

Derechos Digitales’ inputs to the consultation from the UN’s AI Advisory Body (AIAB) on its Interim Report: Governing AI for Humanity¹

Submitted on March 29th, 2024

Opportunities and Enablers

The section starts with a deterministic and techno-solucionist approach that might lead to a mistaken interpretation that technology itself can solve complex social, economic, environmental or political problems. Rather than falling into a binary opposition between good or bad uses of AI, the report should adopt a sociotechnical approach and focus on the enablers that can allow it to benefit all. In the same sense, the AI Advisory Body (AIAB) report should avoid referring to anecdotal cases of potential “beneficial” uses of AI and overcome the idea of “AI for good”, thus completely eliminating the BOX 2.

There are documented cases of AI use in the public sector that not only do not comply with existing requirements regarding public procurement, transparency, accountability and oversight; but have also resulted in serious abuses to human rights. On the contrary, while complying with its mandate to advance “recommendations for the international governance of AI”, the AIAB report should focus on the urgent need to achieve State’s commitment to adopt and support standards that prevent such abuses in all implementations of AI, no matter under which alleged purposes. For this, it should clearly and explicitly recommend the implementation of Human Rights Impact Assessments (HRIA) within the different stages of the AI lifecycle, advancing on the Unesco Recommendations 50 to 53 and A/HRC/48/31 recommendation 60 (a). Such processes should be widely participatory and publicly accessible. Respect to indigenous people’s right to prior consultation and the peasants and other people working in rural areas’ right to meaningful participation should be highlighted when it comes to any intervention (at the infrastructural level or not) that may affect their rights.

Compliance to human rights standards – including social, economic, cultural and environmental rights, that should also be part of HRIAs – is the most relevant enabler for “harnessing AI for humanity”. The report should highlight how global commitments and agreed oversight mechanisms are necessary to assure that AI production attends to environmental and labor law through its full chain of production and propose solutions. To date, it is known that while some countries benefit from AI applications, others are burdened with e-waste, predatory mining, sometimes based in child labor, and precarious work – including for data labeling. To tackle global imbalances in the AI economy, considerations on taxation and competition should be considered, as well as others, such as certification mechanisms.

¹ Please note that the structure of our submission had to be slightly changed to fit character limitations within the form made available for the consultation. The content was fully preserved, although divided in different questions. For the AIAB interim report, please visit: https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf.

Alternative models, such as community networks, should be considered when advancing connectivity infrastructure, as well as non-extractivist alternatives for AI development based on community agreements or others.

Risks and Challenges

This section also needs a review of the binary understanding of AI (as either “good” or “bad”) and, again, a sociotechnical approach would be necessary to overcome the idea of AI risks as either fully technical or human. Historical and cultural factors embedded in databases and models reflect not only pre-existing inequalities, but also world views that become part of the system and which can influence its use and results. Automated content rating and moderation systems present an accessible example of such, facial recognition systems and their flaws in recognizing certain people’s faces and certain faces as people’s is another.

The AIAB report should refrain from listing potential risks and building new categorizations. While categorizations already abound, most of them are incomplete as they don’t capture complexities. A proper alternative would be to prioritize an integral human rights perspective, adopting an approach that considers at least the following levels of possible impact: civil and political, social and economic, cultural and environmental. In that sense, the report should not only expand on the role of human rights compliance throughout the full AI lifecycle, but detail the existing standards to deal with the balance of fundamental rights – escaping the idea of “trade-offs” and offering proper guidance to decision-makers to operate in different situations and contexts. The report should expand on the application of legality, necessity and proportionality as indicated in A/HRC/48/31 recommendation 60 (b).

As risks are never purely technical or human, as those are intertwined in a mutually influential relationship, no meaningful response can come from focusing exclusively on one or other side. Even a system with allegedly less than 1% error rate can be harmful if massively implemented, depending on the level of impact of such error. In this sense, the report should reinforce the recommendations from A/HRC/48/31 (c) and (d) of banning AI applications that “cannot be operated in compliance with international human rights law and impose moratoriums on the sale and use of AI systems that carry a high risk for the enjoyment of human rights, unless and until adequate safeguards to protect human rights are in place”, in particular, imposing a “moratorium on the use of remote biometric recognition technologies in public spaces”.

Incentives should be in place to assure that AI research, design and development consider diversity, equity and inclusion criteria from an intersectional gender perspective. HRIAs should be in place within all the AI lifecycle and also be participatory. Participation shouldn’t be mere formal, but meaningful, encompassing concrete incentives to the representation of different potentially affected groups, with emphasis on historically marginalized populations.

Guiding Principles to guide the formation of new global governance institutions for AI

Considering this document, correctly, attempts to leverage the role of the UN as a legitimate institution to advance the basis for a truly global governance of AI, the UN Charter and the Universal Declaration on Human Rights, should be the main basis adopted and as such the first and cross-cutting principle that crosses all existing and future institutions for AI and AI’s full lifecycle. Similarly, existing advances and mechanisms within the UN to protect and implement international

human rights law should be reinforced and strengthened in any future global AI governance structure. This includes the Human Rights Council, its mechanisms and entities.

On Principle 1, access and capacity building will not be able themselves to address global inequalities regarding AI. Standards should be reinforced and operationalized to ensure business enterprises comply with their responsibilities within the Guiding Principles on Business and Human Rights and make them accountable if not compliant. Similarly, a global AI governance structure will need to advance concrete mechanisms to prevent the circulation of technology that does not comply with international human rights law.

On Principle 2, the need for binding norms at the national level is critical as the report recognizes. However, global divides are not limited to AI production and deployment, but expand to regulatory and oversight capacities. While international cooperation mechanisms should be considered to allow knowledge and capacity exchange, countries should be able to exercise their sovereignty in their regulatory approaches. The AIAB report should recognize that a diversity of regulatory models are possible and recommend countries to adopt legislative responses that align with their contexts and priorities (defined with proper input from participatory processes). This sort of contextualized approach implies the revision of the idea of interoperability, which, if prioritized as a goal, risks suppressing the need for national states to decide on their approaches, and promoting a race to the bottom in terms of regulation. Additionally, the report should recognize the existence of legislation in place that should be observed in the context of AI, including access to information and data protection norms.

Principle 3 mixture of different ideas such as data protection and data commons risks emptying the principle completely. While the latter could be further conceptualized and developed – considering existing criteria to ensure a proper balance between transparency and access to data and privacy –, the report should focus on reinforcing the importance of data protection for AI governance and incentivize the adoption of strong personal data protection rules and enforcement mechanisms by States. It should validate at least the core data protection principles established in different regional and national standards including: legality, necessity, proportionality, minimization, purpose limitation and transparency. Specific limits to the processing of sensitive information should also be recommended. Incentives should be put in place to assure that AI is only trained on databases that comply with such basic data protection principles – something that should also be observed by any data commons initiative, with particular emphasis on the purpose limitation aspect. Data collected within the delivery of public services, for instance, shouldn't ever be repurposed to train AI systems without specific knowledge and consent by affected people and groups.

It is evident that all countries should be able to participate meaningfully with voice and vote in any discussion or decision regarding AI that has impacts on the global level. The report should limit itself to reinforce that a global AI governance structure cannot replicate existing imbalances or prioritize the views of some over others. This should be reflected in its review and weight to existing AI standards, several of them lacking a truly global character. Principle 4 should focus on strengthening the need for civil society participation in decision-making processes at national and global levels throughout all the AI lifecycle, and on the enablers necessary for it, including the guarantee of democracy and a vibrant civic space.

Principle 5 should reiterate the recommendation on the deployment of HRIA throughout the AI lifecycle, as well as considerations on limiting the circulation of AI that cannot comply with human

rights law – which should be reinforced and operationalized. Elements of access to information, transparency, accountability and human oversight should also be included. AI governance frameworks do not grant blanket exemptions from basic rules related to transparency, accountability, accuracy, and quality of AI systems on the grounds of national security and/or counterterrorism.

Institutional Functions that an international governance regime for AI should carry out

A lot of questions emerge for the structure defined in this section, particularly regarding how the governance framework would look like and which new or existing institutions would take which role. Some points of attention are the following:

1. Multistakeholder participation, in terms defined by the WSIS process, should be transversal to the different functions with concrete measures to enable meaningful engagement from civil society actors globally. The ECOSOC Consultative Status should be strengthened to facilitate civil society participation in UN decision making spaces. Hybrid modalities should always be offered in meetings, with proper procedures to assure equal participation from both in-person and online participants, as well as measures for accessibility in different levels. Selection or nomination processes should assure regional, gender and race diversity, but also that participation is meaningful and reflect a commitment with and connection to broader processes, discussions and movements, rather than individual motivations. Tokenist practices of presenting diversity without concrete commitments to make equitable participation possible should be avoided. Established existing mechanisms for nomination by different stakeholders should be recognized and validated by all institutions involved in global AI governance, particularly the processes undertaken within the Internet Governance Forum, its national and regional initiatives and intersectional work. Transparency and accountability regarding consultation processes should be in place by all institutions involved in AI governance.
2. Beyond exercises of futurology, any research or scientific body should depart from existing production and evidence on the impacts of AI in different regions and to different rights in order to properly inform decision making at different levels. While analyzing scientific production on AI, consider that most of the production regarding the impacts of AI implementation to human rights are outside the academia and is often led by civil society. Affirmative action to include members and production from outside Global North, developed by women, women of color and LGBTQIA+ people should also be in place so as to not replicate inequalities coming from the academic and scientific publication sector.
3. No claim on interoperability should impose over States sovereign power to regulate AI at the national level. As indicated above, context is key in the definition of potential harms and benefits from AI. The report should focus on the enablers and criteria to be considered for such regulation, and completely eliminate any reference to legal operability from the text, at least until there is enough clarity on its implications. Policy-making at the national level, while grounded on the respect for human rights law, should be guided by local evidence and informed by meaningful participation of different stakeholders operating at the local level without constraints from any international imposition. States should have, for instance, the power to limit the circulation of certain AI products or applications – depending on the risk they present in their particular context –, establish additional obligations to business enterprises (inspired by the Guiding Principles on Business and Human Rights), as well as to impose taxation measures or apply antitrust law if necessary so they can benefit from the

- digital economy. Derived from its commitment to tackle global imbalances within the AI economy, the AIAB should, thus, additionally recommend against the establishment of limitations to countries policy development in other fora, such as within trade agreements.
4. When it comes to standards, a priority would be to define criteria to analyze at the global level the compliance of certain AI products and applications with international human rights law. The report mentions standardization for access and use of infrastructure, but there is an urgent need to define mechanisms for monitoring, oversight, supervision and certification that allows harm not to be replicated. The UN has a key role to play in assisting countries with information regarding, for instance, environmental and labor compliance, for instance, elements that are harder to trace within the AI production chain from the national level. Similarly, it could offer references as to if a system or model was considered not appropriate in a certain country due to its human rights impacts. Again, care should be taken in attempts to standardize HRIAs at the global level since they may miss contextual nuance, exclude valid concerns from local potentially affected communities and represent a race to the bottom on common minimum standards.
 5. The global governance of AI should allow further dialogue and integration of different actors at national, regional and global levels operating in this issue and offer support to States in dealing with extra-territorial aspects of AI research, design and development. International cooperation, thus, should focus on facilitating better decision-making with information and monitoring compliance. It is not its role to support the adoption and implementation of AI by countries, regardless of the purpose – as no universal idea of “AI for good” can be sustained (see above). The advance of Sustainable Development Goals (SDGs) shouldn’t be used as an excuse to promote non-compliant technology. As it is known, SDGs depend not on AI, but on political will and concrete commitments by States in coalition with other stakeholders at the national and international levels. The report should highlight how investments for digitalization and AI deployment should be carefully considered under necessity and proportionality criteria, so as not to divert necessary investments in more fundamental welfare programs that often threatened by reductions, as the Special Rapporteur on the right of everyone to the enjoyment of the highest attainable standard of physical and mental health has recently recalled (A/HRC/53/65, paragraph 84).
 6. International cooperation for the development or deployment of AI systems should also be subject to prior HRIAs and respond to strict transparency and accountability criteria. A global AI governance structure should have a function to oversee international institutions, including within the United Nations, and ensure that their use and support to AI is aligned with international human rights law.

Other comments on the International Governance of AI section (aside from Principles and Functions, covered in above questions)

Global governance frameworks should not adapt to business enterprises' transnational practices for evading accountability. On the contrary, it should enhance mechanisms and cooperation to ensure that human rights are respected by all stakeholders within the full AI lifecycle and the private sector has a key responsibility in this regard. More than focusing on harmonizing risk-based frameworks or attempting to establish global standards to be imposed on countries under arguments of interoperability, the report should reinforce international human rights as the main basis for assessing AI systems and give proper guidance to States on how human rights criteria should be considered in concrete cases.

Reclaiming the UN's role for its truly global nature is fundamental, as it is to assure equitable participation by States on decision-making, together with mechanisms for multistakeholder participation. Such role should go beyond international security to ensure transnational corporations, international bodies – including development banks –, and States are accountable when it comes to AI development and deployment. UN human rights bodies should be involved in such process and be duly strengthened to fulfill such responsibility.

In the same sense, a reviewed version of this report should account for all the developments and recommendations within the Human Rights Council (HRC) regarding AI and explicitly incorporate them. This includes not only the HRC Resolutions on New and Emerging Technologies and Privacy in the Digital Age, as well as the vast production from the High-Commissioner on Human Rights (A/HRC/48/31) and, particularly, from experts from Special Procedures, who have already addressed impacts of AI on, for example: extreme poverty and welfare programs (A/74/493); health (A/HRC/53/65); discrimination (A/HRC/44/57); migration and border control (A/HRC/48/76), and people with disabilities (A/HRC/49/52).

Any other feedback on the Interim Report (Maximum 3,000 characters)

To be a meaningful addition to existing discussions, this report should advance on possible models and expand on the various references of institutions cited throughout the text – ranging from CERN to ICANN –, justifying why they are considered relevant for AI governance.