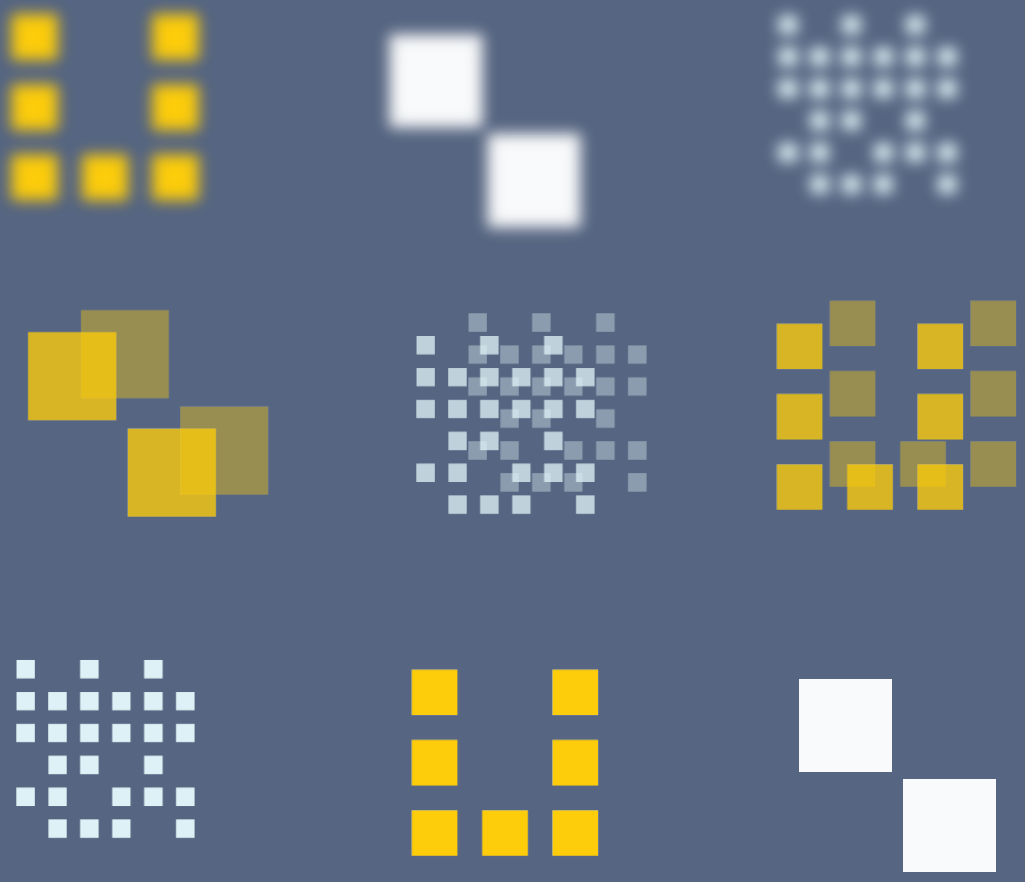
Draft report

AI Global Governance

Assessment of governance mechanisms with a human rights approach

Authored by Global Partners Digital and ECNL



April 2024

Acknowledgments

This study was undertaken by Global Partners Digital (GPD) and the European Center for Non-profit Law (ECNL). The report was written by Maria Paz Canales, Ellie McDonald, and Ian Andrew Barber.

The authors wish to thank all contributors, specifically Marlena Wisniak and Vanja Skoric, who developed the methodology for this research and draft review, and Ruby Khela, for her research and editorial support.

Executive Summary

Whilst the Interim Report produced in December 2023 by the High-Level Advisory Board on Artificial Intelligence (HLAB-AI) highlights existing institutions as examples of global cooperation and coordination, it does not provide any analysis of the effectiveness of these governance mechanisms as they relate to AI governance, or how such models would impact the enjoyment of human rights. This study aims to address this gap and support the considerations of the HLAB-AI and provide recommendations to inform their work, by employing a human rights approach. First, it analyses seventeen governance mechanisms proposed by the Interim report and by academic literature as models for an international AI governance regime. Second, it summarises lessons and mitigation measures for each of the institutional functions of international AI governance proposed by the Interim report, intending to provide insights to guide the design of the international AI governance regime.

This report concludes that there is a pressing need to further interrogate and prioritise particular functions in the Interim Report - specifically scientific research, risk monitoring and coordination, in order to facilitate collaboration, build trust and encourage knowledge-sharing amongst stakeholders. This does not however negate the importance of the other functions, but emphasises the need to build additional consensus first in order for these to be effectively established, and proposes that these functions merit further exploration within other processes such as the Global Digital Compact (GDC) or existing UN mechanisms. It also highlights the important role of any future UN AI governance mechanism to complement and reinforce national regulatory regimes on AI. Finally, it concludes by setting out a revised way forward in terms of function classification and prioritisation.

We hope that this study can serve as a starting point for other stakeholders to advocate for a human rights-based approach to international AI governance, as well as calling on the UN to conduct an in-depth ex ante HRIA, and report externally on their findings, before establishing any new governance entity or mechanism.

Contents

Context

In October 2023, the United Nations Secretary-General announced the creation of a High-Level Advisory Board on Artificial Intelligence (HLAB-AI) to support the international community's efforts to govern artificial intelligence (AI). The HLAB-AI mandate includes building a global scientific consensus on risks and challenges, helping harness AI for the Sustainable Development Goals (SDGs), and strengthening international cooperation on AI governance. The HLAB-AI issued its Interim Report in December 2023, and it is tasked to provide its final recommendations by August 2024, ahead of the Summit of the Future.

From its initial review of the AI governance landscape, the HLAB-AI Interim Report delineates a set of guiding principles to guide the formation of new global governance structures, complemented by the proposed institutional functions which those institutions should perform.

The *guiding principles* identified are the following:

- AI should be **governed inclusively**, by and for the benefit of all;
- AI must be governed in the **public interest;**
- AI governance should be built in step with **data governance** and the promotion of data commons;
- AI governance must be universal, networked and rooted in adaptive **multi-stakeholder collaboration**;
- AI governance should be anchored in the **UN Charter, International Human Rights Law**, and other agreed international commitments such as the SDGs.

In its turn, the Institutional Functions identified are the following:

- Assess regularly the **future directions and implications of AI** (independent, expert-led process that is inclusive, and with multidisciplinary assessments on the future trajectory and implications of AI);
- 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;

- Develop and harmonize **standards, safety, and risk management frameworks;**
- **Facilitate development, deployment, and use of AI** for economic and societal benefit through international multi-stakeholder cooperation;
- Promote **international collaboration on talent development, access to compute infrastructure**, building of diverse high-quality datasets, responsible sharing of open-source models, and AI-enabled public goods for the SDGs;
- **Monitor risks, report incidents, coordinate emergency response;**
- **Compliance and accountability** based on norms.

When examining alternatives for the fulfilment of the identified functions, similar to what has been done by previous academic publications,[1] the HLAB-AI Interim Report looks at existing institutions (including those governing nuclear weapons, financial systems, etc.) to identify examples of global cooperation and coordination; however, it does provide any of its analysis or evaluation of the relative effectiveness of these structures as they relate to AI governance, or how those institutional models would perform in terms of ensuring the exercise of human rights, despite this being recognised as one of the guiding principles that need to be fulfilled. This is a research gap that Global Partners Digital (GPD) and the European Center for Non-profit Law (ECNL) intend to address through the preliminary findings presented in this report.

[1] See Bak-Coleman, Joseph et al. Create an IPCC-like body to harness benefits and combat harms of digital tech, Nature, Vol 617, 18 May 2023, available at: https://www.nature.com/articles/d41586-023-01606-9; Park, Y.J. How we can create the global agreement on generative AI bias: lessons from climate justice. *AI & Soc* (2023), available at: https://doi.org/10.1007/s00146-023-01679-0 ; Hogarth, Ian. We must slow down the race to God-like AI, Financial Times, 12 April 2023, available at: https://on.ft.com/3LeOkan; Afina, Yasmin & Lewis, Patricia. The nuclear governance model won't work for AI, Chatham House, 28 June 2023, available at: https://www.chathamhouse.org/2023/06/nuclear-governance-model-wont-work-ai; Ho, Lewis et al. International Institutions for Advanced AI (2023), available at: https://arxiv.org/abs/2307.04699

Objectives

According to the context explained above, the primary purpose of this study is to assist reflections of the HLAB-AI on the potential impacts of the establishment of alternative models of governance mechanisms and inform the work towards their recommendations with a human rights approach.

While the specifics of a governance structure are not outlined in the Interim Report, it is open to the possibility that the institutional functions are carried out by a single mechanism or a network of institutions. This study includes our preliminary findings in looking into the models mentioned in the Interim Report, as well as other alternatives proposed by recent academic publications referenced above. We also critically reflect on those identified functions to provide insights into their pertinence to guide the design of AI governance from a human rights perspective and draw lessons from the institutional models pre-existent in other fields.

Finally, we are particularly interested in how different stakeholders' responsibilities will differ across institutional functions in the final recommendations to be produced by the HLAB-AI, and we believe this discussion has a reach beyond the concrete choices of the UN systems to address AI governance. Our preliminary findings in this report have further usefulness in supporting a wider range of stakeholders - including governments, international bodies, industry, academics, technical experts and civil society - to engage with AI governance discussions in order to better ensure the incorporation of the human rights considerations in institutional assessment of options moving forward in building global governance for AI.

This study does not take the form of a full Human Rights Impact Assessment (HRIA) process. The takeaways from our assessment are not aimed at recommending mitigation measures for governance mechanisms, but instead towards using the identified elements and lessons learned to tailor any future potential AI governance entity or mechanism with a human rights approach. Considering the large volume of assessed governance mechanisms and the short time framework to

produce initial findings, we have used a combination of procedural, substantive and proxy indicators in this study. Furthermore, we limited our assessment to publicly accessible information from reliable sources, in addition to interviewing several key stakeholders. The methodology for this study is outlined in full in Annex I.

Going forward, we aim to continue gathering stakeholder input to refine our findings, as well as feeding into broader discussions on the functions and modalities of global AI governance. We encourage additional research and evidence gathering related to the advantages and shortcomings of evaluated models and we welcome feedback on these initial findings. We hope that this study can serve as a starting point for other stakeholders to advocate for a human rights-based approach to AI governance. Importantly, we urge the UN to conduct an in-depth *ex ante* HRIA and report the findings externally, before establishing any governance entity or mechanism.

Methodology and scope

A detailed methodology describing the components and sources of information considered for the assessment is provided in Annex I. We acknowledge the limitations of the study as relying primarily on public sources of information and insight from a limited number of interviews with experts on some of the governance mechanisms. Given these research limitations, we welcome further input and feedback on the findings from other stakeholders with relevant experience engaging in the governance mechanisms captured in the study. We intend to provide for further public discussion an updated version of this report to capture this additional information later in 2024.

Due to time and resource constraints, we prioritised the governance mechanisms (GMs) in our assessment, based on the following criteria: a) they were mentioned in the HLAB-AI's Interim Report, or b) they have been discussed in previous academic publications. We also make reference to one additional governance mechanism that has been proposed, but is not currently established, for its role in AI governance or broader digital technology governance. Accordingly, we examine the following GMs which have been grouped according to the primary function they could play in AI governance, while noting that several of them also perform a strong secondary function. We provide further reflection on our proposed function taxonomy in the following section in order to interrogate and dialogue with the GMs proposed in the HLAB-AI Interim Report. We recognise the limitations of this approach and the value of expanding our assessment to encompass additional GMs: the addition of GMs performing the function of technical standard-setting, such as the IEEE Standards Association or the International Organization for Standardisation (ISO), is noted.

The following table outlines the GMs assessed, categorised based on our assessment of their primary function. The GMs were categorised in this manner to facilitate additional, comparative assessment amongst those GMs with a shared primary function.

Primary function	List of Governance Mechanisms (GMs)
Research development	1. European Organization for Nuclear Research (CERN) 2. European Molecular Biology Laboratory (EMBL)
Access	3. The Vaccine Alliance (GAVI)
Risk monitoring	4. Intergovernmental Panel on Climate Change (IPCC) 5. Financial Stability Board (FSB) 6. UK AI Safety Institute
Accountability	7. International Atomic Energy Agency (IAEA) 8. International Civil Aviation Organization (ICAO) 9. Financial Action Task Force (FATF) 10. Multinational Enterprises Guidelines (OECD) 11. UN Treaty bodies (Human Rights Committee/Committee on Economic, Social and Cultural Rights) 12. Universal Periodic Review (UPR) 13. World Trade Organization dispute resolution (WTO)
Coordination	14. High-level Political Forum on Sustainable Development (HLPF) 15. Society for Worldwide Interbank Financial Telecommunication (SWIFT) 16. AI Policy Observatory (OECD) 17. Digital Human Rights Advisory Mechanism (HRAM), facilitated by the Office of the United Nations High Commissioner for Human Rights proposed by the Secretary-General in his Policy Brief on the Digital Compact (Non-established)[2]

[2] As a proposal for a future mechanism which may be influential in determining the future international AI governance regime, we consider it valuable to include HRAM within this assessment. However, because HRAM isn't yet established, it is not possible to assess its functioning as a Governance Mechanism, which is the first part of this assessment. Rather, HRAM is assessed in terms of the contribution it could make to the proposed institutional functions, in the second part of this assessment, based on the information provided to date.

Assessment of governance mechanisms (GMs) with a human rights approach

In this section, we provide an overview of the positive impacts and key concerns of each GM, consisting of a brief assessment of their substantive impacts on human rights (substantive indicators), and how the GM's structure and functions impact the exercise of human rights (procedural indicators). This overview aims to elucidate the cause of particular impacts, to identify lessons and mitigation measures to inform the design of future UN AI international governance with a human rights approach. This overview is underpinned by a full assessment of each GM, described in full in the methodology in Annex I.

European Organization for Nuclear Research (CERN)

Positive impacts on human rights: This GM has a particular impact on economic, social and cultural rights (ESC) through contributing to scientific advancement and fostering development. CERN also provides education and training for researchers and students from all around the world promoting the right to education. The technology transfer and sharing of knowledge can have a positive impact on the right to development and access to the benefits of technology.

Negative impacts on human rights: There is a potential negative impact on health and safety risks due to CERN's high-energy particle accelerator, which has raised concerns about the right to health and the right to a safe and healthy working environment. These risks have led to unsuccessful court challenges. There are also broader concerns about misuse of scientific technology or dual-use technologies, which might have implications for the right to life and security of the person.

Other impacts and commentary: CERN is celebrated in many respects as a mechanism that is able to overcome gridlock, which has been seen to be due to its governance model that is isolated from external influence by particular governments or commercial entities, and instead grounded in independent decision making. Moreover, CERN has a clear mission and overall aim of supporting science and fostering global collaboration. Specifically, its founding Convention mandates that CERN shall have no concern with work for military requirements, and it is committed to the principle of open access to its research findings. This is further supported through dedicated peer review processes, public engagements and partnerships, which ultimately drive transparency and credibility. However, there does not appear to be a means for stakeholders, including civil society, to engage in decision-making.

Recommendations:

- CERN should be considered as a model for UN AI governance where research is to be the primary focus as its governance structure allows for positive impacts on human rights and the benefits for the international community to be maximised, whilst mitigating risks through transparency and effective collaboration.
- CERN has significant partnerships with international and regional organisations, as well as industry bodies and the private sector that should be emulated. However, its engagement with civil society appears to be limited to education and capacity building. Instead, for UN AI governance, there should be mechanisms which allow for CSOs to have some means of providing input to decision-making.
- CERN outputs and research are made public for international use, but they also support innovation with commercial value which can be licensed to member states. This should be considered carefully when adapting to AI-focused governance to ensure the ongoing benefits of open science and availability of information for the global majority, whilst still maintaining an incentive for private sector or specific member state engagement.

European Molecular Biology Laboratory (EMBL)

Positive impacts on human rights: because of its research focus, like CERN, its positive impacts on human rights are centred largely on economic, social and cultural rights. EMBL contributes to scientific advancement which can help public health, development and the right to benefit from scientific progress. Its education and capacity-building work can also support the right to education.

Negative impacts on human rights: As EMBL deals with bio research it could theoretically have a potentially negative impact on the right to life, health or other aspects that stem from the misuse of research leading to bioterrorism or warfare.

Other impacts and commentary: When compared to CERN, the EMBL does not have as many international partnerships and seems to be somewhat less global in nature with the inclusion of states and actors from the Global Majority. It also does not allow for civil society or other stakeholders to engage in decision-making and engagement seems limited to education and some research outreach programmes. Access to the facilities requires applications and self-funding, which inherently limits access.

Recommendations:

- If EMBL is to be used as a research-focused model for UN AI governance, it would be helpful to look instead to CERN for its governance approach, outreach and international partnerships, which appears to provide for more global engagement.
- If EMBL is to be used as a research-focused model for UN AI governance, it would be helpful to have less onerous requirements to access facilities - dedicated funding streams, as well as more engagement with civil society and other relevant stakeholders. While some funding for education and outreach is positive, it should not be a one-way street. This will naturally skew to providing access for the global north of better financially supported entities.

The Vaccine Alliance (GAVI)

Positive impacts on human rights: GAVI contributes to the right to health through vaccine access but also by supporting the strengthening of public health systems. GAVI's clear emphasis on monitoring to ensure accountability increases the perceived cost-effectiveness for donors who not only want to support access to vaccines but who are also committed to supporting national health systems.

Negative impacts on human rights: Gavi has been criticised for the lack of a holistic approach to health systems strengthening by focusing on technical solutions and disease-specific approaches. This can have a disruptive impact on health systems as part of overall social and economic development, which is particularly critical for the developing countries that are recipients of Gavi programs. It has been reported that GAVI has had mixed results at addressing between-country inequities in the utilisation of immunisation services, and it has only more recently put greater emphasis and resources towards addressing within-country inequities in the utilisation of immunisation services.

Other impacts and commentary: The public-private partnership nature of Gavi provides a unique model of collaboration toward more equitable access to a critical resource for health. However, this has also been criticised for the decision-making related to vaccine privileging commercial consideration by supporting investment in new more expensive vaccines rather than wider access to vaccines already developed and cheaper. Despite the criticism, GAVI is often considered a more 'trustworthy' alternative to the traditional, publicly mandated multilateral UN agencies, including the WHO. Gavi has strong transparency policies and information access, it has also implemented a grievance mechanism that allows reporting concerns of wrongdoing related to its funded programs.

Although GAVI is perceived by most partners as being flexible and open to feedback on major issues, direct beneficiaries do not have a direct voice in the governance structure. However, GAVI actively collaborates with civil society organisations, including with a permanent seat on its

Board and participation in a number of advisory bodies and task teams. It has been also reported that GAVI considers feedback from partners when revising its policies and programs.

Recommendations:

- If Gavi is to be used as an access-focused model for UN AI governance, it will be interesting leveraging the concept of investing in strengthening of the governance systems locally rather than exclusively focusing on technology access and transfer.
- If Gavi is to be used as an access-focused model for UN AI governance, the strong focus on transparency and accountability to ensure its support at country level is adequately managed according to the programme objectives as specified in the country agreements, it will be an element to replicate.
- If Gavi is to be used as an access-focused model for UN AI governance, the governance structure that allows for a multistakeholder composition including civil society should be considered as an element to be replicated. However, additional attention should be given to ensuring that the direct beneficiaries of the technology access have a way to engage with the governance structure.

Intergovernmental Panel on Climate Change (IPCC)

Positive impacts on human rights: the IPCC is the UN body that gathers information on the state of scientific, technical and socio-economic knowledge on climate change. As such, it has an impact on the exercise of the rights to life and security.

Negative impacts on human rights: In the IPCC work there is an increasing tension between the focus on physical science evidence, and the focus on adaptation for countries that can be more adversely impacted by climate change. The harms are unevenly distributed between those who can focus on course correcting of global warming and those who need to urgently deal with the negative impacts.

Other impacts and commentary: The work is organised around assessment cycles that last around 5 years. The IPCC does not make concrete policy recommendations, but rather presents the state of the science and projections. The Panel's reports are intended to be policy relevant but not policy prescriptive. They provide key inputs into international climate change negotiations. The experts participate in the review process in their personal capacity and not on behalf of the states or other observers. There are still challenges in the diversity of experts that are engaged in providing input and review to the assessment reports with a majority representing the Global North, and gender diversity is still low. Among the challenges reported by Global Majority scientists are the limited research funding to allow their engagement with the assessment cycles, followed by limited technical capacity, and inadequate institutional support. There is a perception that the Panel's reports are more the result of a political consensus than evidence-driven. There is a procedure for investigating, and if necessary, correcting alleged errors in its published reports.

Recommendations:

- If IPCC is to be considered as a risk monitoring-focused model for UN AI governance, it should preserve and strengthen from this model the ability to engage a diversity of scientific experts across the globe in the process of gathering information. A particular effort should be made to increase the Global Majority and gender diversity participation. It should also preserve the ability of relevant stakeholders to engage as observers.
- If IPCC is to be used as a risk monitoring-focused model for UN AI governance, it should replicate the focus on both the hard science scientific evidence and the social impacts that are part of the risk to be measured in the deployment of technology. Those two elements should be studied with a focus on how unevenly distributed risks can be addressed from the international cooperation perspective between countries that are technology producers and those who need to deal with the negative impacts created by technologies developed somewhere else.
- If IPCC is to be used as a risk monitoring-focused model for UN AI governance, given the fast pace of technology evolution it should consider assessment cycles that are more in line with this reality.

Financial Stability Board (FSB)

Positive impacts on human rights: Even if the FSB mission is to promote the stability of financial markets, the impact of the FSB's work falls well beyond the financial sector impacting the exercise of economic, social and cultural rights and arguably democratic stability of countries.

Negative impacts on human rights: The mechanism lacks a human rights approach. Currently, FSB participants are mostly from the Global North which poses a question in terms of the fulfilment of its objectives around financial stability at the expense of the ESC rights exercised in developing countries which might require a more holistic approach and greater participation from other stakeholders in the mechanism decision-making.

Other impacts and commentary: Members of the FSB commit to pursue the maintenance of financial stability, maintain the openness and transparency of the financial sector, implement international financial standards (including the 15 key International Standards and Codes), and agree to undergo periodic peer reviews. FSB coordination includes international standard-setting bodies and regional bodies like the European Central Bank and the European Commission. The FSB regularly reports to the G20 which regularly endorses the FSB's policy agenda and supports implementation of agreed international standards. The coordination mechanism allows communication directly with national financial authorities shielding them (to some extent) from political pressures. The FSB's decisions are not legally binding on its members, instead, the organisation operates by peer pressure to set internationally agreed policies and minimum standards that its members commit to implementing at the national level.

Recommendations:

- If FSB is to be used as a coordination model for UN AI governance, the coordination participants should be much more inclusive than the model, including Global Majority countries and consultation with societal groups beyond the private sector, as well as national non-sectoral officials.
- If FSB is considered a coordination model for a UN AI governance, it will be useful to retain the ability to connect directly with sectoral

authorities at the national level (to shield from local political pressures), and to have the ability to communicate and endorse its recommendations by other multilateral bodies that can increase the voluntary commitment with them.

UK AI Safety Institute (AISI)

Positive impacts on human rights: AISI has been created with the goal of mitigating risks of advanced AI systems by developing and conducting evaluations to understand their capabilities, safety and security of systems, and assess their societal impacts. This risk mitigation action would have a positive impact on the exercise of civil and political rights, and economic, social and cultural rights that are impacted by the use of advanced AI systems. Defence and national security applications of advanced AI systems are considered in scope for its work.

Negative impacts on human rights: There is no explicit commitment to use the impact on human rights exercise as a benchmark integrated into the evaluation of risks of advanced AI systems. This is problematic as it could lead to an over-reliance on safety and security concerns at the expense of other impacts or rights, which would be contrary to international human rights obligations.

Other impacts and commentary: AISI seems to combine scientific information gathering and the development of its own safety research. The information-sharing channels consider interactions with national and international actors, such as policymakers, international partners, private companies, academia, civil society, and the broader public. However, there was very limited access for selected civil society groups during the first AI Safety Summit. The recent creation of the AISI has not provided an opportunity yet to clarify if its role will lean more towards performing direct safety assessments of AI advanced systems or becoming a standard setter on how the assessment should be performed. Among the principles and procedures embraced by AISI there is an explicit commitment to the open participation of experts across a breadth of views and a diversity of backgrounds. The reports are mandated to clearly state where expert consensus exists or acknowledge disagreement in the wider expert community, and present the debate in an objective manner.

The reports should acknowledge the limitations of evidence, and they are not intended to make policy recommendations.

Recommendations:

- The recent constitution of AISI makes it difficult to identify the full set of characteristics that could be leveraged as components for the UN AI governance, however, the risk monitoring and research functions are essential for the international governance of AI. A mix of the elements from AISI and IPCC could be considered as an option.
- If AISI is to be used as a risk monitoring-focused model for UN AI governance, there should be measures taken to ensure the inclusive participation of scientific and social science experts from a wider range of geographic representation.
- If AISI is to be used as a risk monitoring-focused model for UN AI governance, it should be considered how to enhance its ability to establish standards for safety assessment, which incorporate a human rights approach into its risk assessment, so that they can be widely adopted and applied for institutions beyond the direct assessment of AISI.
- If AISI is to be used as a risk monitoring-focused model for UN AI governance, consideration should be given to how the mechanism can have ready access to the information on the functioning of advanced AI owned by private companies, and how research access under privacy and commercial confidentiality could be granted to experts. Effective information sharing requires a trusted actor with deep connections across all parts of the AI ecosystem. There is currently a lack of clear channels for developers of advanced AI to share information with governments. Competition laws and sensitivities around IP can meanwhile limit information sharing between firms. The AISI could act as a trusted intermediary, enabling responsible dissemination of information as appropriate.

International Atomic Energy Agency (IAEA)

Positive impacts on human rights: the IAEA's work on nuclear safety standards and security measures, particularly its work on nuclear safeguards and non-proliferation, can be said to minimise risks of nuclear accidents and radiation exposure. This directly supports the right to life, the right to health and the right to a healthy environment. The IAEA also supports the use of nuclear technology in development assistance and access, and in doing so supports the right to development and access to benefits of scientific progress. It also supports healthcare via the PACT programme which can help improve access to healthcare and support the right to health.

Negative impacts on human rights: There has been significant criticism that the IAEA's work on nuclear standards and safety may still result in negative impacts on health and the right to a healthy environment due to the inherent nature of nuclear activities, waste disposal and other impacts. The inability to deal with nuclear incidents in an appropriate manner may also have differentiated and disproportionate impacts on vulnerable groups.

Other impacts and commentary: the IAEA's governance structure, run by the Board of Governors and composed of member states, skews very heavily towards the Global North and existing nuclear powers. There has been criticism that there is little oversight by civil society or impacted communities as well. There has been commentary on the IAEA in terms of its overall mandate - that there may be a fundamental conflict in promoting the peaceful use of nuclear energy whilst trying to advance non-proliferation. It is an accountability mechanism but there is little to no enforcement as the IAEA cannot critique member states' nuclear power industries - it can create standards but not enforce them.

Recommendations:

- The IAEA, if used as a model for UN AI governance, needs to be seriously critiqued for its broad ranging and potentially contradictory mandate. It would be difficult for example, for a new mechanism to on one hand create binding safeguards on AI with specific countries

for potentially dangerous uses of AI, whereas advocating for increased use of AI for the rest of the world.
- The IAEA, if used as a model for UN AI governance, would require a revision of its governance structure to not skew so heavily global north and include some means of multistakeholder participation or oversight.
- The IAEA, if used as a model for UN AI governance, would need to have some flexibility in terms of reacting to global events - the equivalent of an AI Fukushima event- in a rapid and non-cumbersome manner.

International Civil Aviation Organization (ICAO)

Positive impacts on human rights: the ICAO activities include the creation and recommended standards of air navigation, and facilities procedures for air accident investigations. Its overall activities focusing on safety can have positive impacts on the right to life, safety and security and the right to freedom of movement. Its investigations can help with the right to remedy. It also has policies on environmental protection that can bolster the right to a healthy environment.

Negative impacts on human rights: by promoting global aviation the ICAO implicitly can negatively impact the right to health and healthy environment due to emissions. Also, aviation security measures can have a negative impact on the right to privacy and freedom of movement through the collection and sharing of information, which can facilitate surveillance. There are also concerns about the ICAO's previous efforts to block online accounts which discuss political issues as they relate to the ICAO.

Other impacts and commentary: There has been some positive commentary on the ability of stakeholders to participate in assembly meetings, but this is limited to aviation-related organisations as opposed to broader multi-stakeholder engagement. Some positive commentary on adaptability, for example, during COVID, but also potential criticism about funding and political manoeuvring re. USA and Taiwan.

Recommendations:

- If ICAO is used as a model for UN AI governance, some elements could be retained or emulated. This includes flexibility - particularly the ability to adapt to changing global circumstances such as COVID-19, as the AI governance will likely need constant adaptation of its standards or policies to address new challenges.
- If ICAO is used as a model for UN AI governance, it would need to ensure lessons learned from political tensions between countries that play different roles in AI development, provisions and use, ensuring that policies are set out clearly
- If ICAO is used as a model for UN AI governance, would be potentially helpful to consider overall funding options and have them more dispersed as currently USA provides ⅛ of funding on top of experts and significant voluntary contributions. Governance structure would need to mitigate against improper influence.

Financial Action Task Force (FATF)

Positive impacts on human rights: the FATF's main activities involve the creation of standards and recommendations to help prevent and combat financial crimes such as money laundering and financing of territory organisations. These have implications for the right to life, right to safety and security and right to own property.

Negative impacts on human rights: the FATF's activities, including its creation of standards, coordination and actions taken by particular countries may have negative consequences for human rights, namely freedom of expression, assembly participation and association. Their requirements for anti-money laundering and combating the financing of terrorism (AML/CFT) can have unintended consequences where financial services deny access to financial services and transfer with customers or regions they perceive as high risk - disproportionately impacting vulnerable groups and CSOs. Burdensome requirements on individuals, businesses and CSOs that can negatively impact freedom of association, privacy, and expression.

Other impacts and commentary: The FATF's evaluation process report has been criticised for a lack of transparency, accountability or public consultation. There are no means of impacting third parties to seek redress or broader CSO engagement.

Recommendations:

- Would not recommend that FATF be used as a model for UN AI governance. This is for several reasons including: (1) lack of accountability and transparency - the accountability body provides little accountability and transparency over its decision-making processes and recommendations; and (2) there are significant negative impacts on human rights and an absence of safeguards for redress.
- If used as a model for UN AI governance, it would need to have a distinct focus on human rights overall and better balance security and human rights. The FATF's focus on AML/CFT objectives does not consider human rights, which has led to an over-securitised approach to law enforcement interest over individuals' human rights.
- Criticism over excessive compliance burdens needs to be addressed and the costs may outweigh the benefits - particularly for smaller entities or countries. The FATF doesn't adequately consider the impacts on Global Majority countries. Any UN AI governance model should carefully consider compliance costs of actions taken, and assess the effectiveness and consequences alongside human rights considerations.

Multinational Enterprise Guidelines (OECD)

Positive impacts on human rights: The OECD Multinational Enterprise Guidelines are enforced through the National Contact Points, which are established by governments to promote the guidelines and handle cases against companies via non-judicial grievance mechanisms - known as specific instance procedures. These can have positive impacts on human rights as they provide access to remedy for individuals or other stakeholders impacted by human rights violations. This promotes corporate accountability and adherence to the OECD Guidelines.

Negative impacts on human rights: The specific instance procedures do not seem to have a distinct negative impact on human rights, only that they fail to live up to the rights it intends to safeguard against the broader right to remedy. The mechanism lacks a specific human rights-based approach.

Other impacts and commentary: The NCPs appear to be positive in that they provide a soft means of accountability and have handled more than 500 cases since 2000. However, the extent of their impact is difficult to assess as some NCPs struggle from lack of visibility, lengthy proceedings and divergent approaches. Commentators have suggested that they could be improved through further resources, stronger institutional arrangements, additional access to expertise, and coordination between NCPs including conditions to accept cases.

Recommendations:

- If the OECD NCPs are used as a model for the UN AI governance that aims to provide accountability, we would recommend that it has stronger enforcement beyond voluntary means, which would bring about more effective change to corporate conduct. It might also be helpful to have such a mechanism targeted at both the private and public use of AI.
- If the OECD NCPs are used as a model for the UN AI governance, it should ensure a more uniform approach to case handling and better access to expertise and centralised leadership. This will ensure heightened coordination and avoid divergent approaches. Centralised leadership is important to avoid bias and conflicts of interest between national authorities and private companies in case of disputes.
- If the OECD NCPs are used as a model for the UN AI governance, it should be given appropriate resources to ensure cases are dealt with in a prompt and effective manner.

Human Rights Committee (HRC) and Committee on Economic, Social and Cultural Rights (CESCR)

Positive impacts on human rights: The HRC and the CESCR contribute to the effective realisation of the rights under their respective Covenants through monitoring and supervising the laws, policies and practices of state parties. Each of their core activities (the assessment of reporting by state parties; examination of individual complaints; development of interpretive guidance in the form of General Comments, etc.) is designed to ensure the respect, protection and fulfilment of Covenant rights.

Negative impacts on human rights: Like the UPR, the HRC and CESCR do not have a "negative" impact on human rights as much as they may experience challenges in executing their mandate to contribute positively to the realisation of rights under their respective Covenants.

Other impacts and commentary: The HRC and CESCR have received praise for their execution of their mandate and for ensuring a predictable, open, and relatively accessible process for civil society participation, where civil society inputs can influence the outcomes of their activities. While there are some linguistic, financial and infrastructural barriers to participation, the Committees have taken some steps to mitigate them. The Committees rely on a constructive dialogue with states to execute their oversight function, which limits their effectiveness and results in insufficient compliance by States. Other challenges included a backlog of individual communications and urgent actions, funding limitations, and diverging working methods among the treaty bodies which limits their capacity to coordinate and exchange expertise.

Recommendations:

- If the HRC and CESCR are used as a model for the UN AI governance, it should retain their focus on supervising the laws, policies and practices of States in respect to their design, development, deployment, evaluation and regulation of AI systems according to the international human rights law framework. The existing work of the treaty bodies to develop guidance for the interpretation of States' obligations under the Covenants would be a

particularly helpful model with respect to UN AI governance envisaged norm-development and enforcement function.

- If the HRC and CESCR are used as a model, the UN AI governance should seek to replicate aspects of their processes for civil society engagement, particularly the predictability of their processes, openness, and accessibility (through partially inclusive meeting formats, e.g. hybrid participation and multiple languages). The predictable nature of their processes which provide for numerous, recurring opportunities for stakeholder engagement, and where civil society input can meaningfully drive the outcomes, is particularly commendable, although more steps should be taken to support financial participation and to ensure accessible modalities.
- If the HRC and CESCR are used as a model, the UN AI governance should strive to ensure a greater degree of independence from the UN system and flexibility to adapt its processes and activities where necessary to respond to new and emerging challenges, and adaptive technologies. The Committees receive their funding from the UN Human Rights income which is a very small proportion of total UN funding. The treaty body strengthening process has also demonstrated some of the challenges of reforming the Committees due to their positioning within the UN system, where reforms may be subject to agreements within appropriate intergovernmental processes (e.g., via resolutions agreed among states via the Human Rights Council or the General Assembly). This can at best delay and at worst stymie the execution of necessary reforms.
- In a similar vein, if the HRC and CESCR are used as a model, the UN AI governance should not replicate their process for electing independent experts, where states nominate representatives. This process has been criticised by civil society for ignoring relevant eligibility criteria and resulting in vote-trading among states.

Universal Periodic Review (UPR)

Positive impacts on human rights: the entire UPR process is designed to have a positive impact on all human rights and improve the situation in each country via assessment, review and the provision of technical assistance and guidance via recommendations.

Negative impact on human rights: the UPR process does not have a "negative" impact on human rights as much as it doesn't fulfil its promise to positively impact human rights.

Other impacts and commentary: The UPR is on one hand celebrated as a dedicated mechanism for the protection and promotion of human rights. It involves significant transparency, openness and active participation from CSOs and other stakeholders. However, it has been criticised for its cumbersome apparatus, labour-intensive and costly manner of operation. There is also criticism over the lack of action on non-compliance overall - states can simply ignore recommendations, or fail to provide information leading to a distorted reality. There is technically a means of discussing cases of persistent non-cooperation within the mechanism but this has never been defined. There is also limited follow-up and implementation as a lack of resources can hinder recommendations into account, as well as there being political manoeuvring where states engage in alliances.

Recommendations:

- If the UPR is used as a model for the UN AI governance, it would be helpful to retain its focus on the improvement of human rights as opposed to security, ethics, etc. However, political and civil rights should be given equal consideration to economic, social and cultural rights.
- If the UPR is used as a model for the UN AI governance, it should emulate its focus on transparency and openness for stakeholders to engage and provide expertise, including from the Global Majority, and continue to have dedicated funds to support this engagement.
- If the UPR is used as a model for the UN AI governance, it should ensure that its openness and transparency do not lead to an overly cumbersome, timely and expensive means of providing

accountability, which is not supported through follow-up and little means to act on non-cooperation or compliance.

- If the UPR is used as a model for the UN AI governance, it should be able to address new and emerging human rights challenges. This could involve updating methodologies or thematic priorities when undertaking assessments for more dynamic evaluations in the AI context.

World Trade Organisation dispute resolution (WTO)

Positive impacts on human rights: The WTO's goal is to ensure that trade flows as smoothly, predictably and freely as possible which is considered to contribute to economic growth and human development. This is indirectly linked to the realisation of economic, social and cultural rights.

Negative impacts on human rights: The mechanism lacks a human rights approach. The proceedings are confidential between the states in dispute even when private parties are directly concerned, they are not permitted to attend or make their own submissions affecting the right to participation. This simplifies the functioning of the procedure but does not provide the opportunity for impacted groups to be part of a dispute settlement process.

Other impacts and commentary: The mechanism provides only a soft mechanism of accountability. However, the dispute resolution mechanism has been useful to avoid unilateral actions in trade. Currently the two-tier mechanism with a panel phase followed by an appeal before an Appellate Body is paralysed by the lack of appointment of the Appellate Body members. This situation shows the mechanism's weakness regarding geopolitical considerations that take over the established rules and conflict resolution mechanisms.

WTO provides developing countries with the possibility to claim 'special and differential treatment' consisting of more favourable terms or extra time to fulfil their commitments, however, approximately two-thirds of WTO members claim developing-country status which has been criticised by other countries looking to establish more objective indicators.

Recommendations:

- It is difficult to foresee what could be the utility of the WTO dispute resolution model as an accountability mechanism for the UN AI governance, but eventually, the already existent mechanism could be used to settle disputes arising from the different regulatory approaches adopted for AI systems regulation across different countries that could be deemed barriers to trade. In this sense, the possibility of providing flexibility on trade rules related to technology transfers is interesting.
- Current criticism of the functioning of the dispute resolution mechanism is linked to the lack of ability to ensure enforcement of decisions, some of the proposed reforms for the mechanism create incentives for countries to implement them through technical cooperation but also through sanctions in case of lack of compliance over the time. These recommendations should be integrated if this mechanism is considered as part of the UN AI governance.

High-Level Political Forum (HLPF)

Positive impacts on human rights: The HLPF, through the Voluntary National Review (VNR) process, provides a framework for review and follow-up on the implementation of the sustainable development goals and targets. Given that aspects of the SDGs correspond to states' international human rights law obligations, it is possible for states and non-governmental stakeholders to monitor their adherence to those obligations, which could result in positive human rights impacts. The HLPF can reaffirm the application of international human rights law in its political and ministerial declarations.

Negative impacts on human rights: The HLPF through the VNR process does not adequately incorporate an assessment of compliance with international human rights law by states in their implementation of the sustainable development goals and targets.

Other impacts and commentary: There has been criticism of the HLPF's ability to integrate with and draw in inputs from the rest of the UN human rights system, and the lack of meaningful coordination with UN human

rights bodies is frequently highlighted. There have been frequent criticisms of the modalities for engagement with the VNR process by non-governmental stakeholders, including the lack of formal recognition of civil society reporting, the lack of time for consideration of state reporting which limits the effectiveness of the review, which has led some to refer to the process as toothless or ineffective. The positioning of the HLPF within the UN and its structure means that its ability to adapt its governance is subject to intergovernmental negotiation by states.

Recommendations:

- Would not recommend that the HLPF be used as a model for the UN AI governance. If it is used as a model there are aspects which should be revised. These include the HLPF's vast mandate, but limited resources, its lack of decision-making power and its structuring as an intergovernmental process which means that its outcomes and governance structure are decided through intergovernmental negotiations. This provides it with limited ability to adapt its operations and activities and means that its decisions are consensus-based, which significantly constrains its mandate.
- While the universal nature of the HLPF's VNR process, and the formal recognition of different stakeholder groups is commendable, the lack of a formal role for non-governmental stakeholders in the reporting process should not be replicated.

Society for Worldwide Interbank Financial Telecommunication (SWIFT)

Positive impacts on human rights: SWIFT can support the exercise of the right to privacy and data protection by providing a secure platform to send and receive financial transactions. Through specifying and publishing rules and best practice guidance (SWIFT standards) on how to comply with applicable regulations and standards (which include the domestic laws in the countries in which it operates), SWIFT has the potential to positively impact human rights protection.

Negative impacts on human rights: SWIFT has contributed through its operations to severe breaches of the right to privacy and data protection

by facilitating the transfer of data without adequate transparency and effective control mechanisms, and in violation of the principles of proportionality and necessity. SWIFT has also been required to disconnect certain banks from its international messaging system to comply with sanctions regimes. The denial of access to financial services could negatively impact the realisation of economic, cultural and social rights.

Other impacts and commentary: SWIFT has received praise for its success in facilitating commercial flows, and is the dominant messaging service through which financial transactions are sent and received. It has developed a sophisticated governance architecture to provide for proportionate representation by its shareholders in each country. This system has also resulted in skewing SWIFT's leadership towards the G-10 countries. SWIFT does not provide any process for engagement with external stakeholders.

Recommendations:

- If SWIFT is considered as a model for the UN AI governance, there are certain positive aspects of SWIFT's governance which could be adapted and warrant further exploration. However, elements of SWIFT by-laws relating to the representation of shareholders in the Board of Directors should not be replicated as these have resulted in the dominance of G-10 countries within SWIFT's leadership. This is in contrast to one of the core aims of the UN AI governance as stated in HLAB's Interim Report, which is to explore forms of governance which allow for "universal buy-in by different member states and stakeholders".
- Relatedly, SWIFT's lack of any process for engagement with external stakeholders is not recommended. SWIFT provides for no external stakeholder engagement, publishes limited information publicly, and has no mechanism for complaints.
- Overall, there is limited publicly available information about SWIFT which makes it difficult to assess how the GM can be useful for UN AI governance. As stated in HLAB's Interim Report, SWIFT can "offer inspiration and examples of global governance and coordination" which is related to SWIFT's ability to evolve existing legal, financial and technical arrangements. However, given that SWIFT has no underpinning in human rights, no possibility for engagement by

affected communities, nor any mechanism for redress, it cannot be recommended based on this assessment.

AI Principles and AI Policy Observatory (OECD)

Positive impacts on human rights: The OECD's AI Principles aim to shape a human-centric, "values-based" approach to AI. While they intend to promote the use of AI which respects human rights, human rights is not their sole or principal frame of reference. The AI Policy Observatory contributes through its research function to monitoring the implementation of human rights-respecting frameworks, which could result in positive impacts. As part of its work to oversee the implementation of the AI Principles, the Policy Observatory undertakes data-collection and research on national laws, policies and oversight of AI. This includes an assessment of the extent to which national frameworks reflect international human rights law and standards.

Negative impacts on human rights: As noted above, the AI Principles do not make human rights their sole or principal frame of reference. By taking an ethical or a human-centric framing, the AI Principles risk undermining adherence to the international human rights law framework, which has been universally agreed to by states and has generated a wealth of guidance on the interpretation of the framework to new and emerging technologies, as well as systems for monitoring and review of compliance, and remedy.

Other impacts and commentary: The AI Policy Observatory is a multi-stakeholder forum which oversees the promotion and implementation of the AI Principles, convenes stakeholders and undertakes and publishes research. The research function of OECD's AI Policy Observatory has received praise for "building consensus on AI opportunities and risks"[3] and it has been highlighted as a model that could be replicated for UN AI governance.

[3] Ho, Lewis et al. International Institutions for Advanced AI (2023), available at: https://arxiv.org/abs/2307.04699.

Recommendations:

- The OECD's AI Principles were one of the nine AI governance initiatives which formed part of the survey and gap analysis undertaken by HLAB-AI for their Interim Report. If the AI Principles are to be considered as a model for the development of principles of international AI governance, forming the basis of the work of the UN AI governance, it should ensure that they are derived from and refer to international human rights law.
- If OECD's AI Policy Observatory is considered as a model for the UN AI governance, its role in providing definitional clarity and taxonomy - which serves to build consensus and can be used as the basis for future policymaking - should be replicated. For example, the OECD's definition of AI has been updated and is regularly cited. The UN AI governance would need to perform a similar function by providing definitional clarity in the case of particularly thorny issues, for example, providing definitions to aid understanding of adaptive technologies, or identifying the norms engaged in particular use cases.
- The intergovernmental and multi-stakeholder nature of the AI Policy Observatory merits further study and could serve as inspiration for the UN AI governance. In its Interim Report, HLAB-AI noted the value of the intergovernmental fora provided by institutions like the OECD in "reinforcing interoperability and regulatory measures across jurisdictions". Its multi-stakeholder nature provides a vehicle for different stakeholders to contribute to its research and monitoring outputs. However, if this model is used, revisions should be made to the process for electing representatives, including limiting the degree of state influence and providing for greater representation of stakeholders from the Global Majority. It should also be clarified how stakeholders participate in its decision-making processes and influence its outcomes.
- The AI Policy Observatory undertakes data-collection and research on national laws, policies and oversight of AI, however, it lacks a formalised process for the review and implementation of its AI Principles. If the UN AI governance seeks to clarify global norms and principles - based on existing frameworks - then this norm-making should be accompanied by a clear and cyclical process for monitoring and reviewing adherence to these norms. As noted

elsewhere, this process should ideally be located within or built upon existing mechanisms which have demonstrated their capacity and expertise to oversee the governance of digital technologies, including the UN human rights mechanisms.

The role of institutional functions in the AI international governance regime

The HLAB-AI Interim Report presents a visually engaging pyramid of seven functions to be considered as potential elements covered by the UN AI governance system. With the purpose of contributing to the HLAB-AI work and based on our initial findings on the GMs examined, we interrogate the proposed functions to determine their pertinence, the risks or opportunities that may arise for the exercise of human rights, as well as observations on which functions should be prioritised. This involves a critique of the taxonomy itself in an attempt to provide clarity over the functions' scope and potential operation, as well as commentary on their potential implementation over time. We believe this will prove helpful in guiding the development of an effective global AI governance framework.

Institutional Function 1: Horizon scanning, building scientific consensus

The HLAB-AI Interim Report provides as part of this function the task of continuously gathering evidence and assessing from a scientific perspective the future directions and implications of AI. Although there is a reference to "building scientific consensus", the experience from IPCC and the recently created AISI make us inclined to recommend that this function leans more toward the value of gathering information from a diverse set of stakeholders and making it widely available, rather than necessarily achieving consensus. The IPCC model of issuing policy relevant but not policy prescriptive advice seems to be at the core of implementing this function, as well as the AISI commitment to being explicit where expert consensus exists or acknowledging disagreement in the wider expert community, and presenting the debate in an objective manner. The consensus adoption of IPCC assessment reports has been precisely a point of criticism as it provides room for geo-political considerations and pressures from states to exclude reference to relevant

scientific evidence or potential impacts on human rights that is gathered during the assessment cycle.

In contrast, CERN has been lauded for its focus on international scientific collaboration and open science policy, which may indirectly support the development of consensus without being marred by political considerations or the need to arrive at policy recommendations. Decision-making at CERN depends on the specific context or structure involved. The key to these more dynamic efforts is that they are grounded in a focus on knowledge sharing across a broad number of stakeholders, and they are decoupled from the creation or emergence of common responses. This approach may better promote the right to benefit from scientific advancement and its applications. This could be implemented in a distributed manner leveraging academic research, but also industry and public research. Such models are more appropriate to consider for this function when compared to those whose outcomes are subject to consensus-based outcomes, as these can result in stasis – particularly where consensus is interpreted as unanimity – and in the predominance of state views in the case of multilateral processes.

It is not clear from the Interim Report how this function will adequately address the need for research-focused capacity building which enables widespread participation by external stakeholders, and particularly those from the Global Majority. This is important to capture with respect to UN AI governance as it will provide a more level playing field and could bolster efforts at the progressive realisation of economic, social and cultural rights. The IPCC work specifically demonstrates that there are relevant challenges for the Global Majority to support their ability to meaningfully contribute to scientific information gathering. This function should therefore account for the task of enabling that participation through capacity building and funding support for Global Majority engagement. If not, it will be unable to fulfil the desired goal of "drawing on expertise and sharing knowledge from around the world".

The shortcomings of many of the examined GMs include limited inclusivity of stakeholders, particularly CSOs and impacted groups in information gathering and decision-making. The mechanism established for fulfilling this function should consider in its design ensuring wide room for engagement by a broad set of non-state actors, including the private

sector, academia and civil society, who should not only be able to share information but also have some role to influence the governance of the mechanism itself. This could, for example, take the form of consultations that enable stakeholders to provide relevant insights on the actions or direction taken by the governance mechanism, the potential impacts on human rights and to hold decision-makers to account. Despite some challenging aspects - explored in greater detail above - the UN Human Rights Committee (HRC) and the Committee on Economic, Social and Cultural Rights (CESCR), and the Universal Periodic Review (UPR), each provide a partially open process for engagement by non-governmental stakeholders, where they can engage predictably and on a recurring basis, and where their views can influence the execution of activities. Another source of inspiration is provided by the Gavi Alliance governance structure with board seats reserved for different stakeholders and geographic representations that provide the opportunity to influence the governance of the mechanism itself.

Important lessons can also be drawn from the assessed GMs in terms of ensuring the independence of the institution, its experts and their diversity. Both the OECD's AI Policy Observatory and the HRC and CESCR demonstrate the limitations where experts are selected by state parties. Eligibility of experts criteria should recognise the full range of disciplines (including social science, psychology, human rights, environment, and sustainability experts, to name a few) which may contribute to AI governance, and measures should be put in place to ensure gender, geographic and disciplinary diversity.

Another key aspect in the design of the mechanism to fulfil this function is the emphasis on covering techno-social aspects and not only hard science ones. This has proven increasingly relevant in the work of IPCC over the years as deals with two relevant aspects, for one the uneven distribution of control of the technology and where impacts are produced, and for the other the link between individual directly traceable AI impacts (such as bias and discrimination), and other collective non-AI specific impacts which are more difficult to trace to specific AI applications (such as democracy erosion).

Finally, a critical point to be considered by HLAB-AI in the fleshing out of the mechanism to fulfil this function is ensuring greater co-ownership in

designing impact assessments, and differentiated responsibility-sharing among stakeholders in the performance of the assessment and continuous monitoring of risks, risk management and mitigation. For this purpose, we agree with the concrete recommendations formulated recently by a Royal Society Report,[4] on the need to focus on answering questions around the frequency of measurement, time frame, indicators (human rights relevant), scope and audience. This function can fulfil a relevant role in standardising in practice the impact assessments that could be leveraged by other parts of the UN AI governance system dealing with accountability.

Institutional Function 2: Interoperability and alignment with norms

This function seems naturally devoted to achieve coordination at the normative level, including required technical standardisation. This appears to be partly captured by the HLAB-AI Interim Report when it briefly cites the work of existing UN organisations and fora such as UNESCO and the ITU in coordination and interoperability of regulatory measures across jurisdictions. However, this section leaves much to be desired in terms of how this will work in practice, and how human rights should play a role as minimum standards universally agreed.

The main challenge linked with the design of the mechanism to fulfil this function as part of the UN AI Governance Framework is ensuring that the "interoperability" of norms is able to deal with a diversity of rules adopted across different jurisdictions, while at the same time ensuring that there is a common understanding of the basic requirements that any regulatory arrangement should embrace. The HLAB-AI Interim Report provides a solid foundation for this as it identifies the need for a mechanism that is "grounded in international norms, such as the Universal Declaration of Human Rights", but it would still benefit from a more concrete suggestion on how to provide guidance to states in pursuing that.

[4] The Royal Society, The United Nations' role in international AI governance. Summary paper of a workshop held on 28 February 2024, Available at: https://royalsociety.org/-/media/policy/publications/2024/un-role-in-international-ai-governance.pdf

It is important to recognise that, in its discussion of universally agreed norms, a UN AI governance would not be starting from zero. There is already a wealth of normative guidance produced by UN human rights bodies and other entities to interpret and apply international law that should serve as the common ground to ensure normative interoperability for AI at the international level. If HLAB-AI seeks to establish a new mechanism to harmonise policies and provide clarity on governance principles and norms, it should have as its basis international law, including international human rights law and international humanitarian law, and the UN Guiding Principles on Business and Human Rights, and the work of existing bodies, in particular OHCHR, and specifically the B-Tech Project, the UN human rights treaty bodies, the Special Procedures mandate-holders. Based on the information available to date, the proposed model of the Human Rights Advisory Mechanism (HRAM) is recommended,[5] given its objective to ensure coherence and complementarity with existing institutions by building on the work of the human rights mechanisms and experts.

One angle of the coordination function that does not seem to be sufficiently addressed by the HLAB-AI Interim Report and merits additional consideration is how to deal with normative, political, social and cultural diversity in the normative landscape for AI governance at the local and regional level. The ability to guide implementation and to avoid AI divides or governance gaps will inherently require an intersectional approach, guided by universal and international standards, and the engagement of multiple, diverse stakeholders. The engagement of experts in legal and regulatory compliance should be specifically acknowledged and welcomed.

[5] This mechanism was first proposed by the UN Secretary General in his Policy Brief on the Digital Compact — an Open, Free and Secure Digital Future for All from May 2023. Available at: https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-gobal-digi-compact-en.pdf.

Institutional function 3: Develop and harmonize standards, safety, and risk management frameworks

When defining this function the HLAB-AI Interim Report again emphasises the "lack of global harmonization and alignment" of the current efforts to create "technical and normative standards, safety, and risk management frameworks for AI". In other words, the function deals at least partially with questions of coordination covered in function 2, and assessment of safety and risk management frameworks that should be captured by function 1. While we understand that there will be some overlap between the functions and underlying sub-functions, we believe these distinct objectives could be better captured as set out at the bottom of this report.

However, putting that aside, we consider that the harmonization of technical standards should be considered an integral part of normative coordination, and therefore be grounded in international human rights law. [6] Decoupling the development of risk identification and management frameworks from a human rights benchmark (e.g., conducting human rights due diligence and addressing broader impacts on the enjoyment of human rights) undermines the ability of technical standards to effectively contribute to safety considerations that are holistic and aligned with states' existing obligations under international law and the fulfilment of their commitments under the SDGs.

As highlighted when commenting on the assessed GMs devoted to risk management, it is essential that technical standards developed for risk identification and management have a strong component of socio-technical evaluation, grounded in human rights, and provide for differentiated responsibility-sharing of stakeholders. This is particularly critical when it comes to the uneven distribution of impacts in the development and deployment of AI systems between the Global North and Global Majority. We therefore reiterate the report's call for active involvement of civil society and transdisciplinary cooperation to develop these standards, including providing the necessary resources to ensure this takes place. We also support the suggestion within the Royal Society's report that one concrete approach to be considered is "mandating the use

[6] As emphasised in 2023 report of the Office of the High Commissioner of Human Rights (OHCHR), Report on the relationship between human rights and technical standard-setting processes for new and emerging digital technologies, available at: https://digitallibrary.un.org/record/4031373?ln=en&v=pdf.

of model cards which explain what AI systems do, how they were constructed, and what data they were trained on."[7] Any such guidance on model cards should be designed in a manner that facilitates transdisciplinary assessment of the socio-technical dimensions of AI systems.

Institutional function 4: Facilitate development, deployment, and use of AI for economic and societal benefit through international multi-stakeholder cooperation

The description of this function in the HLAB-AI Interim Report, as well as the following function 5, stresses the relevance to overcome the access to "enablers" to AI technology. While function 4 seems more focused on the role of "rules or standards" as enablers, function 5 seems to deal with critical components of the designing and functioning of the technology (data, computing infrastructure and talent). As two sides of the same coin, it might be useful to think of a governance mechanism that deals with both aspects in a coordinated manner and under one particular function. As the GMs assessment above shows, it is essential that any access granted to technology is accompanied by the institutional and normative strengthening of local capabilities, in order to avoid mission creep in the technology transfer originally intended to SDGs fulfilment. It is also necessary to fully support the enjoyment of human rights, including the right to development, and mitigate against any risks to human rights that may arise from the transfer of AI technologies, such as the right to privacy.

Putting it in different words, the adherence to technical standards dealing with risk assessment and management, including in the access and management of data used for AI systems, should be considered an accompanying pre-condition for supporting the technology transfer and spread of AI systems. In this regard, a UN AI governance mechanism should not decouple this institutional function from the development of rules or standards, nor from the clarification of norms, grounded in international law. A good model can be found at IAEA, where adherence

[7] The Royal Society. The United Nations' role in international AI governance. Summary paper of workshop held on 28 February 2024. Available at: https://royalsociety.org/-/media/policy/publications/2024/un-role-in-international-ai-governance.pdf.

to normative and technical standards is considered as a condition to benefit from the knowledge sharing and capacity building.

This function is right to highlight how existing arrangements need to evolve to anticipate complex adaptive AI systems of the future, and this requires lessons from forums such as FATF and SWIFT. However, both mechanisms, through their creation of best practices and control mechanisms, have shown negative consequences for human rights, particularly as they disproportionately impact developing countries and vulnerable groups via an excessive focus on punitive measures and lack of flexibility. SWIFT and FATF were created and are still primarily driven by the Global North, which is reflected in their composition, operation and overall exclusion of other stakeholders. A more dynamic and collaborative approach to equitable access, capacity building and standard setting is needed for this function. This differentiated approach is also embraced by WTO agreements that allow for implementation flexibility for "developing" countries.

Institutional function 5: International collaboration in data, compute and talent to solve SDGs

As mentioned in the previous function, this one deals with critical components on the designing and functioning of AI technology (data, computing infrastructure and talent). Each of these components should facilitate international cooperation oriented towards ensuring the conditions for human rights-respecting access to AI technologies to ensure SDGs fulfilment.

The HLAB-AI Interim Report rightly points out in this function to the models provided by CERN, EMLB or IAEA in terms of knowledge-sharing, however our assessment of those mechanisms shows that there is a need for increased attention to engagement with civil society, not only as knowledge recipient but also as a relevant actor in the governance and decision making of the mechanism. Analysis of the experience of the Gavi Alliance also identified the need for governance mechanisms to ensure that the direct beneficiaries of access to the technology have a means of engaging with the governance structure.

For the AI governance mechanism implemented to fulfil this function, addressing how the increased capacity would be achieved will be key.[8] There is a critical role to play for public-private partnerships in terms of enhancing the ability to access and use of existent data sets and AI open source models that can be tailored for local context relevance (including language and concerned population characteristics). Equally, there could be a relevant role to play for the governance mechanism in facilitating international cooperation to support the development of digital public infrastructure that can enhance the ability of the Global Majority to leverage the benefits of AI deployment for SDGs fulfilment in a manner that is mindful and tailored to their needs and context. This may, in turn, support economic, social and cultural rights through potential economic empowerment, social inclusion in the form of education, healthcare and community development, as well as cultural preservation.

A final critical challenge of the mechanism to fulfil this function is addressing the power imbalances in the concentration of power on the few multinational companies that currently are able to produce and control AI advanced models and offer them across jurisdictions. There are relevant elements of the design of the Gavi alliance and specifically its ability to negotiate access to vaccines in the face of the powerful pharmaceutical industry that deserve attention and would merit replication.

On the other side, there are lessons that can be drawn from other governance mechanisms that provide access to their facilities, such as EMBL, which sets potentially onerous conditions - such as independent funding requirements - that could be a limiting factor for enabling equitable access.

Similar to the model offered by CERN, it will be relevant that the governance mechanism that deals with this function has the flexibility to increase the public dissemination of outputs and research for international use, but at the same time also support innovation with commercial value which could be licensed to member states in conditions that attend in a differentiated way to the needs of the Global Majority for technology

[8] We find compelling and support the idea proposed by the Royal Society Report for the UN conducting a "gaps report" on the Global Majority needs on these elements.

transfer. This model will also enhance the incentives for Global Majority-led innovation that could be later on globally licensed.

Finally, we note the stated purpose of this function to facilitate the development, deployment and use of AI as an enabler to fulfil the SDGs, which addresses a vital need to ensure more even distribution of the benefits of AI. While this focus is noteworthy and vital, it is imperative that any AI governance mechanism charged with this function takes a balanced approach and critically assesses the potential benefits and risks that AI development, deployment and use may pose to human rights and SDG realisation. A comprehensive assessment of potential human rights and environmental impacts, including any possible positive impacts for SDG fulfilment, is necessary to ensure a proportionate approach to the delivery of this function. Essentially, it should be considered whether the use of AI is the best means of achieving the desired result and proportionate to the aim pursued (in this case, SDG fulfilment). More specifically, the institution of this function within a future institution should be informed by scanning and risk monitoring (function 1) and undertaken in compliance with harmonised standards (function 2).

Institutional function 6: Monitor risks, report incidents, coordinate emergency response

The HLAB-AI Interim Report correctly cites several risks that may be posed by the design, development and deployment of AI, ranging from the lowering of barriers for access to weapons of mass destruction to the dissemination of harmful information. While not explicitly noted in the Interim Report, these risks may have significant negative impacts on the enjoyment of human rights ranging from the right to life, privacy and freedom of expression,[9] and these impacts should be a central referent in the execution of this function. We therefore commend the recognition of a techno-prudential model as one possibility, which must be grounded in international human rights law: this should be the case for any model established.

[9] See OHCHR. Taxonomy of Human Rights Risks Connected to Generative AI (2023). Available at: https://www.ohchr.org/sites/default/files/documents/issues/business/b-tech/taxonomy-GenAI-Human-Rights-Harms.pdf

We are, however, critical of the suggestion of following the IAEA model as inspiration to address this function. This is because our GMs assessment shows valid critiques of the IAEA for its potentially contradictory mandate - promoting nuclear energy and non-proliferation on one hand, whilst also serving as a watchdog on the other, and inability to respond to emergencies, and its inability to react to pressing events. This was evidenced by the IAEA's inability to respond to the Fukushima disaster in a timely manner, and the backlash it received in terms of reporting on safety standards.

Any function related to risk monitoring and reporting of incidents must carefully devise the scope of what AI risks are included through leveraging interdisciplinary expertise and consider the impacts of these risks on human rights. Related to function 5 (international collaboration to data, compute and talent), risk monitoring could also review the sustainability of AI deployment and other risks to SDG fulfilment.[10] This includes defining what is even meant by a system vulnerability or disruption to international stability. Similarly, it will be critical to unpack how the function will not only respond in a rapidly changing context, but to consider what stabilisation measures will even be considered acceptable by relevant regional or national entities, the private sector or other actors. We believe that the ICAO's response to COVID-19 demonstrates one example of how a global mechanism can respond to such global disruptions in a timely and effective manner.

We believe that the establishment of some form of emergency response mechanism should be developed under this function in the establishment of an AI governance framework. But it should be clearly defined as to whether this response mechanism is one focused on risks and harm prevention, or if it is instead designed to serve as a broader accountability mechanism that ensures compliance with relevant norms and established effective right to redress when emergency happens, as is set out in function 7. This might ultimately take place in two stages: (1) first deciding on and establishing an emergency monitoring framework, which includes devising its scope and institutional capacity to collect real-time information (which could be covered by what is described in function 1);

[10] The Royal Society. The United Nations' role in international AI governance. Summary paper of a workshop held on 28 February 2024. Available at: https://royalsociety.org/-/media/policy/publications/2024/un-role-in-international-ai-governance.pdf.

and (2) creating a response mechanism to ensure compliance and provide accountability, which could be covered by what is described in function 7. This is likely to take place when the governance framework is more firmly established, as discussed in the next section.

Institutional function 7: Compliance and accountability based on norms

The HLAB-AI Interim Report identifies possible scope for the UN to perform a "compliance and accountability" role founded on universally agreed norms and with an enforcement function. While there is a need for an international AI governance framework to contain some mechanism to provide accountability, it might make more sense for other institutional functions to be established beforehand.

We are pleased the Interim Report recognises that "we cannot rule out that legally binding norms and enforcement would be required at the global level", and specifically points to upcoming enforcement efforts which will emerge from the Council of Europe (CoE). This is important as the CoE has already finished elaborating a global treaty on AI and human rights, which will require state reports on compliance with relevant obligations. Another example is the opportunity to leverage the accountability provided by the OECD Multinational Enterprise Guidelines that could be further strengthened according to the recommendations provided above, including a more specific focus on human rights. In all cases, an accountability mechanism at the UN level must therefore not only aim to be the only arbiter of AI governance but should also be careful to not replicate compliance or accountability schemes which already exist and are underway. It therefore requires coherence with the functions related to international coordination and harmonisation.

Rather than duplicating existing compliance or reporting mechanisms, the UN should focus on those areas where it has a unique normative and institutional role to play. In this manner, the OECD AI Policy Observatory may provide a useful precedent, particularly through its work to promote definitional clarity and consensus through its AI principles and OECD definition of an AI system. These have been instrumental in informing the contents and terminology of existing AI frameworks such as the EU AI Act

and CoE Convention on AI. As outlined above, another area in which the UN could be well-placed to contribute to the AI governance landscape is by developing over time a framework for ensuring effective oversight or accountability for uses of AI which pose heightened risks to human rights, and international peace and security.

At the present time, it is our recommendation that the UN gives priority to ensuring harmonisation, in the manner described under institutional function 2. In this manner, the commitment within the Interim Report to ensuring analogous reporting and coordination with other mechanisms is welcome, given the considerable prior work of other mechanisms in norm-making and enforcement. The model of the Human Rights Advisory Mechanism (HRAM) is a useful example of what the UN's role could look like, given its objective to ensure coherence and complementarity with existing institutions by building on the work of the human rights mechanisms and experts to supervise and apply the international human rights law framework in the context of digital technologies.

If, in due time, the UN were to play a role in ensuring the adherence of relevant parties to universally agreed norms, it would require the establishment of an effective mechanism for enforcement. This enforcement should consist of an agreed-upon normative framework, a multistakeholder process for reviewing compliance with the framework, penalties for non-compliance, and avenues for redress. Otherwise, the entity will do little when compared to institutional functions 1, 2 and 6. Our analysis of other governance mechanisms indicates that efforts to address accountability gaps are unlikely to succeed when based solely on reporting. Each of the High-Level Political Forum (HLPF) Voluntary National Review (VNR) process, the Universal Periodic Review (UPR) and the HRC and CESCR rely on a constructive dialogue with state parties to execute their functions, which oftentimes frustrates their abilities to arrive at true accountability. Dispute resolution inspired by the WTO may prove similarly inadequate as the formation of *ad hoc* dispute settlement panels can be timely, costly and ill-suited to resolving the more dynamic issues posed by AI.

The experience of other governance mechanisms demonstrates the need for a more holistic approach to accountability, one that is clear in its mandate, and grounded in principles of independence, transparency, and

multistakeholderism. The UN has a unique role to play in modelling what this should look like. Our assessment indicates that this should incorporate the ability for stakeholders to engage via clear communication channels and the establishment of feedback mechanisms to receive input, suggestions and complaints either from stakeholders themselves or in support of others. A good point of reference for this includes the UPR, the HRC and the CESCR, which ensure a relatively predictable, open, and accessible process for civil society and rights-holders to monitor adherence to established norms. The HRC and CESCR are applauded for having taken some steps to mitigate barriers to access, although linguistic and financial barriers remain. By comparison, the HLPF's VNR process has clear limitations, allocating only limited time to the review of state reporting, and failing to formally recognise the contributions of non-governmental stakeholders.

Finally, the accountability function needs to find a way to deal with the role of industry and the current market power concentration in the design and development of AI. On one hand, the accountability mechanism should leverage collaboration with UN mechanisms that monitor the implementation of UN Guiding Principles on Business and Human Rights, and the workaround its implementation by the B-Tech Project. On the other hand, it should be able to hold enhanced scrutiny for those companies that are dominant in the global market.

Conclusions and recommendations

We wish to conclude by providing our insights and thoughts on the following questions raised by the HLAB-AI interaction with the network of experts:

1. What are the existing strengths of the UN system which are best suited to addressing function implementation?
2. Which of the functions are more urgent to implement and what should be done with the functions that will take longer to take action on?
3. Which of the functions are easier to implement/have a greater level of consensus?

As noted within our analysis, **there are particular functions which are in need of more immediate attention**. These include aspects of **functions 1 and 6** in the Interim Report that focus on scientific research and risk monitoring (**function 1** in our proposed approach below). There is also a pressing need for some form of coordination as envisioned under several functions, notably **functions 2 and 3** in the Interim Report (function 2 in our proposed approach below). These efforts are more urgent to implement due to the fragmentation of existing global mechanisms to assess what global challenges exist, a common space to build trust and confidence among stakeholders, and ultimately facilitate collaboration and knowledge sharing on the issues. The focus on these functions may be beneficial from a more practical perspective as well because they square elements of governance that are the UN's strength, as a critical locus for networking, information exchange, consensus building and coordination.

This does not mean that other functions, particularly **functions 4, 5 and 7** in the Interim Report, focused on promoting access to critical components for enhancing AI benefits or ensuring accountability (including rapid response in **function 6**) are not as urgent, but require additional consensus to be effectively established, and therefore come later, as global governance efforts mature. For example, it is our view that the institution of **function 5** (ensuring a mechanism for access for SDG

fulfilment), must come after the development of a risk monitoring function (**function 1** in our proposed approach) and the coordination of human rights-based normative and technical standards (function 2 in our proposed approach below). Institutional capacities for evidence-based and multidisciplinary risk monitoring and harmonisation of standards should be established prior to the facilitation of access.

These other functions (**functions 4, 5 and 7** of the Interim Report) **would benefit from further discussion within other processes**, including the Global Digital Compact (GDC) or existing UN mechanisms, which could unpack the means of arriving at these functions and their potential operation in a human rights-respecting approach. More specifically, the GDC as a venue is well-placed to establish guidance - grounded in the international human rights law framework - for the responsible and accountable management of digital technologies, including AI. It could usefully set out standards for risk mitigation, including human rights impact assessment, and build upon existing commitments to prohibit the use of AI systems that are impossible to operate in compliance with international human rights law or that pose undue risks to the enjoyment of human rights by specifying technologies or use cases which meet this definition. In this way, the GDC could provide necessary and timely input to the AI governance landscape, which would inform the work of a future UN mechanism.

The interaction between international and national regimes for oversight and accountability of AI is a complex topic which also deserves attention here. A comprehensive global accountability regime is one that combines overlapping national, regional and international enforcement mechanisms. For instance, much of the literature has rightly addressed the need for attention to be given to the establishment of robust data protection frameworks, anti-discrimination legislation, consumer protection regulations, and competition policy for example, as well as to national institutional capacities to implement and oversee these regulations. Future UN AI governance has an important role to play as a useful complement to these national frameworks - particularly in fulfilling the horizon scanning, risk monitoring, international coordination and harmonisation functions - its work could include providing definitional clarity to aid these efforts, or aggregating data on the institutional and legal capacities of member states. However, UN AI governance should not

undermine important national enforcement efforts by, for example, institutionalising access without a human rights benchmark or risk monitoring. It is vital that efforts to establish UN AI governance are designed in a manner that reinforces national regulatory regimes.

Based on our analysis of the respective governance mechanisms and building on the Interim Report categorisation, **we believe that a revised approach for priority aspects of the various functions would streamline the goals and priorities of an international AI governance framework with a human rights approach**. This may also prove easier to navigate for those who lack familiarity with AI itself but have relevant experience in international governance, and additionally could be more future-proofed to anticipate pressing or unforeseen needs.

4. Accountability including rapid response, normative enforcement and access to remedy
3. International cooperation for access to knowledge, data and infrastructure
2. International coordination in normative & technical standards grounded in human rights
1. Horizon scanning, building scientific consensus and risk monitoring

Annex: Methodology

The following provides detail of the methodology used to analyse the Governance Mechanisms (GMs) in focus. It consists of three steps: first, the identification and classification of the GMs according to their primary and, where relevant, secondary functions. Second, an overview of the positive and negative impacts on human rights of the GMs, and the cause of those impacts, from a substantive and procedural perspective, based primarily on proxy sources. Third, based on the prior assessment of impacts and their causes, it draws out recommendations which should be considered if the GM(/s) in question is used as a model for the AI international governance regime. The third part of this assessment informs both the assessment of the GMs from a human rights based approach (part one of our analysis), and the assessment of the role of different institutional functions in contributing to the international AI governance regime (part two of our analysis).

Identification of the GMs

- Name, Acronym
- Scope of work: Thematic, geographic, temporal
- Goal/mandate: The objective that the GM is tasked with achieving, and has the authority and legitimacy to pursue
- Composition: The GM's members and/or constituents (e.g. internal teams, external participants, senior leadership, etc.). This also includes how these members are distributed or internally organised across the GM, and how they are selected (e.g. what criteria/ selection processes exist within the GM).
- Activities: The actions carried out by the GM or its members, including internal mechanisms and modalities of work (e.g. stakeholder engagement, evidence gathering, analysis, internal and external reporting, etc.)
- Function taxonomy:

Categories	Definition
Research development	Activities led by academics or researchers supported by companies or states, and can consider the engagement with other stakeholders (such as civil society organisations or technical experts) to collect information and produce/interchange/share scientific knowledge.
Access	Ensuring the availability of infrastructure, knowledge and skills to benefit from scientific knowledge and technical developments.
Risk monitoring	Collection of information and knowledge about risks impacting human rights coming from technology use to support evidence-based decision-making.
Accountability	A series of actions taken in order to verify states or companies' fulfilment of voluntary or binding commitments.
Coordination	Mechanism agreed by a diversity of stakeholders (such as states, companies, civil society organisations, academics, and technical experts) to interchange information or knowledge to support evidence-based decision-making.

Overview of positive and negative impacts on human rights

To identify and briefly assess the impacts of each GM, we mainly rely on proxy sources due to time and capacity constraints. While we directly consulted with several members of civil society, most of our stakeholder engagement was indirect, by reviewing publications and activities. These include:

- CSO website and reports, open letters, briefers, and other materials
- Journalistic reporting, media stories, academic papers
- Ongoing public litigation and/or previous court cases

For each GM, we map out and do an overview of positive impacts and key concerns based on the proxy sources listed above from a substantive and procedural perspective which are reflected in two sets of indicators.

Substantive indicators (human rights impacted)

As consistent with the UNGPs we have taken a holistic approach to international human rights law. Our analysis is based on the core internationally recognised human rights contained in the International Bill of Human Rights (consisting of the Universal Declaration of Human Rights and the main instruments through which it has been codified: the International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights).

While we recommend when assessing future GMs for AI to consider every right as relevant, below are the most commonly impacted:

- Right to protection against discrimination (art. 2 UDHR)
- Right to life, liberty and security (art. 3 UDHR)
- Right to equality before the law (art. 7 UDHR)
- Right to privacy (art. 12 UDHR)
- Right to freedom of movement and residence (art. 13 UDHR)
- Right to own property (art. 17 UDHR)
- Right to freedom of thought, conscience and religion (art. 18 UDHR)
- Right to freedom of expression and access to Information (art. 19 UDHR)
- Right to freedom of assembly and association (art. 20 UDHR)
- Right to vote and to democracy (art. 21 UDHR)
- Right to social security (art. 22, 25 UDHR)
- Right to work and to gain a living (art. 23 UDHR)
- Right to health (art. 25 UDHR)
- Right to education (art. 26 UDHR)
- Rights to culture, art and science (art. 27 UDHR)

We also considered the right to a clean, healthy and sustainable environment as recognised by the UNGA (A/RES/76/300). For each human right identified, we briefly evaluate the impact to inform the gap analysis or reflection on the cause and effect (see step 3).

Process indicators

Process indicators show how the GM is structured and functions in a way that is effective or not for ensuring the greatest exercise of human rights. The following questions were developed to interrogate the procedural elements of the GMs:

Criteria/ issue	Safeguard	Metrics/ benchmarks
Transparency	Information Access	• What access to information and documentation does the GM provide publicly vs upon request? What information is entirely inaccessible (not even through FOIAs)? • Is there publicly available documentation outlining how to access this information? • What information does the GM share about internal processes and procedures, and their outcomes? • Is there publicly accessible information about funding?
Accountability	Remedy	• What is the object/scope of oversight? • How is the international grievance mechanism structured? • What penalties, if any, are there for non-compliance?
Dispute resolution	*Remedy*	• Can affected parties appeal decisions of the GM? • What redress options exist for external groups/stakeholders (those directly affected and their representatives, third parties and civil society organisations)?

		• What remedies are available (e.g. compensation, reversal of decision, etc)? • Does the GM provide any support to the appellant/affected party?
Openness	Participation	• To what extent is the GM open and accessible to civil society and affected communities or individuals, especially marginalised groups and those in the Global Majority? • What barriers to entry exist for external stakeholders, especially marginalised groups? (e.g. related to resources, language, knowledge and capacity, decision-making power, accreditation or authorization to participate)? • How does the GM address these barriers?
External stakeholder engagement	Participation	• Is the body open to Member-States only, or is it multi-stakeholder? If multi-stakeholder, what sectors have a formal role or representation? Are the rules of engagement and the modalities of decision-making clearly defined? • How can stakeholders from the Global Majority engage in the GM's procedures and activities? How can civil society and affected communities, especially those from marginalised groups, engage in the GM's procedures and activities? • Is there a clearly defined and publicly available process outlining what influence, if any, these stakeholders can

		have on the outcomes of the procedures and activities?
Internal stakeholder engagement	Participation	• What role and influence do Member-States from the Global Majority have? • What role and influence do Member-States with poor human rights records have? • What relations or partnerships (formal or informal) exist between the GM and other UN mechanisms or bodies? Are these publicly disclosed? • What, if any, relationships does the GM have with UN human rights bodies? Is there a clearly defined and publicly available process outlining what influence, if any, human rights bodies can have on the outcomes of the GM's procedures and activities? • What relations does the GM have with regional bodies? • How much decision making power in or concrete influence over the GM's outcomes of the procedures do other UN mechanisms have? Is the consultation mandatory? Do members of these bodies hold any positions or authority in the GM?
Resources and funding	Accountability	• Which countries or actors fund the GM? • Are there potential concerns related to sustainable funding (e.g. funding from 1-2 countries only, funding from countries that have poor records of human rights)? • Is there any funding specifically allocated for civil society participation? If so,

		what amount and how is allocated? • Is there any other support besides funding allocation for civil society participation (e.g. capacity building, training, translation, safety)?
Independence	Influence	• How integrated is the GM with the broader UN system? What is its level of autonomy? • How much can individual Member-States influence the GM? • How much are the processes and outcomes based on, and driven by, evidence-based information vs political agendas?
Flexibility	Context	• From a governance standpoint, how much authority does the GM have to adapt its internal operations, processes and goals to address new risks, concerns, or adaptive technology (if relevant)? • From a logistical and knowledge standpoint, how much capacity does the GM have to adapt its operations, activities and goals given limited resources and expertise?
Monitoring and evaluation	Oversight	• How does the GM monitor its internal activities? • How does it evaluate whether it progresses on its goals?

Recommendations developed according to the gap analysis

Based on the overview analysis in step 1, we surfaced the most salient human rights impacts (procedural and/or substantive)for each GM. For the purpose of this paper, we focus on surfacing the cause and effects of impacts. We try to identify what elements of the GM have caused the adverse impact or what particular gaps prevent the GM from effectively addressing the impact (e.g. lack of transparency, no or ineffective civil society participation, biased funding, etc.).

Based on 'cause and effect' or gap, we provide recommendations to prevent this adverse impact from occurring in the future. However, for the purpose of this study, we are suggesting mitigation measures for the fulfilment of potential functions for the future entity (proposed by the HLAB-AI),and not for each assessed GM. Our recommendations build on the prioritisation and 'cause and effect' analysis, focusing on procedural safeguards. They are intended to form a baseline for establishing "minimum requirements' that must be met for the future AI mechanism to be legitimate and effectively protect and promote human rights.

***

If you have any questions or comments on this report, please contact GPD's Head of Legal, Policy and Research, Maria Paz Canales (mariapaz@gp-digital.org).