Governing AI For Humanity:
Our thoughts on the Final Report of the High-level Advisory Body on Artificial Intelligence

September 2024

Executive Summary

The Final Report of the UN High-level Advisory Body on Artificial Intelligence (HLAB-AI), was published on 19 September 2024.

It marks the conclusion of a yearlong process, set up with the aim of undertaking analysis and providing advance recommendations on the international governance of artificial intelligence. GPD has closely supported the Advisory Body members' work throughout the process. We welcomed the HLAB-AI's work as a potentially generative and useful opening to the UN taking leadership of coordination efforts within the currently fractured ecosystem of AI governance.

The Final Report follows the Interim Report, which was published in December 2023. While the identified principles remain the same, it somewhat departs from its structural design: **de-emphasising the Interim Report's focus on the importance of normative coordination in favour of international cooperation**. In spirit, it feels more like a 'plug-in' proposal than the ambitious intervention promised by the Interim Report.

In this summary, we briefly run through the general direction of the report, and provide a few top-level thoughts on its contribution to the current AI governance landscape. Below this, we present a full analysis of all the recommendations contained in the report.

Key messages of the report

A central message of the Final Report is its emphasis on the "global governance deficit" in AI. The report highlights that the current "patchwork of norms and institutions is still nascent and full of gaps"—exemplified by the exclusion of entire regions from international AI governance discussions, raising the risk of creating "disconnected and incompatible AI governance regimes." Additionally, the report argues, the UN's fragmented approach, due to the specific mandates of its entities, fails to address AI governance comprehensively.

The report underscores the disparity in representation among states involved in AI governance, pointing out that no high-performance computing clusters are hosted in developing countries. This demonstrates the scale of the challenge in ensuring equitable access to advanced AI resources. To mitigate this, the report advocates for supporting distributed and federated AI development models to bridge the AI divide.

Data-related issues also feature prominently in the report, including the misuse of–and lost opportunities around—data for AI, and the lack of data reflecting the world's linguistic and cultural diversity, which contributes to AI bias. The report calls for shared resources, such as open models, to address these gaps and promote inclusivity.

Our analysis

The report presents a series of proposals and recommendations to create a coherent framework for global AI governance, addressing the multifaceted challenges and opportunities AI presents. We welcome the inclusion of some of these proposals.

However, the final report **departs from the original premise of "form follows function" embraced in the Interim Report**, which intended to interrogate the specific functions required to provide robust AI governance and set a roadmap for when and how to advance its implementation. Instead, the Final Report focuses on mechanisms for filling gaps within the existing patchwork. **Such a piecemeal approach cannot provide the**

procedural and substantive elements necessary to ensure the achievement of the intended outcomes. Nor will it strengthen existing AI governance initiatives, which poses the risk of **making the mechanisms proposed by HLAB-AI irrelevant due to the currently crowded AI governance landscape.**

Normative coordination, which was correctly highlighted as critical in functions 2 and 3 of the Interim Report, **receives notably less emphasis in the Final Report—replaced by a focus on the** urgency of international cooperation. . This is, in our view, an error if the aim is building a progressive and strengthened path for accountability in AI governance. After all, normative harmonisation—anchored in human rights standards and bodies—sits at the core of the UN mission. **We believe that the report's proposals could be improved by better aligning with the UN's expertise and clearly prioritising actions.** We continue to believe—as highlighted in our previous research offered as input to the HLAB-AI's work—that institutional capacities for evidence-based and multidisciplinary risk monitoring and harmonisation of standards should be established prior to the facilitation of access.

The final report **centralises the majority of its newly proposed entities and processes within a singular office**, building upon the existing Office of the Technology Envoy, and anticipated to be based in New York. This, when read together with the recently adopted Global Digital Compact (GDC), has raised serious concerns of **a departure from the existing landscape of digital technology policymaking**, which is characterised by the multistakeholder precept set out in the Tunis Agenda and by a distributed ecosystem of UN institutions and multistakeholder forums and venues. While all of these entities individually face challenges around stakeholder engagement, the overall ecosystem, by virtue of its decentralisation, provides a range of avenues for non-governmental actors to engage and shape outcomes. By comparison, a more centralised—and, potentially, more closed and opaque—entity risks hindering non-governmental engagement.

The final report of the HLAB-AI provides a comprehensive overview of the challenges and gaps in the current global AI governance landscape, emphasising the need for cohesive and inclusive strategies. While several of its recommendations are promising and could potentially advance international AI governance, we have concerns regarding their

implementation and alignment with existing human rights frameworks and the bodies charged with overseeing compliance with those frameworks. The desire to fill gaps in AI governance should not sideline a proper consideration of which governance functions the UN is best placed to fulfil. Such an approach risks creating yet another initiative with too little buy-in to be impactful in global AI governance globally.

These recommendations therefore require further refinement to ensure they effectively integrate human rights and facilitate meaningful international cooperation in an open, transparent and inclusive way. A more robust focus on normative coordination and human rights standards will be crucial for achieving a proportionate and effective global AI governance system.

See below for our full analysis of all proposals in the Final Report.

Analysis of Proposals

Recommendation 1: An International Scientific Panel on AI

The report proposes an independent, multidisciplinary International Scientific Panel on AI, emulating and learning from precedents like the Intergovernmental Panel on Climate Change (IPCC) and the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR). Housed under the UN, this scientific panel would provide credible, impartial research to inform scientists, policymakers, member states, and other stakeholders about AI technology and its applications. This proposal from the HLAB-AI has been already included in the recently agreed Global Digital Compact.

The panel would offer expertise on AI opportunities related to the Sustainable Development Goals (SDGs) and serve as a trusted platform for global knowledge exchange. It would be supported by the UN AI Office and other relevant UN agencies such as the ITU and UNESCO, and through partnerships with the OECD and GPAI. It would produce annual reports on AI trends, quarterly digests, and ad hoc reports on emerging AI risks and governance gaps. It would facilitate “deep dives” into applied domains of the SDGs, and risk assessments would draw on work of other research initiatives, with the UN offering a “safe harbour” space for exchanges.

Our assessment

Ensuring diversity and independence

We welcome that the report places emphasis on a diverse and independent panel. However, we maintain that the independence of this panel requires that available resources and their administration be provided in an appropriate and distributed manner. We remain concerned about the voluntary basis of the commitment to serve on the panel, as well as the secretariat support that will be provided by the UN AI Office. Diversity of representation is difficult to ensure in a structure based exclusively on voluntary commitments, since underrepresented actors are less likely to be able to offer this.

Questions around output development

We are also concerned about how outputs would be agreed upon. The report states that the panel should operate independently—however, we recommend making it more explicit how the panel's independence will be guaranteed. To ensure independence, outputs should be produced by the independent body of experts, and not subject to intergovernmental negotiations and consensus-based decision making. The selection process for experts is another area that requires clarification to ensure impartiality and representation. Otherwise, there is a risk that outside influences will be able to improperly influence outputs through the composition of these bodies.

While the annual report's focus on scientific consensus is valuable, it should also highlight areas of disagreement and divergence amongst experts with different socio-technical approaches. The mandate's lack of clarity on where risk assessments would be undertaken and housed—whether in the annual report or ad hoc reports—needs addressing to avoid potential gaps in governance, as well as potential overlap.

Human rights emphasis

We welcome that the outputs of the panel would focus on the SDGs and areas of public interest, particularly the quarterly thematic research digests, but recommend there should be an explicit focus on human rights as well.

As well as exploring areas where AI could help to achieve the SDGs, the panel should also be granted latitude to explore, based on the evidence, whether the use of AI is the best means of achieving the desired result and proportionate to the aim pursued, and any cases where AI may jeopardise the achievement of the SDGs. Finally, we suggest that this panel be supported by relevant human rights mechanisms under the auspices of the United Nations such as the OHCHR and special mandates relevant for the areas of review, in addition to the ITU and UNESCO.

Recommendation 2: Policy Dialogue on AI Governance

The Policy Dialogue on AI Governance aims to share best practices, foster interoperable governance approaches, and address transboundary challenges. It seeks to align industry and national efforts in ways that help Member States work together effectively.

The dialogue would be an intergovernmental and multistakeholder forum, leveraging existing mechanisms like ITU's AI for Good, UNESCO's AI Ethics Forum, and UNCTAD's eWeek. These meetings would take place on a twice-yearly basis on the margins of existing UN gatherings in New York or Geneva—and with one meeting potentially focused on opportunities across diverse sectors, and the other focused on risks. The dialogue would also have connections to the work of the scientific panel, which would enhance that dynamic, comparable to the relationship between the IPCC and the UN Climate Change Conference (UNCCC). It would become a venue for voluntarily sharing info about AI incidents that stretched or exceeded the capacity of state agencies to respond.

Our Assessment

Inclusivity challenges

We commend the commitment to make the dialogue inclusive by design, but the dual focus on intergovernmental and multistakeholder dialogue presents a challenge in ensuring meaningful and substantive exchanges. The restrictive modalities of UNGA, and even increasingly with Geneva-based meetings, particularly for non-governmental actors such as civil society

organisations, further complicates the facilitation of a genuinely inclusive dialogue. There should be a clear commitment to avoiding conversations happening in parallel uncommunicated tracks that take away the benefits of cross pollination of views. The special office, tasked by the Global Digital Compact of implementing the Policy dialogue, should apply the recommendations within the NETmundial+10 outcome document to ensure meaningful and substantive stakeholder engagement in the context of an intergovernmental process.

Grounding in IHRL

While the initiative aims to foster interoperability of governance processes, it lacks explicit reference to grounding this in the International Human Rights Law (IHRL) framework and the expertise of the UN human rights system, including but not limited to the OHCHR, B-Tech Project, and treaty bodies. Promoting common understandings on implementation of AI governance measures by different sectors at the international level with the aim of encouraging best practices sits more closely to international cooperation. This is not the same as supporting a normative coordination grounded in the baseline human rights protections related to AI that are already outlined in the work of UN human rights bodies. This lack of provision for normative coordination could undermine the forum's ability to provide a trusted space for exchanging views to avoid a race to the bottom, particularly with respect to preventing or addressing AI incidents and ensuring accountable governance. As an initiative focused on exchange of practices rather than the implementation of a normative common ground, it risks becoming an exchange with little concrete influence in regional or national governance efforts, as well as duplicating other relevant policy forums such as the IGF or AI Summits.

Risks of political capture

We are also concerned with how the report frames the relationship between the proposed Policy Dialogue and the International Scientific Panel: particularly the suggestion that it would be comparable to the relationship of the IPCC and the UN Climate Change Conference (COP). This requires significant caveats and guardrails to avoid reproducing underlying issues in the IPCC/ COP dynamic. In particular, it's critical that the HLAB-AI addresses potential risks of policymakers selectively using Scientific Panel findings to

support their own agendas, and learns from how COP political negotiations have sometimes served to water down IPCC commitments.

Recommendation 3: AI Standards Exchange

The AI Standards Exchange would have a mandate of developing and maintaining a register of definitions and applicable standards for measuring and evaluating AI systems, debating and evaluating the standards and the process for creating them, and identifying gaps where new standards are needed.

The report notes that the UN system could serve as a global clearing house for AI standards, bringing together representatives from national and international standard-development organisations (SDOs), technology companies, civil society, and the Scientific Panel.

Our Assessment

Transparency and diversity needed to avoid coordination challenges

The focus on bringing together diverse representatives from a range of SDOs is beneficial for comprehensive standard development monitoring, reflecting a multistakeholder approach and the perspectives of diverse actors. We welcome this but stress that standards evaluation and its decision-making processes should be grounded in transparency, making it clear how stakeholder input has been weighted in the development of the standards concerned. It is also critical to consider that AI systems operate across diverse regional and cultural contexts. Socio-technical standards must therefore have global applicability whilst maintaining enough flexibility or adaptability in more specific contexts, which can be achieved through inclusive and diverse stakeholder involvement.

Effective integration of SDO work

The focus on maintaining a register of standards, evaluating existing ones, and identifying gaps appears constructive. However, we are unsure how the standards work undertaken by this exchange would be grounded in a

common understanding of meaning. The current landscape of SDOs is currently quite fragmented —evidenced by numerous struggles connected with the work developed by different SDOs. Coordination across SDOs rarely happens, and when it does it is only due to conscious individual efforts. It is wishful thinking that the AI standards exchange will achieve the intended coordination if SDOs are not integrated within its work in a more structured way. SDOs have not yet been at the forefront of the integration of socio-technical considerations, which requires an additional layer of considerations—among them, the integration of human rights impacts in their evaluation.

Finally, if this is solely intended to be a repository, it is not clear why it would need to be separate from the mandate of the Scientific Panel. Independent evaluation of the scientific value of AI standards would greatly benefit from the expertise of the stakeholder mix proposed for the Scientific Panel. This would also help ensure that socio-technical scientific choices are kept separate from political or economic considerations that could interfere with adoption and deployment.

Recommendation 4: Capacity Development Network

The AI Capacity Development Network aims to link collaborating, UN-affiliated capacity development centres, providing expertise, computing and AI training data to key actors and to "serve as a matching function" that expands the range of possible partnerships to enhance the interoperability of capacity building approaches. The network would align regional and global capacity-building approaches, and build the AI governance capacity of public officials. It aims to promote a bottom-up, cross-domain, open, and collaborative effort—with a focus on applying AI to local public interest use cases, including through protocols, sandboxes, online resources, and a fellowship program.

Our Assessment

Incentives for access and transparency in resource allocation

The initiative lacks detailed information on incentivising access to trainers, computing, and AI training data, or on relationships with external stakeholders. To ensure legitimacy of the network, careful guidance should be provided on making transparent the provenance of the resources that will be part of this pool, and the role and responsibilities of the identified nodes in order to avoid mission creep. When resources are not committed with transparency and independence, their allocation is not assured, and diversity of representation is unlikely to be achieved. The proposed network sounds like it would be relatively resource intensive. In considering this initiative, the implementers of HLAB-AI should consider whether training programmes could be scaled without compromising quality.

Human rights anchor

There is a risk that the creation of capacity development networks without a human rights anchor could lead to opaque and uneven allocation of resources and learning opportunities, potentially facilitating abuses, appropriation for personal benefits, political interference in access to the network or even fragmentation of nodes' approaches to providing access to opportunities. The emphasis on access for local public interest use cases and protocols for appropriate access to computing resources are positive aspects, though the robustness of this model needs further scrutiny and guidance to ensure it operates in a distributed manner.

Recommendation 5: Global Fund for AI

The Global Fund for AI aims to connect talent, computing power, and data, leveraging in-kind support to access AI-related models and curated datasets at lower than market cost. It would have an independent governance structure, with support from public and private sources, and disburse resources via the Capacity Development Network. Its mandate would include sharing computing power, providing sandboxes and

benchmarking tools, and creating a repository of AI models and curated datasets for SDG-related projects.

Our Assessment

Ensuring multistakeholder governance

The reference to an independent governance structure is positive. However, it should also provide for multistakeholder participation in governance or, at minimum, multistakeholder oversight and multidisciplinary expertise. Transparency around the structure's relationship to public and private sources and the process for leveraging in-kind support is essential, particularly with respect to training programmes and distribution within the Capacity Development Network.

Need for a research agenda that upholds human rights by design

We are concerned that the structure and operation of such a fund is not oriented towards a dynamic research agenda which is built and devised with participation of expert bodies including the Capacity Building Network and the Scientific Panel. The report references CERN as a model that lends itself towards actionable practices. However, there are relatively weak commitments in the report on orienting capacity built with resources from the Fund towards a critical assessment of the pertinence of AI deployments for the achievement of the SDGs. Nor is there much said on the governance of AI, and how to avoid unintended harmful consequences for safety and security, beyond references to the "governance stack". We stress the importance of taking a proportionate approach to measure the potential benefits and risks that AI development, deployment and use may pose to human rights and SDG realisation. This should involve a comprehensive assessment of potential human rights impacts, including any possible positive or negative impacts on SDG fulfilment.

It is imperative that funding processes are transparent to ensure human rights compliance and avoid opaque geopolitical and economic practices which do not serve the public interest. To this end, the Global Fund should adhere to established principles for international assistance programming to ensure that its investments promote a rights-respecting agenda and uphold the principle of "do no harm" and report publicly on its investments so that

these can be subject to public scrutiny. The support for open access to curated datasets and models for SDG-related projects is commendable, but it needs to be balanced against concerns related to environmental impacts and techno-solutionism.

Recommendation 6: Global AI Data Framework

The proposed Global AI Data Framework seeks to establish a standardised, international system for managing AI training data. Initiated by agencies such as UNCITRAL and supported by international organisations, the framework aims to define global data-related principles, promote transparency and accountability, and facilitate the exchange of anonymized data. By creating data trusts and well-governed global marketplaces, the initiative intends to support diverse AI ecosystems and address existing gaps in data-sharing and accessibility, thus enabling inclusive and vibrant AI development.

This initiative would emphasise the importance of cultural and linguistic diversity in AI training data and the need for transparency regarding underrepresented people or missing data. It advocates for a "data commons" to incentivize data curation and establish model contracts and techno-legal protocols to ensure privacy, data protection, and interoperability. It would aim to create international "guard rails" and "common rails" that promote equitable and ethical use of AI training data across borders.

Our Assessment:

Non-binding nature = implementation challenge

The proposed framework has the potential to significantly enhance the quality and diversity of AI training data. The emphasis on data stewardship and the creation of a data commons can support equitable access to high-quality training data, contributing to the realisation of the SDGs and the protection of cultural heritage. The development of the Global AI Data Framework aims to be inclusive, involving a wide range of stakeholders from various countries and international organisations. This collaborative process

should ensure that diverse perspectives and needs are considered, enhancing the legitimacy and acceptance of the framework.

However, there is a clear risk that the non-binding nature of the proposed framework will not result in voluntary contributions necessary to ensure the utility of this framework. Moreover, there is also a risk of inconsistent implementation across jurisdictions. Achieving consensus on issues like data privacy and proprietary data access could prove difficult.

Who shepherds the commons, who benefits from it?

There are significant challenges around ensuring the financial and operational sustainability of the proposed data trusts and marketplaces. Not only is there a risk that the framework might inadvertently favour more developed nations or large corporations, as well as fostering greater dependencies on existing powerful players in the landscape, the complexity of implementing and maintaining interoperable systems across diverse legal and technological landscapes could lead to uneven benefits.

Integration with broader data protection frameworks

There is also a broader issue with these data-related principles and the framework being developed in an AI-specific manner, as opposed to a more holistic approach that builds on current data protection and international data transfers— encompassing new and emerging technologies that are equally data-intensive, and which pose similar risks and opportunities.

Recommendation 7: An AI Office Within the Secretariat

The creation of an AI Office in the UN close to the UN Secretary-General is proposed to act as a "glue" to unite the recommended initiatives efficiently and sustainably, helping to avoid fragmentation and missed opportunities in the fast-emerging ecosystem of international AI governance.

Our Assessment

Impact of Institutional choice on effective participation

The impacts of the AI Office will ultimately depend on the mandates and operations of the various initiatives recommended in the report. Careful attention must be paid to how these initiatives are integrated and overseen to ensure coherence and effectiveness in AI governance.

While there is value to be drawn from closer coordination, the positioning of these various initiatives within an overarching office, anticipated to be located in New York, carries the risk of further centralising digital technology policymaking which could hinder the effective participation of civil society and some Member States. Up to now, digital policymaking has been carried out through a range of UN institutions and multistakeholder venues which have, to differing extents, embedded open, inclusive and transparent approaches to engagement of non-governmental actors. At a minimum, the AI Office should incorporate best practices of stakeholder engagement into its ways of working and develop clear structures for cooperation with existing bodies in the design and implementation of the proposed initiatives.

Alternative model for capturing the strengths of the UN

As captured above, an AI Office anchored in the UN Secretariat is not necessarily the most holistic way of approaching the opportunities and challenges presented by new and emerging technologies. Other proposals, such as the Human Rights Advisory Service in the digital space, facilitated by the Office of the United Nations High Commissioner for Human Rights (OHCHR) and originally proposed by the Secretary-General in his Policy Brief on the Digital Compact, seem better suited to leveraging the already existing and relevant experience of UN human rights bodies dealing with the impacts of new and emerging technologies. This alternative proposal would also harness the experience of other UN agencies with existing relevant work on AI—such as the ITU, WHO, UNCTAD, and UNESCO. Relatedly, the GDC—even while recognizing the role of the OHCHR in providing an advisory service on human rights in the digital space—fails to provide it with stronger mandate and resources to play the coordination role it could have related to AI governance.