Response to the UN AI Advisory Body's Interim Report on Governing AI for Humanity

IT for Change

March 2024

IT for Change's Response to the UN AI Advisory Body's Interim Report on Governing AI for Humanity[1]

We welcome the UN Secretary-General's AI Advisory Body's Interim Report on Governing AI for Humanity. The report's proposal to strengthen and enhance the international governance of artificial intelligence (AI) is an important starting point towards bridging the critical global-level digital policy gap that widens with every passing day. As part of the online consultation process, IT for Change wishes to engage with the review of this important document and submit its inputs, with a focus on opportunities and enablers, risks and challenges, and guiding principles (GP) to form global governance institutions for AI.

1. Opportunities and Enablers

The AI innovation path we forge should be a carefully crafted alignment with the assemblage of technological trajectories, market mechanisms, and governance norms. The report notes the potential of AI to "transform access to knowledge and increase efficiency around the world," and stresses on the need for enablers to ensure equitable sharing and distribution of benefits. While we concur with the report's argument to balance growing successful AI ecosystems with establishing and maintaining an effective accountability threshold, we argue for a skew towards the latter.

AI 'enablement' as a starting point is misplaced when we consider how the Generative AI (GenAI) boom mimics historical patterns of the Web 2.0 phenomenon—a sudden and meteoric rise of a new technological paradigm in the absence of regulatory framework, driven by reckless pro-profit motivations. The resultant concentration of tech resources and the significant distortion of the economy and public sphere have had profound implications for questions of fairness, accountability, and equitable transfer of benefits. These are by now well understood trends and offer cornerstone historical lessons in how we must approach AI.

AI is not merely a product, but rather, a paradigm that co-constitutes contemporary social, economic, and political relations:

- As a means of production, its economic governance is deeply implicated in trade and intellectual property (IP), and consequently, the competitiveness and integrity of market structures.

---

[1] This response was authored by Anita Gurumurthy, Deepti Barthur, Eshani Vaidya, Shreeja Sen, Jai Vipra, and Merrin Muhammed Ashraf from IT for Change. If you wish to share any feedback regarding this response, please write to us at itfc@itforchange.net.

- As the ordering principle of communicative rationality, AI is a determinant of the diversity and vitality of our public sphere and the arbiter of our communicative rights.

- Within the emerging architectures of democracy, AI tools play a large part in the architecture of civic spaces and infrastructures.

AI systems must, therefore, be seen as a site for the operation of principles of democratic integrity and distributive justice. To this effect, the enablement of AI must be seen as a fundamental question of furthering justice, above all else.

As a prerequisite to AI enablement, we must fully understand and deliberate the threats that uncritical adoption of AI poses to justice, including to standards of public reason and justification, protection and enhancement of citizen rights, and contribution to fair outcomes and social value, focusing on the most marginalized and vulnerable people.

The heightened faith in the seemingly boundless potential of AI to deliver on a range of solutions has propelled an innovation-first approach that obscures the structural causes of the world's development gaps. We must peer under the hood to interrogate the false solutionism and spurious myth-making that has accompanied inscrutable AI models peddled as sound data science. Outcome distortions in affect-recognition technology in areas like hiring, criminal justice, and more are responsible for direct and indirect harm to individuals and communities. Evidence of the degradation caused by large language models (LLMs) and deepfakes to our global information environment and trust architectures, also come to mind, not to mention the rise of financial speculation around AI innovation.

2. Risks and Challenges

As AI reshapes key institutions and their practices, our conception of risk needs to be broadened significantly. The report rightly highlights that AI-related harms are not experienced in isolation, but are interconnected with structural issues and must be assessed proactively. An understanding of risks and harms must therefore adequately reflect upon AI's potential for actual harm to people, society, and habitat.

While the report provides an early categorization of AI-based risks from the perspective of vulnerable communities and the commons, the language is vague and leaves room for interpretation. We suggest that the final draft of the report map risks and challenges in a systematic, nested, and comprehensive manner. An explicit rights-centric language must form the basis of the proposed risk assessment framework. Within this, rights should be considered non-negotiable, i.e., they must be maintained regardless of the risk associated with external factors.

Derisking is not only about addressing the technical aspects endogenous to the AI system. It is about enabling a continuous recalibration of the inner workings of the system for alignment with the norms and principles adequate to a just and equal AI society. The *a priori* role of shared principles is therefore of paramount importance to the challenge of future global coexistence —among peoples and nations and with technology. Derisking must include rejecting AI use cases that are inimical to society.

Risks and challenges that pertain to economic justice are crucial. The potential concentration/ control and ownership of AI resources in the hands of a few countries and their corporations is one of the biggest risks to a just and equitable AI society. The greed for geoeconomic and geopolitical control has seen a data inequity that inhibits the Majority World's right to innovate; appropriation of the data commons for proprietary AI, delegitimization and lack of room for alternative AI models, exploitation of workers– including peasants and indigenous people through corporate AI value chains,– and a shocking neglect of public finance to build digital public goods including compute power, in developing countries.

Addressing the risks and challenges in/of the AI paradigm also implies putting the right to participate in AI related policy and oversight in the hands of the people. Without this, the democratic deficits in the socialization of AI cannot be addressed. This assumes significance as AI is becoming a core element in all socio-economic and governance infrastructure— from health to education, welfare, urban utilities, and more.

The report highlights that through global cooperation and coordination, it is possible to ensure that risks of AI are addressed and that uncertainty around assigning responsibility should not lead to governance paralysis. Such pragmatism must find its logical path in legally-binding accountability and liability mechanisms to ensure that there is no evasion of responsibility for AI harms by state or private actors.

3. Guiding Principles (GP) to Form Global Governance Institutions for AI

Multiple values abound within the multi-faceted AI ecosystem (varied uses, stakeholders, modes of deployment, etc.). Rather than viewing them as trade-offs, governance mechanisms must allow varied values to coexist within a larger set of principles that further democratic and distributive integrity. The maximization of public interest and the inviolable bottom line of human rights must not be viewed as binaries.

Guiding Principle 1 speaks of inclusive governance for the benefit of all, but the substance in the report speaks only of access to AI. Access to AI products from large technology companies cannot bring empowering inclusion. Equity and self-determination in AI capabilities–compute power, talent and

data prowess– should be the larger aim. It would be useful and important to consider the principle of common but differentiated responsibilities (borrowing from the climate debate) to make the AI resource paradigm inclusive and equitable. This would mean new international regimes to correct for decades of data extractivism and digital colonization, and redistribution of data/data value—including legal obligations for dominant countries in data sharing and relaxation of IP rules in data/AI, along with Official Development Asisstance (ODA)/international public finance for public digital infrastructure development in all countries.

In their current iteration, the guiding principles do not emphasize a rights-based approach to AI design, development, and deployment, except for a reference in the fifth principle. The report must also build towards a framework that addresses historical and contextual injustices. This would involve a cross-cutting/cross-sectoral effort to redefine AI-related rights regimes in areas such as social communications, food sovereignty, health, environment, gender equality, welfare delivery, work/employment, etc.

A specific focus on the right to self-determination is necessary under Guiding Principle 2. A critical element of this is that people are empowered to have a meaningful say in the design, deployment, and purpose of AI technologies, including to reject certain use cases. The report must expressly address questions on the mode of empowerment, specifically with regard to the specific right to access and modify data.

The report highlights the shortcomings of self-regulation and mentions the need for "binding norms enforced by member states". This needs to expressly provide for a public accountability framework legitimated by law and not left to enforcement through private contractual obligations. The report, in its final iteration, must adopt a nuanced approach to determine which duties lie with whom and at what point in time, including the duty of deployers and providers of technology.

Guiding Principle 3 must address the role played by the egregious use of intellectual property regimes and the failures of antitrust regulation in countering monopolies, which have enabled data extractivism and resource and capital evacuation in the Global South. Notably, without a global data governance framework, this guiding principle becomes a mere red herring.

4. Institutional functions in an international governance regime

At the outset, we would like to request further clarification on the operationalizing of principles by individual institutions or a network of institutions. The report suggests that governance functions will be distributed on the basis of "institutional hardness" but does not define this concept adequately. Accountability processes need to be grounded in clear objectives and principles. Ambiguity in the

distribution of governance functions will only create further loopholes in implementation.

The report refers to a multiplicity of agreements, such as the Universal Declaration of Human Rights (UDHR), Global Digital Compact (GDC) Common Agenda, and intergovernmental bodies, such as the Intergovernmental Panel on Climate Change (IPCC), Financial Accountability Task Force (FATF), European Council for Nuclear Research (CERN), among others—all of which approach AI governance with a unique perspective, which may create competing visions about the purpose and nature of AI governance. A uniform approach must be agreed upon to operationalize the guiding principles through a robust implementation framework—addressing misrecognition and maldistribution in the AI paradigm, and the right to self-determination of states, communities, and individuals to pursue their pathways to development, peace, and human flourishing.

With regard to demarcating roles and responsibilities, we would like to emphasize the need to enforce transparency obligations on private entities. For instance, Article 11 of the European Union AI Act provides for "technical documentation of a high-risk AI system" and requires that the specifications of this documentation (Annex. IV) provide for a general (intended purpose, details of provider and deployer, etc.,) and specific description (manner and mode of use of training data) of an AI system with details on monitoring, functioning and control; risk management systems, etc. The voices of Global South actors must feed into the process for developing these benchmarks and obligations so that transparency can serve the interests of the weakest and most vulnerable actors in the AI paradigm.

While the report discusses the low rates of participation of the Global South in international conferences of AI and highlights the need for capacity building in the Global South, stronger language on the structural injustices arising through lack of representation, equal participation, and inclusion of diverse voices must be adopted in the final iteration of this report.

Other Comments on International Governance

The UN General Assembly has adopted a resolution led by the US on safe, secure, and trustworthy AI systems, which is significant and of note for this report as well. However, this resolution is limited in its scope and does not address the paradigm shift that AI has brought. In fact, while it recognizes the importance of respecting human rights, it also places intellectual property rights on the same footing, and sidesteps a much warranted discussion on AI transparency. It is an opportunity for the final report to address these issues to ensure that international governance on AI is fair, accountable, and just.

5. Other Comments

There is currently a conceptual failure in discourse, whereby it is impossible for the vast majority to understand and interpret the impact and consequences of AI meaningfully. Dominant powers frame AI entirely from a technological point of view, in terms of its design and development capabilities. But the final report and the advisory body can play an important role in shifting the frame and mainstreaming the idea of AI as a larger societal process.

It should thus push back against the assumption that deep technical expertize of AI is a prerequisite to engage with its societal impacts. The lack of full understanding should not be a condition for citizen-interest to be upheld. Nor should an undue knowledge burden be shifted to the weakest actors in the ecosystem for states to act in their favor. The final report must outline mechanisms for capacity building around AI that allow for engaged and democratized debate on what AI can do, what it's best suited for, and how it can be a tool for self-determination, rather than loss of agency.

Lastly, questions and considerations of ecological impact have been largely missing from the AI governance conversation, even as its massive carbon footprint looms over the world. The report briefly mentions ecological harms in risk mapping but should probe this further.