Task Force 05
INCLUSIVE DIGITAL TRANSFORMATION

GOVERNING GLOBAL EXISTENTIAL AI RISKS: LESSONS FROM THE INTERNATIONAL ATOMIC ENERGY AGENCY

Shamira Ahmed, Executive Director, Data Economy Policy Hub (DepHUB) (South Africa)

Dr. Peter G. Kirchschlaeger, Professor, University of Lucerne and ETH Zurich (Switzerland)

TF05

Abstract

The rapid advancement of data-based systems (DS) offers unparalleled opportunities for societal progress but also poses significant existential risks. Evidence suggests that the potential dangers of DS surpass those of nuclear weapons, yet regulatory oversight remains insufficient.

The lessons learned from the field of nuclear technology (NT) and the establishment of the International Atomic Energy Agency (IAEA) offer valuable insights into how global cooperation and regulatory mechanisms can effectively manage complex technological risks, while at the same time fostering equitable technological progress.

Cognisant of disparities between DS and NT, this brief proposes the establishment of an International Data-Based Systems Agency (IDA) within the United Nations (UN) system in partnership with the Group of Twenty (G20). By establishing IDA at the UN as a central regulatory authority, the international community can proactively address existential AI risks while harnessing the transformative potential of DS to advance the sustainable development goals (SDGs).

Keywords: AI, International Data-Based Systems Agency (IDA), Existential Risks, Human Rights, Multilateralism.

Potential Global Existential Risks from Data-Based Systems (DS)

There are rising concerns over potential global existential risks (Hendrycks et al., 2023) from artificial intelligence (AI) systems (OECD, 2024), more accurately described as data-based systems (DS) (EU, 2024). Consequently, effective multilateralism is urgently needed to facilitate coordinated, comprehensive, collaborative, and intergenerational promotion of the advantages and challenges of DS (Ahmed et. al., 2023, UNCTAD, 2021).

The proliferation of unilateral action has resulted in a patchwork of national and regional mechanisms and rules that exacerbate regulatory fragmentation in the global digital landscape (Veale et al., 2023; Bradford, 2023), which not only threatens business opportunities and innovative ventures, but will most likely adversely impact low-and middle-income countries (LMICs), with potential intergenerational effects (UNCTAD, 2021). A mission-oriented approach is needed to navigate the geopolitics of responsible AI governance (Ahmed, 2024).

The Need for Global Governance for DS

Data serves as the foundation upon which many frontier technologies are built (EU,2024). Effective management, analysis, and utilisation of data are essential for training models, making predictions, and optimising performance that ultimately drive innovation in various fields. Robust data governance, enusres the socioeconomic benefits of data are not concentrated to benefit a few, as the reliability, accuracy, and ethical use of DS in different data ecosystems can be protected (Ahmed, 2023).

Furthermore, robust data governance is essential to maintain data integrity and prevent security breaches for other data-driven frontier technologies (FT) beyond AI such quantum computing (WEF, 2024), and extended reality (XR) technologies (Egliston et al.,2024) and for maintaining user privacy and data securityin the digital age (CIGI, 2024).

Overall, effective governance of DS is crucial to address the diverse challenges and opportunities presented by FT across various domains (Chon & Alexander, 2023) including to support ethical digital transformation (Kirchschlaeger, 2021).

While there have been regional initiatives to govern DS (mostly from the global North), there is an urgent need to ensure the current status quo (Arora et al.,2023) does perpetuate future governance of DS at the global level, this urgency was reflected in March 21, 2024, when the United Nations General Assembly (UNGA) unanimously adopted the resolution "Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development" which emphasizes: "The same rights that people have offline must also be protected online, including throughout the life cycle of artificial intelligence systems." (UNGA, 2024). This is not the case yet.

Lessons from Existing Global Governance Institutions

The Intergovernmental Panel on Climate Change (IPCC) model consists of a panel of experts, if replicated in the DS domain would provide policymakers and governments with information, scenarios, and models for their decision-making. However, adopting the IPCC model for the global governance of DS would not effectively mitigate the existential extent of the challenges posed by DS, due to the lack of legal authority and legal enforcement-tools (CCEIA, 2023).

As an alternative, the International Civil Aviation Organization (ICAO) model consists of a binding global framework, including implementation initiatives through the development of policies and standards, compliance audits, studies, industry analyses, assistance to states and stakeholders, and contributions to the global alignment of aviation regulations. However, the application of the ICAO model for global DS governance raises concerns due to the non-comparability of a self-contained industry like aviation to a cross-cutting technology like DS causing multifaceted legal and ethical issues (Baker McKenzie, 2023).

Given the urgency (Guterres, 2023; Türk, 2023), the following analysis and the areas of convergence between DS and nuclear technologies (NT)[1], the International Atomic Energy Agency (IAEA) can be an appropriate model for global governance of DS.

The emphasis on safety, transparency, and accountability in nuclear governance serves as a guiding principle for regulating DS in a manner that prioritises environmental stewardship and societal wellbeing. At the same time, there is a need to consider that DS – unlike NT involving tangible, physical assets – DS are not only highly dynamic but operate within the intangible realm of data, the borderless internet, and algorithms. These characteristics of DS, though, allow IDA to monitor DS and to intervene if necessary. Consequently, the oversight model employed by the IAEA at the UN provides a template for international cooperation and coordination in managing emerging technologies (CCEIA, 2023).

---

[1] See Appendix

Recommendations

The following recommendations are proposed for the G20:

i.   Establish the International Data-Based Systems Agency (IDA) at the UN

The G20 should support efforts towards the establishment of an International Data-Based Systems Agency (IDA) at the UN (IDA, 2024). The UN system's experience in promoting more equitable international cooperation and global governance provides valuable insights for responsible global DS governance. The G20 can leverage the UN as a partner to play a central role in convening key stakeholders and advising on progress to achieve the sustainable development goals (SDGs), through responsible governance of DS.

The IDA would also serve as a central authority responsible for access to market approval, setting and monitoring human rights-based global standards (UNHRCl, 2023), which can facilitate cooperation among nations and address emerging challenges in dynamic transnational DSecosystems .

IDA can also ensure that DS are governed in a manner that promotes fairness, transparency, and accountability across borders by emphasising the importance of human rights-centred regulatory frameworks, robust data governance mechanisms, and adaptive technical cooperation that align with the G20 and also integrate existing complementary initiatives to pool resources from multiple investors that range from regional banks, to philanthropists, and international donor assistance organisations.

ii. Establish Robust Global Data Governance Frameworks for DS

The global response to many of humanity's most significant challenges relies on the ability to collect, share, and analyse high-quality machine-readable (and real-time) data that transcend national borders, public and private domains – and increasingly navigate a complex dynamic system that includes non-traditional data sources and types. Data as a digital public good (DPG) should be subject to distinct governance and sharing arrangements that enable specific data to be shared and reused, including across borders and between various stakeholders – robust data governance frameworks are essential for ensuring the ethical and responsible use of DS. G20 member states can leverage IDA at the UN to establish clear policies for data collection, processing, and usage, as well as mechanisms for enforcing compliance and addressing violations.

Robust data governance, framed within the context of just data value creation (JDVC), places paramount importance on cooperation, equity, and inclusivity throughout the data value chain through acknowledging data as a critical factor of production alongside traditional resources in a way that ensures equity in transnational data ecosystems.

iii. Ensure Meaningful International Multistakeholder Collaboration

Addressing the complex transnational challenges DS poses requires the involvement of multiple stakeholders. A sociotechnical lens is crucial to support meaningful international multistakeholder collaboration for sharing best practices and building consensus on critical issues.

As a key stakeholder for the proposed IDA at the UN, the G20 should leverage IDA as a platform for dialogue, cooperation, collaboration, and coordination (CCC approach), among stakeholders, particularly on the following:

- **Foster Transparency and Inclusive Stakeholder Engagement:**

By providing regular updates on operations, performance, and outcomes in DS through various measures that include disclosing relevant information throughout the data value chain including implementing feedback mechanisms to capture insights from stakeholders, conducting post-implementation reviews to assess the effectiveness of interventions, and adjusting strategies and policies based on lessons learned. These continuous improvement mechanisms ensure that DS operations remain aligned with evolving ethical, regulatory, and sustainability requirements in a transparent manner. Transparency builds trust, accountability and fosters a culture of openness and collaboration among various stakeholders, existing transnational networks, and vulnerable communities.

- **Enhance Coordination of Technology Facilitation Mechanisms:** Contextually relevant technology facilitation mechanisms are crucial for maximising the benefits of DS and addressing gaps in global technical capacity differences. Suggested measures include increasing investments in technical cooperation programs, particularly in developing countries, to build DS development and deployment capacity. Additionally, efforts should be made to enhance South-South cooperation, leveraging the expertise and resources of emerging economies to support the adoption and utilisation of DS in regions with limited technical capabilities. The G20 can leverage IDA to generate adaptive technical assistance mechanisms focused on contextual realities and realise that countries have moved from developing basic AI knowledge to applying it in a broad range of fields to improve daily life and wellbeing in order to achieve national development priorities.

- **Define Clear Performance Metrics to Implement Regular Assessments**

The G20 can leverage the proposed IDAto define clear and measurable performance metrics to assess global DS operations' transparency, accountability,risk, and

sustainability. These metrics should encompass critical indicators such as data privacy compliance, ethical data usage, safety, environmental impact, evaluation of language models for low-resourced languages, and social responsibility, to name a few, which can also be used to identify potential threats and vulnerabilities associated with DS operations, within IDA.

- **Promote Technology Innovation for a Just Green-Digital Twin Transition:** NT governance encourages technological innovation to enhance safety, efficiency, and environmental sustainability in nuclear operations. Leveraging DS to optimise resource management, reduce carbon emissions, and support sustainable development initiatives should include measuring the environmental impact of AI. IDA can play a crucial role in facilitating access to diverse datasets relevant to green digital initiatives by promoting open data standards, data-sharing agreements, and interoperability protocols for use in public interest projects aimed at promoting environmental sustainability and social equity.

Explore Scenarios of Outcomes

The G20 can play a leading role in shaping transnational governance frameworks that promote the responsible and equitable use of DS to advance collective wellbeing of humanity and the environment, in the digital age. The IDA at the UN proposed in this brief is a novel and practical solution to the transversal challenges of the 21st century. Without interventions the following scenarios are expected:

i. **DS Governs Humans:** Without targeted action, the development and deployment of DS will perpetuate the current trajectory of significant ethical, legal, social, and policy problems, that are increasingly reported with the use of DS, such as growing multidimensional global inequalities and poverty, increasingly negative climate and environment impacts, constant human rights violations, economic and political manipulation, disinformation, and security-risks that range from threats to mental health of children and young people to the existential consequences for global peace due to DS-based cyber-attacks and dual use military applications of DS.

ii. **Humans Talk – DS Act Differently:** If we establish an institution following the model of IPCC or ICAO, we will continue to experience the discrepancy between "theory" and "practice", humanity and the planet will continue to suffer with the enormous ethical and legal problems the use of DS poses.

Should the G20 adopt the mission-oriented recommendations, the following scenario is expected:

i. **Humans Govern DS:** With the establishment of IDA at the UN, the G20 can ensure positive, constructive, and creative interaction by humans with DS supporting sustainable, human-centered, peaceful use of DS. In this scenario, humans govern DS rather than be governed by DS. By drawing upon the lessons learned from nuclear governance and adapting them to the digital age, we can develop effective strategies for responsible innovations that benefit humanity and the planet and not contribute to their detriment.

References

Ahmed, Shamira. "Advancing Data Justice in Africa through Just Data Value Creation", DepHUB, accessed March 28, 2024.

https://www.dataeconomypolicyhub.org/items/advancing-data-justice-in-africa-through-just-data-value-creation.

Ahmed, Shamira, Dio H.Tobing, and Mohammed Soliman. "Why the G20 Should Lead Multilateral Reform for Inclusive Responsible AI Governance for the Global South", T20 2023, accessed, March 28, 2024. https://t20ind.org/wp-content/uploads/2023/06/T20_PB_TF7_755_AI_Governance.pdf.

Ahmed, Shamira."Geopolitics of responsible AI governance: The need for a mission-oriented approach", DepHUB, accessed May 2024, 2024.

https://www.dataeconomypolicyhub.org/news/dephub-at-stg-talk-on-%22geopolitics-of-responsible-ai-governance%3A-the-need-for-a-mission-oriented-approach%22

Arora, Anmol, Michael Barrett, Elaine Lee, Eivor Oborn, and Karl Prince. "Risk and the Future of AI: Algorithmic Bias, Data Colonialism, and Marginalization." *Information and Organization* (2023).

Bradford, Anu. *Digital Empires: The Global Battle to Regulate Technology*. Oxford: Oxford University Press, 2023.

Carnegie Council for Ethics in International Affairs (CCEIA). "Envisioning Modalities for AI Governance: A Response from AIEI to the UN Tech Envoy." Artificial Intelligence & Equality Initiative, accessed March 28, 2024.

https://www.carnegiecouncil.org/media/article/envisioning-modalities-ai-governance-tech-envoy#gaio.

Centre for International Governance Innovation (CIGI). "Facing Reality: Canada Needs to Think about Extended Reality and AI", accessed March 28, 2024.

https://www.cigionline.org/publications/facing-reality-canada-needs-to-think-about-extended-reality-and-ai/.

Chon, Zee, and Alexander David. "Data Governance Frameworks for Ensuring Data Integrity in Clinical Trials Informatics." 2023. https://www.researchgate.net/publication/373214562_Data_Governance_Frameworks_for_Ensuring_Data_Integrity_in_Clinical_Trials_Informatics/citation/download.

Egliston, Ben, Marcus, Carter, and Kate E. Clark. "Who Will Govern the Metaverse? Examining Governance Initiatives for Extended Reality (XR) Technologies." *New Media & Society* 0, no. 0 (2024). https://doi.org/10.1177/14614448231226172.

European Union (EU). "Open Data and AI: A Symbiotic Relationship for Progress", accessed March 28, 2024. https://data.europa.eu/en/publications/datastories/open-data-and-ai-symbiotic-relationship-progress.

Guterres, António. "Secretary-General Urges Security Council to Ensure Transparency, Accountability, Oversight, in First Debate on Artificial Intelligence." 2023, accessed March 28, 2024. https://press.un.org/en/2023/sgsm21880.doc.htm.

Hendrycks, Dan, Mantas Mazeika, and Thomas Woodside. "An Overview of Catastrophic AI Risks." arXiv, 2023. http://arxiv.orsg/abs/2306.12001.

Kirchschlaeger, Peter. *Digital Transformation and Ethics: Ethical Considerations on the Robotization and Automation of Society and the Economy and the Use of Artificial Intelligence*. Baden-Baden: Nomos, 2021.

Organisation for Economic Co-operation and Development (OECD). "What is AI? Can You Make a Clear Distinction between AI and Non-AI Systems?" accessed March 24, 2024. https://oecd.ai/en/wonk/definition.

United Nations Conference on Trade and Development (UNCTAD). *Technology and Innovation Report 2021*. n.d. https://unctad.org/page/technology-and-innovation-report-2021.

Türk, Volker. "Artificial Intelligence Must Be Grounded in Human Rights, Says High Commissioner", accessed March 28, 2024.

https://www.ohchr.org/en/statements/2023/07/artificial-intelligence-must-be-grounded-human-rights-says-high-commissioner.

United Nations General Assembly (UNGA). "Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development," accessed March 28, 2024. https://daccess-ods.un.org/access.nsf/Get?OpenAgent&DS=A/78/L.49&Lang=E.

United Nations Human Rights Council. "Resolution on New and Emerging Digital Technologies and Human Rights." No. 41/11, July 13, 2023.

https://www.ohchr.org/en/hr-bodies/hrc/advisory-committee/digital-technologies-and-hr.

Veale, Michael, Kira Matus, and Robert Gorwa. "AI and Global Governance: Modalities, Rationales, Tensions." *Annual Review of Law and Social Science* 19 (October 2023). https://www.annualreviews.org/loi/lawsocsci.

World Economic Forum (WEF). *Quantum Economy Blueprint 2024*. n.d. https://www3.weforum.org/docs/WEF_Quantum_Economy_Blueprint_2024.pdf.

Appendix

Data-based Systems (DS)

Nuclear Technologies (NT)

Risks include privacy breaches, algorithmic bias, and manipulation.

Regulatory landscape is still evolving, with fewer international agreements.

Focus on ethics, privacy, and transparency.

Widely accessible, with potential for global deployment.

Data-centric: involves digital information and algorithms.

Requires international cooperation to address global challenges.

Technical expertise and standards are necessary for safety and reliability.

Capacity building and knowledge transfer are needed to ensure equitable access and responsible use.

Clear evidence of potential catastrophic consequences from misuse or accidents.

Highly regulated by international treaties and organizations (e.g., IAEA).

Focus on safety, security, and non-proliferation.

Limited number of state actors with access to technologies.

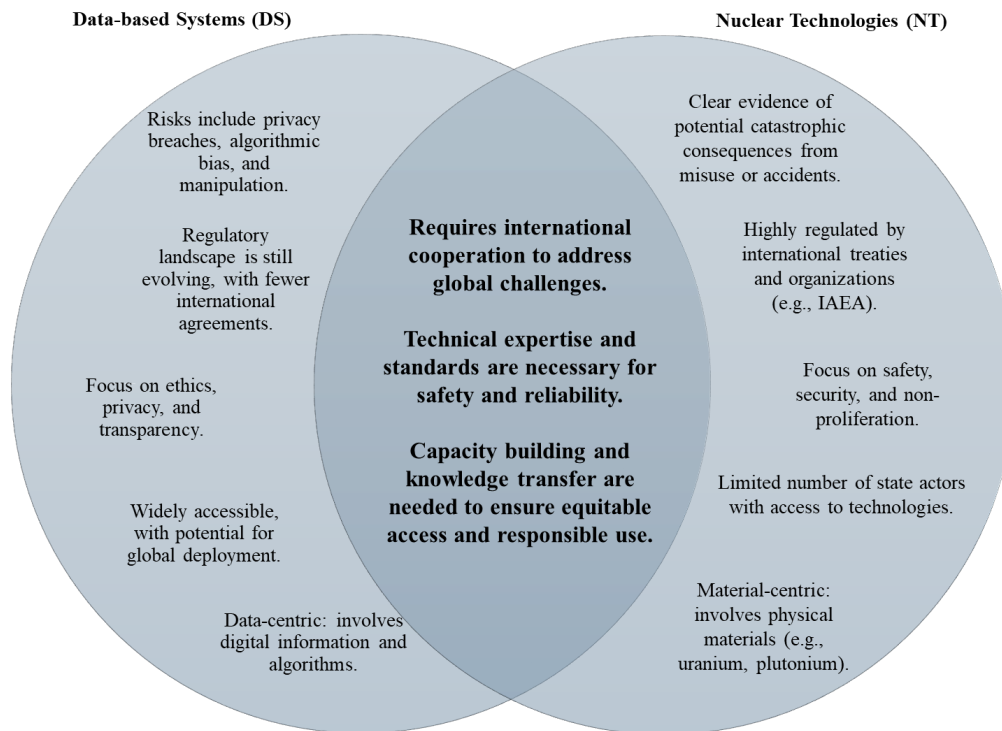Material-centric: involves physical materials (e.g., uranium, plutonium).

FIGURE 1. Comparative Analysis: Governance of DS versus NT. Source: Authors own analysis from various sources

AI and NT both pose significant threats, they differ significantly and have distinct characteristics, as shown in Figure 1.

- **Nature of the Threat:** AI and NT differ significantly in the nature of the threat they pose.

- **Scope and impact**: The scope and impact of AI and NT differ significantly.

- **Regulatory Challenges for Exponential Pace of Innovations:** Regulatory challenges for AI and NT differ significantly.

- **Technical Complexity and Control:** AI and NT differ significantly in technical complexity, geopolitical influence, and controlled interventions, to date.

- Paradoxically, there are various aspects where DS and NT intersect or share common themes as follows:

- **Data Management and Analysis:** Both DS and NT rely heavily on data management and analysis for decision-making and optimisation.

- **Safety and Risk Management:** Safety and risk management are critical in both DS and NT domains.

- **Regulatory Compliance and Governance:** DS and NT are subject to regulatory frameworks and governance structures to ensure ethical, legal, and responsible use.

- **Cybersecurity:** Cybersecurity is a common concern for both DS and NT due to the sensitive nature of data and nuclear facilities.

- **Legal and Ethical Considerations:** Legal and ethical considerations are paramount in DS and NT domains.

Despite the differences between DS and NT, Table 1 highlights key lessons that can be applied to the governance of DS.

TABLE 1: Summary of lessons from the IAEA for IDA

Area	Shortcomings	Successes	Adaptations for IDA
International Cooperation	Limited enforcement mechanisms	Facilitated cooperation among member states	Establish more robust enforcement mechanisms through binding international agreements.
Technical Expertise and Standards	Limited in addressing rapidly evolving technology	Establishment of technical standards and expertise	Develop agile regulatory frameworks to address rapid technological changes.
Safeguards and Inspections	Resource intensive and prone to manipulation	Effective monitoring and verification mechanisms	Implement advanced monitoring technologies and independent verification systems as well as access to pre-market approval-process.
Information Sharing and Transparency	Restricted access to sensitive information	Enhanced transparency and information exchange	Establish transparent information-sharing mechanisms with appropriate safeguards.
Capacity Building	Unequal distribution of resources and knowledge	Improved capacity and skills development	Promote equitable access to resources and knowledge through capacity building,direct investments, and technology transfer.
Education and Outreach	Limited public engagement	Increased public awareness and engagement	Enhance public awareness and involvement through education and outreach initiatives.

Source: Authors own analysis from various sources

G2O
BRASIL 2024
BUILDING A JUST WORLD
AND A SUSTAINABLE PLANET

T2O
Brasil 2024
Let's rethink the world

Let's
rethink
the world

CEBRI
Rethink Tank

FUNDAÇÃO
ALEXANDRE
DE GUSMÃO

ipea
Instituto de Pesquisa
Econômica Aplicada