September 2024

# GOVERNING AI FOR HUMANITY

United Nations | AI Advisory Body

**Governing AI for Humanity: Final Report**



eISBN: 9789211067873

# Governing AI for Humanity

## Final Report

United Nations | AI Advisory Body

# About the High-level Advisory Body on Artificial Intelligence

The multi-stakeholder High-level Advisory Body on Artificial Intelligence, initially proposed in 2020 as part of the United Nations Secretary-General's Roadmap for Digital Cooperation (A/74/821), was formed in October 2023 to undertake analysis and advance recommendations for the international governance of artificial intelligence.

The members of the Advisory Body have participated in their personal capacity, not as representatives of their respective organizations. This report represents a majority consensus; no member is expected to endorse every single point contained in this document. The members affirm their broad, but not unilateral, agreement with its findings and recommendations. The language included in this report does not imply institutional endorsement by the members' respective organizations.

# Table of Contents

# Executive summary

i Artificial intelligence (AI) is transforming our world. This suite of technologies offers tremendous potential for good, from opening new areas of scientific inquiry and optimizing energy grids, to improving public health and agriculture and promoting broader progress on the Sustainable Development Goals (SDGs).

ii Left ungoverned, however, AI's opportunities may not manifest or be distributed equitably. Widening digital divides could limit the benefits of AI to a handful of States, companies and individuals. Missed uses – failing to take advantage of and share AI-related benefits because of lack of trust or missing enablers such as capacity gaps and ineffective governance – could limit the opportunity envelope.

iii AI also brings other risks. AI bias and surveillance are joined by newer concerns, such as the confabulations (or "hallucinations") of large language models, AI-enhanced creation and dissemination of disinformation, risks to peace and security, and the energy consumption of AI systems at a time of climate crisis.

iv Fast, opaque and autonomous AI systems challenge traditional regulatory systems, while ever-more-powerful systems could upend the world of work. Autonomous weapons and public security uses of AI raise serious legal, security and humanitarian questions.

v There is, today, a global governance deficit with respect to AI. Despite much discussion of ethics and principles, the patchwork of norms and institutions is still nascent and full of gaps. Accountability is often notable for its absence, including for deploying non-explainable AI systems that impact others. Compliance often rests on voluntarism; practice belies rhetoric.

vi As noted in our interim report,[1] AI governance is crucial – not merely to address the challenges and risks, but also to ensure that we harness AI's potential in ways that leave no one behind.

## 1. The need for global governance

vii The imperative of global governance, in particular, is irrefutable. AI's raw materials, from critical minerals to training data, are globally sourced. General-purpose AI, deployed across borders, spawns manifold applications globally. The accelerating development of AI concentrates power and wealth on a global scale, with geopolitical and geoeconomic implications.

viii Moreover, no one currently understands all of AI's inner workings enough to fully control its outputs or predict its evolution. Nor are decision makers held accountable for developing, deploying or using systems they do not understand. Meanwhile, negative spillovers and downstream impacts resulting from such decisions are also likely to be global.

ix The development, deployment and use of such a technology cannot be left to the whims of markets alone. National governments and regional organizations will be crucial, but the very nature of the technology itself – transboundary in structure and application – necessitates a global approach. Governance can also be a key enabler for AI innovation for the SDGs globally.

x AI, therefore, presents challenges and opportunities that require a holistic, global approach cutting transversally across political, economic, social, ethical, human rights, technical, environmental

1 See https://un.org/ai-advisory-body.

and other domains. Such an approach can turn a patchwork of evolving initiatives into a coherent, interoperable whole, grounded in international law and the SDGs, adaptable across contexts and over time.

xi In our interim report, we outlined principles[2] that should guide the formation of new international AI governance institutions. These principles acknowledge that AI governance does not take place in a vacuum, that international law, especially international human rights law, applies in relation to AI.

## 2. Global AI governance gaps

xii There is no shortage of documents and dialogues focused on AI governance. Hundreds of guides, frameworks and principles have been adopted by governments, companies and consortiums, and regional and international organizations.

xiii Yet, none of them can be truly global in reach and comprehensive in coverage. This leads to problems of representation, coordination and implementation.

xiv In terms of representation, whole parts of the world have been left out of international AI governance conversations. Figure (a) shows seven prominent, non-United Nations AI initiatives.[3] Seven countries are parties to all the sampled AI governance efforts, whereas 118 countries are parties to none (primarily in the global South).

xv Equity demands that more voices play meaningful roles in decisions about how to govern technology that affects us. The concentration of decision-making in the AI technology sector cannot be justified; we must also recognize that historically many communities have been entirely excluded from AI governance conversations that impact them.

xvi AI governance regimes must also span the globe to be effective — effective in averting "AI arms races" or a "race to the bottom" on safety and rights, in detecting and responding to incidents emanating from decisions along AI's life cycle which span multiple jurisdictions, in spurring learning, in encouraging interoperability, and in sharing AI's benefits. The technology is borderless and, as it spreads, the illusion that any one State or group of States could (or should) control it will diminish.

xvii Coordination gaps between initiatives and institutions risk splitting the world into disconnected and incompatible AI governance regimes. Coordination is also lacking within the United Nations system. Although many United Nations entities touch on AI governance, their specific mandates mean that none does so in a comprehensive manner.

xviii However, representation and coordination are not enough. Accountability requires implementation so that commitments to global AI governance translate to tangible outcomes in practice, including on capacity development and support to small and medium enterprises, so that opportunities are shared. Much of this will take place at the national and regional levels, but more is also needed globally to address risks and harness benefits.

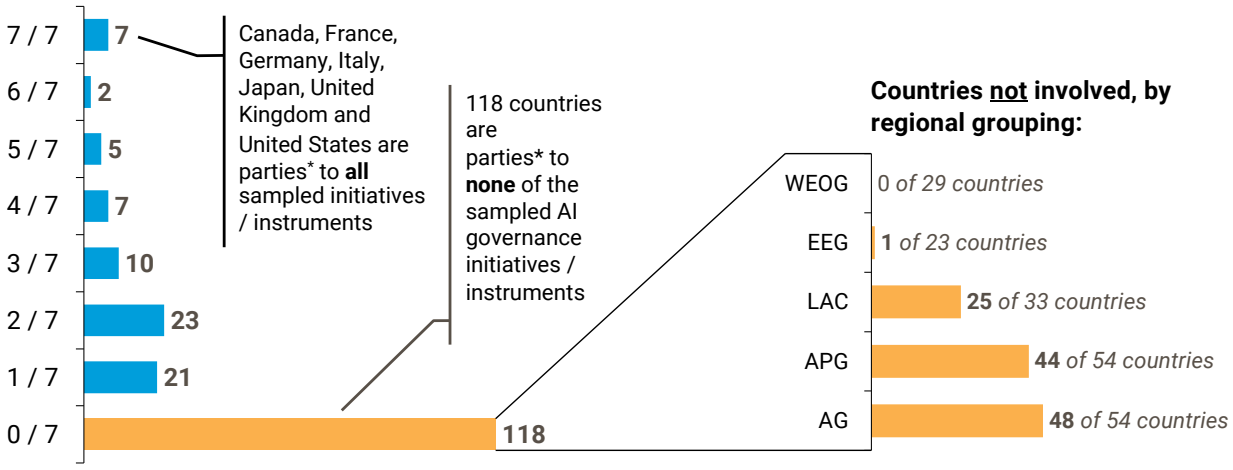
2 Guiding principle 1: AI should be governed inclusively, by and for the benefit of all; guiding principle 2: AI must be governed in the public interest; guiding principle 3: AI governance should be built in step with data governance and the promotion of data commons; guiding principle 4: AI governance must be universal, networked and rooted in adaptive multi-stakeholder collaboration; guiding principle 5: AI governance should be anchored in the Charter of the United Nations, international human rights law and other agreed international commitments, such as the SDGs.

3 Excluding the United Nations Educational, Scientific and Cultural Organization (UNESCO) Recommendation on the Ethics of Artificial Intelligence (2021) and the two General Assembly resolutions on AI in 2024: "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development" (78/265) and "Enhancing international cooperation on capacity-building of artificial intelligence" (78/311).

## Figure (a): Representation in seven non-United Nations international AI governance initiatives

*Sample*: OECD AI Principles (2019), G20 AI principles (2019), Council of Europe AI Convention drafting group (2022–2024), GPAI Ministerial Declaration (2022), G7 Ministers' Statement (2023), Bletchley Declaration (2023) and Seoul Ministerial Declaration (2024).

**INTERREGIONAL ONLY, EXCLUDES REGIONAL**




* Per endorsement of relevant intergovernmental issuances. Countries are not considered involved in a plurilateral initiative solely because of membership in the European Union or the African Union. *Abbreviations*: AG, African Group; APG, Asia and the Pacific Group; EEG, Eastern European Group; G20, Group of 20; G7, Group of Seven; GPAI, Global Partnership on Artificial Intelligence; LAC, Latin America and the Caribbean; OECD, Organisation for Economic Co-operation and Development; WEOG, Western European and Others Group.

# 3. Enhancing global cooperation

**xix** Our recommendations advance a holistic vision for a globally networked, agile and flexible approach to governing AI for humanity, encompassing common understanding, common ground and common benefits. Only such an inclusive and comprehensive approach to AI governance can address the multifaceted and evolving challenges and opportunities AI presents on a global scale, promoting international stability and equitable development.

**xx** Guided by principles established in our interim report, our proposals seek to fill gaps and bring coherence to the fast-emerging ecosystem of international AI governance responses and initiatives, helping to avoid fragmentation and missed opportunities. To support these measures efficiently and to partner effectively with other institutions, we propose a light, agile structure as an expression of coherent effort: an AI office in the United Nations Secretariat, close to the Secretary-General, working as the "glue" to unite the initiatives proposed here efficiently and sustainably.

## A. Common understanding

**xxi** A global approach to governing AI starts with a common understanding of its capabilities, opportunities, risks and uncertainties. There is a need for timely, impartial and reliable scientific knowledge and information about AI so that Member States can build a shared foundational understanding worldwide, and to balance information asymmetries between companies housing expensive AI labs and the rest of the world (including via information-sharing between AI companies and the broader AI community).

**xxii** Pooling scientific knowledge is most efficient at the global level, enabling joint investment in a global public good, and public interest collaboration across otherwise fragmented and duplicative efforts.

## International scientific panel on AI

**xxiii** Learning from precedents such as the Intergovernmental Panel on Climate Change (IPCC) and the United Nations Scientific Committee on the Effects of Atomic Radiation, an international, multidisciplinary scientific panel on AI could collate and catalyse leading-edge research to inform scientists, policymakers, Member States and other stakeholders seeking scientific perspectives on AI technology or its applications from an impartial, credible source.

**xxiv** A scientific panel under the auspices of the United Nations could source expertise on AI-related opportunities. This might include facilitating "deep dives" into applied domains of the SDGs, such as health care, energy, education, finance, agriculture, climate, trade and employment.

**xxv** Risk assessments could also draw on the work of other AI research initiatives, with the United Nations offering a uniquely trusted "safe harbour" for researchers to exchange ideas on the "state of the art". By pooling knowledge across silos in countries or companies that may not otherwise engage or be included, a United Nations-hosted panel can help to rectify misperceptions and bolster trust globally.

**xxvi** Such a panel should operate independently, with support from a cross-United Nations system team drawn from the below-proposed AI office and relevant United Nations agencies, such as the International Telecommunication Union (ITU) and the United Nations Educational, Scientific and Cultural Organization (UNESCO). It should partner with research efforts led by other international institutions, such as the Organisation for Economic Co-operation and Development (OECD) and the Global Partnership on Artificial Intelligence.

*Recommendation 1*

# An international scientific panel on AI

We recommend the creation of an independent international scientific panel on AI, made up of diverse multidisciplinary experts in the field serving in their personal capacity on a voluntary basis. Supported by the proposed United Nations AI office and other relevant United Nations agencies, partnering with other relevant international organizations, its mandate would include:

**a)** Issuing an annual report surveying AI-related capabilities, opportunities, risks and uncertainties, identifying areas of scientific consensus on technology trends and areas where additional research is needed;

**b)** Producing quarterly thematic research digests on areas in which AI could help to achieve the SDGs, focusing on areas of public interest which may be under-served; and

**c)** Issuing ad hoc reports on emerging issues, in particular the emergence of new risks or significant gaps in the governance landscape.

## B. Common ground

**xxvii** Alongside a common understanding of AI, common ground is needed to establish interoperable governance approaches anchored in global norms and principles in the interests of all countries. This is required at the global level to avert regulatory races to the bottom while reducing regulatory friction across borders; to maximize learning and technical interoperability; and to respond effectively to challenges arising from the transboundary character of AI.

## Policy dialogue on AI governance

**xxviii** An inclusive policy forum is needed so that all Member States, drawing on the expertise of stakeholders, can share best practices that are based on human rights and foster development, that foster interoperable governance approaches and that account for transboundary challenges that warrant further policy consideration. This does not mean global governance of all aspects of AI, but it can set the framework for international cooperation and better align industry and national efforts with global norms and principles.

**xxix** Institutionalizing such multi-stakeholder exchange under the auspices of the United Nations can provide a reliably inclusive home for discussing emerging governance practices and appropriate policy responses. By edging beyond comfort zones, dialogue between non-likeminded countries, and between States and stakeholders, can catalyse learning and lay foundations for greater cooperation, such as on safety standards and rights, and for times of global crisis. A United Nations setting is essential to anchoring this effort in the widest possible set of shared norms.

**xxx** Combined with capacity development (see recommendations 4 and 5), such inclusive dialogue on governance approaches can help States and companies to update their regulatory approaches and methodologies to respond to accelerating AI. Connections to the international scientific panel would enhance that dynamic, comparable to the relationship between IPCC and the United Nations Climate Change Conference.

**xxxi** A policy dialogue could begin on the margins of existing meetings in New York (such as the General Assembly[4]) and in Geneva. Twice-yearly meetings could focus more on opportunities across diverse sectors in one meeting, and more on risks in the other meeting.[5] Moving forward, a gathering like this would be an appropriate forum for sharing information about AI incidents, such as those that stretch or exceed the capacities of existing agencies.

**xxxii** One portion of each dialogue session might focus on national approaches led by Member States, with a second portion sourcing expertise and inputs from key stakeholders – in particular, technology companies and civil society representatives. In addition to the formal dialogue sessions, multi-stakeholder engagement on AI policy could leverage other existing, more specialized mechanisms, such as the ITU AI for Good meeting, the annual Internet Governance Forum meeting, the UNESCO Global Forum on AI Ethics and the United Nations Conference on Trade and Development (UNCTAD) eWeek.

4 Analogous to the high-level political forum in the context of the SDGs that takes place under the auspices of the Economic and Social Council.

5 Relevant parts of the United Nations system could be engaged to highlight opportunities and risks, including ITU on AI standards; ITU, the United Nations Conference on Trade and Development (UNCTAD), the United Nations Development Programme (UNDP) and the Development Coordination Office on AI applications for the SDGs; UNESCO on ethics and governance capacity; the Office of the United Nations High Commissioner for Human Rights (OHCHR) on human rights accountability based on existing norms and mechanisms; the Office for Disarmament Affairs on regulating AI in military systems; UNDP on support to national capacity for development; the Internet Governance Forum for multi-stakeholder engagement and dialogue; the World Intellectual Property Organization (WIPO), the International Labour Organization (ILO), the World Health Organization (WHO), the Food and Agriculture Organization of the United Nations (FAO), the World Food Programme, the United Nations High Commissioner for Refugees (UNHCR), UNESCO, the United Nations Children's Fund, the World Meteorological Organization and others on sectoral applications and governance.

*Recommendation 2*

# Policy dialogue on AI governance

We recommend the launch of a twice-yearly intergovernmental and multi-stakeholder policy dialogue on AI governance on the margins of existing meetings at the United Nations. Its purpose would be to:

**a)** Share best practices on AI governance that foster development while furthering respect, protection and fulfilment of all human rights, including pursuing opportunities as well as managing risks;

**b)** Promote common understandings on the implementation of AI governance measures by private and public sector developers and users to enhance international interoperability of AI governance;

**c)** Share voluntarily significant AI incidents that stretched or exceeded the capacity of State agencies to respond; and

**d)** Discuss reports of the international scientific panel on AI, as appropriate.

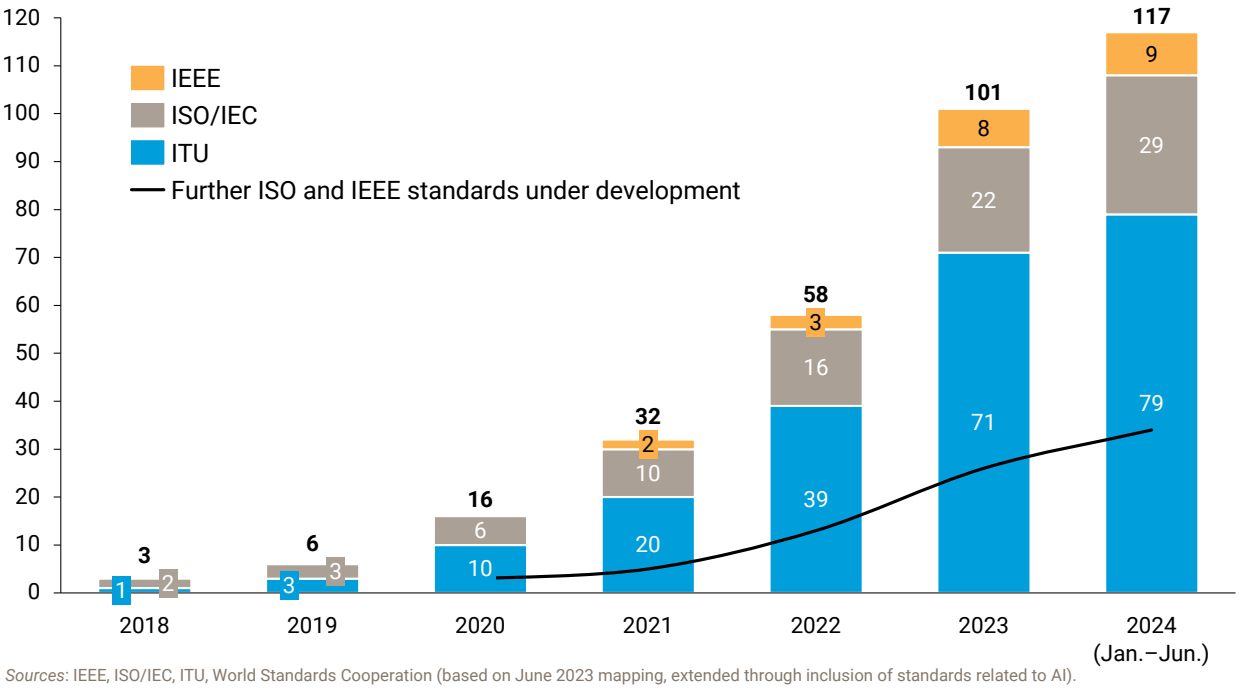
## AI standards exchange

**xxxiii** When AI systems were first explored, few standards existed to help to navigate or measure this new frontier. More recently, there has been a proliferation of standards. Figure (b) illustrates the increasing number of standards adopted by ITU, the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC) and the Institute of Electrical and Electronics Engineers (IEEE).

**xxxiv** There is no common language among these standards bodies, and many terms routinely used with respect to AI – fairness, safety, transparency – do not have agreed definitions. There are also disconnects between those standards that were adopted for narrow technical or internal validation purposes, and those that are intended to incorporate broader ethical principles. We now have an emerging set of standards that are not grounded in a common understanding of meaning or are divorced from the values that they were intended to uphold.

**xxxv** Drawing on the expertise of the international scientific panel and incorporating members from the various national and international entities that have contributed to standard-setting, as well as representatives from technology companies and civil society, the United Nations system could serve as a clearing house for AI standards that would apply globally.

## Figure (b): Number of standards related to AI




*Sources*: IEEE, ISO/IEC, ITU, World Standards Cooperation (based on June 2023 mapping, extended through inclusion of standards related to AI).

*Recommendation 3*

# AI standards exchange

We recommend the creation of an AI standards exchange, bringing together representatives from national and international standard-development organizations, technology companies, civil society and representatives from the international scientific panel. It would be tasked with:

**a)** Developing and maintaining a register of definitions and applicable standards for measuring and evaluating AI systems;

**b)** Debating and evaluating the standards and the processes for creating them; and

**c)** Identifying gaps where new standards are needed.

## C. Common benefits

**xxxvi** The 2030 Agenda for Sustainable Development, with its 17 SDGs, can give clarity of purpose to the development, deployment and uses of AI, bending the arc of investments towards global development challenges. Without a comprehensive and inclusive approach to AI governance, the potential of AI to contribute positively to the SDGs could be missed, and its deployment could inadvertently reinforce or exacerbate disparities and biases.

**xxxvii** AI is no panacea for sustainable development challenges; it is one component within a broader set of solutions. To truly unlock AI's potential to address societal challenges, collaboration among governments, academia, industry and civil society is crucial, so that AI-enabled solutions are inclusive and equitable.

**xxxviii** Much of this depends on access to talent, computational power (or "compute") and data, in ways that help cultural and linguistic diversity to flourish. Basic infrastructure and the resources to maintain it are also pre-requisites.

**xxxix** Regarding talent, not every society needs cadres of computer scientists for building their own models. However, whether technology is bought, borrowed or built, a baseline socio-technical capacity is needed to understand the capabilities and limitations of AI, and harness AI-enabled use cases appropriately while addressing context-specific risks.

**xl** Compute is one of the biggest barriers to entry in the field of AI. Of the top 100 high-performance computing clusters in the world capable of training large AI models, not one is hosted in a developing country.[6] It is unrealistic to promise access to compute that even the wealthiest countries and companies struggle to acquire. Rather, we seek to put a floor under the AI divide for those unable to secure needed enablers via other means, including by supporting initiatives towards distributed and federated AI development models.

**xli** Turning to data, it is common to speak of misuse of data in the context of AI (such as infringements on privacy) or missed uses of data (failing to exploit existing data sets). However, a related problem is missing data, which includes the large portions of the globe that are data poor. Failure to reflect the world's linguistic and cultural diversity has been linked to bias in AI systems, but may also be a missed opportunity for those communities to access AI's benefits.

**xlii** A set of shared resources – including open models – is needed to support inclusive and effective participation by all Member States in the AI ecosystem, and here global approaches have distinct advantages.

## Capacity development network

**xliii** Growing public and private demand for human and other AI capacity coincides with emergent national, regional and public-private AI centres of excellence that have international capacity development roles. A global network can serve as a matching platform that expands the range of possible partnering and enhances interoperability of capacity-building approaches.

**xliv** From the Millennium Development Goals to the SDGs, the United Nations has long embraced developing the capacities of individuals and institutions.[7] A network of institutions, affiliated

6 Proxy indicator since most high-performance computing clusters do not have graphics processing units (GPUs) and are of limited use for advanced AI.

7 Through the work of UNESCO, WIPO and others, the United Nations has helped to uphold the rich diversity of cultures and knowledge-making traditions across the globe. The United Nations University has long had a commitment to capacity-building through higher education and research, and the United Nations Institute for Training and Research has helped to train officials in domains critical to sustainable development. The UNESCO Readiness Assessment Methodology is a key tool to support Member States in their implementation of the UNESCO Recommendation on the Ethics of Artificial Intelligence. Other examples include the WHO Academy in Lyon, France, the UNCTAD Virtual Institute, the United Nations Disarmament Fellowship run by the Office for Disarmament Affairs and the capacity development programmes led by ITU and UNDP.

with the United Nations, could expand options for countries seeking capacity partnerships. It could also catalyse new national centres of excellence to stimulate the development of local AI innovation ecosystems, following interoperable approaches aligned with United Nations normative commitments.

**xlv** Such a network would promote an alternative paradigm of AI technology development: bottom-up, cross-domain, open and collaborative. National-level efforts could continue to employ diagnosis tools, such as the UNESCO AI Readiness Assessment Methodology, to help to identify gaps at the national level, with the international network helping to address them.

*Recommendation 4*

## Capacity development network

We recommend the creation of an AI capacity development network to link up a set of collaborating, United Nations-affiliated capacity development centres making available expertise, compute and AI training data to key actors. The purpose of the network would be to:

**a)** Catalyse and align regional and global AI capacity efforts by supporting networking among them;

**b)** Build AI governance capacity of public officials to foster development while furthering respect, protection and fulfilment of all human rights;

**c)** Make available trainers, compute and AI training data across multiple centres to researchers and social entrepreneurs seeking to apply AI to local public interest use cases, including via:

- **i)** Protocols to allow cross-disciplinary research teams and entrepreneurs in compute-scarce settings to access compute made available for training/tuning and applying their models appropriately to local contexts;
- **ii)** Sandboxes to test potential AI solutions and learn by doing;
- **iii)** A suite of online educational opportunities on AI targeted at university students, young researchers, social entrepreneurs and public sector officials; and
- **iv)** A fellowship programme for promising individuals to spend time in academic institutions or technology companies.

## Global fund for AI

**xlvi** Many countries face fiscal and resource constraints limiting their ability to use AI appropriately and effectively. Despite any capacity development efforts (recommendation 4), some may still be unable to access training, compute, models and training data without international support. Other funding efforts may also not scale without it.

**xlvii** Our intention in proposing a fund is not to guarantee access to advanced compute resources and capabilities. The answer may not always be more compute. We also need better ways to connect talent, compute and data. The fund's purpose would be to address the underlying capacity and collaboration gaps for those unable to access requisite enablers so that:

a. Countries in need can access AI enablers, putting a floor under the AI divide;
b. Collaborating on AI capacity development leads to habits of cooperation and mitigates geopolitical competition;
c. Countries with divergent regulatory approaches have incentives to develop common templates for governing data, models and applications for societal-level challenges related to the SDGs and scientific breakthroughs.

**xlviii** This public interest focus makes the fund complementary to the proposal for an AI capacity development network, to which the fund would also channel resources. The fund would provide an independent capacity for monitoring of impact, and could source and pool in-kind contributions, including from private sector entities, to make available AI-related training programmes, time, compute, models and curated data sets at lower-than-market cost. In this manner, we ensure that vast swathes of the world are not left behind and are instead empowered to harness AI for the SDGs in different contexts.

**xlix** It is in everyone's interest to ensure that there is cooperation in the digital world as in the physical world. Analogies can be made to efforts to combat climate change, where the costs of transition, mitigation or adaptation do not fall evenly, and international assistance is essential to help resource-constrained countries so that they can join the global effort to tackle a planetary challenge.

*Recommendation 5*

# Global fund for AI

We recommend the creation of a global fund for AI to put a floor under the AI divide. Managed by an independent governance structure, the fund would receive financial and in-kind contributions from public and private sources and disburse them, including via the capacity development network, to facilitate access to AI enablers to catalyse local empowerment for the SDGs, including:

**a)** Shared computing resources for model training and fine-tuning by AI developers from countries without adequate local capacity or the means to procure it;

**b)** Sandboxes and benchmarking and testing tools to mainstream best practices in safe and trustworthy model development and data governance;

**c)** Governance, safety and interoperability solutions with global applicability;

**d)** Data sets and research into how data and models could be combined for SDG-related projects; and

**e)** A repository of AI models and curated data sets for the SDGs.

## Global AI data framework

**I** Access to AI training data, via market or other mechanisms, is a critical enabler for flourishing local AI ecosystems — particularly in countries, communities, regions and demographic groups with "missing" data (see the section on "common benefits" above).

**Ii** Only global collective action can incentivize interoperability, stewardship, privacy preservation, empowerment and rights enhancement in ways that promote a "race to the top" across jurisdictions towards protection of human rights and other agreed commitments, data availability and fair compensation to data subjects in the governance of the collection, creation, use and monetization of AI training data. This aim motivates our proposal for a global AI data framework.

**Iii** Such a framework would not create new data-related rights. Rather, it would address issues of availability, interoperability and use of AI training data. It would help to build common understanding on how to align different national and regional data protection frameworks. It could also promote flourishing local AI ecosystems supporting cultural and linguistic diversity, as well as limiting further economic concentration.

**Iiii** These measures could be complemented by promoting data commons and provisions for hosting data trusts in areas relevant to the SDGs, based on templates for agreements to hold and

share data in a fair, safe and equitable manner. The development of these templates and the actual storage and analysis of data held in commons or in trusts could be supported by the proposed capacity development network and global fund for AI (recommendations 4 and 5).

[liv] The United Nations is uniquely positioned to support the establishment of global principles and practical arrangements for AI training data governance and use, in line with agreed international commitments on human rights, intellectual property and sustainable development, building on years of work by the data community and integrating it with recent developments on AI ethics and governance. This is analogous to the role of the United Nations Commission on International Trade Law in advancing international trade by developing legal and non-legal cross-border frameworks.

[lv] Similarly, the Commission on Science and Technology for Development and the Statistical Commission have on their agenda data for development and data on the SDGs. There are also important issues of content, copyright and protection of indigenous knowledge and cultural expression being considered by the World Intellectual Property Organization (WIPO).

*Recommendation 6*

## Global AI data framework

We recommend the creation of a global AI data framework, developed through a process initiated by a relevant agency such as the United Nations Commission on International Trade Law and informed by the work of other international organizations, for:

a) Outlining data-related definitions and principles for global governance of AI training data, including as distilled from existing best practices, and to promote cultural and linguistic diversity;

b) Establishing common standards around AI training data provenance and use for transparent and rights-based accountability across jurisdictions; and

c) Instituting market-shaping data stewardship and exchange mechanisms for enabling flourishing local AI ecosystems globally, such as:

   i) Data trusts;

   ii) Well-governed global marketplaces for exchange of anonymized data for training AI models; and

   iii) Model agreements for facilitating international data access and global interoperability, potentially as techno-legal protocols to the framework.

## D. Coherent effort

lvi The above proposals seek to address the representation, coordination and implementation gaps identified in the emerging international AI governance regime. These gaps can be addressed through partnerships and collaboration with existing institutions and mechanisms to promote a common understanding, common ground and common benefits.

lvii Nevertheless, without a dedicated focal point in the United Nations to support and enable coordination among these and other efforts, the world will lack the inclusively networked, agile and coherent approach required for effective and equitable governance of AI as a transboundary, fast-changing and general-purpose technology.

lviii The patchwork of norms and institutions outlined under the section "Global AI governance gaps" above, reflect widespread recognition that governance of AI is a global necessity. The unevenness of that response demands some measure of coherent effort.

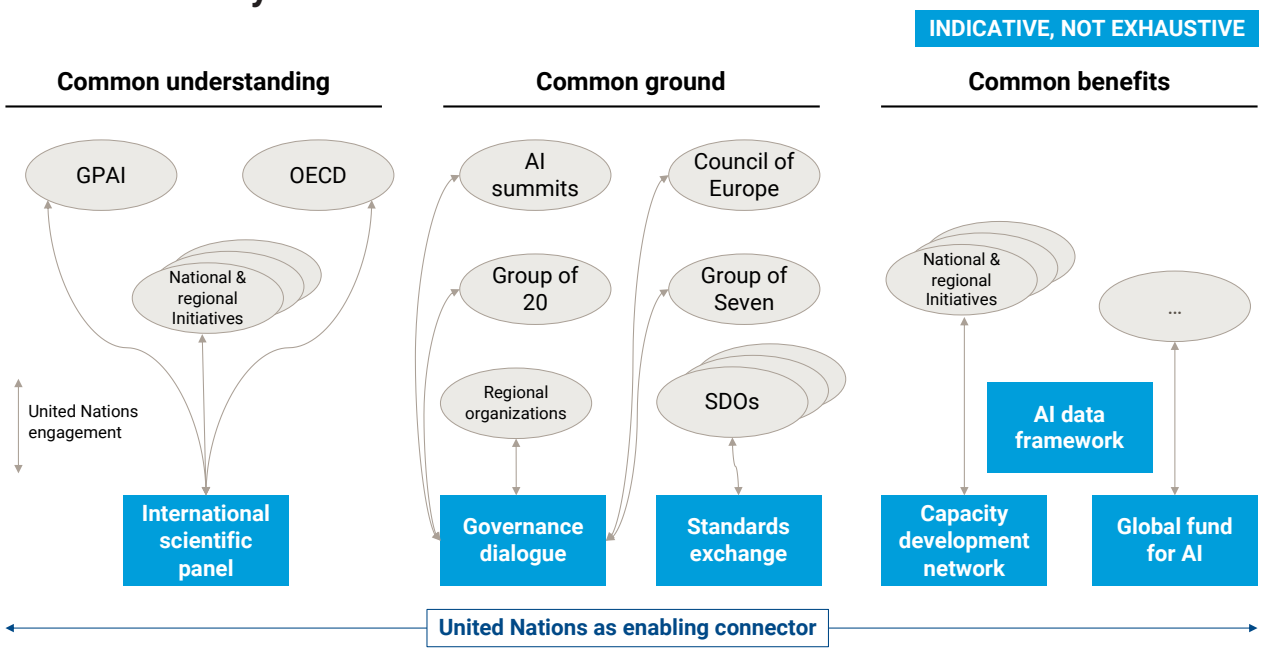
## AI office in the United Nations Secretariat

lix We, therefore, propose a light touch mechanism to act as the "glue" that supports and catalyses the proposals in this report, including through partnerships, while also enabling the United Nations system to speak with one voice in the evolving AI governance ecosystem.

lx This small, agile capacity, in the form of an AI office within the United Nations Secretariat, would report to the Secretary-General, conferring the benefit of connections throughout the United Nations system, without being tied to one part of it. That is important because of the uncertain future of AI and the strong likelihood that it will permeate all aspects of human endeavour.

lxi Such a body should be agile, champion inclusion and partner rapidly to accelerate coordination and implementation – drawing as a first priority on existing resources and functions within the United Nations system. The focus should be on civilian applications of AI.

## Figure (c): Proposed role of the United Nations in the international AI governance ecosystem




*Abbreviations*: GPAI, Global Partnership on Artificial Intelligence; OECD, Organisation for Economic Co-operation and Development; SDOs, standards development organizations.

[lxii] It could be staffed in part by United Nations personnel seconded from specialized agencies and other parts of the United Nations system, such as ITU, UNESCO, the Office of the United Nations High Commissioner for Human Rights (OHCHR), UNCTAD, the United Nations University and the United Nations Development Programme (UNDP). It should engage multiple stakeholders, including companies, civil society and academia, and work in partnership with leading organizations outside of the United Nations (see fig. (c)). This would position the United Nations to enable connections for fostering common understanding, common ground and common benefits in the international AI governance ecosystem.

[lxiii] Recommendation 7 is made on the basis of a clear-eyed assessment as to where the United Nations can add value, including where it can lead, where it can aid coordination and where it should step aside. It also brings the benefits of existing institutional arrangements, including pre-negotiated funding and administrative processes that are well established and understood.

*Recommendation 7*

## AI office within the Secretariat

We recommend the creation of an AI office within the Secretariat, reporting to the Secretary-General. It should be light and agile in organization, drawing, wherever possible, on relevant existing United Nations entities. Acting as the "glue" that supports and catalyses the proposals in this report, partnering and interfacing with other processes and institutions, the office's mandate would include:

a) Providing support for the proposed international scientific panel, policy dialogue, standards exchange, capacity development network and, to the extent required, the global fund and global AI data framework;

b) Engaging in outreach to diverse stakeholders, including technology companies, civil society and academia, on emerging AI issues; and

c) Advising the Secretary-General on matters related to AI, coordinating with other relevant parts of the United Nations system to offer a whole-of-United Nations response.

## E. Reflections on institutional models

lxiv Discussions about AI often resolve into extremes. In our consultations around the world, we engaged with those who see a future of boundless goods provided by ever-cheaper, ever-more-helpful AI systems. We also spoke with those wary of darker futures, of division and unemployment, and even extinction.[8]

lxv We do not know whether the utopian or dystopian future is more likely. Equally, we are mindful that the technology may go in a direction that does away with this duality. This report focuses on the near-term opportunities and risks, based on science and grounded in fact.

lxvi The seven recommendations outlined above offer our best hope for reaping the benefits of AI, while minimizing and mitigating the risks, as AI continues evolving. We are also mindful of the practical challenges to international institution-building on a larger scale. This is why we are proposing a networked institutional approach, with light and agile support. If or when risks become more acute and the stakes for opportunities escalate, such calculations may change.

lxvii The world wars led to the modern international system; the development of ever-more-powerful chemical, biological and nuclear weapons led to regimes limiting their spread and promoting peaceful uses of the underlying technologies. Evolving understanding of our common humanity led to the modern human rights system and our ongoing commitment to the SDGs for all. Climate change evolved from a niche concern to a global challenge.

lxviii AI may similarly rise to a level that requires more resources and more authority than is proposed in the above-mentioned recommendations, into harder functions of norm elaboration, implementation, monitoring, verification and validation, enforcement, accountability, remedies for harm and emergency responses. Reflecting on such institutional models, therefore, is prudent. The final section of this report seeks to contribute to that effort.

# 4. A call to action

lxix We remain optimistic about the future with AI and its positive potential. That optimism depends, however, on realism about the risks and the inadequacy of structures and incentives currently in place. The technology is too important, and the stakes are too high, to rely only on market forces and a fragmented patchwork of national and multilateral action.

lxx The United Nations can be the vehicle for a new social contract for AI that ensures global buy-in for a governance regime which protects and empowers us all. Such a social contract will ensure that opportunities are fairly distributed, and the risks are not loaded on to the most vulnerable – or passed on to future generations, as we have seen, tragically, with climate change.

lxxi As a group and as individuals from across many fields of expertise, organizations and parts of the world, we look forward to continuing this crucial conversation. Together with the many others we have connected with on this journey, and the global community they represent, we hope that this report contributes to our combined efforts to govern AI for humanity.

8 See https://safe.ai/work/statement-on-ai-risk.

## Figure (d): High-level Advisory Body on Artificial Intelligence at its meeting in Singapore, 29 May 2024

# 1. Introduction

1 The Secretary-General's High-level Advisory Body on Artificial Intelligence was formed to analyse and advance recommendations for the international governance of artificial intelligence (AI). Our members are diverse by geography and gender, discipline and age; we draw expertise from governments, civil society, the private sector and academia. Intense and wide-ranging discussions have yielded broad agreement (as reflected in our interim report[1]) that there is a global governance deficit with respect to AI. In that report, we articulated guiding principles for that role and functions that could be required internationally.

2 Over subsequent months, we benefited from extensive feedback and consultations. This included 18 "deep dives" on specific issue areas with more than 500 expert participants, more than 250 written submissions from over 150 organizations and 100 individuals from all regions, an AI risk pulse check with around 350 expert respondents from all regions, an opportunity scan with around 120 expert respondents from all regions, and regular consultations with and briefings of Member States, United Nations entities and other stakeholder groups in more than 40 engagements across all regions.[2] Members of the Advisory Body have also engaged extensively in forums around the world, held more than a hundred virtual discussions and had three plenary in-person meetings, in New York, Geneva and Singapore.

3 The present final report, therefore, has many authors. While it cannot reflect the full richness and diversity of views expressed, it shows our shared commitment to ensuring that AI is developed, deployed and used in a manner that benefits all of humanity, and ensuring that AI is governed effectively and inclusively at the international level.

4 This report reaffirms the findings of the Advisory Body's interim report on opportunities and enablers, risks and challenges; it also reprises the need for global governance of AI and outlines seven recommendations.

5 These include a scientific panel to promote a common understanding of AI capabilities, opportunities, risks and uncertainties. Based on this common understanding, we need mechanisms to find common ground on how AI should be governed at the international levels. Achieving that depends on regular dialogue and the development of standards acceptable and applicable to all.

6 The report also makes recommendations on common benefits, intended to ensure that the benefits of AI are equitably shared, which can depend on access to models or capabilities such as talent, computational power (or "compute") and data. These include a network for capacity development, a global fund for AI and a global AI data framework.

7 To enable those efforts, to partner with other initiatives and institutions on addressing AI concerns and opportunities and ensure that the United Nations system speaks with one voice on AI, we propose the creation of an AI office within the United Nations Secretariat.

8 While we have considered the possibility of recommending the creation of an international agency for AI, we are not recommending this action currently; yet we acknowledge the need for governance to keep pace with technological evolution.

1 See https://un.org/ai-advisory-body.
2 See annex C for an overview of the consultations.

**9** Beyond immediate multilateral debates and processes involving Governments, our report is also intended for civil society and the private sector, researchers and concerned people around the world. We are acutely aware that achieving the ambitious goals that we have outlined can only happen with multisector global participation.

**10** Overall, we believe that the future of this technology is still open. This has been corroborated by our deep dive into the direction of technology and the debate between open and closed approaches to its development (see box 9). Larger and more powerful models developed in fewer and fewer corporations is one alternative future. Another could be a more diverse global innovation landscape dominated by interoperable small to medium-sized AI models delivering a multitude of societal and economic applications. Our recommendations seek to make the latter more likely, while also acknowledging the risks.

**11** From its founding, the United Nations has been committed to promoting the economic and social advancement of all peoples.[3] The Millennium Development Goals sought to establish ambitious targets so that economic opportunities are made available to all the world's people; the Sustainable Development Goals (SDGs) then sought to reconcile the need for development with the environmental constraints of our planet. The expanded development, deployment and use of AI tools and systems pose the next great challenge to ensuring that we embrace our digital future together, rather than widening our digital divide.

**12** Inclusive AI governance is, arguably, one of the most difficult governance challenges the United Nations will face. There is a mismatch between the dominant role of the private sector in AI and the Westphalian system of international politics. States are tempted by AI's potential for power and prosperity, at a time of intense geopolitical competition. Many societies are still at the margins of AI development, deployment and use, while a few are gripped by excitement mixed with concern at AI's cross-cutting impact.

**13** Despite the challenges, there is no opt-out. The stakes are simply too high for the United Nations, its Member States and the wider community whose aspirations the United Nations represents. We hope that this report provides some signposts to help our concerted efforts to govern AI for humanity.

# A. Opportunities and enablers

**14** AI is transforming our world. This suite of technologies[4] offers tremendous potential for good, from opening new areas of scientific inquiry (see box 1) and optimizing energy grids, to improving public health or agriculture.[5] If realized, the potential opportunities afforded by the use of AI tools for individuals, sectors of the economy, scientific research and other domains of public interest could play important roles in boosting our economies (see box 2), as well as transforming our societies for the better. Public interest AI – such as forecasting of and addressing pandemics, floods, wildfires and food insecurity – could even help to drive progress on the SDGs.

3 This included through trade, foreign direct investment and technology transfer as enablers for long-term development.

4 According to the Organisation for Economic Co-operation and Development (OECD), "An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment" (see https://oecd.ai/en/wonk/ai-system-definition-update).

5 We believe, however, that rigorous assessment by domain experts is needed to assess claims of AI's benefits. Pursuit of AI for good should be based on scientific evidence and a thorough evaluation of trade-offs and alternatives. In addition to scientific inquiry, the social sciences are also being transformed.

## Box 1: Potential of AI in advancing science

AI could well be the next major leap in scientific advancement, building on the transformative legacy of the Internet. The World Wide Web facilitated the sharing of vast amounts of experimental data, scientific papers and documentation among scientists. AI is building on this foundation by enabling the analysis of extensive data sets, uncovering hidden patterns, building new hypotheses and associations and accelerating the pace of discovery, including via experiments at scale with automated robotics.

The impact of AI on science spans major disciplines. From biology to physics, and from environmental science to social sciences, AI is being integrated in research workflows, and is accelerating the production of scientific knowledge. Some of the claims today might be hyped, while others have been demonstrated, and its long-term potential appears promising.[a]

For example, in biology, the 50-year challenge of protein-folding and protein structure prediction has been addressed with AI. This includes predicting the structure of over 200 million proteins, with the resulting open-access database being used by over 2 million scientists in over 190 countries at the time of writing, many of them working on neglected diseases. This has since been extended to life's other biomolecules, DNA, RNA and ligands and their interactions.

For Alzheimer's, Parkinson's and amyotrophic lateral sclerosis (ALS), experts using AI are identifying disease biomarkers and predicting treatment responses, significantly improving precision and speed of diagnosis and treatment development.[b] Broadly, AI is helping in advance precision medicine (e.g. in neurodegenerative diseases) by tailoring treatments based on genetic and clinical profiles. AI technology is also helping to accelerate the discovery and development of new chemical compounds.[c]

In radio astronomy, the speed and scale of data being collected by modern instruments, such as the Square Kilometre Array, can overwhelm traditional methods. AI can make a difference, including by helping to select which part of the data to focus on for novel insights. Through "unsupervised clustering", AI can pick out patterns in data without being told what specifically to look for.[d] Applying AI to social science research could also offer profound insights into complex human dynamics, enhancing our understanding of societal trends and economic developments.

In time, by enabling unprecedented levels of interdisciplinarity, AI may be designed and deployed to spawn new scientific domains, just as bioinformatics and neuroinformatics emerged from the integration of computational techniques with biological and neurological research. AI's ability to integrate and analyse diverse data sets from areas such as climate change, food security and public health could open research avenues that bridge these traditionally separate fields, if done responsibly.

AI may also enhance the public policy impact of scientific research by allowing for the validation of complex hypotheses, for example combining climate models with agricultural data to predict food security risks and linking these insights with public health outcomes. Another prospect is the boosting of citizen science and the leveraging of local knowledge and data for global challenges.

a See John Jumper and others, "Highly accurate protein structure prediction with AlphaFold", *Nature*, vol. 596 (July 2021), pp. 583–589; see also Josh Abramson and others, "Accurate structure prediction of biomolecular interactions with AlphaFold 3", *Nature*, vol. 630, pp. 493–500 (May 2024).
b Isaias Ghebrehiwet and others, "Revolutionizing personalized medicine with generative AI: a systematic review", *Artificial Intelligence Review*, vol. 57, No. 127 (April 2024).
c Amil Merchant and others, "Scaling deep learning for materials discovery", *Nature*, vol. 624, pp. 80–85 (November 2023).
d Zack Savitsky, "Astronomers are enlisting AI to prepare for a data downpour", *MIT Technology Review*, 20 May 2024.

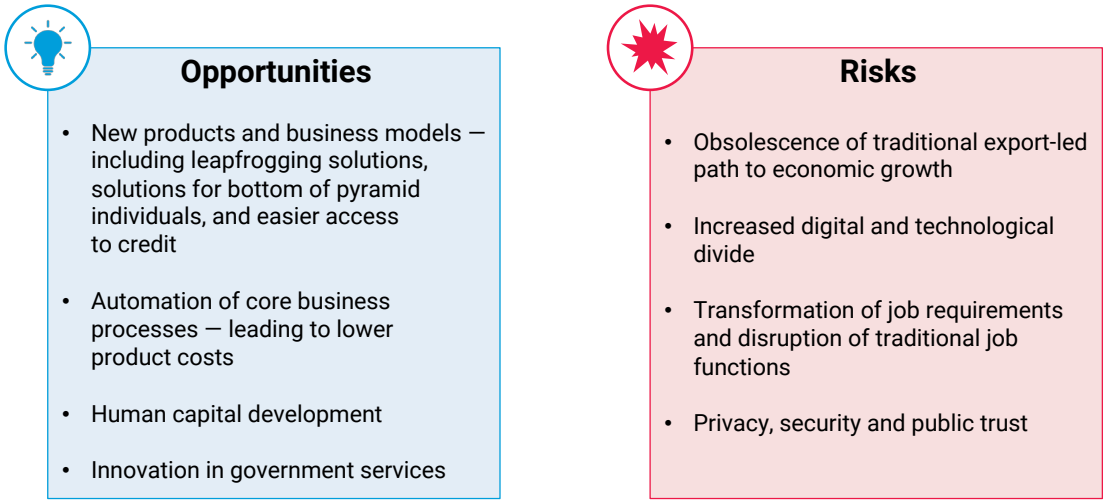
## Box 2: Economic opportunities of AI

Since the Industrial Revolution, a handful of innovations have dramatically accelerated economic progress. These earlier "general-purpose technologies" have reshaped multiple sectors and industries. The last major change came with computers and the digital age. These technologies transformed economies and increased productivity worldwide, but their full impact took decades to be felt.

Generative AI is breaking the trend of slow adoption. Experts believe its transformative effects will be seen within this decade. This quick integration means new developments in AI could rapidly reshape industries, change work processes and increase productivity. The rapid adoption of AI may thus transform our economies and societies in unprecedented ways.

The economic benefits of AI may be considerable. Although it is difficult to predict all the ramifications of AI on our complex economies, projections indicate that AI could significantly increase global gross domestic product, with relevant impacts across almost all sectors. For businesses, especially micro and small and medium-sized enterprises, AI can offer access to advanced analytics and automation tools, which were previously only available to larger corporations. The wide applicability of AI suggests that AI could be a general-purpose technology. As such, AI could enable productivity for individuals, small and large businesses, and other organizations in sectors as diverse as retail, manufacturing and operations, health care and the public sector, in developed and developing economies.[a] They will require broad adoption within and across sectors; application in productivity-enhancing uses; and AI that makes workers more productive and ushers in new economic activities at scale. They will also require investment and capital deepening, co-innovations, process and organizational changes, workforce readiness and enabling policies.

### Figure 1: Selected development opportunities and risks from AI in emerging markets




*Source:* International Finance Corporation.

a James Manyika and Michael Spence, "The coming AI economic revolution: can artificial intelligence reverse the productivity slowdown?", *Foreign Affairs*, 24 October 2023.

## Box 2: Economic opportunities of AI (continued)

Nevertheless, while AI can enhance productivity, boost international trade and increase income, it is also expected to impact work. Research suggests that AI could be assistive to workers in some cases, and job displacement in others cases.[b] Research, including by the International Labour Organization (ILO), suggests that in the foreseeable future, AI is likely to be more worker-assistive than worker-displacing.[c]

Research has also shown that when it occurs, job displacement is expected to occur differently in economies at different stages of development.[d] While advanced economies are more exposed, they are also better prepared to harness AI and complement their workforce. Low- and middle-income countries may have fewer capabilities to leverage this technology. Additionally, the integration of AI in the workforce may disproportionately affect certain demographics, with women potentially facing a higher risk of job displacement in some sectors.

Without focused and coordinated efforts to close the digital divide, AI's potential ability to be harnessed in support of sustainable development and poverty alleviation will not be realized, causing large segments of the global population to remain disadvantaged in the swiftly changing technological environment, with exacerbation of existing inequalities.

To successfully integrate AI into the global economy, we need effective governance that manages risks and ensures fair outcomes. This means among other options creating regulatory sandboxes for testing AI systems, promoting international cooperation on standards and setting up mechanisms to continuously evaluate AI's impact on labour markets and society. Apart from sound national AI strategies and international support, it specifically requires:

- Skills development: Implementing education and training programmes to develop AI skills across the workforce, from basic digital literacy to advanced technical expertise, to prepare workers for an AI-augmented future.
- Digital infrastructure: Significant investment in digital infrastructure, especially in developing countries, to bridge the AI divide and facilitate widespread AI adoption.
- Workplace integration: Leveraging social dialogue and public-private partnerships for managing AI integration in the workplace, ensuring worker participation in the process and protecting labour rights.
- Value chain considerations: Ensuring decent work conditions along the entire AI value chain, including often overlooked areas, such as data annotation and content moderation, for equitable AI development.

b Erik Brynjolfsson and others, "Generative AI at work", National Bureau of Economic Research, working paper 31161, 2023; see also Shakked Noy and Whitney Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence", Science, vol. 381, No. 6654, pp. 187–192 (July 2023).

c Pawel Gmyrek and others, Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality (Geneva: ILO, 2023).

d Mauro Cazzaniga and others, "Gen-AI: artificial intelligence and the future of work", staff discussion note SDN2024/001 (Washington, D.C.: International Monetary Fund, 2024).

## B. Key enablers for harnessing AI for humanity

15 The potential opportunities emerging from the development and use of AI will not necessarily be realized or pursued equitably. In May 2024, an analysis of funding for AI projects to advance progress towards completion of the SDGs found only 10 per cent of grants allocated had gone to organizations based in low- or middle-income countries; for private capital, the figure was 25 per cent (over 90 per cent of which in China).[6]

## C. Governance as a key enabler

16 Enablers need to be in place globally for the benefits of AI to be fully realized and accrued beyond a few people in a few countries. Ensuring that AI is deployed for the common good, and that its opportunities are distributed equitably, will require governmental and intergovernmental action to incentivize participation from the private sector, academia and civil society. Any governance framework should shape incentives globally to promote larger and more inclusive objectives and to help identify and address trade-offs.

## D. Risks and challenges

17 The development, deployment and use of AI bring risks, which can span many areas at the same time. We conceptualize AI-related risks in relation to vulnerabilities; this offers a vulnerability-based way to define policy agendas.

18 Challenges to traditional regulatory systems arise from AI's speed, opacity and autonomy. AI's accelerating technical development and deployment also raise the stakes for international governance, its general-purpose nature having implications across borders for multiple domains simultaneously.

## E. Risks of AI

19 Problems such as bias in AI systems and invidious AI-enabled surveillance are increasingly documented. Other risks are associated with the use of advanced AI, such as the confabulations of large language models, high resource consumption and risks to peace and security. AI-generated disinformation threatens democratic institutions.

20 Putting together a comprehensive list of AI risks for all time is a fool's errand, given the ubiquitous and rapidly evolving nature of AI and its uses; we believe that it is more useful to look at risks from the perspective of vulnerable communities and the commons (see paras. 26–28 below).

21 A snapshot of current expert risk perceptions is illustrated by the results of a horizon-scanning exercise commissioned for our work (AI Risk Global Pulse Check; see annex E), a poll which sourced perceptions on AI-related trends and risks from 348 AI experts across disciplines and 68 countries in all regions.[7] Overall, 7 in 10 experts polled were concerned or very concerned that harms (existing or new) resulting from AI will become substantially more serious and/or widespread in the next 18 months (see annex E).

6 "An analysis of the location of grant recipients' headquarters from a database of US-majority foundations reveals that from 2018 to 2023, only 10 percent of grants allocated toward AI initiatives that address one or more of the SDGs went to organizations based in low- or middle-income countries … Analysis of private capital shows that 36 percent of 9,000 companies addressing SDGs are headquartered in the United States, but these companies received 54 percent of total funding. We also found that while 20 percent of 9,000 companies addressing SDGs are headquartered in lower- or middle-income countries, they received a higher proportion (25 percent) of total funding. One reason for this is that Chinese companies receive a high proportion of investment … The remaining developing countries in the sample received only 3 percent of funding while representing 7 percent of the sample" (Medha Bankhwal and others, "AI for social good: improving lives and protecting the planet", McKinsey & Company, May 2024).

7 The invitee list was constructed from the Office of the Secretary-General's Envoy on Technology (OSET) and the Advisory Body's networks, including participants in deep dives. Additional experts were regularly invited during the fielding period to improve representation. The final n=348 represents a strong, balanced global sample of respondents with relevant expertise to provide an informed opinion on AI risks (see annex E for the methodology).

## Figure 2: Experts' levels of concern about AI risks across multiple domains


*Note:* Excludes "Don't know" / "No opinion" and blank responses.
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

**22** From a list of example AI-related risk areas,[8] a plurality of experts were concerned or very concerned about harms related to:

a. Societal implications of AI: **78 per cent** regarding damage to information integrity [question j], **74 per cent** regarding inequalities such as concentration of wealth and power in a few hands [question l] and **67 per cent** regarding discrimination / disenfranchisement, particularly among marginalized communities [question i];

b. Intentional use of AI that harms others: **75 per cent** regarding use in armed conflict by State actors [question b], **72 per cent** regarding malicious use by non-State actors [question a] and **65 per cent** regarding use by State actors that harms individuals [question c].

**23** In all but two example risk areas, most AI experts polled were concerned or very concerned about harms materializing. Although fewer than half of experts expressed such concern regarding unintended harms from AI [questions e and f], 1 in 6 of those who were very concerned about unintended AI harms mentioned that they expected agentic systems to have some of the most surprising or significant impacts on AI-related risks by 2025.[9]

**24** Expert perceptions varied, including by region and gender (see annex E for more detailed results). This highlighted the importance of inclusive representation in exercises concerning definition of shared risks. Despite the variation, the results did reveal concerns about AI harms over the coming year, highlighting a sense of urgency among experts to address risks across multiple areas and vulnerabilities in the near future.

8 Built on the vulnerability-based risk categorization in box 4, an earlier version of which was in our interim report.
9 Question: "What emerging trends today do you think could have the most surprising and/or significant impact on AI-related risks over the next 18 months?"

**25** Moreover, autonomous weapons in armed conflict, crime or terrorism, and public-security use of AI in particular, raise serious legal, security and humanitarian questions (see box 3).[10]

**26** Risk management requires going beyond listing or prioritizing risks, however. Framing risks based on vulnerabilities can shift the focus of policy agendas from the "what" of each risk (e.g. "risk to safety") to "who" is at risk and "where", as well as who should be accountable in each case.

## Box 3: AI and national and international security

Many AI technologies are not simply dual-use but inherently "re-purposable". AI applications for law enforcement and border controls are growing and raise concerns about due process, surveillance and lack of accountability regarding States' commitments to human rights norms, enshrined in the Universal Declaration of Human Rights and other instruments.

Among the challenges of AI use in the military domain are new arms races, the lowering of the threshold of conflict, the blurring of lines between war and peace, proliferation to non-State actors and derogation from long-established principles of international humanitarian law, such as military necessity, distinction, proportionality and limitation of unnecessary suffering. On legal and moral grounds, kill decisions should not be automated through AI. States should commit to refraining from deploying and using military applications of AI in armed conflict in ways that are not in full compliance with international law, including international humanitarian law and human rights law.

Presently, 120 Member States support a new treaty on autonomous weapons, and both the Secretary-General and the President of the International Committee of the Red Cross have called for such treaty negotiations to be completed by 2026. The Advisory Body urges Member States to follow up on this call.

The Advisory Body considers it essential to identify clear red lines delineating unlawful use cases, including relying on AI to select and engage targets autonomously. Building on existing commitments on weapons reviews in international humanitarian law, States should require weapons manufacturers through contractual obligations and other means to conduct legal and technical reviews to prevent unethical design and development of military applications of AI. States should also develop legal and technical reviews of the use of AI, as well as of weapons and means of warfare and sharing related best practices.

Furthermore, States should develop common understandings relating to testing, evaluation, verification and validation mechanisms for AI in the security and military domain. They should cooperate to build capacity and share knowledge by exchanging good practices and promoting responsible life cycle management of AI applications in the security and military domain. To prevent acquisition of powerful and potentially autonomous AI systems by dangerous non-State actors, such as criminal or terrorist groups, States should set up appropriate controls and processes throughout the life cycle of AI systems, including managing end-of-life cycle processes (i.e. decommissioning) of military AI applications.

For transparency, "advisory boards" could be set up to provide independent expert advice and scrutiny across the full life cycle of security and military applications of AI. Industry and other actors should consider mechanisms to prevent the misuse of AI technology for malicious or unintended military purposes.

10 This list is intended to be illustrative only, touching on only a few of the risks facing individuals and societies.

27 This is significant, as evolving risks manifest differently for different people and societies. A vulnerability-based approach, also proposed in our interim report, offers an open-ended framework for focusing on those who could be harmed by AI, which can be a foundation for dynamic risk management (see box 4).

## Box 4: Categorizing AI-related risks based on existing or potential vulnerability

Individuals

- Human dignity, value or agency (e.g. manipulation, deception, nudging, sentencing, exploitation, discrimination, equal treatment, prosecution, surveillance, loss of human autonomy and AI-assisted targeting).
- Physical and mental integrity, health, safety and security (e.g. nudging, loneliness and isolation, neurotechnology, lethal autonomous weapons, autonomous cars, medical diagnostics, access to health care, and interaction with chemical, biological, radiological and nuclear systems).
- Life opportunities (e.g. education, jobs and housing).
- (Other) human rights and civil liberties, such as the rights to presumption of innocence (e.g. predictive policing), the right to a fair trial (e.g. recidivism prediction, culpability, recidivism, prediction and autonomous trials), freedom of expression and information (e.g. nudging, personalized information, info bubbles), privacy (e.g. facial recognition technology), and freedom of assembly and movement (e.g. tracking technology in public spaces).

Politics and society

- Discrimination and unfair treatment of groups, including based on individual or group traits, such as gender, group isolation and marginalization.
- Differential impact on children, older persons, persons with disabilities and vulnerable groups.
- International and national security (e.g. autonomous weapons, policing and border control vis-à-vis migrants and refugees, organized crime, terrorism and conflict proliferation and escalation).
- Democracy (e.g. elections and trust).
- Information integrity (e.g. misinformation or disinformation, deepfakes and personalized news).
- Rule of law (e.g. functioning of and trust in institutions, law enforcement and the judiciary).
- Cultural diversity and shifts in human relationships (e.g. homogeneity and fake friends).
- Social cohesion (e.g. filter bubbles, declining trust in institutions, and information sources).
- Values and norms (e.g. ethical, moral, cultural and legal).

Economy

- Power concentration.
- Technological dependency.
- Unequal economic opportunities, market access, resource distribution and allocation.
- Underuse of AI.
- Overuse of AI or “technosolutionism”.
- Stability of financial systems, critical infrastructure and institutions.
- Intellectual property protection.

Environment

- Excessive consumption of energy, water and material resources (including rare minerals and other natural resources).

28 The policy-relevance of taking a vulnerability-based lens to AI-related risks is illustrated by examining AI governance considerations from the perspective of a particular vulnerable group, such as children (see box 5).

29 The individuals, groups or entities of concern identified via a vulnerability-based framing of AI risks – and implied policy agendas – can themselves vary. The AI Risk Global Pulse Check also asked experts which individuals, groups, societies/economies/(eco)systems they were particularly concerned would be harmed by AI in the next 18 months. Marginalized communities and the global South, along with children, women, youths, creatives and those with jobs susceptible to automation, were particularly highlighted (see fig. 3).

## Box 5: Focusing on children in AI governance

Ensuring that businesses and schools address the needs and rights of children requires a comprehensive governance approach that focuses on their unique circumstances. Children generate one third of the data and will grow up to an AI-infused economy and world accustomed to the use of AI. This box summarizes some of the measures relating to this topic discussed during our deep dives.

**Prioritizing children's rights and voices**:
AI governance must recognize children as priority stakeholders, emphasizing their right to develop free from the addictive effects of technology and their right to disengage from it. Unlike general human-centric approaches, child-centric governance must consider the long-term impacts on children's perspectives, self-image, and life choices and opportunities. Including children in design and governance processes is crucial to ensuring that AI systems are safe and appropriate for their use.

**Research and policy development**:
We need extensive research to understand how AI affects children's social, cognitive and emotional development over time. This research should inform policy discussions and guide protective measures across countries.

**Protection and privacy**:
Children should not be used as subjects for AI experimentation. Protecting children's privacy is paramount. AI technologies must incorporate stringent data protection protocols and provide age-appropriate content.

**Child impact assessments and child appropriate design**:
Mandating child impact assessments for AI systems is essential to ensuring their suitability and safety. AI systems should be designed with children's needs in mind, incorporating safety and restriction features from the start. Design choices should involve input from children themselves.

**Digital inclusion and equity**:
Access to AI should empower children with agency, choices and voice, emphasizing holistic approaches to digital inclusion. This includes providing AI content in multiple languages and ensuring that it is culturally appropriate for non-English-speaking children.

**International cooperation and standards**:
Global interoperability of rules for children's engagement with AI technologies is needed to protect children across different educational and developmental environments. Global standards will be essential to address cross-border data flows and ethical AI use for children.

## Figure 3: Concerns on vulnerability highlighted in the AI Risk Global Pulse Check

*"Are there specific individuals, groups or societies/economies/(eco)systems that you are particularly concerned may be harmed by AI over the next 18 months?" [free text response]* (n = 188 meaningful responses to this question)

**INDICATIVE**


*Note:* Keywords tagged for each response by OSET. Showing only keywords identified in 2+ responses. Font size is proportional to number of responses mentioned. For scale, "global South" was identified by 46 of 188 respondents who provided meaningful responses to this question; "marginalized communities" by 43 of 188.
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

**30** These results illustrate the importance of inclusive representation when reaching common understandings of AI risks and common ground on policy agendas, as per recommendations 1 and 2. Without such representation, AI governance policy agendas could be framed in ways that miss the concerns of portions of humanity, who will nonetheless be affected.

# F. Challenges to be addressed

**31** Besides near-future risks and harms, the evolution of AI development, deployment and uses also poses challenges in the context of prevailing institutions, which in turn affects strategies for AI governance. The technological pace around advanced AI – and its general-purpose nature – further tests humanity's ability to respond in time.

**32** The race to develop and deploy AI systems defies traditional regulatory systems and governance regimes. Most experts polled for the AI Risk Global Pulse Check expected AI acceleration over the next 18 months, both in its development (74 per cent) and adoption and application (89 per cent) (see fig. 4).

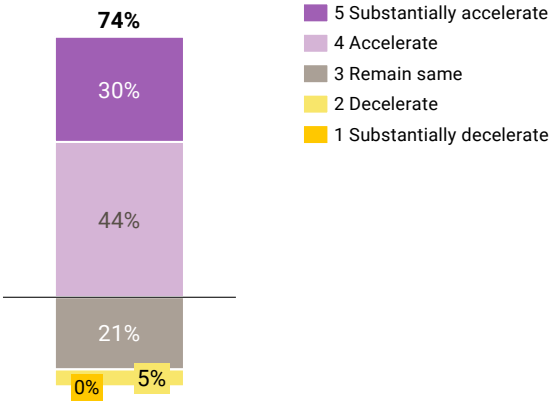
**33** As mentioned in paragraph 23, some experts expect the deployment of agentic systems in 2025. Moreover, leading technical experts acknowledge that many AI models remain opaque, with their outputs not fully predictable or controllable, even as negative spillovers downstream may impact others globally.

**34** Increasing reliance on automated decision-making and content-creation by opaque algorithms can undermine fair treatment and safety. While humans often remain legally accountable for decisions to automate processes that impact others, accountability mechanisms may not evolve quickly enough for such accountability to be given prompt and meaningful effect.

## Figure 4: Experts' expectations regarding AI technological development

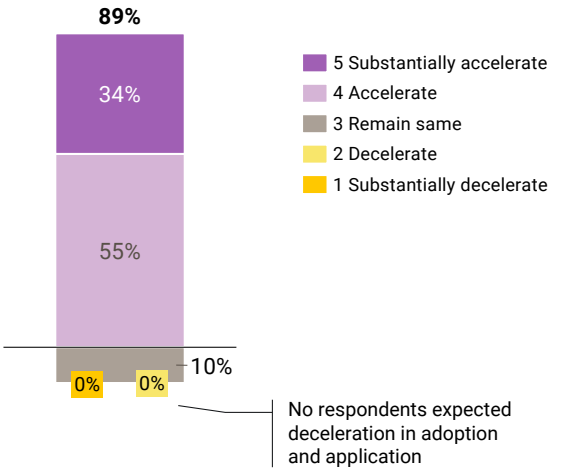
**74% expect pace of technical change to accelerate (30% substantially)**

*"In the next 18 months, compared to the last 3 months, do you expect the pace of technical change in AI (e.g. development / release of new models) to..."* (n = 348)




**89% expect pace of adoption & application to accelerate (34% substantially)**

*"In the next 18 months, compared to the last 3 months, do you expect the pace of adoption and application of AI (e.g. new uses of AI in business / government) to..."* (n = 348)




*Note:* Numbers may not add up to 100% owing to rounding. Excludes "Don't know" / "No opinion" and blank responses.
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

**35** A societal risk thus emerges that ever-fewer individuals end up being held accountable for harms arising from their decisions to automate processes using AI, even as increasingly powerful systems enter the world. This demands agile governance to ensure that accountability mechanisms keep pace with accelerating AI.

**36** If the pace of AI development and deployment challenges existing institutions, so does the breadth. A general-purpose technology with global reach, advanced AI can be deployed across domains affecting societies in manifold ways, with broad policy implications.

**37** The implications and potential impact of AI's intersection with multiple areas, including finance, labour markets, education and political systems, presage broad consequences that demand a whole-of-society approach (see examples in box 6). Existing institutions must mount holistic, cross-sectoral responses that address AI's wide-ranging societal impacts.

**38** The pace, breadth and uncertainty of AI's development, deployment and use highlight the value of a holistic, transversal and agile approach to AI. Internationally, a holistic perspective needs to be mirrored in a networked institutional approach to AI governance across sectors and borders, which engages stakeholders without being captured by them.

**39** On climate change, the world has come to realize only belatedly that a holistic approach to global collective action is needed. With AI, there is an opportunity to do so by design.

**40** The above challenges are compounded by an associated concentration of wealth and decision-making among a handful of private AI developers and deployers, particularly multinational corporations. This raises another question of how stakeholders can be engaged in AI's governance without undermining the public interest.

## Box 6: AI-related societal impacts

As part of its broader engagement, Advisory Body members consulted with a range of stakeholders to discuss the implications of AI on society. This box summarizes key concerns and potential initiatives brought forward as part of deep dives on this topic.

**Social, psychological and community impact**:

As AI becomes more powerful and widespread, its development, deployment and application will become more personalized, with the potential to foster alienation and addiction. To some Advisory Body members, AI trained on an individual's data, and its consequent role as a primary interlocutor and intermediary, may reflect an inflection point for human beings – with the potential to create urgent new societal challenges, while exacerbating existing ones.

For example, future AI systems may be able to generate an endless feed of high-quality video content tailored to individuals' personal preferences. Increased social isolation, alienation, mental health issues, loss of human agency and impacts on emotional intelligence and social development are only a few of the potential outcomes.

These issues are already insufficiently explored by policymakers in the context of technologies such as smart devices and the Internet; they are almost completely unexplored in the context of AI, with current governance frameworks prioritizing risks to individuals, rather than society as a whole.

As policymakers consider future responses to AI, they must weigh these factors as well, and develop policies that promote societal well-being, particularly for youth. Government interventions could foster environments that prioritize face-to-face interactions between humans, making mental health support more readily available, and investing more into sports facilities, public libraries and the arts.

Nevertheless, prevention is better than cure: industry developers should design their products without addictive personalized features, ensure that the products do not damage mental health and promote (rather than undermine) a sense of shared belonging in society. Tech companies should establish policies to manage societal risks on an equal basis to other risks as part of efforts to identify and mitigate risks across the entire life cycle of AI products.

**Disinformation and trust**:

Deepfakes, voice clones and automated disinformation campaigns pose a specific and serious threat to democratic institutions and processes such as elections, and to democratic societies and social trust more generally, including through foreign information manipulation and interference (FIMI). The development of closed loop information ecosystems, reinforced by AI and leveraging personal data, can have profound effects on societies, potentially making them more accepting of intolerance and violence towards others.

Protecting the integrity of representative government institutions and processes requires robust verification and deepfake detection systems, alongside rapid notice and take-down procedures for content that is likely to deceive in a way that causes harm or societal divisions, or which promotes war propaganda, conflict and hate speech. Individuals who are not public figures should have protections from others creating deepfakes in their likeness for fraudulent, defamatory or otherwise abusive purposes. Sexualized deepfakes are a particular concern for women and girls and may be a form of gender-based violence.

## Box 6: AI-related societal impacts (continued)

Voluntary commitments from private sector players – such as labelling deepfakes or enabling users to flag and then take down deepfakes made or distributed with malicious intent – are important first steps. However, they do not sufficiently mitigate societal risks. Instead, a global, multi-stakeholder approach is required, alongside binding commitments. Common standards for content authentication and digital provenance would allow for a globally recognized approach to identify synthetic and AI-modified images, videos and audio.

Additionally, real-time knowledge-sharing between public and private actors, based on international standards, would allow for rapid-response capabilities to immediately take down deceptive content or FIMI before it has a chance to go viral. Nonetheless, these processes should incorporate safeguards to ensure that they are not manipulated or abused to abet censorship.

These actions should be accompanied by preventive measures, to increase societal resilience to AI-driven disinformation and propaganda, such as public awareness campaigns on AI's potential to undermine information integrity. Member States should additionally promote media and digital literacy campaigns, support fact-checking initiatives and invest in capacity-building for the FIMI defender community.

# 2. The need for global governance

**41** There is, today, a global governance deficit with respect to AI. Despite much discussion of ethics and principles, the patchwork of norms, institutions and initiatives is still nascent and full of gaps. Accountability and remedies for harm are often notable primarily for their absence. Compliance rests on voluntarism. There is a fundamental disconnect between high-level rhetoric, the systems being developed, deployed and used, and the conditions required for safety and inclusiveness. As we noted in our interim report, AI governance is crucial, not merely to address the challenges and risks, but also to ensure that we harness their potential in ways that leave no one behind.[11]

**42** The imperative of global governance, in particular, is irrefutable. AI's raw materials, from critical minerals to training data, are globally sourced. General-purpose AI, deployed across borders, spawns manifold applications globally. The accelerating development of AI concentrates power and wealth on a global scale, with geopolitical and geoeconomic implications. Moreover, no one currently understands all of AI's inner workings enough to fully control its outputs or predict its evolution. Nor are decision makers held accountable for developing, deploying or using systems that they do not understand. Meanwhile, negative spillovers and downstream impacts resulting from such decisions are also likely to be global.

**43** Despite AI's global reach, national and regional institutional structures and regulations end at physical borders. This reduces the ability of any single country to govern the downstream applications of AI that result in transboundary harms, or to address issues along complex cross-border supply chains of compute infrastructure, training data flows and energy sources that lie behind AI's development and use. Leading AI companies often have more direct influence over downstream applications (via upstream risk mitigation) than most countries acting alone.

**44** The development, deployment and use of such a technology cannot be left to the whims of markets alone. National governments and regional organizations will be crucial. However, in addition to considerations of equity, access and prevention of and remedies for harm, the very nature of the technology itself – transboundary in structure and application – necessitates a global multisector approach. Without a globally inclusive framework that engages stakeholders, and given the competitive dynamics at play, both Governments and companies might be tempted to cut corners or to prioritize self-interest.

**45** AI, therefore, presents global challenges and opportunities that require a holistic and global approach that cuts transversally across political, economic, social, ethical, human rights, technical, environmental and other domains. Such an approach can turn a patchwork of evolving initiatives into a coherent, interoperable whole, grounded in international law and adaptable across contexts and time.

**46** The need for global governance of AI arises at a time of geopolitical and geoeconomic competition for influence and markets. Yet addressing AI's risks while enabling opportunities to be harnessed equitably requires concerted global action. A widening digital divide could limit the benefits of AI to a handful of States and individuals, with risks and harms impacting many, especially vulnerable, groups.

11 See https://un.org/ai-advisory-body.

# A. Guiding principles and functions for international governance of AI

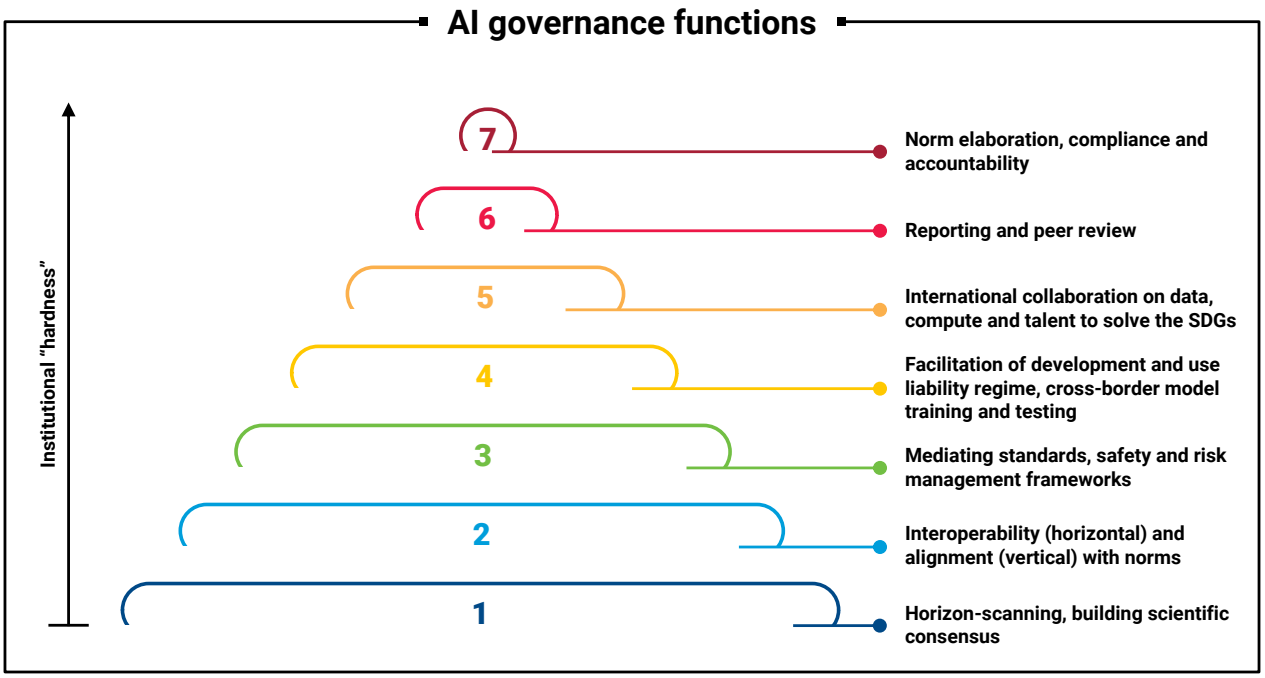
47 In our interim report, we outlined five principles that should guide the formation of new international AI governance institutions:

- **Guiding principle 1**: AI should be governed inclusively, by and for the benefit of all
- **Guiding principle 2**: AI must be governed in the public interest
- **Guiding principle 3**: AI governance should be built in step with data governance and the promotion of data commons
- **Guiding principle 4**: AI governance must be universal, networked and rooted in adaptive multi-stakeholder collaboration
- **Guiding principle 5**: AI governance should be anchored in the Charter of the United Nations, international human rights law and other agreed international commitments such as the SDGs

48 Box 7 summarizes the feedback on these principles, which emphasized the importance of human rights and the need for greater clarity on effective implementation of the guiding principles, including regarding data governance. It challenged us to address the problem of ensuring that support for inclusivity was backed by action, and that marginalized groups would be represented.

49 In our interim report, we also proposed several institutional functions that might be pursued at the international level (see fig. 5). The feedback largely confirmed the need for these functions at the global level, while calling for additional complementary functions related to data and AI governance to translate guiding principle 3 (AI governance should be built in step with data governance and the promotion of data commons) into practice.

## Figure 5: AI governance functions proposed at the international level



## Box 7: Feedback on the guiding principles

**Emphasis on human rights-based AI governance**:

Based on the extensive consultations conducted by the High-level Advisory Body following the publication of its interim report, guiding principle 5 (AI governance should be anchored in the Charter of the United Nations, international human rights law and other agreed international commitments) garnered the strongest support across all sectors of stakeholders, including governments, civil society, the technical community, academia and the private sector. This included respecting, promoting and fulfilling human rights and prosecuting their violations, as well as General Assembly resolution 78/265 on seizing the opportunities of safe, secure and trustworthy AI systems for sustainable development, unanimously adopted in March 2024.

The Advisory Body in its deliberations was convinced that to mitigate the risks and harms of AI, to deal with novel use cases and to ensure that AI can truly benefit all of humanity and leave no one behind, human rights must be at the centre of AI governance, ensuring rights-based accountability across jurisdictions. This foundational commitment to human rights is cross-cutting and applies to all the recommendations made in this final report.

**Specific implementation mechanisms and clarity on guidelines**:

Many stakeholders emphasized the need for detailed action plans and clear guidelines to ensure effective implementation of the Advisory Body's guiding principles for international AI governance. Governmental entities suggested developing clear recommendations for defining and ensuring the public interest, along with mechanisms for public participation and oversight. The need for clear policies and leveraging existing regulatory frameworks to maintain competitive and innovative AI markets was often stressed by private sector entities. Many international organizations and civil society organizations also called for agile governance systems designed to respond in a timely manner to evolving technologies. Some specifically requested a new entity with "muscle and teeth", beyond mere coordination.

**Mechanisms to hold key actors responsible**:

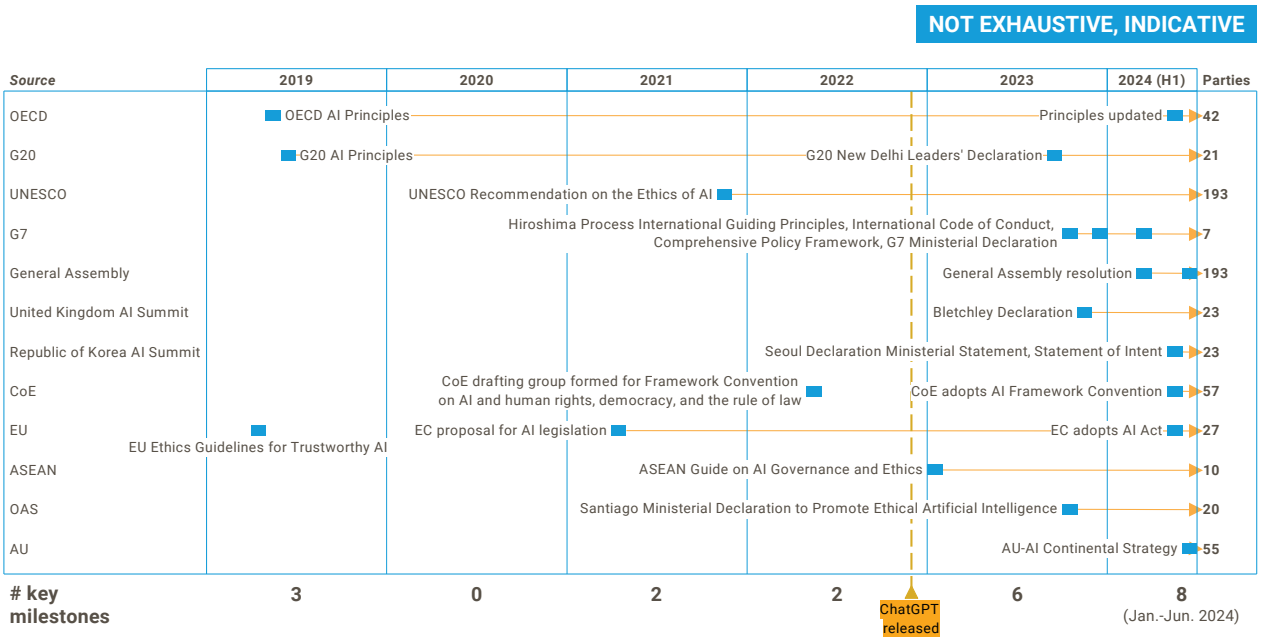
A common concern was accountability for discriminatory, biased and otherwise harmful AI, with suggestions for mechanisms to ensure accountability and remedies for harm and address the concentration of technological capacity and market power. Many organizations highlighted the necessity of addressing unchecked power and ensuring consumer rights and fair competition. Academic institutions recognized the strengths of the guiding principles in their universality and inclusivity, but suggested improvements in stakeholder engagement. Private sector actors emphasized responsible use of AI, along with breaking down barriers to access.

**More specific functions on AI data governance**:

The absence of data governance systems was mentioned in multiple consultations, with stakeholders indicating that the United Nations was a natural venue for dialogue on data governance. Governments emphasized the need for robust data governance frameworks that prioritized privacy, data protection and equitable data use, advocating for international guidelines to manage data complexities in AI development. The frameworks were requested to be developed through a transparent and inclusive process, integrating ethical considerations such as consent and privacy.

Academia highlighted that data governance should be dealt with as a priority in the short term. Private sector entities noted that data governance measures should complement AI governance, emphasizing comprehensive privacy laws and responsible AI use. International organizations and civil society organizations stressed that governance of AI training data should protect consumer rights and support fair competition among AI developers via non-exclusive access to AI training data, underscoring the call for specific and actionable data governance measures. The United Nations was identified as a key venue for addressing these governance challenges and bridging resource disparities.

## Figure 6: Interregional and regional AI governance initiatives, key milestones, 2019–2024 (H1)




*Abbreviations*: ASEAN, Association of Southeast Asian Nations; AU, African Union; CoE, Council of Europe; EU, European Union; G20, Group of 20; G7, Group of Seven; GPAI, Global Partnership on Artificial Intelligence; OAS, Organization of American States; OECD, Organisation for Economic Co-operation and Development; UNESCO, United Nations Educational, Scientific and Cultural Organization.

**50** Regarding the institutionally "harder" AI governance functions of monitoring, verification, reporting, compliance, accountability stabilization, response and enforcement, the feedback noted that first, international treaty obligations would be needed, prior to the institutionalization of such functions, and that the case for institutionalizing such functions in governing AI as a technology was not yet made.

**51** Not all functions need to be performed exclusively by the United Nations. However, if the patchwork of norms and institutions is to be transformed into a safety net that promotes and supports sustainable innovation benefiting all of humanity, then there needs to be a shared understanding of the science and common ground behind the rules and the standards by which we assess whether governance is achieving its objectives.

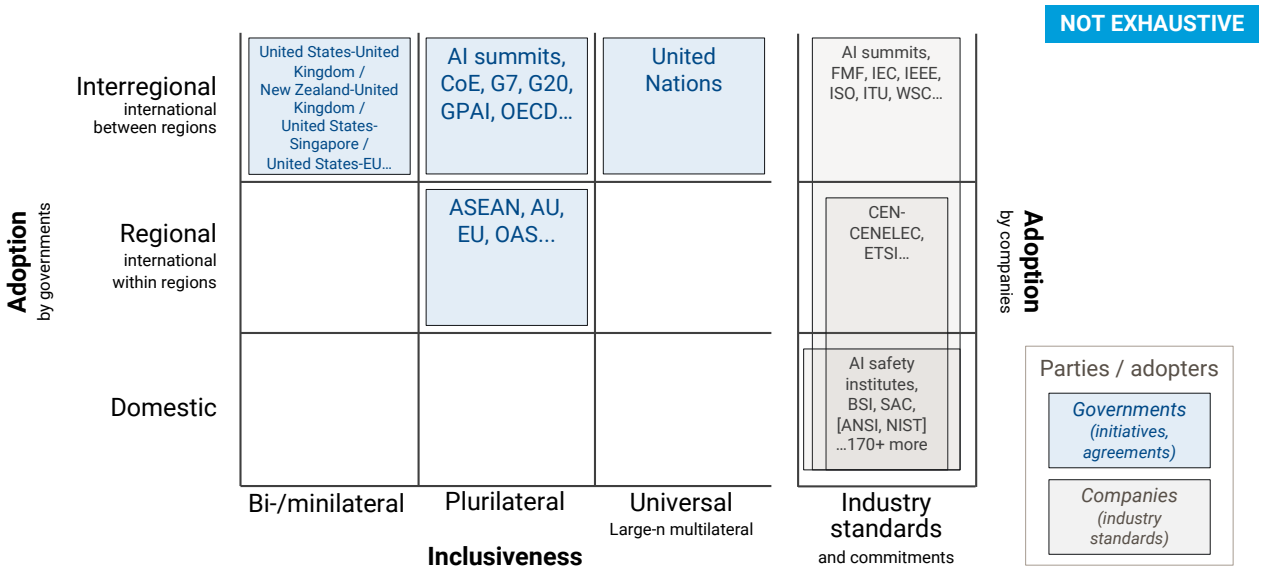
**52** During our consultations, we heard calls for a more detailed landscape analysis of existing and emerging efforts to govern AI internationally, and of gaps needing to be filled for the equitable, effective and efficient international governance of AI.

# B. Emerging international AI governance landscape

**53** There is, to be sure, no shortage of documents and dialogues presently focused on AI governance. Hundreds of guides, frameworks and principles have been adopted by governments, companies and consortiums, and by regional and international organizations. Dozens of forums convene diverse actors, from established intergovernmental processes and expert bodies, to ad hoc multi-stakeholder initiatives. These are accompanied by existing and emerging regulation at the national and regional levels.

**54** International initiatives by Governments are proliferating (see fig. 6). These emerging initiatives increasingly follow a transversal approach to AI governance at the international level, consisting of principles, declarations, statements and other issuances that address AI holistically, rather than in specific domains. They have accelerated sharply

## Figure 7: Sources of governance initiatives that focused on AI specifically




*Abbreviations*: ANSI, American National Standards Institute; ASEAN, Association of Southeast Asian Nations; AU, African Union; BSI, British Standards Institution; CEN, European Committee for Standardisation; CENELEC, European Committee for Electrotechnical Standardization; CoE, Council of Europe; ETSI, European Telecommunications Standards Institute; EU, European Union; FMF, Frontier Model Forum; G20, Group of 20; G7, Group of Seven; GPAI, Global Partnership on Artificial Intelligence; IEC, International Electrotechnical Commission; IEEE, Institute of Electrical and Electronics Engineers; ISO, International Organization for Standardization; ITU, International Telecommunication Union; NIST, National Institute of Standards and Technology; OAS, Organization of American States; OECD, Organisation for Economic Co-operation and Development; SAC, Standardization Administration of China; WSC, World Standards Cooperation.

since 2023, spurred by releases of multiple general-purpose AI large language models following the release of ChatGPT in November 2022.

**55** In parallel, industry standards on AI have been developed and published for adoption internationally. Other multi-stakeholder initiatives have also sought to bridge the divide between the public and private sectors, including in discussion arenas such as the Internet Governance Forum.

**56** A survey of some of the sources of AI governance initiatives and industry standards, mapped by geographical range and inclusiveness, is provided in figure 7 (in listing this recent work, we acknowledge many years of efforts by academics, civil society and professional bodies).

**57** Examples of relevant regional and interregional plurilateral initiatives include those led by the African Union, various hosts of AI summits, the Association of Southeast Asian Nations, the Council of Europe, the European Union, the Group of Seven (G7), the Group of 20 (G20), the Global Partnership on Artificial Intelligence, the Organization of American States and the Organisation for Economic Co-operation and Development (OECD), among others.

**58** Our analysis of current governance arrangements is likely to be outdated within months. Nevertheless, it can help to illustrate how current and emerging international AI governance initiatives relate to our guiding principles for the formation of new global governance institutions for AI, including principle 1 (AI should be governed inclusively, by and for the benefit of all).

# 3. Global AI governance gaps

**59** The multiple national, regional, multi-stakeholder and other initiatives mentioned above have yielded meaningful gains and informed our work; many of their representatives have contributed to our deliberations in writing or participated in our consultations.

**60** Nonetheless, beyond a couple of initiatives emerging from the United Nations,[12] none of the initiatives can be truly global in reach. These representation gaps in AI governance at the international level are a problem, because the technology is global and will be comprehensive in its impact.

**61** Separate coordination gaps between initiatives and institutions risk splitting the world into disconnected and incompatible AI governance regimes.

**62** Furthermore, implementation and accountability gaps reduce the ability of States, the private sector, civil society, academia and the technical community to translate commitments, however representative, into tangible outcomes.

## A. Representation gaps

**63** Our analysis of the various non-United Nations AI governance initiatives that span regions shows that most initiatives are not fully representative in their intergovernmental dimensions.

**64** Many exclude entire parts of the world. As figure 8 shows, looking at seven non-United Nations plurilateral, interregional AI initiatives with overlapping membership, seven countries are parties to all of them, whereas fully 118 countries are parties to none (primarily in the global South, with uneven representation even of leading AI nations; see fig. 8).

**65** Selectivity is understandable at an early stage of governance when there is a degree of experimentation, competition around norms and diverse levels of comfort with new technologies. However, as international AI governance matures, global representation becomes more important in terms of equity and effectiveness.

**66** Besides the non-inclusiveness of existing efforts, representation gaps also exist in national and regional initiatives focused on reaching common scientific understandings of AI. These representation gaps may manifest in decision-making processes regarding how assessments are scoped, resourced and conducted.

**67** Equity demands that more voices play meaningful roles in decisions about how to govern technology that affects all of us, as well as recognizing that many communities have historically been excluded from those conversations. The relative paucity of topics from the agendas of major initiatives that are priorities of certain regions signals an imbalance stemming from underrepresentation.[13]

**68** AI governance regimes must span the globe to be effective – effective in building trust, averting "AI arms races" or "races to the bottom" on safety and rights, responding effectively to challenges arising

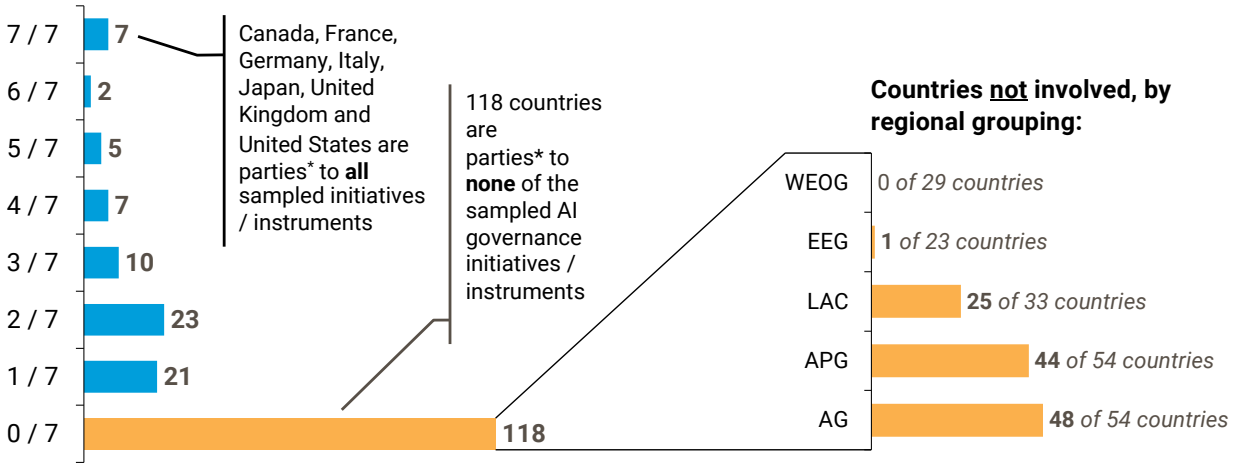
12 The United Nations Educational, Scientific and Cultural Organization (UNESCO) Recommendation on the Ethics of Artificial Intelligence (2021), and two General Assembly resolutions on AI.

13 For example, governance of AI training data sets, access to computational power, AI capacity development, AI-related risks regarding discrimination of marginalized groups and use of AI in armed conflict (see annex E for results of the AI Risk Global Pulse Check, which shows different perceptions of risks by respondents from the Western European and Others Group versus others). Many States and marginalized communities have also been excluded from the benefits of AI or may disproportionately suffer its harms. Equity demands a diverse and inclusive approach that accounts for the views of all regions and that spreads opportunities evenly while mitigating risks.

## Figure 8: Representation in seven non-United Nations international AI governance initiatives

*Sample*: OECD AI Principles (2019), G20 AI principles (2019), Council of Europe AI Convention drafting group (2022–2024), GPAI Ministerial Declaration (2022), G7 Ministers' Statement (2023), Bletchley Declaration (2023) and Seoul Ministerial Declaration (2024).

**INTERREGIONAL ONLY, EXCLUDES REGIONAL**




* Per endorsement of relevant intergovernmental issuances. Countries are not considered involved in a plurilateral initiative solely because of membership in the European Union or the African Union. *Abbreviations*: AG, African Group; APG, Asia and the Pacific Group; EEG, Eastern European Group; G20, Group of 20; G7, Group of Seven; GPAI, Global Partnership on Artificial Intelligence; LAC, Latin America and the Caribbean; OECD, Organisation for Economic Co-operation and Development; WEOG, Western European and Others Group.

from the transboundary character of AI, spurring learning, encouraging interoperability and sharing AI benefits.[14] There are, moreover, benefits to including diverse views, including un-likeminded views, to anticipate threats and calibrate responses that are creative and adaptable.

**69** By limiting the range of countries included in key agenda-shaping, relationship-building and information-sharing processes, selective plurilateralism can limit the achievement of its own goals. These include compatibility of emerging AI governance approaches, global AI safety and shared understandings regarding the science of AI at the global level (see recommendations 1, 2 and 3 on what makes a global approach particularly effective here).

**70** The two General Assembly resolutions on AI adopted in 2024 so far[15] signal acknowledgement among leading AI nations that representation gaps need to be addressed regarding international AI governance, and the United Nations could be the forum to bring the world together in this regard.

**71** The Global Digital Compact in September 2024, and the World Summit on the Information Society Forum in 2025 offer two additional policy windows where a globally representative set of AI governance processes could be institutionalized to address representation gaps.[16]

14 If and when red lines are established – analogous perhaps to the ban on human cloning – they will only be enforceable if there is global buy-in to the norm, as well as monitoring compliance. This remains the case despite the fact that, paradoxically, in the current paradigm, while the costs of a given AI system go down, the costs of advanced AI systems (arguably the most important to control) go up.

15 Resolutions 78/265 (seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development) and 78/311 (enhancing international cooperation on capacity-building of artificial intelligence).

16 Various plurilateral initiatives, including the OECD AI Principles, the G7 Hiroshima AI Process and the Council of Europe Framework Convention on Artificial Intelligence, are open to supporters or adherents beyond original initiating countries. Such openness might not, however, deliver representation and legitimacy at the speed and breadth required to keep pace with accelerating AI proliferation globally. Meanwhile, representation gaps in international AI governance processes persist, with decision-making concentrated in the hands of a few countries and companies.

# B. Coordination gaps

**72** The ongoing emergence and evolution of AI governance initiatives are not guaranteed to work together effectively for humanity. Instead, coordination gaps have appeared. Effective handshaking between the selective plurilateral initiatives (see fig. 8) and other regional initiatives is not assured, risking incompatibility between regions.

**73** Nor are there global mechanisms for all international standards development organizations (see fig. 7), international scientific research initiatives or AI capacity-building initiatives to coordinate with each other, undermining interoperability of approaches and resulting in fragmentation. The resulting coordination gaps between various sub-global initiatives are in some cases best addressed at the global level.

**74** A separate set of coordination gaps arise within the United Nations system, reflected in the array of diverse United Nations documents and initiatives in relation to AI. Figure 9 shows 27 United Nations-related instruments in specific domains that may apply to AI – 23 of them are binding and will require interpretation as they pertain to AI. A further 29 domain-level documents from the United Nations and related organizations focus specifically on AI, none of which are binding.[17] In some cases, these can address AI risks and harness AI benefits in specific domains.

**75** The level of activity shows the importance of AI to United Nations programmes. As AI expands to affect ever-wider aspects of society, there will be growing calls for diverse parts of the United Nations system to act, including through binding norms. It also shows the ad hoc nature of the responses, which have largely developed organically in specific domains and without an overarching strategy. The resulting coordination gaps invite overlaps and hinder interoperability and impact.

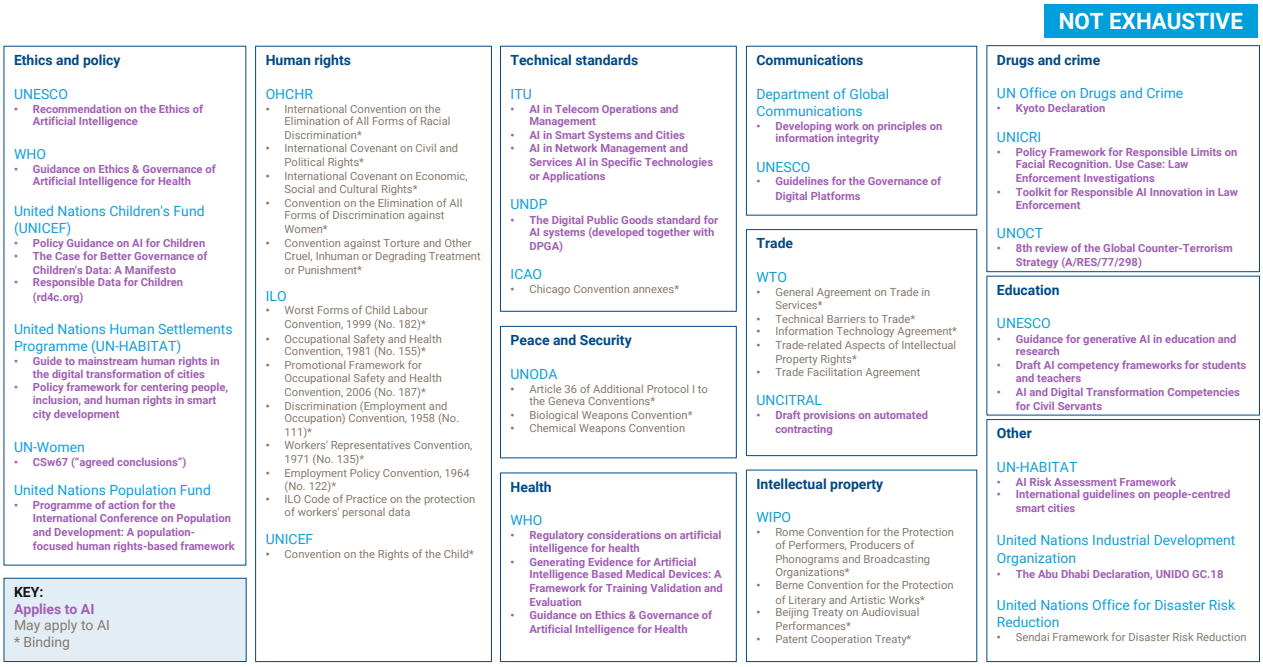
**76** The number and diversity of approaches are a sign that the United Nations system is responding to an emerging issue. With proper orchestration, and in combination with processes taking a holistic approach, these efforts can offer an efficient and sustainable pathway to inclusive international AI governance in specific domains. This could enable meaningful, harmonized and coordinated impacts on areas such as health, education, technical standards and ethics, instead of merely contributing to the proliferation of initiatives and institutions in this growing field. International law, including international human rights law, provides a shared normative foundation for all AI-related efforts, thereby facilitating coordination and coherence.

**77** Although the work of many United Nations entities touches on AI governance, their specific mandates mean that none does so in a comprehensive manner; and their designated governmental focal points are similarly specialized.[18] This limits the ability of existing United Nations entities to address

17 A survey conducted by the United Nations Chief Executives Board in February 2024 of 57 United Nations entities reported 50 documents concerning AI governance; 44 of the 57 entities responded, including the Economic Commission for Latin America and the Caribbean; the Economic and Social Commission for Asia and the Pacific; the Economic and Social Commission for Western Asia; the Food and Agriculture Organization of the United Nations (FAO); the International Atomic Energy Agency (IAEA); the International Civil Aviation Organization (ICAO); the International Fund for Agricultural Development; ILO; the International Monetary Fund; the International Organization for Migration; International Trade Centre; the International Telecommunication Union (ITU); the United Nations Entity for Gender Equality and the Empowerment of Women (UN-WOMEN); the Joint United Nations Programme on HIV/AIDS (UNAIDS); the United Nations Conference on Trade and Development (UNCTAD); the Department of Economic and Social Affairs; the Department of Global Communications; the Executive Office of the Secretary-General; the Office for the Coordination of Humanitarian Affairs; the Office of the United Nations High Commissioner for Human Rights; the Office of Counter-Terrorism; the Office for Disarmament Affairs; the Office of Information and Communications Technology; OSET; the United Nations Development Programme (UNDP); the United Nations Office for Disaster Risk Reduction; the United Nations Environment Programme; UNESCO; the United Nations Framework Convention on Climate Change; the United Nations Population Fund; the United Nations High Commissioner for Refugees (UNHCR); the United Nations Children's Fund; the United Nations Interregional Crime and Justice Research Institute; the United Nations Industrial Development Organization; the United Nations Office on Drugs and Crime/United Nations Office at Vienna; the United Nations Office for Project Services; the United Nations Relief and Works Agency for Palestine Refugees in the Near East; United Nations University; United Nations Volunteers; the World Trade Organization; the Universal Postal Union; the World Bank Group; the World Food Programme; the World Health Organization (WHO); and the World Intellectual Property Organization (WIPO). See "United Nations system white paper on AI governance: an analysis of the UN system's institutional models, functions, and existing international normative frameworks applicable to AI governance" (available at https://unsceb.org/united-nations-system-white-paper-ai-governance).

18 For example, ministries of education, science and culture (UNESCO); telecommunication or ICT (ITU); industry (United Nations Industrial Development Organization); and labour (ILO).

## Figure 9: Selected documents related to AI governance from the United Nations and related organizations




*Source*: "United Nations system white paper on AI governance: an analysis of the UN system's institutional models, functions, and existing international normative frameworks applicable to AI governance", 28 Feb 2024.

the multifaceted implications of AI globally on their own. At the national and regional levels, such gaps are being addressed by new institutions,[19] such as AI safety institutes or AI offices for an appropriately transversal approach.

# C. Implementation gaps

**78** Representation and coordination are not enough, however. Action and follow-up processes are required to ensure that commitments to good governance translate into tangible outcomes in practice. More is needed to ensure accountability. Peer pressure and peer-to-peer learning are two elements that can spur accountability.

**79** Engaging with the private sector will be equally important for meaningful accountability and remedy for harm. The United Nations has experience of this in the United Nations Guiding Principles on Business and Human Rights. Equally, we would need robust engagement of civil society and scientific experts to keep governments and private companies honest about their commitments and claims.

**80** Missing enablers for harnessing AI's benefits for the public good within and between countries constitute a key implementation gap. Many countries have put in place national strategies to boost AI-related infrastructure and talent, and a few initiatives for international assistance are emerging.[20] However, these are under-networked and under-resourced.

**81** At the global level, connecting national and regional capacity development initiatives, and pooling resources to support those countries left out from such efforts, can help to ensure that no country is left behind in the sharing of opportunities associated with AI. Another key implementation gap is the absence of a dedicated fund for AI capacity-building despite the existence of some funding mechanisms for digital capacity (box 8).

19 Including those set up by Canada, Japan, Singapore, the Republic of Korea, the United Kingdom, the United States and the European Union.
20 National-level efforts could continue to employ diagnosis tools, such as the UNESCO AI Readiness Assessment Methodology to help to identify gaps at the country level, with the international network helping to address them.

## Box 8: Gaps in global financing of AI capacity

The Advisory Body believes that there are no existing global funds for AI capacity-building with the scale and mandate to fund the significant investment required to put a floor under the AI divide.

Indicative estimates place the amount needed in the range of $350 million to $1 billion annually,[a] including in-kind contributions from the private sector, mandated to target AI capacity across all AI enablers, including talent, compute, training data, model development and interdisciplinary collaboration for applications. Examples of existing multilateral funding mechanisms include:

**a) Joint SDG Fund**

This fund is broad and encompasses every SDG, as well as emergency response. It supports country-level initiatives for integrated United Nations policy and strategic financing support to countries to advance the SDGs. The fund helps the United Nations to deliver and catalyse SDG financing and programming. Since 2017, 30 participating United Nations entities have received a total of $223 million. It does not fund national governments, communities or entities directly, and it does not fund cross-border initiatives.

In 2023, the fund had around 16 donors for a total of $57.7 million, and an estimated $58.8 million in 2024. The private sector has contributed $83,155 since 2017, and none in 2023 or 2024 to date.

Most of the fund, 60 per cent, go to actions in five SDGs: Goals 2 (zero hunger), 5 (gender equality), 7 (affordable and clean energy), 9 (industry, innovation and infrastructure) and 17 (partnerships).

The fund's Policy Digital Transformation stream (launched in 2023) has funded one project of $250,000, disbursed equally between the International Telecommunication Union (ITU) and the United Nations Development Programme (UNDP). At the end of financial year 2023, its delivery rate was 2.27 per cent. Digital transformation activities form a small part of the fund's activities, and typically in relation to other SDGs (e.g. connectivity and digital infrastructure to support service delivery, such as in small island developing States).

**b) World Bank, Digital Development Partnership**

This fund supports countries in developing and implementing the digital transformation with a focus on broadband infrastructure, access and use, digital public infrastructure and data production, accessibility and use. By the end of 2022, it had invested $10.7 billion in more than 80 countries.

The partnership includes a cybersecurity associated multi-donor trust fund (Estonia, Germany, Japan and the Kingdom of the Netherlands) to support national cybersecurity capacity development.

a Less than 1 per cent of estimated annual private sector AI investment in 2023.

# 4. Enhancing global cooperation

**82** Having outlined the global governance deficit, we now turn to recommendations to address the priority gaps for the near term.

**83** Our recommendations advance a holistic vision for a globally networked, agile and flexible approach to governing AI for humanity, encompassing common understanding, common ground and common benefits to enhance representation, enable coordination and strengthen implementation (see fig. 10). Only such an inclusive and comprehensive approach to AI governance can address the multifaceted and evolving challenges and opportunities AI presents on a global scale, promoting international stability and equitable development.

**84** Guided by the principles listed in our interim report (see para. 47), our proposals seek to fill gaps and bring coherence to the fast-emerging ecosystem of international AI governance responses and initiatives, helping to avoid fragmentation and missed opportunities. To support these measures efficiently and partner effectively with other institutions, we propose a light, agile structure as an expression of coherent effort: an AI office in the United Nations Secretariat, close to the Secretary-General, working as the "glue" to hold these other pieces together.

**85** The United Nations is far from perfect. Nevertheless, the legitimacy arising from its unique inclusiveness, coupled with its binding normative foundations in international law, including international human rights law, presents hope for governing AI for the benefit and protection of humanity in a manner that is equitable, effective and efficient.[21]

## Figure 10: Overview of recommendations and how they address global AI governance gaps

| Purpose | Enhance representation | Enable coordination | Strengthen implementation |
|---|---|---|---|
| **Common understanding**<br>International scientific panel on AI | ✓ | ✓ | |
| **Common ground**<br>Policy dialogue on AI governance<br>AI standards exchange | ✓ | ✓ | (✓) |
| **Common benefits**<br>Capacity development network<br>Global fund for AI<br>Global AI data framework | ✓ | ✓ | ✓ |
| **Coherent effort**<br>AI office within the Secretariat | Advising the Secretary-General on matters related to AI, working to promote a coherent voice within the United Nations system, engaging States and stakeholders, partnering and interfacing with other processes and institutions, and supporting other proposals as required. | | |

21 It should also be inclusive and cohesive, and enhance global peace and security.

# A. Common understanding

**86** A global approach to governing AI starts with a common understanding of its capabilities, opportunities, risks and uncertainties.

**87** The AI field has been evolving quickly, producing an overwhelming amount of information and making it difficult to decipher hype from reality. This can fuel confusion, forestall common understanding and advantage major AI companies at the expense of policymakers, civil society and the public.

**88** In addition, a dearth of international scientific collaboration and information exchange can breed global misperceptions and undermine international trust.

**89** There is a need for timely, impartial and reliable scientific knowledge and information about AI for Member States to build a shared foundational understanding worldwide, and to balance information asymmetries between companies housing expensive AI labs and the rest of the world, including via information-sharing between AI companies and the broader AI community.

**90** This is most efficient at the global level, enabling joint investment in a global public good and public interest collaboration across otherwise fragmented and duplicative efforts.

## International scientific panel on AI

**Recommendation 1: An international scientific panel on AI**

We recommend the creation of an independent international scientific panel on AI, made up of diverse multidisciplinary experts in the field serving in their personal capacity on a voluntary basis. Supported by the proposed United Nations AI office and other relevant United Nations agencies, partnering with other relevant international organizations, its mandate would include:

a. Issuing an annual report surveying AI-related capabilities, opportunities, risks and uncertainties, identifying areas of scientific consensus on technology trends and areas where additional research is needed;
b. Producing quarterly thematic research digests on areas in which AI could help to achieve the SDGs, focusing on areas of public interest which may be under-served; and
c. Issuing ad hoc reports on emerging issues, in particular the emergence of new risks or significant gaps in the governance landscape.

**91** There is precedent for such an institution. Some examples include the United Nations Scientific Committee on the Effects of Atomic Radiation, the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), the Scientific Committee on Antarctic Research, and the Intergovernmental Panel on Climate Change (IPCC).

**92** These models are known for their systematic approaches to complex, pervasive issues affecting various sectors and global populations. However, while they can provide inspiration, none is perfectly suited to assessing AI technology and should not be replicated directly. Instead, a tailored approach is required.

**93** Learning from such precedents, an independent, international and multidisciplinary scientific panel on AI could collate and catalyse leading-edge research to inform those seeking scientific perspectives on AI technology or its applications from an impartial, credible source. An example of one kind of issue to which the panel could contribute is the ongoing debate over open versus closed AI systems, discussed in box 9.

**94** A scientific panel under the auspices of the United Nations would have a broad focus to cover an inclusive range of priorities holistically. This could include sourcing expertise on AI-related opportunities, and facilitating “deep dives” into applied domains of the SDGs, such as health care, energy, education, finance, agriculture, climate, trade and employment.

**95** Risk assessments could also draw on the work of other AI research initiatives, with the United Nations offering a uniquely trusted "safe harbour" for researchers to exchange ideas on the "state of the art". International law, including human rights law, would provide a compass for defining pertinent risks. By pooling knowledge across silos in countries or companies that may not otherwise engage or be included, a United Nations-hosted panel can help to rectify misperceptions and bolster trust globally.

**96** Such a scientific panel would not necessarily conduct its own research but be a catalyst for networked action.[22] It could aggregate, distil and translate developments in AI for its audiences, highlighting potential use cases. It would reduce information asymmetry, help to avoid misdirected investments and keep information flowing across a global network of experts.

**97** The panel would have three key audiences:

a. The first is the global scientific community.[23] The shift of fundamental research on AI to private corporations, driven in part by the cost of computational power, has led to concerns that such research may be unduly driven by financial interests. A scientific panel could encourage greater research in public institutions worldwide focused on the public good.

b. Secondly, regular independent assessments would inform Member States, policymakers and other processes recommended in this report. An annual risk survey from the world's experts would help to shape the agenda of the AI governance dialogues proposed in recommendation 2. The state-of-the-art report would inform the development of standards proposed in recommendation 3, as well as the capacity development network proposed in recommendation 4.

c. Thirdly, through its public reports, it could serve as an impartial source of high-quality information for the public.

**98** The global reach of networks uniquely accessible via the United Nations would enable common understanding across the widest basis, making available findings in ways relevant to various socioeconomic and geographical contexts. The panel can thereby activate the United Nations as a reliable platform for inclusively networked, multidisciplinary stakeholder understanding.

**99** The panel could be established for an initial period of 3–5 years (with extension subject to review by the Secretary-General), and could function according to the following basis:

a. The panel could start with 30–50 members appointed through a mix of Member State- and self-nomination, comparable to how the Advisory Body was established. It should focus on scientific expertise across disciplines, and would need to ensure diverse representation by region and gender, as well as reflecting the interdisciplinary nature of AI. Membership could be rotated periodically within the overall mandate of 3–5 years.

b. The panel would meet virtually (and in-person as a plenary, perhaps twice a year). Meetings could rotate between cities hosting relevant United Nations entities, including in global South locations. It should be encouraged to form thematic working groups, adding additional members as needed and engaging networks of academic partners. It could explore inviting participation in these working groups from relevant United Nations entities.[24]

c. The panel would operate independently, particularly in relation to its findings and conclusions, with support from a United Nations-system team drawn from the proposed AI office and relevant United Nations agencies, such as ITU and the United Nations Educational, Scientific and Cultural Organization (UNESCO).

d. It should partner with and build on research efforts led by other international institutions such as OECD and the Global Partnership on Artificial Intelligence, and other relevant

22 It could build, in particular, upon existing sectoral or regional panels already operating.
23 It could also conduct outreach to broader audiences, including civil society and the general public.
24 For a list of United Nations entities active in this area, see figure 9.

processes such as the recent scientific report on the risks of advanced AI commissioned by the United Kingdom,[25] and relevant regional organizations.

e. A steering committee would develop a research agenda ensuring the inclusivity of views and incorporation of ethical considerations, oversee the allocation of resources, foster collaboration with a network of academic institutions and other stakeholders, and review the panel's activities and deliverables.

**100** By drawing on the unique convening power of the United Nations and inclusive global reach across stakeholder groups, an international scientific panel can deliver trusted scientific collaboration processes and outputs and correct information asymmetries in ways that address the representation and coordination gaps identified in paragraphs 66 and 73, thereby promoting equitable and effective international AI governance.

## Box 9: Open versus closed AI systems

Among the topics discussed in our consultations was the ongoing debate over open versus closed AI systems. AI systems that are open in varying degrees are often referred to as "open-source AI", but this is somewhat of a misnomer when compared with open-source software (code). It is important to recognize that openness in AI systems is more of a spectrum than a single attribute.

One article explained that a "fully closed AI system is only accessible to a particular group. It could be an AI developer company or a specific group within it, mainly for internal research and development purposes. On the other hand, more open systems may allow public access or make available certain parts, such as data, code, or model characteristics, to facilitate external AI development."[a]

Open-source AI systems in the generative AI field present both risks and opportunities. Companies often cite "AI safety" as a reason for not disclosing system specifications, reflecting the ongoing tension between open and closed approaches in the industry. Debates typically revolve around two extremes: full openness, which entails sharing all model components and data sets; and partial openness, which involves disclosing only model weights.

Open-source AI systems encourage innovation and are often a requirement for public funding. On the open extreme of the spectrum, when the underlying code is made freely available, developers around the world can experiment, improve and create new applications. This fosters a collaborative environment where ideas and expertise are readily shared. Some industry leaders argue that this openness is vital to innovation and economic growth.

However, in most cases, open-source AI models are available as application programming interfaces. In this case, the original code is not shared, the original weights are never changed and model updates become new models.

Additionally, open-source models tend to be smaller and more transparent. This transparency can build trust, allow for ethical considerations to be proactively addressed, and support validation and replication because users can examine the inner workings of the AI system, understand its decision-making process and identify potential biases.

a Angela Luna, "The open or closed AI dilemma", 2 May 2024. Available at https://bipartisanpolicy.org/blog/the-open-or-closed-ai-dilemma.

25 International Scientific Report on the Safety of Advanced AI: Interim Report. Available at https://gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai.

## Box 9: Open versus closed AI systems (continued)

Closed AI systems offer greater control to their developers. Additionally, closed-source systems can be more streamlined and efficient, as the codebase is not constantly evolving through public contributions. Many companies regard full openness as impractical and promote partial openness as the only feasible option. However, this viewpoint overlooks the potential for a balanced approach that can achieve "meaningful openness".[b]

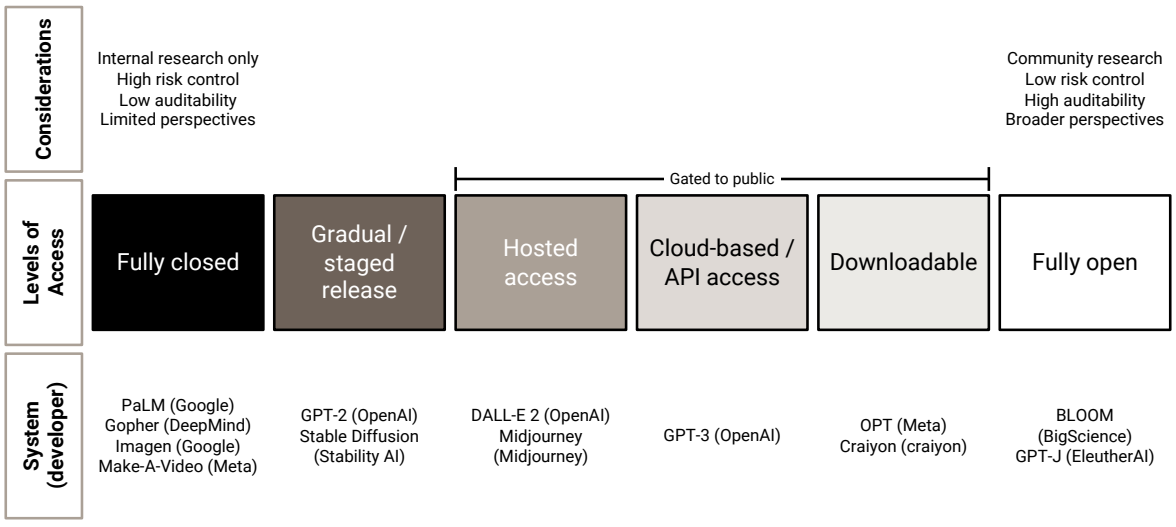
Meaningful openness exists between the two extremes of the spectrum and can be tailored to different use cases. This balanced method fosters safe, innovative and inclusive AI development by enabling public scrutiny and independent auditing of disclosed training and fine-tuning data. Openness, being more than merely sharing model weights, can propel innovation and inclusion, helping applications in research and education.

The definition of "open-source AI" is evolving,[c] and is often influenced by corporate interests as illustrated in figure 11. To address this, we recommend initiating a process, coordinated by the above-proposed international scientific panel, to develop a well-rounded and gradient approach to openness. This would enable meaningful, evidence-based approaches to openness, helping users and policymakers to make informed choices about AI models and architectures.

Data disclosure – even if limited to key elements – is essential for understanding model performance, ensuring reproducibility and assessing legal risks. Clarification around gradations of openness can help to counter corporate "open-washing" and foster a transparent tech ecosystem.

It is also important that, as the technology matures, we consider the governance regimes for the application of both open and closed AI systems. We need to develop responsible AI guidelines, binding norms and measurable standards for developers and designers of products and services that incorporate AI technologies, as well as for their users and all actors involved throughout their life cycle.

## Figure 11: Corporate interests and openness




*Source:* Irene Solaiman, "The gradient of generative AI release: methods and considerations", *Proceedings of the 2023 Association for Computing Machinery (ACM) Conference on Fairness, Accountability, and Transparency* (June 2023), pp. 111–122.

b Inspired by Andreas Liesenfeld and Mark Dingemanse, "Rethinking open source generative AI: open-washing and the EU AI Act", *The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)* (June 2024).

c The Open Source AI Definition – draft v. 0.0.3. Available at https://opensource.org/deepdive/drafts/the-open-source-ai-definition-draft-v-0-0-3.

# B. Common ground

**101** Alongside a common understanding of AI, common ground is needed to establish governance approaches that are interoperable across jurisdictions and grounded in international norms, such as the Universal Declaration of Human Rights (see principle 5 above).

**102** This is required at the global level not only for equitable representation, but also for averting regulatory "races to the bottom" while reducing regulatory friction across borders, maximizing technical and ontological interoperability, and detecting and responding to incidents emanating from decisions along AI's life cycle which span multiple jurisdictions.

## Policy dialogue on AI governance

**Recommendation 2: Policy dialogue on AI governance**

We recommend the launch of a twice-yearly intergovernmental and multi-stakeholder policy dialogue on AI governance on the margins of existing meetings at the United Nations. Its purpose would be to:

a. Share best practices on AI governance that foster development while furthering respect, protection and fulfilment of all human rights, including pursuing opportunities as well as managing risks;
b. Promote common understandings on the implementation of AI governance measures by private and public sector developers and users to enhance international interoperability of AI governance;
c. Share voluntarily significant AI incidents that stretched or exceeded the capacity of State agencies to respond; and
d. Discuss reports of the international scientific panel on AI, as appropriate.

**103** International governance of AI is currently a fragmented patchwork at best. There are 118 countries that are not parties to any of the seven recent prominent non-United Nations AI governance initiatives with intergovernmental tracks[26] (see fig. 8). Representation gaps occur even among the top 60 AI capacity countries, highlighting the selectiveness of international AI governance today (see fig. 12).

**104** An inclusive policy forum is needed so that all Member States, drawing on the expertise of stakeholders, can share best practices that foster development while furthering respect, protection and fulfilment of all human rights, promote interoperable governance approaches and monitor for common risks that warrant further policy interventions.

**105** This does not mean global governance of all aspects of AI (which is impossible and undesirable, given States' diverging interests and priorities). Yet, exchanging views on AI developments and policy responses can set the framework for international cooperation.

**106** The United Nations is uniquely placed to facilitate such dialogues inclusively in ways that help Member States to work together effectively. The United Nations system's existing and emerging suite of norms can offer strong normative foundations for concerted action, grounded in the Charter of the United Nations, human rights and other international law, including environmental law and international humanitarian law, as well as the SDGs and other international commitments.[27]

26 These initiatives are not always directly comparable. Some reflect the work of existing international or regional organizations, while others are based on ad hoc invitations from like-minded countries.

27 See, for example, the Charter of the United Nations (preamble, purposes and principles, and Articles 13, 55, 58 and 59). See also core international instruments on human rights (Universal Declaration of Human Rights; International Covenant on Civil and Political Rights; International Covenant on Economic, Social and Cultural Rights; International Convention on the Elimination of All Forms of Racial Discrimination; Convention on the Rights of the Child; Convention on the Elimination of All Forms of Discrimination against Women; Convention against Torture; Convention on the Rights of Persons with Disabilities; Convention on the Rights of Migrants; International Convention for the Protection of All Persons from Enforced Disappearance); instruments on international human rights law (Geneva Conventions; Convention on Certain Conventional Weapons; Genocide Convention; Hague Convention); instruments on related principles such as distinction, proportionality and precaution and the 11 principles on Lethal Autonomous Weapons Systems adopted within the Convention on Certain Conventional Weapons); disarmament and arms control instruments in terms of prohibitions on weapons of mass destruction (Treaty on the Non-Proliferation of Nuclear Weapons; Chemical Weapons Convention; Biological Weapons Convention); environmental law instruments (United Nations Framework Convention on Climate Change; Convention on the Prohibition of Military or Any Other Hostile Use of Environmental Modification Techniques); the Paris Agreement and related principles such as precautionary principle, integration principle and public participation; and non-binding commitments on the 2030 Agenda for Sustainable Development, gender and ethics, such as the UNESCO Recommendation on the Ethics of Artificial Intelligence.

# Figure 12: Top 60 AI countries (2023 Tortoise Index) in the sample of major plurilateral AI governance initiatives with intergovernmental tracks

party to

| Country / party* | Tortoise Global AI Index Rank (2023) | Number of initiatives party to | Chronological order → OECD AI Principles (2019) | G20 AI Principles (2019) | CoE drafters (2022) | GPAI Ministerial Declaration (2022) | Bletchley Declaration (2023) | G7 Ministerial Statement on Hiroshima AI Process (2023) | Seoul Ministerial Statement (2024) |
|---|---|---|---|---|---|---|---|---|---|
| United States of America | 1 | 7 | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| China | 2 | 2 | | ■ | | | ■ | | |
| Singapore | 3 | 4 | ■ | | | ■ | ■ | | ■ |
| United Kingdom of Great Britain and Northern Ireland | 4 | 7 | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| Canada | 5 | 7 | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| Republic of Korea | 6 | 5 | ■ | ■ | | ■ | ■ | | ■ |
| Israel | 7 | 5 | ■ | | ■ | ■ | ■ | | ■ |
| Germany | 8 | 7 | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| Switzerland | 9 | 4 | ■ | | ■ | | ■ | | ■ |
| Finland | 10 | 2 | ■ | | ■ | | | | |
| Netherlands | 11 | 5 | ■ | | ■ | ■ | ■ | | ■ |
| Japan | 12 | 7 | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| France | 13 | 7 | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| India | 14 | 4 | | ■ | | ■ | ■ | | ■ |
| Australia | 15 | 6 | ■ | ■ | ■ | ■ | ■ | | ■ |
| Denmark | 16 | 3 | ■ | | ■ | ■ | | | |
| Sweden | 17 | 3 | ■ | | ■ | ■ | | | |
| Luxembourg | 18 | 2 | ■ | | ■ | | | | |
| Ireland | 19 | 4 | ■ | | ■ | ■ | ■ | | |
| Austria | 20 | 2 | ■ | | ■ | | | | |
| Spain | 21 | 5 | ■ | | ■ | ■ | ■ | | ■ |
| Belgium | 22 | 3 | ■ | | ■ | ■ | | | |
| Italy | 23 | 7 | ■ | ■ | ■ | ■ | ■ | ■ | ■ |
| Norway | 24 | 2 | ■ | | ■ | | | | |
| Estonia | 25 | 2 | ■ | | ■ | | | | |
| United Arab Emirates | 27 | 2 | | | | | ■ | | ■ |
| Portugal | 28 | 2 | ■ | | ■ | | | | |
| Russian Federation | 29 | 1 | | ■ | | | | | |
| Saudi Arabia | 30 | 3 | | ■ | | | ■ | | ■ |
| Malta | 31 | 2 | ■ | | ■ | | | | |
| Brazil | 33 | 4 | ■ | ■ | | ■ | ■ | | |
| New Zealand | 34 | 3 | ■ | | | ■ | | | ■ |
| Slovenia | 35 | 3 | ■ | | ■ | ■ | | | |
| Hungary | 36 | 2 | ■ | | ■ | | | | |
| Türkiye | 37 | 6 | ■ | ■ | ■ | ■ | ■ | | ■ |
| Iceland | 38 | 2 | ■ | | ■ | | | | |
| Chile | 39 | 3 | ■ | | | | ■ | | ■ |
| Qatar | 40 | 0 | | | | | | | |
| Lithuania | 41 | 2 | ■ | | ■ | | | | |
| Malaysia | 42 | 0 | | | | | | | |
| Greece | 43 | 2 | ■ | | ■ | | | | |
| Indonesia | 44 | 3 | | ■ | | | ■ | | ■ |
| Viet Nam | 45 | 0 | | | | | | | |
| Colombia | 46 | 1 | ■ | | | | | | |
| Argentina | 47 | 4 | ■ | ■ | ■ | ■ | | | |
| Slovakia | 48 | 2 | ■ | | ■ | | | | |
| Mexico | 49 | 5 | ■ | ■ | ■ | ■ | | | ■ |
| Egypt | 50 | 1 | ■ | | | | | | |
| Uruguay | 51 | 1 | | | ■ | | | | |
| Armenia | 52 | 1 | | | ■ | | | | |
| South Africa | 53 | 1 | | ■ | | | | | |
| Tunisia | 54 | 0 | | | | | | | |
| Morocco | 55 | 0 | | | | | | | |
| Bahrain | 56 | 0 | | | | | | | |
| Pakistan | 57 | 0 | | | | | | | |
| Sri Lanka | 58 | 0 | | | | | | | |
| Nigeria | 59 | 2 | | | | | ■ | | ■ |
| Kenya | 60 | 2 | | | | | ■ | | ■ |
| European Union | n/a | 5 | ■ | ■ | ■ | ■ | ■ | | ■ |
| **Total (including those not shown)** | | | **47** | **20** | **58** | **29** | **29** | **7** | **28** |

*Including jurisdictions such as the Holy See and the European Union.

Sources:
- OECD, Recommendation of the Council on Artificial Intelligence (adopted 21 May 2019), available at https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.
- G20, AI Principles (June 2019), available at https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/documents/en/annex_08.pdf.
- GPAI, 2022 ministerial declaration (22 November 2022), available at https://one.oecd.org/document/GPAI/C(2022)7/FINAL/en/pdf.
- Bletchley Declaration (1 Nov 2023), available at https://gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023.
- G7, Hiroshima AI Process G7 Digital & Tech Ministers' Statement (1 Dec 2023), available at https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document02_en.pdf.
- Council of Europe, Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (adopted 17 May 2024), available at https://coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence.
- Seoul Ministerial Statement for advancing AI safety, innovation and inclusivity, AI Seoul Summit (22 May 2024).
- Tortoise Media, Global AI Index (2023), available at https://tortoisemedia.com/intelligence/global-ai/#rankings.

**107** Combined with expertise from the international scientific panel and capacity development (see recommendations 1, 4 and 5), inclusive dialogue at the United Nations can help States and companies to update their regulatory approaches and methodologies to keep pace with accelerating AI in an interoperable way that promotes common ground. Some of the distinctive features of the United Nations can be helpful in this regard:

a. Anchoring inclusive dialogue in the United Nations suite of norms, including the Charter of the United Nations and human rights and international law, can promote a "race to the top" in governance approaches. Conversely, without the universal global membership of the United Nations, international collective action faces greater pressure to succumb to regulatory "races to the bottom" between jurisdictions on AI safety and scope of use.

b. The global membership of the United Nations can also enable coordination between existing sub-global initiatives for greater compatibility between them. Many in our consultations called for the United Nations to be a key space for enabling soft coordination across existing regional and plurilateral initiatives, taking into account diverse values across different cultures, languages and regions.

c. The Organization's predictable, transparent, rule-based and justifiable procedures can enable continuous political engagement to bridge non-likeminded countries, and moderate dangerous contestation. In addition to building confidence, relationships and communication lines for times of crisis, reliably inclusive dialogues can foster new norms, customary law and agreements that enhance cooperation among States.

**108** Operationally:

a. A policy dialogue could begin on the margins of existing meetings in New York, such as the General Assembly,[28] Geneva and locations in the global South.

b. One portion of each dialogue session might focus on national approaches led by Member States, with a second portion sourcing expertise and inputs from key stakeholders – in particular, technology companies and civil society representatives.

c. Governmental participation could be open to all Member States, or a regionally balanced grouping (for more focused discussion among a rotating, representative interested subset), or a combination of both, calibrated as appropriate to different agenda items or segments over time, as the technology evolves and global concerns emerge or gain salience. A fixed geometry might not be helpful, given the dynamic nature of the technology and the policy context.

d. In addition to the formal dialogue sessions, multi-stakeholder engagement on AI policy could also leverage other existing mechanisms such as the ITU AI for Good meeting, the annual Internet Governance Forum meeting, the UNESCO AI ethics forum and the United Nations Conference on Trade and Development (UNCTAD) eWeek, open for participation to representatives of all Member States on a voluntary basis.

e. In line with the inclusive nature of the dialogue, discussion agendas could be broad to encompass diverse perspectives and concerns. For instance, twice-yearly meetings could focus more on opportunities across diverse sectors in one meeting, and more on risk trends in the other.[29] This could include uses of AI to achieve the SDGs, how to protect children, minimize climate impact, as well as an exchange on approaches to manage risks. Meetings could also include a discussion of definitions of terms used in AI governance and AI technical standards, as well as reports of the international scientific panel, as appropriate.

---

28 Analogous to the high-level political forum in the context of the SDGs that takes place under the auspices of the Economic and Social Council.

29 Relevant parts of the United Nations system could be engaged to highlight opportunities and risks, including ITU on AI standards; ITU, UNCTAD, UNDP and the Development Coordination Office on AI applications for the SDGs; UNESCO on ethics and governance capacity; the Office of the United Nations High Commissioner for Human Rights (OHCHR) on human rights accountability based on existing norms and mechanisms; the Office for Disarmament Affairs on regulating AI in military systems; UNDP on support to national capacity for development; the Internet Governance Forum for multi-stakeholder engagement and dialogue; WIPO, ILO, WHO, FAO, the World Food Programme, UNHCR, UNESCO, the United Nations Children's Fund, the World Meteorological Organization and others on sectoral applications and governance.

f. In addition, diverse stakeholders – in particular technology companies and civil society representatives – could be invited to engage through existing institutions detailed below, as well as policy workshops on particular aspects of AI governance such as limits (if any) of open-source approaches to the most advanced forms of AI, thresholds for tracking and reporting of AI incidents, application of human rights law to novel use cases, or the use of competition law/antitrust to address concentrations of power among technology companies.[30]

g. The proposed AI office could also curate a repository of AI governance examples, including legislation, policies and institutions from around the world for consideration of the policy dialogue, working with existing efforts, such as OECD.

**109** Notwithstanding the two General Assembly resolutions on AI in 2024, there is currently no mandated institutionalized dialogue on AI governance at the United Nations that corresponds to the reliably inclusive vision of this recommendation. Similar processes do exist at the international level, but primarily in regional or plurilateral constellations (para. 57), which are not reliably inclusive and global.

**110** Complementing a fluid process of plurilateral and regional AI summits,[31] the United Nations can offer a stable home for dialogue on AI governance. Inclusion by design – a crucial requirement for playing a stabilizing role in geopolitically delicate times – can also address representation and coordination gaps identified in paragraphs 64 and 72, promoting more effective collective action on AI governance in the common interest of all countries.

## AI standards exchange

**Recommendation 3: AI standards exchange**

We recommend the creation of an AI standards exchange, bringing together representatives from national and international standard-development organizations, technology companies, civil society and representatives from the international scientific panel. It would be tasked with:

a. Developing and maintaining a register of definitions and applicable standards for measuring and evaluating AI systems;

b. Debating and evaluating the standards and the processes for creating them; and

c. Identifying gaps where new standards are needed.

**111** When AI systems were first explored, few standards existed to help to navigate or measure this new frontier. The Turing Test – of whether a machine can exhibit behaviour equivalent to (or indistinguishable from) a human being – captured the popular imagination, but is of more cultural than scientific significance. Indeed, it is telling that some of the greatest computational advances have been measured by their success in games, such as when a computer could beat humans at chess, Go, poker or Jeopardy. Such measures were easily understood by non-specialists, but were neither rigorous nor particularly scientific.

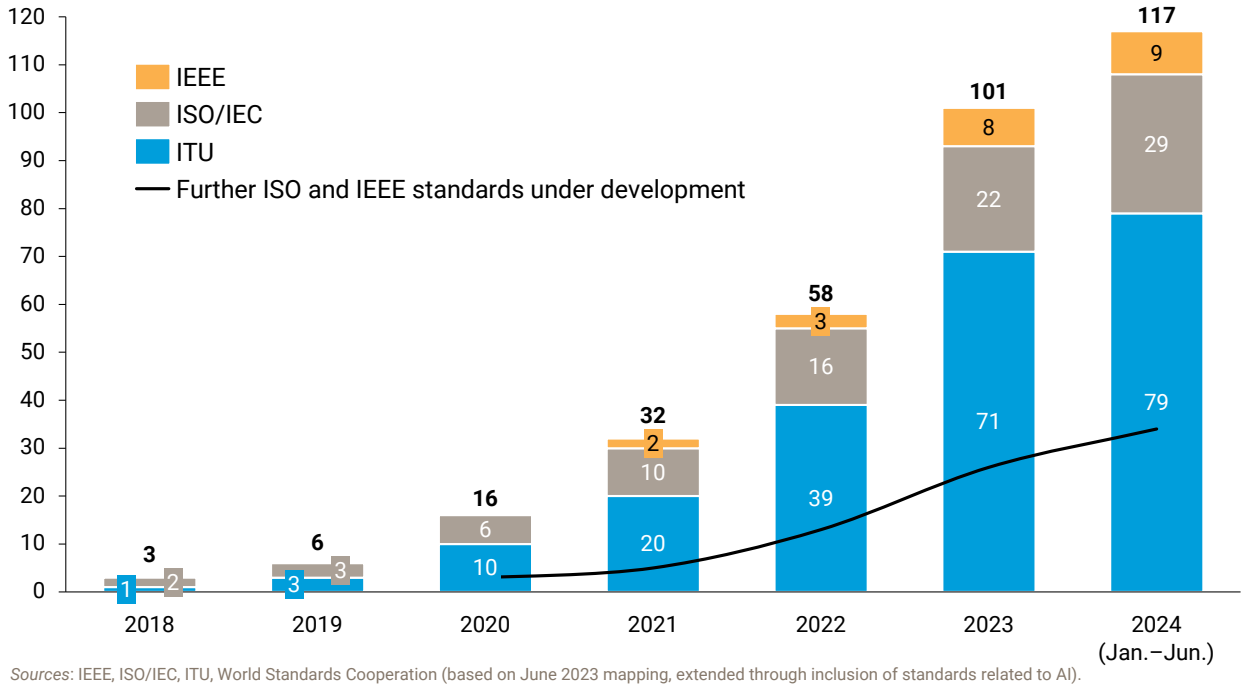
**112** More recently, there has been a proliferation of standards. Figure 13 illustrates the increasing number of relevant standards adopted by ITU, the International Organization for Standardization (ISO), the International Electrotechnical Commission (IEC) and the Institute of Electrical and Electronics Engineers (IEEE).[32]

30 Such a gathering could also provide an opportunity for multi-stakeholder debate of any hardening of the global governance of AI. These might include, for example, prohibitions on the development of uncontainable or uncontrollable AI systems, or requirements that all AI systems be sufficiently transparent so that their consequences can be traced back to a legal actor that can assume responsibility for them.

31 Although multiple AI summits have helped a subset of 20–30 countries to align on AI safety issues, participation has been inconsistent: Brazil, China and Ireland endorsed the Bletchley Declaration in November 2023, but not the Seoul Ministerial Statement six months later (see fig. 12). Conversely, Mexico and New Zealand endorsed the Seoul Ministerial Statement, but did not endorse the Bletchley Declaration.

32 Many new standards are also emerging at the national and multinational levels, such as the United States White House Voluntary AI Commitments and the European Union Codes of Practice for the AI Act.

## Figure 13: Number of standards related to AI




*Sources*: IEEE, ISO/IEC, ITU, World Standards Cooperation (based on June 2023 mapping, extended through inclusion of standards related to AI).

**113** Two trends stand out. First, these standards were largely developed to address specific questions. There is no common language and many terms that are routinely used with respect to AI – fairness, safety, transparency – do not have agreed definitions or measurability (despite recent work by OECD and the National Institute of Standards and Technology adopting a new approach for dynamic systems, such as AI).

**114** Secondly, there is a disjunction between those standards that were adopted for narrow technical or internal validation purposes, and those that are intended to incorporate broader ethical principles. Computer scientists and social scientists often advance different interpretations of the same concept, and a joined-up paradigm of socio-technical standards is promising but remains aspirational (see box 10).

**115** The result is that we have an emerging set of standards that are not grounded in a common understanding of meaning or are divorced from the values they were intended to uphold. Crucially, there are few agreed standards concerning energy consumption and AI. A lack of integration of human rights considerations into standard-setting processes is another gap to be bridged.[33]

**116** This has real costs. In addition to the concerns of Member States and diverse individuals, many of our consultations revealed the concern of businesses (including small and medium-sized enterprises in the developing world) that fragmented governance and inconsistent standards raise the costs of doing business in an increasingly globalized world.

**117** This report is not proposing that the United Nations adds to this proliferation of standards. Instead, drawing on the expertise of the international scientific panel (proposed in recommendation 1), and incorporating members from the various entities that have contributed to standard-setting, as well as representatives from technology companies and civil society, the United Nations system could serve as a clearing house for AI standards that would apply globally.[34]

33 See A/HRC/53/42 (Human rights and technical standard-setting processes for new and emerging digital technologies: Report of the Office of the United Nations High Commissioner for Human Rights) and Human Rights Council resolution 53/29 (New and emerging digital technologies and human rights).

34 Even this may seem a challenging task, but progress towards a global minimum tax deal shows the possibility of collective action even in economically and politically complex areas.

## Box 10: Standards applicable to AI safety

A comprehensive approach to AI safety involves understanding the capabilities of advanced AI models, adopting standards for safe design and deployment, and evaluating both the systems and their broader impacts.

In the past, AI standards focused mainly on technical specifications, detailing how systems should be built and operated. However, as AI technologies increasingly impact society, there is a need to shift to a socio-technical paradigm. This shift acknowledges that AI systems do not exist in a vacuum; they interact with human users and affect societal structures. Modern AI standards can integrate ethical, cultural and societal considerations alongside technical requirements. In the context of safety, this includes ensuring reliability and interpretability, as well as assessing and mitigating risks to individual and collective rights,[a] national and international security, and public safety in different contexts.

A primary objective of the recently established AI safety national institutes is to ensure consistent and effective approaches to AI safety. Harmonizing such approaches would allow AI systems to meet high safety benchmarks internationally, enabling cross-border innovation and trade while maintaining rigorous safety protocols.

As far as "safety" is contextual, involving various stakeholders and cultures in creating such standards enhances their relevance and effectiveness and helps with shared understanding of definitions and concepts. By incorporating diverse perspectives, protocols can more thoroughly address the wide range of potential risks and benefits associated with AI technologies.

a See A/HRC/53/42 (Human rights and technical standard-setting processes for new and emerging digital technologies: Report of the Office of the United Nations High Commissioner for Human Rights) and Human Rights Council resolution 53/29 (New and emerging digital technologies and human rights).

**118** The Organization's added-value would be to foster exchange among the broadest set of standards development organizations to maximize global interoperability across technical standards, while infusing emerging knowledge on socio-technical standards development into AI standards discussions.

**119** Collecting and distributing information on AI standards, drawing on and working with existing efforts such as the AI Standards Hub,[35] would enable participants from across standards development organizations to converge on common language in key areas.

**120** Supported by the proposed AI office, the standards exchange would also benefit from strong ties to the international scientific panel on technical questions and the policy dialogue on moral, ethical, regulatory, legal and political questions.

**121** If appropriately agreed, ITU, ISO/IEC and IEEE could jointly lead on an initial AI standards summit, with annual follow-up to maintain salience and momentum. To build foundations for a socio-technical approach incorporating economic, ethical and human rights considerations, OECD, the World Intellectual Property Organization (WIPO), the World Trade Organization, the Office of the United Nations High Commissioner for Human Rights (OHCHR), ILO, UNESCO and other relevant United Nations entities should also be involved.[36]

35 See https://aistandardshub.org.
36 This could include relevant sectoral, national and regional standards organizations.

**122** The standards exchange should also inform the capacity-building work in recommendation 4, ensuring that the standards support practice on the ground. It could share information about tools developed nationally or regionally that enable self-assessment of compliance with standards.

**123** The report does not presently propose that the United Nations should do more than serve as a forum for discussing and agreeing on standards. To the extent that safety standards are formalized over time, these could serve as the basis for monitoring and verification by an eventual agency.

# C. Common benefits

**124** The 2030 Agenda with its 17 SDGs can lend a unique purpose to AI, bending the arc of investments away from wasteful and harmful use and towards global development challenges. Otherwise, investments will chase profits even at the cost of imposing negative externalities on others. Another signal contribution that the United Nations can make is linking the positive application of AI to an assurance of the equitable distribution of its opportunities (box 11).

## Box 11: AI and the SDGs

AI's potential in advancing science (box 1) and creating economic opportunities (box 2) underlie hope that AI can accelerate progress in achieving the SDGs. A 2023 review of relevant evidence argued that AI may act as an enabler on 134 targets (79 per cent) across all SDGs, generally through technological improvement that may enable certain prevailing limitations to be overcome.[a]

An overview of current expert perceptions is illustrated by the results of an opportunity scan exercise commissioned for our work, which surveyed over 120 experts from 38 countries about their expectations for AI's positive impact in terms of scientific breakthroughs, economic activities and the SDGs. The survey asked only about possible positive implications of AI.

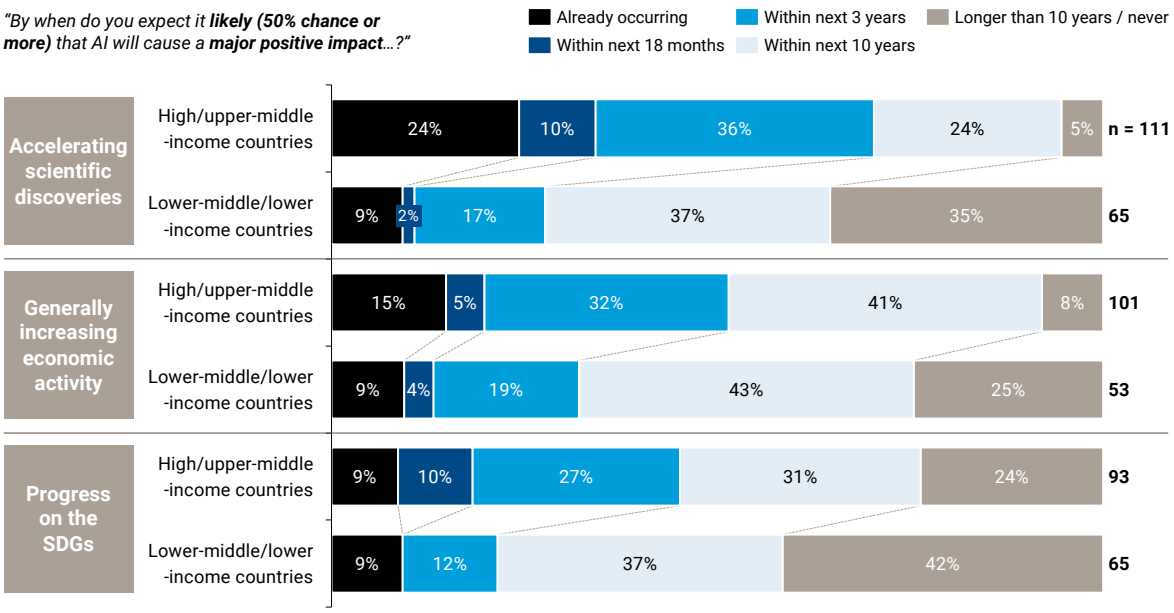
Overall, experts had mixed expectations on how soon AI could have a major positive impact (see also fig. 14):

- They were most optimistic about **accelerating scientific discoveries**, with 7 in 10 saying that it is likely that AI would cause a major positive impact in the next three years or sooner in high/upper-middle-income countries, and 28 per cent predicting the same for lower-middle/lower-income countries.
- Around 5 in 10 expected major positive impact on **increasing economic activity** as likely in the next three years or sooner in high/upper-middle-income countries, and 32 per cent expected the same in lower-middle/lower-income countries.
- A total of 46 per cent expected major positive impact on **progress on the SDGs** as likely in the next three years or sooner in high/upper-middle-income countries. However, only 21 per cent expected this in lower-middle/lower-income countries, with 4 in 10 experts gauging such major positive impact on the SDGs as likely to be at least 10 years away in such places.

a See Ricardo Vinuesa and others, "The role of artificial intelligence in achieving the Sustainable Development Goals". *Nature Communication*, vol. 11, No. 233 (January 2020). This study also argued that 59 targets (35%, also across all SDGs) may experience a negative impact from the development of AI.

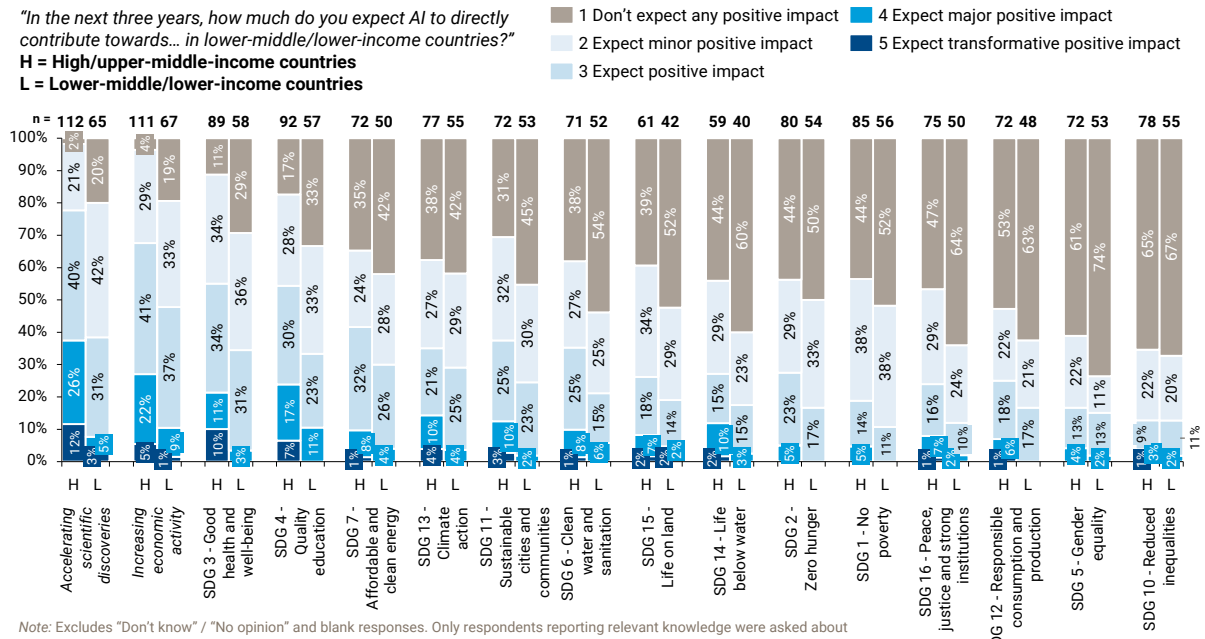
## Box 11: AI and the SDGs (continued)

### Figure 14: Experts' expectations regarding timing of major positive impact of AI by area




*Note:* Excludes "Don't know" / "No opinion" and blank responses. Only respondents reporting relevant knowledge were asked about lower-middle/lower-income countries.
*Source:* OSET AI Opportunity Scan survey, 9-21 August 2024.

### Figure 15: Experts' expectations regarding major positive impact of AI in the next three years, by area and SDG



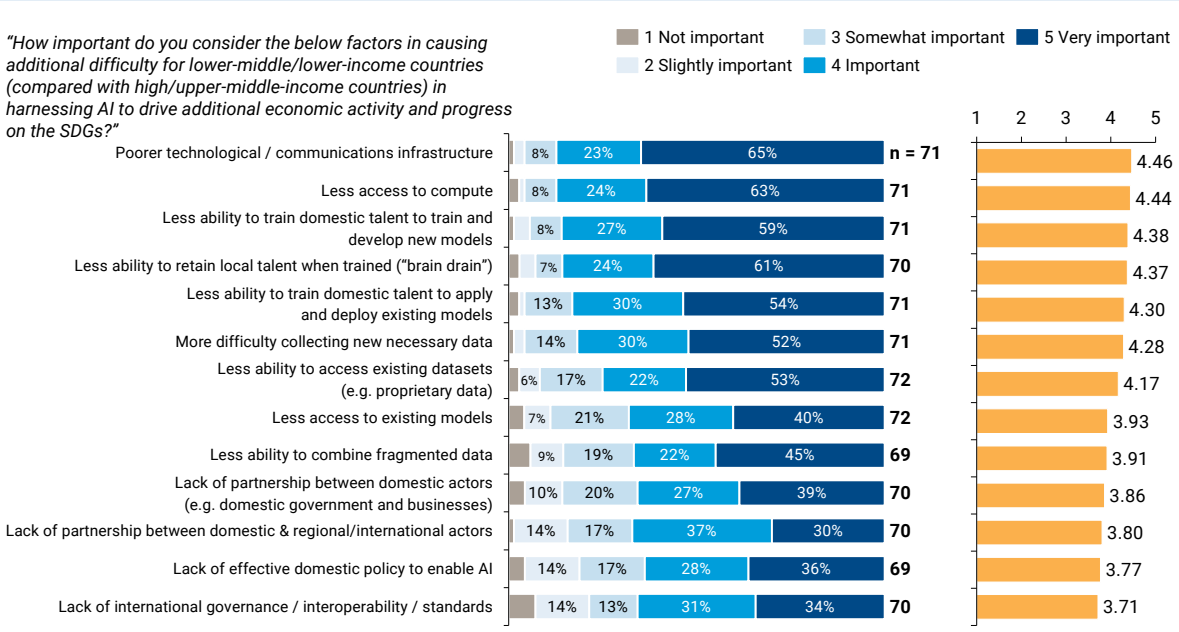

*Note:* Excludes "Don't know" / "No opinion" and blank responses. Only respondents reporting relevant knowledge were asked about lower-middle/lower-income countries. Did not ask about SDGs 8, 9 and 17. *Source:* OSET AI Opportunity Scan survey, 9-21 August 2024.

## Box 11: AI and the SDGs (continued)

Experts expected greater positive impact of AI in the next three years in higher-income countries across all areas surveyed, including accelerating scientific discoveries, increasing economic activity[b] and in the 14 SDG areas asked about (see fig. 15). Experts were most optimistic about AI's positive impact on health and education (SDGs 3 and 4), where 20–25 per cent of experts expected major or transformative positive impact of AI in the next three years in high/upper-middle-income countries. They were least optimistic regarding AI's positive impact on gender equality and inequalities (SDGs 5 and 10), with 2 in 3 expecting AI to have no positive impact on reducing inequalities within or between countries in either higher or lower-income countries.

AI may be expected to have earlier and greater impacts in higher-income countries, in part due to barriers holding back lower-middle and lower-income countries (see fig. 16). Missing enablers – from poorer infrastructure, to lack of domestic policy and international governance – were cited by more than half of respondents as important factors causing additional difficulty for lower-income countries in harnessing AI for economic activity and SDG progress.

### Figure 16: Experts' ratings of barriers to harnessing AI to drive additional economic activity and progress on the SDGs in lower-middle/lower-income countries




*Note:* Excludes "Don't know" / "No opinion" and blank responses. Only respondents reporting relevant knowledge were asked about lower-middle/lower-income countries.
*Source:* OSET AI Opportunity Scan survey, 9-21 August 2024.

These results underline the tentativeness of AI's eventual contribution to the SDGs, and how it remains highly dependent on missing enablers. This is particularly so in less developed countries, which already lack much of what more-developed countries have, from infrastructure to policy. Without cooperation to build capacity and facilitate access to key enablers, existing AI divides could further widen and become entrenched, limiting AI's ability to meaningfully contribute to progress on science, economic benefit and progress on the SDGs before 2030.

b The share of experts expecting "major positive impact" on increasing economic activity and accelerating scientific discovery over three years is higher in the first chart than the second chart. This may be due to the qualifier "by when do you expect it likely (50% chance or more) that AI will cause a major positive impact" (emphasis added) in the question responses depicted in the first chart, which is absent in the second.

**125** As we argued in our interim report, this depends largely on access to talent, compute and data, in ways that help cultural and linguistic diversity to flourish. Governance itself can be a key enabler, aligning incentives, engendering trust and sustainable practices while promoting collaboration across borders and subject domains. Without a comprehensive and inclusive approach to AI governance, the potential of AI to contribute positively to the SDGs could be missed, and its deployment could inadvertently reinforce existing disparities and biases.

**126** During extensive consultations conducted by the Advisory Body on topics such as education, health, data, gender, children, peace and security, creative industries and work, it became evident that AI holds substantial potential to significantly accelerate progress on the SDGs owing to its capabilities to boost innovation and delivery in various critical areas.

**127** However, AI is not a panacea for development challenges; it is one component within a broader set of solutions, and may even exacerbate some of these challenges, such as climate change. To truly unlock AI's potential to address societal challenges, collaboration among governments, academia, industry and civil society is crucial.

**128** The effectiveness of AI solutions depends on the quality and availability of data, and there are significant concerns about quality and representativeness in SDG-relevant data sets, which may fail to reflect relevant realities of certain populations. Further, AI solutions designed by AI experts without full knowledge of the intersecting domains of application often work in silico, and are not robust or impactful enough in actual development settings. That is the reason why AI solutions must be designed collaboratively and implemented with a deep understanding of their social, economic and cultural contexts. They must fit into broader local and national strategies for digital transformation and addressing digital divides.

**129** For example, AI capabilities in low- and lower-middle-income countries cannot be achieved without securing reliable electricity and Internet connectivity for running data centres, maintaining consistent computer operations, accessing global data sets, engaging in international research collaborations and using cloud-based AI tools. Therefore, we align ourselves with calls for investing in basic digital infrastructure, which is a prerequisite for developing countries to participate in and benefit from AI advancements.

**130** Building AI capacity is vital to ensuring that individuals across the globe, regardless of their region's development stage, can benefit from AI advancements. Strategic capacity-building, backed by adequate funding, is also essential to making AI technologies effective, sustainable and in the public interest – key for global development efforts. Below, we examine three critical enablers of national AI capabilities: the availability of technical expertise, access to compute and the availability of quality data. We then recommend specific actions.

## Talent

**131** The ability of societies around the world to participate in the beneficial outcomes of AI depends, first and foremost, on people. It is important to acknowledge that not every society needs cadres of computer scientists for building their own models. However, regardless of whether technology is bought, borrowed or built, human resources are needed to understand the capabilities and limitations of AI and harness AI-enabled use cases appropriately.

**132** Such a capacity – primarily in the public sector, but also in academia, business and civil society – will enhance the effectiveness of AI strategies and their implementation across various sectors. Nurturing AI-related human capacity will also be vital for preserving the world's cultural and linguistic diversity and building high-quality data sets for future AI development. In essence, this is capacity-building for public interest AI.

**133** Fostering human resources in diverse settings with youthful demographics, such as Africa (one third of the global workforce will be African within the first half of this century) will also be vital for the

future global talent pipeline. Enhancing the capacity of women in tech needs to be focused on closing the existing gender gap, on the one hand, and avoiding the gender gap in AI, on the other hand. The AI sector also needs more women in leadership positions to embed gender perspectives in AI governance. This starts with enabling increasing AI talent opportunities for girls.

## Compute

**134** Despite ongoing efforts to develop less compute-hungry approaches to AI, the need for access to affordable compute remains acute for training capable AI models.[37] This is one of the biggest barriers to entry in the field of AI for companies in the global South, but also many start-ups and small and medium-sized enterprises in the global North. Of the top 100 high-performance computing clusters in the world capable of training large AI models, none is hosted in a developing country.[38] There is only one African country among the top 300. Two countries account for half of the world's hyperscale data centres.[39]

**135** Most developers access compute infrastructure through cloud services; many have chosen to partner with the large cloud companies to secure reliable access to compute. It is possible that supply-chain issues may be resolved over time and competition may lead to more diverse sources of hardware, including high-performance chips for training models and AI accelerator chips for deployment on mobile devices. However, for the foreseeable future this constraint will remain a formidable barrier to a more globally inclusive AI innovation ecosystem.

**136** Ironically, compute capacity can lie idle or get outmoded quickly. There is potential value in fully using such capacity across depreciation cycles. However, there are hurdles to be overcome in terms of interoperability of different hardware configurations and scheduling demanding tasks, while preserving priority of time-critical use (such as for meteorological predictions).

**137** Moreover, without talent and data, compute alone is of no value. In the proposed global fund for AI, we consider how to address all three through a combination of financial and in-kind support.

## Data

**138** Although many discussions about the economics of AI focus on the "war for talent" and competition over hardware, such as graphics processing units (GPUs), data are no less vital. Facilitating access to quality training data at scale for training AI models by start-ups and small and medium-sized enterprises, as well as mechanisms to compensate data holders and creators of training data in rights-respecting ways, might be the most important enabler of a flourishing AI economy. Pooling data for the public interest in furthering specific SDGs is one key aspect (outlined in box 12), although it is not enough.

**139** In the context of AI, it is common to speak of "misuse" of data (e.g. infringing on privacy) or "missed" uses of data (failing to exploit existing data sets), but a related problem is "missing" data, which includes the large portions of the globe that are data poor. One example is health care, where around half of the leading data sets can be traced to a dozen organizations, with one in Europe, one in Asia and the rest in North America.[40]

**140** Another example is agriculture, where data are required across a complex interplay of factors (such as climate, soil and crop management practices) to enable useful AI models. Agriculture also often suffers from paucity of data and data-collection infrastructures. Dedicated efforts are needed to curate agriculture data sets particularly in the context of climate change resilience for food systems.

37 The Advisory Body is aware of a recent case where a company based in the global South spent $70 million for a 3-month training run for a large language model. Owning the graphics processing units (GPUs) instead of renting them from cloud service providers would have cost many times less.
38 See https://top500.org/statistics/sublist; proxy indicator since most high-performance computing clusters do not have GPUs and are of limited use for advanced AI.
39 UNCTAD, *Digital Economy Report 2021* (Geneva, 2021).
40 See https://2022.internethealthreport.org/facts.

## Box 12: Pooling data for the public interest in SDG areas

Collaborative data and AI commons – where shared models are cross-trained on pooled data – can play a key role in furthering the public interest where data would otherwise be missing or too sparse for AI benefits. Cross-functional and multi-domain data pools could enable the development of transdisciplinary data sets that encompass various SDG domains, derived from a variety of sources.

As an example, we can consider the complex issue of assessing the health impacts of climate change. To effectively address this challenge, a transdisciplinary approach is essential, integrating epidemiological data on the prevalence of diseases with meteorological data tracking climate variations. By pooling these distinct types of data from countries worldwide, in a privacy-preserving manner, researchers may be able to use AI to identify patterns and correlations that are not evident from isolated data sets.

Including data from all countries ensures comprehensive coverage, reflecting the global nature of climate change and capturing diverse environmental impacts and health outcomes across different regions. The transdisciplinary origins of the data enhance the predictive accuracy of models that aim to forecast future public health crises or natural disasters driven by climate change.

**141** Analogous to the problem of informal capital, those whose data are not captured – from birth records to financial transactions – may be unable to participate in the benefits of the AI economy, obtain government benefits or access credit. Use of synthetic data may only partially offset the need for new data sets.

**142** Feedback on our interim report noted that there was insufficient articulation of how current cross-jurisdictional practices around sourcing, use and non-disclosure of AI training data threaten rights and result in economic concentration. It was recommended that we consider how international AI governance could enable and catalyse more diverse participation in the leveraging of data for AI.

### Building a core public international AI capacity for common benefit

**143** Cutting across the above three enablers, advanced economies have both the capability and duty to facilitate AI capacity-building through international collaboration. In turn, they will benefit from a more broad-based digital economy, as well as quality talent and data flows. Importantly, everyone will benefit from the mainstreaming of good AI governance through such collaboration.

**144** Cooperation should focus on nurturing AI talent, boosting public AI literacy, improving capacity for AI governance, broadening access to AI infrastructure, promoting data and knowledge platforms suited to diverse cultural and regional needs, and enhancing uptake of AI applications and service capabilities. Only such a comprehensive approach can ensure equitable access to AI benefits, so that no nation is left behind.

**145** Many of the stakeholders we consulted emphasized that detailed strategies should be outlined to pool global resources together to build capacity, catalyse collective action towards equitable sharing of opportunities and close the digital divide.

## Capacity development network

**Recommendation 4: Capacity development network**

We recommend the creation of an AI capacity development network to link up a set of collaborating, United Nations-affiliated capacity development centres making available expertise, compute and AI training data to key actors. The purpose of the network would be to:

a. Catalyse and align regional and global AI capacity efforts by supporting networking among them;
b. Build AI governance capacity of public officials to foster development while furthering respect, protection and fulfilment of all human rights;
c. Make available trainers, compute and AI training data across multiple centres to researchers and social entrepreneurs seeking to apply AI to local public interest use cases, including via:
    i. Protocols to allow cross-disciplinary research teams and entrepreneurs in compute-scarce settings to access compute made available for training/tuning and applying their models appropriately to local contexts;
    ii. Sandboxes to test potential AI solutions and learn by doing;
    iii. A suite of online educational opportunities on AI targeted at university students, young researchers, social entrepreneurs and public sector officials; and
    iv. A fellowship programme for promising individuals to spend time in academic institutions or technology companies.

**146** From the Millennium Development Goals to the SDGs, the United Nations has long contributed to the development of capacities of individuals and institutions.[41] Through the work of UNESCO, WIPO and others, the United Nations has helped to uphold the rich diversity of cultures and knowledge-making traditions across the globe.

**147** At the same time, capacity development for AI would require a fresh approach, in particular cross-domain training to build a new generation of multidisciplinary experts in areas such as public health and AI, or food and energy systems and AI.

**148** Capacity would also have to be linked to outcomes through hands-on training in sandboxes[42] and collaborative projects pooling data and compute to solve shared problems. Risk assessments, safety testing and other governance methodologies would have to be built into this collaborative training infrastructure.

**149** Given the urgency and scale of the challenge, we suggest pursuing a strategic approach that pools and brokers access to compute through a network of high-performance computing nodes, incentivizes the development of critical data sets in SDG-relevant domains, promotes sharing of AI models, mainstreams best practices on AI governance and creates cross-domain talent for public interest AI, thus ensuring cross-cutting integration of human rights expertise.

**150** In other words, instead of chasing critical enablers one at a time through disjointed projects, we propose an **all-at-once, holistic strategy implemented through a chain of collaborating centres**. Emerging initiatives on capacity development and AI for the SDGs such as the International Computation and AI Network (ICAIN) initiative launched by Switzerland can help to create the initial critical mass for this strategy.

41 The United Nations University has long been committed to capacity-building through higher education and research, and the United Nations Institute for Training and Research has helped to train officials in domains critical to sustainable development. The UNESCO Readiness Assessment Methodology is a key tool to support Member States in their implementation of the UNESCO Recommendation on the Ethics of Artificial Intelligence. Other examples include the WHO Academy in Lyon, the UNCTAD Virtual Institute, the United Nations Disarmament Fellowship run by the Office for Disarmament Affairs and capacity-development programmes led by ITU and UNDP.

42 Sandboxes have been developed by various national institutions, including financial and medical authorities, such as the Infocomm Media Development Authority of Singapore.

**151** Ideally, there should be at least one or two nodes in each region of the world. The two centres of expertise participating in the Global Partnership on Artificial Intelligence could join the United Nations in supporting the capacity development network. Academic institutions and private sector contributors to capacity development could seek affiliation through the closest regional node or an international organization supporting the network.

**152** We are particularly encouraged by the prospect of cooperation among countries, for example through federated access to compute and related infrastructure. As noted in our interim report, the European Organization for Nuclear Research (CERN) offers useful lessons. A "distributed-CERN" reimagined for AI, networked across diverse States and regions, could expand opportunities for greater access to AI tools and expertise.

**153** We envision the capacity development network as a catalyser of national and regional capabilities and not as a concentrator of hardware, talent and data. By accelerating learning, it could catalyse national centres of excellence to stimulate the development of local AI innovation ecosystems, addressing the underlying coordination and implementation gaps mentioned in paragraphs 73, 80 and 81. National-level efforts could continue to employ diagnosis tools such as the UNESCO AI Readiness Assessment Methodology to help to assess initial maturity of countries, identify gaps and guide how road maps for capacity-building can be tailored per country and region, with the international network helping to address these gaps.

**154** The proposed AI office may be best placed to focus on strategy, partnerships and affiliation to link up nodes with the network, serving to connect rather than reinvent. It could also help to broker access to compute across the network. A node or nodes in the network could serve as leads on specific aspects of training, host sandboxes or high-performance computing clusters for AI model development. Nodes could collaborate on research programmes on topics such as privacy-preserving use of data, new methods to link different types of hardware or data sets for model training, as well as ways to use AI models in combination with each other.

**155** Our hope is that the network would also promote an alternative paradigm of AI technology development: bottom-up, cross-domain, cross-regional, open and collaborative. Given the rising energy and other costs of training and deploying AI models, and the prospect of compute lying unused, it makes sense to link computational resource for access on a time-sharing basis, while leveraging such access for advancing cross-domain talent, data and AI models for the SDGs.

## Global fund for AI

**Recommendation 5: Global fund for AI**

We recommend the creation of a global fund for AI to put a floor under the AI divide. Managed by an independent governance structure, the fund would receive financial and in-kind contributions from public and private sources and disburse them, including via the capacity development network, to facilitate access to AI enablers to catalyse local empowerment for the SDGs, including:

a. Shared computing resources for model training and fine-tuning by AI developers from countries without adequate local capacity or the means to procure it;
b. Sandboxes and benchmarking and testing tools to mainstream best practices in safe and trustworthy model development and data governance;
c. Governance, safety and interoperability solutions with global applicability;
d. Data sets and research into how data and models could be combined for SDG-related projects; and
e. A repository of AI models and curated data sets for the SDGs.

**156** The model of AI development and use proposed here is analogous to the original vision of the Internet: a distributed but connected infrastructure, interoperable and empowering. Public interest would be better served by a marketplace in which AI models and the infrastructure and data that they rely on are interoperable, well-governed and trustworthy. This would not be achieved automatically. Dedicated efforts backed by sufficient resources would be essential.

## Box 13: Global fund for AI: examples of possible investments

A relatively modest fund could help to create a minimum shared compute infrastructure for training small to medium-sized models. Such models have important SDG potential, for example, for training farmers in their local language.

This investment would also create a sandbox environment for developers to fine-tune existing open-source models with their own contextual and high-quality data. Access to the compute and sandbox infrastructure could be on a time share basis with reasonable usage fees contributing to meeting the maintenance and running costs.

A third use of the funding would be to help to curate gold standard data sets for select SDGs where the commercial incentive is absent. The model development, testing and data curation efforts could come together strategically in a powerful hands-on AI empowerment approach linked to concrete outcomes.

Finally, the fund could stimulate research and development, not only for contextually relevant development and SDG-related applications of AI, but also for interlinking of compute and models as well as new governance assessments.

**157** We approach this recommendation with humility, conscious of the powerful market forces shaping access to talent and compute, and of geopolitical competition pushing back against collaboration in the field of science and technology. Unfortunately, many countries may be unable to access training, compute, models and training data without international support. Existing funding efforts might also not be able to scale without such support.

**158** Levelling the playing field is, in part, a question of fairness. It is also in our collective interest to create a world in which all contribute to and benefit from a shared ecosystem. This is true not merely across States. Ensuring diverse access to AI model development and testing infrastructure would also help to address concerns about the concentration of disproportionate power in the hands of a handful of technology companies.

### Fund purpose and objective

**159** Our intention in proposing a fund is not to guarantee access to compute resources and capabilities that even the wealthiest countries and companies struggle to acquire. The answer may not always be more compute. We may also need different ways to leverage existing high-performance computing infrastructures, which are built for peak usage and not necessarily designed for AI. Perhaps there could be better ways to connect talent, compute and data.

**160** The purpose is, therefore, to address the underlying coordination and implementation gaps in paragraphs 73, 80 and 81 for those unable to access the requisite enablers through other means, to ensure that:

a. Countries in need can access AI enablers, putting a floor under the AI divide;
b. Collaboration on AI capacity development leads to habits of cooperation and mitigates geopolitical competition;
c. Countries with divergent regulatory approaches have incentives to develop common templates for governing data, models and applications for societal-level challenges related to the SDGs and scientific breakthroughs.

**161** The capacity built with resources from the global fund would be oriented towards the SDGs and the shared global governance of AI (box 13). It could, for instance, incorporate a “governance stack” for security and safety testing. This would help to mainstream best practices across the user base, while reducing the burden of validation for small users.

**162** This public interest focus makes the global fund complementary to the proposal for an AI capacity development network, to which the fund would channel resources. The fund would also provide an independent capacity for monitoring of impact. In this manner, we ensure that vast swathes of the world are not left behind, but instead empowered to harness AI for the SDGs in different contexts.

**163** It is in everyone's interest to ensure that there is cooperation in the digital world as in the physical world. Analogies can be made to the efforts to combat climate change, where the costs of transition, mitigation or adaptation do not fall evenly, and international assistance is essential to help resource-constrained countries, so that they can join the global effort to tackle a planetary challenge.

**164** Here, the focus is on using financing to help to ensure that a minimum capacity can be created in countries in different regions to understand AI's potential for sustainable development, adapt and build models for local needs, and join international collaborative efforts on AI.

### Fund governance

**165** The fund would source and pool in-kind contributions, including from private sector entities. Coordinating financial and in-kind contributions requires appropriate levels of independent oversight and accountability. Governance arrangements should be inclusive with board members drawn from government, the private sector, philanthropists, civil society and United Nations agencies. They should incorporate scientific and expert inputs, channelled (for example) through the proposed international scientific panel, and engender neutrality and trust for collaboration around data and model development.

### Fund operations

**166** The fund's operating model should be informed by lessons from pooled international research and development collaborations, such as CERN and Gavi, the Vaccine Alliance, as well as lessons from commercial platforms for timeshared infrastructure. It should also draw lessons from bodies such as the Global Fund (established in 2002 to pool resources to defeat HIV, tuberculosis and malaria)[43] and the Complex Risk Analytics Fund (which pools data in support of all stakeholders in crisis anticipation, prevention and response).

## Global AI data framework

**Recommendation 6: Global AI data framework**

We recommend the creation of a global AI data framework, developed through a process initiated by a relevant agency such as the United Nations Commission on International Trade Law and informed by the work of other international organizations, for:

a. Outlining data-related definitions and principles for global governance of AI training data, including as distilled from existing best practices, and to promote cultural and linguistic diversity;
b. Establishing common standards around AI training data provenance and use for transparent and rights-based accountability across jurisdictions; and
c. Instituting market-shaping data stewardship and exchange mechanisms for enabling flourishing local AI ecosystems globally, such as:
    i. Data trusts;
    ii. Well-governed global marketplaces for exchange of anonymized data for training AI models; and
    iii. Model agreements for facilitating international data access and global interoperability, potentially as techno-legal protocols to the framework.

**167** In our consultations, we heard that although there have been plenty of proposals to promote wider access to data and data-sharing arrangements to create more diverse AI ecosystems, not many have materialized so far. This is a critical gap in developing inclusive and vibrant AI ecosystems.

43 See https://www.theglobalfund.org/en/about-the-global-fund.

**168** Part of the answer is in transparency on cultural, linguistic and other traits of AI training data. Identifying underrepresented or "missing" data is also helpful. Related to this is the promotion of "data commons" that incentivize curation of training data for multiple actors. Such initiatives could create best practices by demonstrating how design can embed techno-legal frameworks for privacy, data protection, interoperability and the equitable use of data, and human rights.

**169** The data marketplaces for AI are something of a "wild west" today. The idea of "grab what you can and hide it in opaque algorithms" seems to be one operating principle; another is exclusive contractual arrangements for access to proprietary data enforceable in select jurisdictions. Such exclusive relationships lie behind the United Kingdom Competition and Market Authority's concern that "the [Frontier Model] sector is developing in ways that risk negative market outcomes".[44]

**170** We consider it thus vital to launch a global process that involves a variety of actors, including nations at different levels of development, supported by relevant international organizations from the United Nations family and beyond (OECD, WIPO and the World Trade Organization), to create "guard rails" and "common rails" for flourishing AI training data ecosystems. The outcomes of this process need not be binding law but model contracts and techno-legal arrangements. These facilitative arrangements can be developed one by one, as protocols to a framework of principles and definitions.

**171** While the full details are beyond our scope, key principles for a global AI data framework would include interoperability, stewardship, privacy preservation, empowerment, rights enhancement and AI ecosystem enablement.

**172** We are mindful that antitrust and competition policy remains domains of national and regional authorities. However, international collective action can facilitate cross-border access to training data for local AI start-ups not available domestically.

**173** The United Nations is uniquely positioned to support the establishment of global principles and practical arrangements for the governance and use of AI training data, building on years of work by the data community and integrating it with recent developments on AI ethics and governance. This is analogous to efforts of the United Nations Commission on International Trade Law on international trade, including on legal and non-legal cross-border frameworks, and enabling digital trade and investment via model laws on e-commerce, cloud-computing and identity management.

**174** Likewise, the Commission on Science and Technology for Development and the Statistical Commission have on their agenda data for development and data on the SDGs. There are also important issues of content, copyright and protection of indigenous knowledge and cultural expression being considered by WIPO.

**175** The framework proposed here would be without prejudice to national or regional frameworks for data protection and would not create new data-related rights nor prescribe how existing rights apply internationally, but would have to be designed in a way that prevents capture by commercial or other interests that could undermine or preclude rights protections. Rather, a global AI data framework would address transversal issues of availability, interoperability and use of AI training data. It would help to build common understanding on how to align different national and regional data protection frameworks.

44 Competition and Markets Authority, *AI Foundation Models: Technical Update Report* (London, 2024).

## Box 14: Securing data for training AI models: data empowerment, data trusts and cross-border data flow arrangements

There are many circumstances in which data need to be protected (including for privacy, commercial confidentiality, intellectual property, safety and security), but where there would also be benefits to individuals and society in making it available for training AI models.

Data rights in law are generally rights to prevent actions in relation to data. Data privacy rights are also personal to individuals. The constitution of data rights can make it difficult to exercise data rights in a flexible way that enables data to be used for some purposes without losing the rights, and to do that collectively as a group. Even when it is possible to control permissions flexibly and positively, this tends to require more time, technical expertise and confidence than most people and organizations have.

Mechanisms that enable owners and subjects of data to allow safe and limited use of their data, while maintaining their rights, can be described as means of data empowerment. Data empowerment can make many more people and groups in society into active partners and stakeholders in AI, and not only subjects of data. There are already tools in development for managing access securely, including data trusts and privacy protecting applications for steering cross-border data flows.

Data trusts are mechanisms that make it possible for individuals and organizations to provide access to their data collectively, with access in the control of trustees. The data-owners can set the terms for access, use and purpose, which the trustees exercise. The owners and subjects of the data retain their legal rights while contributing to shared objectives. An AI model trained on this data could be expected to perform more accurately than one that lacked this specific input, and thus better serve the well-being of that particular group or of society more broadly.

Mechanisms for managing access and use, and access across borders in particular, all rely on dedicated legal frameworks. Using these mechanisms in practice also requires adaptation to the needs and contexts of sectors and communities. Gaps in data stewardship should be identified and closed. Successful and widespread use of these mechanisms in the future would depend on technical assurance and maintaining the trust of contributors of data.

We thus propose that more support is given to the further development of these tools, and to identifying the areas where use of them for training AI could deliver the greatest public value.

**176** Steps to address these issues at the national and regional level are promising, with the public and private sector paying more attention to best practices. Yet without a global framework governing AI training data sets, commercial competition invites a race to the bottom between jurisdictions on access and use requirements, making it difficult to govern the AI value chain internationally. Only global collective action can promote a race to the top in the governance of the collection, creation, use and monetization of AI training data in ways that further interoperability, stewardship, privacy preservation, empowerment and rights enhancement.

**177** Equally, such action is necessary to promote flourishing local AI ecosystems and limit further economic concentration. These measures could be complemented by promotion of data commons and provisions for hosting data trusts in areas relevant to the SDGs (see box 14). The development of these templates and the actual storage and analysis of data held in commons or in trusts could be supported by the capacity development network and the global fund for AI.

# D. Coherent effort

**178** By promoting a common understanding, common ground and common benefits, the proposals above seek to address the gaps identified in the emerging international AI governance regime. The gaps in representation, coordination and implementation can be addressed through partnerships and collaboration with existing institutions and mechanisms.

**179** However, without a dedicated focal point in the United Nations to support and enable soft coordination among such and other efforts, and to ensure that the United Nations system speaks with one voice regarding AI, the world will lack the inclusively networked, agile and coherent approach required for effective and equitable governance of AI.

**180** For these reasons, we propose the creation of a small, agile capacity in the form of an AI office within the United Nations Secretariat.

## AI office in the United Nations Secretariat

**Recommendation 7: AI office within the Secretariat**

We recommend the creation of an AI office within the Secretariat, reporting to the Secretary-General. It should be light and agile in organization, drawing, wherever possible, on relevant existing United Nations entities. Acting as the "glue" that supports and catalyses the proposals in this report, partnering and interfacing with other processes and institutions, the office's mandate would include:

a. Providing support for the proposed international scientific panel, policy dialogue, standards exchange, capacity development network and, to the extent required, the global fund and global AI data framework;
b. Engaging in outreach to diverse stakeholders, including technology companies, civil society and academia, on emerging AI issues; and
c. Advising the Secretary-General on matters related to AI, in coordination with other relevant parts of the United Nations system to offer a whole-of-United Nations response.

**181** During our consultations, it became clear that the case for an agency with reporting, monitoring, verification and enforcement powers has not been made thus far, and there has not yet been much appetite on the part of Member States for an expensive new organization.

**182** We, therefore, focus on the value that the United Nations can offer, mindful of the shortcomings of the United Nations system, as well as what could realistically be achieved within a year. In this regard, we propose a light, agile mechanism to act as the "glue" that holds together processes promoting a common understanding, common ground and common benefits, and enables the United Nations system to speak with one voice in the evolving international AI governance ecosystem.

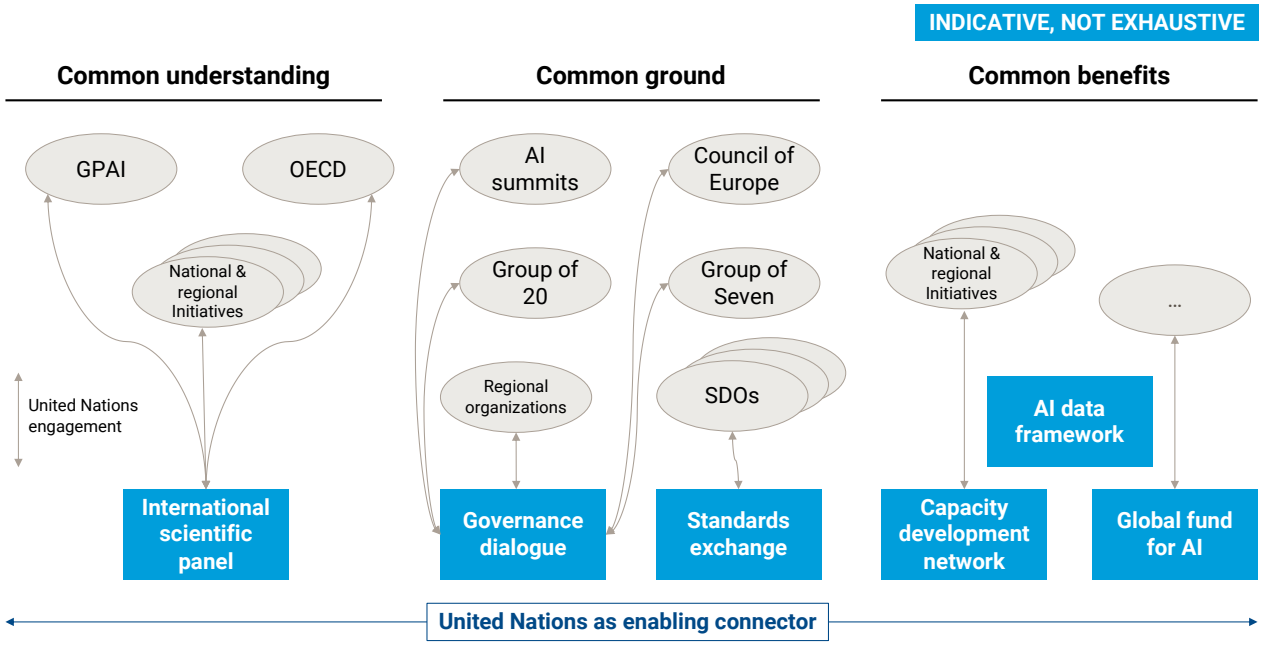
**183** Just as countries have set up dedicated institutes and offices focused on the national, regional and international governance of AI,[45] we see the need for a capacity that services and supports the international scientific panel on AI and AI policy dialogue, and catalyses the AI standards exchange and capacity development network – with lower overheads and transaction costs than if each were supported by different organizations.

**184** An AI office within the United Nations Secretariat, reporting to the Secretary-General, would have the benefit of connections throughout the United Nations system, without being tied to one part of it. That is important because of the uncertain future of AI and the strong likelihood that it will permeate all aspects of human endeavour.

**185** A small and agile AI office would be well positioned to connect various domains and organizations on AI governance issues to help to address gaps dynamically, working to amplify existing efforts within and beyond the United Nations. By bridging

45 Including Canada, Japan, the Republic of Korea, Singapore, the United Kingdom, the United States and the European Union.

## Figure 17: Proposed role of the United Nations in the international AI governance ecosystem




*Abbreviations*: GPAI, Global Partnership on Artificial Intelligence; OECD, Organisation for Economic Co-operation and Development; SDOs, standards development organizations.

and connecting other initiatives, such as those led by regional organizations and other plurilateral initiatives, it can help to lower the costs of cooperation between them.

**186** Such a body should champion inclusion and partner rapidly to accelerate coordination and implementation, drawing, as a first priority, on existing resources and functions within the United Nations system. It could be staffed in part by United Nations personnel seconded from relevant specialized agencies and other parts of the United Nations system. It should engage multiple stakeholders, including civil society, industry and academia, and develop partnerships with leading organizations outside of the United Nations, such as OECD.

**187** The AI office would ensure information-sharing across the United Nations system and enable the system to speak with authority and with one voice. Box 15 lists possible functions and early deliverables of such an office.

**188** This recommendation is made on the basis of a clear-eyed assessment as to where the United Nations can add value, including where it can lead, where it can fill gaps, where it can aid coordination and where it should step aside, working in close partnership with existing efforts (see fig. 17). It also brings the benefits of existing institutional arrangements, including pre-negotiated funding and administrative processes that are well understood.

**189** The evolving characteristics of AI technology should be considered. There is a high probability of technical breakthroughs that will dramatically change the current AI model landscape. Such an AI office should be effectively in place to adjust governance frameworks to the evolving landscapes and respond to unforeseen developments concerning AI technology.

## Box 15: Possible functions and first-year deliverables of the AI office

The AI office should have a light structure and aim to be agile, trusted and networked. Where necessary, it should operate in a "hub and spoke" manner to connect to other parts of the United Nations system and beyond.

Outreach could include serving as a key node in a so-called soft coordination architecture between Member States, plurilateral networks, civil society organizations, academia and technology companies in a regime complex that weaves together to solve problems collaboratively through networking, and as a safe, trusted place to convene on relevant topics. Ambitiously, it could become the glue that helps to hold such other evolving networks together.

Supporting the various initiatives proposed in this report includes the important function of ensuring inclusiveness at speed in delivering outputs such as scientific reports, governance dialogue and identifying appropriate follow-up entities.

**Common understanding**:

- Facilitate recruitment of and support the international scientific panel.

**Common ground**:

- Service policy dialogues with multi-stakeholder inputs in support of interoperability and policy learning. An initial priority topic is the articulation of risk thresholds and safety frameworks across jurisdictions
- Support ITU, ISO/IEC and IEEE on setting up the AI standards exchange.

**Common benefits**:

- Support the AI capacity development network with an initial focus on building public interest AI capacity among public officials and social entrepreneurs. Define the initial network vision, outcomes, governance structure, partnerships and operational mechanisms.
- Define the vision, outcomes, governance structure and operational mechanisms for the global fund for AI, and seek feedback from Member States, industry and civil society stakeholders on the proposal, with a view to funding initial projects within six months of establishment.
- Prepare and publish an annual list of prioritized investment areas to guide both the global fund for AI and investments outside that structure.

**Coherent effort**:

- Establish lightweight mechanisms that support Member States and other relevant organizations to be more connected, coordinated and effective in pursuing their global AI governance efforts.
- Prepare initial frameworks to guide and monitor the AI office's work, including a global governance risk taxonomy, a global AI policy landscape review and a global stakeholder map.
- Develop and implement quarterly reporting and periodic in-person presentations to Member States on the AI office's progress against its workplan and establish feedback channels to support adjustments as needed.
- Establish a steering committee jointly led by the AI office, ITU, UNCTAD, UNESCO and other relevant United Nations entities and organizations to accelerate the work of the United Nations in service of the functions above, and review progress of the accelerated efforts every three months.
- Promote joint learning and development opportunities for Member State representatives to support them to carry out their responsibilities for global AI governance, in cooperation with relevant United Nations entities and organizations such as the United Nations Institute for Training and Research and the United Nations University.

# E. Reflections on institutional models

**190** Discussions about AI often resolve into extremes. In our consultations around the world, we engaged with those who see a future of boundless opportunities provided by ever-cheaper, ever-more-helpful AI systems. We also spoke with those wary of darker futures, of division and unemployment, and even extinction.

**191** We do not know what the future may transpire. We are mindful that the technology may go in a direction that does away with this duality. In this report, we have focused on the near-term opportunities and risks, based on science. The recommendations outlined herein offer our best hope for reaping the benefits of AI while minimizing and mitigating the risks. We are also mindful of the practical challenges to international institution-building on a larger scale. This is why we are proposing a networked institutional approach with light and agile support.

**192** If or when risks become more acute and the stakes for opportunities escalate, however, such calculations will change. The world wars led to the modern international system; the development of ever-more-powerful weapons led to regimes limiting their spread and promoting peaceful uses of the underlying technologies.

**193** Evolving understanding of our common humanity led to the modern human rights system and our ongoing commitments to the SDGs for all. Climate change evolved from a niche concern to a global challenge. AI may similarly rise to a level that requires more resources and more authority than proposed in this report.

**194** Our terms of reference included considering the functions, forms and timelines for a new international agency for AI. We conclude the present report with some reflections on the issue, although we do not currently recommend establishing such an agency.

## An international AI agency?

**195** If the risks of AI become more serious, and more concentrated, it might become necessary for Member States to consider a more robust international institution with monitoring, reporting, verification, and enforcement powers.

**196** There is precedent for such evolution. From the Hague Conventions of 1899 and 1907, to the 1925 Geneva Protocol, and culminating in the Chemical Weapons Convention in 1993, dual-use chemicals have long been subject to limits on access, with protocols for storage and usage, and a ban on weaponization.

**197** Biological weapons have also been banned, along with periodic limits on research, such as the limits on recombinant DNA or gene-splicing in 1975. These emphasized containment as an essential consideration in experiment design, with the level of containment tied to the estimated risk. Certain classes of high-risk experiment for which containment could not be guaranteed were essentially prohibited. Other examples included research that threaten to cross fundamental ethical lines, such as ongoing restrictions on human cloning – an example of the kind of "red line" that may one day be needed in the context of AI research, along with effective cooperation regarding enforcement.

**198** Continued scientific assessments are also a feature of some of these frameworks, for example the Scientific Advisory Board of the Organisation for the Prohibition of Chemical Weapons and article XII of the Biological Weapons Convention.

**199** The comparison between AI and nuclear energy is well known. From the day the atom was split, it was clear to scientists that this technology could be used for good – even though their research was directed at constructing a new and terrible weapon. Then, as now, it was telling that leading scientists were among those who called most ardently for a limit on this new technology.

**200** The grand bargain at the heart of the International Atomic Energy Agency (IAEA) was that nuclear energy's beneficial purposes could be shared – in energy production, agriculture and medicine – in exchange for guarantees that it would not be further weaponized. As the nuclear non-proliferation regime shows, good norms are necessary but not sufficient for effective regulation.

**201** The limits of the analogy are clear. Nuclear energy involves a well-defined set of processes related to specific materials that are unevenly distributed, and much of the materials and infrastructure needed to create nuclear capability are controlled by nation States. AI is an amorphous term; its applications are extremely wide and its most powerful capabilities span industry and States. The grand bargain of IAEA focused on weapons that are expensive to build and difficult to hide; weaponization of AI promises to be neither.

**202** An early idea – pooling of nuclear fuel for peaceful purposes – did not work out as planned. On the pooling of resources for sharing benefits of technology, a more AI-appropriate analogy may be CERN, which pools funding, talent and infrastructure. However, there are limits to the comparison, given the difference between experimental fundamental physics and AI, which requires a more distributed approach.

**203** Another imperfect analogy is organizations such as the International Civil Aviation Organization (ICAO) and the International Maritime Organization (IMO). The underlying technologies of transportation are well established, and their civilian applications can be easily demarcated from military ones – this is not the case with general-purpose AI. The network of national regulatory authorities that apply the international norms developed in the framework of ICAO and IMO is also well established. Safety, facilitation of commercial activity, and interoperability are in focus. Compliance is not handled in a top-down manner.

**204** There are other approaches to compliance that can inspire. Financial risk management benefits from mechanisms such as the Financial Stability Board (FSB) and the Financial Action Task Force (FATF), without recourse to treaties.

**205** Eventually, some kind of mechanism at the global level might become essential to formalize red lines if regulation of AI needs to be enforceable. Such a mechanism might include formal CERN-like commitments for pooling resources for collaboration on AI research and sharing of benefits as part of the bargain.

**206** Given the speed, autonomy and opacity of AI systems, however, waiting for a threat to emerge may mean that any response will come too late. Continued scientific assessments and policy dialogue would ensure that the world is not surprised. Any decision to begin a formal process would, naturally, lie with Member States.

**207** Possible thresholds for such a move could include the prospect of uncontrollable or uncontainable AI systems being developed, or the deployment of systems that are unable to be traced back to human, corporate or State actors. They could also include indications that AI systems exhibit qualities that suggest the emergence of "superintelligence", although this is not present in today's AI systems.

**208** Establishing a watching brief, drawing on diverse and distinguished experts to monitor the horizon, is a reasonable first step. The scientific panel could be tasked with commissioning research on this question, as part of its quarterly research digest series. Over time, the policy dialogue could be an appropriate forum for sharing information about AI incidents, such as those that stretch or exceed the capacities of existing agencies, analogous to the practices of IAEA for mutual reassurance on nuclear safety and nuclear security, or the World Health Organization (WHO) on disease surveillance.

**209** The functions of a proposed international AI agency could draw on the experience of relevant agencies, such as IAEA, the Organisation for the Prohibition of Chemical Weapons, ICAO, IMO, CERN and the Biological Weapons Convention. They could include:

- Developing and promulgating standards and norms for AI safety;
- Monitoring AI systems that have the potential to threaten international peace and security, or cause grave breaches of human rights or international humanitarian law;

- Receiving and investigating reports of incidents or misuses, and reporting on serious breaches;
- Verifying compliance with international obligations;
- Coordinating accountability, emergency responses and remedies for harm regarding AI safety incidents;
- Promoting international cooperation for peaceful uses of AI.

**210** A tailored approach to designing any future AI agency would be required, drawing on lessons from other institutions as appropriate (see box 16).

## Box 16: Lessons learned from past global governance institutions

AI is a unique set of technologies with risks and societal impacts that transcend borders. However, it is not the first set of technologies that have led to global AI governance arrangements. Civil aviation, climate change, nuclear power and terrorism finance are also complex and multidimensional domains that have warranted a global response.

Some of these domains, such as civil aviation, climate change and nuclear power, have led to the creation of new United Nations institutions. Others, notably the protection of global financial flows, have led to bodies that are not treaty-based and yet they have delivered robust normative frameworks, effective market-based enforcement mechanisms and strong public-private partnerships.

As we draw parallels between these institutional responses and nascent efforts to do the same for AI, we should not focus too heavily on which institutional analogue is most suitable for the AI problem set. Our interim report foreshadowed that we should look instead at which governance functions are needed for effective and inclusive global AI governance, and what we can learn from past global governance endeavours.

One lesson is that the development of a shared scientific and technical understanding of the problem is necessary to trigger a commonly accepted policy response. Here, IPCC, which continues to address the risks of climate change, is a useful model. It offers an example of how an inclusive approach to crafting reports and developing scientific consensus in a constantly evolving area can level the playing field for researchers and policymakers and create the shared understanding that is essential for effective policymaking. The process of drafting and disseminating IPCC reports and global stock takes, although not without challenges, has been centrally important to building a shared understanding and common knowledge base, lowering the costs of cooperation and steering the Conference of the Parties to the United Nations Framework Convention on Climate Change towards concrete policy deliverables.

For AI, as the technology evolves, it will be just as important to develop a shared scientific understanding. As the capabilities of AI systems continue to advance and potential risks may exceed known effective approaches to mitigating them, the international scientific panel could be evolved to match emerging needs.

A second lesson is that multi-stakeholder collaboration can deliver strong standards and promote quick responses. Here, ICAO and FATF offer useful examples of how to govern a highly technical issue across borders. In civil aviation, the ICAO safety and security standards, developed by industry and government experts and enforced through market access restrictions, ensure that a plane that takes off from, for example, New York can land in Geneva without triggering new safety audits. A combination of ICAO-led safety audits and Member State-driven audits ensure consistent implementation, even as the technology evolves.

FATF – established by the G7 in 1989 to address money-laundering – offers another example of how soft law institutions can promote common standards and implementation. Its peer review system for monitoring is

## Box 16: Lessons learned from past global governance institutions (continued)

flexible; and widespread acceptance of its recommendations has created reputational costs for those companies and Member States that fail to comply. Even as the risks to international financial flows have evolved, most significantly with the rise of terrorism and proliferation finance, the nimble structure and normative framework of FATF have allowed it to respond quickly and keep pace with complex challenges.

In their own unique ways, both ICAO and FATF have created widely recognized international standards, domestic frameworks for measuring compliance, and interoperable systems for responding to certain classes of risks and challenges that manifest across jurisdictions. ICAO enforces via market access incentives and restrictions, while FATF creates reputational risk for non-compliance. Both offer useful templates for AI, as they demonstrate how governments and other stakeholders can work together to create a web of interconnected norms and regulations and create costs for non-compliance.

A third lesson is that global coordination is often vital for monitoring and taking action in response to severe risks with the potential for widespread impact. FSB and IAEA models offer key examples. Established in 2009, FSB was created by the G20 countries to monitor and warn against systemic risks to the international financial system. Its unique composition of G20 finance officials and international financial and development organizations has allowed it to be nimble, adept and inclusive when coordinating efforts to identify global financial risks.

The IAEA approach to nuclear safeguards offers a different model. Its comprehensive safeguards agreements, signed by 182 States, are part of the most wide-ranging United Nations regime for ensuring compliance. By using a combination of inspections and monitoring – as well as the threat of Security Council action – IAEA offers perhaps the most visible censure of Member States who fail to comply.

Both FSB and IAEA demonstrate how international coordination is fundamental to monitoring severe risks. As the risks of AI become clearer and more pronounced, there may be a similar need to create a new AI-focused institution to maximize coordination efforts and monitor severe and systemic risks, so that Member States can, wherever possible, intervene to stay ahead of those risks.

A fourth lesson is that it is important to create inclusive access to the resources needed for research and development, along with their benefits. The experiences of CERN and IAEA are both instructive. CERN brings together world-class scholars and physicists to perform complex research into particle accelerators and other projects that are meant to benefit humanity. It also offers training to physicists and engineers.

Similarly, IAEA facilitates access to technology, in this case nuclear energy and ionizing radiation. The basic trade-off is simple: Member States comply with nuclear safeguards and IAEA offers technical assistance towards the use of peaceful nuclear power. In this regard, IAEA provides an inclusive approach to spreading the benefits of technology to developing countries. Its facilitation of a network of centres of excellence on nuclear security is similar to our recommendation for a networked approach to capacity-building.

As we have explained above, AI is a set of technologies whose benefits need to be shared in a more inclusive and equitable manner, especially with countries in the global South. This is why we have recommended both an AI capacity development network and a global fund for AI. As we learn more about AI through the work of the international scientific panel, and as the responsible deployment of AI in support of the SDGs becomes even more pressing, United Nations Member States may want to institutionalize this function more widely. If they do so, they should look to draw lessons from CERN and IAEA as useful models for supporting broader access to resources, as part of an overall global AI governance structure.

# 5. Conclusion: a call to action

211 As experts, we remain optimistic about the future of AI and its potential for good. That optimism depends, however, on realism about the risks and the inadequacy of structures and incentives currently in place. We also need to be realistic about international suspicions that could get in the way of the global collective action needed for effective and equitable governance. The technology is too important, and the stakes are too high, to rely only on market forces and a fragmented patchwork of national and multilateral action.

212 We need to be active and purposeful. Beyond the duality of opportunity and risk is the challenge of rapid and cross-cutting change. AI's downstream impact may leave few people untouched. To place its governance in the hands of a few developers, or the countries that host them, will create a deeply unfair situation where the impacts of developing, deploying and using AI are imposed on most people without their having any say in the decisions for doing so.

213 The past year of global attention and discussion on AI governance has given us hope. There are divergences across countries and sectors, but also a strong desire for dialogue. Engaging diverse experts, policymakers, businesspeople, researchers and advocates – across regions, genders and disciplines – has shown us that diversity need not lead to discord, and dialogue can lead to common ground and collaboration.

214 Sometimes we hesitated: Should we be pragmatic and focus on what seems feasible? Or should we aim high with lofty ambition? In the end, we resolved to do both. Our proposals reflect a comprehensive vision for an equitable and effective global AI governance regime, with careful thought on how it can be implemented, step by step.

215 We are grateful to the many people, organizations and Member States that have contributed to our deliberations, including the representatives of United Nations agencies and Secretariat personnel who offered discerning assessments of the capabilities and the limitations of the United Nations in this complex area. The issue of AI governance is not only about managing the implications of this technology. Also at stake is the future of multilateral and multi-stakeholder cooperation.

216 When we look back in five years, the technology landscape could appear drastically different from today. However, if we stay the course and overcome hesitation and doubt, we can look back in five years at an AI governance landscape that is inclusive and empowering for individuals, communities and States everywhere. It is not technological change itself, but how humanity responds to it, that ultimately matters.

217 We believe that the functions and forms recommended in this report, if implemented in good faith, can deliver an agile and adaptable regime that stays in step with AI's march and helps to reap its benefits and address its risks. They can help us to spot problems and opportunities in time, use shared principles and frameworks to align international action, promote international cooperation, and build capacity of individuals and institutions to deal with change.

**218** The implementation of the recommendations in the present report may also encourage new ways of thinking: a collaborative and learning mindset, multi-stakeholder engagement and broad-based public engagement. The United Nations can be the vehicle for a new social contract for AI that ensures global buy-in for a governance regime that protects and empowers us all. Such a contract will ensure that opportunities are fairly accessed and distributed, and the risks are not loaded onto the most vulnerable – or passed on to future generations, as we have seen tragically with climate change.

**219** As a group and as individuals from across many fields of expertise, organizations and parts of the world, we look forward to continuing this crucial conversation. Together with the many we have connected with on this journey, and the global community that they represent, we hope that this report contributes to our combined efforts to govern AI for humanity.

# Annexes

## Annex A: Members of the High-level Advisory Body on Artificial Intelligence

Anna Abramova

Omar Sultan Al Olama

Latifa Al-Abdulkarim

Estela Aranha

Carme Artigas (Co-Chair)

Ran Balicer

Paolo Benanti

Abeba Birhane

Ian Bremmer (Co-Rapporteur)

Anna Christmann

Natasha Crampton

Nighat Dad

Vilas Dhar

Virginia Dignum

Arisa Ema

Mohamed Farahat

Amandeep Singh Gill

Wendy Hall

Rahaf Harfoush

Ruimin He

Hiroaki Kitano

Haksoo Ko

Andreas Krause

James Manyika (Co-Chair)

Maria Vanina Martinez Posse

Seydina Moussa Ndiaye

Mira Murati

Petri Myllymäki

Alondra Nelson

Nazneen Rajani

Craig Ramlal

Emma Ruttkamp-Bloem

Marietje Schaake (Co-Rapporteur)

Sharad Sharma

Jaan Tallinn

Philip Thigo

Jimena Sofia Viveros Álvarez

Zeng Yi

Zhang Linghan

# Annex B: Terms of reference of the High-level Advisory Body on Artificial Intelligence

The High-level Advisory Body on Artificial Intelligence, convened by the United Nations Secretary-General, will undertake analysis and advance recommendations for the international governance of artificial intelligence. The Body's initial reports will provide high-level expert and independent contributions to ongoing national, regional, and multilateral debates.

The Body will consist of 38 members from governments, private sector, civil society, and academia, as well as a member Secretary. Its composition will be balanced by gender, age, geographic representation, and area of expertise related to the risks and applications of artificial intelligence. The members of the Body will serve in their personal capacity.

The Body will engage and consult widely with governments, private sector, academia, civil society, and international organizations. It will be agile and innovative in interacting with existing processes and platforms as well as in harnessing inputs from diverse stakeholders. It could set up working parties or groups on specific topics.

The members of the Body will be selected by the Secretary-General based on nominations from Member States and a public call for candidates. It will have two Co-Chairs and an Executive Committee. All stakeholder groups will be represented in the Executive Committee.

The Body shall be convened for an initial period of one year, with the possibility of extension by the Secretary-General. It will have both in-person and online meetings.

The Body will prepare a first report by 31 December 2023 for the consideration of the Secretary-General and the Member States of the United Nations. This first report will present a high-level analysis of options for the international governance of artificial intelligence.

Based on feedback to the first report, the Body will submit a second report by 31 August 2024 which may provide detailed recommendations on the functions, form, and timelines for a new international agency for the governance of artificial intelligence.

The Body shall avoid duplication with existing forums and processes where issues of artificial intelligence are considered. Instead, it shall seek to leverage existing platforms and partners, including UN entities, working in related domains. It shall fully respect current UN structures as well as national, regional, and industry prerogatives in the governance of artificial intelligence.

The deliberations of the Body will be supported by a small secretariat based in the Office of the Secretary-General's Envoy on Technology and be funded by extrabudgetary donor resources.

# Annex C: List of consultation engagements in 2024

| Engagement | Date, 2024 | Region |
|---|---|---|
| UNESCO Slovenia | 5 Jan. | Europe |
| Secretary-General's Scientific Advisory Board | 8 Jan. | Global |
| Presentation to Member States on the interim report | 12 Jan. | Global |
| World Economic Forum in Davos | 24 Jan. | Europe |
| Association of Southeast Asian Nations (ASEAN) Digital Senior Officials' Meeting | 30 Jan. | Asia |
| World Government Summit | 12 Feb. | Middle East |
| Montreal Institute for Learning Algorithms (Mila - Quebec AI Institute) | 14 Feb. | North America |
| Berlin Consultation | 15 Feb. | Europe |
| Euro-Asian IT Forum | 20 Feb. | Global |
| Mobile World Congress | 26 Feb. | Europe |
| Moscow State Institute of International Relations | 28 Feb. | Europe |
| Royal Society workshop on international AI governance | 28 Feb. | Europe |
| Foreign Ministries Science & Technology Advice Network | 28 Feb. | Global |
| OECD-African Union AI dialogue | 4 Mar. | Europe |
| Brussels Consultation | 5 Mar. | Europe |
| World Bank, Global Digital Summit | 5 Mar. | North America |
| Open Science and Artificial Intelligence: ethical issues webinar | 5 Mar. | Eastern Europe |
| UNESCO Digital Transformation Dialogue | 6 Mar. | Europe |
| Inter-Parliamentary Union | 6 Mar. | Global |
| 47th session of the High-level Committee on Programmes | 11 Mar. | Global |
| Global Youth Summit on Digital Rights | 13 Mar. | Latin America |
| Group of Seven (G7) summit on AI in Trento, Italy | 15 Mar. | Europe |
| Kick-off consultative network meetings, 18–19 March | 18 Mar. | Global |
| 68th session of the Commission on the Status of Women | 21 Mar. | North America |
| Advisory Body update to Member States | 25 Mar. | Global |
| African Observatory on Responsible AI | 25 Mar. | Africa |
| AI for sustainable and inclusive futures conference - French Development Agency | 26 Mar. | Europe |
| Shaping Global Norms: collective feedback | 28 Mar. | Africa |
| Innovate Switzerland | 2 Apr. | Europe |
| OSET visit to China, 9–12 April | 9 Apr. | Asia |
| Russian Internet Governance Forum | 9 Apr. | Eastern Europe |
| Wharton Cypher Days - Finance | 12 Apr. | North America |
| Silicon Valley visit | 15 Apr. | North America |
| Stanford, AI+Policy Symposium: A Global Stocktaking | 16 Apr. | North America |
| United Nations Commission on Science and Technology for Development | 16 Apr. | Europe |
| Group of 20 (G20) Digital Economy, 16–18 April, Brazil | 17 Apr. | Latin America |
| Advisory Body update to Member States | 22 Apr. | Global |
| United Nations University, Macau AI Conference, 24–25 April | 24 Apr. | Asia |
| OSET visit to Brussels and Paris, 25–26 April | 26 Apr. | Europe |
| Advisory Body presentation to National AI Advisory Committee (United States) | 2 May | North America |
| Global Artificial Intelligence (GAIN) Assembly in Riyadh, with the Islamic World Educational, Scientific and Cultural Organization (53 countries, 4 regions) | 14 May | Middle East |
| AI in interests of sustainable development: Kazakhstan's contribution to the 2030 Agenda | 20 May | Asia |
| Group of Latin American and Caribbean States | 21 May | Latin America |
| BRICS Academic Forum | 22 May | Global |
| AI governance session in Seoul | 23 May | Asia |
| Tech Summit Asia, Singapore, 29–31 May | 29 May | Asia |
| AI for Good Global Summit, 29–31 May | 29 May | Europe |

# Annex D: List of “deep dives”

| Domain | Date (Eastern Daylight Time) |
|---|---|
| Education | 29 March |
| Intellectual property and content | 2 April |
| Children | 4 April |
| Peace and security (1) | 12 April |
| Peace and security (2) | 29 April |
| Agriculture (session 1) | 30 April |
| Agriculture (session 2) | 30 April |
| Faith-based | 1 May |
| Open-source and technology direction | 1 May |
| Impact on society | 3 May |
| Gender | 7 May |
| Data | 13 May |
| Future of work | 13 May |
| Standards (session 1) | 14 May |
| Standards (session 2) | 14 May |
| Peace and security (3) | 20 May |
| Environment | 20 May |
| Health | 22 May |
| Rule of law, human rights and democracy | 24 May |

# Annex E: Risk Global Pulse Check responses

On the request of the High-level Advisory Body on Artificial Intelligence, the Office of the Secretary-General's Envoy on Technology (OSET) conducted an AI Risk Global Pulse Check survey, as part of a horizon-scanning exercise on AI to capture perceptions on AI risks from experts from around the world. Experts were asked to respond with their views in their personal capacity (not on behalf of their institution or employer). Experts were asked to rate the degree to which they expected AI technical change and (separately) AI adoption and application to accelerate or decelerate.

They were also asked to rate their overall level of concern that harms (existing or new) resulting from AI would become substantially more serious and/or widespread, and how much that concern had recently increased or decreased. Respondents were given a list of 14 sample areas of harm (such as "Intentional malicious use of AI by non-State actors") to rate their level of concern. Finally, many text-response prompts were provided, inviting experts to comment on emerging trends, and individuals, groups and (eco)systems at particular risk from AI, and to elaborate on their rated answers.

The survey was fielded from 13 to 25 May 2024, with the invitee list constructed from OSET and the Advisory Body's networks, including participants in Advisory Body deep dives. During the fielding period, additional experts were continually invited, particularly from regions often less represented in discussions around AI, based on referrals from initial respondents and outreach to regional networks. More than 340 respondents replied to the survey, providing a rich and diverse perspective (including across regions and gender) on risks posed by AI.

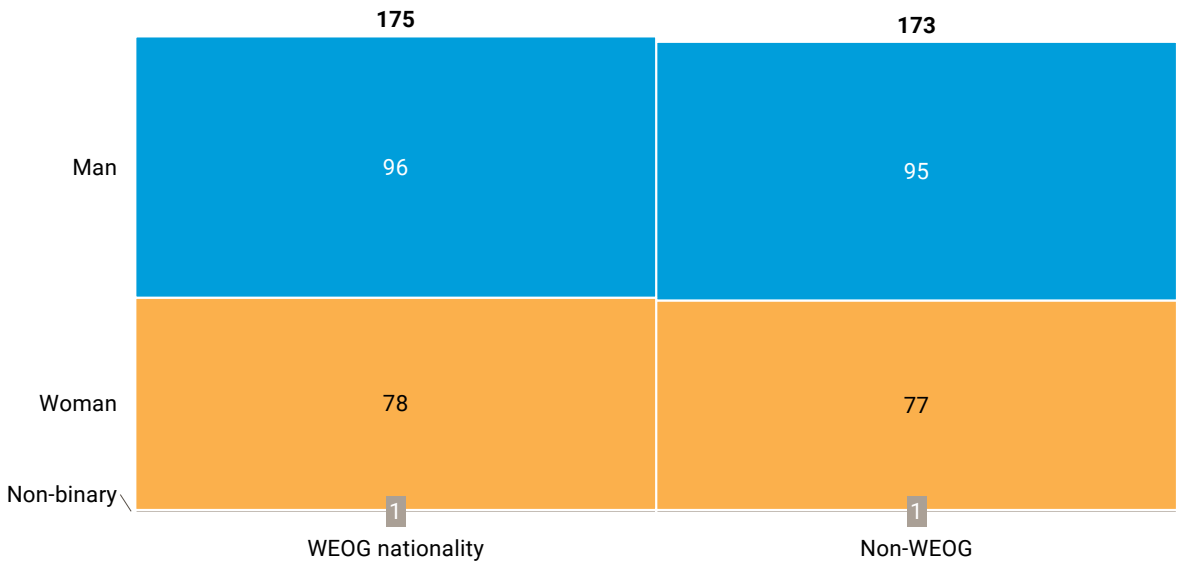
## Overview of sample

### Split by gender and region is evenly balanced

Univariate analysis by gender and region is not immediately contaminated by the other variable.

Respondents by region of nationality* (n = 348)

BY REGION & GENDER

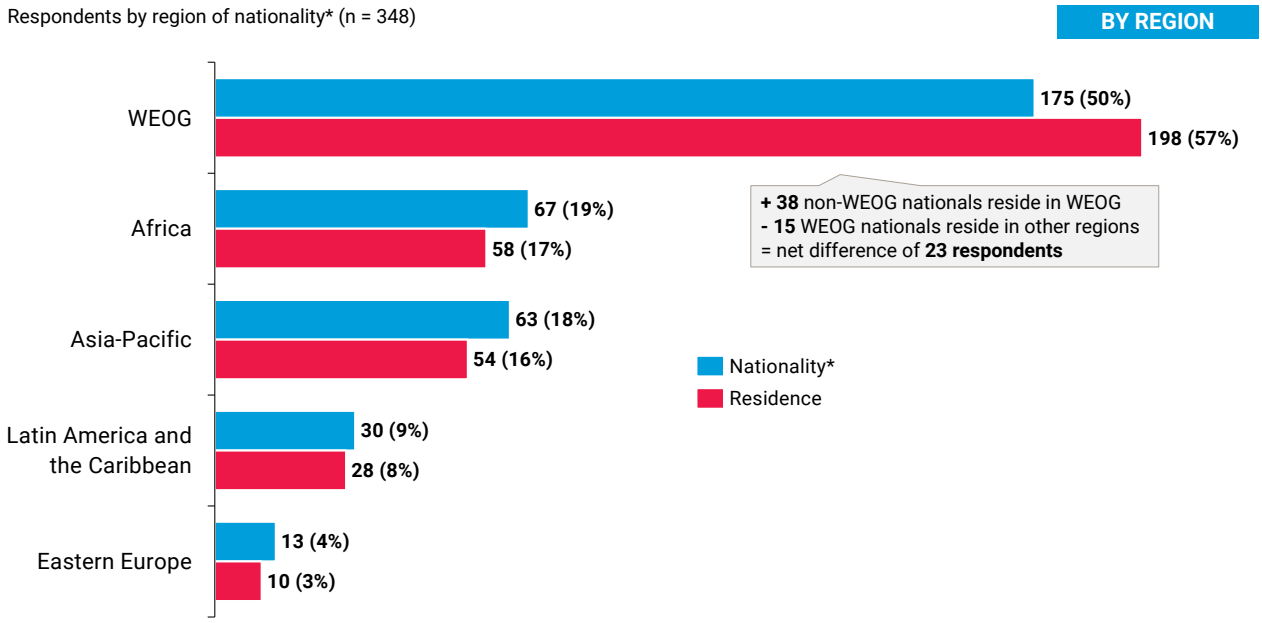



* 43 respondents (12%) indicated multiple nationalities. If respondents were resident in one of their countries of nationality, that nationality was used for analysis (34 of 43). Otherwise, the least represented nationality was used (9 of 43).
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

## Sample remains global if considered by residence

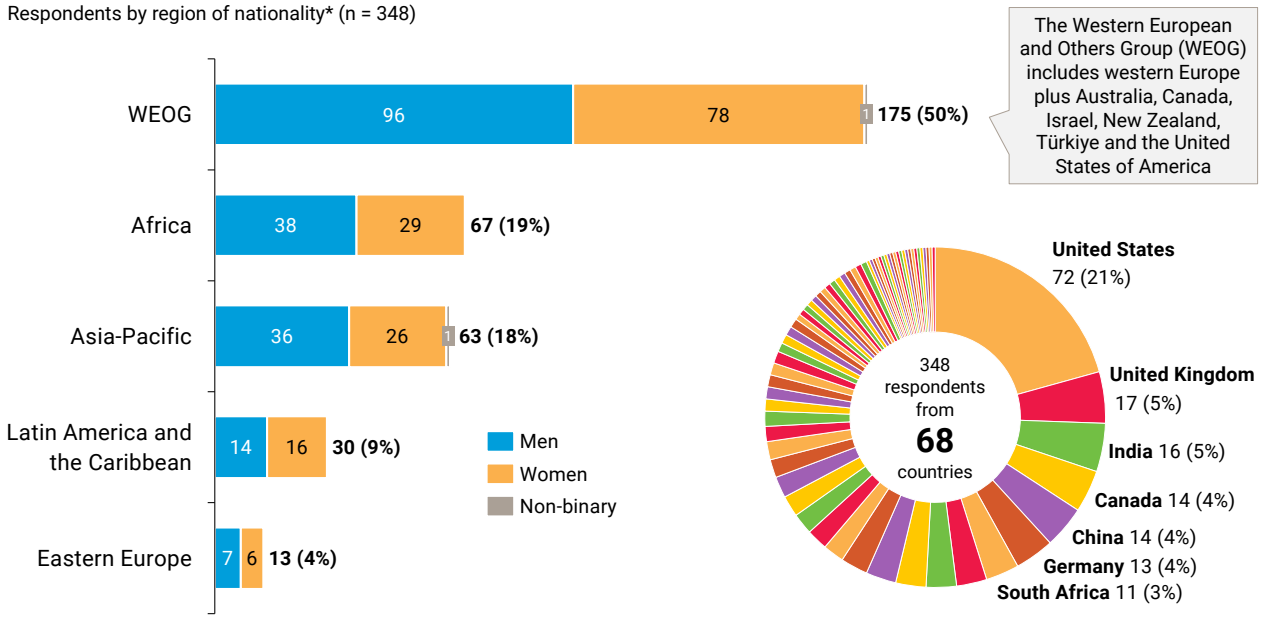
84% of respondents reside in the same region as their nationality.

Respondents by region of nationality* (n = 348)




* 43 respondents (12%) indicated multiple nationalities. If respondents were resident in one of their countries of nationality, that nationality was used for analysis (34 of 43). Otherwise, the least represented nationality was used (9 of 43).
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

Respondents by region of nationality* (n = 348)



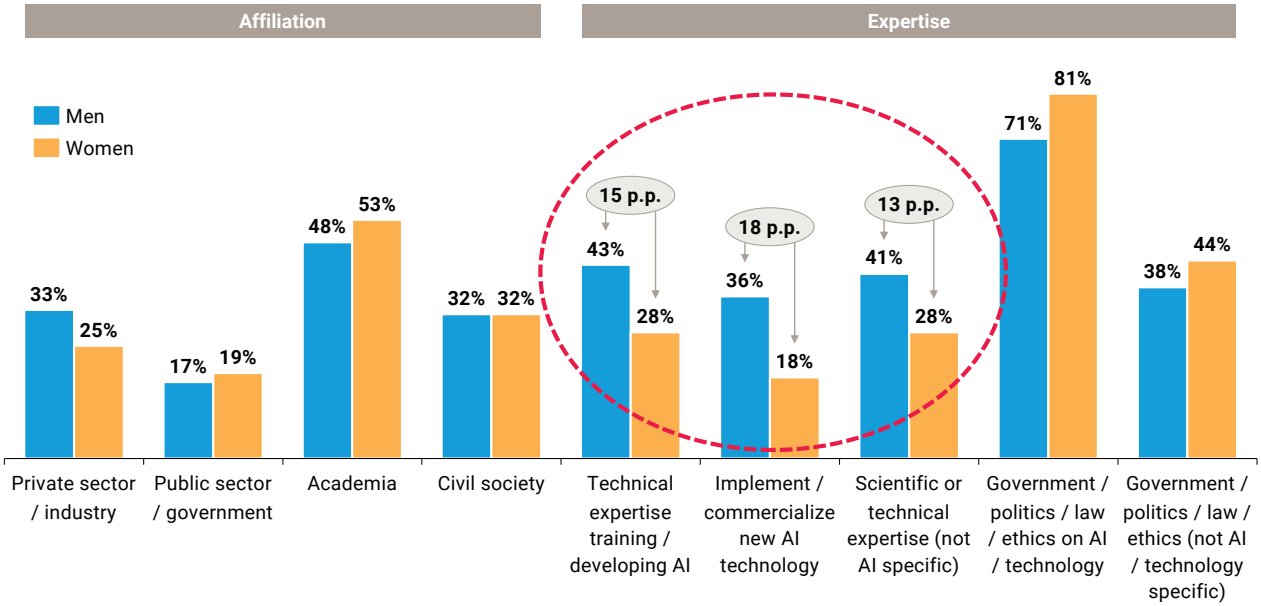

* 43 respondents (12%) indicated multiple nationalities. If respondents were resident in one of their countries of nationality, that nationality was used for analysis (34 of 43). Otherwise, the least represented nationality was used (9 of 43).
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

## Profiles of men and women respondents have some differences

More men report technical expertise; more women report governance, policy, law/ethics.

% of respondents reporting affiliation / expertise by gender (n = 348)

BY GENDER



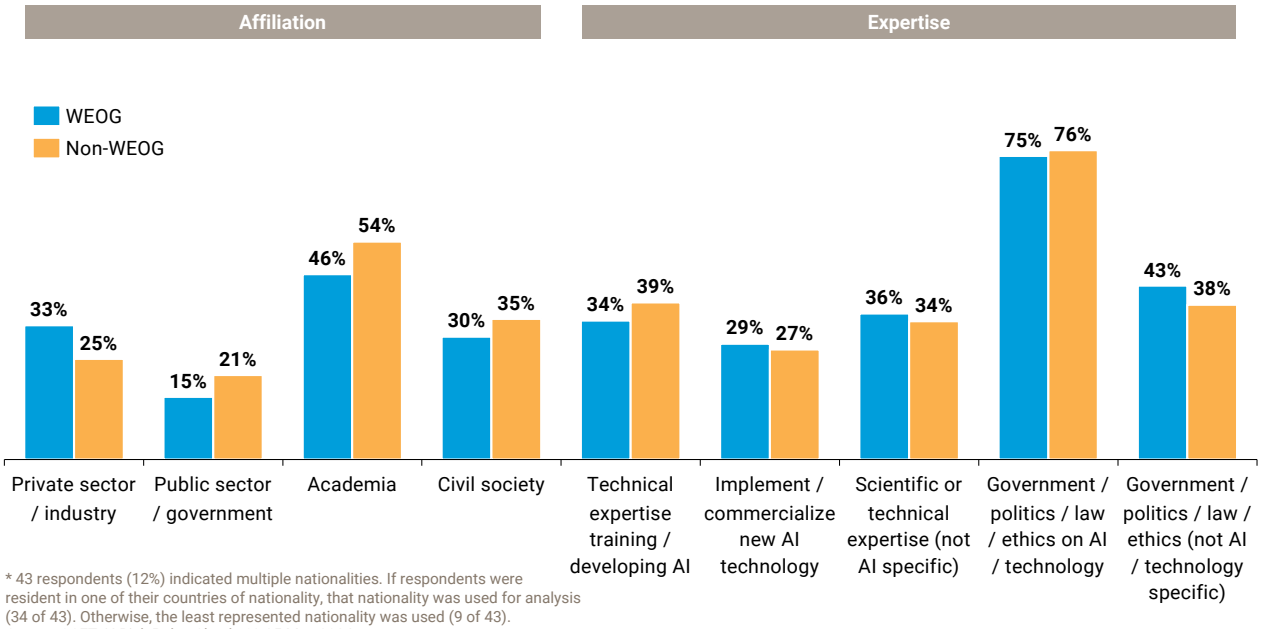

*Note:* p.p. = percentage points.
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

## Profiles of WEOG and non-WEOG respondents are reasonably similar

Non-WEOG respondents are more likely to be in the public sector or academia than in the private sector or industry.

% of respondents reporting affiliation / expertise by region of nationality* (n = 348)

BY REGION



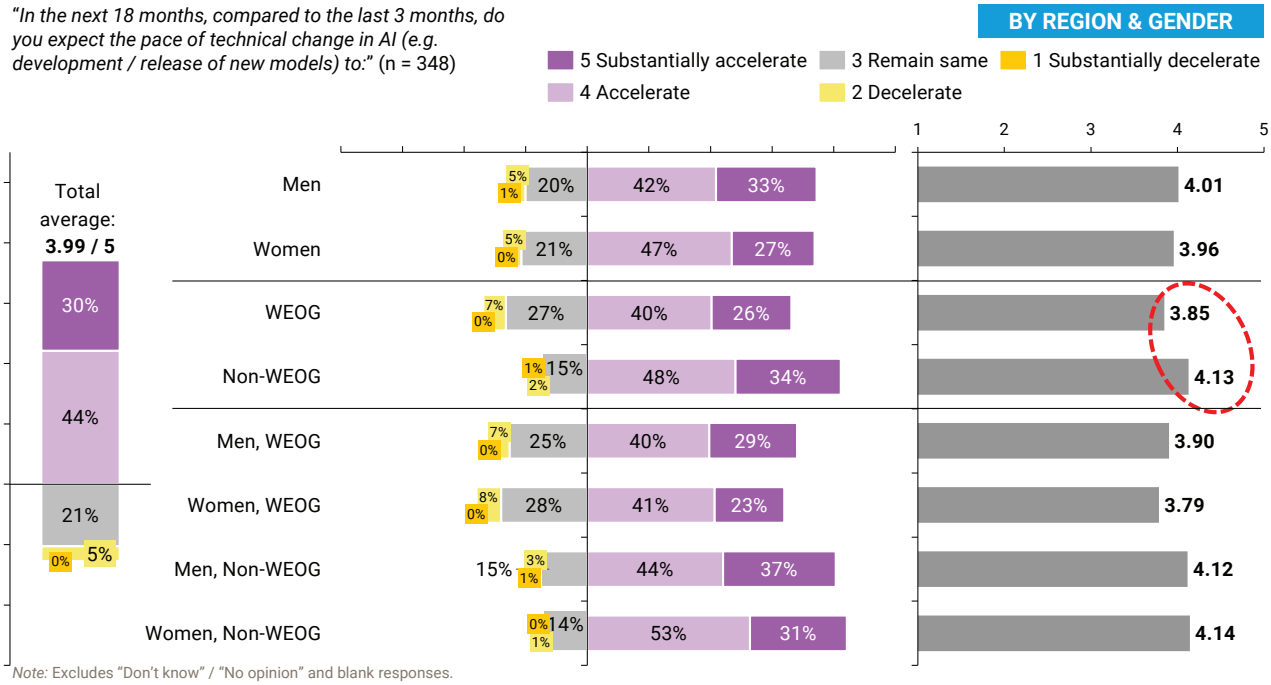

* 43 respondents (12%) indicated multiple nationalities. If respondents were resident in one of their countries of nationality, that nationality was used for analysis (34 of 43). Otherwise, the least represented nationality was used (9 of 43).
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

# Perceptions regarding acceleration of AI

## 74% of respondents expect acceleration of technical change

Higher percentage of non-WEOG respondents expect acceleration compared with WEOG respondents.

*"In the next 18 months, compared to the last 3 months, do you expect the pace of technical change in AI (e.g. development / release of new models) to:"* (n = 348)

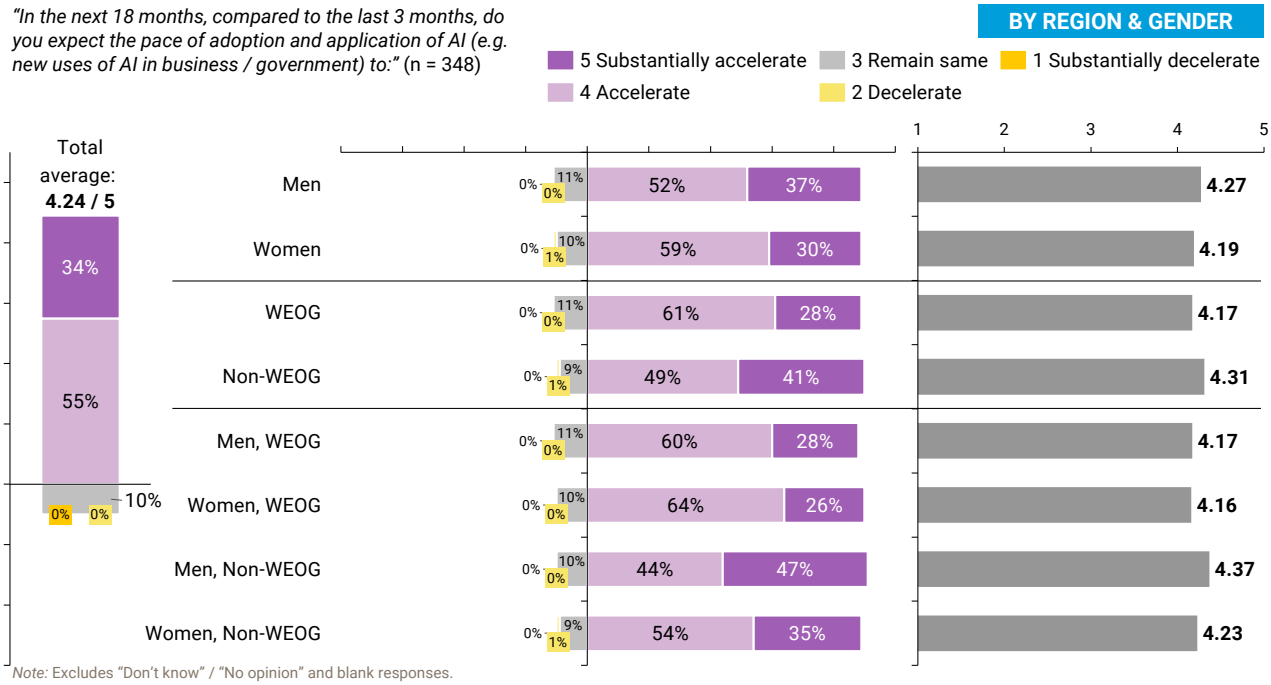



*Note:* Excludes "Don't know" / "No opinion" and blank responses.
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

## 89% of respondents expect acceleration of adoption and application

Slightly more non-WEOG respondents expect substantial acceleration (especially men).

*"In the next 18 months, compared to the last 3 months, do you expect the pace of adoption and application of AI (e.g. new uses of AI in business / government) to:"* (n = 348)




*Note:* Excludes "Don't know" / "No opinion" and blank responses.
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

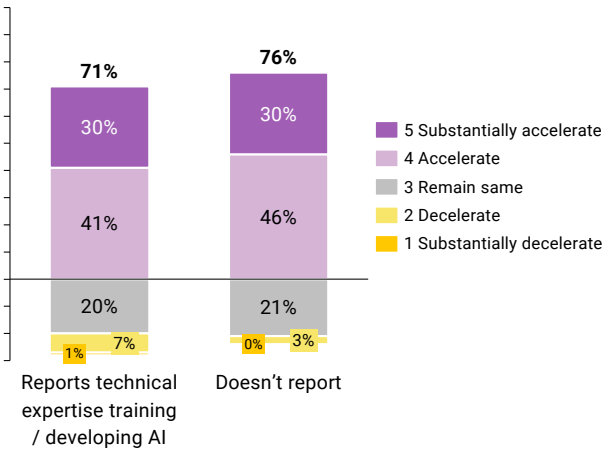
## Limited impact from technical expertise (training / developing AI)

Respondents are slightly more pessimistic on technical change, and slightly more optimistic on adoption and application.

BY EXPERTISE

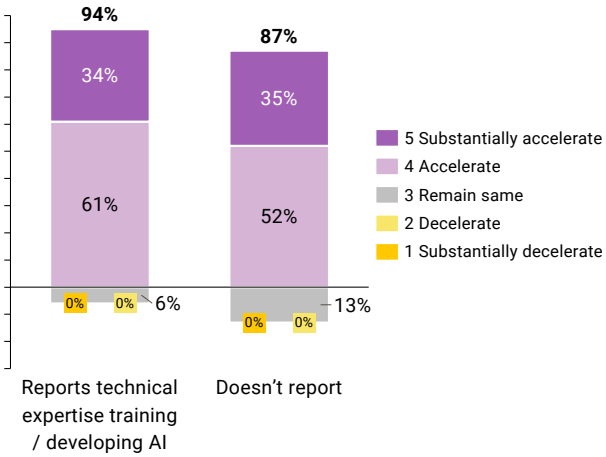
### Technical change

*"In the next 18 months, compared to the last 3 months, do you expect the pace of technical change in AI (e.g. development / release of new models) to..."* (n = 348)




### Adoption & application

*"In the next 18 months, compared to the last 3 months, do you expect the pace of adoption and application of AI (e.g. new uses of AI in business / government) to..."* (n = 348)




*Note:* Numbers may not add up to 100% owing to rounding. Excludes "Don't know" / "No opinion" and blank responses.
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

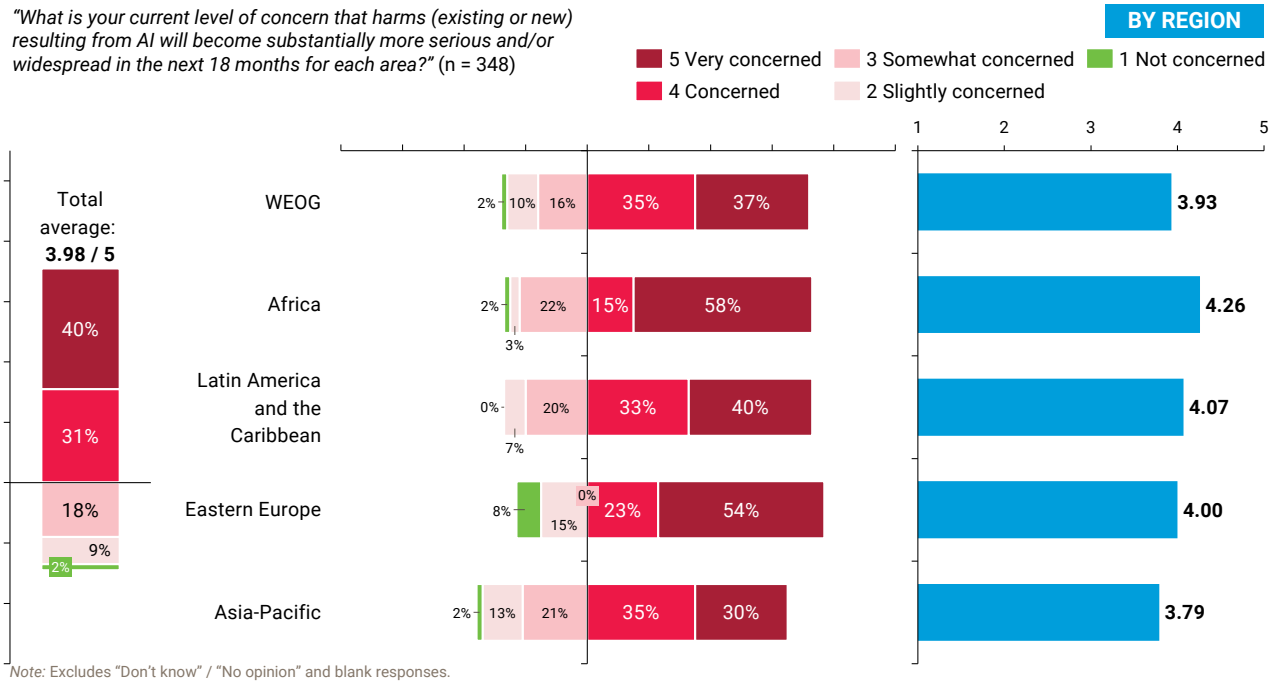
# Perceptions regarding risks of AI harms in the next 18 months (from May 2024)

## 71% concerned/very concerned about AI harms in the next 18 months

African respondents are more concerned than others; Asia-Pacific respondents are less concerned than WEOG.

*"What is your current level of concern that harms (existing or new) resulting from AI will become substantially more serious and/or widespread in the next 18 months for each area?"* (n = 348)

BY REGION

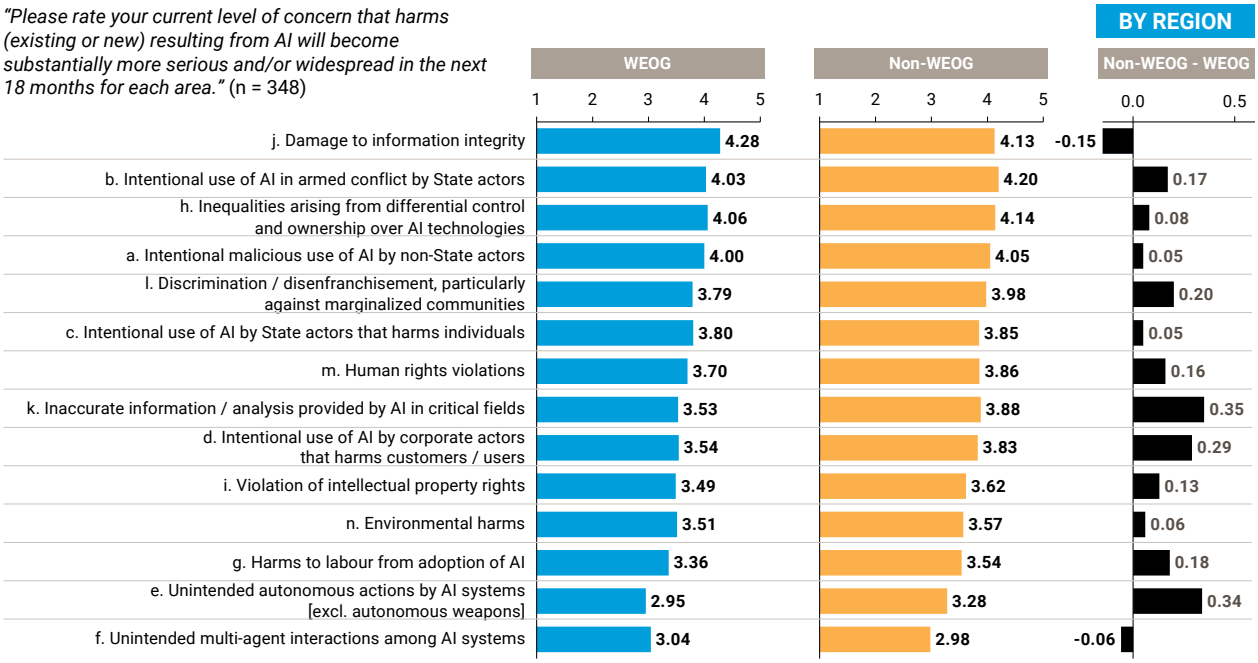



*Note:* Excludes "Don't know" / "No opinion" and blank responses.
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

## Non-WEOG more concerned than WEOG in most example areas

Particularly large gaps in inaccurate information, unintended autonomous actions and intentional corporate use.

*"Please rate your current level of concern that harms (existing or new) resulting from AI will become substantially more serious and/or widespread in the next 18 months for each area."* (n = 348)

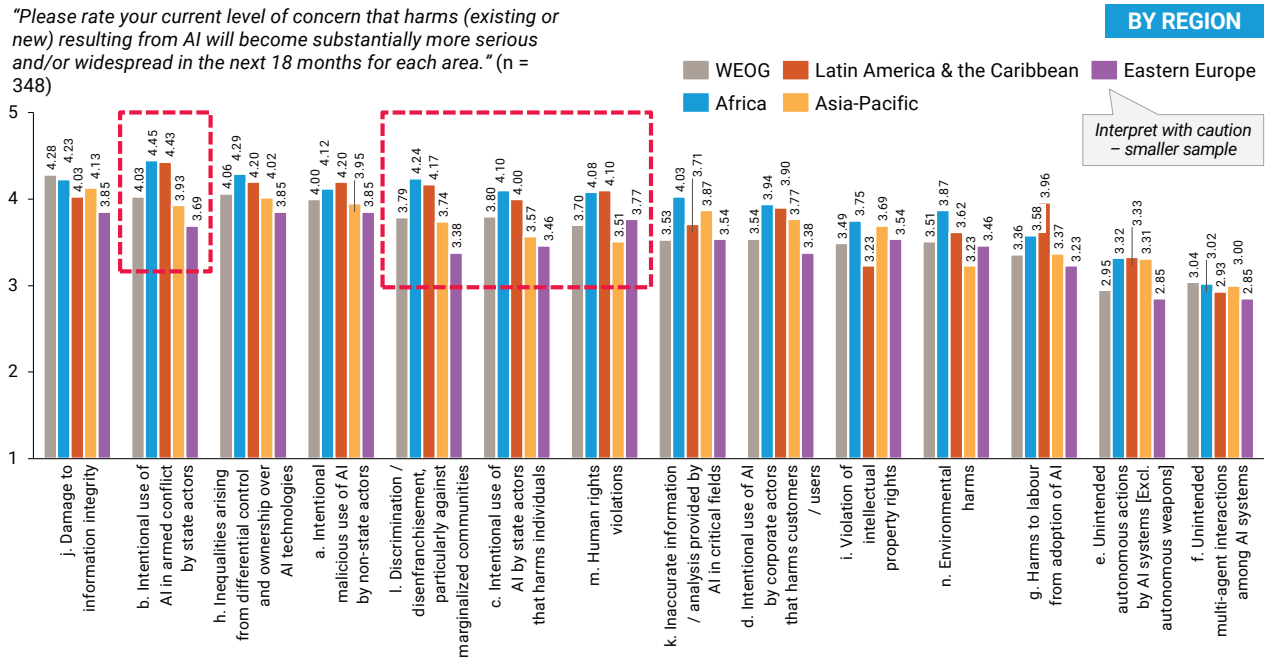



*Shown:* Average, where: 1 = Not concerned, 2 = Slightly concerned, 3 = Somewhat concerned, 4 = Concerned, 5 = Very concerned.
*Note:* Excludes "Don't know" / "No opinion" and blank responses. *Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

## Many concerns highest in Africa and in Latin America and the Caribbean

Especially around State use in armed conflict, enabling discrimination or human rights violations.

*"Please rate your current level of concern that harms (existing or new) resulting from AI will become substantially more serious and/or widespread in the next 18 months for each area."* (n = 348)

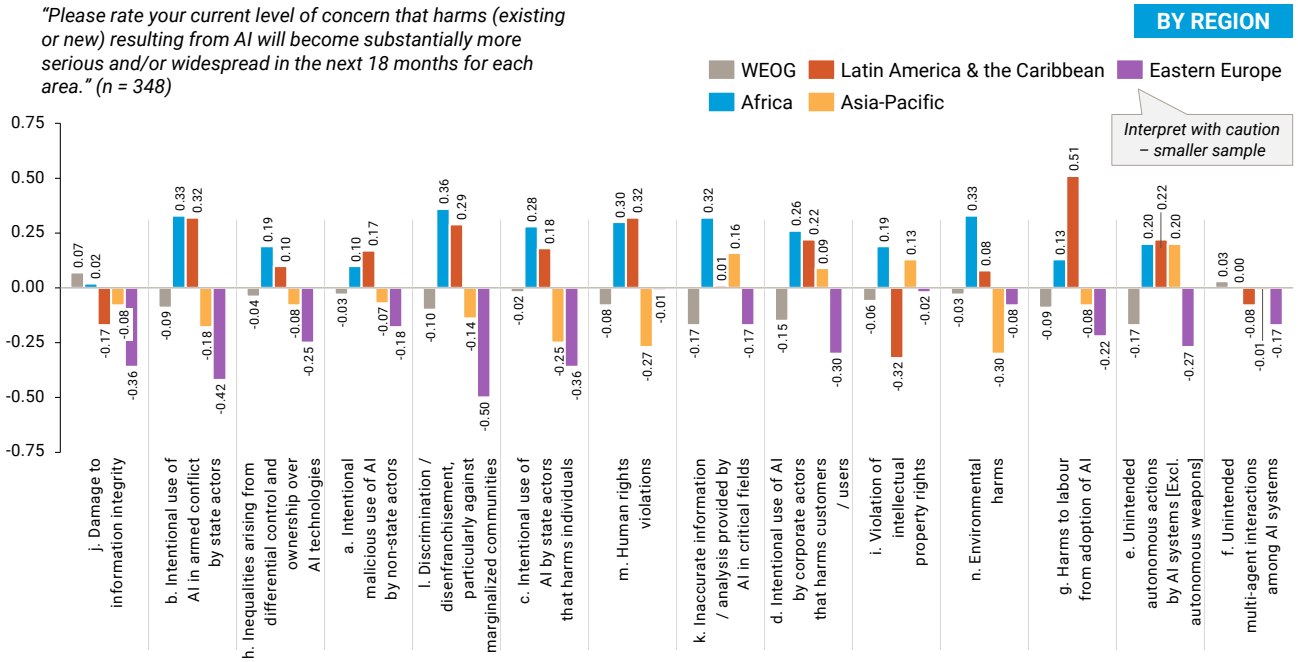



*Shown:* Average, where: 1 = Not concerned, 2 = Slightly concerned, 3 = Somewhat concerned, 4 = Concerned, 5 = Very concerned.
*Note:* Excludes "Don't know" / "No opinion" and blank responses. *Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

## Many concerns highest in Africa and in Latin America and the Caribbean

Especially around State use in armed conflict, enabling discrimination or human rights violations.

*"Please rate your current level of concern that harms (existing or new) resulting from AI will become substantially more serious and/or widespread in the next 18 months for each area." (n = 348)*

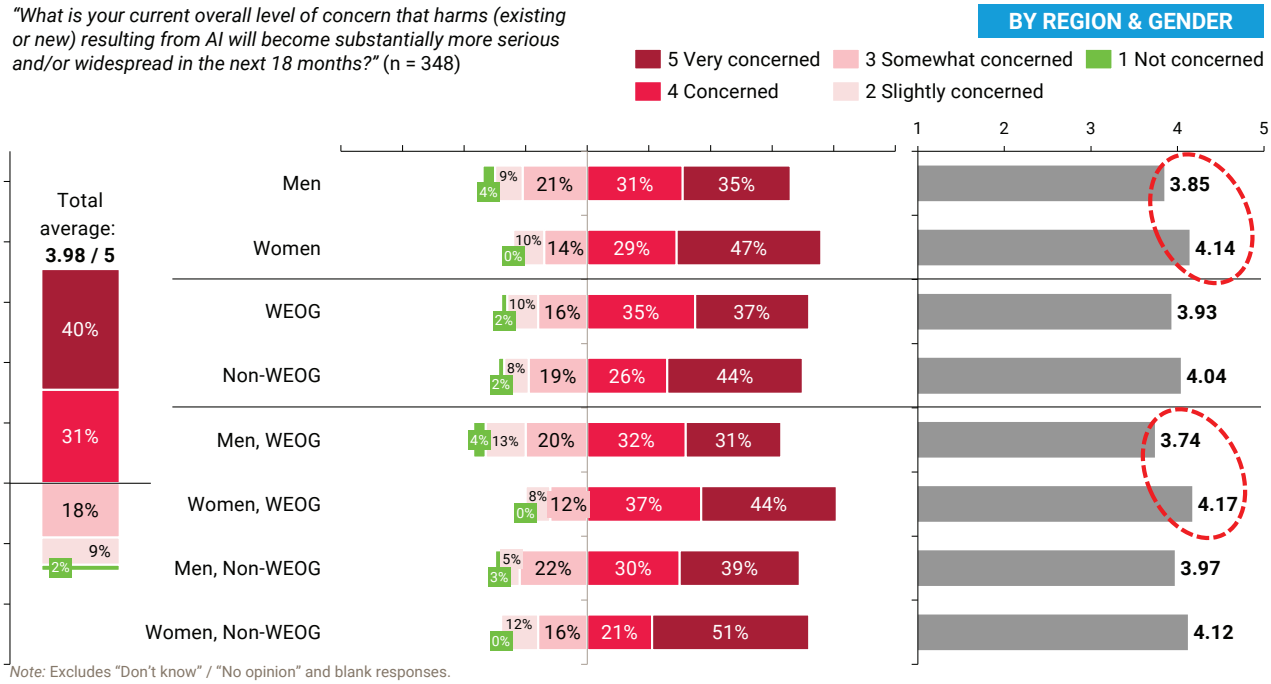



*Shown:* difference between aggregate (all regions) rating and indicated region's rating where: 1 = Not concerned, 2 = Slightly concerned, 3 = Somewhat concerned, 4 = Concerned, 5 = Very concerned. *Note:* Excludes "Don't know" / "No opinion" and blank responses. *Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

## 71% concerned / very concerned about AI harms in the next 18 months

Women more concerned than men, particularly in WEOG.

*"What is your current overall level of concern that harms (existing or new) resulting from AI will become substantially more serious and/or widespread in the next 18 months?"* (n = 348)

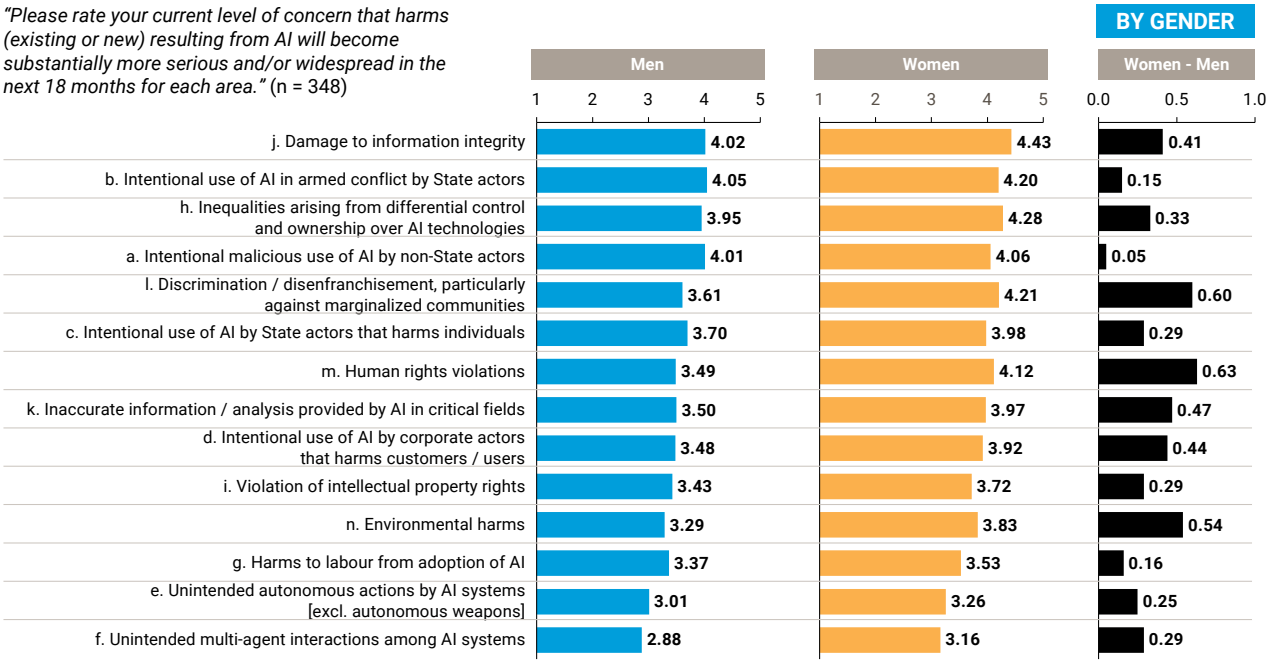



*Note:* Excludes "Don't know" / "No opinion" and blank responses.
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

## Women more concerned than men about all example areas

There are particularly large gaps on human rights violations, discrimination and the environment.

*"Please rate your current level of concern that harms (existing or new) resulting from AI will become substantially more serious and/or widespread in the next 18 months for each area."* (n = 348)

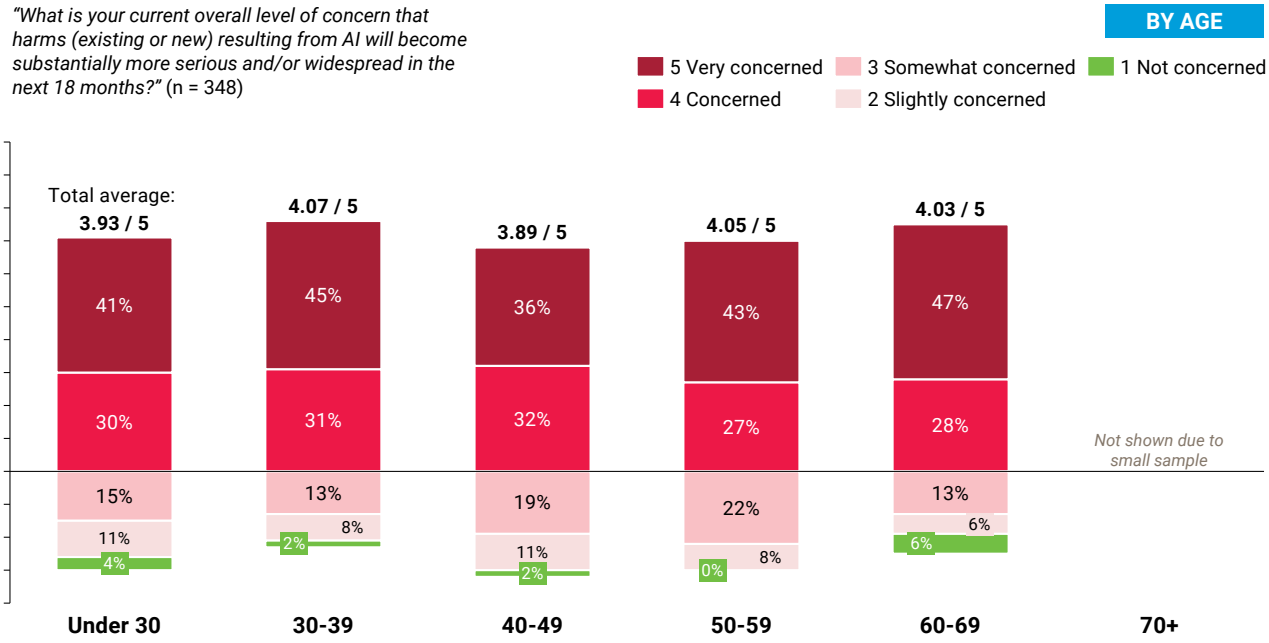



*Shown:* Average, where: 1 = Not concerned, 2 = Slightly concerned, 3 = Somewhat concerned, 4 = Concerned, 5 = Very concerned.
*Note:* Excludes "Don't know" / "No opinion" and blank responses. *Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

## 71% concerned / very concerned about AI harms in the next 18 months

Relatively small differences in concern by age of respondent.

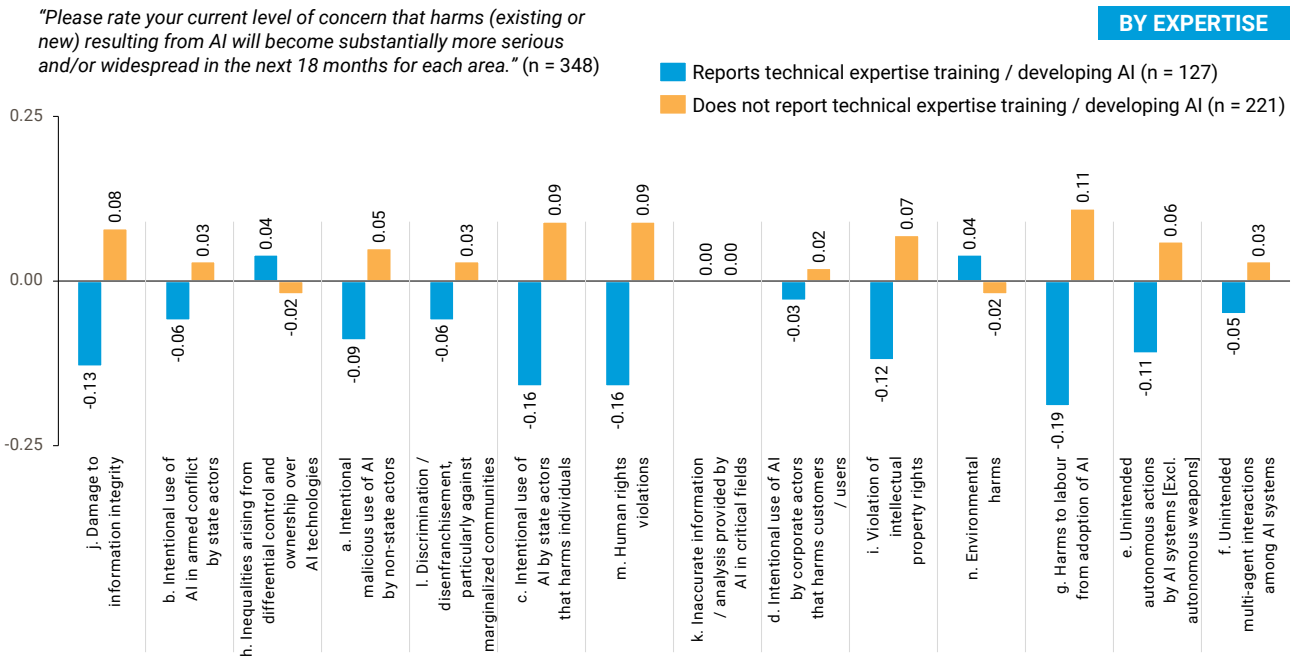
*"What is your current overall level of concern that harms (existing or new) resulting from AI will become substantially more serious and/or widespread in the next 18 months?"* (n = 348)




*Note:* Excludes "Don't know" / "No opinion" and blank responses.
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

## Respondents reporting technical expertise (training / developing AI) less concerned about most example areas

*"Please rate your current level of concern that harms (existing or new) resulting from AI will become substantially more serious and/or widespread in the next 18 months for each area."* (n = 348)

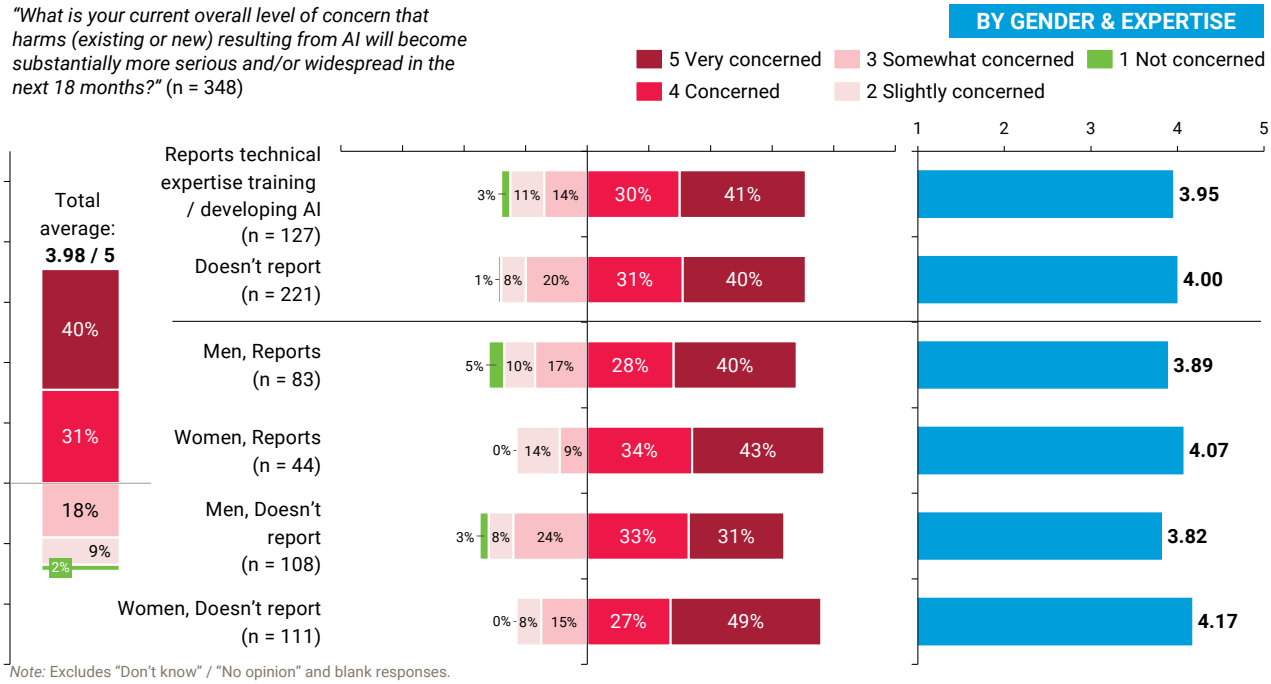



*Shown:* Difference between aggregate (all respondents) rating and indicated group's rating where: 1 = Not concerned, 2 = Slightly concerned, 3 = Somewhat concerned, 4 = Concerned, 5 = Very concerned. *Note:* Excludes "Don't know" / "No opinion" and blank responses. *Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

## Limited impact from technical expertise (training / developing AI)

Men are less concerned than women regardless of reporting status.

*"What is your current overall level of concern that harms (existing or new) resulting from AI will become substantially more serious and/or widespread in the next 18 months?"* (n = 348)

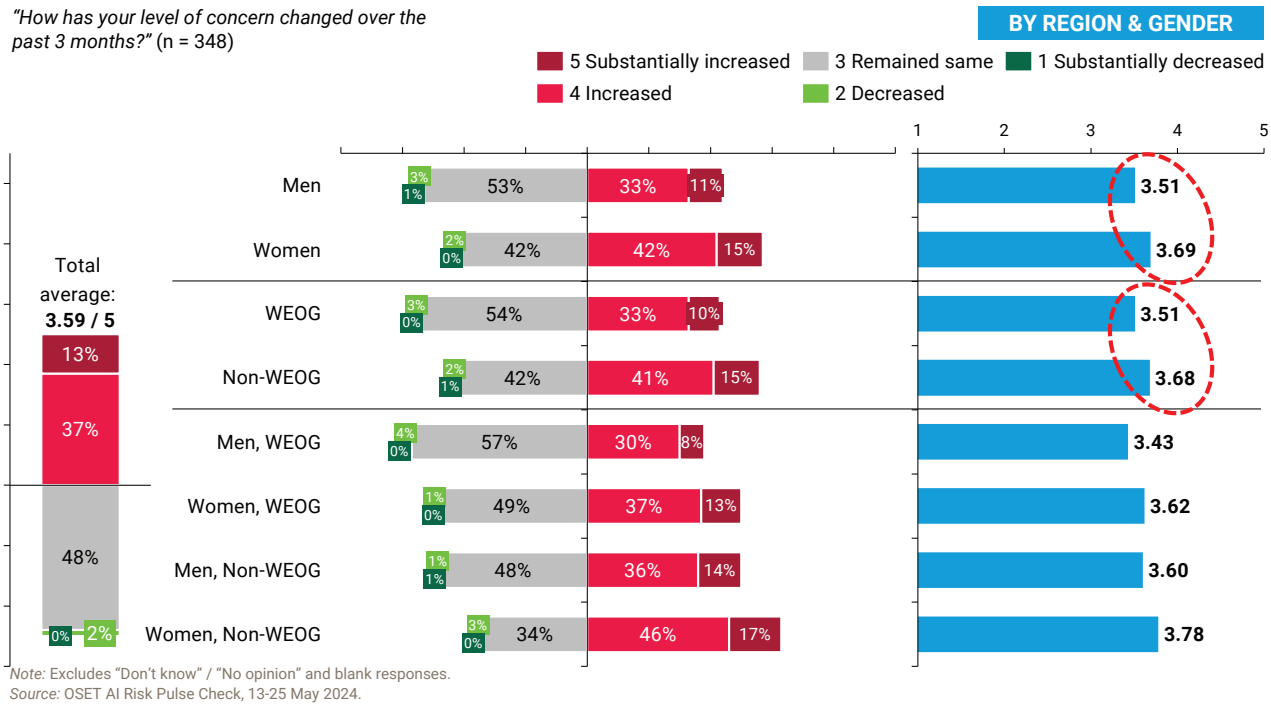



*Note:* Excludes "Don't know" / "No opinion" and blank responses.
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

# Change in perception of level of concern in the past three months regarding risks of AI harms

## 50% of the respondents increased concern in the past three months; 48% remained the same

Almost nobody decreased; more women, non-WEOG respondents have increased level of concern.




*Note:* Excludes "Don't know" / "No opinion" and blank responses.
*Source:* OSET AI Risk Pulse Check, 13-25 May 2024.

# Annex F: Opportunity scan responses

On the request of the High-level Advisory Body on Artificial Intelligence, the Office of the Secretary-General's Envoy on Technology (OSET) conducted a global AI opportunity scan survey. Experts were asked to respond with their views in their personal capacity (not on behalf of their institution or employer). The survey was divided into sections covering opportunities in high/upper-middle-income countries and lower-middle/lower-income countries, with only respondents reporting specific knowledge about lower-middle/lower-income country contexts answering those questions. The survey asked only about possible positive implications of AI.

Respondents were asked to what extent they were aware of specific examples to date of AI increasing economic activity, accelerating scientific discoveries and contributing to progress on individual SDGs.[1] They were asked to provide details including case studies, names of organizations, data and links to relevant articles/publications/papers. Respondents were then asked how much progress they expected in the next three years along the same dimensions.

As an additional view, respondents were asked by when they expected major impact from AI along those dimensions, with 50% confidence/likelihood. Additional questions were asked including which actors were involved in capturing certain opportunities, what barriers contributed to the AI divide between countries, and whether specific groups faced additional limitations harnessing opportunities from AI and how these could be addressed.

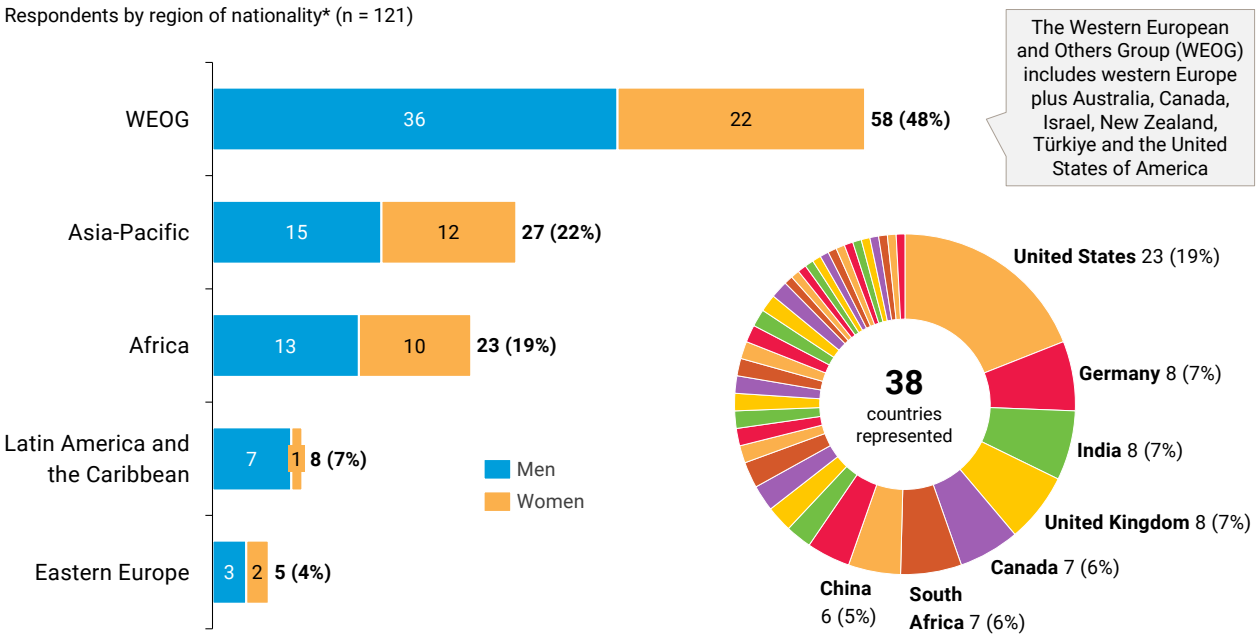
The survey was fielded from 9 to 21 August 2024, with the invitee list constructed from OSET and the Advisory Body's networks, including participants in Advisory Body deep dives. Additionally, both the International Telecommunication Union's AI for Good meeting and the networks of the United Nations Conference on Trade and Development were generously used to field the survey. Over 1,000 individuals were invited overall. More than 120 respondents replied to the survey, providing a rich and diverse perspective (including across regions and gender) on opportunities from AI.

1 SDG 8 (Decent work and economic growth) and SDG 9 (Innovation, industry and infrastructure) were not asked about separately, given their close link to increasing economic activity. SDG 17 (Partnerships for the Goals) was also not asked about specifically.

# Overview of sample

## Regional representation: strong global participation

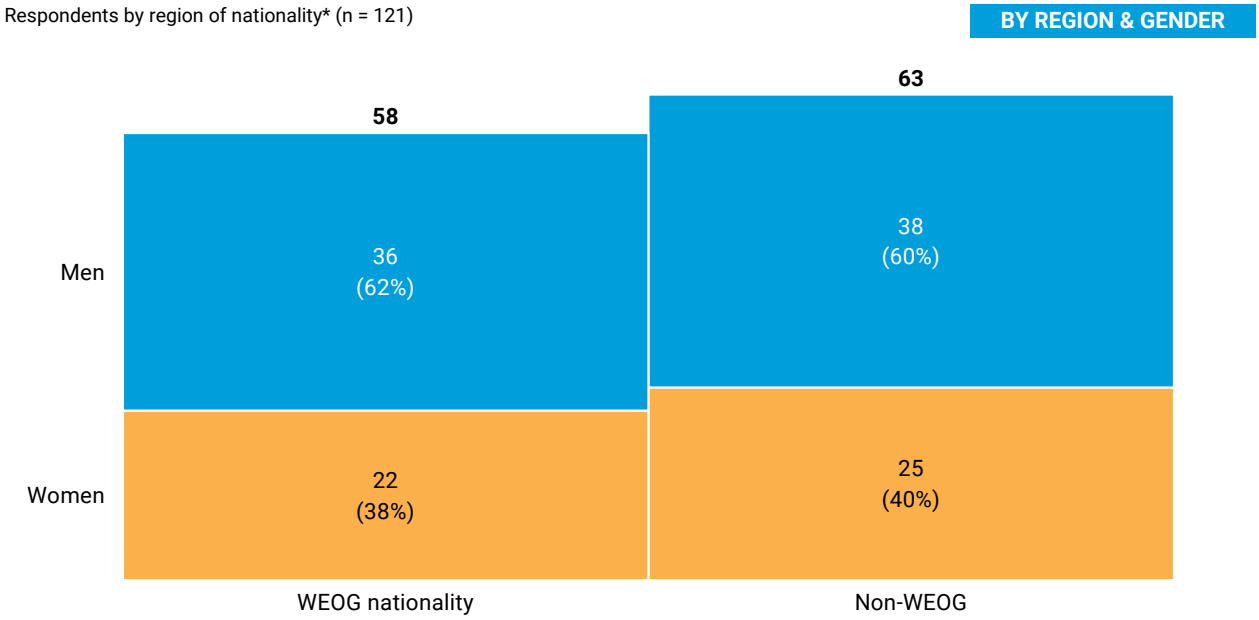
Allows comparison of responses between Western European and Others Group (WEOG) and other regions.




* 9 respondents (7%) indicated multiple nationalities. If respondents were resident in one of their countries of nationality, that nationality was used for analysis (8 of 9). Otherwise, the least represented nationality was used (1 of 9).
*Source:* OSET AI Opportunity Scan survey, 9-21 August 2024.

## Men are ~60% of both WEOG, non-WEOG samples

Consistency means univariate analysis by gender, region is not immediately contaminated.

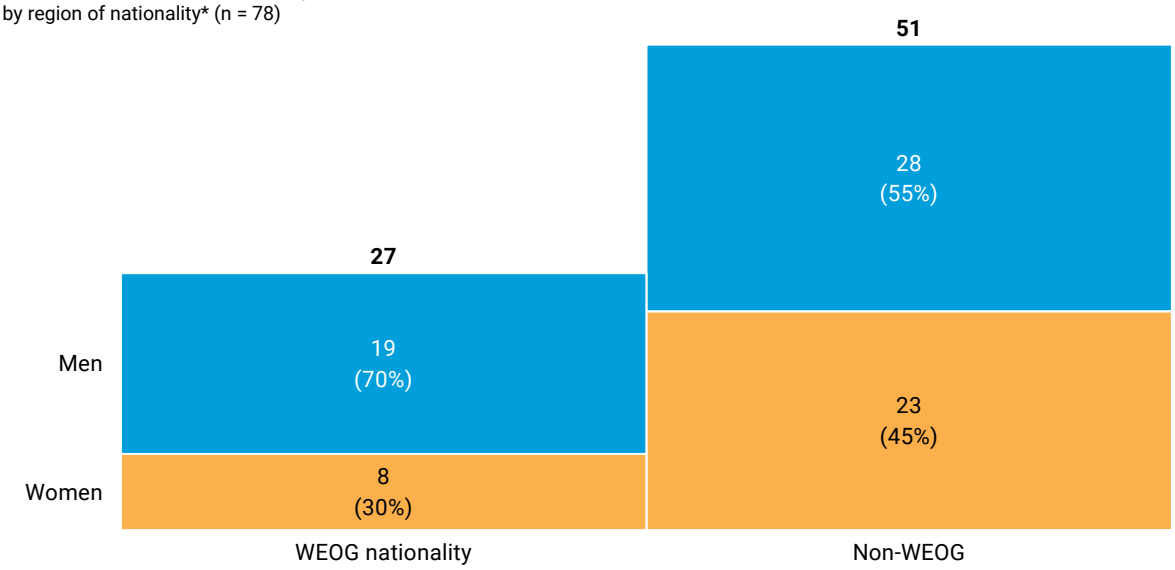



* 9 respondents (7%) indicated multiple nationalities. If respondents were resident in one of their countries of nationality, that nationality was used for analysis (8 of 9). Otherwise, the least represented nationality was used (1 of 9).
*Source:* OSET AI Opportunity Scan survey, 9-21 August 2024.

## Developing-country-knowledgeable sample less balanced

Respondents reporting specific knowledge about lower-middle/lower-income-country-contexts, by region of nationality* (n = 78)

**BY REGION & GENDER**



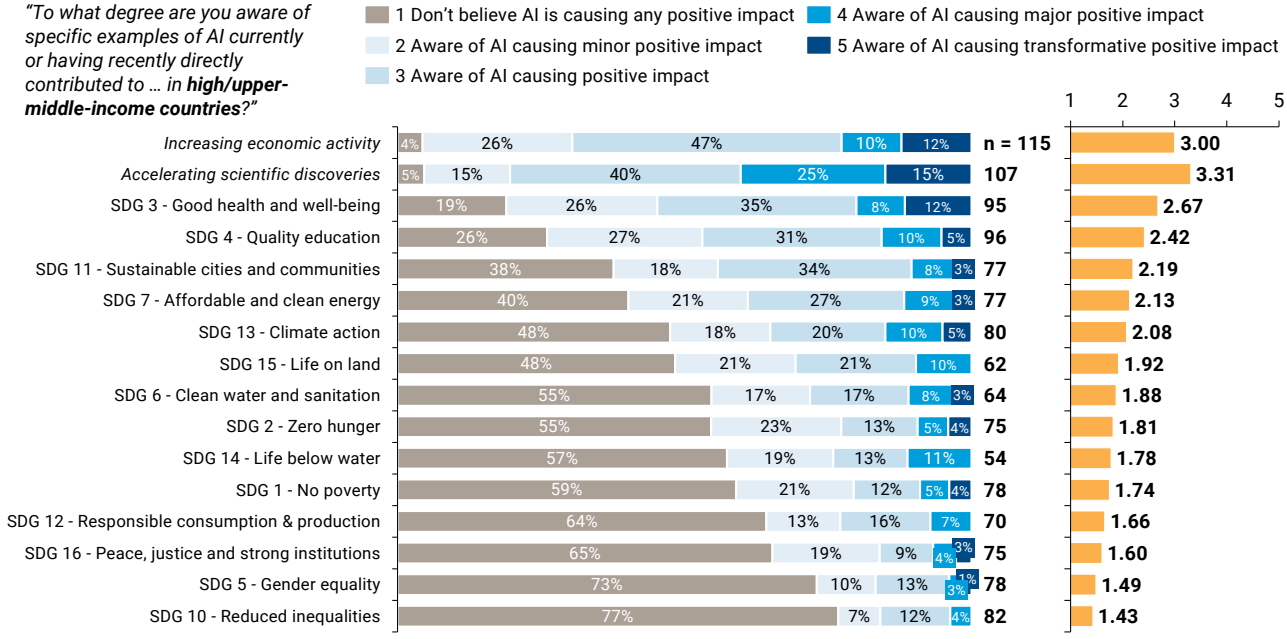

* 9 respondents (7%) indicated multiple nationalities. If respondents were resident in one of their countries of nationality, that nationality was used for analysis (8 of 9). Otherwise, the least represented nationality was used (1 of 9). Only respondents reporting relevant knowledge were asked about lower-middle/lower-income countries.
*Source:* OSET AI Opportunity Scan survey, 9-21 August 2024.

# Perceptions regarding positive impact of AI to date

## Positive impact to date on growth and science, but less on most SDGs

Impact to date in high/upper-middle-income countries.

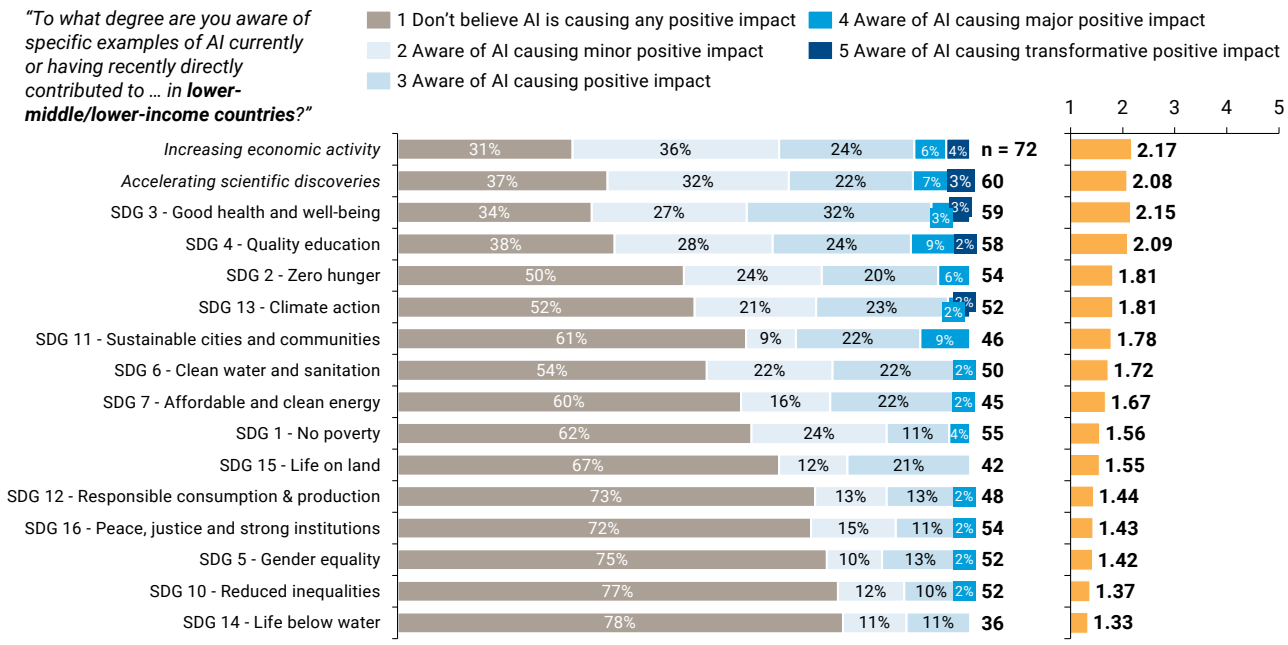
*"To what degree are you aware of specific examples of AI currently or having recently directly contributed to … in **high/upper-middle-income countries**?"*




*Note:* Excludes "Don't know" / "No opinion" and blank responses. Did not ask about SDGs 8, 9 and 17.
*Source:* OSET AI Opportunity Scan survey, 9-21 August 2024.

## Less impact reported in the lower-income world on all fronts

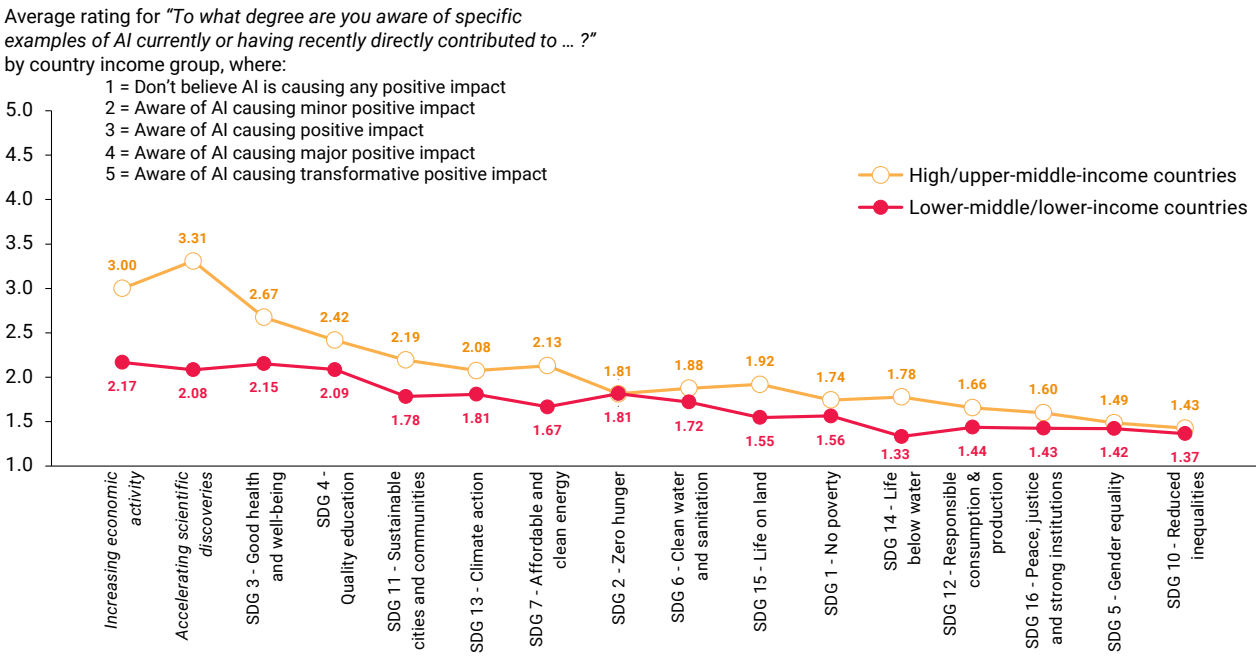
Impact to date in lower-middle/lower-income countries.




*Note:* Excludes "Don't know" / "No opinion" and blank responses. Only respondents reporting relevant knowledge were asked about lower-middle/lower-income countries. Did not ask about SDGs 8, 9 and 17.
*Source:* OSET AI Opportunity Scan survey, 9-21 August 2024.

## Less impact reported in the lower-income world on all fronts

Gap most pronounced on economic growth and science.



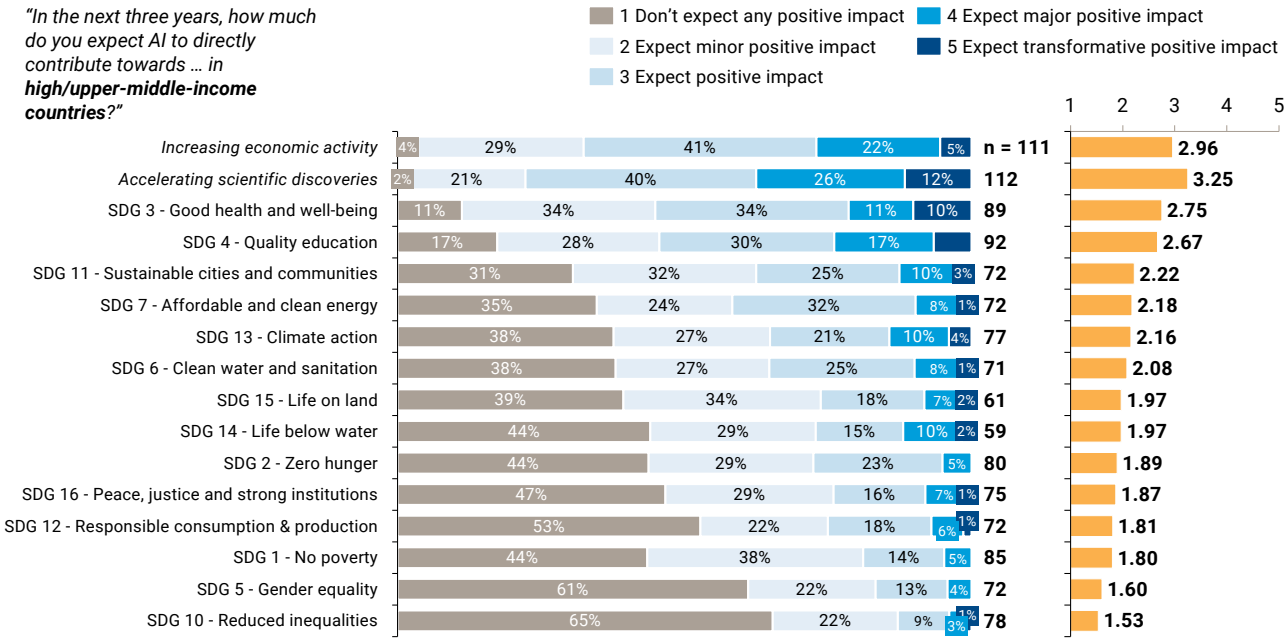

*Note:* Excludes "Don't know" / "No opinion" and blank responses. Only respondents reporting relevant knowledge were asked about lower-middle/lower-income countries. Did not ask about SDGs 8, 9 and 17. *Source:* OSET AI Opportunity Scan survey, 9-21 August 2024.

# Perceptions regarding expected positive impact of AI in the next three years

## Expected impact on growth, science, health, education – less on others

Impact expected in the next three years in high/upper-middle-income countries

*"In the next three years, how much do you expect AI to directly contribute towards … in **high/upper-middle-income countries**?"*

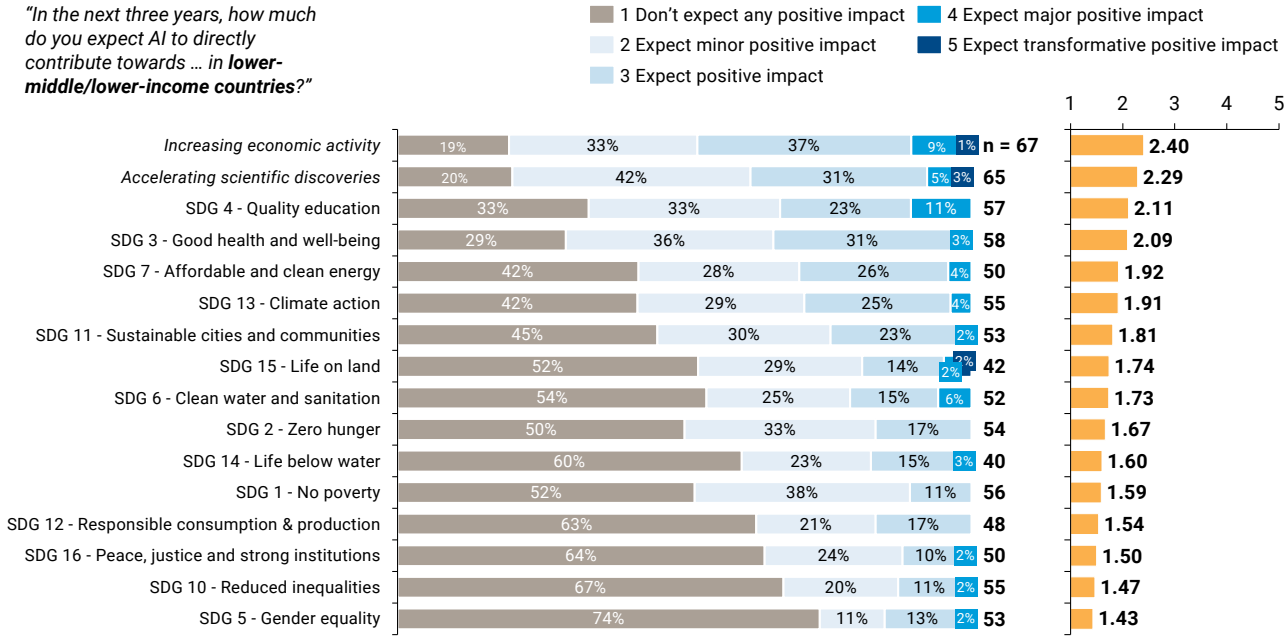



*Note:* Excludes "Don't know" / "No opinion" and blank responses. Did not ask about SDGs 8, 9 and 17.
*Source:* OSET AI Opportunity Scan survey, 9-21 August 2024

## Some expected impact in lower-income world, but again more limited

Impact expected in the next three years in lower-middle/lower-income countries.

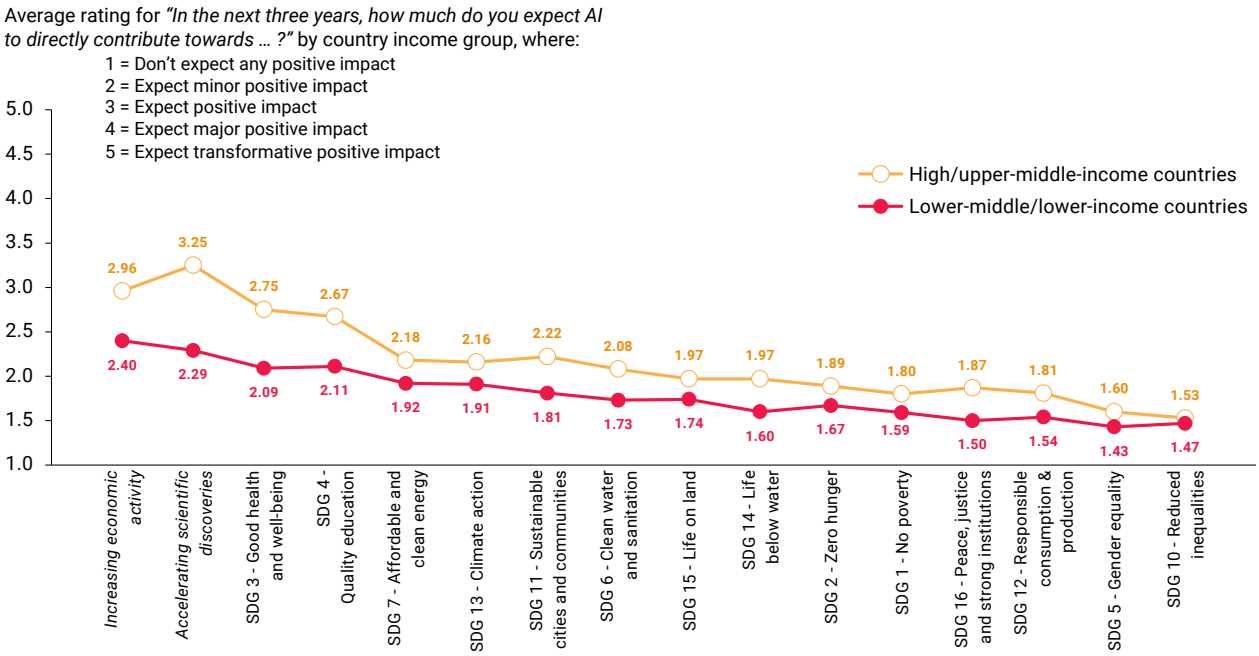
*"In the next three years, how much do you expect AI to directly contribute towards … in **lower-middle/lower-income countries**?"*




*Note:* Excludes "Don't know" / "No opinion" and blank responses. Only respondents reporting relevant knowledge were asked about lower-middle/lower-income countries. Did not ask about SDGs 8, 9 and 17.
*Source:* OSET AI Opportunity Scan survey, 9-21 August 2024.

## Less impact expected in the lower-income world on all fronts

Gap most pronounced on economic growth, science, health and education.




*Note:* Excludes "Don't know" / "No opinion" and blank responses. Only respondents reporting relevant knowledge were asked about lower-middle/lower-income countries. Did not ask about SDGs 8, 9 and 17. *Source:* OSET AI Opportunity Scan survey, 9-21 August 2024.
Charts prepared with think-cell

# Annex G: List of abbreviations

| | |
|---|---|
| **ACM** | Association for Computing Machinery |
| **AG** | African Group |
| **AI** | artificial intelligence |
| **ANSI** | American National Standards Institute |
| **APG** | Asia and the Pacific Group |
| **ASEAN** | Association of Southeast Asian Nations |
| **BSI** | British Standards Institution |
| **CEN** | European Committee for Standardisation |
| **CENELEC** | European Committee for Electrotechnical Standardization |
| **CERN** | European Organization for Nuclear Research |
| **EEG** | Eastern European Group |
| **ETSI** | European Telecommunications Standards Institute |
| **FAO** | Food and Agriculture Organization of the United Nations |
| **FATF** | Financial Action Task Force |
| **FIMI** | foreign information manipulation and interference |
| **FMF** | Frontier Model Forum |
| **FSB** | Financial Stability Board |
| **G20** | Group of 20 |
| **G7** | Group of Seven |
| **GPU** | graphics processing unit |
| **IAEA** | International Atomic Energy Agency |
| **ICAO** | International Civil Aviation Organization |
| **IEC** | International Electrotechnical Commission |
| **IEEE** | Institute of Electrical and Electronics Engineers |
| **ILO** | International Labour Organization |
| **IMO** | International Maritime Organization |
| **IPCC** | Intergovernmental Panel on Climate Change |
| **ISO** | International Organization for Standardization |
| **ITU** | International Telecommunication Union |
| **LAC** | Latin America and the Caribbean |
| **NIST** | National Institute of Standards and Technology (United States) |
| **OECD** | Organisation for Economic Co-operation and Development |
| **OHCHR** | Office of the United Nations High Commissioner for Human Rights |
| **OSET** | Office of the Secretary-General's Envoy on Technology |
| **SAC** | Standardization Administration of China |
| **SDG** | Sustainable Development Goal |
| **UNCTAD** | United Nations Conference on Trade and Development |
| **UNDP** | United Nations Development Programme |
| **UNESCO** | United Nations Educational, Scientific and Cultural Organization |
| **UNHCR** | United Nations High Commissioner for Refugees |
| **UNOCT** | Office of Counter-Terrorism |
| **WEOG** | Western European and Others Group |
| **WHO** | World Health Organization |
| **WIPO** | World Intellectual Property Organization |
| **WSC** | World Standards Cooperation |

**Donors**

The Body gratefully acknowledges the financial and in-kind contributions of the following governments and partners, without whom it would not have been able to carry out its responsibilities:

Government of the Czech Republic
European Union
Government of Finland
Government of Germany
Government of Italy
Government of Japan
Government of the Kingdom of the Netherlands
Government of the Kingdom of Saudi Arabia
Government of Singapore
Government of Switzerland
Government of the United Arab Emirates
Government of the United Kingdom of Great Britain and Northern Ireland
Omidyar Network Fund
L'Organisation internationale de la Francophonie

**Secretariat**

Simon Chesterman
Quintin Chou-Lambert
Eleonore Fournier-Tombs
Sebastian Frank
David Michael Kelly
Brian Shung Seun Lau
Andrew Morritt
Filippo Pierozzi
Mehdi Snène
Isabel de Sola
Lucia Velasco
Rebekah Hayoung Woo

United Nations

AI Advisory Body