

TECHNICAL NOTE

September, 2023

A Global Architecture for Artificial Intelligence

Eleonore Fournier-Tombs, Muznah Siddiqui, Céline Castets-Renard, and Valère Ndior

Introduction

In August 2023, the United Nations Secretary-General announced the creation of a Multistakeholder High-Level Advisory Body on Artificial Intelligence (AI). This Body will meet during the last quarter of 2023 and produce a report at the end of the year to advise the United Nations (UN) and its Member States on the next steps for the global governance of AI.¹ This latest step in global AI governance follows many other multilateral efforts, including a new initiative from the members of the G7, who announced their commitment in May 2023 to urgently push for a global technical standard on AI.² All of this comes after several years of multilateral efforts on AI, including, notably, *Recommendations on the Ethics of AI*, published by the United Nations Educational, Scientific, and Cultural Organization (UNESCO),³ and signed by all UNESCO Member States, which has been hailed as the first global normative effort on AI.

In recent months, there have also been many ideas published by private sector⁴ and academic thought leaders⁵ on an architecture for global AI governance, which according to the frameworks presented, might include several interconnected organizations; a focus on Intergovernmental Panel on Climate Change (IPCC)-style assessments; or various new commissions and bodies.

This paper is an effort to consolidate some of the key questions surrounding global AI governance today, in order to inform the UN's efforts in this domain. It builds out the third recommendation of Shift 6, put forward in the report of the UN's High-Level Advisory Board (HLAB) on Effective Multilateralism, entitled: *A Breakthrough for People and the Planet*,⁶ published in 2023. The recommendation calls for a multilateral agreement on a timeline for a global architecture for AI design, development, and use based on common standards

1 "Multistakeholder High-Level Advisory Body on AI," Office of the Secretary General's Envoy on Technology, last accessed 17 September 2023, <https://www.un.org/techenvoy/content/artificial-intelligence>.

2 Hiroki Habuka, "The path to trustworthy AI: G7 Outcomes and Implications for Global AI Governance," Center for Strategic and International Studies, 6 June 2023, <https://www.csis.org/analysis/path-trustworthy-ai-g7-outcomes-and-implications-global-ai-governance>.

3 United Nations Educational, Scientific, and Cultural Organization (UNESCO), *Recommendation on The Ethics of Artificial Intelligence* (Paris: UNESCO, 2022). Accessible at: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.

4 See, for example: Lewis Ho, Joslyn Barnhart, Robert Trager, Yoshua Bengio, Miles Brundage, Allison Carnegie, Rumman Chowdhury, Allan Dafoe, Gillian Hadfield, Margaret Levi, and Duncan Snidal, "International Institutions for Advanced AI," *arXiv preprint arXiv:2307.04699*.

5 Michael Veale, Kira Matus, and Robert Gorwa, "AI and Global Governance: Modalities, Rationales, Tensions," *Annual Review of Law and Social Science* Vol. 19 (2023).

6 See: High-Level Advisory Board on Effective Multilateralism (HLAB), *A Breakthrough for People and Planet: Effective and Inclusive Global Governance for Today and the Future* (New York: United Nations University, 2023).

This Technical Note is an output of the UNU-CPR initiative, 'A Breakthrough for People and Planet: Building Momentum for the Summit of the Future and Beyond.'



and approaches.⁷ This paper is also informed by a July 2023 workshop at the United Nations University Centre for Policy Research (UNU-CPR) which involved over 30 academics and UN thought leaders, all of whom are listed in the acknowledgements section.

The paper has three parts. First, a rationale for the UN-led global governance of AI. Second, a description of the building blocks of a global architecture for AI, adapted from existing models of global governance. Third, four decision points for members of the High-Level Advisory Body on AI, related to agile policymaking, multistakeholder engagement, interoperability, and scope.

A Rationale for the UN-led Global Governance of AI

There have been many discussions globally focused on the risks posed by AI, which can range from benign to catastrophic. Shevlane et al speak about alignment risks,⁸ in which intended outcomes diverge from actual outcomes. This is particularly the case for the issue of bias in AI, which often is unintended, but which nonetheless has the effect of both permanently embedding old discriminations into our systems and accelerating inequalities of the status quo, and causing negative social effects, such as polarization, erosion of democracy, and misinformation. However, there are also many intentionally harmful uses of AI, such as AI used for surveillance, autonomous weapons systems deployed in conflict, deepfake disinformation, and gender-based digital violence.⁹ Inequality exacerbated by bias and discrimination also risks adding to global instability, which threatens the UN's work in peacekeeping and peacebuilding.

In a July 2023 speech to the Security Council,¹⁰ the Secretary-General discussed the dual nature of AI, highlighting its potential for both risk and opportunity. He identified important concerns around AI in areas like discrimination, surveillance, and peace and security. Conversely, he also noted AI's prom-

ise in sustainable development, climate crisis mitigation, and medical research advancement.

These risks are significant, and are crossborder in nature. In an era of digital globalization, Member States cannot regulate independently, and have increasingly called for a global effort to govern AI. Additionally, many of the risks of AI are closely related to the pillars of the UN: sustainable development, human rights and gender equality, and peace. In this sense, a UN-led effort is critical, not only because AI requires a global multilateral forum for governance, but also because unregulated AI could undermine other multilateral priorities.

Global Coordination

Global coordination of AI governance mechanisms is critical. While national governments led the first AI regulation efforts, they have recently been joined by various multilateral and regional initiatives. Without the UN to coordinate these efforts from a global perspective, AI governance risks having redundancies and gaps. The transboundary nature of AI, touching also on the governance of data, hardware, and the Internet, requires the leadership of the UN and the cooperation of its 193 Member States.

Regional Coordination

One important area of coordination for the UN will be with regional organizations, such as the European Union, the African Union, the Association of Southeast Asian Nations, the Organisation for Economic Co-operation and Development, and the Council of Europe. If a global approach to regulating AI does not appropriately include the concerns and priorities of regional organizations, particularly from the Global South, such organizations may express opposition. The African Union, for example, is developing initiatives to address the impact of AI on their respective regions, particularly in terms of digital transformation, people's rights,¹¹ and human rights. These organizations aim to establish unique legal frameworks that align with

7 "Agreeing on a timeline for the development of a global architecture for AI design, development, and use based on common standards and approaches. This could be taken forward by a series of dialogues between governments, the private sector, and civil society under the aegis of the UN and, specifically, the proposed Global Commission on Just and Sustainable Digitalization. The proposed Secretary-General's Scientific Advisory Board could engage with this process and report regularly on the evolving nature of AI technologies, feeding into the Global Commission's knowledge function (its fourth competency). The objective will be to generate a set of definitions and standards for identifying and mitigating global AI risks. It would be important to ground the process in existing AI standards while developing additional standards that strengthen practices of safety and responsibility among AI practitioners. This activity should take into account the pioneering global normative frameworks recently adopted on the ethics of AI. It could also be linked to a fund that would incentivize research and preparedness on the existential risks that can arise from ungoverned AI evolutions."

8 Toby Shevlane, Sebastian Farquhar, Ben Garfinkel, Mary Phuong, Jess Whittlestone, Jade Leung, Daniel Kokotajlo, Nahema Marchal, Markus Anderljung, Noam Kolt, Lewis Ho, Divya Siddarth, Shahar Avin, Will Hawkins, Been Kim, Iason Gabriel, Vijay Bolina, Jack Clark, Yoshua Bengio, Paul Christiano, and Allan Dafoe, "Model evaluation for extreme risks," *arXiv preprint* arXiv:2305.15324.

9 Branka Marijan and Wanda Muñoz, "Autonomous Weapons and Deepfakes: The Weaponization of Artificial Intelligence and the Urgent Need for Regulation," *Missing Links in AI Governance* eds. Benjamin Prud'homme, Catherine Régis, Golnoosh Farnadi (UNESCO and Mila – Quebec Institute of Artificial Intelligence, 2023).

10 "Secretary General's remarks to the Security Council on Artificial Intelligence," United Nations, 18 July 2023, <https://www.un.org/sg/en/content/sg/speeches/2023-07-18/secretary-generals-remarks-the-security-council-artificial-intelligence>.

11 African Commission, Resolution on the need to undertake a Study on human and peoples' rights and artificial intelligence (AI), robotics and other new and emerging technologies in Africa, 2021, ACHPR/Res. 473 (EXT.OS/XXXI).

their specific interests and promote a unified stance on AI-related issues. This means that any new global governance on AI led by the UN should take into account the work of regional organizations and ensure that mechanisms exist for consultation, thereby avoiding unnecessary overlap or redundancies, ensuring that a global governance mechanism adequately reflects regional concerns and priorities, and supporting vertical interoperability. This would ensure that private sector companies, civil society organizations, and other entities would not face different standards that complicate compliance. In addition, a UN-led global governance process would have legitimacy by garnering the necessary buy-in from regional bodies, and ensuring that global, regional, and national principles reinforce each other. Perhaps most importantly, it would secure the participation of a wide variety of stakeholders and the representative organizations of marginalized groups, who disproportionately experience the negative impacts of AI.¹²

The UN's Governance of its Own Uses of AI

The UN itself is only regulated by international law, and as such does not fall under the jurisdiction of national or regional law when it comes to AI.¹³ However, the UN is increasing the number of AI initiatives to support its activities across all relevant areas.¹⁴ A growing number of agencies are defining standards that enable them to internally implement the best practices they recommend to their members (for example operational policies and internal regulations). In addition, UN agencies are coordinating their approach to AI through the Inter-agency AI Working Group, co-chaired by UNESCO and the International Telecommunication Union (ITU).¹⁵ By adopting operational standards applicable to its internal activities, the UN aims to further reinforce its objectives when it comes to developing an international legal framework for AI, and demonstrate its accountability to stakeholders. As such, in 2022, the AI Working Group published the *Principles for the Ethical Use of AI in the United Nations System*.¹⁶ A next step in this domain would be further efforts on detailing and operationalizing these principles.

Description and Examples of Humanitarian AI

Many instances of AI tools are being used to support humanitarian activities throughout the UN. Types of AI applications have included biometric identification and optimization analytics for humanitarian aid delivery, and predictive analytics to anticipate the effects of severe drought, the spread of waterborne diseases or COVID-19, and the arrival of refugees in camps. Geographic information systems analysis is also increasingly used to understand climate change, natural disaster risk or impact, deforestation, urbanization, and mobility. Faced initially with a lack of guidelines for safe humanitarian AI, many practitioners have begun proposing their own frameworks, including the Humanitarian Data Science Ethics Group,¹⁷ co-chaired by IOM and The Data Science Initiative, and OCHA's Centre for Humanitarian Data Peer Review Framework for Predictive Analytics in Humanitarian Response.¹⁸

Sustainable Development

The role of AI in contributing to the achievement of the Sustainable Development Goals (SDGs) is well documented. In his *Roadmap for Digital Cooperation*, the UN Secretary-General stressed the importance of better harnessing AI for the SDGs. In February 2023, the ITU, in partnership with 40 UN partners, convened an event on the role of AI in achieving the SDGs, specifically highlighting SDG 11 (sustainable cities and communities) and 13 (climate action).¹⁹ The ITU has noted the ability of AI to act as a double-edged sword, by simultaneously contributing to the fulfilment of 134 targets across the SDGs (for example, related to SDG1 on no poverty, SDG7 on clean water and sanitation, and SDG11 on sustainable cities), yet at the same time inhibiting the achievement of 51 targets, resulting, for instance, from a high carbon footprint, gender risks from algorithmic bias, and the uneven distribution of technology.²⁰

12 "Artificial intelligence must be grounded in human rights, says High Commissioner," United Nations Human Rights Office of the High Commissioner (OHCHR), 12 July 2023, <https://www.ohchr.org/en/statements/2023/07/artificial-intelligence-must-be-grounded-human-rights-says-high-commissioner>.

13 Eleonore Fournier-Tombs, "Towards a United Nations Internal Regulation for Artificial Intelligence," *Big Data & Society* Vol 8 Issue 2 (2021).

14 ITU catalogs each year the AI-related initiatives in the UN system, which are listed here: <https://aiforgood.itu.int/about-ai-for-good/un-ai-actions/>.

15 "Interagency Working Group on Artificial Intelligence," UN System Chief Executives Board for Coordination, last accessed 13 September 2023, <https://unsceb.org/inter-agency-working-group-artificial-intelligence>.

16 United Nations System, *Principles for the Ethical Use of AI in the United Nations System* (2022). Accessible at: https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf.

17 See: "Humanitarian Data Science and Ethics Group," Data Science & Ethics Group, last accessed 13 September 2023, <https://www.hum-dseg.org/>.

18 See: The Centre for Humanitarian Data, *Peer Review Framework for Predictive Analytics in Humanitarian Response* (Office for the Coordination of Humanitarian Affairs, 2021). Accessible at: https://data.humdata.org/dataset/2048a947-5714-4220-905b-e662cbcd14c8/resource/76e488d9-b69d-41bd-927c-116d633bac7b/download/peer-review-framework-2020.pdf?_gl=1*1yzbarb*_ga*ODcwNTczMzc0LjE2ODg0ODg5NDA.*_ga_E60ZNX2F68*MTY5MzM0MDEwNS4xMC4xLjE2OTMzNDxMTAuNTUuMC4w.

19 "Event Highlights Potential of Artificial Intelligence in Enabling SDGs." IISD SDG Knowledge Hub, 1 February 2023, <https://sdg.iisd.org/news/event-highlights-potential-of-artificial-intelligence-in-enabling-sdgs/>.

20 "The Role of Artificial Intelligence in Achieving the Sustainable Development Goals," AI for Good, last accessed 13 September 2023, <https://aiforgood.itu.int/event/the-role-of-ai-in-achieving-the-sustainable-development-goals/>; Ricardo Vinuesa, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia Dignum, Sami Domisch, Anna Felländer, Simone Daniela Langhans, Max Tegmark and Francesco Fuso Nerini, "The Role of Artificial Intelligence in Achieving the Sustainable Development Goals," *Nature Communications* Vol 11 Issue 233 (2020). Accessible at: <https://www.nature.com/articles/s41467-019-14108-y>.

The Global Partnership on AI (GPAI)'s *Climate Change and AI* report presents a comprehensive overview of the opportunities of AI for climate change, as well as its risks, especially its high carbon footprint.²¹ The GPAI and The Future Society have also noted several challenges in implementing AI systems to support the SDGs, including a prioritization challenge, given the breadth of possible AI applications for the SDGs.²²

AI governance is a key concern as it pertains to the achievement of the SDGs, particularly as it relates to fairness and accountability. However, existing AI Ethics frameworks often avoid mentioning the SDGs.²³ Given AI's potential to significantly disrupt most societal sectors, thorough consideration and oversight of its impact on the SDGs is critical.

In addition, efforts to enhance the positive impact of AI for the SDGs and minimize risks and damages will need to take capacity building into account, especially as it relates to bridging the global AI divide.²⁴ In fact, AI has increased the importance of the digital divide, with current infrastructure, skills, and capacity related to AI concentrated in the hands of a few major AI players in the Global North. While many emerging economies are interested in harnessing AI's potential for their own sustainable development, they have had to contend with multiple challenges.

For example, the treatment of AI workers in the Global South has emerged as a significant issue, with vulnerable and poorly paid labourers employed as dataset labellers by many leading AI companies. On the other hand, there have been examples of companies in the Global South upskilling their workers to meet growing demand in skilled AI work, in a way facilitated by global digital transformation. Connected Women, a woman-led social enterprise in the Philippines,²⁵ is a prime example of what might be possible at scale with a global AI governance regime that aims to protect workers in the Global South, promote training and capacity development, and allow for collaboration on localization and the decentralization of AI.

Lessons Learned from 'Connected Women' in the Philippines

The Connected Women ELEVATE AIDA initiative was introduced in the Philippines to equip women with AI skills through training and capacity building programmes. The initiative, funded by public and private partners, has been successful in enhancing the digital upskilling of women, and has disbursed 1046 scholarships for women since 2020, resulting in 20 per cent of graduates obtaining work immediately after graduation.²⁶

The initiative has successfully trained women in data annotation skills, prepared women for reintegration in the Philippines labour market, provided women with remote work opportunities, and validated business models that incorporate women into the tech workforce at scale. This approach can contribute to global approaches to AI governance by emphasizing skills development and AI literacy for women, integrating gender-responsive principles, and addressing bias against women at the earliest stages of the AI life cycle by engaging relevant and connected stakeholders in advisory roles.²⁷

Human Rights

In the realm of AI governance, any initiative will need to consider implications for human rights, aligning with the principles enshrined in the UN Charter and associated legal frameworks. UN High Commissioner for Human Rights Volker Türk has stressed that: "any solution – any regulation – must be grounded in respect for human rights."²⁸

To date, there is no universal framework for digital activities covering human rights considerations such as the protection of personal data or automated decision-making processes. However, equality and non-discrimination are protected in both digital and non-digital realms. The absence of a dedicated framework

21 Global Partnership on AI (GPIA), *Climate Change and AI. Recommendations for Government Action* (GPIA, 2021). Accessible at: <https://www.gpai.ai/projects/climate-change-and-ai.pdf>.

22 The Future Society, *Areas for Future Action in the Responsible AI Ecosystem* (The Future Society, 2020). Accessible at: <https://gpai.ai/projects/responsible-ai/areas-for-future-action-in-responsible-ai.pdf>.

23 Osama Nasir, Rana Tallal Javed, Shivam Gupta, Ricardo Vinuesa, Junaid Qadir, "Artificial Intelligence and Sustainable Development Goals Nexus via Four Vantage Points," *Technology in Society* Vol 72 (2023). Accessible at: <https://www.sciencedirect.com/science/article/pii/S0160791X22003128>.

24 Danni Yu, Hannah Rosenfeld, and Abhishek, "The AI Divide between the Global North and the Global South," World Economic Forum, 16 January 2023, <https://www.weforum.org/agenda/2023/01/davos23-ai-divide-global-north-global-south/>.

25 "Bringing Opportunities Home," Connected Women, last accessed 13 September 2023, <https://connectedwomen.com/>.

26 "Building an Inclusive Future: The Intersection of AI Governance and Women's Empowerment," Connected Women, last accessed 13 September 2023, https://drive.google.com/file/d/1KwPnlQ0-SnO_p7MHAQvafujOioWQsf3y/view.

27 Ibid.

28 "Artificial intelligence must be grounded in human rights, says High Commissioner," OHCHR, last accessed 18 September 2023, <https://www.ohchr.org/en/statements/2023/07/artificial-intelligence-must-be-grounded-human-rights-says-high-commissioner>.

does not negate the applicability of pre-existing international legal instruments. All UN Member States have ratified at least one international human rights treaty, such as the United Nations Declaration of Human Rights, the Convention on the Elimination of All forms of Discrimination Against Women, the Convention on the Rights of Persons with Disabilities, and the Convention on the Rights of the Child.²⁹ These instruments serve as the primary existing means to regulate AI activities conducted by both public entities (States, international organizations) and private actors (companies, non-profit organizations).

In light of this, a UN-led structure would acknowledge the direct relevance of human rights, as stipulated in pertinent legal instruments, to the domain of AI. While certain States may voice objections due to differing views on global human rights norms, connecting AI standards to human rights instruments remains essential. Notably, even regional human rights conventions, such as those in Europe, Africa, and the Americas, reference shared principles in the Universal Declaration of Human Rights and the UN Charter. This interconnection enables UN practices to influence regional court assessments of AI with a human rights perspective, fostering a broader understanding of human rights protections across regions.

Maintaining Global Peace, Peacekeeping, and Humanitarian Aid

In recent years, AI systems have increasingly been used to inform a range of UN activities: conflict prevention, peacekeeping, peace mediation, peacebuilding,³⁰ human rights efforts, development work, and humanitarian aid. However, they have also caused damage to peacekeeping and humanitarian operations specifically, and more broadly, to global peace.³¹ In his speech to the Security Council on 18 July 2023, the UN Secretary-General noted the threat that AI-led cyber-attacks pose to peacekeeping missions and critical infrastructures.³² In 2022, the UN peacekeeping operation in the Democratic Republic of Congo (DRC) was attacked, in a lethal incident that was blamed, at least partially, on the spread of disinformation online. Military uses of AI, such as autonomous weapon systems and surveillance systems, provide a looming threat to global peace, especially as frontier generative models are loaded onto highly advanced robots, which could now be deployed by any actor.

Generative AI has the potential to transform the delivery of humanitarian aid, by enhancing access to information during crises, strengthening preparedness and responses to emergencies, and supporting advocacy work for fundraising.³³ However, the use of AI technologies for delivering humanitarian assistance comes with documented risks. Surveillance humanitarianism has emerged as a key risk, since the collection of personal data, such as biometric and migration data, could increase the vulnerability of aid recipients.³⁴ Humanitarian AI, like all types of AI, also runs the risk of bias, such as having outputs that exclude information from a marginalized group. In this context, there is a particular risk that the most vulnerable aid recipients may not benefit from AI tools.³⁵

AI and Disinformation

New developments in AI, especially in the field of Generative AI, threaten to accelerate the spread and severity of disinformation globally. Already, disinformation is a significant issue, which has had an impact on the UN's humanitarian and peacekeeping operations. For example, the Secretary-General's *Policy Brief on Information Integrity of Digital Platforms* showed that the majority of UN Country Offices are concerned about disinformation. Globally, there have been many instances of disinformation spread online, leading to violence.

A research project with global peacebuilding organization Interpeace³⁶ found growing concerns around how AI can be used to create and manipulate online content, with relatively little investment in moderating social media platforms, especially in countries already experiencing humanitarian crises, such as DRC or Myanmar. There is, in fact, growing evidence that Generative AI will contribute to the creation of deepfakes and other types of disinformation used to manipulate elections and contribute to political tensions. Any new initiative on global AI governance should therefore address, as a key priority, the impact of AI on online discourse, hate speech, and disinformation leading to conflict.

29 "Status of Ratification Interactive Dashboard," OHCHR, last accessed 13 September 2023, <https://indicators.ohchr.org/>.

30 Daanish Masood Alavi, Martin Wählisch, Colin Irwin, and Andrew Konya, "Using Artificial Intelligence for Peacebuilding," *Journal of Peacebuilding & Development* Vol 17 Issue 2 (2022): 239-243. Accessible at: <https://doi.org/10.1177/15423166221102757>.

31 United Nations, "International Community Must Urgently Confront New Reality of Generative, Artificial Intelligence, Speakers Stress as Security Council Debates Risks, Rewards," 18 July 2023.

32 Ibid.

33 Jay Mahanand, "Generative AI – A Game Changer for Humanitarian Assistance," LinkedIn, 14 April 2023, <https://www.linkedin.com/pulse/generative-ai-game-changer-humanitarian-assistance-jay-mahanand>.

34 Mark Latonero "Stop Surveillance Humanitarianism," *New York Times*, 11 July 2019. Accessible at: <https://www.nytimes.com/2019/07/11/opinion/data-humanitarian-aid.html>.

35 Ana Beduschi, "Harnessing the Potential of AI for Humanitarian Action: Opportunities and Risks," *International Review of the Red Cross*, No. 919, June 2022. Accessible at: <https://international-review.icrc.org/articles/harnessing-the-potential-of-artificial-intelligence-for-humanitarian-action-919>.

36 Eleonore Fournier-Tombs, Rebecca Brubaker, and Eduardo Albrecht, "Artificial Intelligence-Powered Disinformation and Conflict," *UNU-CPR Policy Brief* (New York: United Nations University, 2023). Accessed at: <https://unu.edu/sites/default/files/2023-09/artificial%20Intelligence%20powered%20disinformation.pdf>.

In this sense, AI has already had a significant impact on the most important facets of the UN’s humanitarian and peace-related work. The UN therefore must play a key role in ensuring that AI supports rather than actively threatens these activities - whether it is maintaining or moving towards global peace, deploying peacekeeping missions, or coordinating humanitarian interventions to support disaster victims.

The Building Blocks of a Global Architecture for AI

The HLAB report defines a global architecture as containing “a set of definitions and standards for identifying and mitigating global AI risks.”³⁷ In this sense, these definitions and standards can be combined to create a structure that will serve an important global function and be resilient to new technological advances and unforeseen policy challenges, as well as protect human rights.

Types of Norms for Global AI Governance

In the table below, we distinguish between different types of norms related to AI, all of which should be addressed by a global governance of AI mechanism.

At a global level, the UN has played an important role in leading States to negotiate and adopt common ethical and technical standards.³⁸ There will likely be pressure on future efforts in global AI governance to contain a legal component, as more and more countries advance their regulatory efforts. In this sense, we can consider ethical norm development to have been the groundwork for further legal and technical norm-setting, all of which could be globally coordinated.

Models for Global AI Governance

In this section, we draw on several existing global governance models that might inspire a new global governance mechanism. **While we do not imply that creating a new body is the only way forward, understanding the structure of these bodies from a modular perspective can help us to identify the tools at our disposal.**

In the figure below, we therefore examine the different modules, or building blocks, of global governance models as developed by Member States for the International Civil Aviation Organization (ICAO), the Intergovernmental Panel on Climate Change (IPCC), and the International Atomic Energy Agency (IAEA).

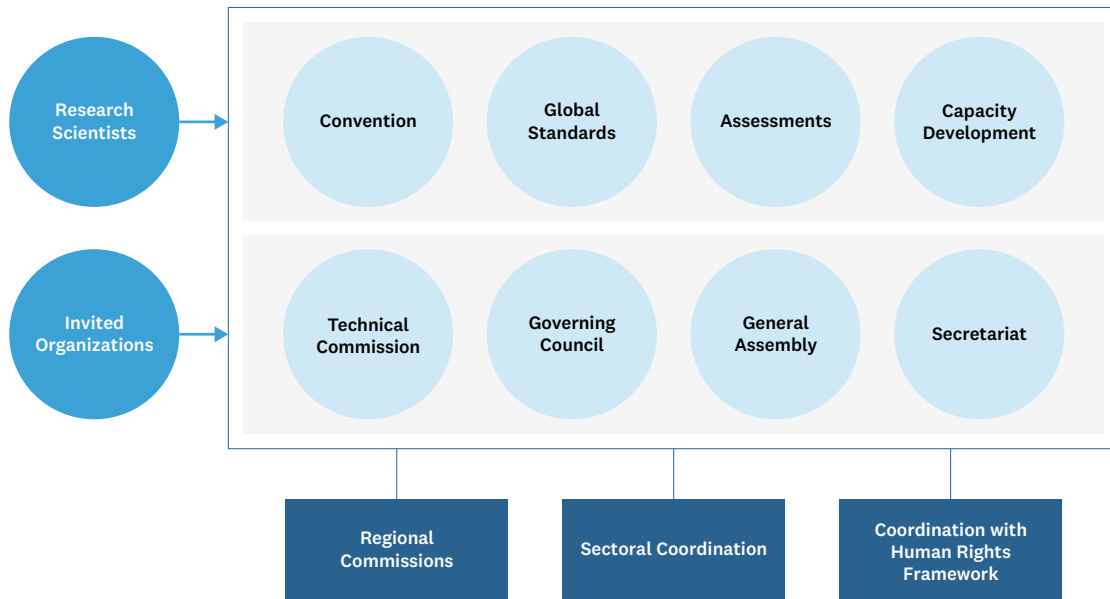
Type of Norm	Description
Ethical Norms	Ethical standards were the first to be developed at early stages of AI governance. ³⁹ These efforts, such as UNESCO’s <i>Recommendations on the Ethics of AI</i> , have become cornerstones for a global AI architecture. The presence of AI ethics as a field has also allowed for an asynchronous global deliberation on AI, with a multitude of exercises at the grassroots, private sector, and government levels aimed at better understanding what global society expects from AI.
Legal Norms	To date, only a handful of countries have chosen to develop AI regulatory projects, with specific laws related to AI deployments and legal consequences for entities violating them. However, many countries have adopted strategies, policies, or white papers, often explicitly explaining how these might be considered first steps towards regulation. These policies and strategies, however, have not always explicitly mentioned human rights risks, or even ethical principles more broadly.
Technical Norms	Finally, many technical standards for AI have been put in place, including from the ITU, International Organization for Standardization (ISO) and the Institute of Electrical and Electronics Engineers (IEEE). The advantage of these standards has been that they have brought different stakeholders together to coordinate solutions to AI risks, and they have also focused on risks with technical solutions, such as homogenous training datasets, which have often been concrete and actionable. Additionally, there has been sectoral norm setting in relation to medicine, biology, and many other disciplines.

³⁷ See: HLAB, *A Breakthrough for People and Planet*.

³⁸ UNESCO, *Recommendations on The Ethics of Artificial Intelligence*.

³⁹ Jessica Fjeld, Hannah Hilligoss, Nele Achten, Adam Christopher Nagy, and Madhulika Srikumar, *Principled Artificial Intelligence: A Map of Ethical and Rights-Based Approaches* (Berkman Klein Centre, 2019). Accessible at: [https:// ai- hr.cyber.harvard.edu/prim-p-viz.html](https://ai-hr.cyber.harvard.edu/prim-p-viz.html).

Figure 1: The Building Blocks of Models for Global Governance.



In the table below, we examine each building block in turn.

Name	Adaptation to AI
Secretariat	The possibility has been raised of a joint secretariat (following the IPCC model, which was established by both the World Meteorological Organization and the United Nations Environment Programme).
Convention	A convention is a type of legally-binding treaty which establishes broader commitments and obligations for its parties and leaves the setting of specific targets either to subsequent more detailed agreements (usually called protocols) or to national legislation, with plans of action supported by budgets and monitoring mechanisms. In essence, this agreement would serve as an umbrella document which lays down the principles, objectives, and the rules of governance of the treaty regime. In the case of AI, it would be based on UNESCO’s <i>Recommendations on the Ethics of AI</i> , which provide a foundation for future binding norms.
Global Standards and Protocols	These would be norms with a nimbler component derived from the framework convention that could be adopted on a regular basis for more specific issues (for example, hardware, trade, labour, and open source technologies).
Assessments	Assessments would be scientifically-informed overviews of the current state of an issue governed by the global AI architecture, including both global and national evaluations of progress and regression. Given the level of agility required with AI, the assessments could be more frequent than IPCC assessments.
Capacity Development	Capacity development initiatives would involve supporting Member States to build both technical capacity (hardware, software, connectivity, etc.) and AI-related skillsets. AI has evolved significantly in the past few years, especially with recent advances in Generative AI, and will likely continue to do so. Part of any UN-led initiative would need to involve building understanding and capacity across Members States so they can harness AI to support sustainable development, human rights, and peace.

Name	Adaptation to AI
Invited Organizations	Typically, a list composed of international organizations, civil society organizations (especially those representing marginalized groups, the Global South, and human rights defenders), and private sector organizations would be drawn up ahead of time. Organizations would apply for membership to this list. Invited organizations would then be invited to meetings as contributors (not observers), but would not have voting power.
Research Scientists	In the case of the IPCC, thousands of research scientists work together to produce scientific assessments every 5–6 years. To support the governance of AI, AI and social scientists would need to produce assessments more frequently.
Technical Commission	The Technical Commission would meet regularly and could be subdivided into different topics.
Governing Council	In practice, this council could be composed of two parts: an executive board and a plenary. ICAO and IAEA councils have 36 and 35 members, respectively, which meet regularly to approve recommendations from their technical commissions and address other governance issues.
Assembly	The Assembly would be composed of all Member States and they would be required to meet every 1–3 years. While some decisions could be taken directly by the Governing Council, others would be voted on by the Assembly. The Assembly would also be responsible for electing the Governing Council, according to a specific rotation and geographic distribution.
Regional Commissions	The Regional Commissions would be responsible for leading the regional and national implementations of the Convention and protocols, including, potentially, direct engagement with Member States.
Sectoral Coordination	A sectoral coordination mechanism would interface with UN agencies engaged in policy, to ensure interoperability between sectoral policies and the framework convention. It would also translate sectoral needs to different bodies in the Secretariat. A possible model for this would be humanitarian coordination, which similarly involves a coordinating body and specific roles and responsibilities for different UN agencies.
Coordination with Human Rights Frameworks	This mechanism would play an advisory and monitoring role and potentially receive complaints. ⁴⁰

Four Decision Points for Global Policymakers

During negotiations around a UN-led global governance of AI there are several areas, in particular, that will require thoughtful deliberation from decision-makers. Each of these areas are outlined in this section, along with tensions that would need to be addressed.

Decision Point 1: Agile Policymaking

One of the most difficult hurdles in regulating AI has been its constant evolution as a technology. While the UN system has

regulated other technologies, such as civil aviation and nuclear energy, AI is unique in that it is constantly evolving. The question raised has therefore been: how can a global governance architecture be put in place when the nature and the use of the technology is in constant flux?

This tension has given rise to a multitude of different interpretations, which have often been linked to a belief that agile policymaking can only be achieved through a non-binding and very broad regulatory framework. In fact, the term agile refers to a well-established technique in software project management

⁴⁰ See, for example: “Introduction to the Committee: Committee on the Rights of Persons with Disabilities,” OHCHR, last accessed 13 September 2023, <https://www.ohchr.org/en/treaty-bodies/crpd/introduction-committee#:~:text=Through%20its%20engagement%20and%20cooperation,provisions%20enshrined%20in%20the%20Convention.>

that iterates through code releases in short cycles.⁴¹ If we take this definition, agile policymaking would involve, among other things, active working groups on specific topics that would meet and review guidelines and standards on a regular basis. As we see in the diagram above, current models for the global governance of technologies do include a provision for standard-setting and reviewing critical topics at the working level.

The main components of an agile policymaking framework for Global AI Governance are outlined below:

Convention and Global Standards: While a Convention would be the more static instrument, the methodology for arriving at global standards could be developed using a more agile methodology, where an active working group would meet regularly to review and adapt standards, based on new technological developments. Expecting an agile method from the beginning would also mean that any Convention would need to be developed in such a way as to allow for rapid technological development that might even replace our current understanding of AI, possibly moving beyond consensus-based models.

Assessments: As we have seen, the assessment method, as used by IPCC, has interesting potential for AI governance. However, there are significant challenges to producing very large and comprehensive research outputs in 6–7 year cycles. Research on the current state of AI, current risks of AI, and the state of governance would likely need to be produced in much shorter sprints, possibly on an annual basis or even more frequently, with the possible addition of more interactive and real-time initiatives.

Decision Point 2: Multi-stakeholder Engagement

Private sector actors are critical in this field, as they are not only leading technological development, but also taking an active role in proposing policy solutions, including at the global level. In other global governance models, private sector actors can participate, just like civil society and academia, in negotiations as invited organizations. They cannot, however, vote, a privilege which is reserved for Member States. While a private sector contribution to AI governance deliberations is critical, it is equally critical that there be a separation between the roles and responsibilities of governments, who are accountable to their citizens, and companies, who are not.⁴² Bremmer and Suleyman, in an article for *Foreign Affairs*,⁴³ succinctly summarize this dilemma as follows: “These actors may not derive legitimacy from a social contract, democracy, or the provision of public goods, but without them, effective AI governance will not stand a chance.”

In addition, a relatively under-considered, but equally critical sector is civil society. Civil society organizations have played a leading role in developing AI ethics, and should participate in decision-making processes that impact the human rights of the stakeholders they are representing. There are therefore several areas of opportunity for multi-stakeholder engagement to consider, such as:

Invited organizations (similar to a model adopted by the IAEA): This model would involve stakeholders applying for membership to a list, after which they could be invited to participate in any meeting convened.

Specific multi-stakeholder consultations: Including the private sector and civil society organizations.

Coordination with multi-stakeholder entities: Such as the Partnership on AI (PAI), or standards-setting organizations, such as ISO and IEEE.

It is important to outline an AI governance effort in such a way that establishes multi-stakeholder participation in a sustainable and engaged way. Member States must consider how to develop consultations with representation from the aforementioned actors, which can be rotated in either an annual, bi-annual, or quarterly basis, and should institute quotas to ensure diversity according to, for example, location, age, race, gender, sexual orientation, ethnic origin, socioeconomic status, and disability.

Decision Point 3: Interoperability and Coordination

The next area of consideration is the interoperability between a global governance framework and national and regional initiatives, as well as other global governance initiatives. As noted above, Regional Commissions of the UN, such as the Economic Commission for Africa (ECA), could offer an opportunity for principles and standards developed globally to cascade down at a regional and country level, while also allowing ideas to come up and be considered globally. They allow for greater regional cooperation amongst relevant stakeholders in sharing best practices and developing capacity.

Additional interoperability with UN-led initiatives is important, not only from a global norms perspective (such as UNESCO’s *Recommendations on the Ethics of AI*), but also from a sectoral perspective, such as the work conducted by the United Nations Children’s Fund on AI for children and the World Health Organization on AI for public health. There is important coordination work already conducted through a ‘One UN’ approach,

⁴¹ See the original use of Agile Software Development, in the Agile Manifesto, available here: <http://agilemanifesto.org/>.

⁴² See, for example: OHCHR, *Guiding Principles on Business and Human Rights* (Geneva and New York: United Nations, 2011). Accessed at: https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf.

⁴³ Ian Bremmer and Mustafa Suleyman, “The AI Power Paradox: Can States Learn to Overcome Artificial Intelligence – Before It’s Too Late?” *Foreign Affairs*, 16 August 2023. Accessible at: <https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox>.

notably the Interagency AI Working Group, which is co-chaired by the ITU and UNESCO. In an ideal scenario, the global governance solution would involve a coordination mechanism between different UN entities, as has been done in the humanitarian sector.

Finally, there are numerous global AI governance efforts that are not driven by the UN. These include, for example, multilateral initiatives outside of the UN, as well as multi-stakeholder initiatives, such as the GPAI and the PAI. There may be a means, as with UN-led initiatives, to establish a coordination structure whereby the contribution of each major global governance effort is coordinated with those of the UN-led effort.

Decision Point 4: The Scope of Governed AI

The final decision point for policymakers is the scope of AI systems under discussion. Given parallel processes in disarmament on autonomous weapons systems, there has been a proposal that any entity tasked with AI governance focus only on *civilian AI*, leaving military uses of AI to other processes. However, rather than choosing between civilian and military AI, a possibility is to establish separate but overlapping bodies.⁴⁴ Nevertheless, many have argued that leaving arms control out of a global governance mechanism would be a mistake, given the important impact that military uses of AI could have on

human rights violations, peace and security, and sustainable development.

Additionally, the European Commission’s risk-based approach, outlined in its draft AI Act in 2021,⁴⁵ remains relevant today. According to this approach, only high-risk and forbidden uses of AI are effectively regulated, while low-risk AI systems are not. Examples of AI uses that pose an unacceptable risk include: “Cognitive behavioural manipulation of people or specific vulnerable groups: for example voice-activated toys that encourage dangerous behaviour in children; social scoring: classifying people based on behaviour, socio-economic status or personal characteristics; and real-time and remote biometric identification systems, such as facial recognition.”⁴⁶ The draft AI Act also lists eight high-risk uses of AI, such as AI used in law enforcement, education, and employment.⁴⁷ There have been criticisms of this approach, however, by feminist and other civil society organizations, who argue that risk levels might differ by population, and that what might be considered high-risk for the general population could be considered unacceptable risk for marginalized populations.

Nevertheless, a scope for global AI governance has not yet been determined. Two possible approaches and their implications are summarized below.

Dimension of AI Governance	Approaches
Civilian versus Military	It has been suggested that civilian and military uses would warrant separate but connected regimes. However, the extent to which a global governance framework would address military uses of AI should still be determined.
High-risk versus Low-risk	This may not need to be defined immediately, but the definition of ‘high-risk’ and ‘forbidden’ uses of AI would be determined by any governing effort, including how these might be experienced differently by diverse populations.

44 Ibid.

45 European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206. Accessible at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>.

46 Ibid.

47 The full list of high-risk uses of AI, according to the European Commission, reads as follows: “Biometric identification and categorisation of natural persons; Management and operation of critical infrastructure; Education and vocational training; Employment, worker management and access to self-employment; Access to and enjoyment of essential private services and public services and benefits; Law enforcement; Migration, asylum and border control management; Assistance in legal interpretation and application of the law.”

What's next for Global AI Governance?

The HLAB report recommended the creation of a Global Commission on Just and Sustainable Digitization which would be the multi-stakeholder forum responsible for agenda-setting and coordination of global digital governance, with a role going beyond AI to include all data, the Internet, and digital innovations. According to the recommendation, the Global Commission would have four core competencies: a) addressing human rights in the digital age; (b) data governance; (c) inclusive and sustainable digitalization, including universal and meaningful connectivity; and (d) a knowledge function.

There are many advantages to having an international effort address connected, current policy issues, as well as future, frontier technologies, and not solely AI. First, the global governance of AI will not occur in a vacuum, but rather will be affected by parallel governance innovations in data, the Internet, and other digital technologies. Second, an international effort such as the proposed Global Commission would provide space to anticipate technological innovation, which will likely generate unforeseen ethical and regulatory concerns in decades to come.

While AI has been at the top of policymakers' minds, it is very likely that new frontier technologies will appear over the next decade, requiring global attention. Advances in synthetic biology, for example, bypass AI completely by growing neurons directly onto circuit boards.⁴⁸ Molecular nanotechnology⁴⁹ aims

to eventually allow objects to be built from machines the size of bacteria, including skyscrapers. Quantum computing⁵⁰ is also rapidly emerging, as are activities related to the Metaverse,⁵¹ including in support of UN activities. Ethical and regulatory concerns will arise when these technologies become more widespread. Ideally, these will be able to be addressed by the governance mechanism that is put in place for AI.

In addition to a potential Global Commission, the HLAB report also proposes that the Secretary-General's Scientific Advisory Board and its network make regular contributions to the knowledge base required for resolving new technological policy challenges.

The work of the multi-stakeholder High-Level Advisory Body on AI is the most immediate venue for the UN to determine the next steps for the global governance of AI. In this sense, many of the decision points discussed in this paper may be issues debated by the Body's newly appointed 32 members, which are due to release their first report at the end of December 2023. However, it is also hoped that this paper will inform other global positions beyond the work of the Body. The shape of the global architecture for AI, the agility of policymaking, the way in which stakeholders and particularly marginalized groups take part in AI governance, the interoperability of governance regimes, and the scope of governed AI are all issues that merit both careful, and urgent, consideration.

48 Brett J Kagan, Andy C Kitchen, Nhi T Tran, Forough Habibollahi, Moein Khajehnejad, Bradyn J Parker, Anjali Bhat, Ben Rollo, Adeel Razi, Karl J Friston, "In vitro neurons learn and exhibit sentience when embodied in a simulated game-world," *Neuron* Vol 110 Issue 23 (2020).

49 Marvel Van de Voorde and Gunjan Jeswani, "Introduction: Overall vision of ethics in nanotechnology developments," *Handbook of Nanoethics* eds. Marvel Van de Voorde and Gunjan Jeswani (De Gruyter, 2021): pp. XXIII-XXVI.

50 United Nations Office of Information and Communications Technology, *A Thought Piece on Quantum Computing: Considerations and Scenarios Version 1.01* (United Nations: 2022). Accessible at: https://unite.un.org/sites/unite.un.org/files/quantumcomputingthoughtpiece_etl_final.pdf.

51 ITU, "UN Tech Agency Seeks Open and Inclusive Metaverse, 19 January, 2023. Accessible at: <https://www.itu.int/en/mediacentre/Pages/PR-2023-01-19-TSB-Focus-Group-metaverse.aspx>.

About the authors: Eleonore Fournier-Tombs is Head of Anticipatory Action and Innovation at UNU-CPR; Muznah Siddiqui is a Global Governance Research and Planning Associate at UNU-CPR; Céline Castets-Renard is a Full Professor at the University of Ottawa Faculty of Law and the University Research Chair in Accountable AI in a Global Context; Valère Ndior is a Full Professor of Law at the University of Bretagne Occidentale and a Junior member of the Institut Universitaire de France.

Acknowledgements: This paper was informed by discussions at a workshop on Global AI Governance, hosted by the United Nations University Centre for Policy Research (UNU-CPR). The workshop was organized to promote and further develop recommendations in the report of the High-Level Advisory Board on Effective Multilateralism (HLAB), and was organized by a UNU-CPR initiative, 'A Breakthrough for People and Planet: Building Momentum for the Summit of the Future and Beyond.' For more information visit: <https://unu.edu/cpr/project/breakthrough-people-and-planet>.

UNU-CPR would like to thank all those who attended the workshop and provided peer reviews and suggestions for this Technical Note: Eduardo Albrecht, UNU Centre for Policy Research; Rebecca Brubaker, Interpeace; Quintin Chou-Lambert, Office of the Secretary General's Envoy on Technology; Mia Dand, Women in AI Ethics; Adam Day, UNU Centre for Policy Research; Aman-deep Gill, UN Secretary General's Envoy on Technology; Peggy Hicks, OHCHR;

Sumaya Hussein, ILINA Program; David Michael Kelly, EOSG; Caitlin Kraft-Buchman, Women at the Table; Ievgen Klymnyk, UN DPPA; Lyse Langlois, OBVIA, Université de Laval; Emmanuel Letouzé, Data Pop Alliance; Brian Lutz, UNDP Human Development Report Office; Preetam Maloor, ITU; Tshilidzi Marwala, Rector of UNU; Omar Mohsine, UN-OICT; Giuseppe Molinaro, World Bank; Wanda Munoz, International Consultant and member of the Red de Investigación Feminista en Inteligencia Artificial; Ana Neves, CSTD; Catharina Nickel, UNU Centre for Policy Research; David Passarelli, UNU Centre for Policy Research; Gina Romero, Connected Women; Carolina Rossini, The Datasphere Initiative; Anneke Schmider, Chatham House; Alexandra Håkansson Schmidt, UN Women; Stefaan Verhulst, The GovLab; Martin Waehlich, UN DPPA; Ursula Wynhoven, ITU; Klara Wyrzykowska, EOSG; Masaru Yarime, The Hong Kong University of Science and Technology (HKUST); Li Zhou, CEB.

Disclaimer: The views and opinions expressed in this paper do not necessarily reflect the official policy or position of the United Nations University.

ISBN: 978-92-808-6609-4 © United Nations University 2023

All content (text, visualizations, graphics), except where otherwise specified or attributed, is published under a Creative Commons Attribution-NonCommercial-ShareAlike IGO license (CC BY-NC-SA 3.0 IGO). Using, reposting, and citing this content is allowed without prior permission.