Policy Network on Artificial Intelligence Input to the UN AI Advisory Body's Open Consultation

March 2024

This multistakeholder submission was drafted under the Policy Network on Artificial Intelligence (PNAI). The Policy Network is a global multistakeholder effort hosted by the United Nations' Internet Governance Forum to foster open global dialogue on AI and contribute to the global AI policy discourse. For more on the PNAI, see www.intgovforum.org/en/pnai

Opportunities and Enablers

To effectively unlock AI's enablers and opportunities, it is imperative to concentrate on advancing digital public infrastructure (DPI), which is crucial for realizing the Sustainable Development Goals (SDGs). DPI underpins a worldwide AI ecosystem rooted in principles of human rights, openness, accessibility, and multistakeholder inclusivity, guaranteeing the broad reach of AI's advantages irrespective of geographical, economic, or experiential disparities. Development of skilled digital talent and s embedding AI competencies at all levels of education is equally essential.To render AI development more inclusive, initiatives should prioritize an Internet ecosystem, shared computational access, and data protection. Data sharing should be promoted by establishing an interoperable common data space model. Strategic investments in digital infrastructure are necessary to facilitate AI's growth and expansion. Promoting universal and meaningful connectivity is vital due to AI's significant data and connectivity requirements, it will help ensure equitable access to computational resources and data sharing. To guide AI's technological development, it is crucial to establish effective governance frameworks rooted, with respect and recognition in our ethical, moral and local cultural values.

Establishing high-quality, adaptable data labelling and testing standards is essential for facilitating broader testing and deployment of AI solutions by SMEs. These standards must be flexible enough to accommodate the diversity of local requirements while promoting the global deployment of AI technologies.

The global dissemination and cultivation of AI talent are contingent upon synergistic North-South and South-South development cooperation, technical assistance, and academic partnerships. This approach encompasses fostering joint tertiary / higher education programs, including PhD programs, student exchanges, and online learning platforms that comprehensively address AI skills enhancement.

The United Nations, regional and national institutions and initiatives catalyse talent development across various AI domains, promoting a balanced, unbiased and sustainable

global AI ecosystem. International collaboration is paramount for harnessing AI in scientific research and innovation benefiting humanity and achieving the SDGs.

Investment strategies related to AI should be comprehensive and inclusive, consider public-private partnerships, international cooperation, and regional and national investments. They should create a robust and sustainable framework for AI skill, capacity, and policy development, especially in economically disadvantaged regions. The involvement of all stakeholders is crucial, and mechanisms such as raising funds for all stakeholder groups to join the dialogue and equal regional participation must be in place.

Risk and Challenges

The risks and challenges associated with the development and deployment of AI systems are multifaceted. Addressing them requires a comprehensive approach that integrates principles of fairness, protection of human rights, and proactive measures.

AI-linked technologies should be categorised through a process determining the local and global risks implications, including risks to the information space, and posting details and recommendations to policymakers on how to mitigate the identified potential risks. A multistakeholder panel composed of at-risk groups from the Global North and the Global South industrialised and underdeveloped regions should be convened to discuss the specific implications of AI technologies on a local and global basis. The panel should come to a recommendation and be open for public viewing. Effective mechanisms should be established for feedback and complaining of problems of AI applications on a continuous basis, along with effective mechanisms to address properly potential harms, redress harms, and lead to improvements in AI systems.

Establishing red line standards for AI developers, deployers and users, and a red line assessment mechanism through Impact Assessments Evaluations, will help to obtain the best out of the AI technologies. A global "denylist" mechanism for people or organizations that have violated the red line needs to be installed, and the list should be published openly in a timely manner. Red line can include axioms like: no AI system should be able to copy or improve itself without explicit human approval, accountability and assistance; no AI systems should substantially increase the ability of actors to design weapons of mass destruction; no AI system should be able to autonomously execute cyberattacks resulting in loss of human life, serious financial or institutional losses or equivalent harm; no AI system should be able to consistently cause its designers or regulators to misunderstand its likelihood or capability to cross any of the preceding red lines.

AI deployers cannot provide systems with major hidden harms, or intentionally mismanage or misuse the data bases or logic for the systems training; developers cannot agree in algorithms that embed bias based on gender, race, ability, language, culture, region, political or religious values differences etc.; users must not abuse AI systems to cause major harm through techniques such as deep fakes, malicious robots, or illegal undertakings, including invading privacy, security and safety or violating any individual's rights, granted by the United Nations.

Mandating a global recall mechanism for contaminated large models should be implemented at local and global levels, where they must be evaluated by an independent testing agency before they would be available in the market again.

A code of conduct for developers and deployers to provide transparency of their AI systems will help to mitigate such risks with a tiered approach.

Guiding Principles to guide the formation of new global governance institutions for AI

The proposed principles and functions need to incorporate more specific guidance on how to implement these principles effectively, particularly in diverse socio-cultural contexts. Some of the gaps can be improved through a reflective approach in the following ways:

Addressing Data Gaps and Biases

Addressing data gaps and biases in AI systems is crucial to ensure inclusivity, effectiveness, and just data value creation (JDVC). It can involve incorporating more inclusive methodologies such as data commons and considerations of the contextual realities in less structured data ecosystems, which can facilitate the creation of quality datasets that represent various communities to avoid perpetuating historical biases and underrepresentation of marginalised groups.

Harmonising Ethical Considerations

Apart from UNESCO's Recommendations, most AI ethics frameworks do not adequately reflect the social, cultural, and ethical contexts of diverse populations, especially in the Global South. Reflective AI ethics involves challenging perspectives, values, and norms that are often commonly accepted, recognizing other forms of knowledge systems, institutions, and ensuring that ethical frameworks align with various common ethical viewpoints that reflect the rich cultural diversity of our interconnected world.

Mitigating Algorithmic Power Imbalance

Algorithmic dispossession, exploitation, and oppression highlight how power imbalances are perpetuated through AI technologies. A reflective approach can involve rethinking labor practices, scientific experimentation, and international governance standards to prevent the centralization of power and assets in the hands of a minority.

Meaningful Dialogue

Promoting inclusive dialogue between stakeholders in AI development is essential to prevent predatory inclusion and algorithmic oppression. A reflective approach to AI governance requires creating meaningful avenues for marginalised groups to influence decision-making processes and avoid perpetuating algorithmic exploitation and dispossession.

Balancing Innovation with Responsibility

Clear lines of responsibility, liability frameworks, and regulatory oversight are necessary as AI systems become more autonomous, striking a balance between rapid innovation and responsible governance is central to a reflective approach to AI governance.

A Reflective Approach for Horizon Scanning, Scientific Consensus, and Foresight

Integrating reflective principles into horizon scanning processes, can ensure stakeholders anticipate future trends more inclusively, by considering the impact on marginalized communities and regions often overlooked in traditional foresight practices.

Behavioural Science

Behavioural science is paramount in the drafting process of AI governance. By understanding how individuals interact with AI systems, helps creating user-centric frameworks that address biases and promote fairness hence aligning AI with societal values and preferences.

Institutional Functions that an international governance regime for AI should carry out

An inclusive international AI Governance Body with representatives from government, academia, the private sector, and civil society is needed to set global standards, promote data governance interoperability, and align AI governance in keeping with the UN Charter and Sustainable Development goals. Sub-bodies with sectoral expertise will address salient risks and challenges pertaining to: standards for AI system-training that maintain personal data privacy, intellectual property; equity and fairness in AI outcomes; public-private partnerships for a resilient workforce; cybersecurity risks; infrastructure and global existential risks; environmental impacts; standards for use in military/defence contexts. Given the pace of private sector AI advancement and its imminent convergence with quantum computing, this effort must be a robust public-private partnership, fostering an industry animated by the values articulated within the UN SDGs.

Governance will support a safe and fair AI industry. This body's funding should be substantially obtained via levies from AI companies, commensurate with their computing resources and engagement in the marketplace, with independent oversight to prevent regulatory capture. Public-private funding models to consider include the International Accounting Standards Board (IASB) and the International Air Transport Association (IATA).

The deployment of AI systems should be conditional upon the successful completion of validated ethical and social impact assessments, using established technology assessment mechanisms with attention paid to systems' underlying training data. This body will establish standards for independent AI audits to mitigate harmful practices and outcomes

This institution will foster partnerships between governments, academics and the private sector to ensure AI innovations are not only socially responsible — a baseline — but to advance AI's progress in tackling societal needs as articulated in the UN SDGs.

Experts within this body may advise on legal frameworks to establish AI developers' liability for harms resulting from negligence. Given the impossibility of a monitoring structure encompassing all evolving risks, it is essential for legal liability to stand as an incentive for adequate safeguards.

PNAI suggests rigorous consideration of taxation proposals designed to account for the societal and environmental costs associated with AI's progression. Proposed taxation strategies include consumption-based taxes targeting substantial environmental impacts and automation taxes to address the potential mass displacement of human labour.

We suggest that the necessary institutional functions for Governing AI can be built on existing structures, such as the WSIS framework that encompasses a wide array of institutions, initiatives, mechanisms aimed at addressing various aspects of the digital ecosystem.

Other comments on the International Governance of AI section (aside from Principles and Functions, covered in above questions)

In the context of continued and rapid development of AI, we acknowledge the strategic importance of strengthening interoperability of AI governance while fostering a pro-innovation environment in the international governance of AI.

PNAI proposes following steps to increase interoperability of AI governance:

- Define and agree exactly what needs to be addressed at the global level. This could include existing and emerging risks related to AI, with focus on issues that have occurred or been observed in practice. To achieve interoperability in AI governance, we propose that the development of regional and/or global regulatory policies, guidelines and principles should be agile, reflexive, and inclusive, and evolve according to the AI maturity level.

- International funding programs with focus on interoperability of AI governance should be established.

Advocate for the establishment of international standards and certification processes for AI technologies.

- Strengthen legislative cooperation. This can be achieved using various instruments that promote international cooperation. National regulators should strengthen cross-border and pan-industry cooperation. They should ensure AI governance frameworks facilitate inclusiveness and a level playing field for all to benefit from AI. Unnecessary costs and fragmentation due to different regional requirements should be avoided as far as possible. AI legislation should always be in line with human rights principles, norms and international standards.

- Foster regional multi-stakeholder initiatives and interlink them globally to strengthen cooperation.We need to allow different speeds of cooperation based on different levels

of maturity and public policy needs while not lose sight of the goal of increased interoperability of AI governance.

- Reduce regional disparities to encourage maturity increases. This requires a comprehensive political, scientific and industrial exchange and cooperation. Proven best practices from regions should be selected with interoperability in mind for adaptability and usability for regions with lower levels of maturity.

- Monitor and evaluate progress in reaching policy goals set on national, regional and global levels to advance interoperability of AI governance. It is essential to continuously track progress made against set out goals , identify areas requiring improvement, adjust strategies accordingly, and evaluate overall effectiveness of implemented measures.

- Uphold and strengthen the Internet Governance Forum (IGF) process, its regional and global multistakeholder initiatives including the Policy Network on AI. We need to foster spaces for open, transparent, inclusive and transborder consensus and capacity building of AI governance. That includes in particular policy advocacy skills, hence enabling stakeholders to engage in policy discussions and influence decision-making processes related to AI governance.

Any other feedback on the Interim Report

- Regarding capacity development for AI (technical deployment, policymaking, and implementation) and the development of individual competences (computer literacy, privacy safeguards, etc.).

- Just-in-time learning

- AI governance and related topics in the curriculum of academic post-graduate studies.

- Holistic capacity development

- Those strategies will develop and assimilate the benefits of AI into their economies.

- The range of policy issues AI encompasses complicates reaching international agreements for financing and capacity building. There is little agreement among stakeholders as to which policy problems should be prioritised. Two pathways forward are notable: (a) developing a new centralised AI institution or (b) strengthening the coordination, capacities, and legitimacy of existing institutions on a national level. A necessary first step for assessing potential regimes is developing normative criteria.

- An AI regulation complex would allow for the advantages of a decentralised approach to materialise. A strong regulation complex would also have the flexibility to adapt to new policy challenges. Clarifying institutional remits and strengthening communication channels would help define which nodes(s) should address problems provoked by new AI developments. Perhaps most importantly, the “polycentric ordering” of the regulation

complex means no external authority can impose coordination, as it is only generated by committed participants.

- It is important to strengthen ties between nodes in the network like international bodies whose remits related to AI overlap, such as OSET, UNESCO, WHO, ITU and the OECD. This could be followed by negotiations to agree on remits, as is common when international organisations have overlapping jurisdiction. Therefore, some recommendations can be offered for national and local governments to coordinate a successful AI strategy:

- Wise, calibrated interventions are critical

- A private public strategy for financing the AI strategy should be sought according to local contexts.

- Also public investment initiatives to support development, maintenance and innovation in the field of AI are needed.

- Support AI business models: AI SaaS, AI Platform as a Service (PaaS), Licensing AI, AI Professional Services, AI Data Monetization, AI Pay-as-you-go Model, AI Partnerships and Joint Ventures, through special taxes and incentives.

- Leverage AI to innovate in the public interest and deliver breakthroughs to improve quality of life for people

- Improve transparency, interpretability and accountability for AI models for companies, and deployers through inclusive, rights-respecting AI governance frameworks

- Support the development of international AI rules and standards.

- Development of policy frameworks and research on discrimination and bias, and advocacy efforts to ensure civil society.

- Support the continuous evaluation and adaptation of governance frameworks

- Share best practices and successful initiatives.