Data Considerations for the UN AI Advisory Body's Report: Governing AI for Humanity

IT for Change

May 2024

Data Considerations for the UN AI Advisory Body's Report:

Governing AI for Humanity

IT for Change's Anita Gurumurthy and Merrin Muhammed Ashraf's Submission

As the UN AI Advisory Body's (UN AIAB) interim report acknowledges, data is critical for many major AI systems. It is the lifeblood of innovation in today's times. However, a corporate-led digital economy has seen the vital resource of data locked up for private profit. The *laissez-faire* regime of data extractivism has allowed transnational platform companies free rein to enclose the data collected from developing countries and entrench themselves as the harbingers of digital innovation and development in the Global South. Thus, the primary deterrent for developing countries on their path to structural transformation, through AI, is the governance deficit in global rules on cross-border flows of data.

The global governance framework must straddle the imperative for accessibility of data and nurturance of data commons as humanity's collective heritage, and the need to ensure transparency and explainability of systems built on such data. To realize this, **the data sharing regime should be differentiated– which requires states, private corporations, and other actors to access and share data in specific ways.** This can ensure there is no enclosure of public data/ insights from public data, and private data is available for states to perform their public functions in the public interest. An example of this is the European Union (EU) Data Act which enables the sharing of both inferred and derived data, as well as aggregated datasets from multiple users to ensure that such data sharing provisions are successful.

To reclaim the non-rivalrous nature of aggregate data resources and encourage their availability for the creation of public and social value, it is often suggested that a 'global public goods' approach or a data commons approach is necessary and well-suited. Even the recent zero draft of the Global Digital Compact calls for open data and data commons.

While this is significant, placing a resource under a more open, rather than standard private property, regime without corresponding institutional processes for regulating the terms of data access and use will not make valuable data resources available for public benefit. **Controls on appropriate access to global**

data public are needed so that conditions are instituted to prevent free-riding and consolidation of intellectual monopolies at the root of inequalities in the digital economy.

Also necessary are guarantees to recognize the sovereignty of communities from whom data is aggregated and mechanisms for equitable benefit-sharing (monetary and non-monetary) from data processing with such communities, akin to the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity.

It is also important to note that access to AI products from large technology companies cannot bring empowering inclusion. Equity and self-determination in AI capabilities—computing 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 redistribute data/data value—including, legal obligations for dominant countries in data sharing and relaxation of IP rules in data/AI, along with ODA/international public finance for public digital infrastructure development in all countries.**

Another important issue is monopolies built on the misuse of Intellectual Property protection. Trade secrets in the digital economy are increasingly becoming the go-to system of protection since data is not patentable and has limited copyright coverage. The consequence is that large amounts of aggregate data collected by Big Tech from transaction activity of users and public datasets also become proprietized. This results in the wasteful underuse of data resources and adversely affects the realization of SDGs. Further, trade secrets in data also render AI systems built on such data unexplainable, which can result in the entrenchment of misrecognition harms (identity-based exclusion and harm) in the data economy. To address these concerns, a reform of the IP regime that includes the following elements is necessary:

- strong institutional safeguards to protect social sector data sets, especially where there is a risk of proprietization of core development functions through AI models (such as in health, education, and welfare);
- conditional access to the public domain and open government data, with the inclusion of

purpose limitations and clear sunset clauses on use;

- fair use limitations on how models learn from and use training data to specifically prevent profiteering through reuse, including through strict stipulations against free-riding and the development of substitutive value propositions;
- reciprocity guarantees in common data pools, where private model developers who build on public data layers must share back and enrich the commons

IPR in general, and trade secrets protections, in particular, must not be used to restrict data accessibility and transparency of AI systems. As such, the IP regime should be limited to existing types, like patents, copyrights, etc, and must not be expanded judicially. National courts and parliaments are crucial authorities who can ensure overexpansion of trade secrets in non-personal or aggregate data can be stymied, as well as balance the rights of non-owners of data. Such restriction and balancing through access to aggregate data can be on the grounds of freedom of information, freedom of scientific research, and freedom of free movement of data to enable the provision of welfare services.