G2O BRASIL 2024

T2O

Brasil 2024
Let's rethink the world

Task Force 05
INCLUSIVE DIGITAL TRANSFORMATION

AI Governance Mechanisms with a Human Rights Approach

Maria Paz Canales, Head of Legal Policy and Research for Global Partners Digital (United Kingdom)

Ellie McDonald, Lead engagement and advocacy for Global Partners Digital (United Kingdom)

Ian Andrew Barber, Lead legal for Global Partners Digital (United Kingdom)

TF05

Abstract

During the last year, AI governance has proven an urgent priority. As mapped by Kaspar et al. (2023), there is currently more than 50 active international AI governance initiatives, nearly a quarter of them within the UN system itself - including the UN High-level Advisory Body on Artificial Intelligence (HLAB-AI) - joined by other multilateral initiatives or state-led initiatives, including the Council of Europe's AI Convention, the EU AI Act, and the G7 Hiroshima Process.

However, there is a concerning absence of reference to the international human rights law (IHRL) when looking at the different institutional options that have been presented as an alternative for global AI governance. There is a risk that the AI governance discussions sidelines - or worse, neglects to account for - the application of IHRL on AI governance.

The authors have conducted a study to assess the human rights impacts of different mechanisms proposed in the public discussion for AI governance and mentioned in the HLAB-AI interim report (2023). The research has collected valuable evidence on what institutional formats and functional elements could provide a strong foundation to ensure the AI global governance mechanisms are underpinned by human rights and support the realisation of the SDGs.

This policy briefing leverages the conclusions and recommendations from this study which analyses seventeen governance mechanisms as models for international AI governance. We use this opportunity to summarise lessons and mitigation measures for addressing institutional functions of international AI governance mechanisms, intending to provide insights to guide the design of an international AI governance regime as part of the G20 Framework for Responsible Human Centric AI Governance, but that can also

be useful in the engagement of its members states in international AI governance discussions at the UN and other forums.

Diagnosis of the Issue[1]

Discussions on responsible AI global governance has become a regular part of the G20 agenda since 2019. While the G20 AI Principles provide a framework to develop and deploy AI in a way that is beneficial and addresses concerns related to ethics, privacy, and security, there is limited consideration of the distributional aspects and existing multidimensional power dynamics that shape global AI governance. In 2023, the G20 presidency called for a framework for Responsible Human-Centric AI governance.

There is a critical opportunity to take a proactive approach to ensure that the G20 agenda on AI governance takes IHRL into account from the start, and to influence the direction that other stakeholders can take in supporting different institutional options when engaging in other global AI governance processes. Whatever form this global AI governance takes, there is the urgent need for it to be shaped in a more open, inclusive and transparent manner, ensuring the Global Majority countries perspectives are captured, and G20 is in a unique position to ensure this.

Accordingly, we examine the following governance mechanisms (GMs) which have been grouped by the primary function they could play in AI governance, while noting that several of them also perform a strong secondary function. We recognise the limitations of this approach and the value of expanding our assessment to encompass additional GMs such as technical standard-settings. This research does not pretend to be exhaustive but rather spark additional interrogation of governance mechanisms from a human rights perspective.

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	GMs list
Research development	CERN European Organization for Nuclear Research EMBL European Molecular Biology Laboratory
Access	GAVI The Vaccine Alliance
Risk monitoring	IPCC Intergovernmental Panel on Climate Change FSB Financial Stability Board UK AI Safety Institute
Accountability	IAEA International Atomic Energy Agency ICAO International Civil Aviation Organization FATF Financial Action Task Force OECD Multinational Enterprises Guidelines UN Treaty bodies (Human Rights Committee/Committee on Economic, Social and Cultural Rights) UPR Universal Periodic Review WTO dispute resolution
Coordination	HLPF High-level Political Forum on Sustainable Development SWIFT Society for Worldwide Interbank Financial Telecommunication OECD AI Policy Observatory HRAM the digital Human Rights Advisory Mechanism, 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, based on the information provided to date.

The full findings on the human rights impacts of the GMs assessed can be consulted in Canales et al. (2024).

Recommendations

Our study concludes that there is a pressing need to further interrogate and prioritise particular functions in AI governance - 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 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), existing UN mechanisms and G20 Framework for Responsible AI. It also highlights the important role of any future global AI governance mechanism to complement and reinforce national regulatory regimes on AI.

We recommend the AI global governance effort from G20 focus in the first place on supporting: i) scientific research and risk monitoring, and ii) international coordination in normative and technical standards grounded in human rights. 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 can benefit from G20 policy coordination, as a critical locus for networking, information exchange and consensus building.

Our research has also shown that there are relevant challenges for the Global Majority to support their ability to meaningfully contribute to scientific information gathering. **G20 support of AI governance should therefore account for the task of enabling that participation through capacity building and funding support for Global Majority engagement**.

It is important to recognise that related to normative coordination that global AI governance is not 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.** The G20 should support that any new mechanism to harmonise policies or provide for normative interoperability 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 the Office of High Commissioner of Human Rights, 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 model of the Human Rights Advisory Mechanism (HRAM) proposed by the UN Secretary General (2023) is recommended, given its objective to ensure coherence and complementarity with existing institutions by building on the work of the human rights mechanisms and experts.

The interaction between international and national regimes for oversight and accountability of AI is a complex topic which also deserves attention. **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. The G20 should ensure moving forward in its AI governance approach to support its members in enhancing their domestic policies – for example could include providing definitional clarity to aid these efforts, or aggregating data on the institutional and legal capacities of member states. However, AI global 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 from G20 to support AI governance are designed in a manner that reinforces national regulatory regimes. Further, the coordination function should deal with political, social and cultural diversity in the normative landscape for AI governance at the local and regional level.

The ability to avoid AI divides or governance gaps will inherently require an intersectional approach and the engagement of diverse stakeholders. The experience of examining GMs 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. 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.

Finally, **the G20 should be a place to advance the dialogue on the access to AI as an enabler to fulfil the SDGs ensuring a more even distribution of the benefits of AI**. For this purpose, a comprehensive assessment of potential human rights and environmental impacts, including any possible positive impacts for SDGs fulfilment, is

necessary to ensure a proportionate approach to developing the access mechanisms and the evaluation of each country of the pertinence of an AI deployment.

For AI access enabling there might be a critical role to play for public-private partnerships as well as 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. In order to enhance the opportunity of international cooperation for access to knowledge, data and infrastructure to address the geopolitical power imbalances, the G20 should work towards flexible conditions and a differentiated approach to the needs of the Global Majority for technology transfer. This model will also enhance the incentives for Global Majority-led innovation that could be later on globally licensed.

Scenario of Outcomes

The recommendations made above to prioritise governance efforts on scientific research and risk monitoring, and international coordination in normative and technical standards grounded in human rights do not mean that other functions, such as promoting access to critical components for enhancing AI benefits or ensuring accountability (including rapid response and remedy) are not as urgent, but require additional consensus to be effectively established, and therefore might come later, as global governance efforts mature. For example, it is our view that ensuring a mechanism for access for SDGs fulfilment, must come after the development of a risk monitoring function and the coordination of human rights-based normative and technical standards. Institutional capacities for evidence-based and multidisciplinary risk monitoring and harmonisation of standards should be established prior to the facilitation of access.

The failure to sufficiently address the socio-technical components in the governance mechanisms for risk evaluation and mitigation can be an additional source of geopolitical unbalance between the Global North and Global Majority considering the uneven distribution of impacts in the development and deployment of AI systems and the places where the control of the technology is held. We therefore call for active involvement of civil society and transdisciplinary cooperation to develop these standards, and therefore we encourage the G20 to provide the necessary resources to ensure this takes place.

AI technologies can have both positive and negative impacts on a range of human rights, including civil and political rights, as well as cultural, economic, social, and environmental rights. AI systems may also have unique impacts on groups whose rights are protected in specific international legal instruments, including women, ethnic minorities, children, people with disabilities, refugees, and migrants. We need to

remember that all human rights are indivisible and interdependent. This means that violation of a civil and political right (such as privacy or non-discrimination) might also result in a violation of the right to access to an economic, social, cultural, and environmental right (such as access to health, work or education). G20 Framework for Responsible AI to be truly human centric needs to account for the application of IHRL, otherwise AI deployment might end facilitating discriminatory outcomes which negatively impact vulnerable and traditionally marginalised groups, violating the civil, political, economic, social and cultural rights, and lastly even deepening the power asymmetries inside and between countries.

References

Canales, Maria Paz, McDonald Ellie, and Ian Barber. April 2024. "AI Global Governance: Assessment of Governance Mechanisms with a Human Rights Approach." Available at: https://www.gp-digital.org/publication/ai-global-governance-assessment-of-governance-mechanisms-with-a-human-rights-approach/

Kaspar, Lea, Maria Paz Canales, and Michaela Nakayama. October 2023. "Navigating the Global AI Governance Landscape."Available at: https://www.gp-digital.org/navigating-the-global-ai-governance-landscape/

UN High-level Advisory Body on Artificial Intelligence. December 2023. "Interim Report: Governing AI for Humanity." Available at: https://www.un.org/en/ai-advisory-body

UN Secretary General. May 2023. "Policy Brief on the Digital Compact — an Open, Free and Secure Digital Future for All." Available at:

https://www.un.org/sites/un2.un.org/files/our-common-agenda-policy-brief-gobal-digi-compact-en.pdf.