



AI
Assurance
Club

A Global Digital Foundation Initiative

AI Governance and Assurance

Global Trends 2023-24



Contents

Executive Summary	2
Introduction	3
The Global Landscape	4
United Nations	5
OECD	6
Global Partnership on AI	6
Hiroshima AI Process	7
AI Safety Summit	7
International Trade	7
International Standardisation	8
Regional Landscapes	11
Africa	12
Asia and Asia Pacific	14
Europe and Central Asia	16
Latin America and the Caribbean	18
The Middle East	20
North America	22

This report was written by Catriona Gray and Dr Robert H Wortham, March 2024

Correspondence: catriona.gray@globaldigitalfoundation.org

This paper represents the independent opinions of its authors. It has been prepared following independent scholarly research and consultation with the AI Assurance Club.

Executive Summary

We are at a watershed moment in AI governance. Across the globe, government and industry actors, along with citizens, are weighing up the options available. Many different tools are being developed to help take advantage of the promises of AI while minimising its risks. These range from formal regulation through new AI laws or increased powers for sector based regulators, to assurance processes such as voluntary standards conformity and certification schemes. This report surveys the global AI governance landscape, identifying six trends:

1. While the European Union is set to enact AI-specific regulation, it is far from guaranteed that the AI Act will influence global norms.
2. There is considerable convergence around OECD definitions and benchmarks, and nominally risk-based regimes.
3. Large generative models have taken centre stage, but this has also laid bare the difficulties of contending with concentrated market power.
4. Foundational AI process standards have now been published, but further expertise and capacity to implement are still needed.
5. There have been important shifts in US public policy towards a more interventionist approach, and this will have wider global effects.
6. Geopolitical rifts are now more visible, with more countries pressured to position themselves in relation to the US and China.

Introduction

Across the globe, the development and use of artificial intelligence (AI) has accelerated. Policymakers, businesses, and other stakeholders are trying to develop strategies to unleash the full potential of AI technologies while addressing the risks they pose. Yet, divergent approaches are emerging. In the European Union, AI-specific regulation will come into force within the next couple of years. Elsewhere, the burgeoning AI governance landscape is made up of various voluntary commitments and standards as well as sectoral rules.

As the legal scholar Anu Bradford puts it in *Digital Empires*, the digital society is at an inflection point:

“The cascade of regulation that is being drafted, and that will be drafted in the coming years, will be crucial in shaping the digital economy and digital society for years to come.”

At the centre of the AI governance puzzle is the issue of *trust*. As AI systems become more embedded in our lives – in infrastructure, business processes, public services, or in consumer goods – users need to have confidence that systems will consistently operate as intended, without causing harm. This requires not only subjective confidence but also some mechanism to verify and validate the factors contributing to that confidence. We increasingly see this need for trust in AI development and use addressed by governments throughout the world. However, these Government initiatives vary widely, ranging from voluntary codes of practice through to hard regulation. In this report we cover these initiatives from two perspectives. Firstly, we describe the latest initiatives from global organisations such as the United Nations and OECD. Secondly, we look at regional and selected national progress around the world.

The Global Landscape

United Nations

Recent years have seen growing recognition of the need for multilateral cooperation on AI governance. In October 2023, the United Nations Secretary-General António Guterres announced the creation of a High-level Advisory Body on AI to analyse and advance recommendations for the international governance of AI. The Advisory Body will publish its final report ahead of the Summit of the Future to be held in 2024.

An interim report, Governing AI for Humanity, was published in December 2023. It identified seven key AI governance functions, each requiring international cooperation among actors:

1. Regularly assess the future directions and implications of AI.
2. Reinforce interoperability of governance efforts emerging around the world and their grounding in international norms through a Global AI Governance Framework endorsed in a universal setting (UN).
3. Develop and harmonise standards, safety, and risk management frameworks.
4. Facilitate development, deployment, and use of AI for economic and societal benefit through international multi-stakeholder cooperation.
5. Promote international collaboration on talent development, access to compute infrastructure, building of diverse high-quality datasets, responsible sharing of open-source models, and AI-enabled public goods for the Sustainable Development Goals (SDGs).
6. Monitor risks, report incidents, coordinate emergency responses.
7. Compliance and accountability based on norms.

Implementation of the UNESCO Recommendation on the Ethics of AI, adopted in November 2021, has continued. The Recommendation mandated the development of two key tools: the Readiness Assessment Methodology (RAM) and the Ethical Impact Assessment (EIA), both published in August 2023. These tools are designed to support those who procure and adopt AI to better align their decisions with ethical principles, and to help governments assess how robust their laws, policies and institutions are in addressing AI risks.

More recently, on 21 March 2024, the United Nations General Assembly adopted its first resolution on the promotion of safe, secure, and trustworthy AI.

OECD

The Organisation for Economic Co-operation and Development (OECD) continues to be a key forum for the development of AI governance and assurance. The OECD AI Policy Observatory provides research and analysis that guides the work of governments and policymakers. For example, its definition of AI systems, updated in November 2023, will be used to shape ongoing work in Brazil, the EU, the US, and elsewhere. Another main output in 2023 was the Catalogue of Tools and Metrics for Trustworthy AI. Entries include almost 700 tools and 100 metrics, including training and guidelines, standards, and software code. In November, the OECD also launched the AI Incidents Monitor (AIM) which will inform work towards a common incident reporting framework for AI, and provide real-time evidence for policymakers and regulators.

Global Partnership on AI

Global Partnership on AI (GPAI) is a multistakeholder initiative for sharing research and identifying key issues among AI practitioners, built around a shared commitment to the OECD Recommendation on Artificial Intelligence. It aims to bridge the gap between theory and practise through the sharing of multidisciplinary research and the work of its working groups. Its Multistakeholder

Expert Group (MEG) published its annual report in November 2023, reporting on activities and priorities, including providing expert support to governments under the Hiroshima Process, and analysis of the implications of generative AI.

Hiroshima AI Process

In April 2023, Japan hosted a meeting of the G7 Digital and Technology Ministers meeting. Leaders then agreed to prioritise collaboration for inclusive governance of the most advanced AI technologies. In October 2023, they agreed on the Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems, and the Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems. Both the Principles and the Code of Conduct will provide guidance for organisations developing and deploying the most advanced AI systems, including foundation models and generative AI systems.

On 15 March 2024, G7 governments agreed to collaborate further on innovation and technology adoption, including the creation of a joint report looking in detail at the factors involved in AI uptake among businesses.

AI Safety Summit

In November 2023, the UK government brought together actors from across governments, leading AI companies, civil society, and academia to set the agenda for future international cooperation on AI safety. The Bletchley Declaration was signed by all countries attending the summit, including China and the United States. A second AI Safety Summit will be held online and hosted by South Korea in May 2024, followed by the next in-person Summit in France in November 2024.

International Trade

Amongst the many important shifts that occurred in 2023 was a dramatic shift in US digital trade policy. Until recently, the US had advocated for a more laissez-

faire approach to digital trade which effectively limited the permissibility of trade barriers, including forms of regulation. The decision by the US to withdraw its position on digital trade at the World Trade Organisation (WTO) follows its pivot towards a more interventionist stance. According to the Centre for Inclusive Trade Policy, controlling “potentially harmful developments of artificial intelligence – particularly Generative AI - cybersecurity, and the monopolistic power of large platforms requires public policies that may impact digital trade.”

Leaders from both sides of the Atlantic have continued efforts to strengthen ties in trade and technology through the US-EU Trade and Technology Council (TTC). Activities include the implementation of a Joint Roadmap on Evaluation and Measurement Tools for Trustworthy AI and Risk Management, and the launch of three dedicated expert groups: AI terminology and taxonomy, Cooperation on AI standards and tools for trustworthy AI and risk management, and Monitoring and measuring existing and emerging AI risks. The TTC is increasingly seen as a key forum in which the United States can develop strategies for dealing with intensifying rivalry with China. Some commentators have warned that this rivalry could precipitate a “race to the bottom” in AI governance.

International Standardisation

Technical standards are set to play a central role in the development and operationalisation of AI regulations, including the forthcoming European Union AI Act (see below).

Now that the AI Act has been finalised, attention will turn to its implementation. Industry-led organisations will develop and adopt harmonised standards against which conformity will be assessed. As with other EU regulations, high-level requirements are outlined in the legislation, but detailed specifications are left to standards. Three standards bodies are recognised by the European Union and European Free Trade Association (EFTA) as responsible for developing and defining voluntary standards at European level:

- European Committee for Standardization (CEN)
- European Electrotechnical Committee for Standardization (CENELEC)
- European Telecommunications Standards Institute (ETSI)

It is the first two of these bodies, CEN and CENELEC, who have been tasked with producing the relevant standards for the AI Act. AI standards, however, are still relatively immature compared to those in comparable industries. For businesses, this could push up the costs of compliance and make enforcement more challenging for regulators. A 2023 report by the European Commission's Joint Research Centre assessed a set of eight IEEE standards on AI, finding that all had only poor or moderate levels of maturity and detail.

Concerns have also been raised about the potential misalignment between the technical expertise of standardisation bodies, and the task of making decisions about interpreting human rights law and other public policy goals. Moreover, there are high barriers to effective civil society participation in standardisation processes. These include the time commitment required to take part, the complexity of standardisation processes, and the persistent dominance of larger industry actors.

In 2023, at least two significant standards offering strategic frameworks for organisations to manage risks associated with AI were published:

- ISO/IEC 23894 — Risk management standard was published. It offers strategic and sector-neutral guidance for managing risks connected to the development and use of AI, and guidance on how organisations can integrate risk management into their AI-driven activities and functions. It is possible that this standard will form the basis of one of the harmonised

standards to be developed by European Standards Organisations (ESOs) for the implementation of the proposed EU AI Act.

- The publication of ISO/IEC 42001 in December 2023 marked a milestone in the development of AI standards. It is a foundational management system standard providing guidelines for managing AI systems within organisations. It establishes a framework for organisations to systematically address and control risks related to AI development and deployment. Many commentators expect that CEN-CENELEC will use ISO/IEC 42001 as the basis of a European harmonised standard.

Regional Landscapes



Africa

AI governance became a central focus of the African Union (AU) in 2023. In August, leaders developed a Draft Conceptual Framework of the Continental Strategy on Artificial Intelligence. The framework will address the technological, ethical, economic, security and social perspectives of AI, and promote responsible, safe, and beneficial use.

The draft document was discussed and refined at the 5th Ordinary Session of the Specialized Technical Committee on Communication and Information Communications Technology in November 2023. Following the meeting, an [AU spokesperson](#) commented “AI is important to Africa because of its economic, social, political and geopolitical impact. AI technologies can stimulate economic growth by creating new industries, driving innovation, and generating employment opportunities. It can also support education and the preservation of African languages.”

Further key developments in 2023 included the Egyptian National Council for AI's publication of the first version of its Charter for Responsible AI. The charter aims to articulate "Egypt's interpretation of the various guidelines on ethical and responsible AI, adapted to the local context." It includes both general guidelines and implementation guidelines.

Policymakers in Kenya are continuing to consider how best to regulate AI technologies. Proposed legislation, known as the Kenya Robotics and Artificial Intelligence Society Bill, would create a body to oversee the activities of AI practitioners and would impose licence fees. According to reporting by Semfor, Kenya's tech sector remains largely opposed to the proposal.



Asia and Asia Pacific

Until last year, there were few signs of coordination on AI governance by Asian governments. In February 2024, however, the Association of Southeast Asian Nations (ASEAN) released its Guide on AI governance and ethics. The approach adopted is voluntary, and, unlike the EU's AI Act, does not prohibit or restrict the development and use of AI systems based on levels of risk.

In recent years, several significant AI governance initiatives have been adopted in China. Although often overlooked in the West, China's efforts to regulate AI are some of the most comprehensive instruments seen to date. Before 2023, China had already enacted regulation covering recommendation algorithms (2021), and deep fakes (2022). In August 2023, China became the first country to pass legislation specifically targeting generative AI. But, as Professor Angela Zhang has recently argued, these measures may be more interventionist on paper:

“despite maintaining strict information control over public-facing AI services, China’s overall approach to AI regulation has been markedly business-friendly.”

Most other countries in the region have been less inclined to bring hard regulation into force. Singapore, for example, has pursued a relatively industry-friendly approach. AI governance initiatives have tended to be limited to voluntary schemes like a practical [AI ethics toolkit for industry](#) and, in June 2023, [AI Verify](#), an AI governance testing framework and software toolkit that “validates the performance of AI systems against a set of internationally recognised principles through standardised tests.” Singapore’s revised national AI Strategy, released in December 2023, indicates no sign of a shift towards stronger regulation.

Japan has similarly favoured a less interventionist approach than that of the European Union. Writing for the Center for Strategic and International Studies, [Hiroki Habuka](#) suggests that rather than opting for strict and specific requirements, Japan has taken the approach of “respecting companies’ voluntary governance and providing non binding guidelines to support it, while imposing transparency obligations on some large digital platforms.” As such, Japan has no AI-specific legislation.

In recent months, the Australian government has been undertaking a [consultation on Safe and Responsible AI](#). Its interim response outlines some of the potential steps that will be taken, including mandatory guardrails for high-risk AI systems, a voluntary AI Safety Standard, and a voluntary labelling and watermarking scheme for AI-generated materials. For the time being, however, Australia has no specific regulatory regime addressing AI.



Europe and Central Asia

The most significant development in AI governance has been the development of the AI Act. Following months of negotiations, political agreement on the regulation was reached between the Parliament (made up of elected MEPs) and Council (made up of member state governments) in December 2023. The regulation aims to promote European innovation in AI while ensuring that values of fundamental rights, democracy, the rule of law and environmental sustainability are protected. It establishes obligations for market actors based on the potential risks and level of impact posed by different AI systems. The European Parliament formally adopted the text on 13 March 2024, beginning a phased two-year transition period.

The final text of the instrument maintains much of the original 2021 proposal, including its risk-based approach. One key departure is provisions on GPAI models, which include large generative AI models. These will be subject to transparency obligations and obligations to respect copyright law during training.

Additional requirements are also placed on models if they meet certain criteria based on compute power.

In 2023, the Council of Europe published a draft text for an international convention on AI focussed on human rights, the rule of law, and democracy. The definition of AI will be the same as that adopted by the OECD and the EU. Countries outside of the Council of Europe's membership, and the text will be heavily influenced by observer states like the United States, Japan, and Canada. At present, there is significant disagreement over the scope of the convention. It has been reported that the United States delegation has been pushing for a private sector exemption from the treaty. The European Union has rejected this position, arguing that it would undermine the protection of fundamental rights.

At the national level, EU member states will be bound by the final outcome of the AI Act. Outside of the EU, the UK has been pursuing an alternative, more tentative strategy. The AI regulation White Paper, published in March 2023 proposed a piecemeal, sectoral approach in which individual regulators will be tasked with incorporating principles into their work. To ensure regulatory coordination, there will be some central AI regulatory functions, but no new legal requirements are proposed. More recently, in its formal response to its consultation on the White Paper, the UK government signalled a possible change in direction, even going as far as to say "AI technologies will ultimately require legislative action in every country once understanding of risk has matured." With a general election looming, and a new government likely, it seems highly possible that UK policy will eventually shift its current non-statutory and primarily voluntary approach towards tougher rules.



Latin America and the Caribbean

Outside of Europe, Brazil represents an example of one of the most detailed regulatory proposals on AI. [Bill No. 2338/2023](#) aims to establish rules and principles to regulate the development and application of AI technologies. In many ways, the proposal closely reflects the EU AI Act. Both texts incorporate the OECD's expanded definition of AI systems, and both adopt a risk-based approach that creates tiers of obligations based on levels of risk. The Brazilian law arguably goes further in guaranteeing rights to individuals. Writing in *Internet Policy Review*, [Laura Schertel Mendes and Beatriz Kira](#) explain these protections: “While establishing fundamental rights for all interactions between machines and humans, such as information and transparency, it imposes additional obligations when the AI system has a significant impact or produces relevant legal effects. This includes the right to contestation and human intervention, ensuring a fair and comprehensive defence – an informational due process – for individuals affected by automated decisions.”

In 2023, Chile followed Brazil in taking the first steps towards discussing legislation to regulate AI. The proposed Bill seeks to incorporate legal and ethical issues into the development, distribution, commercialisation, and has been seen as another example of the EU's AI Act influencing regulatory policy around the world. Alongside these efforts to regulate the market in AI products and services, new guidelines on the use of AI in the public sector came into effect in January 2024.

Chile has also led on regional initiatives, including hosting the first Latin American and Caribbean Ministerial and High Level Summit on the Ethics of Artificial Intelligence. The event brought together ministers from over 30 countries to promote the implementation of UNESCO's Recommendation on the Ethics of Artificial Intelligence in Latin America and Caribbean. It culminated in the signing of the Santiago Declaration, and the proposal to establish a Council on Artificial Intelligence for Latin America and the Caribbean.



The Middle East and North Africa

While there are few regional AI governance processes and fora, several countries in the Middle East region are developing AI governance regimes. The United Arab Emirates (UAE) has long made clear its ambitions to become an AI leader by 2031. In 2023, the Technology Innovation Institute (TII), a scientific research centre and the applied research pillar of Abu Dhabi's Advanced Technology Research Council (ATRC), launched Falcon, a foundational large language model (LLM) with 40 billion parameters. New AI laws may not be far behind. According to a survey conducted by California-based Alteryx, Inc, around 87% of UAE business leaders believe that AI regulations and standards should be developed within their sector. Meanwhile, UAE is collaborating with University of Oxford on a project to educate public officials on AI governance, auditing AI systems, and the exploration of potential use cases.

UAE is, like Saudi Arabia, considered a battleground in the competition between China and the US over AI and computing hardware. In December 2023, one of its leading AI companies, G42, announced it is to cut ties with Chinese hardware suppliers in favour of partnerships with US companies.

Riyadh has pursued several domestic initiatives to enhance its AI capabilities, including by developing language models in Arabic and English. In December 2023, the Saudi Data and Artificial Intelligence Authority (SDAIA) announced the publication of a guide on generative AI.



North America

In January 2023, the National Institute of Standards and Technology (NIST) launched the first version of the Artificial Intelligence Risk Management Framework (AI RMF 1.0). This is a voluntary framework that helps AI actors to govern, map, measure and manage risks. It allows organisations to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems. This framework has been widely endorsed by key US-based industry actors, including: Microsoft, IBM, Google, Amazon Web Services, Partnership on AI, the Information Technology Industry Council, and the Alliance for Automotive Innovation. According to [Brookings](#), the framework “adds coherence to evolving U.S. policy on AI and contributes to ongoing international debate about AI policy and development.”

The AI RMF 1.0 has already been used to assess efforts by companies to make their AI systems safer. The [Federation of American Scientists](#), for example, used the framework to assess how well [efforts by OpenAI to test and improve GPT-4’s safety](#) before release conform to current best practice. They found some alignment

in the process to map, measure, and manage risks, and in specific measures used such as *red teaming*. However, they found that, whilst NIST's resources provide a helpful overview of considerations and best practices to be considered when managing AI risks, "they are not currently designed to provide concrete standards or metrics by which one can assess whether the practices taken by a given lab are "adequate.""

Later in 2023, President Biden issued a much-anticipated Executive Order on Safe, Secure and Trustworthy AI. The Order sets out requirements for AI transparency, testing, and cybersecurity measures, and directs a major programme of work across the federal government. In a White House statement release on 29 January 2024, the Biden administration confirmed it had delivered on the 90-day actions contained in the Executive Order. These included the Secretary of Commerce (SoC) compelling companies developing or demonstrating an intent to develop potential "dual use foundation models" to report and provide information to government.

Alongside White House initiatives, regulators have shown increasing willingness to intervene. In January 2024, the Federal Trade Commission (FTC) launched an inquiry into five AI companies: Alphabet, Amazon, Anthropic, Microsoft, and OpenAI. Each is required to provide information regarding recent investments and partnerships involving generative AI companies and major cloud service providers. FTC chair Lina Khan says the study "will shed light on whether investments and partnerships pursued by dominant companies risk distorting innovation and undermining fair competition."

Canada has proposed AI-specific legislation in the form of the Artificial Intelligence Data Act (AIDA); one of three pieces of legislation contained within the omnibus Bill C-27, the Digital Charter Implementation Act. AIDA has two main aims: promoting interprovincial trade in AI systems by establishing common requirements, applicable across Canada; and prohibiting conduct that may result

in serious harm to individuals or their interests. In this way, AIDA bears some resemblance to the EU AI Act. While policymakers decide on the scope, reach and content of any future regulation, the federal government has introduced a voluntary code of conduct on generative AI.