AI & EQUALITY Input Comment

Reacting to a call for input by the Hamburg Sustainability Conference & UNDP regarding a Declaration outlining how AI can advance or hinder SDGs

Our comment focuses on the three Ps of People, Planet, and Partnership.
digital@undp.org

People

Human Rights

AI has the potential to impact almost every recognized human right. Examples where AI is harming the fundamental rights of individuals are increasing, whether in equal access to healthcare (Obermeyer, 2019), work and education (Burke, 2020), safety (Del Bosque, 2023), or an adequate standard of living (Purtil, 2016; Gerchick et al., 2023; Booth, 2024 - for further examples see AI Incident Database. Further, such harms disproportionately affect marginalised communities; both at a local level as well as across the globe: since marginalised communities are less present in datasets and in teams creating AI systems, their perspectives and experiences are often out of scope of what AI systems are created for. The resulting systems range from applications with a low fit to their lives and needs up to systems that increase existing inequalities by enforcing existing power asymmetries, e.g. who systems are designed to serve. Especially developing countries are at-risk as they are often swamped with technology from Western suppliers that are not fit for their context and culture, thus more likely to result in harms due to e.g. cultural bias and exclusion from non-Western/Global Minority perspectives, further enhanced through the linguistic asymmetries enhanced by LLMs (Miller, 2024). The Global Report on Human Rights & AI found that the Global South (except China), will accrue only 10% of the gains from AI by 2030 since many countries in the Global South lack access to basic digital infrastructure and foundational elements of AI like data and compute capacity: in 2024, 2.6 billion people still do not have access to the Internet, including about 65% of households in the least developed countries (ITU, 2024, SDG fulfillment by countries in the Sustainable Development Report). Thus, it is urgent to act now before existing inequalities are further exacerbated and this is everyone's duty: under international law, governments are obliged to respect, protect, and fulfill human rights and fundamental freedoms, while corporations and

business enterprises are required to comply with all applicable laws and to respect human rights.

Specific targets can rebalance the decision power around which systems are developed, thus helping to create a more equal AI future. More concretely, following commitments are a starting point:

- **Going beyond not harming Human Rights:** Setting standards that incentivise AI that is *based on* Human Rights principles such as the AI & Equality Framework. This can help to distribute power and benefits more equally across societies and the globe. This includes:
    - Assessing not only Human Rights harms and risks, but also how systems contribute to the fulfillment of Human Rights.
    - Develop Human Rights-based AI governance that integrates existing international Human Rights frameworks into AI policy approaches , including mechanisms for grievance handling and effective redress for individuals and affected communities.
    - Impact based funding - focussed on and learning from DPI framework (The DPI Approach) as a way to incentivise digital ready systems based on open, secure, and interoperable standards.
- **Rebalancing Who Shapes AI:**
    - Establishment of regional centers and AI safety Institutes and other independent notified bodies focused on localizing and governing AI applications for solving SDG challenges, inclusively and sustainably.
    - Encouraging multilateral co-operation and establishing partnerships fostering inclusive opportunities - for both representations and impact - locally and globally.
- **Developing principles for comprehensive Human Rights Impact Assessments:**
    - Developing Human Rights Impact Assessments (HRIAs), a proactive measure that aims at integrating the human rights legal framework into AI governance strategies whilst allowing tailoring to the deployment of AI systems in a particular state or region (Nonnecke & Dawson, 2021; World Bank, 2013).
        - Conducting such assessments would enable developers of AI systems to appreciate the contexts that are relevant to a particular state or region and to mitigate any biases that could potentially infringe on human rights in those states or regions.
        - Example & potential starting point: The Council of Europe's HUDERIA Methodology
- **Use education as a driver for public awareness of human rights as an important priority in the tech age and as a tool to enhance human flourishing through AI:**

- Evaluate what constitutes AI literacy through a variety of international frameworks (for example, Article 4 of the EU AI Act, the UNESCO Competency Framework for Teachers and the UNESCO Competency Framework for Students). Foster partnership and collaboration between state, regional and global organisations to share educational best practices that foster AI Literacy and durable AI skills for educators, professionals, learners and students (https://digital-strategy.ec.europa.eu/en/library/living-repository-foster-learning-and-exchange-ai-literacy).

Inclusion

AI has the transformative potential to foster a more equitable and sustainable future. AI can accelerate progress toward the SDGs by optimizing resource management through big data analytics and enhanced decision-making. However, to ensure that AI solutions do not perpetuate existing inequalities (see above), it is essential to embed inclusion into principles, concepts, and design. By embedding AI with principles of human rights, diversity and inclusion, it can foster more equitable settings. This is essential across settings, but we especially so for vulnerable populations, ultimately leading to a more sustainable and equitable world in which everyone has "the right to AI" (Mushkani, 2025) – or the "right to NOT AI". Concrete threshold towards such responsible AI are:

- Introducing temporary special measures to ensure representations from marginalized groups (women, people with disabilities, indigenous and migrant people etc.), especially for AI policy makers and AI creators.
- Promoting develop mechanisms that do not only require consent capture from the communities that will be affected by it and involve them throughout, but also that allow them to opt out of the AI system without being disadvantaged.
- Strict regulations and oversight to ensure the responsible development and deployment of AI technologies, ensuring the respect of human rights, accountability, accessibility, redress and transparency to the public on high risk systems in areas such as recruitment software, digital infrastructure for public goods/services, policing and judicial processes, among others. (Following the EU AI Act's measures for high-risk systems, however, without exceptions to open-source models).
- Clear guidance on the procurement of AI systems used in government agencies since the communities most reliant on government services are usually the communities most at risk of being discriminated against by AI systems.
- Global AI governance frameworks are required for many reasons, including to ensure that Global Majority nations have an active role in creating the future of AI

    - Resource outlining how the Global South can harness AI regulation to advance innovation
- Prepare guidelines for developers and deployers of AI on how they need to structure diverse teams within their organisations. Importantly, this includes that representatives of affected communities are team members, i.e. involved and empowered all the way through the ideation, design and development process (not just "consulted").
    - Each team should have a designated role to support accountability, e.g. moderating training datasets, red teaming to mitigate biases, and tweak AI system data to align with a state or region's particular context before deployment of an AI system takes place (WEF, 2022)
- Develop principles and guidelines for the sharing of AI best practices, aspirations and improvement that can constitute a bridge between Global North and Global South narratives concerning AI adoption and implementation, using conferences, symposia, public awareness and education as drivers for inclusive AI in achieving a decreased AI-divide between the two global positions (WEF, 2023). Special focus given on inclusion of broader communities through advocacy, outreach, public consultations and virtual mode of communications.
- Establish a definition and scope for a "right to algorithmic transparency" as a digital right and "algorithmic transparency" as a digital jurisprudential value for ensuring inclusion, fairness and creating accountability among developers of AI systems where algorithmic bias may reveal itself in AI systems (Oxford Insights, 2022)

Gender Equality

Women and girls are underrepresented in the AI workforce and in data (UN Women, 2024). Further, in many low- and middle-income countries, there's still a clear gap in mobile internet use between women and men. About 66% of women use mobile internet, compared to 78% of men. This difference might not appear detrimental, however, it implies that around 785 million women are still not connected (GSMA, 2024). This gap is even wider in places like South Asia and Sub-Saharan Africa, where access remains a serious challenge. Additionally, women and girls still face unequal access to knowledge and resources and face barriers to thriving in the AI field (and generally in the STEM sector).

Who and how AI is developed, and which data is used (or not) has gender implications. It can perpetuate and amplify harmful gender biases and stereotypes as well as increase gender disparities. It is essential to rebalance the genders that are creating AI – and thus the future of technology; a start are the following goals:

- Introducing and mainstreaming gender and intersectional perspectives in AI development and deployment as well as executive boardrooms, positioning these considerations as a key to the achievement of the 2030 agenda.
- Increase efforts to foster inclusive and equitable opportunities for girls and women in STEM:
    - Tackling gender stereotypes, negative gender norms and inequalities in education and workplaces.
    - Equal representations and access to capability building initiatives for providing a level playing field and ensuring that women can be 'employment ready', following approaches such as the 30 steps in Chile's Gender Equality Policy for Science, Technology, Knowledge and Innovation until 2030 (Gob.cl, 2021)
- Gender based financing and Temporary Special Measures: increase targeted public and private investment to close the gender digital divide, foster inclusive innovation ecosystems, and support women's participation and retention in the digital economy.
- Envision an independent commitment to a solitary goal of gender equality based on the following three elements as elucidated in the Global Digital Compact 2024:
    - Freedom from technology-facilitated gender-based violence and discrimination - By creating awareness, enhancing AI literacy and developing effective redressal mechanisms.
    - Equitable educational and economic opportunities - By creating strong internal and external controls for high risk systems to mitigate gender bias risks.
    - Equal voice, leadership and participation - By targeting 50/50 representation across all levels and developing supporting policies to address challenges like non-paid labour, workforce safety and benefits to increase women participation

Planet

Green AI

The training of AI models requires a lot of compute power and electricity, enhancing the climate crisis. Further, in developing countries, the prioritization of AI projects in the distribution of electricity (vs for households) could put an additional burden on fragile infrastructure, resulting in power cuts and increased electricity prices. Not surprisingly, the

benefits of AI are often unequally distributed between high- and middle- or low-income countries, thus further increase global inequality (Freire, 2025)

At the same time, AI serves as a key enabler in the energy sector by supporting the delivery of energy services to populations, fostering low-carbon systems through circular economies and resource-efficient smart cities, and enhancing the integration of variable renewables with smart grid technology. Thus, AI helps reduce fossil fuel consumption and supports the development of sustainable electricity production from renewable sources like solar, hydroelectric, and wind energy (Shang et al., 2024). This leads to increased energy efficiency and reduced carbon dioxide emissions (Shang et al., 2024).

Examples for concrete goals are:

- Ensure access to affordable, reliable, sustainable and modern energy for all by:
    - 1) Establishing standards for promoting energy-efficient algorithms and hardware development
    - 2) Tracking and reporting energy efficiency throughout the AI supply chain.
    - 3) Open Exchanges for Carbon Trading
    - 4) Real time data availability for energy usage and consumption; deployment of AI systems for smart grid integrations
        - Here, affected communities should have decision agency, i.e. they are not only informed about resource usage, but also have a mechanism to hold companies accountable (complain, legal action, redress, reparation, ...) for disproportionate use

AI for Environmental Solutions

Reliable water sources remain a precarious and concerning prospect for myriad communities and we believe that AI can play a crucial role here: by utilizing predictive analytics and harnessing satellite imagery, AI can monitor natural water sources to better conserve WASH resources as well as identify leaks in municipal water infrastructure. However, limited access to technology and infrastructure makes it difficult for impoverished regions to benefit from AI-driven WASH solutions, widening the digital divide. To address this, the development community must introduce existing technologies or invest in locally developed AI solutions in less technologically advanced nations (the latter are likely to have a better fit with the needs of affected communities and local contexts). To achieve this, we recommend (based on the Universal DPI Safeguards Framework):

- The UN should take a lead in introducing open source technologies/models that local governments and civil society groups can deploy to monitor and improve the climate

and environmental conditions (especially useful for countries that will house a large number of data centres)
    - This includes alternative approaches aimed at greater digital sovereignty (including data sovereignty), decentralization, and equitable value distribution. These include sovereign 'open' cloud initiatives (such as Gaia-X in Europe, India's MeghRaj, and Africa's AfriCloud) which aim to establish regionally controlled infrastructure that reduces reliance on external cloud providers. Open-source cloud platforms enable organizations to develop and control their own cloud environments, avoiding vendor lock-in, while cooperative and community-owned services can provide an alternative to corporate-owned infrastructure
- Enabling the global tracking and reporting on water systems - including water quality.
- Open Water Exchanges
- Scalable Smarter Cities Initiatives - Open networks for data from traffic and mobility networks, air quality, disaster management etc.
- Setting up monitoring centres, observatories and environment situation rooms like those envisioned by the United Nations Environment Programme (UNEP, 2022), especially to ensure environmental protection and environmental sustainability where data centres and AI factories are built and used.
- Integrating AI into the evaluation of Environmental Impact Assessments (EIA), with human oversight, and factoring the impact of AI on environments as part of EIAs.

Partnership

Multi-Stakeholder Collaboration

Participatory AI development is key to responsible AI: only if communities impacted by a system are empowered to shape it (before, during & after deployment), can we ensure that these systems are in service of the people whose lives they affect (Kallina, 2024) – the goal of the first UNECE PPP model in 2015 (meeting the needs of People-first and delivering value for people). Importantly, it is associated with (i) rebalancing the decision-power around which and how systems are built, (ii) a better understanding of a systems socio-technical context (increasing a system's fit with its environment), and (iii) a better detection of risks and harms before they occur. For concrete, actionable guidance on how

such participatory development could be implemented, see the AI & Equality Framework. Thus, we recommend participatory AI development for all AI systems.

Participatory development is even more pressing in developing countries as external (especially if global) AI suppliers almost always lack context and understanding of the local culture, needs, and application contexts (Lin et al., 2024 - very valuable resource!). Importantly, such involvement has to start at the objective setting stage and actually empower involved communities to change key aspects of the system. Otherwise the risk that resulting technology does not actually or appropriately address the community's most pressing needs is significant, i.e. the system cannot be successfully deployed: Lin et al. (2024) found that "only two out of 14 projects we studied reached the deployment stage". To harness these benefits, we recommend following goals:

- Establish accountability for AI creators and deployers by demanding a report that outlines (i) which affected (and especially at-risk) communities have been involved, (ii) which insights have been gathered, as well as (iii) which actions resulted from these (example template, see Barker et al., 2023).
    - Such reports should make a difference when assessing AI harms: if AI creators took precautionary actions by involving communities (especially at-risk groups!), harms created by the final system should be penalized differently from AI harms resulting from systems by creators that did not use participatory development to minimize harms, i.e. did not involve affected at-risk communities.
    - It is important that reactions to community involvements are clearly justified, especially if no reaction took place
    - Demanding a stakeholder mapping report (see e.g. our AI & Equality Framework or HUDEIRA page 19) to enforce a clear consideration of affected communities, including and especially at-risk groups.
    - Subsequent report on how those were involved.
- AI policy recommendations, frameworks, or declarations (such as this one) should be up for public consultations for a minimum of 90 days, including awareness and advocacy efforts to ensure maximum participation and better representation.
    - Extra efforts to include voices of marginalized groups or technically challenged communities (alternate methods say, mobile voting, radio promotions etc.)
    - Following this:
        - Bodies, associations and individuals consulted to be made publicly available on the website.
        - A clear summary of received input and resulting actions / amendments

- Enforce Human Rights Impact Assessments throughout the lifecycle of an AI system, see HUDEIRA for a concrete assessment (following the EU AI Act)
    - Defining accountabilities of all actors including developer, deployer, end user to ensure impact group is included and there is no violation of fundamental rights.
    - Investing XX budget for capacity building initiatives aligned to AI transformation agenda.

Open Access and mutual learning

AI and related technologies have a significant potential for causing disruption at a large scale, hence providing the necessary resources for open innovation not only accelerates innovation and minimizes risks but also adds a layer of transparency and accountability enhancing the overall governance frameworks. AI is at the risk of perpetuating existing biases and widening the digital divide if power is confined to a few players dominating the space, driving the innovation and also its purpose. It is therefore essential to base AI innovation on open source and wider transparency and interconnectedness principles and engage wider participation so that technological advancements can be used to move the needle on SDGs and reduce the disparities. It is only through mutual learning and transparent mechanisms we can build, deploy and scale beneficial AI systems using the collective wisdom, especially from the underrepresented groups and developing economies. Thus, we recommend:

- Promote Multilateral Cooperation and best practice sharing through Global Summits and Coalitions - Further refinement through Research and Academic Inputs along with independent expertise (as needed)
- International cooperation for developing and implementing AI standards, based on open principles. Resources:
    - What are standards? - AI Standards Hub
    - AI Standards Search - AI Standards Hub
    - Universal DPI Safeguards Framework
- Review the regulatory frameworks in light of emerging technologies like AI, Digital twins, IoT and Blockchain and ensure public consultations for inclusive, transparent and fair governance and redress mechanisms
- Intellectual Property Protection: Aligning existing indices related to digital innovation to human development.
    - Example, the Global Innovation Index (GII) led by the World Intellectual Property Organization (WIPO) measures innovation across pillars like science and innovation investment, technology adoption, and other areas. A high

score on this index but a lower score on the Responsible AI Index would highlight misalignment between ethics, governance, and innovation. (source)

- Technology/ Innovation grants: Local development of models for fair representations, added context and reducing linguistic barriers.
- Public Private Partnerships: Development and availability of open source models that decentralize the access to technology and provide training to ethically use the same.
    - Examples:
        - IndiaAIKosh - Platform for enabling MSMEs/ startups to access compute power (economically)
        - Competency framework for AI integration in India
- **Strengthening Institutional Research Leadership in Responsible AI**: Promote institutional leadership in AI ethics and governance by creating dedicated research centers and fostering collaborations to drive progress in responsible AI development. These centers should focus on the following:
    - **Algorithmic Transparency and Explainability:** Develop frameworks and methodologies to enhance the transparency of AI models and improve explainability across diverse applications.
    - **Bias Detection and Mitigation:** Advance research into the identification, detection, and elimination of biases (including binary discrimination) within AI systems, ensuring fair outcomes for all user demographics.
    - **Cybersecurity Standards:** Establish cybersecurity frameworks and standards specifically tailored to AI systems, safeguarding them from manipulation, vulnerabilities, and malicious attacks.
    - **Human Rights Impact Assessment:** Develop tools and methodologies for assessing the potential impact of AI systems on human rights, ensuring that systems are aligned with international human rights standards, also including ethical impact assessments.
    - **Local Context Adaptation Guidelines:** Promote the creation of guidelines for adapting AI systems to local cultural, social, and legal contexts, ensuring that solutions are relevant, equitable, and context-sensitive.

AI & EQUALITY Authors

All authors are part of the AI & Equality Community (link to platform and website).

Emma Kallina, 4th year PhD in Stakeholder Involvement for Responsible AI Development in Practice at the University of Cambridge, UK

Geetika Bhatnagar Marketing and Communications Professional with expertise in Emerging Technologies and Digital Transformation Initiatives. Also,an Advocate for DEI, TechforGood, Sustainability, and Climate Action.

Gabriela Del Barco - Renard, Consultant and facilitator, working at the intersection of DEI and Equitable and Responsible AI. Based in Paris, France.

Ozge Caglar - Independent consultant and psychologist working in GBV, MHPSS, project evaluations, and AI integration into humanitarian action, based in Türkiye.

Chandrashekar Konda - A seasoned technology leader experienced in architecting and implementing Enterprise scale Gen AI & predictive AI models/solutions.

Cinthya Vergara - student PhD in Data Science at Universidad Adolfo Ibáñez, Chile AI/ML Project Management Specialist at Research Group for Trustworthy Systems (R3), supported by the ANID Fondef IT24I0144.

Ann Borda - Dr Ann Borda is an Ethics Fellow at The Alan Turing Institute (UK). She is involved in the development of public sector guidance on AI ethics, safety and governance, including a Human Rights Impact Assessment for public sector agencies.

Warren Bowles - AI researcher, writer, AI ethicist and legal academic. CAIDP Research Group Member. EDSAFE AI Alliance Catalyst Fellow. Based in South Africa.

Resources

Sustainable Development Report, Rankings The overall performance of all 193 UN Member States, https://dashboards.sdgindex.org/rankings

James Purtill, 2016, Students accused of welfare fraud say Centrelink's sums are wrong, https://www.abc.net.au/triplej/programs/hack/students-accused-of-welfare-fraud-say-centrelink-is-wrong/8127550

Marissa Gerchick et al., 2023, How Policy Hidden in an Algorithm is Threatening Families in This Pennsylvania County, https://www.aclu.org/news/womens-rights/how-policy-hidden-in-an-algorithm-is-threatening-families-in-this-pennsylvania-county

Robert Booth, 2024, DWP algorithm wrongly flags 200,000 people for possible fraud and error, https://www.theguardian.com/society/article/2024/jun/23/dwp-algorithm-wrongly-flags-200000-people-possible-fraud-error

Lilah Burke, 2020, The Death and Life of an Admissions Algorithm, https://www.insidehighered.com/admissions/article/2020/12/14/u-texas-will-stop-using-controversial-algorithm-evaluate-phd

Jayant Narayan, 2025, Human Rights and AI: A Global Index on responsible AI, https://giraianalyticalreportseries.tiiny.site/

Hongjin Lin, Naveena Karusala, Chinasa T. Okolo, Catherine D'Ignazio, and Krzysztof Z. Gajos. 2024. "Come to us first": Centering Community Organizations in Artificial Intelligence for Social Good Partnerships. https://doi.org/10.1145/3687009

How Leaders in the Global South Can Devise AI Regulation That Enables Innovation, 2025, https://institute.global/insights/tech-and-digitalisation/how-leaders-in-the-global-south-can-devise-ai-regulation-that-enables-innovation

The Universal Digital Public Infrastructure Safeguards Framework, 2024, https://dpi-safeguards-framework.org/frameworkpdf?_gl=1*q66mp7*_ga*MTI1MjQ0NjIyMi4xNzQwMzkyODc0*_ga_N2QB2GX10S*MTc0MDk0MDU5OC44LjEuMTc0MDk0MTM3OS4wLjAuMA..

What would a human rights-based approach to AI governance look like?, 2023, https://www.gp-digital.org/what-would-a-human-rights-based-approach-to-ai-governance-look-like/

Artificial Intelligence and gender equality, 2024, https://www.unwomen.org/en/articles/explainer/artificial-intelligence-and-gender-equality

Public policies for gender equality in science, technology, engineering and mathematics (STEM), 2023, https://repositorio.cepal.org/server/api/core/bitstreams/6b7dd3c2-8639-46c9-b7db-2cab26fab283/content

Government launches Gender Equality Policy for Science, Technology, Knowledge and Innovation, as well as the Regional Alliance for the Digitalization of Women in Latin America, 2021, https://www.gob.cl/en/news/government-launches-gender-equality-policy-science-technology-knowledge-and-innovation-well-regional-alliance-digitalization-women-latin-america/

Accelerating The SDGs Through Digital Public Infrastructure: A Compendium of The Potential of Digital Public Infrastructure | United Nations Development Programme, https://www.undp.org/publications/accelerating-sdgs-through-digital-public-infrastructure-compendium-potential-digital-public-infrastructure

The DPI Approach: A Playbook | United Nations Development Programme, https://www.undp.org/publications/accelerating-sdgs-through-digital-public-infrastructure-compendium-potential-digital-public-infrastructure

AI Standards Hub, https://aistandardshub.org/resource/main-training-page-example/1-what-are-standards/

Freire, C. 2025. Is this time different? Impact of AI in output, employment and inequality across low, middle and high-income countries. https://doi.org/10.1016/j.strueco.2024.12.016

Brandie Nonnecke and Philip Dawson. 10/21/2021. “Human Rights Implications of Algorithmic Impact Assessments: Priority Considerations to Guide Effective Development and Use.” https://www.hks.harvard.edu/centers/carr/publications/human-rights-implications-algorithmic-impact-assessments-priority

Human rights impact assessments : a review of the literature, differences with other forms of assessments and relevance for development (English). http://documents.worldbank.org/curated/en/834611524474505865