



The Hague, September 29, 2023

Subject: Global AI Governance – Input for the UN High-Level Advisory Body on AI

To the attention of the Office of the UN Tech Envoy:

ECNL is pleased to respond to the UN Tech Envoy's call for papers on global AI governance and provide input as part of the creation of the UN the High-Level Advisory Body on AI. We outline herein key issues on global AI governance that the Advisory Body prioritize. The below recommendations and insights are based on ECNL's 20 years of work in the area of enabling environment for civic space, participation and recently, digital rights, following substantial research and practical experience collaborating with civil society, AI developers and deployers, and policymakers around the world.

We look forward to engaging further with the UN Tech Envoy and the Advisory Body and are available to discuss or elaborate on any of the issues addressed in this paper. Please do not hesitate to contact us at vanja@ecnl.org (Vanja Skoric) or marlena@ecnl.org (Marlena Wisniak).

Best regards,

Vanja Skoric and Marlena Wisniak





I. A human rights-based approach aligned with the UN values

Policies, standards and regulation should be regulated for AI from a human rights-based approach. Framing AI solely from a risk-based approach that focuses solely on operational and business risks and those related to external factors such as security fails to adequately protect people and society, as well as undermines the meaningful implementation of the Sustainable Development Goals (SDGs). It has been generally accepted in the recent multilateral and multistakeholder discussions¹ that AI systems do not operate in isolation. Moreover, a socio-technical approach that includes universally accepted rights and societal values is needed to detect and mitigate harms and risks. Technical and operational measures alone cannot resolve issues and tensions that are inherently social, political, or legal. Such measures must therefore be complemented by the universally accepted international human rights framework, based on the core UN covenants and SDGs.²

We therefore urge the UN Tech Envoy and the Advisory Body to consider risks within a rights-based framework, i.e., risks to a broad range of human rights, including political, economic, and social and cultural rights, within future efforts to develop AI risk taxonomies and responses. In addition, these efforts must be aligned with the implementation of the SDGs.

¹ Human Rights Council, Resolution on new and emerging technologies, A/HRC/RES/53/29, 14 July, 2023, < <https://undocs.org/Home/Mobile?FinalSymbol=A%2FHRC%2FRES%2F53%2F29&Language=E&DeviceType=Desktop&LangRequested=False>>; UN Special Procedures, (2023), Joint Statement: New and emerging technologies need urgent oversight and robust transparency, < <https://www.ohchr.org/en/press-releases/2023/06/new-and-emerging-technologies-need-urgent-oversight-and-robust-transparency>>; Human Rights Council, The right to privacy in the digital age, A/HRC/51/17, 4 August 2022, < <https://www.ohchr.org/en/documents/thematic-reports/ahrc5117-right-privacy-digital-age>>

² Human Rights Council, The Core International Human Rights Instruments and their monitoring bodies, <https://www.ohchr.org/en/core-international-human-rights-instruments-and-their-monitoring-bodies>; United Nations, Sustainable Development Goals, < https://www.un.org/en/footballforthegoals?gclid=CjwKCAjwyNSoBhA9EiwA5aYlboxygwOhxGUBjuvObEqvwJFAMqIJaoHgXHG3PUvJx_y0Tnmnmn25ZpdRoCTWAQAvD_BwE>.



To operationalise this rights-based framework within overall AI governance and management, we recommend that the Advisory Body includes mechanisms for detecting AI risks and harms in a way that is future-proof. AI developers, decision-makers and policymakers can make informed decisions that balance competing interests and contribute to the responsible, just, rights-based and sustainable development of AI by conducting a comprehensive, inclusive *ex ante* impact assessment and meaningfully including relevant stakeholders. Such an approach is increasingly recognized and standardised within ongoing regulatory and standard-setting efforts globally.³

Assessing the impact of AI systems on human rights should be conducted using the already established methods of human rights impact assessment, in line with the UN Guiding Principles on Business and Human rights (UNGPs) and broadly accepted human rights due diligence process. We recommend that the UN Tech Envoy and Advisory Body include human rights due diligence, and within those, human rights impact assessment, as a key component of Global AI Governance, as consistent with the UNGPs.

Where AI systems are fundamentally incompatible with human rights, the Advisory Body should call for UN Member States to ban, or at least set a moratorium, for the development or deployment thereof, in line with the ongoing requests by the UN special procedures and bodies.⁴

³ E.g., the U.S. National Institute of Standards and Technology (NIST) AI Risk Management Framework (2023); International Organization for Standardization ISO proposed ISO/IEC 42001 and 42005; Brazil draft Artificial Intelligence Bill; European Union AI Act; Council of Europe Convention on AI; US Blueprint for an AI Bill of Rights; Canada Algorithmic impact assessment directive; Costa Rica AI Law proposal.

⁴ UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, (2019), Surveillance and human rights, A/HRC/41/35, <<https://www.ohchr.org/en/press-releases/2019/06/un-expert-calls-immediate-moratorium-sale-transfer-and-use-surveillance>>; Human Rights Council, (2022), Resolution on the Promotion and Protection of Human Rights in the Context of Peaceful Protests, A/HRC/50/L.16 at para 30-31.



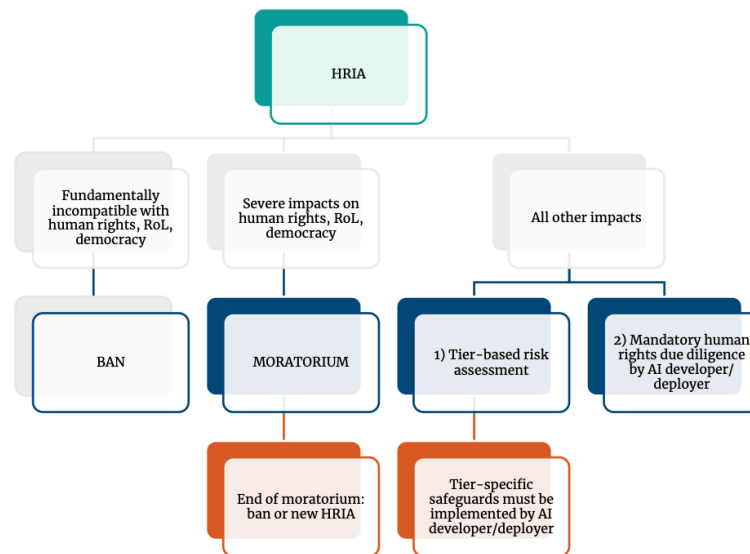
II. Criteria for determining the risk level of AI systems through human rights impact assessment

A thorough, inclusive, and transparent human rights impact assessment (HRIA) must be the starting point for all subsequent regulatory actions or governance of any AI system. Although HRIAs are no silver bullet for addressing the risks of AI systems, they can contribute to AI accountability and transparency, reducing the potential future liability concerns as well as determining the AI risk level, if certain key conditions are met.

a) AI risk levels and appropriate actions

In ECNL's research paper (2021),⁵ we outline key criteria for evaluating the risk level of AI systems from a human rights-based approach. We recommend determining a level of risk based on an assessment of the product design, the severity of the impact, any internal due diligence or compliance mechanisms, a causal link between the AI system and an adverse human rights impact, and the potential for remedy, among other factors. We note that stakeholder engagement and internal and external reporting are cross-cutting measures that should apply throughout the process.

⁵ ECNL, 'Evaluating the Risk of AI Systems to Human Rights from a Tier-based Approach', 2021, < https://ecnl.org/sites/default/files/2021-06/Evaluating%20the%20Risk%20of%20AI%20Systems%20to%20Human%20Rights_ECNL%20proposal.pdf>.



Map of determining AI risk level

Depending on the risk of the AI system, and the specific needs raised during the HRIA process, AI developers and end-users should take appropriate measures to mitigate and prevent risk. These include:

1. Ban. AI systems that are inherently incompatible with human rights should be permanently banned.
2. Moratorium. AI systems that have severe impacts on actual and potential human rights should be temporarily paused. At the end of the moratorium period, a new HRIA should be conducted, and further actions re-evaluated.
3. Regulation prior to deployment. To prevent adverse impacts on human rights, all AI systems that do not fall into the ban or moratorium categories are subject to further appropriate levels of regulation prior to deployment, including tier-based risks assessments and iterative human rights due diligence.

b) Meaningful human rights impact assessments for AI

In a joint paper with Data & Society (2021),⁶ we outline 10 key components of meaningful HRIAs:

⁶ ECNL, Data & Society, 'Recommendations for Assessing AI Impacts to Human Rights, Democracy, and the Rule of Law', November 22, 2021, <





1. **Sources of Legitimacy:** Impact assessments need to be legitimized either through legislation or within a set of norms that are officially recognized and publicly valued. Human rights principles should be the root source of legitimacy for HRIAs, and made concrete through legislative mandates.
2. **Actors and Forum:** Impact assessments are rooted in establishing an accountability relationship between actors that design or deploy a system and a forum that can allocate responsibility for potential consequences of such systems and demand changes in their design, deployment, and operation. Forum should be empowered by law or similarly robust social, political, and economic norms.
3. **Catalyzing Event:** Points in the development and/or procurement process that trigger a requirement to conduct impact assessments. These can be mandated by law, or solicited voluntarily at any stage of a system's development life cycle.
4. **Time Frame:** Once triggered, the time frame is the period often mandated through law or mutual agreement between actors and the forum within which an impact assessment must be conducted. Most impact assessments are performed *ex ante*, before developing a system, but they can also be done *ex post* as an investigation of what went wrong.
5. **Public Access:** Achieving genuine transparency and accountability requires the ability of the public to scrutinize and contest an impact assessment's process and documentation. The broader the public access, the stronger is its potential to enact accountability.
6. **Public Consultation:** The conditions for solicitation of feedback should be from the broadest possible set of stakeholders in a system. Who constitutes this public and how they are consulted are critical questions for the success of an impact assessment.
7. **Methods:** These are standardized techniques of evaluating and foreseeing how a system would operate in the real-world. Most impact assessments have a roster of well-developed techniques that can be applied to foresee the potential consequences of deploying a system as impacts.
8. **Assessors:** Impact assessments are conducted by assessors. The independence of assessors from the actor as well as the forum is crucial to how an assessment process identifies impacts, how those impacts relate to

https://datasociety.net/wp-content/uploads/2021/11/HUDIERA-Full-Paper_FINAL.pdf>.



tangible harms, and how it acts as an accountability mechanism that avoids, minimizes, or mitigates such harms.

9. **Impacts:** Impacts are proxies for the effects of the deployment of a system in the real world. They are what assessors document through their assessment methods, and are used to identify, measure, and ameliorate potential harms. When communicated to a forum, they enable the forum to mandate changes to the system being assessed, or to otherwise hold actors accountable.
10. **Harms and Redress:** Harms are lived experiences of the adverse consequences of a system's deployment and operation in the real-world. Some of the harms can be anticipated through impact assessments as impacts, others cannot be foreseen. To secure justice, redress procedures must be developed to complement any anticipated harm to secure justice.

III. Stakeholder engagement

Stakeholder engagement is a necessary and cross-cutting component of AI governance, development, and use. We understand stakeholder engagement as a collaborative processes where diverse stakeholders (both internal and external) have can meaningfully influence decision-making.

We urge the UN Tech Envoy and the Advisory Body to prioritize and properly resource stakeholder engagement of civil society and affected communities, both in their own activities and in broader global AI governance, centering at-risk and marginalized groups.

In the context of HRIAs of AI systems, stakeholder engagement is particularly effective: a) to understand potential problems or opportunities in which products or services using AI, machine learning or algorithm-based data analytics may potentially contribute to improvement, or where it might not; and b) to identify specific potential positive or adverse impacts, implications, benefits and harms of these products or services on people's individual and collective human rights, especially marginalized and already vulnerable groups.

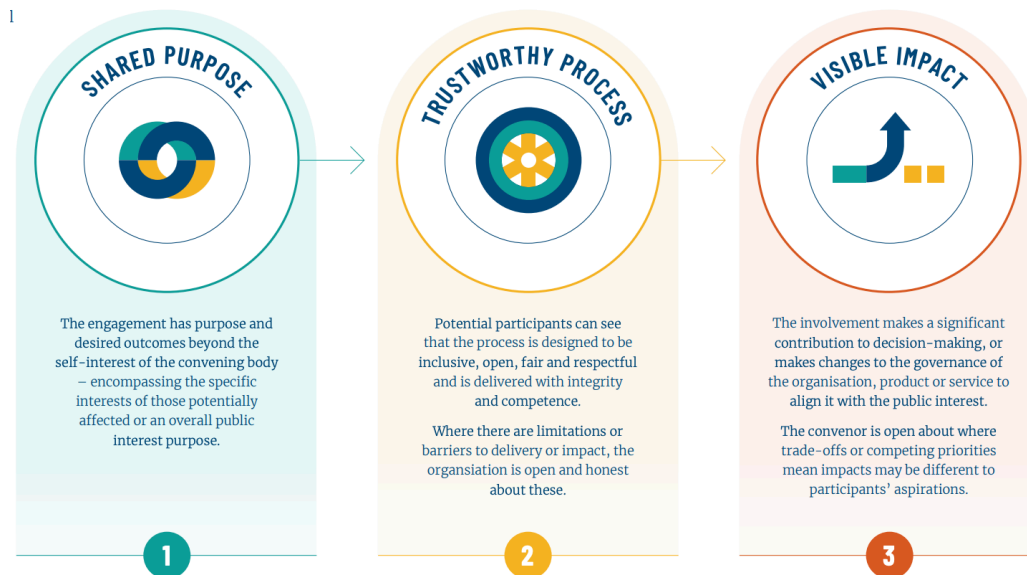
In ECNL's tool framework for meaningful engagement (2023),⁷ we provide concrete recommendations for engaging internal and external stakeholders in AI

⁷ ECNL, Framework for meaningful engagement: Human rights impact assessments of AI, August 2023, < <https://ecnl.org/publications/framework-meaningful-engagement-human-rights-impact-assessments-ai>>.





systems in a way that truly protects and promotes human rights. We find the three elements of meaningful engagement are the following:



In the framework, we explain why creating a shared purpose is important. We then provide evidence-based recommendations for understanding and addressing barriers and limitations, deciding when and who to engage, and choosing engagement methods. We also define minimum conditions for an engagement process to be considered trustworthy. These include inclusion, openness, fairness, respect, integrity, and competence. Finally, we unpack how AI developers and deployers can demonstrate the impact of engagement on product design and use.

IV. Transparency measures

Meaningful transparency is needed to build public trust in the use of AI, empower policymakers, researchers, users and affected communities to scrutinize and enhance their engagement with AI developers and deployers. This requires centering the users and affected groups as the target audience and providing information about AI systems to them in an understandable way. **The Advisory Board is well positioned to suggest global and harmonized metrics for algorithmic transparency.**

We recommend that the Advisory Body provide guidelines to outline what specific information AI developers and deployers should report on. These guidelines can inform national regulation by Member-States on transparency provisions of AI





development and use, a prerequisite for AI accountability and rights-based global AI governance.

For instance, AI developers or deployers should report on datasets (including training datasets), performance and accuracy metrics, false positive and false negatives, human in the loop and human review, and access to remedy. A non-exhaustive list of more detailed information to report on can be found in a joint research paper with Access Now on risk assessments for the EU Digital Services Act (see part 22, algorithmic systems and design).⁸

We note that AI consumer-facing products, such as AI-driven social media or education platforms, often require less technical information. **The Advisory Board could for example establish a set of metrics for AI systems in the form of “nutrition labels,”** which are user-friendly notices that can operationalize transparency goals at scale. A study on algorithmic amplification of social media platforms shows how nutrition labels could enable meaningful transparency and what metrics could apply.⁹

Labels are no solution on their own to ensure transparency and accountability; many challenges related to business and profit maximization, implementation, operationalization, and digital literacy can severely limit their effectiveness and impact. Nonetheless, labels can be a helpful tool in the broader algorithmic transparency toolkit and are worth exploring further. Civil society and affected communities are key stakeholders in the design and content of the labels and are critical to ensuring that meaningful transparency is attained.

⁸ ECNL, Access Now, ‘Towards meaningful fundamental rights impact assessment under the DSA, September 2023, <<https://ecnl.org/publications/human-rights-impact-assessments-are-key-effective-dsa-enforcement>>.

⁹ Marlena Wisniak, Luca Belli, ‘What’s in an Algorithm? Empowering Users Through Nutrition Labels for Social Media Recommender Systems’, August 22, 2023, <<https://knightcolumbia.org/content/whats-in-an-algorithm-empowering-users-through-nutrition-labels-for-social-media-recommender-systems>>.