



CEB
Chief Executives Board
for Coordination

High-Level Committee on Programmes (HLCP)

Inter-Agency Working Group on Artificial Intelligence (IAWG-AI)

**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**

* HLCP approved the UN system white paper on AI governance on 16 April 2024, following a system-wide consultation process.

Table of contents

Executive summary	3
A. Introduction	7
1. Risks, challenges, and considerations for global AI governance	7
(a) The AI value chain.....	7
(b) The AI life-cycle.....	8
(c) The broader macro risks and challenges associated with AI	8
(d) The diverse levels and avenues of impact that AI has on different sectors and aspects of society	9
(e) Governance must go hand-in-hand with sustainable development opportunities	9
2. Context for the paper	9
3. Overview of the paper	10
(a) Focus areas of the paper	10
(b) Methodology.....	11
B. Focus area I – Overview of existing normative and policy instruments in the UN system to inform and shape AI governance.....	12
1. Overview of instruments	12
(a) Specific instruments on AI governance	13
(b) Instruments on areas inter-related with AI governance	14
(c) International legal frameworks that could be applied to aspects of AI governance.....	15
(d) International instruments for governance of global public goods and sector-specific approaches	16
2. Key takeaways.....	16
C. Focus area II - Institutional functions that inform global AI governance.....	21
1. Scientific consensus through technical and authoritative assessments and research	22
2. Consensus-building and norm-setting around risks and opportunities	23
3. Regulatory coordination, monitoring, and enforcement	24
4. Addressing the developmental needs of AI, including capacity building, technological equity, and technology diffusion & dissemination	26
5. Coordination across the UN system for cross-cutting issue areas	28
D. Focus Area III - Learnings from existing governance structures, inclusive normative processes and agile & anticipatory approaches	29
1. Learnings from existing governance structures and normative processes	29
2. Learnings from agile and anticipatory governance mechanisms within the UN system	31
(a) Agile and reflexive governance	31
(b) Anticipatory governance and foresight-based research	32
E. Recommendations	34
Annex 1: AI Risks	37
Annex 2: Methodology details.....	39
Annex 3: Highlights from the survey.....	40
Annex 4: Instruments from the UN system on AI or inter-related with AI or those that could apply to AI.....	41

Executive summary

Artificial Intelligence (AI) governance has gained significant importance in recent years, especially since late 2022, due to the rapid development and expansion of generative AI and large language models, coupled with numerous global leaders voicing concerns about AI's existential risks. There have been calls to ensure the effectiveness and coherence of the multiple international AI governance efforts and to understand them in the broader context of governance of digital technologies, factoring in the existing digital divide, including the gender digital divide.

This paper¹ outlines the UN System's work on AI governance, focusing on current institutional models and related functions, and existing international normative frameworks in the UN system that could be applied or leveraged for international AI governance. This paper is intended to inform the United Nations (UN) System deliberations on AI and is also a UN System-wide contribution to the Secretary-General's High-level Advisory Body on AI (HLAB-AI).

The data and findings generated during the research present a high-level overview of the UN System's extensive resources, expertise, and experience with normative governance processes and structures related to AI governance. It demonstrates that the UN System has been proactive in addressing the many challenges derived from the rapid development of Artificial Intelligence and has responded with diverse mechanisms to support Member States. In particular, the UN System plays a unique role as a convener for scientific and political consensus building and a platform for norm-setting, engaging with governments, the private sector, academia, and civil society, and helping Member States develop technical and policy capacities. Additionally, the UN System champions an ethical, human-centred, and human-rights-based approach to bridging the digital divide.

Further deliberations are required on the critical efforts/expertise needed to fully address emerging AI challenges and ensure that AI governance approaches balance economic, social, and environmental goals. It is also essential for the System to strengthen implementation and coordination mechanisms, enhance working modalities and resource mobilization efforts, leverage its expertise, build policy/programmatic coherence, and streamline communications and outreach. A Task Force under the High-level Committee on Management (HLCM) has been progressing to develop an internal normative guidance/model policy for the UN system on the use of AI and to identify and promote mechanisms for pooling technical capacity and knowledge sharing on AI.

The key takeaways from the three focus areas of the paper are presented below, followed by recommendations for the UN System.

Focus Area I – Existing normative and policy instruments in the UN System to inform and shape AI governance

Key observation: Several instruments were identified by the IAWG-AI members as illustrative examples of existing multi-layered and multi-faceted instruments that provide a strong foundation for normative efforts. Reflecting the specific and intersectional mandates of UN System entities, the instruments vary in scope from cross-sectoral/overarching efforts to more sector-specific ones, providing an array of governance levers that cater to varying maturity levels of AI governance efforts and capacities, including at the national, regional and international levels. The examples cover comprehensive instruments that already include implementation mechanisms, as well as more targeted instruments. As a starting point, AI governance efforts must be anchored in international law, including international human rights law.

- The UN System's technical instruments, such as international standards, play an effective role in facilitating norm-setting and interoperability to address the opportunities and risks of AI.

¹ This paper been developed by the Inter-Agency Working Group on Artificial Intelligence (IAWG-AI), co-led by ITU and UNESCO, comprising of over 40 UN System entities. This paper was developed in response to a request from the Joint Session of the High-level Committee on Programmes (HLCP) and the High-level Committee on Management (HLCM) for the development of a white paper on current institutional models and related functions and existing international normative frameworks in the UN system that could be applied or leveraged for international AI governance, for consideration by HLCP at its 47th session. CEB subsequently welcomed this outcome.

- Including key stakeholders from the beginning of the process, including relying on their support in piloting the frameworks under development, brings legitimacy, helps demonstrate early results, and improves the adoption rate.
- In the discussions on compliance, monitoring, and enforcement for AI systems, it is essential to consider mechanisms such as audits, inspections, and certifications and to learn from the experiences of the UN system in addressing other complex global challenges.
- The implementation of normative instruments governing global public goods provides important lessons for transparency, accountability, and redress mechanisms, which are essential for AI governance efforts.
- Tracking tools and data observatories can help with global alignment and sharing of information and best practices.
- Capacity development for policy and AI practitioners is key to supporting the implementation of relevant instruments.

Focus Area II – Institutional functions that inform global AI governance

Key observation: There is no one-size-fits-all approach in what concerns the institutional functions for AI governance, including scientific research and consensus building, norm and standard setting, enforcement, and monitoring, as well as development and diffusion of AI technology across all the diverse areas of application of AI. A continuation of a networked approach would benefit the UN System, leveraging inter-agency coordination and cooperation while enhancing capacities to strengthen its work and deliver effectively on global AI governance. Institutions have developed different ways of addressing this and benefit from cooperation.

- Scientific assessments driven by an independent, multidisciplinary, and multi-stakeholder group of experts brings legitimacy to an issue area and international alignment on actions needed.
- Standards can influence the development and deployment of specific AI systems through product specifications in areas such as explainability, robustness, and fail-safe design. The involvement of key stakeholders in the standard-setting process is important to ensure that the necessary guardrails are appropriately implemented.
- While drawing lessons related to auditing and monitoring procedures from the models of entities focused on coordinating global governance of civil aviation, maritime operations, or nuclear energy, it is essential to distinguish between these specific subject areas and the decentralized nature of AI systems.
- The UN's experience in development work, including capacity-building support to broader strategy work like adoption of national AI strategies, can provide avenues for tailored development programs catering to the specific needs of Member States and stakeholders. Such capacity development can support AI development that is grounded in fairness, gender equality, reliability, safety, interpretability and accountability.
- AI requires a 'dual-action' approach to address safety risks as well as opportunities for sustainable development presented by the technology, especially in developing countries.
- The UN System's efforts in areas such as cybersecurity provide lessons for effectively leveraging the complementary strengths and mandates of multiple entities to address the cross-cutting nature of technology.

Focus Area III – Learnings from existing governance structures, inclusive normative processes, and agile and anticipatory approaches from the UN system

Key observation: As highlighted in the [Governing AI for Humanity interim report](#) prepared by the HLAB-AI, developing and implementing AI governance should be inclusive, universal, rooted in multi-stakeholder collaboration, adaptive, and anchored in the UN Charter. In addition, agility in normative process and instrument design and enhanced foresight focus on technological developments can boost existing processes.

- In the normative efforts focused on AI and global public goods, the UN System champions inclusivity in its norm-making processes through multistakeholder participation, tripartite structures, and internal coordination frameworks.

- Engaging the private sector is key for global AI governance efforts, including in the normative and operational processes of the UN. The System can also advance private sector accountability by supporting and advocating for reporting mechanisms across voluntary and mandatory requirements.
- Providing capacity development and detailed technical guidance can help develop a trusted platform for assessments of requirements, like *ex-ante* assessments of AI systems.
- Flexible and dynamic decision-making mechanisms are essential in specific contexts, especially those related to emergency/safety.
- Aligned with [UN 2.0](#), existing and emerging normative processes can be further strengthened with foresight and lessons from the system's anticipatory governance in food, climate, and humanitarian work.

General recommendations for consideration, including by the (HLAB-AI:

- 1a. Global AI governance efforts in the UN System are anchored in international law, including the UN Charter, International Human Rights Law, and other agreed international commitments such as the 2030 Agenda for Sustainable Development. To adequately cater to the specific requirements and economic, social, and environmental priorities of different sectors, UN instruments and frameworks providing sector-specific guidance are key to these governance efforts.
- 1b. The UN System's convening power, normative and policy instruments, institutional functions and frameworks, stakeholder networks and resources, and diverse expertise and experience should be leveraged to help address global AI governance challenges, including scientific and political consensus building, monitoring and enforcement, capacity development on technical, policy and regulatory aspects, enhanced multistakeholder collaboration, and technology diffusion and dissemination.
- 1c. Well-conceived AI governance should provide appropriate incentives and guardrails commensurate with the particular characteristics of different AI systems and applications to advance ethical and human rights-based governance while maximizing the technology's positive impact on society and mitigating its risks.
- 1d. Effective AI governance could be delivered through an ecosystem of critical functions, including but not limited to technology development and consensus building through research and analysis, stakeholder engagement and coordination, standards and norm-setting, capacity building, and monitoring and accountability. These functions have already been tested by the entities surveyed, and this experience can deliver tailored approaches on the basis of the specific networks of focus areas and diverse stakeholder groups within the UN System, facilitated by established governance structures and coordination mechanisms.
- 1e. Adopting a pragmatic approach and building on the current governance initiatives in the UN system when designing global AI governance efforts is crucial. In addition, the fast pace of technology development, compared to the relatively slow processes to develop new international law instruments of institutional structures, the need for regional or industry/sector-specific approaches, as well as the level of agility of mechanisms and processes that exist at an institutional level also need to be factored in. The UN System has launched various initiatives to adapt to this changing pace of technology development.
- 1f. Link international AI governance to ongoing efforts on international data governance to ensure complementarity and avoid fragmentation. Effective AI governance requires an integrated approach across related or adjacent governance efforts; it relies on the principles and practices of data governance as availability of quality data is a key enabler for the development of AI.

Summary of specific recommendations for the consideration by the UN system:

- 2a. Present the UN System's tools and instruments in AI governance as a combined toolbox for the benefit of the Member States and stakeholders. This toolbox should encompass cross-cutting and sector-specific instruments. In this regard, consider building upon the more impactful initiatives to enhance the effectiveness of the System's response to global AI governance.

- 2b. Expand taxonomies for existing instruments, including international human rights law, to facilitate technical and normative guidance for leveraging existing instruments and complement these efforts with capacity-building support.
- 2c. Put in place or update internal policies to govern the use of AI within UN system entities, building on the work of the HLCM Task Force on AI to develop a normative guidance or model policy for the UN system on the use of AI.
- 2d. Leverage, enhance, and scale observatories on AI to disseminate best practices, use cases, and lessons learned on the use, identification, adoption, and implementation of existing instruments.
- 2e. Leverage existing multidisciplinary global networks in the UN ecosystem for building consensus and communicating key technological milestones and developments in AI, enabled by effective information flow and exchange.
- 2f. Invest in and develop in-house granular and comprehensive AI expertise to support Member States effectively, engage with stakeholder groups, and build trust.
- 2g. Enable sandboxes to facilitate the development of internationally harmonized approaches for AI risk assessments and monitoring efforts.
- 2h. Proactively manage risks and mainstream foresight capabilities across all system efforts on technology and AI governance through cross-sectoral and multi-stakeholder partnerships, including with academic institutions and think tanks focused on technology foresight.
- 2i. Invest in talent, data, compute resources, and regulatory and procurement capacity, through initiatives aimed at addressing the AI divide under the broader context of the digital divide and enhancing technical capacities for policymakers and practitioners of AI.

A. Introduction

1. With a transformative impact across many industries and sectors, Artificial Intelligence (AI) can spark global innovation, enhance data-driven decision-making across all countries, and boost progress toward the Sustainable Development Goals (SDGs). Amidst rapid innovation and advancements in AI, there is growing recognition that the technology pairs opportunities for extraordinary growth and inclusive, sustainable development with the potential for significant disruption and risk. The discussion on risks is especially pertinent for areas in which AI technologies could have adverse implications for peace and security, such as autonomous weapons systems, as well as for the role of AI in misinformation, its use by law enforcement, or public service delivery.
2. AI is already the focus of intense geopolitical competition. States with the financial and technical resources and capacity are prioritizing AI systems as a strategic objective. Meanwhile, technology companies wield significant influence and control over data, algorithms, and computational resources. Harnessing the benefits of AI systems while minimizing their risks and potential harms requires a collaborative effort among stakeholders through shared global responsibility. The rapid emergence of generative AI applications has further accelerated the push toward exploring adaptable governance models and mechanisms that evolve with technological progress. These models should combine openness to experimentation and innovation with risk-conscious and responsible technology adoption to advance inclusive and sustainable development and ensure trust. Such models also need to factor in the limited resources that small and medium-sized enterprises or stakeholders have for compliance and regulatory efforts, and they should be designed proportionately.
3. These trends reaffirm the need for global AI governance to be anchored in the United Nations (UN) Charter and human rights framework and for the UN System to continue leveraging its unique convening power to facilitate international cooperation and multi-stakeholder engagement in governance efforts.

1. Risks, challenges, and considerations for global AI governance

4. Given the horizontal, transversal, and cross-sectoral nature of AI, it has the potential to impact all aspects of humanity. Mitigating associated risks will require multiple lenses to address any issues comprehensively – from economics and markets to the social fabric and digital rights to inclusivity. A global AI governance discussion, therefore, needs to factor in the following dimensions – i) the AI value chain and its socio-economic and geopolitical impacts, as some countries vie for technological dominance while others risk being confined to sources of data, cheap labor or raw materials; ii) the AI life-cycle, which looks at the technical dimensions of AI development and deployment; iii) the broader macro risks and challenges associated with AI; iv) the diverse levels and avenues of impact that AI has on different sectors and aspects of society – from precision agriculture supporting higher agricultural yields, to lethal autonomous weapon systems introducing new threats to peace and security; and v) that governance must go hand-in-hand with the sustainable development opportunities presented by AI.
5. These dimensions have also helped guide the structure of this paper and the selection of examples researched and presented across the focus areas.

(a) The AI value chain

6. An AI value chain is typically comprised of the following elements: computer hardware → cloud platforms → data and AI models → applications → services. As AI use and innovation gain momentum, an equity gap and unequal concentration of power are emerging across all elements of this value chain. For instance, the economic benefits of AI applications are accrued primarily by three stakeholders: businesses building AI models and applications, hardware companies supplying the chips and compute capacity that power AI models, and cloud service providers. Most of these companies are currently located in a handful of countries.
7. Conversely, workers in developing countries are delivering on a substantial chunk of the work related to aspects such as data labelling and content moderation while accruing limited economic

benefits. While such divides have increasingly been observed with technological advancements over the past several years, the scale and pace of the impact of AI technologies and applications on the digital divide is unprecedented. It has the potential to exacerbate the gap between developing and developed countries while adversely affecting women and children and people in marginalized situations, who are both at risk of facing bias from AI systems as well as having limited control over how their data is collected and processed. Geopolitical factors are also essential to consider, as the design and manufacturing of hardware chips are restricted to a few companies and countries, while a number of other countries have their role confined to being sources of essential minerals and metals needed to produce such chips and other hardware.

8. It is also important to consider the environmental cost of building and training large AI systems as well as the resource-intensive manufacturing of the hardware-powering AI systems. While some emerging AI regulations reflect environmental considerations, and some studies are assessing and measuring the energy consumption of AI systems, there is a gap in global accountability regarding the overall carbon footprint of the AI value chain. As per one estimateⁱ, generating an image using a large AI model can use as much energy as fully charging a smartphone. In addition, mining essential minerals and metals to manufacture chips for AI-related hardware leads to massive greenhouse gas (GHG) emissions and water consumption in the manufacturing process. The environmental dimension also reflects another equity gap, as countries facing energy poverty might have limited energy resources to power data centres, a **foundational layer in AI systems**.

(b) The AI life-cycle

9. As described in the United Nations Educational, Scientific and Cultural Organization's (UNESCO) Recommendation on the Ethics of Artificial Intelligence, another dimension is the AI life cycle, which ranges from research, design, and development to deployment and use, including maintenance, operation, trade, financing, monitoring and evaluation, validation, end-of-use, disassembly and termination of AI. When viewed from a life cycle lens, the technical dimensions of AI governance would involve governance across these various elements, including aspects like the interpretability of models, the quality of training data, licensing and certification for AI applications and services, and other factors. Data governance, data protection, privacy, and cybersecurity remain essential aspects throughout the life cycle of AI development.

(c) The broader macro risks and challenges associated with AI

10. On the dimension of risk, while it is challenging to define every possible risk from AI, especially from general-purpose AI systems, some key ones have been highlighted in the diagram below, which include those identified by UN System entities during research for this paper, those broadly highlighted by experts, as well as those highlighted in the [Governing AI for Humanity interim report](#) by the HLAB-AI. The Annex of the report contains further details on these risks.

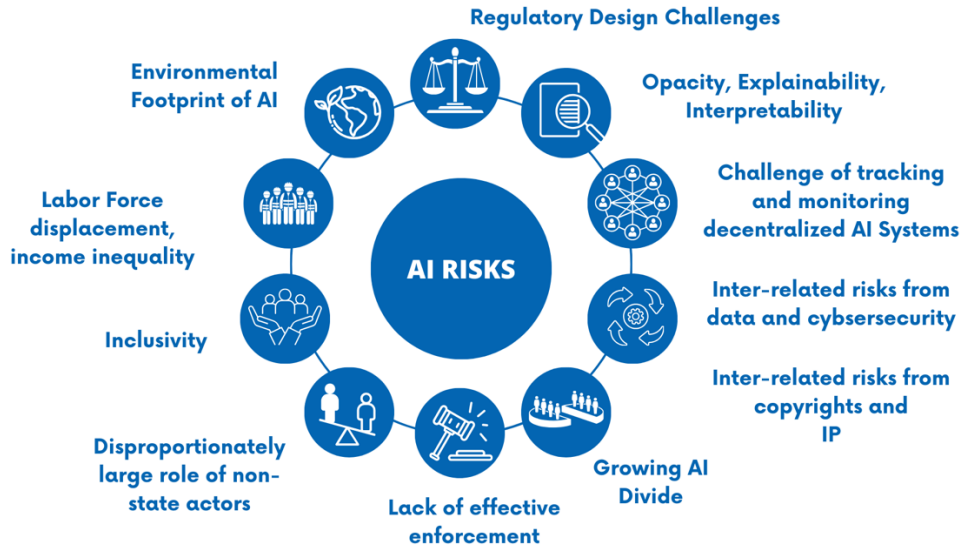


Illustration 1: Current and evolving global AI risks

(d) The diverse levels and avenues of impact that AI has on different sectors and aspects of society

11. AI presents diverse levels and avenues of impact across different sectors and aspects of society – from precision agriculture supporting higher agricultural yields to lethal autonomous weapon systems introducing new threats to peace and security. AI governance discussions and approaches need to be designed, taking into consideration overarching principles as well as the varying levels of risks and opportunities across sectors.

(e) Governance must go hand-in-hand with sustainable development opportunities

12. AI governance needs to factor in the capacity development needs, including facilitating and scaling AI use cases for societal benefit, like in the area of climate or health. Scaling up these use cases requires investments and concerted policy efforts across the entire value chain of AI, particularly for compute capacity and cloud infrastructure, data access, and democratizing AI.

2. Context for the paper

13. The UN System has been closely tracking and responding to developments in the AI space, including the growing digital divide, divergent regulatory approaches, and a proliferation of principles and tools. There have been recent calls to enhance international cooperation on AI and to address the growing challenges in AI governance, including by effectively engaging all stakeholder groups. In this regard, some experts have suggested creating new mechanisms and related bodies to govern AI, while others have called for concerted efforts within existing institutions, including the UN System, to tackle AI governance. The UN System has a unique role in norm-setting and consensus building, capacity development, and international cooperation on topics ranging from human rights to climate, disarmament, health, and technology, ready to be leveraged to address new challenges or new areas of governance. The System recognizes the importance of ensuring a human rights-centred approach to AI governance through effective, diverse, and multistakeholder engagement.

14. AI’s complex, multi-dimensional nature demands both overarching efforts as well as focused actions within specific sectoral verticals and a decentralized network of networks governance approach. The UN System’s structural strengths provide robust avenues and mechanisms for AI governance and acknowledge the necessity for continual improvements to meet the evolving challenges posed by AI technologies. Certain UN System entities have primarily sectoral

mandates encompassing areas such as health, agriculture, economics and finance, trade, industry, education, labour, nuclear, aviation, environment, and others; while the mandates of other entities are focused on adherence to overarching principles, such as human rights, and others yet on the well-being of vulnerable groups, such as women and children. Each brings specific networks of focus areas and diverse stakeholder groups facilitated by established governance structures and coordination mechanisms.

15. Recognizing the importance of leveraging existing mandates of UN System entities in their particular areas of knowledge and expertise, in October 2023, the Joint Session of the [High-level Committee on Programmes](#) (HLCP) and the [High-level Committee on Management](#) (HLCM) of the [United Nations System Chief Executives Board for Coordination](#) (CEB)² on “The use and governance of Artificial Intelligence (AI) and related Frontier Technologies” (Joint Session) requested [the Inter-Agency Working Group on AI \(IAWG-AI\)](#) comprising over 40 UN entities³, with input from HLCM as relevant, to develop a White Paper on AI governance by analyzing current institutional models and related functions and existing international normative frameworks in the UN System that could be applied or leveraged for international AI governance, for consideration by HLCP at its 47th session. CEB subsequently welcomed this outcome. This paper is intended to inform UN System deliberations on AI and is also a UN System-wide contribution to the Secretary-General’s High-level Advisory Body on AI. In December 2023, preliminary findings from the research and survey were shared with the co-chairs of the HLAB-AI. At the joint session, members also decided to develop a “*system-wide normative and operational framework on the use of AI in the UN System, to establish appropriate knowledge sharing mechanisms for ideas and experiences, including the possibility of developing a generative AI platform for the UN system, as well as to pool the necessary technical capacity,*” as a complementary and internally focused effort. Subsequently, a HLCM Task Force on AI was established to develop an internal normative guidance/model policy for the UN System on the use of AI and to identify and promote mechanisms for pooling technical capacity and knowledge sharing on AI.

3. Overview of the paper

(a) Focus areas of the paper

16. The paper is focused on three key areas as outlined below:
- **Focus Area I** – Existing normative⁴ and policy instruments within the UN System to inform and shape AI governance
 - **Focus Area II** – Institutional functions that inform global AI governance
 - **Focus Area III** – Lessons learned from existing governance structures, normative processes, and agile and anticipatory approaches from the UN System
17. These focus areas are based on the request of the Joint Session highlighted above, as well as the ongoing global dialogue and debate on the governance of AI. Several experts and civil society

² CEB is the longest-standing and highest-level coordination forum of the United Nations system. The 31-member body is chaired by the UN Secretary-General. It seeks to enhance UN System-wide coherence and coordination and provides broad guidance and strategic direction to the UN System on issues of system-wide concern. The work of the Board is supported by two High-level Committees: The High-level Committee on Programmes (HLCP) and the High-Level Committee in Management (HLCM), which are principal mechanisms for UN System coordination and policy coherence in the areas of programmes and management. Under each Committee, inter-agency mechanisms and networks coordinate on specific topics.

³ The Inter-Agency Working Group on Artificial Intelligence (IAWG-AI), co-led by UNESCO and ITU, was created by the HLCP in 2020. It brings together United Nations system expertise on AI, in support of the work of CEB and HLCP, to enhance inter-agency cooperation in the area of AI.

⁴ Normative work is defined as “the support to the development of norms and standards in conventions, declarations, regulatory frameworks, agreements, guidelines, codes of practice and other standard setting instruments, at global, regional and national level. Normative work also includes the support to the implementation of these instruments at the policy level, i.e. their integration into legislation, policies and development plans, and to their implementation at the programme level.” (UNEG, 2012, p. 5)

groupsⁱⁱ have emphasized the importance of leveraging existing instruments for technology governance, especially the international human rights framework, as a binding framework for governance. While technologies keep evolving, human rights are technology-neutral, even as their interpretation in the digital context might require guidance to adapt to the changing realities of a technology and AI-driven era.

18. The study of existing institutional models and governance structures of the UN System is to help understand how the system has adapted not only its strategy and programmatic delivery over the past several decades to the changing realities and geopolitics of the world but also its experience as a trusted fora to build consensus through normative and multistakeholder processes, in addition to developing tailored capacity development programs in support of Member States on complex and challenging areas.
19. Broadly, the paper will elaborate upon the abovementioned focus areas, review existing UN capacities and resources, highlight examples, and draw lessons from research and findings.

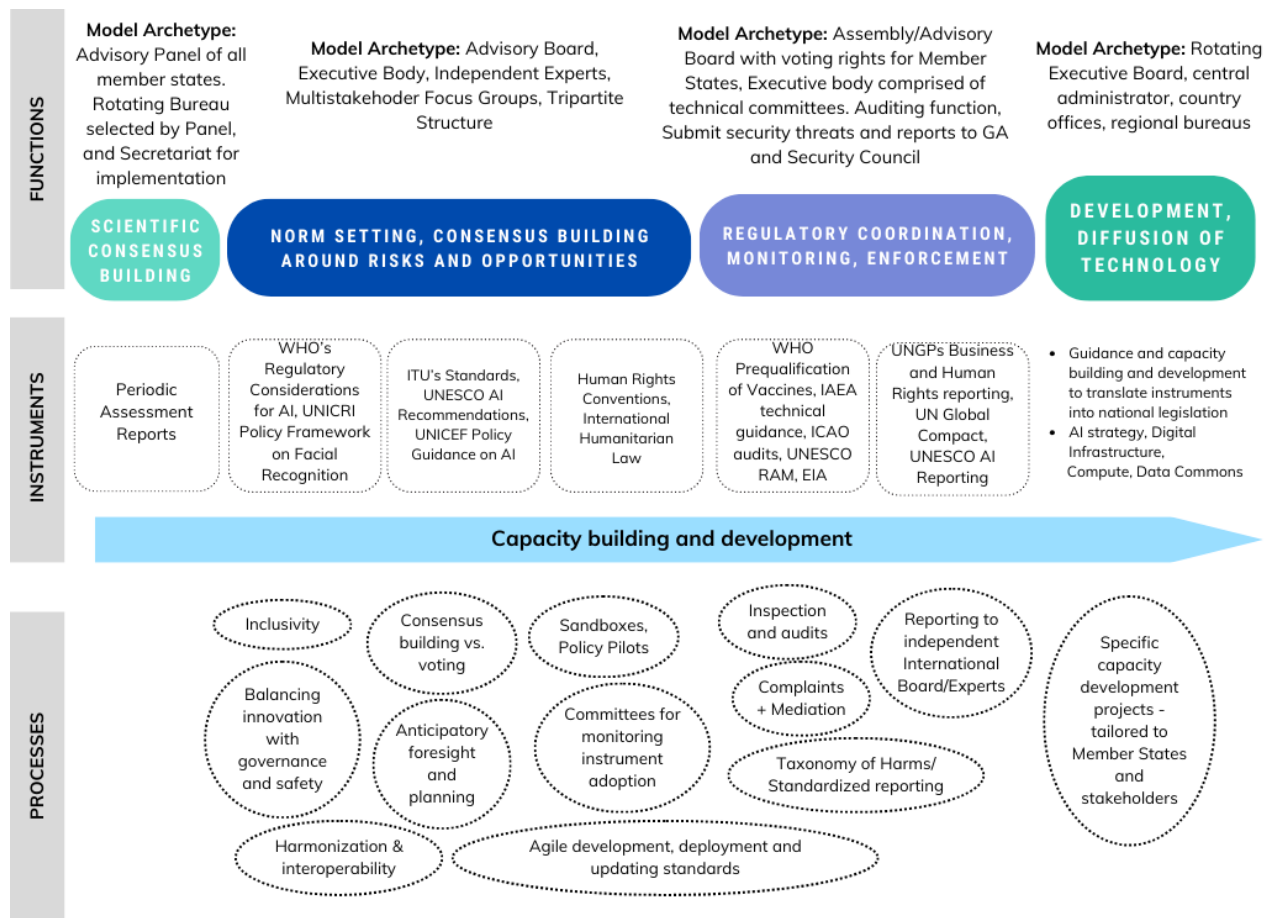


Illustration 2: Focus areas of the White Paper (examples are indicative and not exhaustive)

(b) Methodology

20. This paper has been developed by the IAWG-AI, co-led by ITU and UNESCO, and comprised of over 40 UN entities. The findings presented in this paper are based on the results of a UN System-wide survey, follow-up interviews with select entities, and a desk review of global trends and existing literature^{iii,iv,v,vi}, including the [Governing AI for Humanity](#) interim report from the HLAB-AI. Details on the research methodology can be found in the Annex of this paper, with the illustration below providing a brief overview of the methodology and process.

In addition, the Annex also includes highlights from the early analysis of the survey, which was also shared at a CEB meeting on 9th November 2023.

21. The survey received inputs from 44 UN System entities. It was followed by expert interviews conducted with 10 UN System entities based on their relevance to AI governance or experience in addressing challenges similar to the ones encountered in AI governance.



Illustration 3: Overview of methodology

B. Focus area I – Overview of existing normative and policy instruments in the UN system to inform and shape AI governance

22. The UN System deploys various normative and other instruments across its entities, including treaties, conventions, resolutions, standards, guidelines, policy guidance, and other instruments.⁵ These instruments cater to different policy and issue areas and can be cross-cutting or sector-specific. The first focus area of the paper presents an illustrative overview of existing international legal frameworks and normative and policy instruments within the UN System that could be applied or extended to AI, as identified by the UN System entities via the survey (for the full list, please visit the Annex). The classification of instruments, as presented in the sections below, has been developed solely to facilitate ease of analysis of these instruments in the context of the AI governance risks and opportunities discussed in the paper. This paper does not aim to present a qualitative assessment of the instruments, a comparative assessment vis a vis other international or regional instruments, or a legal classification and analysis.

1. Overview of instruments

23. Over 50 laws and instruments have been identified by the survey respondents, including both binding and non-binding instruments, that are either directly applicable to AI or are applied in inter-related areas like ethics, data, cybersecurity, copyrights, patents, information integrity, disarmament, human rights, international labour standards and codes of practice, international humanitarian law, and others. These include treaties, conventions, recommendations, compliance and certification procedures, policy instruments, technical standards, and other instruments. The terminologies and legal effects of these international laws and other instruments vary from one organization to another. In particular, laws or instruments negotiated and approved by Member States carry significant weight and standing in terms of expressing their commitments or directing implementation efforts.

⁵ United Nations Treaty Collection - refer to this [link](#) for more information

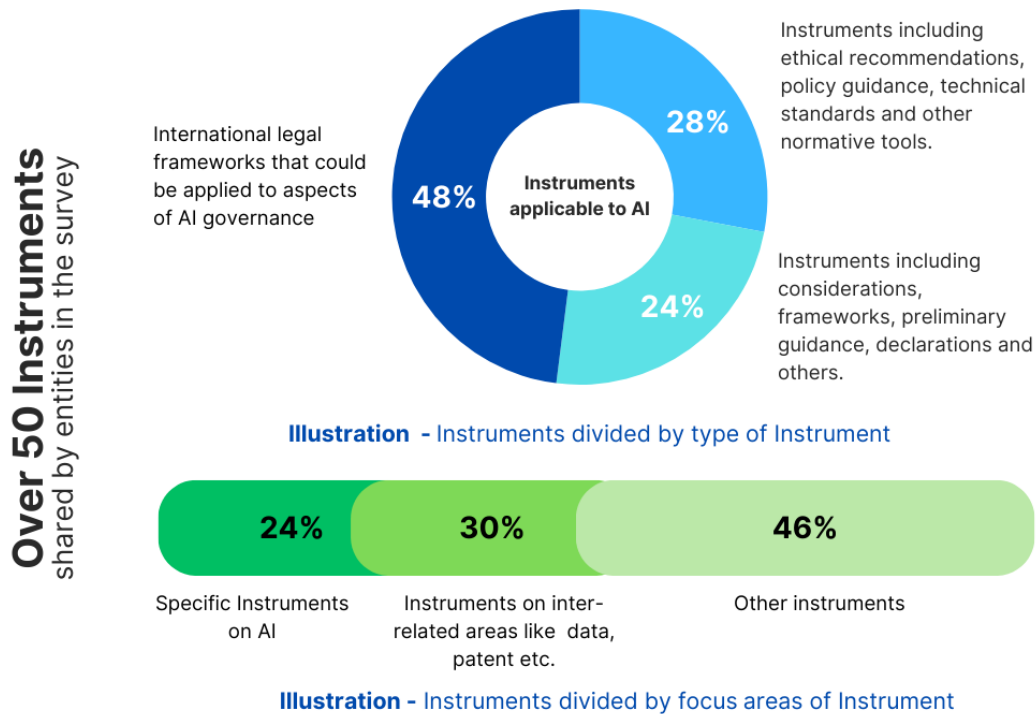


Illustration 4: Instruments within the UN system applicable to AI – based on survey responses

24. An analysis of these instruments shows that some have provisions that already apply to the digital sphere, particularly those relating to human rights, international labour standards, and humanitarian law. In contrast, there are some instruments in areas related to AI whose application could be extended to AI. Some instruments that apply to other forms of global public goods are relevant as possible examples from which lessons can be drawn for AI governance. An overview of all these instruments is presented below, followed by additional details and key takeaways.

(a) Specific instruments on AI governance

25. UNESCO’s [Recommendation on the Ethics of AI](#) (UNESCO Recommendation) serves as a comprehensive framework for ensuring ethical governance of AI throughout all stages of the AI system life cycle (i.e., from research, design, and development to deployment and use, including maintenance, operation, trade, financing, monitoring and evaluation, validation, end-of-use, disassembly, and termination). The UNESCO Recommendation was developed through a global multistakeholder process and adopted by 193 Member States in November 2021 – an undertaking that demonstrates that the UN System is responding to a gap in establishing globally shared values and principles around ethical and responsible AI focused on promoting human rights and human dignity, fairness, and sustainability. Its design encompasses values and principles, including the rule of law, human determination, accountability, transparency, and various policy areas, including ethical governance, gender, education, environment, and others.
26. Furthermore, the UN Department of Economic and Social Affairs (UN-DESA), the World Health Organization (WHO), the United Nations International Children’s Emergency Fund (UNICEF), the United Nations Development Programme (UNDP) and the United Nations Interregional Crime and Justice Research Institute (UNICRI) have instruments directly focused on AI governance. The International Telecommunications Union (ITU) is working on international AI-related technical standards across specific sectors, many of which have been developed in collaboration with other UN agencies, and has published over 100 standards on AI

already, with 120 more in development as of 2024. The [report](#) from DESA's Internet Governance Forum (IGF) Secretariat, produced through an expert, multistakeholder group, the IGF Policy Network on AI, focuses on AI interoperability in the global regulatory landscape, AI biases in gender and race, as well as current use cases for environmental security. With more than forty co-authors, including from the ITU, UNESCO, and UN Women, it provides a broad international lens on some of the most contentious and promising aspects of AI application. The WHO, UNICEF, and UNICRI's instruments have been tailored to the specificities of their sectors - they range from policy guidance on AI for children to toolkits focused on specific use cases like facial recognition in law enforcement. UN-Habitat's upcoming International Guidelines on People-Centred Smart Cities will set key principles for AI development and application across cities globally, ensuring adherence to human rights and inclusion⁶.

27. The UN System's instruments on AI are helping shape norms and providing resources and tools for Member States during a period of flux in the global AI governance space. For instance, the UNESCO Recommendation has accompanying implementation chapters and tools like the [Readiness Assessment Methodology](#) (UNESCO's RAM), which is currently being piloted in the first cohort of 50 countries from all regions of the world. The RAM helps countries identify gaps in different dimensions related to a country's AI ecosystem, including the legal and regulatory, social and cultural, economic, scientific and educational, and technological and infrastructural ones. This exercise comprises the establishment of a steering committee, with the Ministers engaged in AI, along with the private sector, academics, and civil society. It contributes to shaping or strengthening the country's national AI strategies and enables tailored support for governments on specific needs, such as institutional and legal reforms. With the results of UNESCO's RAM published on [UNESCO's Global AI Ethics and Governance Observatory](#) and discussed at the [Global Forum on the Ethics of AI](#), such a tool facilitates peer-to-peer learning among countries.
28. [UNICRI's Policy Framework for Responsible Limits on Facial Recognition, Use Case: Law Enforcement Investigations](#), is being considered for adoption into national legislation by some Member States. Its [Toolkit for Responsible AI Innovation in Law Enforcement](#) has been tested in a practical setting by INTERPOL and 15 law enforcement agencies. [UNICEF's Policy Guidance on AI for Children](#) has also been integrated into Scotland's National AI Strategy. Built upon UNESCO's Recommendation on ethics and human rights and aligned with UNDP's broad mandate to work across sectors, the [UNDP's AI Readiness Assessment \(AIRA\)](#) assists governments in understanding the AI landscape, both as users and enablers across various sectors and has been implemented in three countries, with plans to extend it to 35 countries in 2024. [UNIDO's Gender, Digital Transformation and Artificial Intelligence](#) assessment reviews the current state of policies and initiatives related to the promotion and strengthening of global efforts toward gender-transformative strategies for artificial intelligence.
29. As existing instruments on AI are adopted or new ones are introduced, they should continue to be implemented/designed in a manner that supports Member States, with appropriate guidance on how to respect, protect, and promote human rights and fundamental freedoms throughout the life cycle of AI systems.

(b) Instruments on areas inter-related with AI governance

30. In the context of AI governance, the UN's work on international data governance, data protection, and privacy is important to consider – given the close relationship between data and AI and the role diverse data sets can potentially play in reducing bias in AI systems and building trustworthy AI. Particular attention is needed towards building foresight on the potential human rights impacts of data collection, retention, processing, and transfer, especially for AI-driven tools deployed in a security context. In addition, access to data and data commons for enabling AI applications focused on the SDGs is also an important consideration. Multiple entities, including UNDP, UNESCO, UNICEF, the World Food Programme (WFP), the United Nations

⁶ 193 Member States adopted the resolution mandating UN-Habitat to develop International Guidelines on People Centred Smart Cities in June 2023 at the UN Habitat Assembly in its second session.

Commission for Trade and Development (UNCTAD), ITU, the Universal Postal Union (UPU), the United Nations High Commissioner for Refugees (UNHCR), UN-Habitat, the International Labour Organization (ILO) and WHO have developed instruments and efforts focused on data, ranging from data policy to data protection to digital public infrastructure to address these issues.

31. Specific data-related instruments, for example, in migration, human rights, or health, demonstrate the UN System's recognition of the varying sensitivities and nuance related to data governance, protection, and privacy of different sectors. The HLCP Working Group on International Data Governance, co-led by WHO, as Chair of the Committee of Chief Statisticians of the United Nations System, and the United Nations Office on Drugs and Crime (UNODC), has also developed a [paper](#) that articulates a vision for the governance of data at the international level and its annexes can serve as an analytical resource to support Member States' efforts. The World Bank's World Development Report 2021 – *Data for Better Lives* - included a number of [data resources](#) to assist policy-makers in elaborating digital development strategies. The World Bank's Global Data Regulation Diagnostic⁷ further builds on the survey and the trust framework to provide a more expansive geographic coverage.
32. In addition, some instruments under development, like the UN System's [Code of Conduct for Information Integrity](#), are closely related to AI governance as they pertain to information flow on digital platforms. These digital platforms have observed several cases of misinformation fuelled by deep fakes and misuse of generative AI tools. It's also important to take note of UN instruments related to cybersecurity, which will be discussed in a later section of the paper.

(c) International legal frameworks that could be applied to aspects of AI governance

33. International law, including international human rights law, is of fundamental importance for designing and implementing AI governance frameworks, institutions and process, both domestically and internationally, and various existing international legal frameworks can readily be extended or applied to mitigate harms emerging from AI. International human rights law reflects universally agreed values and encompass a vast range of civil, political, social, economic, and cultural rights. They clearly define the harms to address and the goals to achieve and can guide the prioritization of risks and actions. Crucially, international human rights law defines binding obligations for States to respect, protect and fulfil human rights. Businesses have a responsibility to respect human rights. Principles and recommendations from numerous reports and resolutions across the UN human rights system are relevant to the AI context and include extensive language adopted by consensus. The General Assembly and Human Rights Council, for example, have underscored that human rights should be respected, protected, and promoted throughout the lifecycle of AI systems. They have also specifically highlighted the importance of applying human rights due diligence⁸.
34. Among binding obligations are those arising from the International Covenant on Civil and Political Rights, with relevant obligations pertaining, for example, to the right to freedom of expression, the right to privacy, and fair trial rights. The International Covenant on Economic, Social and Cultural Rights contains provisions defining states' duties in sectors such as social security, employment, and health, which are relevant to AI. The obligations under the [Convention on the Rights of the Child](#) and under the Optional Protocol on the Sale of Children, Child Prostitution, and Child Pornography, especially when read in light of the [General Comment No. 25 \(2021\) on children's rights in relation to the digital environment](#), are specifically relevant to digital harms⁹.

⁷ To be published imminently.

⁸ [A/HRC/RES/53/29; A/C.3/78/L.49/Rev.1.](#)

⁹ The Committee on the Rights of the Child published [guidelines regarding the implementation of that protocol \(revised in 2019\)](#). They do not mention AI explicitly. However, by encouraging State parties in para.63 to include in their legal provisions on child sexual abuse “representations of non-existing children”, the guidelines gain direct relevance to synthetic images, such as deep fakes. Para. 75 emphasizes the need to apply article 7 of the Optional Protocol to online

35. It is also important to take note of treaties under the auspices of WIPO¹⁰, factoring in the concerns about copyrighted materials and their use in training large language models. In addition, there may be legal uncertainty in the ability to patent innovations ‘co-developed’ with AI and open questions around the disclosure requirements of the patent system in the context of black-box algorithms.
36. AI’s impact on labour and the associated instruments is also an important factor in this discussion. As highlighted in the risks section, the impact of AI on the workforce cannot be understated. Automation and job displacement have been a feature of all past industrial revolutions. However, the pace and complexity of change that AI brings forth are staggering. [As per recent analysis from the IMF](#), almost 40 percent of global employment is exposed to AI, with advanced economies at greater risk but also better poised to exploit AI benefits than emerging market and developing economies. The [ILO Declaration on Fundamental Principles and Rights at Work](#) applies to all working environments impacted by AI, irrespective of whether ILO member States have ratified all fundamental standards. Ongoing research provides examples of both AI’s positive and negative impact on human rights at work. It will develop global standards, enhancing decent work in the platform economy.
37. In addition, ICAO has also highlighted current deliberations related to AI safety and how it applies to their existing normative instrument on aviation safety ([Chicago Convention – Convention on International Civil Aviation](#)).

(d) International instruments for governance of global public goods and sector-specific approaches

38. Through the survey, UN entities have shared several instruments that span a range of sectors, including health, aviation, labour, climate, education, communications, disarmament, and others. While a detailed analysis of these instruments is beyond the scope of this paper, they have been reviewed with the objective of drawing lessons from their monitoring, reporting, and enforcement mechanisms, their success factors, and the UN’s role in facilitating their adoption by stakeholders, especially given the varying capacities and interests of Member States.

2. Key takeaways

39. **Numerous instruments developed within the UN system are highly relevant for the governance of AI.** International law, including the UN Charter and international human rights law, provide the fundamental frameworks that should underpin the design, implementation and operation of governance instruments, mechanisms, institutions and processes. Instruments that are directly applicable to AI demonstrate the UN System’s engagement and presence across all layers of governance – ranging from broader normative frameworks to sector-specific ones. Instruments on AI within the UN System align with what is being observed in current national, regional, and local AI governance efforts that are witnessing both cross-sectoral or overarching regulatory efforts and sector-specific regulation. As an overarching instrument, the [UNESCO Recommendation](#) provides a policy-friendly blueprint for the ethical governance of AI across sectors. The accompanying [Ethical Impact Assessment Tool for AI](#) can be applied to procurement cases across different sectors for public or private entities. Similarly, the accompanying [Readiness Assessment Methodology tool](#) can be used by Member States with the flexibility to adapt it to their needs. In addition, the policy chapters of the UNESCO

chats and forums. The data collection recommendation in para. 20 could be a tool for consistent monitoring of the implementation of the Optional Protocol. Similarly, discriminatory or biased automated hiring systems could be under the purview of the Convention on the Elimination of All Forms of Discrimination against Women and its relevant articles. The Convention on the Elimination of All Forms of Racial Discrimination contains similar non-discrimination obligations, and the Committee on the Elimination of Racial Discrimination has adopted a General Recommendation on preventing and combating racial profiling by law enforcement officials (CERD/C/GC/36) with recommendation concerning AI.

¹⁰ WIPO’s treaties - [Rome Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organizations](#), [Berne Convention for the Protection of Literary and Artistic Works](#), the [Beijing Treaty on Audiovisual Performances](#), as well as the [Patent Cooperation Treaty](#)

Recommendation can provide inputs to policies in critical sectors like health, employment, environment, gender, and education.

40. Member States are also implementing methodologies from UNDP, UNCTAD, and other agencies to support their development of AI governance and enable the use of AI for sustainable development. Broad-based normative frameworks and tools are essential for setting a baseline and for international alignment, especially when national and international legislative efforts and capacities are still evolving and adapting to the continuous changes in AI. The UN System can also leverage experiences from local and regional governance practices, which can inform national and global frameworks.
41. On the other hand, instruments from ITU, UNESCO, WHO, UNICEF, UNIDO, UN-Habitat, or UNICRI support the nuanced needs of their sectors (e.g., the specificities of health data or AI in medical devices, or AI in education or AI in cities and urban context) and allow for specific technical details to be developed, in collaboration with stakeholders. These details can be further nuanced through deep dives on specific use cases (e.g., facial recognition use cases in law enforcement). As highlighted earlier in the paper, UNICRI's tools were developed and deployed in collaboration with INTERPOL and piloted or trialed with law enforcement agencies. ITU's technical standards are developed through a consensus-driven multi-stakeholder platform, often in partnership with other UN agencies and/or led by industry from a specific sector. UN-Habitat's AI Risk Assessment Framework for Cities provides recommendations for developing AI strategies.
42. The case for a sector-specific approach also emerges from the review of multilateral instruments addressing other global public goods, such as the Montreal Protocol on substances that deplete the Ozone layer and WHO's Convention on Tobacco Control. Scientific and political consensus on the issue area was key to the success of these instruments. The specificity of the issue (e.g., substituting CFCs in refrigeration equipment, banning smoking in public places, or banning the sale of tobacco to minors) helped facilitate international agreement on the scientific evidence, urgency, timeline, and action plan, which were vital to the success of these instruments. Learnings from this can be translated by delineating use cases of AI in the public good domain (for example, AI for climate modelling, AI use cases for addressing public health challenges, and pre-competitive AI research) from the more commercial or military uses of AI. This approach also might allow for shorter time frames to build consensus on targeted topics and achieve small but material wins.
43. **The UN System's technical instruments, such as international standards, play an effective role in facilitating norm-setting and interoperability around the opportunities and risks of AI.** Technical standards can encourage competition and innovation, facilitate compatibility and interoperability, improve cost efficiency, and promote national development across a range of sectors, including health, financial services, transportation, energy, agriculture, smart cities, and aviation. On AI specifically, technical standards can influence the development and deployment of specific AI systems through product specifications with requirements for explainability, robustness, and fail-safe design. They can also affect the larger context in which AI is researched, developed, and deployed through process specifications. Particularly, safeguards, including those related to privacy and data protection, oversight mechanisms, and transparency reporting requirements, can be incorporated to facilitate human rights compliance in the development of technology. These safeguards are especially pertinent in the context of technology in high-risk settings, including when technology affects people in marginalized or vulnerable situations, such as children or people on the move.
44. Within the UN System, work on technical standards for AI is already underway in the ITU, with over 100 standards already published on AI and 120 more in development as of 2024. The standardisation development work often takes place in collaboration with other UN agencies that have lead mandates and domain expertise in specific sectors. Examples include ITU's work with WHO and WIPO on AI for health, the Food and Agriculture Organization (FAO) on AI for agriculture, the United Nations Environment Programme (UNEP) and the World Meteorological Organization (WMO) on AI for natural disaster management, and the United Nations Economic

Commission for Europe (UNECE) on intelligent transport systems and automated driving. ITU also works closely with other SDOs to strengthen and advance the voluntary consensus-based international standards system including through the World Standards Cooperation (WSC), a high-level collaboration between the IEC, ISO and ITU.

45. Towards these efforts, the [AI for Good platform](#) has been instrumental in facilitating the development of agile multi-stakeholder working environments to develop specifications rapidly to address industry needs as they emerge. Standards Development Organizations (SDOs) such as the ITU have set up mechanisms to augment the standards-making processes by providing an alternative working environment to develop specifications in their relevant areas more flexibly and rapidly. These groups are open to all stakeholders to participate on an equal footing, regardless of whether they are members of the ITU or not.
46. However, standards can have some limitations: for example, market forces may be insufficient to incentivize the development and adoption of standards that govern fundamental research and other transaction-distant systems and practices. In addition, standard-setting processes need to benefit from the key expertise of a wide range of stakeholders, including all countries, especially the Global South, civil society, and potentially affected communities, to ensure adequate consideration of critical aspects such as human rights.¹¹ Efforts are needed to ensure the necessary level of inclusivity.
47. There is a need and opportunity to build the capacity of civil society and academia not just on the standards processes themselves but also on digital technologies so that a wider pool of experts can effectively participate. Similarly, there is a need and opportunity to build the capacity of technical experts from governments across all levels, as well as for business and academia, on the potential human rights impacts of their standards work.
48. Concerted efforts among the technical community and all stakeholders might be needed to achieve such standards in practice, as well as a potential monitoring mechanism to assess adoption and implementation. Existing international treaties, national mandates, government procurement requirements, market incentives, and global harmonization pressures can contribute to the spread of standards once established – in that regard, standards are part of a broader array of governance levers and need to be deployed in conjunction with them.
49. **Including key stakeholders from the beginning of the process, including relying on their support in piloting the frameworks under development, brings legitimacy, helps demonstrate early results, and improves the adoption rate.** While endorsed broadly across the UN System, multistakeholder approaches often face significant challenges in effectively including civil society voices and end-users of AI systems.¹² Including key stakeholders from the beginning of the process or helping in piloting the frameworks under development brings legitimacy and improves the adoption rate. WHO's work on AI is linked to practical use cases like AI's use for cervical cancer screening, resulting in buy-in from Member States and stakeholders on specific use cases that also help address the demands of the Member States concerning certain diseases and health needs. The deployment of UNESCO's RAM in specific countries is accompanied by the engagement of the public, private, academic, and civil society organizations.
50. **In the discussions on compliance, monitoring, and enforcement for AI systems, it is essential to consider mechanisms such as audits, inspection, and certifications and to learn from the experiences of the UN System in addressing other complex global challenges.** For instance, within the UN System, the International Atomic Energy Association (IAEA) safeguards play a central role in preventing the proliferation of nuclear weapons through the independent verification of States' compliance with nuclear non-proliferation undertakings.

¹¹ Refer to A/HRC/53/42, available at [link](#)

¹² Refer to A/HRC/53/42, available at [link](#)

IAEA safeguards are embedded in legally binding agreements concluded between States and the IAEA. These agreements provide the legal basis for the implementation of safeguards.

51. In its vaccine prequalification process, WHO has provisions for comprehensively evaluating vaccines against international standards, which includes site inspection, targeted visits, and other mechanisms. Similarly, the International Civil Aviation Organization conducts on-site and off-site audits to assess States' oversight capabilities and the degree of States' compliance with relevant ICAO standards, and the International Maritime Organization (IMO) audits signatory states against IMO standards and develops recommendations and audit reports. ICAO, IMO, or WHO do not have legal enforcement capacity. However, as is highlighted in this study¹³ⁱⁱⁱ recommendations from these entities, based on audits, are considered essential, and many national legislations refer to these standards as a minimum requirement that must be adhered to.
52. There is a case for globally coordinated minimum safety standards and certifications of AI systems, especially for AI systems posing a high risk. Given the nuances of each sector, certification is expected to involve sector-specific approaches and inter-agency collaborations, especially for general-purpose AI systems that cut across sectors. For example, the [Secretary-General's Policy Brief on a New Agenda for Peace](#) calls for an agreement on a global framework regulating and strengthening oversight mechanisms for using data-driven technology, including AI, for counter-terrorism purposes.
53. **The implementation of normative instruments governing global public goods provides important lessons for transparency, accountability, and redress mechanisms, which are essential for AI governance efforts.** Currently, except for voluntary efforts to monitor AI incidents, there are no internationally coordinated avenues specifically aimed at redress mechanisms for AI harms once they have been reported and recorded.
54. However, the UN System has many examples of reporting mechanisms for tracking an issue area or avenue for receiving complaints, which are potentially relevant to AI harms, as well as examples from other areas, which provide lessons for AI governance.
 - Multiple human rights reporting and accountability mechanisms systematically address human rights issues. The [Universal Periodic Review, and treaty-based consideration of States Parties' reports with concluding observations](#) track developments in the human rights situation in countries. The [Human Rights Council Complaint Procedure and treaty-based monitoring and complaints procedures serve](#) as avenues for individuals and groups to initiate proceedings once relevant eligibility criteria have been met¹⁴. The Special Procedure Mechanisms of the Human Rights Council may also act on individual cases of allegations of human rights violations by sending communications to Member States, and the Working Group on Arbitrary Detention is mandated to consider individual complaints.
 - [UNODC's Convention against Illicit Traffic in Narcotic Drugs and Psychotropic Substances, 1988](#), involves compliance with both mandatory and voluntary reporting requirements and reporting to the [International Narcotics Control Board \(INCB\)](#). The Convention allows governments to update estimates and assessments in their reporting mid-cycle, and the Board can rapidly confirm such estimates or assessments.
 - ITU's elected [Radio Regulation Board](#) helps monitor compliance with complex radio regulations and sets out regulatory and technical procedures to establish compatible, non-interfered working of various radio services. It clarifies the application of specific regulations,

¹³ The referenced study and analysis present different examples. The US government's Federal Aviation Administration and Transportation Security Administration can prohibit a country's flight from entering US airspace if they aren't compliant with ICAO standards. Similarly, National regulatory agencies (NRAs) and national control laboratories (NCLs) play a vital role in WHO vaccine prequalification since they are responsible for regulatory oversight, testing and release of WHO-prequalified vaccines. In addition, the work of UNESCO on bio-ethics led to the establishment of the Bio-ethics committees in many countries, along with the development of hand books and guidance for users.

¹⁴ Refer to <https://www.ohchr.org/en/treaty-bodies/what-treaty-bodies-do>

considers reports of unresolved interference investigations carried out by the ITU's Radiocommunication Bureau, and formulates recommendations.

- The ILO supervisory mechanism provides for regular monitoring of the application of ratified Conventions as well as intermittent supervision of the effect given to non-ratified Conventions and Recommendations both by independent experts and tripartite political bodies ([Committee of Experts on the Application of Conventions and Recommendations](#) (CEACR) and [Conference Committee on the Application of Standards](#) (CAS)). Complaints-based mechanisms, notably [representations](#), [complaints](#), and a [special procedure](#) to examine complaints alleging violations of freedom of association complement this provision.
 - UNESCO's Recommendations include quadrennial reporting, where member states need to submit a report on the actions taken for the implementation of the standards.
55. Given the existing mechanisms and committees to address human rights concerns and provide redress, the capacity needs of these existing mechanisms to address human rights risks from AI could be supported and enhanced. While a committee that receives complaints related to AI at the international or regional level cannot substitute for the legal processes at the Member State level, it can provide additional capacity and mechanisms for redressal, especially after domestic remedies have been exhausted. Such redress mechanisms would be essential for cases like discriminatory harm emerging from AI, which has been identified in the survey as one of the top areas for the UN System to address.
56. **Tracking tools and data observatories can help with global alignment and sharing of information and best practices:** Most of the entities surveyed indicate that they are using tools and observatories for information sharing, transparency, peer learning, and knowledge exchange. Examples of such tools in areas that are not directly related to AI include [IOM's Displacement Tracking Matrix](#), which gathers and analyzes data to disseminate critical multi-layered information on the mobility, vulnerabilities, and needs of displaced and mobile populations, [UNDP's Electricity Access Mapping](#) that uses high-resolution satellite imagery in conjunction with household-derived data for more than 100 countries to identify electricity access gaps and enable the design of electrification strategies, ILO's [Forced Labour Observatory](#) that provides national data on protection, prevention, and remedies for victims of forced labour, and the United Nations Economic Commission for Latin America and the Caribbean's (ECLAC) [Gender Equality Observatory of Latin America and the Caribbean](#) that provides updated information on gender equality indicators. [ITU's ICT Regulatory tracker](#) helps decision-makers and regulators make sense of the rapid evolution of ICT regulation. UN-Habitat's model of urban observatories provides technical support and builds capacity with local and regional governments in data governance, a cornerstone of AI governance.
57. On AI specifically, the [AI for Good Neural Network](#) facilitates the sharing of information and best practices among a multi-stakeholder professional community of over 25,000 participants. UNIDIR's [AI Portal](#) is another example of a tracking tool that lists countries and their relevant AI policies. UNESCO also recently launched the [Global AI Ethics and Governance Observatory](#) with ITU and the Alan Turing Institute. This Observatory will publish the reports that are prepared with UNESCO's RAM applied to over 50 countries and the lessons learnt. An example of this is the [report](#) on Chile's deployment of the RAM, which resulted in concrete recommendations like updating data protection and cybersecurity legislation. It also includes use cases, and analytical papers from the network of experts (AI Experts without Borders, or Women4Ethical AI) that UNESCO has established to support the implementation of the Recommendation. [The Global Observatory of Urban AI](#) maps ethical AI initiatives in local governments. It has been developed in collaboration between UN-Habitat and the cities of Barcelona, London, Amsterdam, and the Cities Coalition for Digital Rights.
58. While many of these observatories are already providing information on AI governance efforts worldwide, it would be worth extending such repositories or observatories to include a comprehensive list of existing UN System laws and instruments that can be extended to AI or any under development for AI governance. In addition, an overview of best practices on legal capacities and the adoption and implementation of laws and instruments across countries would

aid in learnings from such best practices. It would also be helpful for the adoption and enforcement of such instruments. Results from the UN System's work on AI¹⁵ can further feed into such efforts.

59. **Capacity development is key to supporting the implementation of relevant instruments.** In this regard, the UN System has a twofold role: i) Develop technical guidance and tools that assist Member States in translating instruments into national/sub-national legislation, and ii) Provide capacity development support for legislative and enforcement capacities through development and capacity building programs, training and other avenues.
60. In addition to the examples highlighted in the earlier sections on AI and AI-related instruments under Focus area I, there are examples from UNEP's work in chemical control, which provides [detailed guidance](#) on translating instruments into national legislation. There are also examples like [WHO's guidance documents for Pre-qualification of Medical Products, including vaccines](#) (which include information for manufacturers, regulatory agencies, laboratories, and procurement agencies), [IAEA's tools for regulatory bodies on radiation sources](#), [ITU's guidance](#) on global use of radio-frequency spectrum and satellite orbits, ILO's [Codes of Practice](#) providing guidance on safety and health at work in specific economic sectors, [UNESCO's handbooks for bioethics](#) and its bio-ethics committees including their contributions to jurisprudence in certain countries and regions as well as UNOCT's Cyber and New Technologies Programme that has provided technical guidance on potential solutions for the use of AI in counter-terrorism and an exploratory assessment of the probability of misuse or abuse of AI for terrorist purposes.
61. In each case, the respective UN entities supplement these tools with capacity development engagements, including tailored capacity development programs, workshops, online courses, and training. These tools carry both normative and technical weight as they are widely regarded and adopted by Member States. The wide adoption reflects the demand from Member States for guidance on highly technical and sensitive areas like vaccines, chemical control, technology for countering terrorism, and nuclear safety, as well as their trust in the UN System for providing such guidance. UNOCT and UNICRI guidance is sought particularly by Member States' law enforcement bodies to ensure capacities and measures are effectively responding to evolving threats. WHO's work on AI in health shows technical demand from Member States and stakeholders on specific issues such as AI for drug discovery and the associated governance aspects of model explainability and interpretability. The World Bank often provides specific institutional capacity building in its digital development technical support and financing operations. These examples also necessitate further deliberation on the skills required internally in the UN System to address technical nuances in AI systems. The assembly of cross-functional teams that combine technical and social science knowledge, to guide responsible AI governance and the prevention and elimination of AI discrimination and bias, including gender bias, is also essential.

C. Focus area II - Institutional functions that inform global AI governance

62. This focus area relates to recent proposals for international governance institutions for AI inspired by institutional models of existing entities or processes like the Intergovernmental Panel on Climate Change (IPCC), the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), IAEA, and others. A suitable institutional governance model, a robust governance structure, and an effective norm-making process for AI governance need to deliver a variety of functions. These functions can span consensus building on norms, regulations, rules, policies, scientific consensus building and research, harmonizing standards and certifications, monitoring and reporting, building capacity for member states, and development-focused work, including the application of AI for SDGs and others. There is no 'one size fits all' approach that can be adopted across all these functions, and hence, multiple entities have been studied to understand their models and related functions and processes. In

¹⁵ The annual [United Nations Activities on Artificial Intelligence Report](#) partners with over 40 UN entities to highlight ongoing use cases and projects on AI run by the UN System, covering all 17 SDGs. The 2022 edition has reported nearly 300 projects.

addition, the strengths and limitations of the UN System’s models have also been reviewed to understand what could be most effectively adapted to the requirements of global AI governance.

63. The overview below is based on a combination of existing literature, UN entity expert interviews, and desk research on existing institutions. It has also factored in the institutional functions¹⁶ highlighted in the HLAB-AI’s report on AI. This paper focuses on the following institutional functions related to AI governance where currently some gaps are observed – i) Scientific consensus through technical and authoritative assessments and research; ii) Consensus-building and norm-setting around risks and opportunities; iii) Regulatory coordination, monitoring, and enforcement; iv) Addressing developmental needs of AI, including capacity building, technology diffusion, and dissemination. It is important to note that, for an effective AI governance regime, it is necessary that any institutional functions identified are closely interlinked and mutually reinforce each other. The structure of this focus area, as presented below, comprises an overview of an institutional model linked to a specific institutional function, followed by key takeaways for AI governance.

1. Scientific consensus through technical and authoritative assessments and research

64. Concerted and aligned global effort on any issue area is challenging if there isn’t empirical and scientific evidence to back it. Hence, scientific consensus building is almost a precursor to global cooperation. The UN System has different examples of scientific consensus building, such as the Scientific Assessment Panel assessing ozone layer depletion, IPCC, IPBES and others. While the IPCC was created for scientific consensus building, it was in response to growing environmental consciousness at an international level and some political momentum leading up to its creation, including the 1972 United Nations Conference on the Human Environment in Stockholm, the UNEP, and IMO facilitated Villach conference in 1985 and the pivotal ‘Our Common Future’ report from The World Commission on Environment and Development, led by Gro Harlem Brundtland in 1987. The political momentum and willingness built from the Montreal Protocol were also contributing factors.

IPCC’s membership comprises all Member States, and its model includes a bureau comprising 34 members selected by its panel. IPCC’s technical papers and assessment reports are globally recognized as the global observatory on climate-change information and cut across all essential elements, including the physical science elements of climate change and the adaptation and mitigation measures needed. The IPCC model also involves national focal points, which provide and update the list of national experts to help implement the IPCC’s work programme and coordinate between the IPCC and its member governments, thereby providing avenues for a decentralized approach.

IPBES is an intergovernmental body established by Member States to strengthen the science-policy interface for biodiversity and ecosystem services for the conservation and sustainable use of biodiversity, long-term human well-being and sustainable development. It was established in 2012 by 94 Governments, and UNEP provides secretariat services. It provides a mechanism to enable both the scientific and policy communities to synthesize, review, assess and critically evaluate relevant information and knowledge generated worldwide by governments, academia, scientific organisations, non-governmental organisations and indigenous communities. Like IPCC, it produces globally recognized assessment reports.

¹⁶ The HLAB-AI report outlines the following functions: i) Assess regularly the future directions and implications of AI ii) Reinforce interoperability of governance efforts emerging around the world and their grounding in international norms through a Global AI Governance Framework endorsed in a universal setting (UN) iii) Develop and harmonize standards, safety, and risk management frameworks iv) Facilitate development, deployment, and use of AI for economic and societal benefit through international multi-stakeholder cooperation v) Promote international collaboration on talent development, access to compute infrastructure, building of diverse high-quality datasets, responsible sharing of opensource models, and AI-enabled public goods for the SDGs vi) Monitor risks, report incidents, coordinate emergency response vii) Compliance and accountability based on norms

Key takeaways

65. **Scientific consensus driven by assessments from an independent, multidisciplinary, and multi-stakeholder group of experts brings legitimacy to an issue area and international alignment on actions needed.** Evidence-based decision-making and enhanced scientific consensus-building mechanisms relating to AI will be foundational to global governance efforts.
66. Concerns have been raised globally about how sources of information related to AI scientific research, safety, or important milestones are being dictated and controlled by a limited number of entities, primarily non-State actors like the private sector. For example, public announcements related to AI milestones, such as Artificial General Intelligence (AGI), can result in market hype and exacerbate problems of the AI divide. In addition, any efforts to develop scientific research on the capabilities and risks of frontier AI through multilateral or international efforts should ensure inclusive participation of all countries, especially those that are already lagging behind in AI development and research, with adequate support for capacity building related to scientific knowledge and tools.
2. **Consensus-building and norm-setting around risks and opportunities**
67. Consensus-building and norm-setting around risk and opportunities are core to many issue areas and have many facets, including harmonizing the needs and interests of stakeholders around norms and standards. Technical standards can play a key role in facilitating international governance through expert multi-stakeholder and multidisciplinary consensus. Technical standards have been effectively used to create and globally disseminate best practices in areas such as cybersecurity, environmental sustainability, and safety in numerous industries, including autonomous vehicles and nuclear energy. When it comes to emerging technologies, they are a path towards effective global solutions, laying down the ‘rules of the game’ for markets looking to develop and adopt emerging technologies, such as defining interoperability, facilitating trust through safe and reliable standards, and streamline the development of technologies and related products through processes and systems which are known to work.

As an SDO, ITU offers a contribution-driven and consensus-based environment open to all stakeholders. Work is often driven by the private sector with close involvement from Member States and academia. The model for the development of standards includes Study and Focus groups divided by technical areas such as speech quality, multimedia, security/infrastructure connectivity, and emerging technologies such as AI, Metaverse, Quantum etc. ITU facilitates such standard setting through a holistic lens, balancing the interests of the private sector with the needs of Member States and is working towards strengthened inclusion of human rights considerations. Conformity with ITU standards is voluntary unless mandated by national law or as part of the ITU-facilitated, internationally agreed Radio Regulations. A key policy, the [ITU-R/ITU-T/ISO/IEC Common Patent Policy, and related Guidelines](#), calls for intellectual property covered by ITU standards to be made available to all standards implementers on reasonable and non-discriminatory terms and conditions. ITU also works closely with other SDOs to strengthen and advance the voluntary consensus-based international standards system, including through the World Standards Cooperation (WSC) which is a high-level collaboration between the IEC, ISO and ITU.

UNESCO's model and decision-making process incorporates the findings of expert advisory bodies and intergovernmental institutions, in particular, the International Bioethics Committee (IBC), the World Commission on Scientific Knowledge and Technology (COMEST), and the Intergovernmental Bioethics Committee (IGBC). Especially in bioethics and science and technology ethics, these bodies facilitate consensus-building between experts and member states. Their joint efforts lead to informed standards and recommendations that actively shape ethical (best) practices globally. In 2018, COMEST produced a report on the “Ethics of Artificial Intelligence” that led member states to mandate UNESCO to develop the first global gold standard on AI ethics through its recommendation on AI. More recently, COMEST produced a report on the ethics of neurotechnology, and this was the basis for member states to request UNESCO to develop a normative instrument in this area.

ILO's adoption of international labour standards for social justice and decent work is typically preceded by years of research by the secretariat, technical and expert meetings in which stakeholders (notably trade unions and employer organizations) participate on an equal footing, a law and practice report, and two rounds of tripartite consultations led by national labour administrations.

Under the auspices of UNEP, the Global Framework on Chemicals, adopted by the Fifth International Conference on Chemicals Management (ICCM) in Bonn in 2023, sets concrete targets and guidelines for key sectors across the entire lifecycle of chemicals. A unique international negotiating process – where representatives from governments, the private sector, Non-Governmental Organizations, intergovernmental organizations, youth, and academia participated at the same level – resulted in the historic decision to establish the “Global Framework on Chemicals – For a planet free of harm from chemicals and waste.”

The Montreal Protocol, finalized in 1987, is a global agreement to protect the stratospheric ozone layer by phasing out the production and consumption of ozone-depleting substances (ODS). The Montreal Protocol has proven to be innovative and successful, and it is the first treaty to achieve universal ratification by all countries in the world. It has spurred global investment in alternative technologies and placed the ozone layer on a path to repair. The success of the Montreal Protocol was the result of an unprecedented level of cooperation by the international community, and collaboration between public and private sectors.

Key takeaways

68. **For standards, especially in areas that could be considered critical infrastructure or for technologies that have significant potential for harm to the fundamental rights of citizens, the involvement of all stakeholders, including governments, in the standard-setting process is particularly important to ensure that the necessary guardrails are appropriately implemented.** This multistakeholder involvement is especially important where such technology crosses borders, given the fragmented responses that often occur due to discrepancies between domestic frameworks.

69. **On AI specifically, technical standards can influence the development and deployment of specific AI systems through product specifications, including by introducing requirements for explainability, robustness, and fail-safe design.** They can also affect the larger context in which AI is researched, developed, and deployed through process specifications. For example, standard developments for AI are underway at ITU, the Institute of Electrical and Electronics Engineers (IEEE), the International Organization for Standardization, and other standards bodies. Many UN System entities partner and participate in the standard-setting process internally and externally. In addition, with evolving technology, technical standards may require further refinement depending on sectors and use cases, necessitating agile development processes that can keep pace with technological developments in AI.

3. Regulatory coordination, monitoring, and enforcement

70. UN System entities, including ICAO, IMO, and IAEA, provide examples of internationally coordinated models across Member States with regulatory and policy functions. These entities demonstrate the UN System's expertise and background in coordinating highly complex global systems like global aviation or maritime systems, which are the backbone of globalization and international trade and transport systems as we know them today. In addition, the evolution of focus areas within these entities also shows their response to the changing realities of their specific sectors. For instance, ICAO's focus on aviation sustainability and security has evolved in response to climate change and the need to secure aviation systems in response to unlawful interference and emerging threats to international civil aviation.

The establishment of ICAO predates even the UN and was done to fill a gap in global cooperation on international civil aviation post World War II - where most countries had their own rules (or

limited rules), leading to inefficiencies in travel and transport as well as the risk of not being able to assess safety and security across minimum standards. ICAO's model is inter-governmental, and it works with Member States. ICAO develops and maintains international Standards and Recommended Practices (SARPs) and Procedures for Air Navigation (PANS), which are at the heart of global aviation safety and efficiency. ICAO operates under the Chicago Convention and comprises an Assembly, the supreme body with each Member State carrying one vote, and a resident Council, a permanent body elected by and responsible to the Assembly. One of the mandatory functions of the Council is to adopt international standards and recommended practices by a two-thirds majority. Member States have the obligation to comply with international standards and recommended practices unless they register disapproval or notify the ICAO of differences. While ICAO doesn't have enforcement capacity, adherence to its standards is widely accepted as the international norm and translated into national civil aviation legislation. It's also important to note that while the private sector can provide inputs, it is not part of the voting process for finalizing standards, even as it must comply with them.

IAEA is an institution with provisions for providing its report to the General Assembly and reporting noncompliance to the Security Council, a UN body with enforcement functions. In addition, IAEA's board involves representation of Member States from across all regions but factors in their capabilities in nuclear science and technology. The Board mostly makes decisions based on consensus, but voting by majority or 2/3rds majority is a possibility.

IMO's model comprises Member States with voting rights and Associate Member States without full voting rights. Its structure consists of three tiers – the highest-level Assembly, an executive council, and various technical committees and sub-committees. Committees draw stakeholders' input to propose or amend standards, which are then sent to the Council or Assembly for adoption.

Key takeaways

71. **While drawing lessons related to auditing and monitoring procedures from the models of ICAO, IMO, or IAEA, it is essential to distinguish between these specific subject areas and the decentralized nature of AI systems.** Firstly, there is significant international consensus on safety and security related to aviation, maritime operations, and nuclear energy, and this consensus-building has involved several decades of global coordination. Safety and security are 'preliminary and primary' considerations in these systems before they can be operational or be placed on the market – planes are not allowed to fly unless they meet all the safety criteria. Regarding AI, global political consensus on its safety aspects is gaining momentum. Still, countries have different understandings of safety thresholds and regulatory approaches, and new AI models and systems are being developed and released into the market at record rates. Secondly, the ex-ante assessments of planes/ships/fissile material are different from ex-ante assessments of AI systems, which are more diffused, decentralized, and, in some cases, open source.
72. **However, there are still valuable lessons learned from auditing and monitoring procedures** (such as the number of staff required, their skills, the auditing checklist, post-audit remedial measures, and enforcement measures, etc.), how the Member States discuss and agree on standards (and the associated voting mechanism), how these priorities shift and evolve, how the institutional model allows adapting to these changes (e.g., evolution in standards related to airplane noise over the years, the enhanced focus on sustainability standards for aviation, or the more recent evolving focus on peaceful nuclear fusion energy reactors and other similar examples) as well as the time needed to introduce a standard, test against it and have it adopted internationally, a process which could take several years for some standards.
73. Further deliberations are also needed on the role of the UN System in addressing issues such as existential AI risks and potential complex AI-driven attacks on State assets by other State or non-State actors.

4. Addressing the developmental needs of AI, including capacity building, technological equity, and technology diffusion & dissemination

74. Given the tremendous opportunities that AI presents in domains like health, climate, new material design and discovery, and other vital areas, there is a need for development-focused use cases of AI, especially for developing countries, in a wide variety of areas including climate modelling, drug discovery, and improving access to healthcare information. Given the gaps observed in the global AI space, the emphasis on development opportunities provided by AI is critical. UNCTAD's [Technology and Innovation Report 2023](#) reveals that less than five countries today possess nearly half of all AI-related publications and patents. This AI divide leaves many countries, particularly least developed countries, lagging and often relegated to the role of data providers and subjected to extractive data practices.
75. To understand existing institutional models focused on developmental aspects and needs, the models of the Global Fund, IAEA, the European Organization for Nuclear Research (CERN), and UNDP were studied – keeping in mind the nuanced/specific needs of different countries that might be at varying levels of AI maturity. While CERN plays a broader role in scientific research, its model also highlights the importance of pooling resources for cost and resource-intensive research and innovation.

CERN¹⁷ is an intergovernmental organization with membership from 23 European countries and several associate members, established through a resolution by UNESCO in 1950. One of the key success factors behind CERN has been its neutral and open-science approach¹⁸, including its [knowledge transfer activities](#), which range from the public to academia to industry and start-ups. The birth of the World Wide Web, as we know it today, happened at CERN. CERN's model also outlines the importance of pooling resources. With an annual cost exceeding a billion CHF (Swiss Francs)¹⁹, it would be impossible to cover these costs for one or just a handful of countries without international collaboration.

The Global Fund is a public-private partnership established in 2002 to combat AIDS, tuberculosis, and malaria in low- and middle-income countries. The history of the Fund can be traced back to 1999 when WHO called for a "Massive Attack on Diseases of Poverty," followed by the UNGA endorsing the creation of a fund to fight HIV/AIDs in 2001. The Global Fund's model comprises a 22-member board representing governments, foundations, and civil society organizations. The voting rights of the board for strategic matters are weighted in proportion to their level of contribution to the fund. The programs of the Global Fund are managed at a country level by country coordinating mechanisms (CCMs) and implementing partners comprised of NGOs/Civil society. This diffused governance approach allows local/in-country partners to develop tailored programs that reflect the realities of the respective countries across prevention, treatment, care, and support aspects of fighting HIV, tuberculosis, and malaria. The Fund's approach also involves partnerships with stakeholders for co-financing, technical, procurement, and philanthropic partnerships. In addition, the Global Fund also follows a results-based financing model, which provides accountability and transparency and sets up an incentive structure of funding based on demonstrated progress.

¹⁷ CERN is another example of the UN's role in facilitating international cooperation and research on a complex topic, recognizing the need for neutral scientific research with pooled resources. UNESCO's General Conference in 1950 adopted a resolution calling for a centre on particle physics research to propel neutral scientific research for the benefit of humanity. Subsequently, UNESCO also helped facilitate intergovernmental meetings and negotiations with Member States and scientists and lent support to developing the legal framework for CERN, including the CERN convention signed in 1953, leading up to the eventual establishment of CERN in 1954.

¹⁸ In the [CERN convention](#), Article 11: Publications "CERN shall facilitate the publication of all results obtained at CERN in scientific journals or in other publications accessible to the world scientific community" and Article 12: Technical Information "CERN shall make available to the world scientific community, upon request, technical information concerning equipment and methods used at CERN" demonstrate the open-science approach

¹⁹ Based on CERN's annual report of 2022, accessed at the following [link](#)

UNDP's model reiterates the importance of developing tailored programs in line with the needs of a country or a thematic area. Its governance structure at the apex level consists of a board of 36 Member States, which the UN General Assembly selects for a 3-year term. UNDP's decisions are consensus-driven, but each member state has one vote in the case of voting. UNDP's model also provides examples of decentralization and specific focus – it consists of country offices and thematic centres focused on specific policy issue areas like Climate, Food security, Gender, Human development, Urban development, etc. In addition to [UNDP's AI Readiness Assessment Tool](#), highlighted earlier in the report, UNDP's AI work is also focused on developing concrete AI applications either at the Member State level or a thematic area or sometimes combining both. [The eMonitor+](#) system relies on AI models that help identify and analyze online content harmful to information integrity, and the fact-checking tool [iVerify](#) is used to identify false information and mitigate its spread. Furthermore, in Rwanda, UNDP's Accelerator Labs partnered with the Ministry of ICT and Innovation to deploy five intelligent [anti-epidemic robots](#) in two COVID treatment centres and at Kigali International Airport.

Key takeaways

76. **AI research, development, and training are cost and resource-intensive and would require the pooling of resources to spur responsible innovation in AI.** The cost and resource-intensive nature of AI is also well-reflected in [Stanford HAI's AI Index](#), which showcases that in 2022, there were 32 significant industry-produced machine learning models compared to just three produced by academia. Even as the efficacy of large language models is still evolving, training a large language model with 65-170 billion parameters could cost between \$2.4 to 4 million^{vii}.
77. For any public-focused AI system or model built for the larger benefit of humanity, cost is a definitive factor, especially if it focuses on the development of models, and as such, similar to CERN's model, international cooperation is critical to help pool resources and ensure that such research is both neutral and distributed under the spirit of open-science, especially for use cases related to AI for SDGs like health and climate.
78. In addition, the analysis of CERN's model and its open innovation approach also necessitates further discussions on the use of open-source AI Systems. There is an opportunity for the UN System to consider further its role in ongoing international discussions on open-source AI, including research on its governance aspects. Open-source AI should not be seen as a silver bullet, but it is an essential consideration in the equity, access, and power concentration debate. Such systems can also potentially provide avenues for AI development for stakeholders with limited resources to develop AI systems from the ground up.
79. **The UN's experience in development work, including capacity-building support to broader strategy work like the adoption of National AI strategies, can provide avenues for tailored development programs catering to the specific needs of stakeholders and Member States that take a holistic view of the AI value chain.** Such capacity development can support AI development that is grounded in fairness, gender equality, reliability, safety, interpretability and accountability. As is reflected in practically all National AI strategies developed in the past few years, governance is an important consideration. It has a dedicated chapter or section in most of these strategies. However, it is considered a part of a broader umbrella of AI opportunities and needs in a country, not a stand-alone objective. Developing countries, especially, see AI as a leap-frogging opportunity, and they must balance available resources to leverage the opportunities presented by AI while ensuring its effective governance. The development and capacity-related needs of a country to leverage AI could include AI's compute/hardware requirements, access to data, the existing AI ecosystem and skills in a country, assessment of sectors that can benefit from AI, and how the country can continue to leverage AI for domestic and global value addition as well as better public service delivery. A holistic approach that views

AI-related development efforts in a country in conjunction with any existing UN efforts on digital infrastructure and digital transformation would be beneficial.

80. **AI requires a ‘dual-action’ approach to address safety risks as well as the development opportunities presented by the technology, especially in developing countries.** Learnings for this can be found in the treaties and related international organizations against weapons of mass destruction, where it is recognized that there are substantial developmental opportunities from physics, chemistry, and biology, as well as potential risks of accidents and deliberate misuse. Instruments such as the Biological Weapons Convention, Chemical Weapons Convention, and Nuclear Non-Proliferation Treaty contain provisions to manage this balance and ensure that security concerns do not hamper beneficial applications of science and technology. Recognizing such benefits, for example, in civil nuclear energy, can also help bring a much broader set of countries to the table. This also speaks to the dual nature of an issue area being a public good in addition to being an issue of national security.
81. This tension is also observed in the case of environmental treaties and instruments. Efforts to reduce emissions and associated targets must be balanced with the economic and social development needs of countries, which need to be met as progress in clean energy and sustainable production is achieved. Similarly, for AI, there is a need to strike a balance between innovation for development and governance while being able to provide resources for both adequately. The UN System, through its agencies and their specialized mandates, can help address this duality by providing dedicated support and guidance on ethically developed and governed AI systems that cater to developmental needs in areas like access to justice, health, and energy efficiency.

5. Coordination across the UN system for cross-cutting issue areas

82. The analysis of institutional models also reaffirms that for a cross-cutting and transversal topic like AI, the various governance functions are expected to be distributed across multiple entities. This section highlights examples of UN System-wide coordination and collaboration in cross-cutting and complex issue areas. The UN System’s efforts in areas such as cybersecurity provide lessons for effectively leveraging the complementary strengths and mandates of multiple entities to address the cross-cutting and multifaceted nature of technology.
83. The ITU supports Member States by enhancing their technical cybersecurity capabilities, including establishing Member State level [Computer Incident Response Teams CIRTs](#), conducting cyber drills for better coordination, sharing best practices and developing skills, and providing a platform for developing technical standards. Issues related to national security concerning cyber threats from States or non-State actors are addressed in forums like the [GGE](#) or [OEWG](#) within the [First Committee](#)²⁰. Where UNODC addresses cybercrime, and the HRC/OHCHR typically handles human rights implications of cyber activity, UNODA addresses the State use of information and communications technologies in the international peace and security context. Regular sharing of experiences between Member States and the UN Secretariat, as well as among the different UN entities, helps ensure that the relevant forums are utilized for specific aspects of the complex topic of cybersecurity.
84. Another example is the [Global Counter-Terrorism Coordination Compact](#), which ensures coordination and cohesion across 46 UN and non-UN entities as a platform to strengthen joint UN action in support of Member States to deliver on the [United Nations Global Counter-](#)

²⁰ GGE – Group of Governmental Experts. In GA resolution 73/266, the Secretary-General was requested to establish a Group of Governmental Experts on Advancing responsible State behaviour in cyberspace in the context of international security.

OEWG - Through resolution 73/27, the General Assembly established an Open-Ended Working Group (OEWG), in which all UN Member States are invited to participate.

First Committee - The First Committee deals with disarmament, global challenges and threats to peace that affect the international community and seeks out solutions to the challenges in the international security regime.

[Terrorism Strategy](#) while reducing duplication of UN efforts, including in the field of new technologies such as AI and cybersecurity.

Focus Area III - Learnings from existing governance structures, inclusive normative processes and agile & anticipatory approaches

85. The analysis in this focus area is based on select examples of governance structures and normative processes taken from the results of the survey and interview responses that demonstrate the challenges and gaps observed in global AI governance. It includes aspects such as the disproportionately large role of the private sector and the related concentration of market power, as well as the need to balance their innovation capabilities with safety aspects, legal liability, the growing AI divide, the need for inclusivity, the proliferation of principles and guidelines related to AI but accompanied by lack of enforcement, the need to regularly update standards (technical or safety) in line with rapid AI development and others.
86. There is also a need to “future-proof” regulation. In this regard, human rights experts have highlighted the added value of risk and opportunity assessments that also duly incorporate human rights considerations and the adoption of the precautionary principle when it comes to developing new technologies and AI. The data shows that the UN has extensive experience in addressing several of these challenges, both from its work on AI and other areas. Further reflection is needed on important factors like agility in the normative process and further strengthening engagement with the private sector.

1. Learnings from existing governance structures and normative processes

87. **The UN System champions inclusivity in its norm-making processes** by seeking inputs from a wide variety of stakeholders, which could help address the challenges of inclusive AI design and development. The first instance of such convening in the digital space was the two phases of the World Summit of the Information Society (WSIS), which was supported by more than a hundred Heads of States and world leaders and attracted thousands of representatives from all stakeholder groups. The Outcome Documents, which include the Geneva Plan of Action and the Tunis Agenda (where the concept of multistakeholder-ism was elaborated), continue to drive digital discussions and work programmes across UN agencies and multistakeholder communities.
88. The [AI for Good platform](#), hosted by the ITU in collaboration with over 40 UN agencies and a 20,000+ AI community, continues to play a critical role in hosting global multi-stakeholder conversations on AI and identifying common ground and solutions. The [Commission on Science and Technology for Development](#) and the [Multistakeholder Forum on Science, Technology and Innovation for the Sustainable Development Goals](#) are also important multistakeholder platforms for advancing understanding on science and technology.
89. Other select examples from the survey results and UN entities expert interviews include - ITU’s diverse membership, which comprises over 1000 private sector, civil society, and academia members, and the role of different stakeholders in technical standards development along with experts and Member State representatives, UNAIDS’ governance model involving board representation for HIV patients so they are directly involved in decision-making, and UNICEF’s [U-Report platform](#) for engaging the voice of youth in policy discussions as well as the [Youth Foresight Fellowship programme](#). ILO’s model involves both trade unions and employers’ organizations in their Standard-Setting Conference and executive committees for shared decision-making and governance on an equal footing with governments. UNESCO’s Recommendation on the Ethics of AI was developed by multidisciplinary experts from all over the world, received over 50,000 diverse inputs in an online consultation, and is supported by multiple networks, including the private sector and civil society, for implementation.
90. The success of all the UN human rights mechanisms is rooted in meaningful, consistent, and inclusive stakeholder engagement, including at the Human Rights Council and Treaty Bodies, in OHCHR and Special Procedures reports as well as General Comments and Recommendations of Treaty Bodies, the UN Forum on Business and Human Rights and the Forum on Minority Issues. Stronger attention to public participation and participatory process in AI governance can

ensure inclusion and human rights-based approaches, with special consideration to vulnerable groups such as women and girls, children and youth, people with disability, indigenous groups, and forcibly displaced people. It is also crucial to include developing countries, redressing not only past imbalances in their participation in global governance processes but also as a means for global AI governance to effectively include the development perspective.

91. **Engaging the private sector is key for global AI governance efforts:** The UN System’s private sector engagement was analyzed across a range of sectors (ICTs, aviation, health, disaster risk management, climate, counter-terrorism, and others) and modalities – i.e., engaging the private sector in the standard-setting or harmonization process, seeking inputs in shaping and developing instruments or in some cases, in *ex-ante* or prequalification assessments. Some of these examples include:

- ITU is one of the few UN entities with multi-stakeholder membership that includes industry, universities, research institutes, and international and regional organizations. The private sector plays an important role in decision-making processes at ITU, including standards development, where decisions are primarily made through a consensus-driven process involving different stakeholder groups.
- For the implementation of the UNESCO Recommendation, a Business Council has been launched, considering that it is the business sector that is developing the largest share of AI technologies. This Council includes Microsoft, Telefonica, Mastercard, Salesforce, Lenovo, Innit, and others.
- UNIDO’s Global Alliance on AI for Industry and Manufacturing ([AIM Global](#)) brings together stakeholders from the private sector along with those from the public sector, academia, and civil society to ensure access to AI technologies and their responsible and fair use.
- OHCHR’s B-Tech project has developed a range of policy papers and recommendations that have been developed with inputs from and are applicable to the technology sector, including analysis and recommendations regarding generative AI²¹.

92. Given the current global AI context, some considerations for the UN System for further strengthening engagement with the private sector are shared below:

- **Ensuring neutrality in the normative and operational process:** The UN System has guidelines in place for engaging with the private sector and other non-State actors in normative and operational processes. In the case of WHO, this is observed in their Framework of Engagement with non-State Actors ([FENSA](#)) and also their ‘two-tiered’ normative process while developing their regulatory considerations for AI in health. This process gathers important technical inputs from the private sector (through advisory boards) but filters these inputs through a group of experts and also factors in inputs from Member States to ensure alignment with standards and policies that are being developed without bias, conflict of interest, or risk of regulatory capture. In addition, the final decision-making process is led by Member States without voting/decision-making from the private sector.
- **Bringing accountability by supporting and advocating for reporting mechanisms across voluntary and mandatory requirements:** The UN System has many examples of developing or supporting standardized reporting, such as the [UN Global Compact](#) and [reporting from businesses on UN Guiding Principles on Business and Human Rights](#). While the UN Guiding Principles (UNGPs) have helped make significant progress on business and human rights, there is a long way to go, as reporting is voluntary. Similarly, in the field of AI, reporting from the private sector is voluntary in nature. In addition, while some leading technology companies have

²¹ See Opening Statement of the High Commissioner for Human Rights at the Generative AI and Human Rights Summit on 30 November 2023, <https://www.ohchr.org/en/statements-and-speeches/2023/11/turk-calls-attentive-governance-artificial-intelligence-risks> and OHCHR B-Tech Note regarding the United Nations' plan towards AI Governance at the UN General Assembly 78, September 2023, <https://www.ohchr.org/sites/default/files/documents/issues/business/b-tech/unga-b-tech-gen-ai-governance-note.pdf>

published their transparency reports, there are no standardized metrics in place for reporting on AI responsibility and safety. Standardized disclosures related to AI safety and governance across the AI life cycle can be deliberated while recognizing that such approaches should be differentiated depending on the size of companies - large technology companies have more policy and legal resources than small and medium enterprises. Environmental, Social, and Governance (ESG) led investments in AI companies can further support such efforts, which factor in human rights and responsible AI metrics in investment decisions. Efforts like [UN Principles for Responsible Investment](#) can be extended for a similar discussion on disclosure and investments, considering the nuances of AI and other frontier technologies.

- **Providing capacity development and detailed technical guidance can help develop a trusted platform for assessments of requirements.** When it comes to a trusted body for pre-qualification and *ex-ante* requirements for AI systems, the UN System has multiple examples to draw from its work in different sectors. Member States deliver the ultimate regulatory function, but the UN System plays a key role in providing guidance. For AI, [WHO's Generating Evidence for Artificial Intelligence Based Medical Devices: A Framework for Training Validation and Evaluation](#) delivers guidance on deployment and pre- and post-market surveillance of AI in software as a medical device. Other examples of the UN's trusted role in *ex-ante* assessment include WHO's work on the [Prequalification of Medical Products](#), which is supported by an array of instruments (the legally binding [International Health Regulations](#), WHO's Prequalification Programme Manual, Prequalification Standard Operating Procedures, and others). Similar examples can be found in [WFP's Food Quality and Safety](#), [ICAO's guidance on aircraft airworthiness assessment](#), [ITU's G5 benchmark for fifth generation digital collaborative regulation](#), and [IAEA's safety standards on the safety of nuclear power plants](#). UNESCO's Ethical Impact Assessment and UNDP's AI Readiness Assessment are also useful *ex-ante* assessment tools for AI systems before they are procured and deployed. These examples show the depth of technical support from the UN System to address the need for prequalifying assessments.

93. **Flexible and dynamic decision-making is essential in specific contexts, especially related to emergency/safety:** Agility in the decision-making process, rapid allocation or release of funds, or emergency responses are important lessons for AI safety, particularly for cases related to malicious use of AI or extreme AI incidents (e.g., large scale AI-driven attack on critical infrastructure). Research sheds light on relevant examples in this context, such as the UNHCR's experience in emergency fund and resource mobilization and the continuous need to deploy resources in light of the unpredictable nature of crisis and conflicts. The Universal Postal Union (UPU) structure allows for lower-level regulations of the Universal Postal Convention to be amended bi-annually to keep up with the fast-paced developments in the postal sector. ITU has an Alternative Approval Process (AAP), a fast-track approval procedure for technical standards that allows for a standard to be approved in as little as five weeks.

2. Learnings from agile and anticipatory governance mechanisms within the UN system

94. The rapid evolution of AI technology requires agile governance approaches. It also highlights the need for anticipatory mechanisms, i.e., to predict unforeseen risks and develop an action plan that can be triggered if the risk materializes. This section has examples and lessons from the UN System's work on agile and anticipatory governance approaches, aligned with [UN 2.0](#)²², which envisions accelerating towards agile, diverse, responsive, and impactful UN entities.

(a) Agile and reflexive governance

95. Agile governance, as a concept, borrows from agile software development and involves tenets of fluidity, adaptiveness, and flexibility in governance design, introducing mechanisms and

²² "UN 2.0" encapsulates the Secretary-General's vision of a modern UN family, rejuvenated by a forward-thinking culture and empowered by cutting-edge skills for the twenty-first century – to turbocharge UN's support to people and planet. The approach will involve powerful fusion of innovation, data, digital, foresight and behavioural science skills and culture – a dynamic combination, called the "Quintet of Change".

processes that can be updated in an expedited manner to stay abreast with rapid changes in technology or socio-political-economic levers. Reflexive governance allows for increased participatory processes, incorporating feedback on technology impacts through policy iterations and fostering the capacity of governments to mitigate and respond to risks faster. When it comes to agile governance of emerging technologies, there has been ongoing work globally. Analysis of the survey results has yielded the following broad categories that are relevant to the governance of AI:

- Agility in policy design or instruments (e.g., the ability to add optional protocols or add annexes or update standards in an expedited manner instead of modifying the whole instrument)
- Agility in normative processes and governance mechanisms – i.e., both in developing/modifying instruments or also governance mechanisms allowing for expedited decision-making when relevant
- Agile mechanisms for expedited response in case of emergencies (e.g., disease outbreak or climate/humanitarian crisis)

96. Based on the above, some examples of agile governance within the UN System include:

- UNDP's global network of Accelerator Labs focuses on applying collective intelligence to accelerate development in Member States in specific issue areas, experimenting in testbeds/sandboxes informing national policies, and developing innovative ways of engaging citizens/community engagement at a large scale for policy formulation while capacitating member states and local ecosystems of global south with Open R&D capabilities.
- The WFP's AI Sandbox (under development) aims to provide streamlined access to data and compute resources to experiment with AI solutions and offer an expedited understanding of effective AI solutions.
- ILO has existing examples for regular and expedited updates to amendments. The [Maritime Labour Convention](#) has a mechanism for regular updates by a Special Tripartite Committee and an "opt-out" mechanism for adherence to subsequent amendments, contrasting the time-consuming ratification process. [R194 - List of Occupational Diseases Recommendation, 2002 \(No. 194\)](#) has an accelerated process for updating the list of diseases, achieved through a tripartite meeting of experts with representation from all relevant stakeholder groups.
- UNHCR's policy-making approach enables the expedited adoption of 'interim' guidance when there is a need for an institutional approach while system-wide standards are under development. This approach has been used to introduce/adapt requirements to fast-moving operations or emergencies and to regulate the dissemination of anonymized microdata.

Key takeaways

97. While reviewing these examples, it's important to acknowledge that agility at a project or a smaller-scale programme level has a different magnitude than agility at an entity or institutional level or even in a normative process involving all Member States. The technical complexity of the work, the diversity of views among decision-makers, and also their varying levels of technology adoption maturity call for reflection on the true agility of normative processes, which might necessitate regional approaches or industry/sector-specific approaches. An important next step for the UN System would be to consider ways to integrate agility in its current processes to keep pace with changes in AI and emerging technologies.

(b) Anticipatory governance and foresight-based research

98. Applying foresight in AI is critical for navigating the complex interplay of technology, society, and the environment, crafting adaptive strategies that align with ethical, inclusive, and sustainable values. It serves as an indispensable tool, not only for forestalling unintended consequences but also for uncovering implicit assumptions, ensuring that the development of AI systems comprehends their full potential impact. Moreover, foresight fosters a nuanced understanding of AI's interaction with other emerging technologies, promotes a forward-looking

view, and underscores the significance of human agency in directing AI technologies to enhance societal well-being without exacerbating inequalities.

99. Entities have shared several examples of anticipatory governance, especially in the food, health, climate, conflict, and humanitarian relief and response space – all sectors that have high volatility and risk owing to multi-fold crises like climate change and erratic weather patterns, natural disasters, conflict, unforeseen disease outbreaks, calling for both anticipatory response planning as well as emergency resource mobilization. However, with AI and emerging technologies impacting practically all sectors, specialized entities like ICAO have also been working on anticipatory governance. Some examples of anticipatory governance from within the UN System include:
- ITU's pre-standards work for alignment on an emerging technology/issue area, called "Focus groups," enhances formal standards development processes by swiftly developing specifications in emerging technology areas, addressing emerging industry needs. Some recent areas of focus include Quantum computing, Metaverse, AI for Health, Precision Agriculture and IoTs, Autonomous and Assisted Driving, and many others. ITU's 'G5 Benchmark' addresses the dynamic and intricate nature of the ICT landscape by providing metrics to assess gaps, suggesting adaptable roadmaps amid evolving regulatory environments, monitoring progress, and proposing solutions for overcoming challenges in achieving the Sustainable Development Goals.
 - WFP's work on anticipating food crisis can be exacerbated by unforeseen risks related to climate and natural disasters, conflict, or economic and political shocks. In this regard, the WFP has worked on [common principles](#) to address challenges related to anticipatory action in foreseeing food crises. A key finding here was the lack of uniform methods across agencies to assess the efficacy of anticipatory action. To overcome this lack of uniformity, the principles developed included agreeing on common outcome indicators and alignment on acute food insecurity projections, which are evidence-based and consensus-based. In addition, it's also recommended that these projections on food and the indicators be aligned with forward-looking indicators like climate and economic situation. In this regard, platforms such as [Anticipation Hub](#) are also gaining traction, which outlines that the success of anticipatory governance is contingent upon clearly defining the roles of all stakeholders, common forecasting and threshold triggers, and financing for action.
 - ECLAC – has done work on [foresight for development](#) spanning ten countries in the region, which includes training more than 1,400 government officials and members of civil society and academia in foresight concepts and tools. The methodology deployed by ECLAC involves scenario building for multiple futures based on strategic variables (like competitiveness, investment, economic and social infrastructure, security and justice, etc.) that enable organizations to define risks, opportunities, or threats that may develop in the future. Hypotheses are developed based on the changes in variables and the interaction between them over a 10-, 20-, or 30-year time horizon.
 - ICAO's Global Air Navigation Plan explores the possibility of using emerging technologies in aviation safety and security and sets up timelines for their implementation, considering a differentiated rate of development among member States. The anticipated technologies have all been partially developed by some Member States and are ranked in categories based on their implementation timeline.
 - The UNHCR has several case studies that engage with anticipatory governance. For example, [anticipatory governance in refugee protection from Macedonia](#), the [Sahel Predictive Analytics project](#), which brought together a global consortium of leading academic institutions specializing in predictive analytics and strategic foresight approaches to work on the nexus between climate change and other megatrends as factors contributing to new or exacerbated vulnerabilities, and a [Brazil Predictive Analytics](#) project for contingency planning scenarios. UNHCR is also scoping a global early warning system relating to forced displacement based on Inter-Agency Standing Committee (IASC) contingency planning.

- UNODA, supported by the EU, engages young AI practitioners, civil society, academia, and professional associations on responsible AI and the mitigation of the risks that misuse of civilian AI presents for peace and security - as a form of "upstream arms control." This work builds a civilian capacity to support anticipatory governance, which is critical in a field like AI, where practitioners must be engaged for governance efforts to be appropriately agile and meaningful.
- UNFCCC's Resilience Frontiers initiative embodies the essence of strategic foresight coupled with human agency. Engaging a wide array of international stakeholders and experts from diverse disciplines, it directly confronts the challenges of short-termism and linear thinking, structuring its approach around eight transformative pathways. These pathways focus on leveraging frontier technology in a way that enables humanity and the natural world to thrive in harmony, fostering a sustainable, equitable, and resilient world for all. Unlike theoretical frameworks, these pathways offer actionable plans designed to catalyze an irreversible shift toward a world characterized by permanent resilience. This collaborative effort exemplifies a model of anticipatory governance that values multidisciplinary collaboration, innovative thinking, and transformational future-making.
- The work of the ethics committees of UNESCO (the Committee on the Ethics of Science and Technology and the International Bioethics Committee) is anticipatory in nature. For many decades, it has identified emerging trends in the technological innovation ladder and informed Member States and stakeholders. There are many technological developments that are usually scanned and debated. Examples include the human genome, the ethics of climate engineering or the Internet of Things, and the work on artificial intelligence.

Key takeaways

100. Developing an effective anticipatory governance system for AI requires different stakeholders to be aligned on 'outcome indicators' related to safety, well-being, prosperity, etc. when building forecasting models. Global alignment on all these indicators is expected to be challenging, and a sectoral or regional approach is recommended. In addition, it is also recommended to conduct anticipatory governance from a holistic perspective of the AI value chain (ranging from environmental and energy considerations, hardware for compute capacity to cloud services to data to model design, development, training, and deployment and beyond). For instance, further geopolitical rifts impacting trade curbs on essential minerals and metals or semiconductor chip shortages could affect AI's hardware and compute requirements, which depend on semiconductors fabricated using these essential minerals and metals. These rifts could disrupt the development of future AI applications and exacerbate the equity gap in AI.

D. Recommendations

101. Based on the research and analysis presented in this paper, the following general recommendations are presented to the UN System, as well as the HLAB-AI, for consideration:

1a. Global AI governance efforts in the UN System should be anchored in international law, including the UN Charter, International Human Rights Law, and other agreed international commitments such as the 2030 Agenda for Sustainable Development. To adequately cater to the specific requirements and economic, social, and environmental priorities, UN instruments and frameworks providing sector-specific guidance are key to these governance efforts.

1b. The UN System's convening power, normative and policy instruments, institutional functions and frameworks, stakeholder networks and resources, and diverse expertise and experience should be leveraged to help address global AI governance challenges, including scientific and political consensus building, monitoring and enforcement, capacity development on technical, policy and regulatory aspects, enhanced multistakeholder collaboration, and technology diffusion and dissemination.

1c. Well-conceived AI governance should provide appropriate incentives and guardrails commensurate with the particular characteristics of different AI systems and applications to advance ethical and human rights-based governance while maximizing the technology's positive impact on society and mitigating its risks.

1d. Effective AI governance could be delivered through an ecosystem of critical functions, including but not limited to technology development and consensus building, research and analysis, stakeholder engagement and coordination, standards and norm-setting, capacity building, and monitoring and accountability. These functions have already been tested by the entities surveyed, and this experience can deliver tailored approaches on the basis of the specific networks of focus areas and diverse stakeholder groups within the UN System, facilitated by established governance structures and coordination mechanisms.

1e. Adopting a pragmatic approach and building on the current governance initiatives in the UN system when designing global AI governance efforts is crucial. In addition, the fast pace of technology development, compared to the relatively slow processes to develop new international law instruments of institutional structures, the need for regional or industry/sector-specific approaches, as well as the level of agility of mechanisms and processes that exist at an institutional level also need to be factored in. The UN System has launched various initiatives to adapt to this changing pace of technology development.

1f. Link international AI governance to ongoing efforts on international data governance to ensure complementarity and avoid fragmentation. Effective AI governance requires an integrated approach across related or adjacent governance efforts; it relies on the principles and practices of data governance as availability of quality data is a key enabler for the development of AI.

102. The following specific recommendations are made for consideration by the UN System in further enhancing its AI governance efforts:

2a. Present the UN System’s tools and instruments in AI governance as a combined toolbox to better inform Member States and stakeholders on the tools available: To effectively leverage the existing and developing wide range of multi-layered and multi-faceted tools and instruments on AI and those applicable to AI, there is an opportunity to present all these instruments in a systematic way. Such a “toolbox” could encompass both comprehensive, cross-cutting instruments and those that are more sector-specific or focused on specific aspects of AI governance to help facilitate holistic understanding, adoption, monitoring, and reporting as relevant. In this regard, consider building upon the more impactful initiatives to enhance the effectiveness of the System’s response to global AI governance.

2b. Expand taxonomies that can facilitate technical and normative guidance for existing instruments: Given the rapid evolution of AI and emerging technologies, normative and technical guidance based on existing instruments, including international human rights law, complemented by capacity-building support from the UN System, in the form of a taxonomy of the relevant principles, instruments and technical aspects could be useful, including at a sector-level that can also factor in the related nuances.

2c. Put in place or update internal policies to govern the use of AI within UN system entities, building on the work of the HLCM Task Force on AI to develop a normative guidance or model policy for the UN system on the use of AI.

2d. Leverage, enhance, and scale observatories on AI: Supporting and linking the various observatories that currently exist within the UN System to disseminate best practices, use cases, and lessons learned on the use, identification, adoption, and implementation of existing instruments could be beneficial. In addition, potentially mapping out key stakeholder journeys and steps towards developing specific AI governance frameworks and mechanisms could be useful for Member States and other stakeholders.

2e. Leverage existing networks for building consensus and serve as a platform for communicating key technological milestones and developments in AI: It is recommended that existing multidisciplinary global networks in the UN ecosystem be strengthened and adequately resourced to facilitate scientific consensus building and enable effective information flow and exchange on the rapid developments related to AI.

2f. Invest in and develop in-house granular and comprehensive AI expertise to support Member States effectively, engage with stakeholder groups, and build trust: As can be observed at various specialized technical agencies, in-house core technical and multidisciplinary skills and expertise can lead the entity to be the trusted authority for that topic and also help develop benchmarks and guidance that are received and interpreted globally as ‘norm-setting’ or ‘code of conduct’ in that issue area. Further discussion is needed on the incentives to hire and retain AI talent and expertise within the system, particularly given the competition from the industry. Within the UN system, efforts are ongoing as part of the work of the HLCM Task Force on AI, to identify and promote mechanisms for pooling technical capacity and knowledge sharing on AI.

2g. Enable sandboxes to facilitate the development of internationally harmonized approaches for AI risk assessments and monitoring efforts: Globally, there is a growing recognition of the need for AI safety-related institutions aimed at risk assessment and monitoring of AI systems, as also demonstrated by a few countries recently launching such institutions. The UN System can provide platforms in the form of “sandboxes” that enable relevant stakeholders to convene, develop, and test these assessment approaches, including at sectoral levels, leveraging guidance, frameworks, and instruments from the UN System and its entities.

2h. Proactively manage risks and mainstream foresight capabilities: Aligned with [UN 2.0](#) and the [UN 2.0’s Policy brief on Foresight](#)²³, the UN System’s efforts in technology and AI governance can be bolstered by foresight. Such a foresight function could be embedded into entities’ existing activities. It can be supported through cross-sectoral and multi-stakeholder partnerships, including academic institutions and think tanks focused on technology foresight. Additionally, with an eye to the future, further deliberations are needed on whether the UN System can play a role in addressing issues such as existential AI risks and complex AI-driven attacks on State assets by State or non-State actors.

2i. Invest in talent, data, and compute resources, as well as regulatory and procurement capacity. UN System entities, within their respective mandates and in collaboration with each other, must continue to prioritize, invest in, and support coordination on initiatives aimed at reducing the AI divide in the context of addressing the broader digital divide, including the gender digital divide. Capacity development is key to this effort, with increased focus needed on enhancing regulatory and technical capacities for policymakers and practitioners of AI, considering work that is underway within national AI strategies, as well as sector-specific efforts targeted at key stakeholders, particularly those belonging to developing countries and traditionally marginalized populations. Capacity building on AI for women is also needed to ensure women can equally enjoy the benefits and opportunities of AI, while recognizing the risks and threats that can be detrimental to their well-being.

²³ [UN 2.0’s Policy brief on Foresight](#) shows that 34 percent of UN System entities have designed a strategy for strategic foresight, and climate and crisis situations are primary focus areas.

Annex 1: AI Risks²⁴

1. **The technical nature of AI systems poses regulatory design challenges.** Technology governance approaches involve multiple dimensions, including risk, performance-based governance, safety thresholds, and others. AI, especially the general-purpose nature of foundation models, poses challenges to regulatory design because it's difficult to foresee all permutations and combinations of outcomes from these applications, which in turn makes it challenging to define risk and safety thresholds or align on standards, which would have to be continuously updated.
2. **Opacity and lack of explainability and interpretability of AI systems** hinder effective accountability and decision-making and make it challenging to design governance approaches, as all modalities of the system aren't well understood. Effective guardrails, such as accountability, transparency, equity, and alignment with international human rights law, must be in place to ensure human agency and oversight and to protect human rights.
3. **The decentralized nature of AI applications and models**, and even their quantized versions,²⁵ makes tracking and monitoring every instance and usage of models difficult. This decentralized nature can pose risks around the models being accessed by malicious actors, especially when keeping in mind nefarious use cases of AI, e.g., misinformation troll bots. In addition, there has been a big push for open-sourced AI, which democratizes AI innovation but can also be put to malicious use by certain actors.
4. **Risks from inter-related areas like data, copyright, patents, and cybersecurity can compound AI risks.** The problem of data bias in AI algorithms is also a reflection of inter-generational and historical biases in data collection across gender, race, culture, and other factors. It is often exacerbated in AI Systems, which can lack data diversity. In addition, there is an increasing divergence in interpretations of fair use of data and associated copyright challenges, as well as issues in patenting when the application/system is auto-generated or developed by the AI system. There are also inherent risks emerging from data protection and privacy issues, as well as whether the data being used for AI Systems has been obtained through consent or involves extractive practices. Cybersecurity is a dual risk in terms of adversarial prompt injection (deliberate manipulation of the system for malicious use) or even AI for large-scale complex cyber-attacks.
5. **The market for AI is growing, but so is the AI divide:** Calls to pause giant AI experiments^{viii} have not impacted the growth of AI applications. According to estimates^{ix}, AI investment will reach \$200 billion globally by 2025. While AI innovation is welcome, the risk of a global divide in AI is also evident. A [report from PWC](#) highlights how the most significant economic gains from AI will be in China (26% boost to GDP in 2030) and North America (14.5% boost), equivalent to a total of \$10.7 trillion and accounting for almost 70% of the global economic impact.
6. **The proliferation of principles without adequate accountability, effective enforcement, or redress mechanisms:** The issue of ethics and responsibility of autonomous systems, robots, and AI is several decades old - dating back to Isaac Asimov's Three Laws of Robotics^x. In the past few years, hundreds of principles related to AI governance have emerged, but there is concern about the lack of effective accountability for AI-driven decision-making and associated adequate redress mechanisms.
7. **The disproportionately large role of non-State actors across the value chain and concentration of market power:** non-State actors, i.e., the private sector, are driving most of the developments and investments in AI. The UN System engages the private sector through several multistakeholder groups, but its primary focal point and stakeholders are the Member

²⁴ As highlighted in the paper, this list is not exhaustive but covers major risks that are being observed globally.

²⁵ Quantization as a concept broadly covers techniques to convert input values from a large set to output values in a smaller set. In the case of AI models, a quantized model refers to a 'smaller' version of an AI model, which can be run from a regular laptop/desktop without the need for extensive hardware or computing resources.

States, and enforcement of instruments is contingent upon the legal capacity, resources, the willingness to regulate, and the intent of governments at various levels.

8. **Risk of inadequate inclusion:** The development of AI technologies and developments in AI governance are driven primarily by the Global North. Unequal distribution of technology can impact transparency and confidence between and among States. Further, an accurate representation of disadvantaged groups in the AI development and governance process is missing, resulting in discriminatory or biased outputs. This issue also necessitates adopting a gender and minority lens for the governance of AI.
9. **The dual challenges of AI in the labour force:** AI presents a dual risk to labour force – on the one hand, large-scale automation driven by AI poses risks to the future of work, and the least technical and repetitive jobs are at the highest risk of displacement. On the other hand, concerns are being raised^{xi} that over-reliance on AI systems can result in deskilling in the longer term as people become over-reliant on these systems.
10. **Environmental footprint of AI:** There is growing concern^{xii} about the large environmental footprint of AI, especially with the advent of foundation models with trillions of parameters. While there are efforts to build smaller and more efficient models, the compute requirements of AI applications are increasing the demand for hardware that contains several rare essential minerals and metals and the growing need for cloud computing and data centres, with increasing energy and water consumption needs.

Annex 2: Methodology details

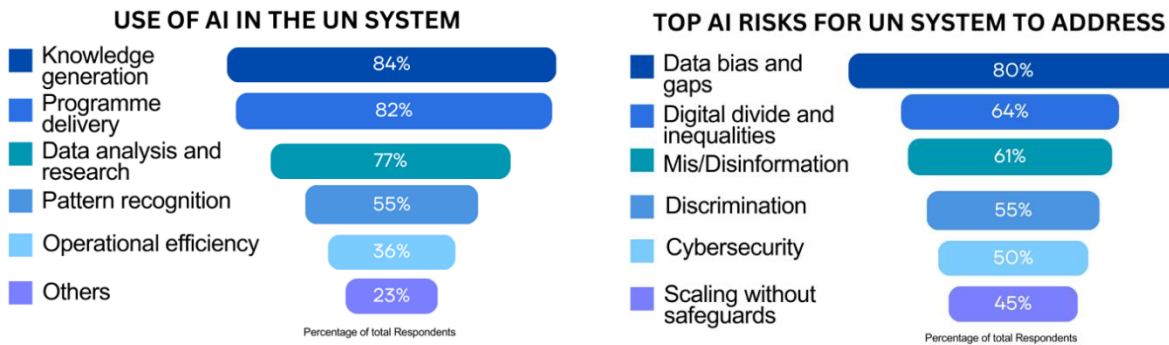
1. The starting point of this research and White Paper was the initial concept note developed by the IAWG-AI and the mandate provided by HLCP to IAWG-AI to develop a White Paper to analyze existing instruments, institutional models, governance structure, and agile and anticipatory approaches to AI governance.
2. The first phase of research, starting in early October 2023, involved desk research, covering some of the initial ideas in the concept note as well as an analysis of extant literature studying different institutional models and governance mechanisms for AI governance, particularly from the viewpoint of international deliberations around the need for a new 'international organization or entity.' The selection of organizations for this desk research was done in consultation with the co-leads of IAWG-AI and the CEB Secretariat. The desk review examined models for developing laws, regulations, norms, and mechanisms for monitoring or enforcing compliance in the context of AI governance and considered the opportunities to integrate anticipatory, agile, or iterative approaches. This desk research was based on the recent papers outlining proposed institutional models and approaches for AI governance.
3. Concurrently, a survey was created and administered to 57 entities within the United Nations system. The survey was designed to gather insights on the three focus areas of the paper. In addition, it also gathered insight on the perspective of entities related to the top AI risks for the UN System to prioritize, as well as the different areas of AI they are exploring in their entities. Out of the 57 entities surveyed, 44 responses²⁶ were received.
4. From the received responses, ten entities were selected based on the relevance and depth of their answers for follow-up interviews. The selection criteria for the follow-up interview were based on the following four factors: i) entities with specific instruments on AI; ii) entities with instruments on areas inter-related with AI or those that could be extended to aspects of AI governance; iii) entities with experience in dealing with the globally complex areas, especially on governance dimensions relevant to AI governance (for example, engaging private-sector for governance issues, globally coordinated sector-specific standards and others); and iv) entities with existing/planned efforts in the top AI risks identified in the survey. The interviews, conducted via videoconference, ranged from 45 to 60 minutes each, allowing for in-depth exploration of identified key themes and issues that had not been fully addressed in the survey responses. The interviews followed a semi-structured format. A set of predefined questions that were shared with the interviewees prior to the interviews provided a foundation for exploring core themes, while the semi-structured nature allowed for flexibility, nuanced perspectives, and open discussion during the interviews.
5. The analysis employed a mixed-methods approach. Qualitative content analysis was used to categorize themes from the desk research, while quantitative analysis identified trends from the survey responses. Thematic coding (qualitative approach) was used for the interview notes. This combination ensured a comprehensive understanding of this complex topic.

²⁶ List of entities that responded to the survey - ECLAC; ESCAP; ESCWA; FAO; IAEA; ICAO; IFAD; ILO; IMF; IOM; ITC; ITU; UN WOMEN; UNAIDS; UNCTAD; UN-DESA; UN-DGC; UN-EOSG; UN-OCHA; UN-OHCHR; UN-OCT; UN-ODA; UN-OICT; UN-OSET; UNDP; UNDRR; UNEP; UNESCO; UNFCCC; UNFPA; UNHCR; UNICEF; UNICRI; UNIDO; UNODC/UNOV; UNOPS; UNRWA; UNU; UNV; UNWTO; UPU; WBG; WFP; WHO; WIPO

Annex 3: Highlights from the survey

The following trends emerged based on the data received from the survey and the interviews:

- **AI is already widely used in the UN System, but entities are still developing guidance for internal usage.** From the respondents, over **80%** of organizations are already using or exploring potential applications of AI. Out of the 37 entities using or exploring AI, **76%** have developed an internal normative/policy guideline on the usage of AI, demonstrating an increasing maturity with which UN agencies are addressing AI while also highlighting that a quarter of entities using or exploring AI lack internal guidance.
- **Data bias, discrimination, and digital divide are top AI risks the UN System should prioritize:** A majority of respondents identified data bias, discrimination, and a deepening of the digital divide as among the top three AI risks that the UN System should address, followed by AI-generated **misinformation and disinformation**, rapid AI deployment without adequate **guardrails**, and **cybersecurity risk**.
- **Developing public administration capacities is critical:** Virtually all organizations identified capacity building for public administration as a top action area for the UN System to facilitate international AI governance, followed by supporting AI literacy and digital public infrastructure.



Annex 4: Instruments from the UN system on AI or inter-related with AI or those that could apply to AI²⁷

Supporting Entity	Name of Instrument	Binding Instrument ²⁸	On AI	Inter-related to AI	Applicable to AI
DESA	The Convention on the Rights of Persons with Disabilities (A/RES/61/106) (disability inclusion and promoting accessibility in digital and physical environments)	✓			✓
DESA	A/HRC/49/70 Digital deficits and access by older persons to information technology			✓	
DESA	Strengthening multistakeholder approach to global AI governance, protecting the environment and human rights in the era of generative AI: A report by the Policy Network on Artificial Intelligence		✓		
ECLAC	Regional AI Index		✓		
ICAO	Chicago Convention Annex	✓			✓
ILO	The Worst Forms of Child Labour Convention, 1999	✓			✓
ILO	Occupational Safety and Health Convention, 1981 (No. 155)	✓			✓
ILO	Promotional Framework for Occupational Safety and Health Convention, 2006 (No. 187)	✓			✓
ILO	Discrimination (Employment and Occupation) Convention, 1958 (No. 111)	✓			✓
ILO	Workers' Representatives Convention, 1971 (No. 135)	✓			✓
ILO	Employment Policy Convention, 1964 (No. 122)	✓			✓
ILO	Discrimination (Employment and Occupation) Convention, 1958 (No. 111)	✓			✓

²⁷ This list is indicative and non-exhaustive

²⁸ Legally binding for Member States that have ratified the relevant instrument.

ILO	ILO Code of Practice on the Protection of workers' personal data				✓
ITU	Over 100 approved standards, and 120 under development. Some examples: AI in Telecom Operations and Management AI in Smart Systems and Cities AI in Network Management and Services AI in Specific Technologies or Applications		✓		
Office of the High Commissioner for Human Rights (OHCHR)	International Convention on the Elimination of All Forms of Racial Discrimination	✓			✓
OHCHR	International Covenant on Civil and Political Rights	✓			✓
OHCHR	International Covenant on Economic, Social and Cultural Rights	✓			✓
OHCHR	Convention on the Elimination of All Forms of Discrimination against Women	✓			✓
OHCHR	Convention against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment	✓			✓
OHCHR	International Convention on the Protection of the Rights of All Migrant Workers and Members of Their Families	✓			✓
OHCHR	International Convention for the Protection of All Persons from Enforced Disappearance	✓			✓
OHCHR	Convention on the Rights of Persons with Disabilities	✓			✓
UNDP	Digital Standards that provide a framework for innovation and improvement, ensuring that new technologies are used effectively, efficiently, and ethically				✓
United Nations Office of Counter-Terrorism	8 th review of the Global Counter-Terrorism Strategy (A/RES/77/298)				✓
UNOCT and UNICRI	Report "Algorithms and Terrorism: The Malicious Use of Artificial Intelligence for Terrorist Purposes"			✓	

UNOCT and UNICRI	Report “Countering Terrorism Online with Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in South Asia and South-East Asia”			✓	
UN Office for Disaster Risk Reduction	Sendai Framework for Disaster Risk Reduction	✓			✓
UN Population Fund	The Programme of Action for the International Conference on Population and Development (ICPD POA) - A population-focused human rights-based framework				✓
United Nations Department of Global Communications (UNDGC)	Code of Conduct Information integrity			✓	
UNESCO	Recommendations on the Ethics of AI		✓		
UNESCO	Guidance for generative AI in education and research		✓		
UNESCO	AI Competency Frameworks for Students and Teachers		✓		
UNESCO	AI and Digital Transformation Competencies for Civil Servants		✓		
UNESCO	Guidelines for the Governance of Digital Platforms			✓	
UNHCR	UN Secretary-General’s Guidance on Human Rights Due Diligence for Digital Technology Use			✓	
UNICEF	The Convention on the Rights of the Child	✓			✓
UNICEF	Policy Guidance on AI for Children		✓		
UNICEF	The Digital Public Goods standard for AI systems (developed together with DPGA)		✓		
UNICEF	Strategic Framework on Child Online Protection			✓	
UNICEF	The Case for Better Governance of Children’s Data: A Manifesto			✓	

UNICEF	Policy on Ethics in Evidence Generation. Procedure on evidence generation utilizing new/emerging/novel technologies or innovations for data collection or processing (directly or through third parties), including predictive or other “black box” analytics.			✓	
UNICEF	Policy on Personal Data Protection and Privacy, which details the permissible use of personal data and required protections			✓	
UNICEF	Responsible Data for Children (rd4c.org) - a set of principles and tools to sensitize audiences to opportunities and risks across the data lifecycle, including data for AI systems			✓	
UNICRI	Toolkit for Responsible Artificial Intelligence Innovation in Law Enforcement		✓		
UN-Habitat	International Guidelines on People-Centred Smart Cities			✓	
UN-Habitat	AI Risk Assessment Framework		✓		
UN-Habitat	Guide to mainstream human rights in the digital transformation of cities				✓
UN-Habitat	Policy framework for centering people, inclusion, and human rights in smart city development				✓
UNICRI	Policy Framework for Responsible Limits on Facial Recognition. Use Case: Law Enforcement Investigations		✓		
United Nations Industrial Development Office (UNIDO)	The Abu Dhabi Declaration, UNIDO GC.18				✓
UNIDO	Standards and Digital Transformation: Good Governance in a Digital Age;			✓	

UNIDO	Global Alliance on AI for Industry and Manufacturing		✓		
UNIDO	GENDER, DIGITAL TRANSFORMATION AND ARTIFICIAL INTELLIGENCE (2023)		✓		
UNIDO	Empowering SMEs through 4IR Technologies: Artificial Intelligence (2024)		✓		
UNIDO	Propelling LDCs in the Digital Age: A 4IR Perspective for Sustainable Development (2021)			✓	
UNIDO	Practical Guide: ISO 56002:2019 Innovation Management Systems				✓
UNIDO	DIGITAL KAIZEN LEAN MANUFACTURING, KAIZEN AND NEW TECHNOLOGIES TO INCREASE BUSINESS PRODUCTIVITY			✓	
UNIDO (and partners)	Empowering Digital Transformation in Small Enterprises Through National Policies: An International Benchmarking			✓	
UNIDO / ESCAP / UNCTAD	Asia-Pacific Trade and Investment Report 2023/2024: Unleashing Digital Trade and Investment for Sustainable Development			✓	
UNIDO	Digital Maturity Assessment tool for National Standardization Bodies (NSBs)			✓	
UNODA	Article 36 of Additional Protocol I to the Geneva Conventions – best practice sharing around ‘new weapons review’	✓			✓
UNODA	Biological weapons convention – AI could reduce barriers to the development of biological weapons	✓			✓
UNODC	The Kyoto Declaration				✓

UN Women	CSw67 Agreed conclusions				✓
UPU	UPU Multilateral Data Sharing Agreement	✓		✓	
WHO	Regulatory considerations on artificial intelligence for health		✓		
WHO	Guidance on Ethics & Governance of Artificial Intelligence for Health		✓		
WHO	Generating Evidence for Artificial Intelligence Based Medical Devices: A Framework for Training Validation and Evaluation		✓		
WIPO	Rome Convention for the Protection of Performers, Producers of Phonograms and Broadcasting Organizations	✓		✓	
WIPO	Berne Convention for the Protection of Literary and Artistic Works	✓		✓	
WIPO	Beijing Treaty on Audiovisual Performances	✓		✓	
WIPO	Patent Cooperation Treaty	✓		✓	
World Bank	Environmental and Social Framework	✓ ²⁹			✓

²⁹ Applicable to all World Bank Investment Policy Financing (IPF) operations.

Endnotes

ⁱ Luccioni, Alexandra Sasha, Yacine Jernite, and Emma Strubell. "Power Hungry Processing: Watts Driving the Cost of AI Deployment?" *arXiv preprint arXiv:2311.16863* (2023)

ⁱⁱ Jones, Kate. "04 Principles of AI Governance: The Contribution of Human Rights." Chatham House – International Affairs Think Tank, January 10, 2023. <https://www.chathamhouse.org/2023/01/ai-governance-and-human-rights/04-principles-ai-governance-contribution-human-rights>

ⁱⁱⁱ Trager, R. F., Harack, B., Reuel, A., Carnegie, A., Heim, L., Ho, L., ..., & Villalobos, J. J. "INTERNATIONAL GOVERNANCE OF CIVILIAN AI: A Jurisdictional Certification Approach." White paper, August 2023. Oxford Martin School. https://www.oxfordmartin.ox.ac.uk/downloads/academic/International_Governance_of_Civilian_AI_OMS.pdf

^{iv} Maas, M., & Villalobos, J. J. "International AI Institutions: A Literature Review of Models, Examples, and Proposals." AI Foundations Report 1, Legal Priorities Project, September 2023. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4579773

^v Fournier-Tombs, E., Siddiqui, M., Castets-Renard, C., & Ndior, V. "Technical Note: A Global Architecture for Artificial Intelligence." United Nations University Centre for Policy Research, 2023. https://collections.unu.edu/eserv/UNU:9284/global_architecture_AI.pdf

^{vi} Jason Hausenloy, Claire Dennis. "Working Paper: Towards a UN Role in Governing Foundation Artificial Intelligence Models." United Nations University, September 2023. <https://unu.edu/cpr/working-paper/towards-un-role-governing-foundation-artificial-intelligence-models>

^{vii} Vanian, Jonathan. "ChatGPT and generative AI are booming, but at a very expensive price." CNBC, March 13, 2023. <https://www.cnbc.com/2023/03/13/chatgpt-and-generative-ai-are-booming-but-at-a-very-expensive-price.html>

^{viii} Future of Life Institute. (2023, March 10). Open letter on the need for control over large artificial intelligence experiments. Retrieved from <https://futureoflife.org/fli-open-letters/>

^{ix} Briggs, Joseph, and Devesh Kodnani. "Generative AI Could Raise Global GDP by 7%." Goldman Sachs. April 5, 2023. <https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>

^x "Three Laws of Robotics." Encyclopaedia Britannica. Encyclopaedia Britannica, Inc. <https://www.britannica.com/topic/Three-Laws-of-Robotics>

^{xi} Varwandkar, Ajit. "Is Chat GPT the End of Thinking Skills?" The Times of India Blog, Feb 10, 2023. <https://timesofindia.indiatimes.com/blogs/the-next-step/is-chat-gpt-the-end-of-thinking-skills/?source=app&frmapp=yes>

^{xii} Cho, Renee. "AI's Growing Carbon Footprint." State of the Planet. Columbia University's Earth Institute, June 9, 2023. <https://news.climate.columbia.edu/2023/06/09/ais-growing-carbon-footprint/>