the algorithms are "run" on.
[9] World Economic Forum, "A Framework for Developing a National Artificial Intelligence Strategy" (August 2019) (Accessible here).
[10] Organisation for Economic Cooperation and Development, "Recommendation of the Council on Artificial Intelligence" (May 2019) (Accessible here).
[11] UNESCO "Recommendations on the Ethics of Artificial Intelligence" (November 2021) (Accessible here.)

South Africa, Kenya, and Uganda were involved,[12] and the 2022 Global Partnership on Artificial Intelligence's "Minister's Declaration".[13]

Further along are national policy papers and strategy documents, such as the United Kingdom's 2023 policy paper "A pro-innovation approach to AI regulation",[14] which lays out the government's plan for responding to the regulatory challenges by AI, which is to provide for principles that should guide the pre-existing UK regulators working in this space. The UK government is also considering "introducing a statutory duty on regulators requiring them to have due regard to the principles", although they have yet to do this.[15]

Another form of regulation is a voluntary commitment made by relevant stakeholders (notably AI companies). The best example of these right now are the voluntary commitments secured by the United States' Biden administration from seven major AI companies - Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI – to manage AI risk, particularly in relation to safety, security, and trust.[16] While these commitments (discussed in further detail below) are impressive, given that they are voluntary, they are by themselves unenforceable, which potentially limits their effectiveness.

National or regional AI legislation would be more binding, but at present no notable dedicated AI legislation exists at either the national or international levels. The European Union's AI Act (discussed below) is expected to be finalised by the end of 2023, after which efforts will turn to its implementation.[17] On the international level, while the United Nations Secretary-General recently noted the need to "urgently confront the new reality of generative and other artificial intelligence",[18] there is at present no dedicated AI legislation being contemplated by the UN – although they are in the process of establishing a 'Multistakeholder Body on Artificial Intelligence'.[19]

It is also worth noting that multistakeholder forums – such as the Global Summit on AI Safety, being convened in Britain in November 2023 – can play an important role in producing the above-mentioned declarations, principles, and related soft law instruments, rendering them another source from which AI regulation flows.[20]

---

[12] University of Hyderabad, "Hyderabad Declaration on Artificial Intelligence (AI) and Digital Wellness (DW) 2021" (2021) (Accessible here).
[13] The Global Partnership on Artificial Intelligence, "GPAI Ministers' Declaration 2022", (2022) (Accessible here).
[14] United Kingdom Department for Science, Innovation, and Technology, "A pro-innovation approach to AI regulation" (2023) (Accessible here).
[15] See above at page 6.
[16] The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI" (July 2023) (Accessible here).
[17] European Parliament, "EU AI Act: first regulation on artificial intelligence" (June 2023) (Accessible here). For the full text of the 2021 EU AI Act, see here. The 2023 amendments can be accessed here.
[18] United Nations Press, "International Community Must Urgently Confront New Reality of Generative, Artificial Intelligence, Speakers Stress as Security Council Debates Risks, Rewards", (July 2023) (Accessible here).
[19] United Nations Office of the Secretary-General's Envoy on Technology. "Multistakeholder Advisory Body on Artificial Intelligence" (2023) (Accessible here).
[20] United Kingdom Government Press Release, "UK to host first global summit on Artificial Intelligence" (June 2023) (Accessible here).

Regional Approaches

Given that regulation is quickly evolving across the world, this section looks at the high-level approach being taken by different significant global powers.

European Union

Although the European Union (EU) already has some related legislation that implicates the regulation of AI – notably the GDPR, the Digital Services Act, and the Digital Markets Act[21] –the centrepiece of its regulatory efforts is its proposed AI Act. The Act aims to provide a harmonised legal framework for the development and use (within the EU) of AI systems.[22] To this end, it takes a risk-based approach, carving out four levels of risk AI systems may pose, and creating different requirements and obligations for each one.

- **Unacceptable risk:** Certain practices are deemed unacceptably risky, and thus are prohibited within the EU market. These include systems that employ harmful manipulative 'subliminal techniques', systems used by public authorities for social scoring, and real-time remote biometric identification systems, such as facial recognition.

- **High risk:** AI systems that are either used as a safety component in products falling under the EU's health and safety legislation, or which are deployed in various specified areas (such as education, migration, law enforcement, or the management of infrastructure).[23] are designated as 'high risk'. They would have to be registered in an EU-wide database managed by the European Commission and would have to comply with a range of measures related to testing, data governance, transparency, human oversight, and cybersecurity.

- **Limited risk**: Systems that interact with humans (such as chatbots), as well as systems that generate audio, visual, and other types of content are deemed to be low-risk and are only subject to limited transparency obligations (such as the requirement to disclose themselves to affected persons).

- **Low or minimal risk:** All other AI systems considered to pose low or minimal risk are not bound by any obligations, although the Act envisions the creation of codes of conduct to encourage the AI labs that develop them to voluntarily abide by the measures required of high-risk systems.

---

[21] European Union, "The Digital Services Act package"(2023) (Accessible here).
[22] European Parliament, "EU AI Act: first regulation on artificial intelligence" (June 2023) (Accessible here). For the full text of the 2021 EU AI Act, see here. The 2023 amendments can be accessed here.
[23] The full list of specified areas is as follows: Biometric identification and categorisation of natural persons; management and operation of critical infrastructure; education and vocational training; employment, worker management and access to self-employment; access to and enjoyment of essential private services and public services and benefits; law enforcement; migration, asylum, and border control management; assistance in legal interpretation and application of the law. See reference above.

Practically, the Act would require all member states to designate one or more competent authorities to oversee the Act's operation – including a national supervisory authority, to supervise the application of the regulation, and a national market surveillance authority, to assess AI providers' compliance with the requirements relevant to high-risk systems. These authorities would have access to confidential information (such as the source code of the relevant systems), and the power to impose corrective measures for non-compliance – including a lofty fine of €30 000 or 6% of a company's global annual turnover.

After the European Parliament agreed to a position on the AI Act in June 2023, it will now be negotiated between EU member states and the European Commission, which is expected to take until the end of the year.

The Act has been met with strong opposition from the global business community, on the basis that it could jeopardise Europe's competitiveness and technological sovereignty without necessarily solving the relevant AI-related challenges.

The United States

In contrast to the European Union's centrally coordinated and expansive approach, the United States' approach is at this stage more piecemeal, sector-specific, and distributed across various federal agencies; although both regions take a risk-based approach.[24] For example, the US Copyright Office has issued rulings that suggest that most text, images, and videos created by AI systems cannot be copyrighted as original works; while an Algorithmic Accountability Act has been proposed, which would require companies to evaluate the bias and effectiveness of their AI systems, and would require the country's Federal Trade Commission to enforce this requirement. [25]

A 2019 Executive Order ("Maintaining American Leadership in Artificial Intelligence") from the Trump administration required the country's various federal agencies to develop plans to regulate AI applications; however, by December 2022, only one of the 41 major agencies (the Department of Health and Human Services) had meaningfully created such a plan.[26] Rather than pursue the implementation of this order, the Biden administration took a different approach to AI risk in the form of their 2022 AI 'Blueprint for an AI Bill of Rights', a non-binding document that sets out five principles[27] and associated practices to guide the development and use of AI. It tasks different federal agencies, responsible for different sectors (like health, labour, and

[24]Engler, "The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment", *Brookings Institute* (April 2023) (Accessible here).
[25] Piper, "There are two factions working to prevent AI dangers. Here's why they're deeply divided", *Vox* (August 2022) (Accessible here).
[26] Engler, "The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment", *Brookings Institute* (April 2023) (Accessible here).
[27] The principles include: safe and effective systems, algorithmic discrimination protections, data privacy, notice and explanation an human alternatives, consideration, and fall back. More about these can be accessed here.

education) with implementation.[28] Responses from federal agencies has been highly uneven, although some, like the Food and Drug Administration and the Consumer Financial Protection Bureau, have made some strides in incorporating AI considerations into their regulatory frameworks.[29]

Discussion is ongoing around questions such as whether there is a need to create a new dedicated federal AI agency. In January 2023, the country's National Institute of Standards and Technology released its 'Artificial Intelligence Risk Management Framework', which is "designed to equip organizations and individual [...]with approaches that increase the trustworthiness of AI systems, and to help foster the responsible design, development, deployment, and use of AI systems over time."[30] In tandem with these efforts, the country is taking steps to "advance the research, development, and deployment of responsible artificial intelligence", as evidenced for example by its "National Artificial Intelligence Strategic Development Plan". The proliferation of these documents alternatively focusing on governance and on fostering innovation is representative of the contemporary American approach.

China

In July 2023, China published its "Interim Measures for the Management of Generative AI Services", which provide a window into how the country is approaching AI regulation.[31] They hit many of the same topics as in the rest of world (discussed further below), although they are notable for their authoritarian slant. Thus, in addition to discussions on the need to promote fairness, transparency, and international cooperation, one finds the following clause:[32]

> "The provision and use of generative artificial intelligence services shall abide by laws and administrative regulations, respect social morality and ethics, and abide by the following provisions:
> (1) Adhere to the core values of socialism, and must not generate incitement to subvert state power, overthrow the socialist system, endanger national security and interests, damage national image, incite secession, undermine national unity and social stability, promote terrorism, extremism..."

AI governance in China is being led by the Cyberspace Administration of China, the "clear bureaucratic leader in governance to date" – although this may change as the scope of AI expands.[33] The public/private distinction is also much less clear in China, as "the Chinese

---

[28] The White House, "Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People" (October 2022) (Accessible here).
[29] Engler, "The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment", *Brookings Institute* (April 2023) (Accessible here).
[30] National Institute of Standards and Technology "Artificial Intelligence Risk Management Framework" (January 2023) (Accessible here).
[31] Cyberspace Administration of China, "Interim Measures for the Management of Generative Artificial Intelligence Services" (July 2023) (Accessible here).
[32] Cyberspace Administration of China, "Interim Measures for the Management of Generative Artificial Intelligence Services" (July 2023) (Accessible here) at article 4(1).
[33] Sheehan, "China's AI Regulations and How They Get Made" *Carnegie Endowment for International Peace* (July 2023) (Accessible here).

government plays a much more prominent role in China's AI ecosystem, often directly facilitating industry-academia cooperation and providing significant compute funding."[34]

Other notable Chinese AI regulations include its 2021 Regulation on Recommendation Algorithms and its 2022 Rules for Deep Synthesis (synthetically generated content).[35]

Trends and Themes

As we have seen above, there is considerable variation in how countries intend to implement their various commitments in relation to AI, and uncertainty remains for most countries about what institutional architecture is most appropriate. Despite this practical divergence and uncertainty, there is a surprisingly high degree of consensus on what the major issues in this space are. One can imagine that almost every country has identified virtually the same basic ingredients, although they are creating distinct dishes. This section serves to highlight these different ingredients. Note that this overview is intended to be illustrative, rather than comprehensive – and that the same concepts are sometimes used with different meanings in different regions.

Transparency and explainability

The idea that AI systems must be transparent about how they work, and that the decisions and actions taken by AI systems must be explainable crops up across the AI regulatory landscape (even though in practice, explaining why an AI system makes a certain decision can be incredibly challenging, as even the creators of these systems frequently refer to them as "black boxes").[36] Transparency is seen as one way to guard against risks from misinformation – although it is likely on its own insufficient.

For example, the 2021 version of the EU AI Act states: [37]

> "High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately."

America's Blueprint for an AI Bill of Rights includes "Notice and Explanation" as one of its principles, noting:[38]

> "You should know that an automated system is being used and understand how and why it contributes to outcomes that impact you. Designers, developers, and deployers of

[34] Ding and Xiao, "Recent Trends in China's Large Language Model Landscape" *Centre for the Governance of AI* (April 2023) at 3. (Accessible here).
[35] Sheehan, "China's AI Regulations and How They Get Made" *Carnegie Endowment for International Peace* (July 2023) at 4 (Accessible here).
[36] Xiang, "Scientists Increasingly Can't Explain How AI Works" *Vice* (November 2022) (Accessible here).
[37] Article 13(1).
[38] The White House, "Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People" (October 2022) at 6 (Accessible here).

> automated systems should provide generally accessible plain language documentation including clear descriptions of the overall system functioning and the role automation plays, notice that such systems are in use, the individual or organization responsible for the system, and explanations of outcomes that are clear, timely, and accessible".

China's interim measures also address the topic, requiring AI service providers to "take effective measures to improve the transparency of generative artificial intelligence services and improve the accuracy and reliability of generated content."[39] Even the UK's "pro-innovation approach" policy paper is guided by this principle, noting that "AI systems should be appropriately transparent and explainable."[40]

AI companies have also committed to transparency-related actions, to foster public trust:[41]

> "The companies commit to developing robust technical mechanisms to ensure that users know when content is AI generated, such as a watermarking system. This action enables creativity with AI to flourish but reduces the dangers of fraud and deception."

Algorithmic discrimination

Another universal concern is that AI systems reproduce and perpetuate the biases latent in the data they are trained on. This problem is known as algorithmic bias or discrimination. As per the US Blueprint:

> "Algorithmic discrimination occurs when automated systems contribute to unjustified different treatment or impacts disfavoring people based on their race, color, ethnicity, sex (including pregnancy, childbirth, and related medical conditions, gender identity, intersex status, and sexual orientation), religion, age, national origin, disability, veteran status, genetic information, or any other classification protected by law."

It is widely addressed across the different types of AI regulation. UNESCO's recommendations, for example, note that:[42]

> "AI actors should make all reasonable efforts to minimize and avoid reinforcing or perpetuating discriminatory or biased applications and outcomes throughout the life cycle of the AI system to ensure fairness of such systems."

China's interim measures state that:[43]

---

[39] Cyberspace Administration of China, "Interim Measures for the Management of Generative Artificial Intelligence Services" (July 2023) at article 4(5) (Accessible here).
[40] United Kingdom Department for Science, Innovation, and Technology, "A pro-innovation approach to AI regulation" (2023) at 28. (Accessible here).
[41] The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI" (July 2023) (Accessible here).
[42] UNESCO "Recommendations on the Ethics of Artificial Intelligence" (November 2021) at 20. (Accessible here.)
[43] Cyberspace Administration of China, "Interim Measures for the Management of Generative Artificial Intelligence Services" (July 2023) (Accessible here).

> "In the process of algorithm design, training data selection, model generation and optimization, and service provision, take effective measures to prevent discrimination based on ethnicity, belief, country, region, gender, age, occupation, health, etc."

America's Blueprint states that "You should not face discrimination by algorithms and systems should be used and designed in an equitable way".[44]

The United Kingdom's policy paper ties this to the broader principle of fairness, declaring:[45]

> "AI systems should not undermine the legal rights of individuals or organisations, discriminate unfairly against individuals or create unfair market outcomes."

Safety and security

The idea that AI systems must be safe and secure finds expression across regulatory instruments, both as a general principle and in specific measures designed to enhance safety and security – particularly in the form of both internal and external audits of AI systems. For example, the US Blueprint discusses the importance of pre-deployment tests:[46]

> "Systems should undergo pre-deployment testing, risk identification and mitigation, and ongoing monitoring that demonstrate they are safe and effective based on their intended use, mitigation of unsafe outcomes including those beyond the intended use, and adherence to domain-specific standards. Outcomes of these protective measures should include the possibility of not deploying the system or removing a system from use."

Safety and security also find expression in the voluntary commitment the AI companies themselves have made:[47]

> "The companies commit to internal and external security testing of their AI systems before their release. This testing, which will be carried out in part by independent experts, guards against some of the most significant sources of AI risks, such as biosecurity and cybersecurity, as well as its broader societal effects."

> "The companies commit to investing in cybersecurity and insider threat safeguards to protect proprietary and unreleased model weights. These model weights are the most essential part of an AI system, and the companies agree that it is vital that the model weights be released only when intended and when security risks are considered. The companies commit to facilitating third-party discovery and reporting of vulnerabilities in their AI systems. Some

[44]The White House, "Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People" (October 2022) (Accessible here).
[45]United Kingdom Department for Science, Innovation, and Technology, "A pro-innovation approach to AI regulation" (2023) at 29. (Accessible here).
[46] The White House, "Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People" (October 2022) (Accessible here).
[47] The White House, "FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI" (July 2023) (Accessible here).

> issues may persist even after an AI system is released and a robust reporting mechanism enables them to be found and fixed quickly."

China's interim measures state:[48]

> "Providers of generative artificial intelligence services with public opinion attributes or social mobilization capabilities shall conduct security assessments in accordance with relevant national regulations, and perform algorithm filing, modification, and cancellation filing procedures in accordance with the "Internet Information Service Algorithm Recommendation Management Regulations".

Safety and security is also an important principle for the UK:[49]

> "Regulators may need to introduce measures for regulated entities to ensure that AI systems are technically secure and function reliably as intended throughout their entire life cycle."

Data Privacy

Data privacy is a critical topic that is widely addressed by existing regulatory efforts. The UNESCO recommendations note that:[50]

> "Privacy, a right essential to the protection of human dignity, human autonomy and human agency, must be respected, protected and promoted throughout the life cycle of AI systems... Algorithmic systems require adequate privacy impact assessments,[51] which also include societal and ethical considerations of their use and an innovative use of the privacy by design approach."

The US Blueprint also emphasises privacy, arguing:[52]

> "You should be protected from abusive data practices via built-in protections and you should have agency over how data about you is used. You should be protected from violations of privacy through design choices that ensure such protections are included by default, including ensuring that data collection conforms to reasonable expectations and that only data strictly necessary for the specific context is collected."

Privacy also find expression in the Chinese interim measures, albeit with a different emphasis:[53]

---

[48] Cyberspace Administration of China, "Interim Measures for the Management of Generative Artificial Intelligence Services" (July 2023) (Accessible here).
[49] United Kingdom Department for Science, Innovation, and Technology, "A pro-innovation approach to AI regulation" (2023) at 27. (Accessible here).
[50] UNESCO "Recommendations on the Ethics of Artificial Intelligence" (November 2021). (Accessible here.)
[51] Impact assessments can also be used in the context of AI. For more on this see here.
[52] The White House, "Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People" (October 2022) (Accessible here).
[53] Cyberspace Administration of China, "Interim Measures for the Management of Generative Artificial Intelligence Services" (July 2023) (Accessible here).

> "Relevant institutions and personnel involved in the safety assessment and supervision and inspection of generative artificial intelligence services shall keep state secrets, commercial secrets, personal privacy and personal information known in the performance of their duties confidential in accordance with the law, and shall not disclose or illegally provide them to others."

Human oversight and accountability

It is widely accepted that humans must remain "in the loop" and in control of the AI systems they create. Since AI systems cannot be held liable in their own right, it is vital that regulations identify mechanisms by which developers can be held accountable for any liabilities incurred by their systems. The question of AI liability is a complex and ultimately unsettled one in the international space.

Human oversight is a principle in the UNESCO recommendations, which state:[54]

> "Member States should ensure that it is always possible to attribute ethical and legal responsibility for any stage of the life cycle of AI systems, as well as in cases of remedy related to AI systems, to physical persons or to existing legal entities. Human oversight refers thus not only to individual human oversight, but to inclusive public oversight, as appropriate."

'Governance and accountability' is also a principle in the UK's policy paper, where they state:[55]

> "Governance measures should be in place to ensure effective oversight of the supply and use of AI systems, with clear lines of accountability established across the AI life cycle… Regulators will need to look for ways to ensure that clear expectations for regulatory compliance and good practice are placed on appropriate actors in the AI supply chain, and may need to encourage the use of governance procedures that reliably ensure these expectations are met."

Meanwhile, the EU's proposed AI Act aims to address the question of accountability comprehensively, creating a host of obligations that apply to various humans and are backed by threat of penalties. These are in addition to all the preexisting ways that AI service providers can be held accountable in both criminal and civil law.

Licensing

Another common proposal is for AI service providers to have to register their models with some public authority. We saw this in relation to the EU AI Act, above. It also finds expression in China's interim measures:[56]

---

[54]UNESCO "Recommendations on the Ethics of Artificial Intelligence" (November 2021) at 22. (Accessible here.)
[55] United Kingdom Department for Science, Innovation, and Technology, "A pro-innovation approach to AI regulation" (2023)at 30. (Accessible here).
[56] Cyberspace Administration of China, "Interim Measures for the Management of Generative Artificial Intelligence Services" (July 2023). (Accessible here).

> "Where laws and administrative regulations stipulate that the provision of generative artificial intelligence services shall obtain relevant administrative licenses, the providers shall obtain licenses in accordance with the law."

Licensing is one way that public authorities might control the development of AI systems. Outside of Europe, Which institutions might have the authority to administer such licenses remains an open question.

Miscellaneous issues

While the above analysis has provided a broad overview of relevant issues in this space, it is by no means exhaustive. For example, there are open questions around how AI systems interact with intellectual property law, which are currently being litigated in courts across the world. Many frameworks also speak to the need for international and multi-stakeholder collaboration, and for globally agreed interoperable standards.

There is also a focus on the ethical development and use of AI more generally, and on grappling with the effects of AI on the labour market. And, as noted above, for those in the AI safety community, the focus of regulation should be on curtailing the catastrophic risks that might arise once AI systems significantly more advanced than the current crop are developed.[57] This worldview focuses on different parts of AI – such as the regulation of raw computing power.

For others, the environmental impacts of AI are also important to consider. And more broadly, there are specific issues relevant to virtually every economic sector – for example, how AI should be governed in the context of educational institutions or the provision of healthcare.

The age of generative AI has also spurred concerns around the proliferation of disinformation and misinformation. AI-generated deepfake technology makes it increasingly difficult to discern real from fabricated content, raising fears of its potential for political manipulation and misinformation campaigns.

Closing Commentary

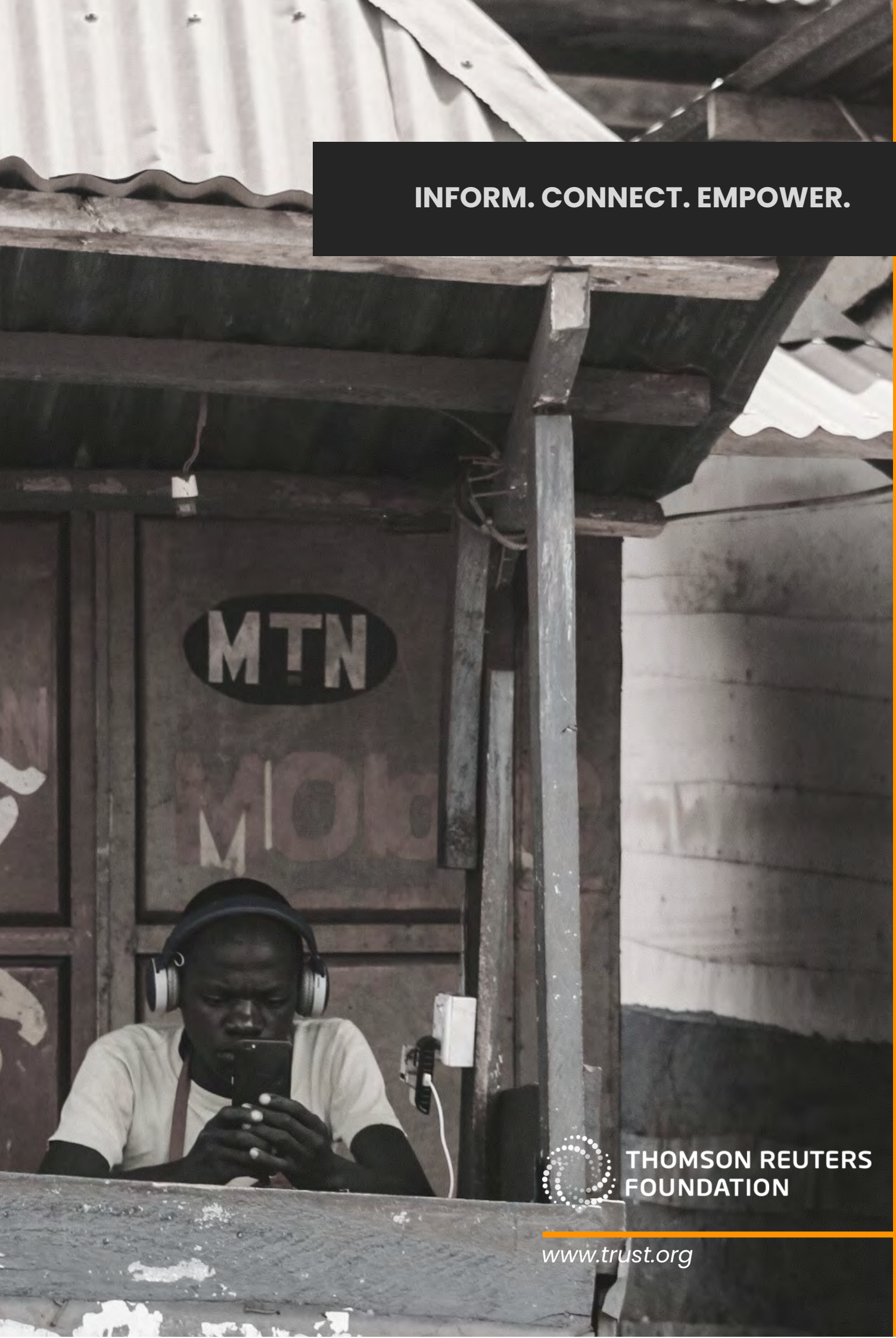
It is interesting to notice how existing public interest concepts like transparency and accountability are being adapted to support the governance of AI. Readers should consider to what extent these adaptions have been successful, and to what extent there are gaps in current frameworks. Although there are already an abundance of frameworks, declarations, and related regulatory instruments, we are still in the very early days of AI governance. New regulations are being developed and propagated monthly, often displacing, or overwriting what came before.

[57] This notion of extremely advanced AI is often captured in the concept of "artificial general intelligence", or AGI. Although this may sound like a fringe concern, the mission of most of the major AI labs involves AGI – for example, OpenAI's stated mission is to "ensure that artificial general intelligence benefits all of humanity" – see OpenAI, "About" (Accessible here).

A close look at existing documents reveals that while there is relative consensus on which values and principles are most important in this space, there is considerable divergence on how those values and principles should be operationalised – with Europe, America, and China all taking distinct approaches. It remains to be seen which approach is most effective – and indeed, the most appropriate approach for a given region may largely depend on its local context.

In the face of all this regulatory uncertainty and development, there is a huge opportunity for civil society and journalists to exert influence on the future of AI governance. Given the complexity of this space, it is thus also important for civil society and related actors to understand what outcomes most desirable, and what methods are might achieve them.

In the Part 3 of this toolkit, we explore approaches to advocacy.



INFORM. CONNECT. EMPOWER.

THOMSON REUTERS FOUNDATION

www.trust.org