



Internet
Governance
Forum
RIYADH
2024

The AI Governance We Want, Call to Action: Liability, Interoperability, Sustainability & Labour

PNAI Policy Brief 2024

DECEMBER 2024
INTERNET GOVERNANCE FORUM
POLICY NETWORK ON ARTIFICIAL INTELLIGENCE

Authors

PNAI Sub-group on Liability as a mechanism for supporting AI accountability

Drafting team leaders: Caroline Friedman Levy (Center for AI and Digital Policy, United States); Umut Pajaro Velasquez (Independent AI Ethics and Governance Researcher, Colombia)

Penholders: Aji Fama Jobe (Women Techmakers Banjul & ISOC Youth Standing Group, The Gambia); Antara Vats (ARTICLE19, India); Azeem Sajjad (Ministry of IT & Telecom, Pakistan); Brian Scarpelli (ACT | The App Association, United States); Debby Kristin (EngageMedia, Indonesia); Dr. Patrick M. Bell (College of Information and Cyberspace, National Defense University, United States); Fadi Salem (MBR School of Government, UAE); Jasmine Ko (DotAsia, Hong Kong China); Karen Mulberry (IEEE, United States); Lika Døhl Diouf (ECLAC, Trinidad and Tobago); Lufuno T Tshikalange (Orizur Consulting Enterprise, South Africa); Omor Faruque (Dynamic Teen Coalition & Project OMNA, Bangladesh); Saba Tiku Beyene (Independent AI Researcher, Ethiopia); Sameer Gahlot (National Internet Exchange of India, India); Sana Nisar (Habib Bank Limited, Pakistan); Seth Hays (Digital Governance Asia); Yasir Zunair (Ostbayerische Technische Hochschule Amberg-Weiden, Germany); Yik Chan Chin (Beijing Normal University, China)

Sub-group on Environmental Sustainability within Generative AI Value Chain

Drafting team leaders: Shamira Ahmed (Data Economy Policy Hub, South Africa); Mohd Asyraf Zulkifley (Universiti Kebangsaan Malaysia, Malaysia)

Penholders: Dr. Muhammad Shahzad Aslam (Xiamen University, Malaysia); Lufuno T Tshikalange (Orizur Consulting Enterprise, South Africa); Mary Rose Ofianga (Wadhvani Foundation, Philippines); Nabuzale Doreen Nandutu (eLAAB Limited, Uganda); Siti Raihanah Abdani (Universiti Teknologi MARA, Malaysia); Somya Joshi (Stockholm Environment Institute, Sweden)

Sub-group on AI governance, Interoperability, and Good practices

Drafting team leaders: Dr. Yik Chan Chin (Beijing Normal University, China); Dr. Chafic Chaya (RIPE NCC, United Arab Emirates); Sergio Mayo (Aragon Institute of Technology, Spain); Amado Espinosa (Medisist, Mexico)

Penholders: Debora Irene Christine (Tifa Foundation, Indonesia); Dooshima Dapo-Oyewole (College of Engineering and College of Business and Economics, Boise State University, USA); Dr. Allison Wylde (Glasgow Caledonian University, UK); Dr. Jin Cui (ITU, China); Antara Vats (National Law School of India University, India); Dr. Patrick M. Bell (National Defense University, USA); Heramb Podar (Center for AI and Digital Policy, India); Jochen Michels (Kaspersky, Germany); Lufuno T Tshikalange (Orizur Consulting Enterprise, South Africa); Muriel Alapini (Benin IGF, Benin); Poncelet Ileleji (Jokkolabs Banjul/NRI, Gambia, Gambia); Xitao Jiang (CNNIC, China); Concettina Cassa (AgID, Italian Government, Italy)

PNAI Sub-group on Labour issues throughout AI's life cycle

Drafting team leaders: Olga Cavalli (University of the Defense of Argentina, Argentina); Germán López Ardila (Javeriana University, Colombia)

Penholders: Azeem Sajjad (Ministry of IT & Telecom, Pakistan); Christian Gmelin (GIZ, Germany); Julian Theseira (Center for AI and Digital Policy CAIDP, Malaysia); Lufuno T Tshikalange (Orizur Consulting Enterprise, South Africa); Mary Rose Ofianga (Wadhvani Foundation); Mondli Mungwe (Mungwe Industries International, South Africa); Muhammad Shahzad Aslam (Assistant Professor, Xiamen University, Malaysia); Saba Tiku Beyene (Independent AI Researcher, Ethiopia); Samuel Jose (Children and Youth Major Group to UNEP as individual contributor, Indonesia); Shantanu Prabhat (Individual contributor, India); Shobhit S. (Individual contributor, India)

Proofreading and editing team: Patrick Bell, Doreen Nandutu Nabuzale, Eleanor Sarpong, Muriel Alapini

Editor: Maikki Sipinen

The PNAI report drafting teams and authors would like to thank PNAI MAG facilitators Amrita Choudhury and Audace Niyonkuru for their support in steering the drafting process.

Disclaimer

The views and opinions expressed herein do not necessarily reflect those of the United Nations Secretariat. The designations and terminology employed may not conform to United Nations practice and do not imply the expression of any opinion whatsoever on the part of the Organization. Some illustrations or graphics appearing in this publication may have been adapted from content published by third parties. This may have been done to illustrate and communicate the authors' own interpretations of the key messages emerging from illustrations or graphics produced by third parties. In such cases, material in this publication does not imply the expression of any opinion whatsoever on the part of the United Nations concerning the source materials used as a basis for such graphics or illustrations. Mention of a commercial company or product in this document does not imply endorsement by the United Nations or the authors. The use of information from this document for publicity or advertising is not permitted. Trademark names and symbols are used in an editorial fashion with no intention of infringement of trademark or copyright laws. We regret any errors or omissions that may have been unwittingly made. © Tables and Illustrations as specified

Suggested Citation: Policy Network on Artificial Intelligence (2024) The AI Governance We Want, Call to action: Liability, Interoperability, Sustainability & Labour Sipinen, M. (Ed.) Internet Governance Forum

This publication may be used in non-commercial purposes, provided acknowledgement of the source is made. The Internet Governance Forum Secretariat would appreciate receiving a copy of any publication that uses this publication as a source.

List of contents

PART 1 The Policy Network on AI in 2024

Liability as a Policy Lever in AI Governance: Amplifying the Global Discussion (p.3)

Environmental Sustainability and the Generative AI Value Chain (p.5)

Legal, Technical and Data Interoperability: Gaps and Effective Instruments in Current AI Governance Landscape (p.8)

Promoting Multistakeholder Dialogue on Artificial Intelligence Related Labour Issues (p.12)

PART 2 Policy Reports

Liability as a Policy Lever in AI Governance: Amplifying the Global Discussion (p.1)

Environmental Sustainability and the Generative AI Value Chain (p.17)

AI Governance, Interoperability, and Good Practices (p.54)

Promoting Multistakeholder Dialogue on Artificial Intelligence Related Labour Issues (p.83)

The Policy Network on AI in 2024

This report outlines the 2024 work conducted by the Policy Network on Artificial Intelligence (PNAI)¹, highlighting key findings, sharing outcomes, and presenting policy recommendations. PNAI is a global, multistakeholder effort hosted by the United Nations' Internet Governance Forum (IGF). It provides a platform for stakeholders and changemakers in the AI field to contribute their expertise, insights, and actionable recommendations.

In 2024, the PNAI community concentrated its discussions and efforts on four key topics:

- Liability as a mechanism for supporting AI accountability
- Environmental sustainability within the Generative AI value chain
- AI governance, interoperability, and good practices
- Labour issues throughout AI's life cycle

This report is divided into two parts:

Part 1 provides an overview of PNAI's activities in 2024, summarises key findings and selected recommendations across these four topics.

Part 2 contains detailed thematic reports prepared by dedicated PNAI subgroups. These reports feature in-depth analysis, examples, and insights on progress, policies, and regulations from different regions. Each thematic report concludes with a set of multistakeholder recommendations for further action.

The insights and recommendations derived from PNAI's 2024 work will be presented and discussed at the 19th annual IGF meeting, to be held in Riyadh, Kingdom of Saudi Arabia, in December 2024.

Development of the 2024 Policy Report and recommendations by PNAI

This report and its recommendations were developed through extensive exploration and multi-stakeholder discussions within the PNAI community. In April 2024, PNAI established four Sub-groups, each tasked with drafting a report on one of the focal topics. These Sub-groups, led by volunteer team leaders, conducted information gathering, consulted with experts from the PNAI community and beyond, and held team meetings to debate and refine their topics. Their work culminated in the preparation and editing of discussion papers. PNAI's monthly Multistakeholder Working Group online meetings guided the overall process. These meetings provided updates, reviewed draft reports, and offered feedback and advice. In October 2024, the draft discussion papers were shared with the PNAI community for review and feedback. Two online workshops were held to discuss the drafts, and additional comments were submitted in writing. Based on this input, the sub-groups finalised their reports.

¹ IGF, [Policy Network on AI website](#), 2024

PNAI's Mission and earlier work

The PNAI's core mission is to act as a platform to foster dialogue among different stakeholders, ensure representation from the Global South, and contribute to the global discourse on AI policy. It addresses key issues related to Artificial Intelligence (AI) and Data Governance. Participation in the PNAI's community, dialogues, and meetings is open to all stakeholders. PNAI was established in response to a request from the IGF community. The 2022 IGF meeting in Addis Ababa highlighted the potential for the IGF to act as a platform for cooperation on AI policy. The meeting's concluding messages suggested that "a policy network on artificial intelligence could be considered for the upcoming work streams to review the implementation of different principles with appropriate tools and metrics."²

Launched in May 2023, the PNAI began its second year of activities in spring 2024. During its inaugural year, the PNAI focused on three main topics:

- Interoperability of Global AI Governance
- AI, Gender, and Race
- AI and the Environment

This work resulted in a report and recommendations addressing these areas.³

The PNAI's efforts on AI policy and data governance draw from previous IGF discussions, reports, and the collective expertise of its global community. The network aims to leverage this foundation to drive impactful dialogue and actionable policy recommendations on emerging AI challenges. The 2024 report focuses on four areas: accountability, interoperability, labour and sustainability in the AI life cycle.

² IGF, [Addis Ababa IGF Messages](#), 2022

³ Policy Network on AI, [Strengthening multi-stakeholder approach to global AI governance, protecting the environment and human rights in the era of generative AI - A report by the Policy Network on Artificial Intelligence](#), 2023

HIGHLIGHTS OF PNAI WORK IN 2024:

Liability as a Policy Lever in AI Governance: Amplifying the Global Discussion

Who is financially responsible when an autonomous vehicle causes an accident? How should damages be allocated if an AI trading algorithm triggers panic in financial markets? Or when an AI-powered medical diagnostic system produces skewed outcomes, harming patients?

Policymakers and stakeholders focused on AI governance consistently emphasise the importance of holding AI developers and deployers accountable for the harms caused across the AI lifecycle. However, mechanisms to ensure this accountability remain poorly defined. Liability in this document refers to the legal and financial responsibility for harm or damage caused by AI systems. Establishing clear liability guidelines offers policymakers an opportunity to promote AI accountability, and to incentivize responsible AI development and deployment. Policymakers and stakeholders, however, are yet to come to a consensus regarding key questions on how to assign liability within the AI lifecycle.

Liability frameworks are a crucial component of AI governance because they address the challenge of constantly advancing technologies, as well as new types of risks that emerge. These frameworks can help mitigate AI risks over time as they provide flexibility, adapting to novel harms without requiring constant regulatory updates. Yet, legal experts argue that traditional liability frameworks are not fit-for-purpose for AI systems.

The unique characteristics and complexity of AI systems pose significant challenges for existing liability regimes. The opacity of AI systems poses significant challenges, as jurisdictions will need access to information about how AI systems factored inputs and characteristics into their decision-making. Clear and rigorous liability frameworks could incentivize algorithmic transparency compliance and innovation. Specialized courts with technical expertise or additional training might be needed to make informed decisions on AI liability. Neither the opacity nor the autonomy of AI systems ought to exempt developers and deployers from accountability.

AI harms can occur at various stages of its lifecycle and supply chain. The complex web of interwoven activities of actors - for example, data providers, model developers, practitioners, and end-users - makes identifying the flaw that caused harm and assigning liability very difficult. Placing a greater responsibility for safeguards and quality controls on foundational players could incentivize risk assessment of actors downstream.

By adhering to the standards for ethical and safe AI development and deployment, companies can significantly reduce their liability risks and demonstrate due diligence. A robust liability regime incentivizes the adoption of industry standards and provides a framework for accountability when those standards are not met.

The questions surrounding AI liability across the value chain are intricate, challenging, and likely to **cross national boundaries**. Amplifying the conversation about AI liability within the **global AI governance community** is critical. Coordinated global principles for AI liability would incentivize developers and deployers to adhere to consistent standards, reducing risks of “regulatory arbitrage” and “forum shopping”, where companies exploit jurisdictions with lenient regulations. Global harmonisation would also create a level playing field and help protect vulnerable individuals, communities, and Global Majority countries from bearing disproportionate harm.

Most countries and regions are in the early stages of their attention to AI liability. While attention to AI liability has been prominent in the European Union for several years, Brazil's draft bill on AI governance featuring liability levers is another notable example. The Council of Europe's *Framework Convention on Artificial Intelligence*, which covers AI systems across the lifecycle, represents the first legally binding international treaty in this area and serves as a promising starting point. International frameworks will be particularly useful in managing cross-border issues and supporting nations with limited resources to manage complex AI liability regimes. Addressing AI liability is essential for filling a critical gap in global AI governance. Robust liability frameworks are indispensable for promoting safe and ethical AI outcomes and providing recourse for harm. By prioritising this issue, policymakers can help build a more accountable and equitable AI ecosystem.

RECOMMENDATIONS

- Establish a **Global AI Liability Task Force**, bringing together experts from diverse jurisdictions to develop harmonized principles that can be adapted across frameworks, and exploring the development of an international framework for AI liability.
- Formalize adherence to **ethical AI industry standards** into liability frameworks – to incentivize AI companies to rigorously implement the voluntary safeguards outlined by standards-setting bodies.
- Investigate the potential applicability of a "**chain of responsibility**" framework for AI liability to clarify accountability across the complex AI lifecycle.
- Develop **capacity-building initiatives focused on AI liability in Global Majority countries**. Address both the technical and legal aspects of enforcement, coupled with initiatives to bridge this divide. Invest in digital infrastructure and promotion of digital literacy to ensure effective implementation and enforcement of AI regulations.

Read the full report by the PNAI Sub-group on Liability as a mechanism for supporting AI accountability throughout AI's life cycle in **PART 2**

HIGHLIGHTS OF PNAI WORK IN 2024:

Environmental Sustainability and the Generative AI Value Chain

The exponential expansion of generative artificial intelligence (Gen-AI) platforms and models is driving innovation across industries. However, the environmental costs of these advances are often overlooked. The current global dialogue on AI governance and environmental risk mitigation tend to focus on improving energy efficiency. This narrow focus fails to address the **broader sustainability and socio-technical challenges tied to the Gen-AI value chain**. Each stage of the Gen-AI value chain contributes to carbon footprint and resource depletion – from computer hardware and cloud platforms to foundation models, model hub and machine learning operations, and finally to applications and services.

Unlike traditional AI, which emphasises accuracy and efficiency in completing tasks, Gen-AI prioritises creativity and producing novel outputs. These differences require distinct governance approaches. Currently, there is limited **governance of Gen-AI at the international level** and the existing initiatives often lack the coordination and capacity needed to address its complex environmental impact. Assessing and mitigating the environmental impact of Gen-AI technologies is particularly important for the Global Majority, who are disproportionately affected by climate change driven by unsustainable practices in the digital economy. Many existing standards and best practices for AI are rooted in the socio-technical contexts of the Global North, making them poorly suited to the realities and needs of the Global Majority.

Efforts to measure Gen-AI's environmental impact and ensure its compliance with global sustainability standards are in their early stages. **Establishing accurate and comprehensive metrics** is crucial for assessing the full environmental impact of Gen-AI. Clear metrics provide AI developers and policymakers a framework for understanding, managing and assessing energy consumption, resource utilisation, and emissions associated with AI development, deployment, and usage. **Defining indicators** is vital for measuring Gen-AI progress toward environmental sustainability goals and helps assess the effectiveness of sustainability initiatives and identify areas for improvement.

Robust data governance is critical in minimising the environmental harm of Gen-AI. Establishing clear policies, standards, and processes for data management ensures that data used to train and deploy Gen-AI models is collected, processed, and stored responsibly. **To effectively assess Gen-AI environmental impact, stakeholders need various types of data, including for example:** data on energy usage, resource consumption, and emissions; socioeconomic data (to understand socioeconomic implications); and contextual data (such as data on political and economic conditions or characteristics of affected populations).

Integrating data governance with environmental impact assessments (EIAs) in the Gen-AI value chain faces several challenges: **Data complexity** (Gen-AI requires vast volumes of structured

and unstructured data from various sources which undergo continuous retraining, making assessments difficult), **Regulatory and Ethical Framework Gaps** (There is a lack of policies to measure Gen-AI's energy consumption, carbon emissions, and electronic waste, as well as frameworks that link digital and environmental concerns), **Complexity of Gen-AI Value Chain** (the Gen-AI value chain involves multiple stakeholders, different environmental standards across jurisdictions and industries) and **Bias and (Un)Fairness** (biases in training data can result in skewed environmental assessment if the data does not represent diverse ecological contexts and stakeholder perspectives).

By developing comprehensive metrics, fostering multistakeholder dialogue, and leveraging high-quality data, stakeholders can collaboratively reduce the ecological footprint of Gen-AI technologies. Integrating data-driven approaches with responsible practices is essential for steering the Gen-AI value chain toward sustainability. This approach will ensure that the benefits of Gen-AI are realised without compromising environmental integrity.

RECOMMENDATIONS

- **Develop Comprehensive Sustainability Metrics for Gen-AI.** Governments and international organizations should create standardized metrics that measure Gen-AI's environmental impact across the entire value chain.
- **Support Regionally Relevant Innovation Ecosystems.** Policies should incentivize Gen-AI applications in climate change mitigation, adaptation, and loss and damage while fostering regionally relevant innovation ecosystems. Local entrepreneurs, businesses, and academia in the global Majority need to be central to developing green digital economies.
- **Leverage Official Development Assistance (ODA) for Sustainable Gen-AI.** ODA can support Low and Middle-Income Countries (LMICs) in developing sustainable AI infrastructure, building local capacity, and creating green jobs, promote self-sustaining innovation by providing and facilitating digital public goods (such as open-source AI tools and data access), support the investment in local talent, including growth of local policy and technical expertise to create an enabling policy and regulatory environment that fosters sustainable, autonomous digital economies in LMICs.
- **Integrate Circular Economy Principles.** Establish policies that promote circular economy practices in the Gen-AI value chain, such as governance to reduce e-waste through hardware reuse and recycling.
- **Implement Environmentally Focused Data Governance.** Data governance frameworks should prioritize a just twin transition approach, that prioritizes environmental and social impacts, ensures equitable data access, transparency in AI models, and integration of environmental data to mitigate the environmental impacts of Gen-AI.

- **Apply Decolonial Socio-Technical Foresight.** A decolonial socio-technical foresight approach can empower LMICs to envision and shape their Gen-AI futures according to local priorities and enable countries in the global Majority to design Gen-AI ecosystems that align with self-determination, sustainability, and intergenerational justice.
- **Collaboration, coherence and coordination within the UN system.** Collaboration, coherence and coordination within the UN system would foster stronger alignment between the digital and climate agendas, enabling multistakeholder dialogue on leveraging AI and digital technologies for environmental sustainability while addressing their growing energy consumption and resource intensity. For example: The 29th United Nations Climate Change Conference (COP29) Declaration on Green Digital Action has underscored the critical role of digital technologies (including AI) in achieving global climate objectives, the Internet Governance Forum (IGF) Policy Network on AI and other Internet governance forums must encourage knowledge exchange with other UN organisations such as the UN Framework Convention on Climate Change (UNFCCC) and the International Telecommunication Union (ITU) to consolidate efforts and advance a just green digital (twin) transition, for people and the planet.

To illustrate the benefits, we present case studies on the use of Gen-AI for environmental conservation, resource optimization, and climate change mitigation:

Case study 1 - **Environmental Sustainability and AI for Forest Fire Management in the ASEAN Countries,**

Case study 2 - **Environmental Sustainability and AI for Climate Change Management in African Countries,**

Case study 3 - **Improving Air Quality with Generative AI in Ghana, and,**

Case study 4 - **A Study on the Environmental Impact of Generative AI**

Read **the full report, recommendations** and **case studies** by the PNAI Sub-group on Environmental Sustainability within Generative AI Value Chain **in PART 2**

HIGHLIGHTS OF PNAI WORK IN 2024:

Legal, technical and data interoperability: gaps and effective instruments in current AI Governance landscape

A concerted effort in governing AI is vital to harness the opportunities while managing the risks that AI brings, **interoperable AI systems and interoperable AI governance frameworks become imperative**. While interoperability is often understood as the ability of different systems to communicate and work seamlessly together, our definition of interoperability is broader and includes the **ways through which different initiatives, including laws, regulations, policies, codes, standards, etc to regulate and govern AI across the world could work together** to become more effective and impactful.

An assessment of AI policies, strategies, frameworks, guidelines, principles, standards, and regulations implemented in 2024 at both national and international levels reveals a growing focus on interoperability and international cooperation to address the challenges posed by AI. These policies and efforts focus on various areas, such as: exchanging standards across standards bodies, establishing frameworks for AI training data, ensuring AI safety, protecting personal data, facilitating cross-border data transfers, mitigating existing and emerging risks, and imposing transparency obligations on AI developers and deployers. There are many regional and multilateral AI governance frameworks, such as the European Union's AI Act or the Global Digital Compact, but there is no comprehensive global interoperability framework to coordinate the different AI governance frameworks. Significant gaps remain in achieving effective interoperability in AI governance. These include the absence of a globally accepted mechanism to coordinate regional and multilateral efforts, inconsistencies in AI interoperability frameworks, limited input from the Global South, and a lack of coordination and consensus on values, principles, and objectives for AI regulation.

The analysis indicated increasing interoperability requires addressing three areas: **Legal interoperability**, which ensures that AI systems operating under different regulatory frameworks, policies and strategies can work together. Increasing legal interoperability enables different frameworks to coexist and communicate with one another, reduces regulatory friction between jurisdictions, advances common policy goals, and balances global integration with domestic regulatory autonomy. **Interoperability among technical standards** focuses on ensuring AI systems can communicate and work together across jurisdictions and sectors by adopting uniform standards across software, hardware components, and platforms. **Data and Privacy Interoperability** facilitates efficient data sharing and collaboration. For example, adopting shared privacy standards and principles, common data formats, metadata standards, and data governance models.

A cohesive global AI governance framework requires concrete mechanisms for regulatory, governance, technical, and data interoperability to overcome existing barriers and tensions. The already existing effective interoperability instruments should be the starting point for lessening barriers and tensions that hinder interoperability efforts in each area. Global

multistakeholder cooperation and input are crucial for promoting inclusive governance frameworks and coordinating AI interoperability efforts across different regions and parts of the world. It is critical that global AI governance frameworks encourage interoperability to promote a safe, secure, fair, and innovative AI ecosystem. Strengthening international cooperation and focusing on shared goals will be vital as we build an interoperable, safe, and sustainable global AI ecosystem.

The following table depicts three aspects of AI interoperability: Effective instruments, barriers, and tensions. Recommendations are focused on pathways to address these aspects.

Legal Interoperability	Interoperability Among Technical Standards	Data and Privacy Interoperability
<p><i>Effective Instruments</i></p> <ul style="list-style-type: none"> Regional and international frameworks Unified AI regulators Collaboration in AI safety Governance The multilateral resolution of the UN General Assembly <p><i>Barriers</i></p> <ul style="list-style-type: none"> Regulatory fragmentation and divergent requirements Inadequate multistakeholder involvement Lack of details on implementing interoperability <p><i>Tensions</i></p> <ul style="list-style-type: none"> Differences in AI governance maturity level Differences in nature of enforcement Differences in regulatory approaches Differences in risk categorization The tension between Global cooperation and local autonomy 	<p><i>Effective Instruments</i></p> <ul style="list-style-type: none"> International collaboration Regional and National Variations Technical Industry Self-Regulation and Technical Integration <p><i>Barriers and Tensions</i></p> <ul style="list-style-type: none"> Absence of widely adopted standards and shared governance frameworks for AI interoperability Inconsistencies in the adoption of AI standards across regions Disparity between top-down and bottom-up models of AI standard frameworks Difference between binding and non-binding standards Unequal Distribution of AI technology between countries and regions 	<p><i>Tensions</i></p> <ul style="list-style-type: none"> Operational Burden of Data Compliance Absence of Data Protection Laws in countries and regions Disproportionate Influence of AI Powerhouses Siloed Data and Resource Limitations Fragmented Global Security Standards and Geopolitical Tensions

SELECTED RECOMMENDATIONS

- **Define priority of interoperability needs at the global level.** Given the heterogeneous nature of AI development, there is a need to agree which interoperability issues need (and need not) to be addressed on the global level.
- **Establish compatibility mechanisms.** Respect regional diversity in AI governance, establishing compatibility mechanisms can help to reconcile divergence in regulation.
- **Meet Local Needs, Establish Cross-Regional Partnerships, and Interlink Them Globally.** Ensure that AI interoperability frameworks are inclusive, adaptable, and capable of addressing specific local challenges while coordinating regional initiatives on a global scale. The UN should collaborate closely with regional bodies, particularly those in the Global South, to develop interoperable mechanisms that foster regional cooperation, mitigate existing disparities, and align efforts at the global level.
- **Commit to diverse and open global multistakeholder engagement in all processes to develop and adopt AI ethics, regulations, and standards** in all global platforms. Strengthen and fully utilize the Internet Governance Forum and its multistakeholder structures and mechanisms as a platform for discussion on the implementation, monitoring and follow-up of the Global Digital Compact in collaboration with all UN agencies active in AI governance.
- Develop **Semantic Interoperability**, this involves a common understanding of key legal definitions and concepts as well as the meaning, intent, nuances, and context of data and actions.
- Create safe, secure, controlled **Cross-Border AI Testing and Simulation Environments** for simulating AI deployments under varying regulatory frameworks and technical standards to identify interoperability issues before real-world implementation, this can help establish best practices and ensure that AI technologies perform safely and ethically across diverse jurisdictions.
- *To strengthen legal interoperability* - Promote the **use of global and international regulatory principles in bilateral, regional, and multilateral agreements**, ensure that local regulations can adapt to cross-border challenges and opportunities, and take international solutions into account. Increase **international regulatory cooperation and develop global standards for categorizing AI risks** across jurisdictions to jointly define risk levels for different types of AI systems.

- *To strengthen technical interoperability* - Promote **global alignment on AI standards and use of AI technologies in interoperability initiatives**, for example, to standardize, clean, and structure data.
- *To strengthen data and privacy interoperability*, develop a **global data framework for sharing AI training data** and create an **international data commons for AI research**, where countries can share anonymized, sector-specific datasets (for example in healthcare or transportation) under secure conditions.

Read the **full analysis**, highlights of **2024 AI governance and interoperability developments in different parts of the world** and **full list of recommendations** by the PNAI Sub-group on AI governance, Interoperability, and Good practices in **PART 2**

HIGHLIGHTS OF PNAI WORK IN 2024:

Promoting Multistakeholder Dialogue on Artificial Intelligence Related Labour Issues

As artificial intelligence (AI) evolves and becomes a fundamental part of modern society, it holds both opportunities and challenges for the workforce all over the world. AI's impact on labour and employment is of critical concern. We highlight the importance of workers-led AI governance, that promotes workers' rights in the AI era as well as innovation and productivity.

AI has the potential to boost worker productivity and competitiveness, create new roles and new career paths, empower education and reskilling for workers. AI can be used to address inequalities in the workplace, for example by reducing harmful biases in recruiting processes. AI systems could help integrate traditionally excluded populations into the workforce (for example speech to text AI tools enabling differently abled persons to participate in the labour market) and empower Diversity, Equity and Inclusion (DEI) processes. AI holds the potential to accelerate the achievement of Sustainable Development Goals (SDGs), as well as create jobs in emerging sectors such as renewable energy.

The transformative capabilities of AI and its capacity to complement or substitute tasks previously handled by humans raise concerns of job loss and decrease of income, reskilling or upskilling large parts of the workforce around the world as the use of AI proliferates in the workspace. There are also mental health effects due to job insecurity and stress from reskilling. Issues also arise regarding the role of AI oversight over workers and guaranteeing workers' rights. For example, algorithmic deployment to manage and gather real-time data of workers in their daily tasks or the increasing use of AI-powered Applicants Tracking Systems (ATS) in resume screening. Wage polarization between workers and inequitable geographical distribution of AI capacities that aggravate the productivity differences between Global North and Global South. Workers in countries of the Global South are increasingly engaged in data work such as data labelling, cleaning, moderation, and tending to the ever-increasing demand for training data for AI systems.

Unions have the potential to play a pivotal role in retraining workers by leveraging their bargaining power and involvement in the processes of technological adaptation across various economic sectors, as well as in the internal management of personnel within organizations. Trade unions themselves need to follow the AI development landscape closely, build understanding of AI to ensure they have the resources and internal AI expertise in their own organizations. Particularly, unions have the potential to play a role in the design, implementation, use and integration of technologies at the workplace.

Even though there is not a globally binding agreement on AI in the workplace, we recognize advances made in countries and regions around the world. This includes, for example, the European Union AI Act, the AI Principles for developers and employers established by the US Department of Labor, and China's Provisions on the Management of Algorithmic

Recommendations in China. Internet Information Services that include provisions for content control and provides protections for workers impacted by algorithms.

KEY RECOMMENDATIONS

- **Establish frameworks** that enable workers and trade unions to actively engage in AI decision-making processes at the national, regional, and multilateral levels.
- **Ethical Frameworks:** Organizations should create codes of conduct that outline responsibilities and accountability for both workers and management in AI usage.
- **Incorporate Worker Feedback in designing AI systems:** Involve workers in the design and testing phases of AI systems that they will interact with in their work. Incorporating this feedback would also help to improve the efficiency and usability of AI systems developed for the workplace.
- Establish **Joint Committees** with equal representation of workers and management to oversee AI integration in organizations, addressing concerns related to labour issues, and encourage **Sectoral Open Dialogue** forums for workers to voice concerns and suggestions about AI use in different sectors.
- **Strengthen governance frameworks:** Develop comprehensive, human-centred, international AI governance standards that include clear ethical guidelines and labour protections to apply to algorithmic management and the protection of workers' personal data. Promote AI transparency and accountability to ensure that workers are aware and understand how AI affects their employment.
- **Mainstream AI use in safeguarding the workers' rights:** Develop guidelines for using AI to address intersectional issues (gender, religious, and cultural context) in workplaces.
- Promote the **development of global and uniform standards** for monitoring AI's impact on the labour market.

Read the **full report and detailed list of recommendations** by the PNAI Sub-group on Sub-group on Labour issues throughout AI's life cycle in **PART 2**

PNAI Policy Brief 2024

PART 2

**DECEMBER
2024** | **INTERNET GOVERNANCE FORUM
POLICY NETWORK ON ARTIFICIAL
INTELLIGENCE**

Liability as a Policy Lever in AI Governance: Amplifying the Global Discussion

Policy Network on Artificial Intelligence (PNAI)
Sub-group on liability as a mechanism for supporting AI accountability

1. Introduction

Policymakers and stakeholders focused on AI governance consistently highlight the need to hold AI developers and deployers accountable for the harms caused across the AI lifecycle.¹ Yet, to date, mechanisms to ensure accountability remain poorly defined.² Many of the accountability frameworks under consideration in various jurisdictions and through intergovernmental agencies – for example, algorithmic impact assessments and audit-based monitoring – will have difficulty keeping pace with the quickly morphing risks that will inevitably accompany AI's evolution. Frameworks aimed at promoting AI safeguards will need to evolve rapidly to address dynamic risks – a particularly tricky proposition for quickly and autonomously changing systems.³ Nevertheless, policymakers have an opportunity to stake a resilient approach to AI accountability, and to incentivize responsible AI development processes and outcomes, by establishing clear guidelines regarding **legal liability for harms**.⁴ Liability can be a critical lever in mitigating AI-related risks ranging from algorithmic bias leading to discriminatory outcomes in hiring or lending to AI-driven misinformation campaigns that can destabilize democracies, to malfunctions in AI-controlled critical infrastructure. This can jeopardize public safety.⁵

With this discussion paper, we hope to amplify the conversation about AI liability within the **global AI governance community**. While attention to AI liability has been prominent in the European Union for several years,⁶ a globally coordinated approach to AI liability principles will be necessary

¹ OECD, [Advancing accountability in AI](#), Feb. 2023

² G. Noto La Diega & L.C.T. Bezerra, [Can there be responsible AI without AI liability? Incentivizing generative AI safety through ex-post tort liability under the EU AI liability directive](#), Sept. 2024

³ Ibid.

⁴ H. Zech, [Liability for AI: Public Policy Considerations](#), Jan. 2021

⁵ C. Wendehorst, [Liability for Artificial Intelligence: The Need to Address Both Safety Risks and Fundamental Rights Risks](#), 2022

⁶ C. Novelli, F. Casolari., P. Hacker, G. Spedicato & L. Floridi, [Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity](#), March 2024

to incentivize safeguards across the digital divide, protecting individuals and communities worldwide from potential AI-related harms.⁷

Liability frameworks can fill critical gaps in AI governance, as they have an inherent capability of keeping pace with evolving risks. Unlike auditing systems which evaluate foreseeable and pre-specified categories of harm, liability frameworks can **adaptively respond to any type of damage that emerges** – whether anticipated or novel – and align incentives with responsible development and deployment, providing a path to accountability and recourse for damages⁸ Rather than advocating for a specific framework, we highlight scholarship and existing policy work, in the hopes of amplifying a discussion of **AI liability principles** among global policymakers to guide jurisdictions in responding to the unique complexities that accompany these determinations.

2. Liability and AI Harms: Unique and Urgent Challenges

Liability refers to the legal and financial responsibility for harm or damage caused by AI systems, encompassing obligations from developers and deployers to compensate affected parties for losses resulting across the AI lifecycle. Such harms can be incurred within the AI-training phase (e.g., web-scraping to train AI systems in a manner that sweeps up personal/private data or intellectual property) or about harmful AI outputs. Affected parties might include individuals, private organizations, or public entities.⁹

The universe of potential AI-related harms ranges from facial recognition systems leading to wrongful arrests, to financial panic caused by faulty AI-driven market analyses, to discriminatory hiring based on AI-enabled human resource applications, to hazards caused by AI-led oversight of physical infrastructure such as water supply systems. **The need to incentivize AI developers and deployers** to proactively safeguard against such potential harms is both clear and immediate. Liability frameworks have helped create powerful economic incentives for innovative safeguards and risk mitigation strategies across industries as varied as food and beverage (e.g. improved food traceability and labeling for allergens), automotive (seatbelts, airbags, advanced driver assistance systems), and pharmaceutical industries (improved drug labeling and warning systems).¹⁰ As is the case for these industries, liability frameworks pertaining to AI-related harm can be a critical component of a broader regulatory framework and a force for safety innovations.

As noted, liability frameworks offer **unique advantages** in AI governance.¹¹ As new risks emerge with advancing AI technologies, liability frameworks can naturally adapt to address these harms

⁷ UN AI Advisory Body, [Governing AI for Humanity](#), Sept. 2024

⁸ G. Weil, [Tort Law as a Tool for Mitigating Catastrophic Risk from Artificial Intelligence](#), June, 2024

⁹ UN AI Advisory Body, [Governing AI for Humanity](#), Sept. 2024

¹⁰ C.M. Sharkey, [The Irresistible Simplicity of Preventing Harm](#), July, 2023

¹¹ G. Weil, [Tort Law as a Tool for Mitigating Catastrophic Risk from Artificial Intelligence](#), June, 2024

without requiring constant regulatory updates. This inherent scalability will make liability an indispensable tool in mitigating risks over time.¹² However, legal experts argue that **traditional liability frameworks are not fit-for-purpose** for AI systems.¹³ As these systems' capacities progress autonomously, models produce outcomes that are not fully predictable – even to their creators. This autonomy blurs conventional lines of responsibility, complicating efforts to assign liability. The complexity will be further compounded by the web of contractual obligations and varied risk management approaches within the AI ecosystem.

Amid these built-in complexities, ***policymakers and stakeholders have yet to come to a consensus regarding key questions on assigning liability within the AI lifecycle.***¹⁴ For example, if an AI-powered medical diagnostic system misses a critical, treatable condition due to underlying biases in its training data, should liability be assigned to the healthcare provider, the medical application developer, to the underlying foundational system, or apportioned to some degree across these players?¹⁵

Given the countless, thorny, similar questions that will arise, it will be vitally constructive for global AI governance stakeholders to coordinate regarding **AI liability principles and standards.**¹⁶ Neither the opacity nor the autonomy of AI systems ought to exempt developers and deployers from accountability. The lack of transparency about these systems elevates the need to incentivize rigorous safeguarding, in part through ensuring companies will be held responsible for harm via clear liability frameworks. Without such mechanisms, damages caused by AI systems will be borne by faultless individuals, communities, and the public at large.¹⁷

2.1 Types of Liability

Our subgroup's work and this discussion paper focus on AI **product liability** and **civil liability**, leaving **criminal liability** out of scope.¹⁸ Product liability is a legal concept that would hold developers and deployers responsible for harm caused by defects in the AI products or services they have made available to the public. Civil liability represents a broader legal category allowing not only individuals and organizations but also states and governments to seek remuneration for harms to protect public interests or recover damages on behalf of their citizens.¹⁹ While product

¹² M.H. Pfeiffer, [First Do No Harm: Algorithms, AI, and Digital Product Liability](#), Sept. 2023

¹³ H. Zech, [Liability for AI: Public Policy Considerations](#), Jan. 2021

¹⁴ Center for Humane Technology, [A Framework for Incentivizing Responsible Artificial Intelligence Development and Use](#), Sept. 2024

¹⁵ W.N. Price II, S. Gerke, G. Cohen, [Potential Liability for Physicians Using Artificial Intelligence](#), 2019

¹⁶ C. Frattone, [Reasonable AI and Other Creatures. What Role for AI Standards in Liability Litigation?](#), 2022

¹⁷ UN AI Advisory Body, [Governing AI for Humanity](#), Sept. 2024

¹⁸ As a multistakeholder group representing diverse nations with vastly different criminal justice systems, the focus of our shared discussion has been on financial penalties versus *criminal liability*. While out of scope for this paper, criminal liability represents another potential avenue for addressing flagrant misconduct.

¹⁹ European Parliamentary Research Service, [Proposal for directive on adapting non-contractual civil liability rules to artificial intelligence](#), Sept. 2024

liability generally operates on a *strict liability* standard – requiring only proof of defect and resulting harm without a need to prove negligence – other civil liability mechanisms generally require proof of negligence or breach of duty.²⁰

Administrative liability refers to penalties imposed by regulatory bodies or Government agencies for non-compliance with AI regulations.²¹ For example, Article 99 of the EU AI Act²² establishes administrative fines of up to 35 million euros or 7% of global annual turnover for violation of prohibited practices including for example deploying subliminal techniques to exploit behaviour; exploiting vulnerabilities of specific groups; certain kinds of AI-enabled biometric identification systems to monitor public; failure to engage comprehensive risk management systems; failure to use high quality, validated training data which have been thoroughly examined for biases; failure to maintain sufficient transparency that allows proper evaluation of high-risk systems; failure to ensure ongoing human oversight of high-risk systems.²³

Complex questions that will need to be navigated in developing clarity about liability for AI include: Who is financially responsible when an autonomous vehicle causes an accident? How should damages be apportioned if an AI trading algorithm causes panic in financial markets? Who bears financial responsibility if a medical diagnostic system produces skewed outcomes that harm patient health?²⁴

2.2 The Rationale for Harmonizing Frameworks

Coordinating AI liability principles on a global scale would incentivize developers and deployers to adhere to consistent standards, preventing “regulatory arbitrage” and “forum shopping” by companies seeking lenient jurisdictions and leveling the playing field for deployment worldwide.²⁵ To make good on Global Digital Compact commitments to closing the digital divide, Global Majority countries that lack resources for comprehensive AI governance must benefit from the collective expertise and enforcement capabilities of the international AI governance community.²⁶ Harmonizing AI liability will help protect vulnerable individuals, communities, and Global Majority countries from bearing the brunt of AI-related harms. Additionally, the questions surrounding AI liability across the value chain are intricate, challenging to navigate, and likely to **cross national boundaries**. Harmonized liability principles can drive uniform requirements for AI/algorithmic transparency, making it easier for vulnerable individuals and communities to seek redress and recourse for harms, reducing the likelihood that those most at risk will be exploited

²⁰ H. Zech, [Liability for AI: Public Policy Considerations](#), Jan. 2021

²¹ A. Bertolini, [Artificial Intelligence and civil liability](#), Jan. 2020

²² European Union, [EU Artificial Intelligence Act, Article 9: Risk Management System](#), 2024

²³ European Union, [EU Artificial Intelligence Act, Article 16: Obligations of Providers of High-Risk Systems](#), 2024

²⁴ H. Zech, [Liability for AI: Public Policy Considerations](#), Jan. 2021

²⁵ G. Noto La Diega & L.C.T. Bezerra, [Can there be responsible AI without AI liability? Incentivizing generative AI safety through ex-post tort liability under the EU AI liability directive](#), Sept. 2024

²⁶ UN Office of the Secretary-General’s Envoy on Technology, [Global Digital Compact](#), Sept. 2024

or harmed, and increasing the likelihood the benefits of AI are shared within a safer, more inclusive market ecosystem.²⁷

2.3 AI - Specific Complexities for Liability Frameworks

The role of transparency and explainability. The opacity of AI systems, particularly large language models, poses significant challenges for any liability regime. To effectively assess potential *bias-related harms*, for example, it will be necessary to access valid indicators about the factors that contributed to decisions.²⁸ For liability frameworks to be meaningful and enforceable, jurisdictions will need access to information about how systems factored inputs and characteristics into their decisions – *algorithmic transparency* – a standard that has been thus far promised, far more generously than it has been provided. Clear, rigorous liability frameworks have the potential to incentivize compliance with algorithmic transparency commitments, enabling regulators and adjudicators to determine whether AI systems are functioning in an unbiased manner.²⁹

Liability Across the AI Lifecycle and Supply Chain. Harms can occur at various stages across the AI lifecycle, from development to deployment and ongoing use: during data collection (e.g., from the improper use of personal data and/or intellectual property), in the deployment phase, or as a result of the AI system's ongoing learning and adaptation.³⁰ These lifecycle complexities are compounded by interwoven activities within an AI supply chain involving data providers, model developers, the software companies that incorporate these AI models, practitioners, and end users – all of which can make identifying the defects that cause harm unusually difficult.³¹ Underlying biases in training data can interact with flaws in the life cycle (e.g., inadequate model training) or the supply chain to produce discriminatory outcomes.³² To prevent downstream harms, liability frameworks might implement a greater emphasis at the source – a “chain of responsibility” approach – such that a greater onus is placed on foundational players to implement robust safeguards and quality controls.³³ An emphasis on the responsibilities of foundational companies will help incentivize a more careful risk assessment of partners and providers downstream.

Adjudicating AI Liability. The “black box” nature of AI systems presents significant challenges in adjudicating liability cases, and the opacity of AI decision-making processes makes it difficult

²⁷ C. Wendehorst, [Liability for Artificial Intelligence: The Need to Address Both Safety Risks and Fundamental Rights Risks](#), 2022

²⁸ H. Zech, [Liability for AI: Public Policy Considerations](#), Jan. 2021

²⁹ G. Noto La Diega & L.C.T. Bezerra, [Can there be responsible AI without AI liability? Incentivizing generative AI safety through ex-post tort liability under the EU AI liability directive](#), Sept. 2024

³⁰ R. Ashmore, R. Calinescu & C. Paterson, [Assuring the machine learning lifecycle: Desiderata, methods, and challenges](#), 2021

³¹ S. Burton et al. [Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective](#), Feb 2020

³² H. Zech, [Liability for AI: Public Policy Considerations](#), Jan. 2021

³³ Y. Bathaee, [The Artificial Intelligence Black Box and the Failure of Intent and Causation](#), 2017

for traditional courts to properly assess fault and causation. Some legal experts have argued that specialized courts or tribunals might be necessary, equipped with the technical expertise necessary to make informed decisions on liability.³⁴ More immediately, members of any judicial system adjudicating AI harms will need to be adequately educated about this technology's unique complexities.³⁵

Indemnity, Contractual Liability, and AI. In many industries involving significant risks, businesses use contracts to allocate responsibilities and liabilities – which might include *indemnification clauses*, such that one party agrees to compensate the other for specific types of losses or damages. Indemnification and contractual liability played significant roles in the establishment of the nuclear power industry in the United States, under a federal indemnity scheme established by the Price-Anderson Act.³⁶ Internationally, the Vienna Convention on Civil Liability for Nuclear Damage created a framework combining private liability, state guarantees, and pooled industry resources.³⁷ The use of such contractual arrangements in an AI context is still evolving.³⁸

2.4 The Role of AI Standards in Mitigating Liability Risks

By adhering to rigorous industry standards for ethical and safe AI development and deployment, such as those developed by IEEE³⁹, ISO⁴⁰, or national standards bodies,⁴¹ companies can significantly reduce their liability risks. Firstly, by meeting such standards, companies demonstrate their commitment to due diligence and duty of care, which can serve as compelling evidence in liability litigation. As courts grapple with the complexities of AI-related harms, standard-setting bodies are likely to serve as guideposts for determining what constitutes reasonable care. Even more critically, by taking meaningful steps to adhere to rigorous standards regarding transparency⁴², accountability⁴³, and algorithmic bias⁴⁴, companies will actively mitigate potential harms. Yet the leverage provided by liability will be critical: robust liability regimes will create powerful

³⁴ S. Chesterman, [Artificial intelligence and the limits of legal personality](#), 2020

³⁵ T. Sourdin, [Judge v Robot?: Artificial intelligence and judicial decision-making](#), 2018

³⁶ M. Kovac, [Autonomous Artificial Intelligence and Uncontemplated Hazards: Towards the Optimal Regulatory Framework](#), 2022

³⁷ R. Trager et al., [International governance of civilian AI: A jurisdictional certification approach](#), Aug. 2023

³⁸ Hannes Claes & Maarten Herbosch, M. [Artificial Intelligence and Contractual Liability Limitations: A Natural Combination?](#), 2023

³⁹ IEEE Standards Association, [The Ethics Certification Program for Autonomous Intelligent Systems \(ECPAIS\)](#), Retrieved September, 2024

⁴⁰ ISO, [The International Organization for Standardization, ISO/IEC 42001: 2023, Information Technology - Artificial Intelligence Management System](#), 2023

⁴¹ NIST, [Artificial Intelligence Risk Management Profile: Generative Artificial Intelligence Profile](#), July 2024

⁴² IEEE [Ontological Specification for Ethical Transparency](#)

⁴³ IEEE [Ontological Specification for Ethical Accountability](#)

⁴⁴ IEEE [Ontological Specification for Ethical Algorithmic Bias](#)

incentive structures, encouraging companies to adopt and implement industry standards while providing a framework for accountability when those standards are not met.⁴⁵

2.5 International Frameworks for Cross-Border AI Liability

International AI governance frameworks have the potential to clarify and even enforce liability decisions and will be particularly useful in cross-border issues as well as to nations with limited resources to manage complex AI liability regimes.⁴⁶ As the first legally binding international treaty covering AI systems across the lifecycle, the Council of Europe's *Framework Convention on Artificial Intelligence and Human Rights, Democracy, and the Rule of Law* offers a promising starting point in its provisions for remedies for harms (Article 14) and international cooperation (Article 25).⁴⁷ Relevant international models with governance and enforcement capabilities for other industries include the *Paris Convention on Third Party Liability in the Field of Nuclear Energy*⁴⁸ and the *Montreal Convention* negotiated by the International Air Transport Association (IATA), which created a "universal liability regime for international carriage by air."⁴⁹

3. A Global View: Existing AI Liability Policy Across Jurisdictions

This chapter presents an overview of AI liability developments in different countries and regions, assembled by our international team to assess the state of play. In evaluating current developments and initiatives in different parts of the world, our researchers confirmed that this critical governance conversation has been most fully developed in the European Union – with Brazil as an additional exception with its draft bill on AI governance featuring liability levers – while the majority of countries and regions are in the early stages in their attention to the topic.

African Union

To date, the African Union (AU) and African nations have not focused on frameworks for addressing liability for harms related to AI and digital technologies. The African Union AI Strategy⁵⁰ and the AU Digital Compact⁵¹ emphasize the need for accountability to protect consumers and promote ethical AI practices. They encourage member states to contemplate the ethical ramifications and legal obligations of AI technologies, but neither framework highlights liability as

⁴⁵ C. Frattone, [Reasonable AI and Other Creatures. What Role for AI Standards in Liability Litigation?](#), 2022

⁴⁶ G. Noto La Diega & L.C.T. Bezerra, [Can there be responsible AI without AI liability? Incentivizing generative AI safety through ex-post tort liability under the EU AI liability directive](#), Sept. 2024

⁴⁷ Council of Europe, [Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law](#), Sept. 2024

⁴⁸ Nuclear Energy Agency, [Paris Convention on Third Party Liability in the Field of Nuclear Energy](#), revised Feb. 2024

⁴⁹ International Air Transport Association, [A Universal Liability Regime for International Carriage by Air](#), 1999.

⁵⁰ African Union, [Continental Artificial Intelligence Strategy](#), July 2024

⁵¹ African Union, [African Digital Compact](#), August, 2024

a governance tool to account for the impact of AI on consumers. The legal structures governing consumer and product liability differ significantly from one country to another, with some nations having consumer protection laws that encompass elements of product liability and others significantly lacking in these protections.

The African Union's Continental AI Strategy⁵² highlights and emphasizes the importance of ensuring responsible AI use, particularly when addressing fairness and accountability in decision-making. The strategy calls for regulatory frameworks that can address biases, ensure inclusivity, and hold the right stakeholders accountable—whether that be developers, service providers, or financial institutions. However, while the AU highlights the importance of consumer protection in AI, there is a critical accountability gap, as no framework has been established yet.

Meanwhile, AI-related risks are growing rapidly. AI-driven lending algorithms in some countries promise greater financial inclusion, yet may inadvertently exacerbate inequalities as they are built on biased, historical data.⁵³ As the training data for the lending algorithm heavily reflects urban, male users who are more digitally active, people from rural areas with limited digital footprints or less access to mobile technology may be deemed less creditworthy, even if they have a history of responsible financial behaviour. This kind of bias can deepen financial exclusion and perpetuate inequalities, as marginalized groups, including women and members of rural communities, may be less likely to receive loans or other financial services. Questions of liability arise regarding who is responsible for correcting these errors—the AI developer, the service provider, or the financial institutions. Deepfakes are increasingly prevalent in some African countries and present an additional category of AI-promoted risk. Without any mechanism for recourse or accountability, these threaten not only defamation but also have the potential to incite social unrest and disrupt political stability.

Thus, the need for harmonization of existing legal approaches across the African continent remains an urgent priority. The primary challenge in developing an effective liability framework is achieving uniformity in laws across different jurisdictions, alongside establishing robust mechanisms for implementation, compliance, and enforcement.

Australia

The Australian Government is in the midst of a public, comprehensive consultation in the effort to provide effective governance, and “best practice for safety”.⁵⁴ While Australia’s current AI Safety Standards are voluntary, the Department of Industry, Science and Resources has developed a draft document on “mandatory guardrails for AI in high-risk settings”⁵⁵ which notes, “some AI

⁵² African Union, [Continental Artificial Intelligence Strategy](#), July 2024

⁵³ B. E. Abikoye & C. Agorbia-Atta, [How artificial intelligence and machine learning are transforming credit risk prediction in the financial sector](#), 2024

⁵⁴ Australian Government, [Promoting safe and responsible AI](#), retrieved Sept. 2024

⁵⁵ Australian Government, [Introducing mandatory guardrails for AI in high-risk settings](#), Sept 2024

characteristics are limiting the ability of existing laws to effectively prevent or mitigate risks. Examples include clarifying accountability and ensuring legal responsibility is distributed appropriately to developers and deployers best placed to manage the causes of potential harms from AI decisions and applications, particularly as many existing laws were originally drafted on the presumption that humans are taking actions and making decisions.”

Extrapolating from another Australian industry, the global AI governance community might take inspiration to help in assigning liability for AI harms from Australia's "chain of responsibility" model within its Heavy Vehicle National Law.⁵⁶ Under this model, each party in the value chain is responsible for ensuring that the next party can meet established safety and quality standards. The clarity of this framework might help to lighten some of the complexities surrounding AI harms, establishing a duty of care such that each entity in the AI system's lifecycle takes responsibility for verifying the capabilities and standards of the next.

Bangladesh

Bangladesh faces a stark digital divide, with a significant percentage of the population lacking access to the internet.⁵⁷ The government data from Bangladesh Sample Vital Statistics shows that the prevalence of internet usage among the rural population is around 37 percent and it is around 54 percent among urban population, implying a gap of 17 percent. Similarly, it finds that such a gap also persists between males and females by around 13 percent. This stark digital divide has far-reaching implications for Bangladesh's development, limiting access to information, education, and economic opportunities, and exacerbating existing profound inequalities. Addressing the digital divide is crucial for Bangladesh's progress and its ability to harness the potential of AI. The lack of internet connectivity also poses challenges for enforcing AI-related regulations, as it can make it difficult to monitor and regulate AI activities in remote areas.

Any AI liability framework in Bangladesh must be coupled with initiatives to bridge the digital divide, including investments in digital infrastructure, promotion of digital literacy, and efforts to make internet access more affordable and widespread.

Brazil

Brazil's draft bill on AI governance, PL 2338/2023, establishes a clear framework for civil liability related to AI systems.⁵⁸ For high-risk or excessive-risk AI systems, the bill specifies that suppliers or operators are strictly liable for damages caused, to the extent of their participation in the damage, regardless of the system's degree of autonomy. For AI systems not classified as high-risk, the bill establishes a presumption of fault, with the burden of proof shifted in favour of the

⁵⁶ B. Walker-Munro & Z. Assaad, Z, [The Guilty \(Silicon\) Mind: Blameworthiness and Liability in Human-Machine Teaming](#), Oct. 2022

⁵⁷ Khawaja Sazzad Ali & Anisur R. Faroque, [Addressing the Complexity of the Digital Divide and the Role of Government in Addressing It: Role of Government in Bridging the Digital Divide](#), 2023.

⁵⁸ Brazil, [Bill 2338/2023](#), 2023

victim. There are exemptions through which AI actors may not be held liable, such as when they can prove they did not deploy the AI system or when damage results exclusively from a victim or third-party action. These provisions create a comprehensive liability framework across the AI lifecycle, with stricter standards for high-risk systems and maintaining protections for consumers.

China

In July 2023, the Cyberspace Administration of China (CAC), along with six other Chinese regulators, jointly issued Interim Measures for the Management of Generative AI Services⁵⁹ reflecting feedback from different stakeholders on previously released draft measures and setting out the rights and responsibilities of providers and users of AI. Together, these measures establish compliance requirements for generative AI service providers, including obligations related to data sourcing, intellectual property rights, personal information protection, and content accuracy. Service providers must ensure the legitimacy of their data sources, obtain consent for using personal information, and take measures to improve training data quality. The framework also mandates labelling of AI-generated content, particularly for “deep synthesis” services, and requires providers to prevent the generation or transmission of illegal content. Violations of these obligations can result in administrative or criminal penalties, effectively creating a form of administrative liability for AI service providers.⁶⁰

Hong Kong SAR authorities have actively sought changes to update copyright law to bolster AI development to keep pace with AI developments as the city aims to become a regional IP trading center.⁶¹ The bureau added that it has reviewed the relevant legislation in Hong Kong and other jurisdictions, as well as the prevailing market situation.

European Union

The European Union has invested considerable study in confronting the complexities of AI liability as a lever of AI governance. By revising its *Product Liability Directive* (PLD), the EU has explicitly begun to address harms caused by AI software. PLD, for example, lowers the burden of proof, to allow for redress for harms created by opaque and autonomous AI systems⁶². By clarifying that software falls within the scope of 'product' and extending liability to cases of cybersecurity vulnerabilities, the revised PLD creates a more comprehensive framework for addressing AI-related harms. This approach not only incentivizes developers and deployers to adhere to consistent standards but also prevents regulatory arbitrage across different regions. Importantly,

⁵⁹ PwC: Tiang and Partners, [Regulatory and legislation: China's Interim Measure for the Management of Gen AI Services](#), August, 2023

⁶⁰ Ibid.

⁶¹ The Government of Hong Kong S.A. Region of China Intellectual Property Department, [Public Consultation on Copyright and Artificial Intelligence](#), July 2024

⁶² European Parliament, [New Product Liability Directive - O4 2020](#). Sept, 2024

PLD alleviates the burden of proof for victims and extends compensable damage to include psychological harm and data loss, making it easier for affected individuals to seek redress.⁶³

The EU's standalone *AI Liability Directive* (AILD) has lingered in the proposal stage. A recent study by the European Parliamentary Research Service suggested that the AILD should be broadened to encompass a more comprehensive software liability framework, "to prevent market fragmentation and enhance clarity across the EU".⁶⁴ The study recommended a mixed framework: for AI systems that have been legally banned under the AI Act, strict liability should be assumed for damages caused; elsewhere, the strict liability standard was recommended for high-risk AI systems causing "illegitimate" harms.⁶⁵ The EPRS recommended expanding the scope of the AILD so it covers not only "high-risk" but also "high-impact" AI systems to encompass general-purpose AI, autonomous vehicles, and other applications not classified as high-risk under the AI Act. The study calls for more explicit liability coverage for AI discrimination cases; closer attention to liability for built-in biases, privacy, and intellectual property violations in general purpose AI systems, and greater harmonization of definitions with the already ratified AI Act.

Assigning liability across the AI value chain is challenging. The EPRS endorses further study of three policy options: 1) Presumption of an equal share of liability; 2) Exempting or protecting small and medium enterprises (SMEs) from the more rigorous liability expectations; 3) Protecting downstream parties such that upstream actors (particularly those with highly dominant market positions) are deemed more responsible for harms and for providing financial recourse.⁶⁶

India

While India has not yet enacted AI-specific regulations, existing legal frameworks provide some basis for addressing AI-related liability. The Information Technology Act, 2000, which establishes liability for content on websites, could potentially extend to AI service providers – holding them responsible for content available through their platforms. Additionally, the Digital Personal Data Protection Act, 2023, introduces liability for the misuse of personal data, which could apply to AI systems processing such information. While not explicitly targeting AI, these laws create a framework where AI developers and deployers could be held liable for harmful or unlawful outcomes in areas of content moderation and data protection.

⁶³ Ibid.

⁶⁴ European Parliamentary Research Service, [Proposal for a directive on adapting non-contractual civil liability rules to Artificial Intelligence](#), Sept. 2024

⁶⁵ As distinguished from "legitimate harm" models which might result in an individual rightfully being excluded from an award or benefit

⁶⁶ European Parliamentary Research Service, [Proposal for a directive on adapting non-contractual civil liability rules to Artificial Intelligence](#), Sept. 2024

Indonesia

In 2020, Indonesia reached a milestone in formally recognizing AI as a distinct business sector⁶⁷ via *The Indonesian National Strategy on Artificial Intelligence*.⁶⁸ The strategy designated the Ministry of Communication and Informatics to formulate ethical guidelines for AI. While further regulations are anticipated,⁶⁹ Indonesia has not yet established specific regulations overseeing AI. However, several existing legal frameworks can be leveraged for this purpose, including the Personal Data Protection Bill.⁷⁰ The liability mechanisms under this framework are divided into two parts: first is a criminal accountability mechanism, which applies solely to individuals deliberately engaging in acts intended to breach and/or misuse personal data. Breaches resulting, instead, from negligence are subject exclusively to administrative sanctions, with ambiguous remedy mechanisms. To date, this law does not comprehensively address accountability measures for potential violations and abuses carried out by the State.

Iran (Islamic Republic of)

According to Iran's *Civil Liability Law*,⁷¹ anyone who, without legal authorization, intentionally or negligently causes harm to another person's life, health, property, freedom, reputation, commercial reputation, or any other right established by law, resulting in material or moral damage, is responsible for compensating the damage caused. In cases where the harmful act has caused material or moral damage to the injured party, the court, after investigation and proof of the matter, will order the perpetrator to compensate for the damages. If the harmful act has caused only one type of damage, the court will order the perpetrator to compensate for that specific type of damage. So, "anyone" in this law could be interpreted as "any AI machines".

The National Policy of AI of I.R. Iran⁷² established a set of ethical principles that will serve as guidance in the responsible and value-based development and use of AI technology, based on Islamic values. These principles are observed by professionals and others involved in the design, production, and utilization of AI, creating mutual rights. Examples of AI ethical issues include respecting privacy, upholding individual and social rights, ensuring social security, fairness, explainability, transparency, non-discrimination and bias, accountability, alignment with the values and norms of Islamic society, responsibility, trust, and preventing misuse of technology. The goal of AI ethics is to optimize the beneficial impact of AI on society and human life while

⁶⁷ This recognition was actualised by introducing Ministerial Regulation No. 3/2021 issuance by the Ministry of Communication and Information Technology, designating it as the governing body for emerging technologies such as AI, Blockchain, and IoT. The implementation of Government Regulation No. 5/2021 further solidified this recognition.

⁶⁸ Stranas AI, [Indonesian National Strategy on Artificial Intelligence](#), 2020.

⁶⁹ Government Regulation Number 71 the Year 2019, regarding implementing Electronic Systems and Transactions, or cloud computing and its procurement regulation

⁷⁰ The [PDP framework](#) encompasses notice and consent mechanisms, the right to be forgotten, transparency and documentation requirements, and emphasizes special considerations for children and individuals with disabilities. The PDP is complemented by [consumer protection law](#) and [human rights law](#) to address any privacy or human rights breaches.

⁷¹ Islamic Republic of Iran, [Civil Code of the Islamic Republic of Iran](#), amended in 1982

⁷² Center for AI and Digital Policy (CAIDP), [Artificial Intelligence and Democratic Values 2023: Iran](#), April, 2024

reducing the risks and unintended consequences of its use, based on Islamic values and beliefs. Article 5 of the national policy underscores attention to justice, dignity, rights, and the physical, mental, and psychological well-being of individuals in the mechanisms of training and utilizing artificial intelligence.⁷³

Japan

In February 2024, the ruling party of Japan issued an AI white paper that proposed an AI Basic Law in February 2024 promoting AI safety.⁷⁴ However, the proposal stresses voluntary measures (soft law), applying hard law to extreme risks presented by high-risk AI. It proposes the establishment of an AI Safety Institute (AISI) as being key to addressing harm by AI. The AISI will undertake the following measures: investigations, standards, creation, developing human talent to address AI safety, fostering third-party certifications and international harmonization. The proposed policy is aimed at promoting Japan as the “world’s most AI-friendly country.” The white paper does not refer to liability laws or frameworks.

Pakistan

As of 2023, Pakistan has completed its Personal Data Protection Bill (PDPB), which is set to serve as the main legal framework for data protection in the country.⁷⁵ Although the bill does not specifically target AI, it contains provisions relevant to AI systems, especially those that manage personal data. Developed by the Ministry of Information Technology and Telecommunication, the bill is now pending approval from the Cabinet and Parliament. Once it becomes law, it will impose significant responsibilities on data controllers and processors, including those using AI technologies.

Alongside the PDPB, Pakistan is investigating the implications of artificial intelligence through various initiatives and discussions focused on AI ethics and governance.⁷⁶ While there is currently no formal AI liability framework, Pakistan's approach to data protection may increasingly align with international standards, such as the GDPR, particularly as the government aims to improve its technological environment. Involving a range of stakeholders, including tech companies and civil society, will be crucial for developing a balanced and effective regulatory framework.

Singapore

In 2020 and 2021, the Singapore Academy of Law issued two reports on AI liability: “Report on the Attribution of Civil Liability for Accidents Involving Autonomous Cars”; and “Report on Criminal Liability, Robotics, and AI Systems.”⁷⁷ These propose that for intentional AI harms, existing laws

⁷³ Tehran Times, [Govt starts implementing national document on AI development](#), July 24, 2024

⁷⁴ LDP Japan, [AI White Paper 2024](#), April, 2024

⁷⁵ Pakistan Ministry of Information Technology & Telecommunication, [Draft of the Personal Data Protection Bill](#), 2023

⁷⁶ International Bar Association, [Pakistan's Draft National AI Policy: fostering responsible adoption and economic transformation](#), July 2023

⁷⁷ Singapore, [Report Series: The Impact of Robotics and Artificial Intelligence on the Law](#), 2021

could be amended to be fit-for-purpose. For civil harms that are non-intentional, the report notes several potential approaches: framing AI systems as legal personalities (such as with corporations or nation-states); creating a new category of legal offence for computer programs that commit harms; applying workplace safety legislation as a model – imposing liability on specified entities as determined within a chain of responsibility.

South Korea

The South Korean legislature proposed a bill focused on AI liability in February 2023.⁷⁸ The proposed law would hold high-risk AI business operators liable for damages caused to users when they violate the obligations of the Act. The obligations include risk assessments, user notifications, and human oversight for both developers and deployers. The bill includes exemptions for defects causing harm that could not be anticipated given the current state of science. The law would establish an “Artificial Intelligence Dispute Mediation Committee” to handle liability disputes and compensation claims. The bill also promotes insurance coverage for high-risk AI businesses to balance accountability with support for emerging technologies.

United Arab Emirates

In July 2024, the UAE Government issued *The UAE Charter for the Development & Use of Artificial Intelligence*,⁷⁹ which includes 12 principles, emphasizing the importance of governance and accountability in AI to ensure the technology is used ethically and transparently. Furthermore, in October 2024, the UAE government announced an official national stance related to the global governance of AI, in which it emphasizes the “critical importance of transparency and establishment of checkpoints within AI tools, enabling governments to ensure compliance with ethical standards and to place the necessary accountability mechanisms to address any potential violations”.⁸⁰ This national stance is articulated as an official foreign policy approach. Overall, the U.A.E. is seeking to become a global hub for AI in compliance with its AI strategy 2031. However, neither of its foundational AI policy documents explicitly addresses AI liability.^{81 82}

United States

Liability laws in the U.S. have not been updated to date to address harms from AI and algorithmic systems. However, there has been increased discussion in the U.S. tech policy world regarding the potential for liability laws to play a critical role in AI governance. An influential US tech policy think tank, the Center for Humane Technology, recently published a proposal which placed liability

⁷⁸ Korea, [Bill on Artificial Intelligence Liability](#), Feb. 2023

⁷⁹ United Arab Emirates Minister of State for Artificial Intelligence, Digital Economy and Remote Work Applications Office, [The UAE Charter for the Development and Use of Artificial Intelligence](#), July 2024

⁸⁰ United Arab Emirates Minister of State for Artificial Intelligence, Digital Economy and Remote Work Applications Office, [The UAE Position on AI Policy](#), October 2024

⁸¹ United Arab Emirates Minister of State for Artificial Intelligence, Digital Economy and Remote Work Applications Office, [The UAE Charter for the Development and Use of Artificial Intelligence](#), July 2024

⁸² United Arab Emirates Minister of State for Artificial Intelligence, Digital Economy and Remote Work Applications Office, [The UAE Position on AI Policy](#), October 2024

at the centre of its legislative efforts for many of the same reasons described in this discussion paper.⁸³ The authors note, “Current legal precedent does not define the status of AI concerning product liability law.... Liability would provide a framework for protection and legal recourse to address immediate and emerging harms from unregulated, highly powerful AI systems, especially as capabilities increase and use proliferates.”

Proposals include assigning a “duty of care” to AI developers and deployers, establishing a legal obligation to prioritize safety and harm prevention in their product design and deployment. Such efforts could integrate with independently established and ratified standards, for example, through the IEEE Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)⁸⁴, such that liability risks are mitigated by careful compliance.

A proposal on AI liability reform from the Center for Urban Policy Research at Rutgers University⁸⁵ makes the case for this mechanism as a means to leverage market forces and familiar legal mechanisms in the interest of safer, more ethical AI outcomes, arguing that expanding liability laws to take algorithmic and autonomously advancing harms into account will incentivize companies to integrate safeguards in their design and deployment. The Center advocates for clearer legal standards and enforcement mechanisms, including the ability for regulatory agencies to bring liability complaints against developers for negligence.

In addition to legislative proposals, U.S. federal agencies are using their existing statutory authority to regulate AI, exerting their power to investigate and carry out enforcement actions including fines for those adjudicated to have violated established standards. The United States Federal Trade Commission (FTC) is among the most active US regulatory bodies in this regard in terms of consumer protection. It announced five enforcement actions as a result of an investigation denoted “Operation AI Comply”. This investigation was undertaken by authority already granted to the FTC to protect consumers from “fraud, scams, and deceptive business practices.”⁸⁶

4. Conclusion and Recommendations

Through this discussion paper, we hope to amplify the conversation about a notable gap in global AI governance – applying liability frameworks as an indispensable lever to incentivize safe and ethical outcomes and to offer recourse for harm. While researchers and policymakers around the world have acknowledged the need to clarify the legal complexities that accompany AI liability,

⁸³ Center for Humane Technology. (2024). [A Framework for Incentivizing Responsible Artificial Intelligence Development and Use](#). Retrieved September 24, 2024

⁸⁴ IEEE Standards Association, [The Ethics Certification Program for Autonomous Intelligent Systems \(ECPAIS\)](#), Retrieved September, 2024

⁸⁵ M.H. Pfeiffer, [First Do No Harm: Algorithms, AI, and Digital Product Liability](#), Sept. 2023

⁸⁶ J.S. Ensor, “Operation AI Comply: continuing the crackdown on overpromises and AI-related lies”, FTC Business Blog, Sept. 2024

this conversation has thus far been most fully explored within the European Union. Gross disparities in liability frameworks will create highly risky vulnerabilities, particularly for individuals and communities in the Global South where AI-related harms, from biased lending algorithms to unchecked deepfakes, are already pervasive. To address this governance gap, we propose four immediate actions:

1. Establish a **Global AI Liability Task Force**, bringing together experts from diverse jurisdictions to develop harmonized principles that can be adapted across frameworks, and exploring the development of an international framework for AI liability.
2. Formalize **adherence to ethical AI industry standards** – such as those by IEEE and ISO – into liability frameworks – to incentivize AI companies to rigorously implement the voluntary safeguards delineated by standards-setting bodies.
3. Investigate the potential applicability of a **“chain of responsibility” framework for AI liability** to clarify accountability across the complex AI lifecycle.
4. Develop **capacity-building initiatives focused on AI liability** in Global Majority countries, addressing both the technical and legal aspects of enforcement, coupled with initiatives to bridge this divide, including investments in digital infrastructure and promotion of digital literacy, to ensure effective implementation and enforcement of AI regulations.

By platforming conversations on AI liability as a lever of governance, global policy stakeholders would take significant strides towards helping level the playing field for AI, mitigating risks, establishing fair means for obtaining recourse for harms, and increasing the likelihood that the benefits of AI are realized responsibly and equitably around the world.

Environmental Sustainability and the Generative AI Value Chain

Policy Network on Artificial Intelligence (PNAI)
Sub-group on Environmental Sustainability within Generative AI Value Chain

Acknowledgments

Team leads

Shamira Ahmed (Data Economy Policy Hub, South Africa); Mohd Asyraf Zulkifley (Universiti Kebangsaan Malaysia, Malaysia)

Pen holders

Dr. Muhammad Shahzad Aslam (Xiamen University, Malaysia); Lufuno T Tshikalange (Orizur Consulting Enterprise, South Africa); Mary Rose Ofianga (Wadhvani Foundation, Philippines); Nabuzale Doreen Nandutu (eLAAB Limited, Uganda); Siti Raihanah Abdani (Universiti Teknologi MARA, Malaysia); Somya Joshi (Stockholm Environment Institute, Sweden)

List of Abbreviations and Acronyms

AWS	Amazon Web Services
DPG	Digital Public Goods
DS	Data-Based Systems
FM	Foundation Models
FT	Frontier Technologies
Gen-AI	Generative Artificial Intelligence
GPU	Graphics Processing Unit
JT	Just Transition
JTCE	Just Transition to the Circular Economy
LCA	Life Cycle Assessment
LMICs	Low and Middle-Income Countries
PNAI	Policy Network on Artificial Intelligence
SDGs	Sustainable Development Goals
VCA	Value Chain Analysis

1. Introduction

Digital transformation is a global phenomenon that enhances productivity, disrupts traditional business models, and fosters a wide range of innovations with significant implications for humanity's future (UNIDO, 2024). Driven mainly by data-based systems (DS), the transformation of socioeconomic, political, and market activity holds tremendous potential to advance sustainability. The current global trajectory spurred by digital transformation often perpetuates the use of frontier technologies (FT), to exacerbate unsustainable practices that harm natural ecosystems, worsen multidimensional inequality, and threaten human well-being (CODES, 2022, UNCTAD, 2024a).

Since their prolific explosion in 2022, the ubiquity and exponential expansion of generative artificial intelligence (Gen-AI) platforms have become a global phenomenon (Harlin et al., 2023). While Gen-AI models such as OpenAI's Generative Pre-Trained Transformer (GPT) models and other related foundation models (FM) present opportunities for innovation across industries, there is a growing realization that not every Gen-AI application will be inherently beneficial or realize its anticipated advantages (Bender et al. 2021). Beyond existential risks that could exacerbate long-standing ethical and socio-economic issues, such as surveillance, privacy violations, multidimensional inequality, and discrimination, to name a few — DS such as Gen-AI also pose significant risks associated with global environmental sustainability concerns (Bashir et al. 2024; Ahmed & Kirchschräger, 2024; Kalantzakos 2020).

Companies are incentivized to prioritize AI performance, efficiency, and scalability, often overlooking the environmental costs of Gen-AI innovations while negating social and environmental impacts (Domínguez Hernández et al, 2024; Varoquaux et al.,2024). At present, efforts to enhance computing sustainability are primarily centered on improving efficiency through boosting hardware energy efficiency, optimizing AI algorithms, and increasing the carbon efficiency of computing workloads through techniques like spatiotemporal workload shifting (Bashir et al, 2024).

Furthermore, the narrow focus on efficiency and scalability—driven by the relentless demand for Gen-AI fails to address the broader environmental challenges tied to the Gen-AI value chain (Bashir et al., 2024). Concerted interventions are essential to shift beyond a purely technical approach toward integrated thematic and topic modeling analysis (Raman et al., 2024), establish robust global Gen-AI governance frameworks (Ahmed and Kirchschräger, 2024), and ensure widespread access to digital public goods (DPGs)^[1].

The Policy Network on AI (PNAI) is dedicated to integrating environmental considerations into the responsible global governance of Gen-AI, that aligns with best practices to support a just green-digital (twin) transition for the Global Majority (PNAI, 2023; WEF, 2024). PNAI's commitment aligns with ongoing broader sustainability goals, including the United Nations Sustainable Development Goals (SDGs), which encompass objectives broadly related to adaptation, mitigation, and loss and damage, including biodiversity loss, wildlife conservation, and the sustainable use of natural resources.

By raising awareness of the global governance dimensions needed to support sustainable practices throughout the Gen-AI value chain, PNAI aims to raise awareness regarding the environmental impact associated with the development, deployment, and disposal of Gen-AI technologies, through prioritising transversal policy action for climate justice while simultaneously minimizing the overall negative environmental impact of Gen-AI.

Assessing and mitigating the environmental impact of Gen-AI technologies is particularly important for the Global Majority, who may disproportionately bear the consequences of climate change linked to unsustainable digital economy practices (UNCTAD, 2024a). Communities in low and middle-income countries (LMICs)^[2] often already bear the brunt of environmental degradation and the extraction of labour and natural resources that are associated with technological transitions (UNCTAD, 2021). To achieve planetary health and human well-being, a shift in perspective is needed to address the grand challenges of our time, which requires analysis beyond the triple planetary crisis to include other dimensions such as institutionalised inequality, decolonialisation, shifts in social and demographic dynamics, advancements in FT, geopolitical tensions, trust in multilateral social and institutional frameworks, and migration, among other factors (Ahmed and Kirchschräger, 2024; UNDP, 2024).

Communities who have been and continue to be marginalised need to be empowered to move beyond narratives as mere 'victims' and should be considered as holders of valuable and legitimate knowledge in times of multidimensional transitions, to ensure that the potential benefits of Gen-AI are not realized at the cost of LMIC's ecological ecosystems, livelihoods, and natural resources (Lehuedé 2024) and simultaneously perpetuate historical and ongoing patterns of inequity (Elia 2023; Guerrero 2023), where wealthier nations and their corporations may benefit from the efficiencies generated by Gen-AI, while poorer regions bear negative externalities, including the environmental costs without reaping similar rewards (UNCTAD, 2024a).

Extraterritorial decisions regarding the development, deployment, and regulation of Gen-AI technologies are often made by high-income countries and large corporations, sidelining voices from marginalized communities that are most likely to be affected (WEF, 2024). For example, the lack of representation in many global climate governance decisions further entrenches inequalities and diminishes the ability of many communities in the Global Majority to advocate for their needs (Ren & Wierman, 2023).

While the field of sustainable AI has developed and has been put forward as a way to address environmental justice issues associated with AI throughout its lifecycle (Luccioni et al., 2024; Robbins and van Wynsberghe 2022; Strubell, et.al.,2020), there is limited literature focused on the Gen-AI value chain or a value chain analysis (VCA) of the environmental toll of Gen-AI from a Global Majority lens.

This discussion paper was created to facilitate much-needed multistakeholder dialogue, providing insights into the opportunities and environmental externalities underpinning the Gen-AI value chain. The development of this discussion paper is based on insights from multidisciplinary stakeholders from diverse regions, ensuring a holistic global perspective on environmental sustainability and the Gen-AI value chain^[3].

2. Environmental Sustainability and the Generative AI Value Chain

2.1. State of Global Gen-AI Governance

The current global governance of AI faces several critical issues that reflect the complexities of the technology such as, structural limitations, global imbalances, and navigating a complex geopolitical landscape characterized by rapid technological advancements, cross-border impacts, ethical considerations, and the need for balance between innovation and regulation (Ahmed et al., 2023; WEF, 2024).

Furthermore, current practices of global environmental safety and risk mitigation for governing AI often focus narrowly on improving energy efficiency without adequately addressing the broader sustainability and sociotechnical challenges, leading to an incomplete understanding of the environmental costs associated with Gen-AI development and deployment (Domínguez Hernández et al. 2024). The development of larger and more complex models is often prioritized for competitive reasons, without fully accounting for the carbon cost of training and deploying these models at scale (Bashir et al. 2024; Varoquaux et al., 2024). As investment in the development and application of Gen-AI technologies continues to grow, it becomes increasingly crucial to understand their impact on the environment.

Furthermore, discussions regarding the balance between the potential benefits of Gen-AI systems and their environmental costs must be based on concrete data and evidence (Bashir, et al., 2024). Unfortunately, most developers and operators of these systems are not currently providing the necessary data. The lack of publicly available information hinders the formulation and implementation of effective evidence-based policies (PNAI, 2023).

Moreover, while Gen-AI is paradoxically increasingly heralded for its potential to contribute to climate mitigation, many of the documented solutions rely on data and case studies primarily from the Global North. As Gen-AI models optimize datasets and computational power to generate outputs that often lack diversity, their solutions may fall short of addressing the unique challenges of the Global Majority (Ahmed, 2023). Analysis of the Gen-AI value chain highlights that true innovation must grow from a strong local digital ecosystem where businesses, entrepreneurs, and academic institutions actively contribute. Multistakeholder collaboration will be essential in co-creating agile, adaptive governance that promotes the development of green digital jobs and sustainable livelihoods, fostering economic growth while enhancing environmental resilience (Bashir et al., 2024).

2.2. Exploring the Generative AI Value Chain

The Gen-AI value chain outlines the various stages and components that comprise the development, deployment, and utilization of Gen-AI technologies (Harlin, et al., 2023).

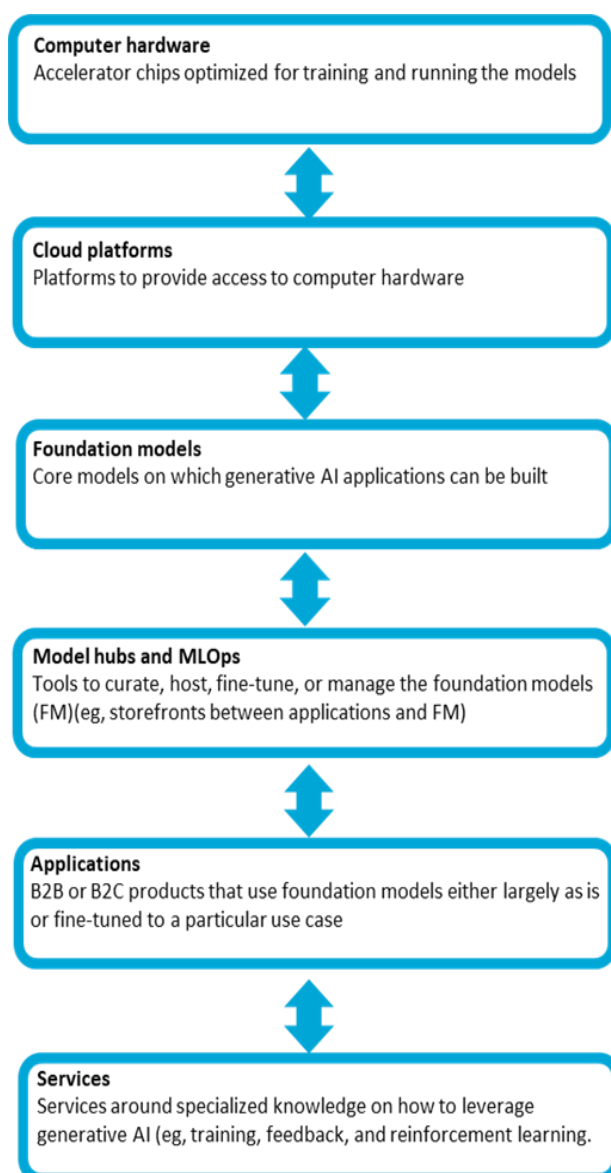
The Gen-AI value chain involves hardware in the form of devices and sensors to capture the data and data centers to store them, cloud platforms and networks for communicating data, FMs, model hubs that act as repositories for storing and accessing foundation models, applications

such as end-user interfaces, and existing AI service providers and new niche players that specialize in Gen-AI applications (Harlin et al.,2023).

The Gen-AI value chain reflects a complex ecosystem that supports the creation and deployment of innovative AI solutions. As Gen-AI continues to evolve, understanding the value chain is essential for identifying investment opportunities, anticipatory governance, and assessing the potential impact of these technologies to enhance environmental stewardship.

As shown in Figure 1, the Gen-AI value chain consists of several non-linear and dynamic key elements that contribute to the overall functionality and effectiveness of Gen-AI systems, including natural resources and energy to build and transport the devices and products, which emit greenhouse gases throughout the value chain (Bashir et al. 2024).

Figure 1: The Generative AI Value Chain



2.3. Key Differences Between Gen-AI and Traditional AI Applications

Gen-AI and traditional (weak or narrow) AI applications differ primarily in their goals and methods of operation, traditional AI typically focuses on recognizing patterns, making predictions, or automating tasks based on pre-existing data (Bond-Taylor et al., 2022).

While traditional AI applications emphasize accuracy and efficiency in task completion, Gen-AI prioritizes creativity and the ability to produce novel outputs that didn't exist before. For instance, chatbots powered by traditional AI may provide factual responses, while those using generative AI can engage in more human-like, creative conversations (WEF, 2024). Gen-AI represents a paradigm shift by enabling the autonomous creation of novel content and adapting to complex scenarios. However, both traditional AI and Gen-AI are complementary, with traditional AI methods still holding immense value for many applications (Harlin et al.,2023).

The differences between Gen-AI and traditional AI applications require distinct governance approaches because of the inherent risks and implications each type of AI presents (Bender et al. 2021; Brundage et al. 2020)[4].

2.4. The Gen-AI Value Chain and the Environment

2.4.1. Mapping the Environmental Toll of the Gen-AI Value Chain

Each stage of this value chain contributes to the overall carbon footprint and resource depletion. The following summaries the environmental toll at each stage of the suggested Gen-AI value chain model:

i. Computer Hardware

Computer hardware provides the foundational layer of the Gen-AI value chain, which includes semiconductors and specialized processors, such as Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs). These provide the necessary computational power to handle the extensive data processing and complex algorithms required by Gen-AI models (Harlin et al.,2023). The production of computer hardware involves substantial energy consumption, contributing to carbon emissions. For instance, manufacturing a single high-performance GPU can emit over 200 kg of CO₂, highlighting the energy-intensive nature of the process (Bashir et al. 2024).

While Gen-AI dominates global technological discussions, it cannot be viewed in isolation, as the raw materials and components driving its progress are controlled by a few, globally dispersed chip manufacturers. The dependency on a few semiconductor companies underscores the interconnectedness of AI with the broader technological and supply chain ecosystem (Burkacky et al., 2024). Many governments have designated semiconductors as 'critical technologies'. Due

to their foundational role at the base of the technology stack. This makes them essential to the advancement of nearly every emerging technology, since semiconductors enable the processing of large amounts of data and the rapid execution of complex calculations that form the foundation of AI systems, including Gen-AI (Janjeva et al.,2024).

Left unchecked, the semiconductor boom could exacerbate huge levels of toxic waste in the form of air pollutants and groundwater contamination, due to certain chemicals and gases that are essential for semiconductor manufacturing (Perkins 2024). As far back as the 1980s, the use of chemicals in semiconductor operations has long been a challenge. The updated F-gas regulation and the proposed ban on per- and polyfluoroalkyl substances (PFAS) reflect growing regulatory efforts to mitigate the environmental impacts associated with semiconductor production (Hess, 2024).

However, the increased integration of Gen-AI systems into various sectors creates new complexities and risks that require coordinated international efforts to ensure their safe and responsible use (WEF, 2024). In the context of growing concerns over climate change and biodiversity loss, addressing the environmental footprint of AI technologies is crucial. As Gen-AI applications evolve and proliferate, it becomes essential to address their ecological footprint to ensure sustainable digital development and a just green digital twin transition (Bashir et al.,2024).

In addition, the computational power to train Gen-AI models requires significant energy that contributes to the overall carbon footprint of AI technologies and water use (Ren and Wierman, 2023). There are projections that the current computationally intensive training process for models like Gen-AI, such as GPT-3 and the demand for high-performance semiconductor components, including logic chips (CPUs, GPUs, AI accelerators), memory chips (HBM, DDR), and data storage chips (NAND) will skyrocket to unprecedented levels by 2030 contributing to a significant carbon footprint, which poses challenges in achieving net-zero greenhouse gas emissions and accelerates depletion of natural resources (Bashir et al.,2024; Burkacky,et al.2024).

The data centres housing the hardware necessary for training and running Gen-AI models require significant cooling to maintain optimal operating temperatures. This cooling process often involves substantial water usage and consumes large amounts of electricity. With estimates indicating that data centres account for approximately 20 percent of electricity consumption in some regions. This raises concerns about water scarcity in regions where these data centres are located (Bashir et al. 2024; Ren, 2023).

As the world transitions to low-carbon technologies, such as electric vehicles and renewable energy systems, the demand for critical minerals like lithium, cobalt, and nickel is surging (Kalantzakos, 2020). These minerals are essential for batteries and other components of so-called green technologies, leading to intensified mining activities. The rising demand for low-carbon technologies escalates the need for critical minerals, as the production of computer hardware involves the extraction of various natural resources, including silicon and rare earth metals (EPA, 2012). The processing of raw rare earth minerals can lead to environmental degradation, habitat destruction, and increased carbon emissions associated with mining and manufacturing; mainly for the global majority. Where there are higher geographic concentrations

of reserves and processing for five critical minerals: cobalt, copper, lithium, nickel, and rare earth elements (UNCTAD, 2024b).

The environmental costs associated with extracting resources for AI infrastructure are significant and multifaceted (UNCTAD, 2024b). Additionally, the rapid obsolescence of hardware technology leads to the discarding of outdated equipment, contributing to the growing problem of electronic waste (e-waste). The Global E-waste Monitor 2024 report indicates that e-waste generation in 2022 reached a record 62 million metric tonnes, with only 22 percent being officially collected and recycled. The annual generation of e-waste is rising by 2.6 million tonnes annually, on track to reach 82 million tonnes by 2030 (UNITAR, 2024). Sustainable disposal and recycling practices are essential to mitigate the environmental impact of outdated equipment.

ii. Cloud Platforms

Gen-AI is revolutionizing industries by empowering machines to produce content, tackle complex challenges, and fuel innovations once thought impossible. From generating human-like text to creating realistic images, the capabilities of generative AI are vast and transformative (Intel, 2024).

Gen-AI requires substantial computational power to process and generate data. Reliable cloud environments offer the scalability needed to handle these demands (Harlin et al.,2023). Cloud platforms facilitate flexibility so that resources can be scaled up or down based on the workload. Ensuring that Gen-AI models run efficiently without hardware limitations. This is particularly beneficial for training large models, which may require varying levels of resources at different stages of the value chain (Harlin et al.,2023).

Major cloud service providers, such as Amazon (AWS), Microsoft Azure, and Google Cloud, are increasingly offering scalable infrastructure that enables the deployment of Gen-AI applications and models. Facilitating access to vast computational resources and new database capabilities for storing and rapidly retrieving the unstructured and semi-structured data used in Gen-AI systems (Accenture, 2023).

However, cloud platforms that host Gen-AI applications consume substantial energy for both processing and cooling. Data centres must maintain optimal temperatures to ensure efficient operation, leading to increased electricity and water usage (Bashir et al. 2024). The ongoing maintenance of cloud infrastructure, including hardware upgrades and system monitoring, can contribute to resource depletion and environmental impact. A 2019 study estimated that training a large natural language model like GPT-3 on cloud platforms could result in carbon emissions equivalent to the lifetime emissions of five cars (Strubell, et.al.,2019). While there are efforts to stem the ecological externalities from increased energy use, there remains a lot more to do given the world's increasing need for computing power (Bashir et al. 2024).

The operational costs associated with running Gen-AI systems—particularly regarding cloud computing and energy consumption—are significant. Smaller organizations or those in developing regions may not have the financial capacity to sustain such expenses. This economic

barrier leads to increased reliance on established providers who can absorb these costs, thereby limiting the ability of local entities to pursue independent innovation (Lynn et al., 2023).

iii. Foundation Models

Gen-AI models rely on extensive datasets to learn patterns and generate realistic outputs. For instance, during the training phase, Gen-AI models absorb petabytes of data—from diverse sources like books, websites, and other machine-readable digital content. The quality and diversity of the training data directly impact the performance of the AI (Wu & Higgins, 2023)

Foundation Models (FM) are large pre-trained models, such as BERT, OpenAI's GPT-4, and DALL-E, which serve as the core building blocks for various Gen-AI applications, (also called large language models or LLMs) and are meant for general use (Harlin et al.,2023).

FMs require extensive computational power and training periods. Once FM are trained, they require continuous energy for inference and processing tasks. The ongoing energy demand can significantly add to the carbon footprint of AI applications, particularly as their usage scales (Patterson et al. 2021). Once AI models are deployed, they require ongoing operational energy for inference and processing tasks. The exact environmental cost of Gen-AI activity is not known, since the developers of the latest models do not provide detailed emissions figures. A thorough assessment of the environmental costs involved in maintaining Gen-AI technologies is urgently required (Bashir et al. 2024).

While there are efforts being made towards enhanced algorithmic efficiency and reduced computational requirements, there is growing demand for Gen-AI applications. Such as the development of more innovative efficient transformer models, which aim to decrease the number of operations and memory needed during training (Burkacky, et al., 2024) to mitigate hallucinations (IBM, 2023). There is not enough global governance considerations on the categories of risks and harms related to environmental sustainability and natural resource management externalities associated with the most urgent negative impacts of FM and their downstream applications (Domínguez Hernández et al., 2024).

Effective governance of Gen-AI requires robust data governance frameworks to ensure interoperability, transparency, and accountability. However, the lack of clear guidelines on data usage, unequal access to high quality datasets, and the potential for misuse complicate the establishment of reliable indicators for sustainability (PNAI, 2023).

iv. Model Hubs and Machine Learning Operations

Model hubs act as repositories for storing and accessing FMs. While machine learning operations (MLOps) encompass the tools and practices used for managing and deploying FM in real-world applications. This stage is crucial for ensuring that models are effectively integrated into user-facing applications (Harlin et al.,2023).

Beyond the products themselves, energy consumption for storing, managing, and deploying models is also required for model hubs and MLOps. For example, regular updates and versioning

of models can lead to increased resource consumption and waste generation, necessitating sustainable practices in the lifecycle management of AI models (Patterson et al., 2021).

The reliance on substantial computational resources for MLOps not only contributes to higher energy consumption but also raises concerns about the environmental impact of these technologies. The operational emissions linked to running Gen-AI systems can exacerbate climate change, necessitating thorough assessments of their environmental costs. Moreover, as model hubs and MLOps become more prevalent, the potential for monopolistic behaviour by a few dominant Gen-AI service providers increases. Which can limit access for smaller players and marginalize communities in the Global Majority. This dependency on established providers can stifle innovation and exacerbate inequalities, as local developers may lack the resources to compete effectively in a landscape dominated by major tech companies.

v. Applications

The applications layer includes the end-user interfaces and solutions, such as chatbots, content generators, and creative tools, that utilize Gen-AI models to perform specific tasks. The application layer is expected to see rapid growth and innovation, offering significant value-creation opportunities due to the demand from both business-to-consumer (B2C) and business-to-business(B2B) applications (Harlin et al.,2023). Gen-AI applications have created a hyper-competitive tech ecosystem that requires Gen-AI platforms to develop constant improvements to the quality of their Gen-AI algorithms. This demand presents a wide range of issues such as high resource-intensity, which often requires vast amounts of quality data and computational power, aided as much by big data as it is by software and hardware (Bashir et al. 2024).

The use of Gen-AI applications, such as chatbots and content generators, requires significant computational resources and energy for real-time processing (Bashir et al. 2024). The environmental implications of these applications span from the depletion of natural resources to the contribution of carbon emissions (Crawford 2024; Strubell et.al., 2019).

vi. Services

The services component involves existing AI service providers and new niche players that specialize in generative AI applications. These services are geared to help organizations navigate the complexities of implementing Gen-AI technologies and often provide tailored solutions for specific industries or functionalities (Harlin et al.,2023). The provision of Gen-AI services, including consulting and support, often relies on substantial computational resources, contributing to energy consumption and environmental impact. The operational emissions associated with running Gen-AI systems can exacerbate climate change, necessitating a thorough assessment of the environmental costs involved in maintaining these technologies (Bashir et al. 2024).

The data required to train effective AI models is often controlled by these dominant firms, which hoard vast datasets that are crucial for innovation. This concentration restricts access for emerging players from the global Majority, who may lack the means to acquire or generate comparable datasets. Consequently, this reinforces a cycle of dependency where local innovators

cannot compete effectively, further entrenching the power of established multinational “Big Tech” corporations.

Overall, the Gen-AI value-chain is dominated by a few large tech companies that control substantial computational resources, data, and infrastructure necessary for developing and deploying AI technologies. These companies leverage their existing market power to dictate terms for access to essential services, creating barriers for smaller players and startups. As a result, organizations in the global Majority often find themselves reliant on these monopolistic entities for critical resources, stifling their ability to develop independent solutions and innovations (Lynn et al., 2023).

3. Need for Gen-AI Environmental Impact Metrics and Indicators^[5]

With the increased demand and use of Gen-AI so do the significant environmental costs associated with the development, training, and deployment of large-scale AI models (Strubell et.al., 2019). As Gen-AI becomes more widespread in applications such as content generation, virtual assistants, and creative tools, its energy consumption, resource use, and carbon footprint grow (Bashir, et al.,2024). Training state-of-the-art models such as GPT-3 or image-generating generative adversarial networks (GAN) involves processing massive datasets across large clusters of GPUs or TPUs, which consume substantial amounts of energy (Patterson et al.,2021).

The need for robust environmental impact metrics and indicators for Gen-AI is underscored by the challenges tech giants face in reducing emissions. Despite pledges to reach net-zero, companies like Google, Microsoft, and Meta have seen increases in greenhouse gas emissions, with Google reporting a 48% rise in 2023 compared to 2019 (Google, 2024). Although they have scaled up low-carbon energy use, growing demands for computational power drive higher overall energy consumption, inadvertently increasing fossil fuel reliance. Establishing targeted environmental metrics for Gen-AI can help address this gap, enabling more precise tracking and management of its carbon footprint.

Furthermore, without a consistent way to measure emissions, it becomes difficult for companies and institutions to track or reduce their Gen-AI-related environmental impacts, including supporting regulations aimed at limiting the carbon emissions of tech companies and holding tech companies accountable (Bashir et al. 2024).

3.1. Development of Metrics

The establishment of accurate and comprehensive metrics is crucial for assessing the environmental impact of Gen-AI. Accurate metrics are essential to enable the comparison of energy footprints between different models and optimization strategies and foster more energy-efficient practices (Bashir et al. 2024).

Without clear metrics to quantify Gen-AI energy use, it's difficult to gauge the full environmental impact of Gen-AI systems. Metrics can help AI developers and policymakers better understand,

manage and provide a framework for assessing energy consumption, resource utilization, and emissions associated with AI development, deployment, and usage. Without reliable metrics, it becomes challenging to identify areas for improvement, track progress toward sustainability goals, and for stakeholders to evaluate the sustainability of AI technologies effectively and inform decision-making processes aimed at minimizing ecological harm (OECD, 2022).

Furthermore, the existence of clear metrics can direct research and development efforts toward reducing environmental impact, encouraging innovations that make Gen-AI technology more sustainable.

Lessons from existing initiatives reveal that various types of metrics can be employed to measure the environmental impact of Gen-AI for a wide range of reporting requirements such as environmental, social and governance (ESG), used to evaluate a company's sustainability and ethical impact (EY, 2023). While not an exhaustive list, a summary of metrics can include:

- i. **Carbon Footprint:** This metric quantifies the total greenhouse gas emissions produced directly and indirectly by Gen-AI systems, providing insight into their contribution to climate change (Bashir et al. 2024).
- ii. **Energy Efficiency:** This metric assesses the amount of energy consumed per unit of output generated by Gen-AI models, helping to identify opportunities for reducing energy usage.
- iii. **E-Waste:** This metric can assess the amount of e-waste produced as compute hardware becomes obsolete due to Gen-AI hardware requirements (UNITAR, 2024).
- iv. **Resource Utilization:** This metric evaluates the extraction and consumption of natural resources, such as water and minerals, associated with the production and operation of Gen-AI infrastructure (Robbins and van Wynsberghe 2022).
- v. **Socio-economic Impact:** This metric can evaluate the sustainability of Gen-AI technologies effectively and inform decision-making processes aimed at minimizing ecological harm and facilitating a just transition (PNAI, 2023).

3.2. Indicators for Sustainability

Defining indicators is vital for measuring progress toward environmental sustainability goals in the context of Gen-AI. These indicators can help stakeholders assess the effectiveness of sustainability initiatives and identify areas for further improvement. For example, establishing standardized indicators of carbon output, such as kilograms of CO₂ per hour of computation or per inference task, would drive accountability and encourage the adoption of carbon-neutral or lower-emission energy sources (OECD, 2022).

Indicators should be clear, measurable, and relevant to the specific environmental goals being pursued. They should also consider the unique challenges and opportunities presented by Gen-AI technologies. Examples of effective indicators can include:

Reduction in energy consumption. Tracking the decrease in energy usage associated with Gen-AI applications can demonstrate progress toward improving energy efficiency (Bashir, et al.2024).

Use of renewable energy sources. Measuring the percentage of energy sourced from renewable resources at different stages of the Gen-AI value chain can indicate the commitment to sustainable energy practices (Bashir, et al.2024).

Volume of recycled materials and minerals. Monitoring the number of recycled materials used in the production of AI hardware can help assess the effectiveness of circular economy (CE) initiatives and leverage AI to support a just transition to the circular economy (JTCE) (Ahmed, 2022).

3.3. Challenges in Developing and Governing Indicators

The fast-paced development of AI technologies poses a challenge for regulators and standard-setting bodies, the velocity of AI-related innovations often outstrips the ability of governance frameworks to adapt, resulting in outdated or ineffective measures that fail to capture the evolving nature of Gen-AI, including understanding and mitigating its environmental implications (Domínguez Hernández et al. 2024).

Developing and governing sustainability indicators to assess the environmental impact of Gen-AI involves navigating a complex landscape of challenges, particularly in the context of global governance, ecological inequities, geopolitical power dynamics, and the influence of socio-technical imaginaries that predominantly shape innovation ecosystems in the Global North (Ahmed, et al, 2023).

There is a notable governance deficit in the current international landscape concerning the governance of Gen-AI and DS, in general (Ahmed and Kirchschräger, 2024; Domínguez Hernández et al. 2024). Existing initiatives often lack the coordination and capacity necessary to address the complexities of Gen-AI's environmental impacts and the fragmentation of governance structures complicates the establishment of coherent and inclusive indicators that can effectively measure sustainability across different contexts (Bashir et al., 2024).

Furthermore, many existing standards and best practices for AI are rooted in the socio-technical contexts of the Global North, which often do not reflect the realities or needs of the Global Majority, which can lead to the development of indicators that are not universally applicable or that overlook critical environmental, political economy, and sociotechnical factors relevant to developing global standards that mitigate risks and support the flourishing of diverse ecosystems (Bashir et al., 2024).

Geopolitical tensions and competition hinder cooperation on global AI governance. Long-standing first-order cooperation problems, combined with second-order issues stemming from dysfunctional international institutions, complicate the establishment of effective governance frameworks for Gen-AI that are equitable and reflective of global needs (Bashir et al., 2024)

The development and governance of indicators for the environmental impact of generative AI face significant challenges, particularly due to biases in existing frameworks, geopolitical barriers, and the rapid evolution of technology. Underrepresentation of the Global Majority in discussions about standards and best practices, and the overall global ethical, legal, social, and policy (ELSP) aspects also contribute to the aforementioned challenges and require a collaborative approach among diverse stakeholders.

Measuring Gen-AI's sustainability, such as its carbon footprint, and ensuring compliance with global sustainability standards is still in its infancy. Tools like the ESG Digital and Green Index are emerging to help, but widespread adoption is needed (Raman et al. 2024; Thelisson et al. 2023). Nevertheless, the development of metrics and indicators for assessing the environmental impact of Gen-AI is essential for promoting sustainability in the technology sector (Bashir et al., 2024). By establishing comprehensive metrics, engaging in multistakeholder dialogue, and leveraging high-quality data, stakeholders can work collaboratively to minimize the ecological footprint of AI technologies and ensure that their benefits are realized without compromising environmental integrity.

4. Role of Data Governance in Assessing Environmental Impacts

Data and its underlying foundations are the determining factors to leverage the potential of Gen-AI (Caserta et al., 2023). Effective governance of DS such as Gen-AI requires robust data governance frameworks to ensure transparency, accountability, and to facilitate just data value creation (JDVC) (Ahmed and Kirchschräger, 2024).

Robust data governance plays a crucial role in establishing clear policies, standards, and processes for data management. Data governance ensures that the data used to train and deploy Gen-AI models is collected, processed, and stored in a responsible manner that minimizes environmental harm (PNAI,2023; Bashir et al. 2024).

Effective data governance also enables transparency and accountability in reporting on the environmental footprint of Gen-AI, allowing organizations to identify areas for improvement and track progress toward sustainability goals (OECD, 2022). This includes measures such as tracking energy consumption and emissions from data centres, managing the use of natural resources like water and minerals, and ensuring data quality and integrity to avoid the need for excessive retraining of models (OECD, 2022). Without robust data governance, it's hard to measure the true environmental cost at each stage of data processing at each stage of the Gen-AI value chain.

Furthermore, robust data governance is essential for ensuring equitable access and distribution of data dividends when treating data as a DPG (UNICEF, 2023)

4.1. Importance of Data

High-quality machine-readable data plays a crucial role in assessing the environmental impacts of Gen-AI. In the 2023 PNAI Report, we highlight how robust data governance facilitates high-quality, accessible datasets, which are necessary for accurate measurement and evaluation of sustainability metrics and climate justice (PNAI, 2023). Reliable data enables stakeholders to quantify the environmental effects of AI technologies, including energy consumption, emissions, and resource utilization. Data is essential for making informed decisions and implementing effective sustainability initiatives (CODES, 2022).

However, the lack of clear guidelines on data usage and the potential for misuse complicate the establishment of reliable indicators for sustainability (OECD, 2022).

4.2. Types of Data Required

To accurately assess the environmental impact of Gen-AI, various types of data are needed, and a comprehensive data collection effort is required, not an exhaustive list but focus on the following key areas is crucial:

Data on Energy Usage. Information on the energy consumption throughout the Gen-AI value chain, particularly during training, deployment, and inference is critical for evaluating their carbon footprint and identifying opportunities for efficiency improvements. This data should include electricity usage by data centres and cloud infrastructure supporting Gen-AI, fuel consumption by backup generators and transportation related to Gen-AI operations, and energy usage per model training run and per inference, to name a few (Patterson et al. 2021; Strubell et al., 2019).

Resource Consumption Data. Data on the extraction and use of natural resources, such as water and critical minerals, is necessary to understand the broader environmental implications of the Gen-AI value chain, which includes water usage for cooling data centres, mineral and metal consumption for manufacturing Gen-AI hardware, and use for data centre construction and siting (Bashir et al., 2024).

Emissions Data. Tracking greenhouse gas emissions associated with Gen-AI operations is essential for measuring progress toward climate goals and identifying areas for reduction, such as direct emissions from on-site fuel combustion, indirect emissions from purchased electricity and heat, and emissions from upstream activities like manufacturing and transportation, to name a few (Kemene et al., 2024).

Socioeconomic Data. Assessing the holistic impact of Gen-AI requires understanding its socioeconomic implications, particularly in underrepresented regions and to realise a just green digital “twin transition”. For example, sex-disaggregated data is crucial for identifying differential impacts at the nexus of climate injustice and AI on women and men (Ahmed, 2022). Women often have distinct roles and responsibilities in resource management and consumption, which can influence how AI technologies are adopted and their subsequent environmental effects (Ahmed, 2024). Sex disaggregated data allows for a nuanced understanding of how technologies

affect different genders, particularly in terms of resource consumption, energy usage, and emissions (GEDA, 2024). Other relevant data can include employment and income effects of Gen-AI adoption, access and use of Gen-AI-enabled services by marginalized communities, and representation of diverse perspectives in Gen-AI development and governance (ILO, 2023; Ahmed et al., 2023; PNAI 2023).

Contextual Data. Given the global AI divide, to adequately interpret the environmental and social impacts of the Gen-AI value chain, contextual data is needed on factors such as: Local climate and environmental conditions, existing infrastructure and resource constraints, political economy dynamics, and socioeconomic and demographic characteristics of affected populations, to name a few (Ahmed, et al., 2023).

Collecting and integrating diverse data will enable a holistic assessment of Gen-AI's environmental footprint and help guide the development of sustainable practices (Bashir et al., 2024). Collaboration among Gen-AI developers, data providers, and domain experts is essential to establish comprehensive data collection frameworks and ensure data quality and different types of interoperability (World Bank 2022; Gonzalez Morales and Orrell, 2018).

For example, the energy usage of Gen-AI models and infrastructure is often opaque, with limited real-time transparency. Models trained in various geographical locations may utilize different sources of energy (renewable vs. non-renewable), complicating the ability to track carbon footprints (Ren and Wierman 2024).

4.3. Key Challenges in Integrating Data Governance with Environmental Impact Assessments in Gen-AI

Integrating data governance with environmental impact assessments (EIAs) in the Gen-AI value chain presents several key challenges that stem from the rapid development of overall AI technologies, the massive amounts of data involved, and the increasingly important focus on sustainability (Bashir et al., 2024). Challenges include the following:

i. Data Complexity (Quality, Transparency, Volume, and Integrity)

Gen-AI relies heavily on both structured and unstructured data, which can be stored in various formats and siloed systems (Harlin et al., 2023). Effective data governance is needed to ensure that unstructured data is appropriately labeled, categorized, and utilized in environmental evaluations and the integration of environmental metrics into assessments (UNCTAD, 2024a).

Managing large-scale AI systems requires significant data from various sources, while ensuring transparency and accountability in data governance, and aligning with environmental impact standards, which is a difficult task (Raman et al. 2024). The complexity of this task is particularly challenging for frontier technologies such as Gen-AI, where there's often a lack of clear

frameworks for assessing how environmental impacts are calculated across complex, multi-stakeholder data environments (Bashir et., al. 2024).

Gen-AI models also undergo continuous retraining and fine-tuning, which implies repeated cycles of data usage, requiring significant energy consumption with each retraining cycle (Luccioni et al., 2024). Effective EIAs must account for the repeated energy demands of retraining models. If data governance structures don't extend to model lifecycle management, the environmental impact of maintaining large-scale Gen-AI models can be underestimated.

In addition, the complexity of sourcing data from multiple channels makes it challenging to establish clear data lineage and traceability. A lack of transparency regarding data origins can lead to inaccuracies in environmental assessments, highlighting the need for comprehensive data governance practices. Tracking the lifecycle of data used in the Gen-AI value chain is essential for understanding its environmental implications (Thelisson et al. 2023).

Addressing the multidimensional aspects of interoperability is critical in ensuring the accuracy and reliability of data used in environmental assessments, inconsistencies in data quality can arise from disparate sources, leading to incomplete or misleading evaluations of Gen-AI's environmental impact (World Bank, 2022). Without robust data governance frameworks, organizations may struggle to maintain high standards of data integrity, resulting in flawed assessments and decision-making (OECD,2022).

Regulatory and Ethical Framework Gaps

In many regions, there are clear regulations regarding AI ethics, but few that tie AI development to environmental sustainability goals such as carbon neutrality. Many organizations have established data management systems that may not seamlessly integrate with new data governance frameworks required for EIAs, as a result, compatibility issues can hinder the effective implementation of data governance practices, making it difficult to incorporate environmental metrics into existing workflows implications (Thelisson et al. 2023). Navigating the complex regulatory landscape surrounding data governance and EIA requires that organizations ensure compliance with various transnational data protection laws while also adhering to environmental regulations. This dual requirement can create challenges in aligning data governance strategies with the specific needs of environmental assessments in the context of Gen-AI, particularly since EIAs depend on well-defined regulatory standards for environmental impact. In the AI domain, the regulatory gaps in measuring energy consumption, carbon emissions, and e-waste can hinder comprehensive environmental assessments (Thelisson et al. 2023).

Existing frameworks often treat AI and environmental governance separately, governance structures focus primarily on privacy, security, and ethical use, but less on sustainability and environmental impact (Bashir et al., 2024). The incoherence leads to a lack of coordinated policies that can address both the digital and environmental aspects together. For example, the European Green Deal emphasizes climate neutrality, but there are no dedicated regulatory bodies focused on aligning AI systems with these environmental goals (Raman et al., 2024).

Furthermore, fragmented data localization and sovereignty laws can create challenges in terms of balancing local regulations with global Gen-AI value chain operations. Accurate environmental impact assessments require transparency in energy consumption data. Without data governance frameworks that enforce energy-use reporting, particularly in cloud computing and distributed systems, it becomes difficult to account for emissions in the value chain (OECD, 2022).

ii. Gen-AI Value Chain Complexity

The Gen-AI value chain involves multiple stakeholders, including data providers, cloud service operators, and hardware manufacturers (Harlin et al., 2023). Governing data across such a complex supply chain is difficult, particularly when environmental standards differ across jurisdictions and industries.

Fragmented supply chains complicate efforts to conduct comprehensive EIAs. For example, data centres in different countries may have varying energy standards, with some relying heavily on non-renewable energy sources.

Without unified data governance, measuring the overall environmental impact across the supply chain becomes inconsistent, data governance must be standardized across stakeholders to ensure accurate and cohesive reporting on environmental impacts at each stage of the Gen-AI value chain (Sebestyén et al., 2021).

iii. Bias and (Un)Fairness

Data used to train Gen-AI models often reflects historical and societal biases, which can be perpetuated in decision-making (Buolamwini and Gebru 2018). Biases present in the training data of Gen-AI models can skew EIA if the data used does not adequately represent diverse ecological contexts or stakeholder perspectives, the resulting assessments may be biased. These biases can lead to climate apartheid, where wealthier nations are better equipped to mitigate and adapt to climate change, while poorer communities suffer disproportionately (Guerreo 2023). Effective data governance must address these biases to ensure fair and equitable evaluations of Gen-AI's environmental impacts (UNCTAD, 2024a).

Environmental datasets may overlook regions in the Global Majority or marginalized communities, this lack of data equity can result in skewed environmental assessments, reinforcing climate injustices where poorer communities, who contribute least to climate change, face the most severe consequences (Dosemagen and Williams 2022).

Additionally, the unequal distribution of resources and AI's reliance on energy-intensive infrastructure create disparities in climate adaptation, favouring wealthier nations with better technological and data governance infrastructures (Thelisson et al. 2023).

Furthermore, global climate governance frameworks, often driven by high-income countries, tend to exacerbate inequalities (Islam and Winkel, 2017). Gen-AI models used in environmental policies may prioritize regions with comprehensive data and advanced infrastructures, leaving vulnerable populations behind (Ahmed 2023). Furthermore, the carbon footprint and e-waste

generated by AI development often affect the Global South, reinforcing existing environmental injustices and imbalances in global climate governance (UNEP 2024; Guerrero 2023).

5. Conclusion

The generative AI (Gen-AI) value chain significantly impacts environmental sustainability through its various stages, each contributing to energy consumption, resource utilization, pollution, and carbon emissions. As the demand for Gen-AI continues to grow, its associated electricity demand is rising, which runs counter to the necessary efficiency gains needed to achieve net-zero greenhouse gas emissions. Gen-AI's relentless demand for computing power and the resulting larger carbon footprints highlight the urgent need for a comprehensive evaluation of the environmental implications of Gen-AI technologies.

While generative AI holds potential benefits for various sectors, its environmental impacts pose significant risks—particularly for marginalized communities in the Global Majority. Addressing these challenges requires a concerted effort to ensure equitable access to technology and participation in decision-making processes that consider the unique needs and vulnerabilities of these populations. To enhance environmental sustainability within the Gen-AI value chain, it is essential to establish robust metrics and indicators that accurately assess its environmental impact. Creating robust metrics includes tracking energy usage, resource consumption, and emissions throughout the value chain of Gen-AI systems.

Addressing bias and fairness in integrating data governance with environmental impact assessments in Gen-AI requires equitable representation in datasets, transparent AI models, and inclusive global climate governance frameworks. The intersection of Gen-AI and environmental policy must prioritize the needs of vulnerable populations to ensure that technological innovation does not exacerbate global climate inequalities or contribute to further climate injustice.

The integration of data-driven approaches and responsible practices will be crucial in steering the Gen-AI value chain towards a more sustainable trajectory, ultimately contributing to a greener and more resilient planet.

6. Multi-stakeholder Recommendations for Policy Action

Stakeholders can work towards a future where the benefits of Gen-AI are realized without compromising ecological integrity or exacerbating social inequalities[6]. By fostering collaboration and prioritizing sustainability in the design, deployment, and governance of Gen-AI technologies, stakeholders can create holistic dialogues that can facilitate the development of comprehensive frameworks that balance socioeconomic growth with environmental stewardship, in a way that mitigates the existing risks and inequities. We recommend the following policy actions:

Develop Comprehensive Sustainability Metrics for Gen-AI: Governments and international organizations should create standardized metrics that measure Gen-AI's environmental impact across the entire value chain, from energy consumption and resource extraction to carbon

emissions and e-waste. These metrics must align with the UN Sustainable Development Goals (SDGs) and consider impacts on biodiversity, ecosystems, and resource sustainability. Tailored to the needs of low- and middle-income countries (LMICs) in the Global Majority, these metrics should be accessible and not add regulatory burdens that reinforce dependency. Policies should also foster local innovation, equitable resource use, and support sustainable digital economies by addressing supply-side and demand-side deficits that exacerbate digital divides.

Support Regionally Relevant Innovation Ecosystems: Policies should incentivize Gen-AI applications in climate change mitigation, adaptation, and loss and damage while fostering regionally relevant innovation ecosystems. Local entrepreneurs, businesses, and academia in the global Majority need to be central to developing green digital economies. National governments should prioritize locally relevant policies that address environmental challenges in LMICs, with support from global organizations like the World Bank, UNCTAD, and UNDP, offering funding, capacity building, and technical assistance to develop sustainable Gen-AI infrastructure.

Strengthen Global AI Governance Frameworks: Establish global governance frameworks that integrate environmental sustainability into AI development. Frameworks must address risks, such as surveillance, privacy concerns, and climate inequalities, to prevent Gen-AI from exacerbating socio-economic or environmental challenges. The Internet Governance Forum (IGF) should facilitate multistakeholder dialogues to include LMICs in shaping global AI standards. Meaningful representation at climate governance discussions such as the United Nations Framework Convention on Climate Change (UNFCCC) should also be addressed to ensure that legal instruments that the Conference of the Parties (COP) adopts support the effective implementation of the Convention, including institutional and administrative arrangements, that support developmental needs of LMICs. This can enhance LMICs' capacity for green-digital transitions, co-creating sustainability standards that respect their unique environmental and socio-economic contexts.

Leverage Official Development Assistance (ODA) for Sustainable Gen-AI: A coordinated and collaborative approach is crucial for meaningful governance structures that address Gen-AI's environmental costs. ODA can support LMICs in developing sustainable AI infrastructure, building local capacity, and creating green jobs. Development assistance should promote self-sustaining innovation by providing and facilitating digital public goods (DPGs) such as open-source AI tools and data access. Development assistance must be revitalised to evolve beyond perpetuating dependency on foreign consultants and exacerbating tied aid, ODA should support the investment in local talent, including the growth of local policy and technical expertise to create an enabling policy and regulatory environment that fosters sustainable, autonomous digital economies in LMICs.

Integrate Circular Economy Principles: Establish policies that promote circular economy practices in the Gen-AI value chain, such as governance to reduce e-waste through hardware reuse and recycling. Circular economy standards should support sustainable sourcing of materials, incentivizing the use of recycled or responsibly sourced alternatives, and partnering with organizations focused on eco-friendly mining. Public awareness campaigns are essential to

educate stakeholders on the environmental benefits of recycling Gen-AI hardware and reducing e-waste.

Implement Environmentally Focused Data Governance: Data governance frameworks should prioritize a just twin transition approach, that prioritises environmental and social impacts, ensures equitable data access, transparency in AI models, and integration of environmental data to mitigate the environmental impacts of Gen-AI. A just twin transition framework merges social equity with environmental responsibility, aiming to guide the development of Gen-AI and frontier technologies in a way that benefits both people and the planet. A just twin transition framework ensures that the advancement of Gen-AI is balanced with social justice and environmental stewardship. It guides Gen-AI innovation in a way that promotes equitable economic opportunities and helps mitigate climate and ecological impacts, fostering a sustainable and fair digital future for all. This governance approach will enforce accountability, and encourage mitigation through risk assessments, with the potential impact of minimizing Gen-AI's ecological footprint throughout its value chain.

Apply Decolonial Socio-Technical Foresight: A decolonial socio-technical foresight approach can empower LMICs to envision and shape their Gen-AI futures according to local priorities. Moving beyond reactive responses to global trends, this approach enables countries in the Global Majority to design Gen-AI ecosystems that align with self-determination, sustainability, and intergenerational justice. It amplifies voices historically marginalized in tech governance and promotes autonomous, resilient digital futures rooted in local contexts and aspirations. Adopting a decolonial approach to sociotechnical foresight is crucial as it acknowledges that multilateral institutions, global governance, and international hardware and software supply chains that power the Gen-AI value chain are not operating in a vacuum. A decolonial approach can address the entrenched inherent biases in technology development that often overlook or harm vulnerable populations. By centering historically marginalised communities in the foresight process, stakeholders can ensure that technological advancements contribute to justice and well-being rather than perpetuating historical and existing inequalities. Developing global Gen-AI systems with an awareness of socio-political implications, ensures that ethical considerations of reflexivity and positionality are embedded in technical processes.

GLOSSARY

Term	Definition
AI (Artificial Intelligence)	According to the OECD, Artificial Intelligence (AI) refers to "a machine's ability to perform tasks that typically require human intelligence. These tasks include reasoning, learning, problem-solving, perception, language understanding, and the ability to adapt to new situations." AI encompasses a variety of technologies, including machine learning and natural language processing.
AI Value Chain	The AI value chain describes the different stages involved in the development, deployment, and utilization of AI technologies. It includes components such as data collection, model training, deployment, and application, highlighting the interconnected processes that contribute to the creation and implementation of AI systems.
Computational Power	Computational power is a measure of a computer's ability to process data and perform calculations. It is typically quantified in terms of operations per second (OPS) or floating-point operations per second (FLOPS). Higher computational power enables faster processing of complex tasks such as simulations, data analysis, and machine learning algorithms. Factors influencing computational power include the architecture of the hardware (CPUs, GPUs), the efficiency of software algorithms, and the overall system design.
Data-Based Systems	Data-based systems are frameworks or architectures designed to store, manage, and process data efficiently. These systems include databases, data warehouses, and data lakes, facilitating the organization and retrieval of data for analysis and decision-making. They are essential for businesses and organizations to leverage data effectively.
Data Quality	Data quality refers to the condition of a set of values of qualitative or quantitative variables. It encompasses several dimensions, including accuracy, completeness, consistency, reliability, and timeliness. High-quality data is essential for effective decision-making and analysis, as poor data quality can lead to incorrect conclusions and actions. Ensuring data quality involves processes like validation, cleansing, and regular audits to maintain the integrity and usefulness of data in various applications.
Decolonial Theory	This theory critiques colonial legacies and seeks to dismantle structures of power that perpetuate inequality. It advocates for recognizing historical injustices and incorporating diverse perspectives, particularly from those who have been historically

	<p>oppressed. By applying this lens, stakeholders can better understand how technology impacts different social groups and can work towards more equitable outcomes.</p>
<p>Generative Artificial Intelligence (Gen-AI)</p>	<p>Generative Artificial Intelligence (Gen-AI) is a subset of AI technologies that can create new content, such as text, images, music, or video, by learning patterns from existing data. Models like OpenAI's GPT and DALL-E exemplify this technology, which has applications across various fields, including entertainment, marketing, and design.</p>
<p>Indicators</p>	<p>Indicators are calculated measures of performance consisting of a set of different metrics. Indicators can be used to evaluate organizational performance, assist in trend analysis, promote continuous improvement and proactive performance, besides transparent management of processes and staff. They are usually expressed in percent rate or frequency formats.</p>
<p>Just Data Value Creation</p>	<p>Just data value creation emphasizes the ethical and equitable generation of value from data. It advocates for practices that ensure data is used responsibly, promoting fairness, transparency, and inclusivity in data-driven decision-making processes. This approach centers data as a factor of production and seeks to balance the benefits of data utilization with the rights and interests of individuals and communities.</p>
<p>Just Green Digital (Twin) Transition</p>	<p>The term "Just Green Digital (Twin) Transition" typically refers to an equitable and inclusive approach to digital transformation that emphasizes sustainability. It advocates for integrating green technologies and practices into digital initiatives, ensuring that the transition to a digital economy does not exacerbate social inequalities. This concept aims to balance economic growth with environmental stewardship, promoting sustainable practices that benefit all stakeholders in society.</p>
<p>Metrics</p>	<p>Metrics are crude, atomic, and simple composition measures, such as value and quantity formats. They are the basis of any operational follow-up. They are not designed to be used as a basis for strategic decision-making, since they measure more than they actually point to some concrete result.</p>
<p>Socio-technical Foresight</p>	<p>This involves anticipating the social and ethical implications of technological advancements. It requires a holistic approach that considers not just the technical aspects but also the societal dynamics at play. Foresight methodologies aim to identify potential risks and benefits associated with new technologies, enabling proactive measures to mitigate negative impacts.</p>

**Sustainable
Digital
Development**

Sustainable digital development refers to the integration of digital technologies in a manner that promotes economic growth while ensuring environmental protection and social equity. This concept emphasizes the responsible use of digital resources to achieve sustainable development goals (SDGs), addressing issues such as climate change, biodiversity loss, and social inclusion.

APPENDIX 1: Case Studies

The environmental impact of generative AI (Gen-AI) raises significant concerns. While AI and machine learning (ML) can optimize processes and provide solutions for various environmental and climate-related challenges—such as natural disasters, greenhouse gas emissions, biodiversity monitoring, agriculture, and weather modeling—these technologies can also result in negative externalities, including increased resource extraction in certain sectors. To illustrate the benefits and inform best practices while addressing these risks, the following case studies on the use of Gen-AI for environmental conservation, resource optimization, and climate change mitigation are essential:

CASE STUDY 1: Environmental Sustainability and AI for Forest Fire Management in the ASEAN Countries

Fire-Net, code-named KK-2022-026, is a collaborative initiative by Malaysian and Indonesian researchers aimed at automating forest fire detection through satellite imagery. Funded by the Asia Pacific Telecommunity with support from the Republic of Korea, this project addresses the critical issue of forest fires in the Association of Southeast Asian Nations (ASEAN) region, which threaten ecosystems and contribute to harmful transboundary haze affecting public health, particularly in children.

By leveraging AI and remote sensing, Fire-Net enhances the efficiency of fire detection, reducing human error associated with fatigue and enabling continuous updates on affected areas during extended fire incidents. Ultimately, this initiative aims to protect the environment and preserve biodiversity by improving the speed and accuracy of forest fire responses. (ASEAN) region, where uncontrolled fires devastate flora and fauna and create harmful transboundary haze affecting respiratory health, especially in children. Early detection of these fires, while still small, is crucial to minimize environmental damage. Manual satellite image observation, however, is impractical due to the vast areas involved. To address this, an AI-based monitoring system has been developed to automatically detect forest fires by using pixel-based segmentation of satellite images, identifying small fire patches at a resolution of 3 meters, enabling faster response and reducing environmental impact. This AI approach based on Landsat-8 satellite data with three reduced channels of information to allow a faster detection rate, enables the monitoring system to have wide forest coverage with relatively lower operating costs. The system has been developed to enable multiple-size (large and small) forest fire patches to be detected and quantified.

CASE STUDY 2: Environmental Sustainability and AI for Climate Change Management in African Countries

The AI for Climate Action Innovation Research Network, part of AI4D Africa, is dedicated to creating AI solutions for climate adaptation and mitigation throughout Africa, supported by funding

from the Swedish International Development Cooperation Agency and Canada's International Development Research Centre, totaling CA\$1,158,100. The initiative encompasses projects in nine African countries, featuring notable applications such as:

§ Uganda: Utilizing AI and drones to estimate greenhouse gas emissions by mapping livestock and agricultural areas through cost-effective remote sensing.

§ Kenya: Developing an AI mobile app for rapid disease detection in commercial crops, capable of identifying diseases like Taro Leaf Blight via smartphone images.

§ Benin: Applying machine learning techniques to evaluate the vulnerability of mangrove ecosystems to climate change, particularly concerning sea level rise.

§ Cameroon: Researching AI applications to predict renewable energy potential in response to increasing electricity demands, analyzing spatial trends in deforestation and surface water resources.

Overall, these projects aim to enhance environmental sustainability, strengthen research capacity, and support AI policy in sub-Saharan Africa, demonstrating the diverse applications of AI in promoting a sustainable environment.

CASE STUDY 3: Improving Air Quality with Generative AI in Ghana

Ghana, facing significant air pollution challenges, has joined other African countries in adopting low-cost air quality sensors, with The Sensor Evaluation and Training Centre for West Africa (Afri-SET, 2024) leading efforts to enhance data quality. Afri-SET's approach tackles three key challenges: 1) standardizing disparate data from various sensors using generative AI for a unified, manufacturer-agnostic database, 2) automating data integration to reduce manual processing, and 3) embedding human-in-the-loop validation to maintain data integrity. The workflow involves three phases: ingestion via Amazon S3, transformation through AI-generated Python code, and storage in a standardized format for analysis using AWS tools like Glue and Athena. This system enhances efficiency, reduces costs by generating reusable code, and improves air quality monitoring, empowering communities with reliable data to support policy and social impact across Ghana and potentially beyond.

CASE STUDY 4: A Study on the Environmental Impact of Generative-AI

Berthelot et al. (2024) used a multi-criteria life cycle assessment (LCA) to evaluate the environmental impact of generative AI services, especially high-energy applications like conversational agents and image generation models. Their findings highlight substantial environmental costs, with the stable diffusion model emitting 360 tons of CO₂ equivalent and consuming 8.93 million megajoules of energy annually. The study reveals that not only the training phase but also the inference phase and related infrastructure, like data centers, networks, and user devices, contribute significantly to environmental impact. Sensitivity analyses show that

frequent model retraining and heavy data center usage exacerbate these effects. The authors recommend an integrated approach between hardware developers and AI providers to create more efficient AI systems, emphasizing full life-cycle assessments to better understand and mitigate generative AI's ecological footprint.

APPENDIX 2: Gen-AI Value Chain Analysis (VCA) vs Life Cycle Assessment (LCA)

The data collection processes for Value Chain Analysis (VCA) and Life Cycle Assessment (LCA) differ significantly in their approaches, methodologies, and the types of data they prioritise, as follows:

1. Data Collection in Value Chain Analysis (VCA)

VCA emphasizes economic activities and value creation, i.e. the interrelated activities that create value from raw material extraction to the final product delivery. Data collection in VCA typically involves gathering information on each stage of the value chain, including production, distribution, marketing, and sales, to name a few (Investopedia, 2024). This includes data on input costs (e.g., raw materials, labour) and output prices to assess where value is added. While VCA can include environmental considerations, its primary focus is on socioeconomic value creation. It may not comprehensively account for the environmental impacts of each stage unless explicitly integrated into the analysis (DEFA, 2017).

2. Data Collection in Life Cycle Assessment (LCA)

LCA follows a cradle-to-grave approach, collecting data on every stage of a product's life cycle—from raw material extraction through production, use, and disposal. This thorough approach allows for a more nuanced understanding of a product's environmental footprint, ensures that all environmental impacts are considered. LCA has been widely adopted for the AI ecosystem (Luccioni, et al., 2022), given that it follows standardized methodologies (e.g., ISO 14040 and 14044) that provide guidelines for data collection, ensuring consistency and comparability across assessments (DEFA, 2017). LCA requires extensive data on energy consumption, emissions, resource use, and waste generation for each life cycle stage. This includes primary data from manufacturers and suppliers, as well as secondary data from databases and literature through the collection of both quantitative data (e.g., CO₂ emissions in kg) and qualitative data (e.g., potential environmental impacts). LCA requires detailed, product-specific data on materials, energy use, emissions, and waste for each life cycle stage.

While both LCA and VCA aim to assess the environmental and economic aspects of a product or service, LCA has a more comprehensive environmental focus throughout the entire life cycle, while VCA emphasizes the value-adding activities and economic distribution along the supply chain. LCA provides a comprehensive environmental assessment and helps identify potential trade-offs between different environmental impacts, while VCA emphasises economic value-adding activities and may rely on aggregated data, while LCA adopts a comprehensive approach to assess environmental impacts across the entire life cycle of a product (DEFA, 2017). VCA may rely more on aggregated data and industry averages, especially for upstream and downstream activities. VCA considers the broader economic and social aspects in addition to environmental factors and can identify opportunities for value creation and competitive advantage along the value chain, which may be particularly useful for governance in ecosystems with less AI maturity such as the Global Majority.

Understanding these differences is crucial for effectively utilizing each analysis method to inform sustainability decisions and strategies. Table xx summarizes the key similarities and differences between LCA and VCA in capturing the environmental footprint of Gen-AI models.

Table 1: Summary of Gen-AI Value Chain Analysis (VCA) vs Life Cycle Assessment (LCA)

Aspect	Life Cycle Assessment (LCA)	Value Chain Analysis (VCA)
Scope	Comprehensive, covering the entire life cycle from raw material extraction to disposal.	Focuses on value-adding activities along the supply chain from raw materials to market.
Perspective	Product-oriented, tracing environmental burdens associated with a specific product or service.	Value chain-oriented, considering all activities and actors involved in bringing a product to market.
Data Requirements	Requires detailed, product-specific data on materials, energy use, emissions, and waste for each life cycle stage.	May rely on aggregated data and industry averages, especially for upstream and downstream activities.
Methodology	Follows standardized methodologies (e.g., ISO 14040 and 14044) for consistency and comparability.	Lacks a universally accepted standard methodology, making comparisons between studies challenging. However, VCA may utilize more flexible and varied approaches depending on the specific context.
Pros	<ul style="list-style-type: none"> - Comprehensive environmental assessment - Identification of trade-offs - Comparability across products 	<ul style="list-style-type: none"> - Broader perspective including economic and social technical, and aspects of political economy - Identification of value creation opportunities - Flexibility across industries
Cons	<ul style="list-style-type: none"> - Data-intensive and time-consuming - Sensitive to assumptions and data quality - Limited scope regarding economic and social aspects 	<ul style="list-style-type: none"> - Limited environmental focus - Lack of standardization - Often relies on aggregated data, missing nuances
Application in Gen-AI	Quantifies specific environmental impacts associated with Gen-AI development and use.	Provides insights into broader economic and social implications of Gen-AI technologies across the value chain.

Source: Authors own adapted from various sources

For the aim of this discussion paper, the focus remains on capturing multidimensional dynamics associated with the environmental toll of Gen-AI. However, we acknowledge that in the context of Gen-AI models, a combination of LCA and VCA can provide a more comprehensive understanding of the environmental footprint. By leveraging both approaches, stakeholders can make more informed decisions to minimize the environmental impact of Gen-AI models while maximizing associated benefits.

APPENDIX 3: Comparing Gen-AI vs. Traditional AI Environmental Impacts

Table 2: Summary of Gen-AI vs Traditional AI Impacts

Aspect	Generative AI (Gen-AI)	Traditional AI	Environmental Applications & Impact
Core Functionality	Creates new content (e.g., images, text, videos) from large datasets.	Analyzes existing data to make predictions, decisions, or classifications.	Gen-AI requires significantly larger datasets and compute power, leading to higher energy consumption and resource use.
Model Complexity	Typically involves larger, more complex models (e.g., GPT, DALL-E).	Uses simpler, task-specific models (e.g., recommendation systems, classifiers).	Larger models require more computational power, leading to greater carbon emissions during training and deployment.
Energy Consumption	High energy demand due to the need for massive GPU/TPU clusters during both training and inference.	Energy consumption varies, often lower, with more efficient task-specific models.	Gen-AI's intensive training processes result in higher carbon footprints, contributing to environmental degradation.
Data Requirements	Requires extensive and diverse datasets for effective model training.	Generally requires smaller, more specific datasets.	Gen-AI's demand for large datasets exacerbates resource extraction for data storage and processing infrastructure.

Deployment Needs	Ongoing compute power needed for real-time generation of new content.	Often requires less real-time computation after deployment.	Gen-AI's continuous content generation uses more energy during deployment, increasing overall carbon emissions.
E-Waste	High due to frequent hardware upgrades and obsolescence, especially in data centers.	Relatively lower, depending on hardware requirements.	The fast turnover of hardware in Gen-AI models contributes to increased electronic waste (e-waste).
Environmental Mitigation Potential	Limited in direct environmental applications but can support creativity in climate awareness.	More practical for operational tasks like optimizing resource efficiency or energy grids.	Traditional AI can be more directly applied in reducing energy consumption, optimizing resources, and monitoring ecosystems.
Sustainability Practices	Largely undeveloped, with limited focus on energy optimization.	Some established practices for energy-efficient algorithms and hardware use.	Traditional AI systems are more likely to incorporate energy-saving mechanisms, while Gen-AI is still evolving in this regard.

Source: Authors own adapted from various source

Table 2 outlines how Gen-AI's larger datasets, energy needs, and complex models often lead to higher environmental costs compared to traditional AI, while the potential for sustainable applications is more developed in traditional AI systems.

APPENDIX 4: Priorities for Environmental Sustainability & Responsible Global Gen-AI Governance Initiatives

Figure 2: Framework for Just Twin Transition & Responsible Global Gen-AI Governance

Source: Adapted from CODES, 2022

We propose three priorities to ensure a Just Twin (green-digital) Transition framework for environmental sustainability and responsible global governance of generative AI (Gen-AI) that emphasises aligning digital transformation with sustainable development.

Enablers: The first priority focuses on ensuring the requisite enablers are present to align vision, values, and objectives. Prioritising enablers to support the just twin transition calls for reorienting digitalization's purpose to reflect common values, visions, and objectives that advance environmental sustainability. This priority requires a concerted effort to redefine digital technology's role in supporting the broader goals of social and environmental well-being in the digital age.

Mitigation: The second priority emphasises the importance of sustainable digitalization to minimize environmental and social costs. Addressing the ecological impacts of energy and material consumption in Gen-AI and other data-based systems is crucial. Additionally, a holistic approach to mitigation under a Just Twin Transition approach highlights the need to address social challenges, including unsustainable consumption patterns, unequal access to digital tools, and inequalities in digital skills and capabilities. This priority aims to prevent the exacerbation of multidimensional inequalities and protect against targeted human rights violations, ensuring that digital transformation benefits all communities equitably through concerted efforts such as :

Reverse Tutelage and Pedagogies: Engaging in knowledge exchange where marginalized communities are empowered to influence technology design and implementation, rather than being passive subjects of technological change.

Renewal of Affective and Political Communities: Building solidarity-based networks that can advocate for equitable technology practices and challenge existing power dynamics within technological development

Impact: The third priority, focuses on avoiding a short term approach to global governance of responsible Gen-AI and instead on developing future oriented sustainable impacts that encourage positive intergenerational effects through fostering resilient and equitable digital innovation ecosystems. By integrating sustainability into digital innovation, this priority fosters a convergence of digitalization and sustainable development. The goal is to accelerate environmentally, socially, and economically sustainable progress, leveraging the transformative power of digital technologies for a more just and resilient future.

REFERENCES

Ahmed, S. (2023). AI and the circular economy in Africa: Key considerations for a just transition. <https://www.dataeconomypolicyhub.org/post/aiandthecirculareconomy>

Ahmed, S., & Kirchschräger, P. G. (2024). Governing global existential AI risks: Lessons from the International Atomic Energy Agency. T20 Brazil. https://www.t20brasil.org/media/documentos/arquivos/TF05_ST_05_GOVERNING_GLOBAL_E_X66d7093af049f.pdf

Ahmed, S., Tobing, D. H., & Soliman, M. (2023). Why the G20 should lead multilateral reform for inclusive responsible AI governance for the Global South. <https://www.dataeconomypolicyhub.org/items/why-the-g20-should-lead-multilateral-reform-for-inclusive-responsible-ai-governance-for-the-global-south>

AI4D Africa. (2024). AI for climate action innovation research. <https://africa.ai4d.ai/project/ai-for-climate-action-innovation-research/>

Afri-SET. (2024). <https://afrisets.org/>

Bashir, N., Donti, P., Cuff, J., Sroka, S., Ilic, M., Sze, V., Delimitrou, C., & Olivetti, E. (2024). The climate and sustainability implications of generative AI. MIT Exploration of Generative AI. <https://mit-genai.pubpub.org/pub/8ulgrckc/release/2>

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? 🐦. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAcCT '21), 610–623. <https://doi.org/10.1145/3442188.3445922>

Berthelot, A., Caron, E., Jay, M., & Lefèvre, L. (2024). Estimating the environmental impact of generative-AI services using an LCA-based methodology. *Procedia CIRP*, 122, 707–712.

Bond-Taylor, S., Leach, A., Long, Y., & Willcocks, C. G. (2022). Deep generative modelling: A comparative review of VAEs, GANs, normalizing flows, energy-based and autoregressive models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11), 7327–7347. <https://doi.org/10.1109/TPAMI.2021.3116668>

Brundage, M., Avin, S., Wang, J., Belfield, H., Krueger, G., Hadfield, G., Khlaaf, H., et al. (2020). Toward trustworthy AI development: Mechanisms for supporting verifiable claims. *arXiv*. <https://arxiv.org/abs/2004.07213v2>

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (PMLR), 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html>

Burkacky, O., Patel, M., Pototzky, K., Tang, D., Vrijen, R., & Zhu, W. (2024, March 29). Generative AI: The next S-curve for the semiconductor industry? McKinsey & Company. <https://www.mckinsey.com/industries/semiconductors/our-insights/generative-ai-the-next-s-curve-for-the-semiconductor-industry>

Coalition for Digital Environmental Sustainability (CODES). (2022). Action plan for a sustainable planet in the digital age. *Zenodo*. <https://doi.org/10.5281/ZENODO.6573509>

Crawford, K. (2024). Generative AI's environmental costs are soaring — and mostly secret. *Nature*, 626(8000), 693. <https://doi.org/10.1038/d41586-024-00478-x>

Digital Public Goods Alliance. (2023). Exploring data as and in service of the public good. <https://digitalpublicgoods.net/PublicGoodDataReport.pdf>

Domínguez Hernández, A., Krishna, S., Perini, A. M., Katell, M., Bennett, S. J., Borda, A., Hashem, Y., et al. (2024). Mapping the individual, social and biospheric impacts of foundation models. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, 776–796. <https://doi.org/10.1145/3630106.3658939>

Dosemagen, S., & Williams, E. (2022). Data usability: The forgotten segment of environmental data workflows. *Frontiers in Climate*, 4. <https://doi.org/10.3389/fclim.2022.785269>

Elia, M. (2023). Climate apartheid, race, and the future of solidarity: Three frameworks of response (Anthropocene, Mestizaje, Cimarronaje). *Journal of Religious Ethics*, 51(4), 572–610. <https://doi.org/10.1111/jore.12464>

Ernst and Young (EY). (2023). Sustainable Value Study. https://www.ey.com/en_gl/insights/sustainability/how-can-we-accelerate-climate-action

Gender and Environment Data Alliance (GEDA). (2024). Gender and environment data alliance. <https://genderenvironmentdata.org/>

Gmyrek, P., Berg, J., & Bescon, D. (2023). Generative AI and jobs: A global analysis of potential effects on job quantity and quality. <https://www.ilo.org/publications/generative-ai-and-jobs-global-analysis-potential-effects-job-quantity-and>

González Morales, L., & Orrell, T. (2018). Data interoperability: A practitioner's guide to joining up data in the development sector (60 pp.). *United Nations Statistics Division*. <https://doi.org/10.25607/OBP-1772>

Google. (2024). Environmental Report. <https://sustainability.google/reports/google-2024-environmental-report/>

Guerrero, D. (2023). Colonialism, climate change and climate reparations. *Global Justice Now*. <https://www.globaljustice.org.uk/blog/2023/08/colonialism-climate-change-and-climate-reparations/>

Härlin, T., Rova, G. B., Singla, A., Sokolov, O., & Sukharevsky, A. (2023). Exploring opportunities in the generative AI value chain. McKinsey & Company. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/exploring-opportunities-in-the-generative-ai-value-chain>

IBM. (2023). What are AI hallucinations? <https://www.ibm.com/topics/ai-hallucinations>

Intel. (2024). Generative AI. <https://www.intel.com/content/www/us/en/developer/topic-technology/artificial-intelligence/training/generative-ai.html>

International Science Council. (2024). Climate inequality: The stark realities and the road to equitable solutions. <https://council.science/blog/climate-inequality-the-stark-realities-and-the-road-to-equitable-solutions/>

Islam, S. N., & Winkel, J. (2017). Climate change and social inequality (Working Paper No. 152). United Nations, Department of Economic and Social Affairs.

Janjeva, A., et al. (2024). Semiconductor supply chains, AI and economic statecraft: A framework for UK-Korea strategic cooperation. CETAS. <https://coilink.org/20.500.12592/tmpg9rd>

Kalantzakos, S. (2020). The race for critical minerals in an era of geopolitical realignments. *The International Spectator*, 55(3), 1–16. <https://doi.org/10.1080/03932729.2020.1786926>

Lehuedé, S. (2024). An elemental ethics for artificial intelligence: Water as resistance within AI's value chain. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-024-01922-2>

Luccioni, S., Trevelin, B., & Mitchell, M. (2024). The environmental impacts of AI: Policy primer. *Hugging Face*. <https://doi.org/10.57967/hf/3004>

Lynn, B., von Thun, M., and Montoya, K.(2023). AI in the Public Interest: Confronting the Monopoly Threat. Open Markets Institute. <https://www.openmarketsinstitute.org/publications/report-ai-in-the-public-interest-confronting-the-monopoly-threat>

Meta. (2024). Sustainability Report. https://sustainability.atmeta.com/wp-content/uploads/2024/08/Meta-2024-Sustainability-Report.pdf?_hsmi=331641542

Organisation for Economic Co-operation and Development (OECD). (2022). Measuring the environmental impacts of artificial intelligence compute and applications. https://www.oecd.org/en/publications/2022/11/measuring-the-environmental-impacts-of-artificial-intelligence-compute-and-applications_3ddddd5.html

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M., & Dean, J. (2021). Carbon emissions and large neural network training. <https://doi.org/10.48550/arXiv.2104.10350>

Perkins, T. (2024). Industry acts to head off regulation on PFAS pollution from semiconductors. *The Guardian*. <https://www.theguardian.com/environment/article/2024/aug/24/pfas-toxic-waste-pollution-regulation-lobbying>

Policy Network on AI. (2022). Responsible AI in Africa: Bridging the policy gap through multistakeholder partnerships. *AI4D Africa*. <https://ai4d.ai/blog/responsible-ai-in-africa-policy-network/>

Raman, R., Pattnaik, D., Lathabai, H. H., Kumar, C., Govindan, K., & Nedungadi, P. (2024). Green and sustainable AI research: An integrated thematic and topic modeling analysis. *Journal of Big Data*, 11(1), 55. <https://doi.org/10.1186/s40537-024-00920-x>

Ren, S., & Wierman, A. (2024). The uneven distribution of AI's environmental impacts. *Harvard Business Review*. <https://hbr.org/2024/07/the-uneven-distribution-of-ais-environmental-impacts>

Robbins, S., & van Wynsberghe, A. (2022). Our new artificial intelligence infrastructure: Becoming locked into an unsustainable future. *Sustainability*, 14(8), 4829. <https://doi.org/10.3390/su14084829>

Strubell, E., Ganesh, A., & McCallum, A. (2020). Energy and policy considerations for modern deep learning research. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(09), 13693–13696.

Sebestyén, V., Czvetkó, T., & Abonyi, J. (2021). The applicability of big data in climate change research: The importance of system of systems thinking. *Frontiers in Environmental Science*, 9. <https://doi.org/10.3389/fenvs.2021.619092>

The Global E-Waste Monitor. (2024). The global E-waste monitor 2024. <https://ewastemonitor.info/the-global-e-waste-monitor-2024/>

Thelisson, E., Mika, G., Schneider, Q., Padh, K., & Verma, H. (2023). Toward responsible AI use: Considerations for sustainability impact assessment. <https://doi.org/10.48550/arXiv.2312.11996>

United Nations Conference on Trade and Development (UNCTAD). (2021). Technology and innovation report: Catching technological waves—Innovation with equity. https://unctad.org/system/files/official-document/tir2020_en.pdf

United Nations Conference on Trade and Development (UNCTAD). (2024a). Digital economy report 2024: Shaping an environmentally sustainable and inclusive digital future. <https://unctad.org/publication/digital-economy-report-2024>

United Nations Conference on Trade and Development (UNCTAD). (2024b). Critical minerals boom: Global energy shift brings opportunities and risks for developing countries. <https://unctad.org/news/critical-minerals-africa-holds-key-sustainable-energy-future>

United Nations Environment Programme (UNEP). (2024). Navigating new horizons: A global foresight report on planetary health and human wellbeing. <https://council.science/wp-content/uploads/2024/07/Global-Foresight-Report-2024-FINAL.pdf>

United Nations Industrial Development Organization (UNIDO). (2024). Development dialogue on digital transformation and artificial intelligence. <https://www.unido.org/news/development-dialogue-digital-transformation-and-artificial-intelligence-1>

United Nations Institute for Training and Research (UNITAR). (2024). The global E-waste monitor 2024. <https://ewastemonitor.info/the-global-e-waste-monitor-2024/>

Varoquaux, G., Luccioni, A. S., & Whittaker, M. (2024). Hype, sustainability, and the price of the bigger-is-better paradigm in AI. <https://doi.org/10.48550/arXiv.2409.14160>

World Economic Forum. (2024). *Generative AI governance: Shaping a collective global future*. https://www3.weforum.org/docs/WEF_Generative_AI_Governance_2024.pdf

[1] Digital public goods refer to open-source software, open AI models, open standards, open content, and open data that adhere to privacy and other applicable international and domestic laws, standards, best practices, and do no

harm. The concept of DPGs stems from the economic term “public good” referring to resources and services individuals cannot (or should not) be excluded from. <https://digitalpublicgoods.net/PublicGoodDataReport.pdf>

[2] Designations such as “global Majority”, “low-income and middle-income countries”, “high-income countries” or “developing” are intended for statistical convenience and do not necessarily express a judgement about the stage reached by a particular country or area in the economic development process

[3] See Annex 1 for motivation for a Gen-AI value chain analysis vs life cycle assessment

[4] See Appendix 3

[5] In the context of assessing the environmental impact of the Gen-AI value chain, it is essential to distinguish between metrics and indicators, as both play critical roles but serve different purpose.

[6] See Appendix 4 .

AI Governance, Interoperability, and Good Practices

Policy Network on Artificial Intelligence (PNAI)
Sub-group on AI governance, Interoperability, and Good practices

1. Introduction

Development and uptake of artificial intelligence (AI) systems are proliferating at an unprecedented pace and across sectors. A concerted effort in governing AI is vital to harness the opportunities while managing the challenges and risks as a result of new technologies. As AI is increasingly embedded in our society, **interoperable systems and interoperable governance frameworks that effectively address the risks and impacts become imperative**. It is critical that global governance frameworks encourage interoperability to promote a safe, secure, fair and innovative AI ecosystem.

Interoperability is often understood as the ability of different systems to communicate and work seamlessly together. In this paper, we adopt the PNAI's definition of interoperability in its 2023 report (see appendix). The PNAI **definition of interoperability is broader and includes the ways through which different initiatives including laws, regulation, policies, codes, standards etc to regulate and govern AI across the world could work together** in legal, semantic and technical layers. It is through the interaction of these layers that AI interoperability becomes more effective and impactful.

In this paper we **assess the landscape of AI in view of the interoperability among governance frameworks and give actionable recommendations. Three critical aspects of interoperability** are assessed in this paper: **Legal frameworks** - which strengthen the existing world-wide AI regulatory ecosystem through enhanced coordination and align regulatory frameworks across regions. **Technical standards and interfaces** - which ensure that world-wide AI systems are compatible across platforms and regions, with a focus on aligning technological standards. **Global data frameworks** - that develop a unified world-wide data framework to facilitate sharing of AI training data, while ensuring robust protections for personal data and privacy.

Fostering interoperability among AI systems and governance frameworks will be key to enhancing collaboration and building consensus among diverse global stakeholders. Building on the 2023 work on global AI governance by the Policy Network on AI (PNAI)⁸⁷, we aim to strengthen the

⁸⁷Policy Network on AI, [Strengthening multi-stakeholder approach to global AI governance, protecting the environment and human rights in the era of generative AI - A report by the Policy Network on Artificial Intelligence](#), 2023

multistakeholder voice in the global dialogue on AI by providing critical analysis with local evidence from both the Global North and South.

2. AI governance and Interoperability: Highlights in 2024

Among the numerous AI policies, strategies, frameworks, guidelines, principles, standards and regulations developed and implemented at national and international levels. In 2024, there is an observable strategic blend of innovation-driven and regulation-focused approaches with an increasing emphasis on interoperability, ethical standards, and international cooperation pertaining to AI governance. In this section we review existing AI interoperability policies and highlight common governance issues to be addressed at the global level. Building on the October 2023 PNAI report on AI governance⁸⁸, this report focuses on new developments that took place in the last months of 2023 and in 2024.

[See the 2024 highlights in Appendix](#)

3. AI Governance and Interoperability Policies: Patterns and Gaps Across Jurisdictions

Similar patterns have been observed across jurisdictions while scanning existing policy documents related to AI governance and interoperability. The year 2024 witnessed **a surge of national and regional AI governance interventions**, potentially due to the recent advances in the practical applications of generative AI. There is an increasing emphasis on interoperability and international cooperation to address the challenges posed by AI. AI policies aimed at facilitating interoperability (see appendix) were developed by the private sector, UN agencies, regional organizations, the OECD, and national governments, with an increasing number of them now emerging from the Global South.

These interoperability efforts focus on exchange of standards across various standards bodies; data framework for AI training data; AI safety; privacy and personal data protection; cross-border data transfers; common frameworks for mitigating existing and emerging risks; and transparency obligations for AI system developers and deployers. In addition, establishing oversight bodies⁸⁹ is a growing trend to evaluate AI systems, particularly for high-risk applications. However, building consensus on the global governance of AI and strengthening multistakeholder forums, such as the Internet Governance Forum (IGF), are scarcely mentioned. Most importantly, as mentioned in the UN AI advisory body's final report, efforts towards inclusion of the Global South in ongoing

⁸⁸ Ibid.

⁸⁹Such as an International scientific panel, the Arab AI Council (does not have enforcement power), and EU AI office and ASEAN Working Group on AI Governance and (have enforcement power)

policy conversations, and in the development and adoption of international AI governance frameworks remains limited⁹⁰.

3.1 Gaps

There are many regional and multilateral AI governance frameworks such as the EU AI Act or Global Digital Compact, but there is no **comprehensive global interoperability framework to coordinate the different AI governance frameworks**. This could lead to the erosion of collective efforts and partnerships, and the reduction of resource-sharing to address the need for collaborative support in expertise, funding, and knowledge transfer, ensuring that AI leaves no one behind. National and regional efforts driven by local priorities could lead to fragmented and divergent requirements that are likely to create friction, undermine common governmental objectives, and result in interoperability barriers⁹¹. Cohesive and responsive governance frameworks are critical for tapping into the full benefits of AI systems to society and managing AI-related risks and impacts effectively. **Significant gaps remain in the current efforts that target effective interoperability in AI governance**. Identifying the gaps and reflecting on best ways to close them is the foundation for recommendations for effective international cooperation. Key gaps identified include (for details see Appendix):

1. Lack of a globally accepted mechanism for coordinating regional and multilateral efforts.
2. Inconsistencies in AI interoperability frameworks span ethical, legal, technical and policy domains. This inconsistency is compounded by the content of many interoperability policies, which often lack clear definitions, frameworks, and measures that are essential for practical implementation. Achieving compatibility amongst these frameworks may be feasible in the near term, but establishing true consistency could face long term political challenges.
3. Lack of input from the Global South, this may increase disparity in how AI technologies and governance develop globally, leading to an uneven distribution of benefits.
4. Lack of coordination and consensus on values, principles and objectives for regulating AI.
5. Lack of active collaboration of multiple key stakeholder groups to facilitate effective interoperability of AI governance.
6. Technical incompatibilities prevent cross-border data sharing and hinder international AI collaboration.
7. Ethical inconsistencies exist due to the lack of a shared understanding of AI's societal functions and implications.

⁹⁰UN AI Advisory Body, [Governing AI for Humanity](#), September 2024

⁹¹ Cejudo, G.M., Michel, C.L. [Addressing fragmented government action: coordination, coherence, and integration](#). *Policy Sci* **50**, 745–767 (2017); Mohieldin, M., Wahba, S., Gonzalez-Perez, M.A., Shehata, M. (2023). [The Role of Local and Regional Governments in the SDGs: The Localization Agenda](#). In: Business, Government and the SDGs. Palgrave Macmillan, Cham.

FRAMEWORK FOR COMPARING AI INTEROPERABILITY INITIATIVES

Building on our findings, **three critical aspects of interoperability** warrant further consideration:

Legal frameworks - there is a need to strengthen the existing world-wide AI regulatory ecosystem to enhance coordination and align regulatory frameworks across regions.

Technical standards and interfaces - there is a need to ensure that worldwide AI systems are compatible across platforms and regions, with a focus on aligning technological standards.

Global data frameworks - There is a need to develop a unified worldwide data framework to facilitate sharing AI training data, while ensuring robust protections for personal data and privacy.

In addressing these key aspects, critical questions arise: *What elements of the existing interoperability policies are effective, and which aspects are lacking? What tensions exist within current interoperability models?* This angle to interoperability explores the friction between different frameworks, standards, and regulatory approaches that may hinder effective interoperability.

We adopt a framework to compare interoperability policies and models effectively. The framework builds on recurring patterns that we have observed across different initiatives⁹². These patterns provide a structured approach for identifying differences between them.⁹³

The **key patterns for comparison** include: 1) Objectives of interoperability; 2) Principles and values of interoperability; 3) Top-down vs. bottom-up approaches; 4) Binding nature; 5) Level of integration; and 6) Components of interoperability frameworks. (See Appendix for more information on the patterns⁹⁴)

3.2 Interoperability Framework: Key Requirements

In the following sections, we analyse **legal, technical and data interoperability** and present **effective interoperability instruments, barriers and tensions** under these three areas. These findings are based on our analysis of existing AI interoperability regulations, strategies and initiatives in different parts of the world.

⁹²Cedric (Yehuda) Sabbah, [Framework Interoperability: A New Hope for Global Digital Governance](#), 2024

⁹³It is important to note that the framework can be adapted during the analysis to fit the specific context under consideration.

⁹⁴See definitions in Appendix

3.2.1 Legal Interoperability

Legal interoperability ensures that AI systems operating under different regulatory frameworks so that policies and strategies can work together. This can be achieved for example through clear agreements on managing differences in legislation, or by introducing new legislation⁹⁵. A legal interoperability framework can be the common denominator for interoperability policies in different jurisdictions.^{96 97}

Legal interoperability frameworks define the scope of interoperability, particularly regarding data exchange, privacy and data protection requirements. Legal framework interoperability is the ability of different frameworks to coexist and communicate with one another. It reduces regulatory friction between jurisdictions, advances common policy goals and balances global integration with domestic regulatory autonomy⁹⁸. “Interoperability checks” by policy makers and regulators are key in formulating regulatory interoperability frameworks. The first step in addressing legal interoperability is screening existing legislation to identify interoperability barriers⁹⁹. The second step is evaluating compatibility between the enabling legislation of different organizations and countries to ensure there is coherence between legislations. This will facilitate interoperability between AI systems at lower levels (semantic and technical) and reduce cost and implementation time.

Effective interoperability instruments

Regional and international frameworks provide a degree of policy consistency and governance coherence. Legal AI interoperability efforts in the Global South developing countries including in the regions of Latin America, Africa, Southeast Asia and China are increasingly influenced by regional or international regulations and standards.¹⁰⁰ Reasons for this include concentrated regulatory leadership, soft power and diplomatic forces, economic power, and asymmetry of influence between the Global North and South in shaping international norms. Singapore and Malaysia align with the non-OECD AI Principles, many countries in Latin America refer to the EU AI Act, IOS and U.S. NIST as guidelines and benchmarks. African AI governance initiatives consider best practices both within the region and globally¹⁰¹. China’s AI standards are based on analysis of domestic and foreign AI laws, and strategies¹⁰². Council of Europe (CoE)’s Convention on AI’s Interoperability efforts include technical standards, transparency and accountability of AI

⁹⁵European Commission, [New European Interoperability Framework: Promoting seamless services and data flows for European public administrations](#), 2017

⁹⁶The Regulatory Review, [Improving International Regulatory Cooperation](#), 2022

⁹⁷See Appendix

⁹⁸Cedric (Yehuda) Sabbah, [Framework Interoperability: A New Hope for Global Digital Governance](#) article in Lawfare, 2024

⁹⁹Such as sectoral or geographical restrictions in the use and storage of data and AI systems, different and vague data or AI licence models, over-restrictive obligations to use specific digital technologies or delivery modes to provide service, contradictory requirements for the same or similar business processes, outdated security and data protection needs etc.

¹⁰⁰Particularly the OECD, EU, CoE, UNESCO, African Union, ISO and U.S. NIST.

¹⁰¹For example EU AI Act, Canadian AI and Data Act, UK AI Regulation, UNESCO’s Ethical Impact Assessment etc. See: AU, [Continental Artificial Intelligence Strategy](#), August 2024

¹⁰²Policy Network on AI, [Strengthening multi-stakeholder approach to global AI governance, protecting the environment and human rights in the era of generative AI - A report by the Policy Network on Artificial Intelligence](#), 2023

systems and compliance¹⁰³, and strengthening cooperation to prevent and mitigate risks and adverse impacts on human rights, democracy and the rule of law.

Unified AI regulators are set up or proposed in national, regional and global levels to coordinate AI governance effectively. For example, Singapore has designated its Personal Data Protection Commission (PDPC) as a key regulator for AI, the EU has set up its AI Office and The Arab AI Council coordinates AI initiatives across Arabic member states. The African Union is building intra-African coordination and cooperation mechanisms. The **first-ever international legally binding treaty** CoE AI Convention¹⁰⁴ established the Conference of the Parties to monitor the implementation of the convention. Different jurisdictions have established a **shared understanding of principles and terms** that are central to AI governance.

Collaboration in AI safety Governance. 2024 saw increased coordination in promoting AI safety. The Seoul Declaration established an international network of publicly backed AI Safety Institutes to work on complementarity and interoperability between technical work and approaches to safety.¹⁰⁵ The U.S and EU worked on a shared understanding of AI safety, working together on research, standards and testing to promote safe, secure, responsible and trustworthy AI.¹⁰⁶ China set up AI Safety and Governance Institutes as a platform for dialogue, interoperability, and collaborations within and beyond China.¹⁰⁷

The multilateral resolution of the UN's general assembly. The two UN resolutions, i.e. “Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development” and “Enhancing international cooperation on capacity-building of artificial intelligence” mark a significant milestone of global multilateral collaboration on AI governance¹⁰⁸. They define the principles, values, and scope of the global AI governance framework, including values of safety, security and trustworthy. The UN plays central role in coordinating various regional and national AI governance and starting global dialogues.

[Interoperability barriers](#)

Regulatory fragmentation and divergent requirements. Efforts to develop AI interoperability and regulations at international, regional, and national levels are shaped by differing principles, values, objectives, and priorities, resulting in potential regulatory fragmentation and divergence. For example, the OECD advocates for outcome-based approaches that allow flexibility in cooperation within and across jurisdictions and fosters an environment conducive to interoperable governance. While entities like the Council of Europe and the U.S. Department of State

¹⁰³CoE, [The Framework Convention on Artificial Intelligence](#)

¹⁰⁴ The Council of Europe, [The Framework Convention on Artificial Intelligence](#), 2024

¹⁰⁵UK, [Global leaders agree to launch first international network of AI Safety Institutes to boost cooperation of AI](#), May 2024

¹⁰⁶ European Parliamentary Research Service, [United States approach to artificial intelligence](#), January 2024

¹⁰⁷Chinese AI Safety Network, [Chinese AI Safety Network](#) information website

¹⁰⁸ The 78th session on [Enhancing international cooperation on capacity-building of artificial intelligence](#), July 2024; United Nations General Assembly, A/78/L.49 Seventy-eighth session, Agenda item 13, [Integrated and coordinated implementation of and follow-up to the outcomes of the major United Nations conferences and summits in the economic, social and related fields](#), 11 March 2024

emphasize AI's impact on international human rights¹⁰⁹¹¹⁰. This variety in approaches could lead to fragmented requirements, creating friction in AI development and deployment and compromising interoperability¹¹¹. To mitigate these challenges, it is essential to identify areas of convergence and establish mechanisms for coherent compatibility among various regulatory efforts. The pursuit of having an interoperable framework must also not compromise on the protection of universal fundamental human rights by bringing legal standards to the lowest common denominator.¹¹²

Inadequate multistakeholder involvement. Data and AI governance plans in the Global Digital Compact are not clear on how a truly multilateral and democratic process will be achieved, raising concerns on diverse stakeholder representation and representation of Global South countries. The reformed UN OEWG process and NETmundial+10 multistakeholder statement can be used to guide the design and evaluation of effectiveness of multi-stakeholder participations.¹¹³ UN initiatives have been criticized for not adequately considering the already existing multistakeholder frameworks, such as the Internet Governance Forum. Assessing these frameworks' learnings, current shortcomings and ways to improve could be a first step toward effective multistakeholder involvement in global AI governance.

Lack of details on implementing interoperability. This report did find an increase in concrete and specific interoperability measures, models and legislations in 2024. Most frameworks stay at an abstract level and offer little concrete details. For example how suggested actions such as international collaboration, best practice sharing and capacity building, can be realized and implemented¹¹⁴.

Tensions

Differences in AI governance maturity level create disparity in enforcing of international and regional frameworks. For instance, while there is broad agreement on guiding principles for AI governance (such as fairness, transparency, accountability, or protection of human rights), the level of detail and enforcement varies with some countries offering more robust guidance than others. This disparity can impact the interoperability of AI governance, particularly when principles are interpreted or applied differently in different countries.

Differences in the nature of enforcement - binding or non-binding frameworks. Most interoperability frameworks are non-binding. Using a soft law approach in the form of declarations,

¹⁰⁹ OECD, [Recommendation of the Council on Artificial Intelligence](#), 2019

¹¹⁰ U.S. Department of State, [Risk Management Profile for Artificial Intelligence and Human Rights](#), July 2024

¹¹¹ World Economic Forum, [ChatWTO: An Analysis of Generative Artificial Intelligence and International Trade 2024](#), 2024

¹¹² OECD, [Promoting innovation, protecting privacy](#), March 2017

¹¹³ NETmundial, [NETmundial+10 Multistakeholder Statement: Strengthening Internet governance and digital policy processes](#), 2024

¹¹⁴ One example of interest here is the [use of MoUs by countries such as Singapore](#) to achieve some degree of interoperability on AI:

guidelines and mutual recognition agreements provides flexibility and lower implementation costs but of course also reduces the enforcement power.

Differences in regulatory approaches. Responsibilities and powers of regulators differ from country to country. Lack of a consistent regulatory approach complicates efforts to achieve legal interoperability. In ASEAN, Singapore’s PDPC is the key regulator for AI, Malaysia and Thailand rely on existing agencies (such as data protection authorities and sector-specific regulators) to oversee AI-related issues. While the Philippines’s National Privacy Commission largely oversees AI governance. AI companies emphasise the need for “harmonized, consistent, quick, and clear decisions” on data privacy regulation¹¹⁵.

Differences in risk categorization. Countries are beginning to diverge in the ways they assign risk levels of AI systems. Risks and impacts are inevitable with the use or development of new technologies. Given that this is the case; a risk and impact framework is needed. Even if most countries agree with this point, finding agreement on how to categorize these risks is a challenge. Absence of a unified and widely accepted international or cross-industry risk categorization framework presents a challenge. Disagreements can arise if AI risks are defined vaguely¹¹⁶. AI systems are often classified and regulated differently in different countries, leading to inconsistencies in compliance requirements and audits. Objective and legally tenable standards for deciding when an AI system is determined to pose a risk are needed¹¹⁷.

Operationalizing AI risk is an important aspect of addressing risk categorization. One approach to address this issue would be to develop a taxonomy based on the nature of the risk incurred during the AI life cycle from design, development, deployment and use of generative AI platforms and programs. Singapore categorizes AI risk systems based on their potential impact on individuals and society. China identifies and addresses AI safety risks based on a broad categorization of AI’s inherent and AI associated applications’ safety risks. These risks range from models, algorithms, data, systems, cyberspace, cognitive and ethical risks.¹¹⁸ The US Commerce Department’s National Institute for Standards and Technology (NIST) released a risk management framework (RMF) that addresses risks posed to data privacy and environmental impact as well as to information integrity and security¹¹⁹.

The CoE’s Committee on Artificial Intelligence also develops a legally non-binding methodology for Risk and Impact Assessment of AI Systems from the point of view of Human Rights, Democracy and Rule of Law (HUDERIA) to support the implementation of the Framework Convention on AI¹²⁰.

¹¹⁵EU needs AI, [Europe needs regulatory certainty on AI](#) open letter (Accessed in September 2024)

¹¹⁶AI Risk Repository, [The AI Risk Repository](#) information web page (Accessed in September 2024)

¹¹⁷Costanza-Chock et al., 2022

¹¹⁸ <https://www.tc260.org.cn/upload/2024-09-09/1725849192841090989.pdf>

¹¹⁹ NIST AI 600-1, *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, (July 2024)

¹²⁰ <https://rm.coe.int/20240704-ecm-9-2024-webinar-huderia/1680b0d26c>

The tension between Global cooperation and local autonomy. Global initiatives can foster global AI governance collaboration. Regional frameworks, such as ASEAN's Guide on AI Governance and the African Union's AI Strategy, emphasize local priorities, which may not always align with global frameworks. Latin America's Santiago Declaration emphasizes the region's aspiration to influence global AI governance, but also highlights local challenges and notes that dependence on foreign technologies may create friction.

3.3.2. Interoperability Among Technical Standards

Technical interoperability enables machine-to-machine communication. For this to occur systems have to adopt the same technology standards for software, physical hardware components.¹²¹ The Open Systems Interconnection (OSI) model created by ISO standardizes communications to ensure interoperability between diverse computing systems¹²². In 2024, several AI initiatives have established standard interoperability frameworks focusing on sustainable development, safety, human rights, and responsible governance of AI systems¹²³.

Technical interoperability focuses on ensuring AI systems can communicate and work together by adopting uniform standards across software, hardware components, and platforms. The key requirement for establishing a technical interoperability framework is adopting common standards across jurisdictions and sectors¹²⁴. Another critical aspect is alignment between international standardization organizations (ISO, IEC, IEEE, and ITU), and ensuring the framework is flexible and adapts to future technological developments. Regular third-party testing, certification, and validation processes are also needed to guarantee that systems from different providers meet these technical interoperability standards.

Effective interoperability instruments

International Collaboration. The UN AI resolutions and EU AI Act encourage Member States to facilitate the development and deployment of internationally interoperable technical tools, standards or practices to seize the opportunities of AI for sustainable development. The Global Digital Compact and the UN High-Level AI Advisory Body also emphasize inclusive international collaboration and ensuring AI standards are adaptable and globally applicable. The first international network of AI Safety Institutes fosters common understanding of AI safety. Additionally, the International Electrotechnical Commission (IEC), International Organization for Standardization (ISO), and International Telecommunication Union (ITU), have cooperated to map AI/machine learning standardization activities to facilitate coordination, mitigate overlaps,

¹²¹World Bank, [Interoperability frameworks](#)

¹²²UK CMA, [Joint statement on competition in generative AI foundation models and AI products](#), July 2024. ISO/IEC 2382:2015 - Information Technology Vocabulary. ISO/IEC 24765:2010 - Systems and Software Engineering Vocabulary.

¹²³ OECD: <https://oecd.ai/en/wonk/evolving-with-innovation-the-2024-oecd-ai-principles-update>

¹²⁴Ibid. Use of open protocols, APIs, and system architecture that enable machine-to-machine communication and data exchange.

and prevent duplicating efforts.¹²⁵ The first ever International AI Standards Summit¹²⁶ hosted by ITU, together with ISO and IEC, responded to a call to action by the UN to enhance AI governance through international standards. Additionally, organizations such as ISO and IEC have developed robust vocabularies to standardize terminology, helping to enhance interoperability across regions and sectors.¹²⁷

Regional and National Variations. CEN-CENELEC develops European standards to help manufacturers conform with the upcoming Artificial Intelligence Act¹²⁸. The standards will be the first legally binding technical standards for AI systems that could have an international influence. The US-Singapore Dialogue on Critical and Emerging Technologies (CET Dialogue) is a platform for information-sharing and consultations on international AI standards development between the two countries. Interoperability of the countries' frameworks was achieved through a joint mapping exercise between Singapore IMDA AI Verify and US NIST AI RMF.¹²⁹ China has set a goal to participate in the formulation of more than 20 international standards by 2026¹³⁰. US NIST released a plan for global engagement on AI standards. The (draft) Kenya AI standard -Code of Practice for AI Applications respects internationally recognized human rights and labour practices.

Technical Industry Self-Regulation and Technical Integration. The US AI Safety Institute Consortium has brought together over 280 organizations (including for example OpenAI, Google, Anthropic, Microsoft, Meta, Amazon and Nvidia) to develop science-based and empirically backed guidelines and standards for AI measurement and policy. The widespread adoption of machine learning and natural language processing technologies has improved interoperability through better data exchange and better understanding across platforms. These technologies allow systems to interact at multiple levels (both technical and semantic) enhancing communication and data usability.¹³¹

Barriers and Tensions

The absence of widely adopted standards and shared governance frameworks for AI interoperability creates friction between different approaches. Key challenges include inconsistent data quality, lack of standardization, and integration difficulties that can hinder implementation. However, unlike many technologies that rely on interoperable standards (such as railway tracks or the Internet), it is not always necessary to have a common international standard for all technical issues related to AI systems. For example, a joint mapping exercise between Singapore's IMDA AI Verify and the US NIST AI RMF can serve as a compatibility mechanism to achieve interoperability. Additionally, excessive regulation and standardization may limit the deployment of new innovations. For instance, AI learning models and algorithms will

¹²⁵See [World Standards Cooperation](#) information page and [AI/ML landscape of ISO/IEC/ITU-T](#) document (August 2024)

¹²⁶ International AI Standards Summit, 14-18 October 2024, New Delhi, India

¹²⁷See for example ISO/IEC 2382:2015 and ISO/IEC 24765:2010.

¹²⁸ [Artificial Intelligence - CEN-CENELEC](#)

¹²⁹[Singapore and the US to Deepen Cooperation In AI](#) and NIST AI 600-1

¹³⁰Directive on AI standards points the way to 'intelligent' tomorrow.

¹³¹ See Appendix

continue to evolve as computing power advances and research progresses. Therefore, we need to establish a consensus on the scope of technical standards that should be integrated at the global level.

Inconsistencies in the adoption of AI standards across regions. The EU AI Act provides strict binding regulations, while other regions focus mostly on non-binding standards. The Global South faces challenges with infrastructure and connectivity that can limit their ability to meet high foreign AI standards.

Disparity between top-down and bottom-up models of AI standard frameworks. Top-down approaches may lack flexibility, especially in accommodating rapidly advancing technologies. Bottom-up approaches are more flexible but can create governance gaps, particularly concerning ethical and human rights issues.

Difference between binding and non-binding standards. Binding standards provide stronger regulatory enforcement but may conflict with more voluntary frameworks in other regions. However, the two approaches can complement each other: non-binding standards can serve as a foundation for innovation and initial alignment, while binding standards ensure accountability and adherence to essential ethical and regulatory requirements

Unequal Distribution of AI technology. All countries are not developing AI applications at the same rate, this situation creates "AI haves" and "have-nots", further complicating burden-sharing and interoperability efforts. The Global South faces challenges with infrastructure and connectivity that can limit their ability to meet high foreign AI standards.

3.3.3 Data and Privacy Interoperability

While data offers immense benefits for innovation and economic growth, privacy concerns are a major challenge. Collecting personal data brings risks of unauthorized access and misuse. Data security has become a critical issue as especially large data repositories attract cybercriminals. High-profile data breaches have resulted in legal and financial consequences for affected companies, highlighting the need for strong interoperable security protocols and shared incident response strategies¹³².

Shared interoperable privacy standards can ensure that, as personal data is processed, it adheres to a common set of privacy principles everywhere in the world. AI training data often comes from diverse sources (different countries, industries, or formats) and must be usable across multiple AI models and platforms. Standardized data formats, consistent labelling practices, and data quality controls allow AI systems to learn from datasets regardless of origin. Lack of interoperability presents obstacles to efficient data sharing and collaboration. The different regulatory environments add another layer of complexity for MSMEs through varying data protection laws across jurisdictions. Organizations must navigate a complex compliance landscape. Inconsistent privacy standards create barriers that must be overcome to allow global operations. The Global

¹³² <https://ico.org.uk/media/about-the-ico/documents/4031620/ai-in-recruitment-outcomes-report.pdf>

Digital Compact emphasizes the urgent need for strengthened data governance cooperation to maximize the benefits of data use while safeguarding privacy and security¹³³.

Data interoperability ensures that data can be shared and reused across different systems while maintaining consistency, quality, and security. The key requirement for setting up a data interoperability framework is adopting common data formats, metadata standards, and protocols that enable seamless data exchange across platforms. It also requires the establishment of data governance models that define the rules for data access, sharing, and protection, particularly regarding privacy and security concerns. Furthermore, the framework must ensure semantic interoperability (data that is exchanged between systems is understood in the same way) regardless of the systems used or organizations involved. This could be achieved by creating a common benchmark for definitions, such as those that have been applied successfully in data interoperability in the e-invoice framework¹³⁴, or by developing common ontologies and taxonomies through regulatory coordination. The proposed frameworks should consider existing frameworks in different regions and industries. Examples include the African Union (AU) Convention on Cybersecurity and Data Protection¹³⁵, China and France's Joint Statement on Artificial Intelligence and Global Governance¹³⁶, the ASEAN Guide on AI Governance and Ethics¹³⁷, the Santiago Declaration, and the ICC Digital Standards Initiative (DSI)¹³⁸. Any proposed frameworks must be inclusive and context based. Finally, the framework should promote compliance with international data protection regulations and ensure that data interoperability supports cross-border data flows while respecting privacy and security requirements¹³⁹.

We have identified five objectives to address the challenges of global data privacy and interoperability: Prevent Data Protection Disparities and Legal Arbitrage; Harmonize Regulatory Environments; Enhance Transparency; Improve Consumer Redress Mechanisms; and Cross-Border Interoperability for AI Training Data Sharing.¹⁴⁰

Current tensions¹⁴¹

Operational Burden of Data Compliance. Strict data privacy regulations impose significant compliance costs, which can be particularly challenging for Micro, Small, and Medium Enterprises (MSMEs)¹⁴². As a result, MSMEs may be excluded from global AI ecosystems or face non-

¹³³ Global Digital Compact, https://www.un.org/global-digital-compact/sites/default/files/2024-09/Global%20Digital%20Compact%20-%20English_0.pdf

¹³⁴ <https://businesspaymentscoalition.org/wp-content/uploads/20191031-bpc-overview.pdf>

¹³⁵ Malabo Convention. Less than 20 countries in the African continent have ratified it.

¹³⁶ https://www.gov.cn/yaowen/liebiao/202405/content_6949586.htm

¹³⁷ https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf

¹³⁸ <https://www.dsi.iccwbo.org/about-us>

¹³⁹ GDPR.EU, [What is GDPR, the EU's new data protection law?](#), information page (Accessed in September 2024)

¹⁴⁰ See Appendix

¹⁴¹ See Appendix

¹⁴² [Competitive Effects of the GDPR | Journal of Competition Law & Economics | Oxford Academic; Achieving Privacy: Costs of Compliance and Enforcement of Data Protection Regulation;](#)

contextualized regulations. This can hinder innovation, particularly in sectors where AI could contribute to SDG initiatives (such as healthcare, education, or agriculture) and where data sharing is essential. Privacy and security could be inculcated in the system design with the DevSecOps model. DevSecOps makes it possible to move away from the practice of checking ready-made code for compliance with security policies, by introducing control mechanisms at all development stages.

Absence of Data Protection Laws. Many countries, particularly in the Global South, lack comprehensive data protection laws. This creates a barrier to interoperability. AI training data from these regions may not meet the standards required for cross-border data flows with countries that have stricter laws. The absence of international or national legal frameworks limits these countries' ability to participate in global AI research, undermining trust in international data sharing initiatives. This situation prevents these regions from fully leveraging the benefits of AI-driven innovation. In the countries and regions that have strong privacy laws, there is a lack of common understanding of personal data, shared partially open data and public data.¹⁴³ This further complicates international interoperability of data and anonymised data.

Disproportionate Influence of AI Powerhouses. Countries with major AI research hubs may exert disproportionate influence over global standards and frameworks for AI data interoperability. This situation can result in interoperability standards favouring technological capabilities and regulatory frameworks of powerful nations. Potentially at the expense of smaller countries or the Global South. The imbalance of influence might also result in data governance policies that prioritize the commercial and innovation interests of the Global North over global ethical concerns or privacy needs of countries with fewer resources. This dynamic may risk creating an unequal AI ecosystem where only the most powerful nations set the terms for data flows and privacy protections.

Some countries struggle to adopt advanced data and privacy standards or regulations due to limited resources and differing legal infrastructures. This creates a challenge for interoperability in AI data flows, leading to fragmented global data sharing practices. The imposition of one-size-fits-all regulations may also overlook the specific needs of these countries. Stifling AI innovation and hindering the progress toward reaching SDGs; where flexible data usage is critical.

Siloed Data and Resource Limitations. Many organizations across the public and private sectors lack the infrastructure and expertise to implement interoperability solutions. Leading to siloed data and inadequate resources. This limits the overall effectiveness of AI systems¹⁴⁴. Countries face difficulties in developing interoperable AI systems and sharing the data that underpins the technology. Data sharing is often politically sensitive, countries are reluctant to share sensitive information.

¹⁴³Pew Research Center, <https://www.pewresearch.org/internet/2019/11/15/americans-and-privacy-concerned-confused-and-feeling-lack-of-control-over-their-personal-information/>

¹⁴⁴Artificial Intelligence for Interoperability in the European Public Sector

Fragmented Global Security Standards and Geopolitical Tensions: Security concerns are a significant barrier to data interoperability. For instance, recent U.S. policies restrict access to bulk U.S. data by Countries of Concern due to cybersecurity threats¹⁴⁵. Countries may prioritize control over data generated within their borders to protect national interests and citizen privacy. While these measures strengthen local governance of data, they can also create challenges for cooperative data sharing and interoperability efforts. Balancing data sovereignty with international data flows and backing it up by appropriate regulations are essential to foster trust and encourage broader adoption of interoperable frameworks¹⁴⁶.

Best Practice of Compatibility Mechanism

Sharing an understanding of principles and terminology by EU, UK and USA

The three jurisdictions developed a Joint effort on competition in generative AI foundation models and AI products in July 2024 to share concrete understanding of Risks to competition and Principles for protecting competition in the AI ecosystem¹⁴⁷.

Recommendations

A combination of concrete regulatory, governance, technical, and data interoperability mechanisms is needed to support AI interoperability. Here are the recommendations of our multistakeholder group:

General Recommendations

Reaffirm the common objectives and principles of AI governance. AI development and AI interoperability for the safe, secure and trustworthy artificial intelligence systems as outlined in the Global Digital Compact (GDC). Underpinned by the UN General Assembly’s Resolution on “Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development”. These principles include AI that is human-centric, reliable, explainable, ethical, inclusive with full respect to the promotion and protection of human rights and international law, privacy preserving, sustainable development oriented, and responsible¹⁴⁸.

¹⁴⁵ White House, [Executive Order on Preventing Access to Americans’ Bulk Sensitive Personal Data and United States Government-Related Data by Countries of Concern | The White House](#)

¹⁴⁶ Maia Hamin, Alphaeus Hanson, [User in the Middle: An Interoperability and Security Guide for Policymakers](#)

¹⁴⁷ [Joint statement on competition in generative AI foundation models and AI products - GOV.UK](#)

¹⁴⁸ <https://documents.un.org/doc/undoc/ltd/n24/065/92/pdf/n2406592.pdf>

Define priority of interoperability needs on global level. Define and agree what interoperability issues need or need not to be addressed on the global level. Develop a concrete plan to tackle them.¹⁴⁹ Focusing on areas such as AI safety and risk governance, technical standards, data privacy, ethics, AI training datasets and capacity building. Unlike many technologies that rely on interoperability standards (eg railway tracks, the internet), theoretically a lot of issues related to AI can see some degree of divergence. It's more about translating national requirements than necessarily having a common international standard. Singapore IMDA's is mentioned in the report. Likewise, the types of harms associated with fragmentation are very different (ie, if China and the US use different types of risk management framework, this is theoretically workable, unlike different internet standards).

Consider effective interoperability mechanisms identified in this report¹⁵⁰. Use these already existing mechanisms to create a foundation for more cohesive global AI governance.

Establish compatibility mechanisms. Respect regional diversity in AI governance by establishing compatibility mechanisms that can help to reconcile divergence in regulation¹⁵¹. These mechanisms can include mutual recognition of regulatory outcome agreements; reliance on international standards; recognition of comparable protection afforded by domestic law or certificate; and, joint AI safety testing or aligning mandates¹⁵². They can also involve harmonising regulatory frameworks and creating a shared understanding of AI principles and terminology.¹⁵³

Meet Local Needs, Establish Cross-Regional Partnerships, and Interlink Them Globally. Ensure that AI interoperability frameworks are inclusive, adaptable, and capable of addressing specific local challenges while coordinating regional initiatives on a global scale. The UN should collaborate closely with regional bodies, particularly those in the Global South, to develop interoperable mechanisms that foster regional cooperation, mitigate existing disparities, and align efforts at the global level. This approach will strengthen both regional and global cooperation, accommodating varying speeds of collaboration based on regional differences in maturity and public policy priorities.

Combine soft-law and hard law approaches. Introducing the co-regulation model involves using both approaches. Instead of relying solely on either soft or hard law, a combination of

¹⁴⁹ This could include current and emerging safety or security risks related to AI, data and privacy protection, sharing AI training datasets, capacity building etc, focused on issues that have occurred or been observed in practice, and providing specific, consistent, clear mechanisms and methodology to address regulatory gaps, disparities and facilitate inclusiveness, certain, fair and a level playing field for all to benefit from AI. See PNAI's 2023 report.

¹⁵⁰ Such as inclusive multistakeholder platforms at the UN for data governance discussions, where all countries have equal representation and decision-making power, ensuring that the concerns of smaller under-resourced nations are addressed with regards to data flow; interoperability with widely accepted global and international "meta-frameworks"; creation of unified AI regulators; international collaboration in AI safety Governance; technical industry self-regulation and technical integration etc.

¹⁵¹ Yik-Chan Chin and Jingwu Zhao, [Governing Cross-Border Data Flows: International Trade Agreements and Their Limits](#), 2022

¹⁵² Marta Ziosi, Claire Dennis, Robert Trager, Simeon Campos, Ben Bucknall, Charles Martinet, Adam L. Smith, Merlin Stein, [AISIs' Roles in Domestic and International Governance](#), 2024. Jane Drake-Brockman et al., [Digital Trade and the WTO: Negotiation Priorities for Cross-Border Data Flows and Online Trade in Services](#); 2021

¹⁵³ For best practices see Appendix

the two—in the form of multistakeholder, participatory co-regulation with technical AI solutions—is the preferred approach for AI governance.

Commit to diverse and open global multistakeholder engagement in all processes to develop and adopt AI ethics, regulation and standards in all global platforms.

Decentralised multilateralism complemented by multistakeholderism should be enforced to achieve inclusive, transparent and accountable dialogue that can deliver legitimate and effective outcomes¹⁵⁴.

Strengthen the Internet Governance Forum. The IGF, and its multistakeholder structures and mechanisms, should be fully utilized as a platform to support and facilitate discussion on the implementation, monitoring and follow up of the Global Digital Compact.¹⁵⁵ This should be done in collaboration with all UN agencies active in AI governance. To maximize IGF’s potential for delivering concrete outcomes,¹⁵⁶ long-term sustainability needs to be ensured through increased financial, technical and human resources support.

Establish a global AI policy dialogue in the margins of existing structures. This will facilitate exchanges and foster mutual understanding of AI policies, legislation, and best practices across countries and regions. Given that the IGF already supports multistakeholder exchanges and promotes mutual understanding in digital governance, this report recommends positioning the IGF as the ideal platform for this global AI dialogue, with enhanced support from the UN for this process.

Establish Robust Feedback Mechanisms for Continuous Improvement.

AI interoperability governance requires ongoing assessment and iteration to adapt to emerging challenges and technological advancements. Establish feedback mechanisms that enable stakeholders to report practical challenges, regulatory issues, and unintended consequences in real-time. This can be implemented as a structured process within existing multistakeholder platforms such as IGF. Allowing continuous refinement and updating of interoperability guidelines and standards based on real-world insights.

Enhance capacity building in countries that lack resources or expertise. Implement capacity-building programs that provide training and resources to countries and organizations with limited AI development capabilities, in line with the UN’s AI resolution on this matter¹⁵⁷. This will help ensure that all regions can participate in and benefit from AI interoperability efforts. Strengthen UN capacity-building initiatives, especially for the Global South. Create a global capacity-building initiative focused on data governance to help under-resourced countries develop robust data protection frameworks.¹⁵⁸

¹⁵⁴ The UN process of the Global Digital Compact with its open consultations model can serve as best practices.

¹⁵⁵ United Nations, [Pact for the Future, Global Digital Compact, and Declaration on Future Generations](#), September 2024

¹⁵⁶ For example evidence-based policy recommendations, best practice guidelines and pilot projects.

¹⁵⁷ The 78th session on [Enhancing international cooperation on capacity-building of artificial intelligence](#), July 2024

¹⁵⁸ This could be funded by a coalition of governments, international organizations, and private sector partners or the proposed GDC AI Fund.

Assess methodologies of AI risk governance. As AI and its related technologies diffuse across the international system, the volatile, uncertain, complex and ambiguous nature of AI risk necessitates a multistakeholder approach which allows diverse stakeholders to collaborate as the nature of AI risk evolves. The increasing prevalence of Adversarial Artificial Intelligence in the form of AI generated disinformation is one example of this evolution. This necessitates a robust multi-stakeholder dialogue on present and future measures to manage AI risk governance.

Develop Semantic Interoperability. Policy and rule makers must achieve semantic interoperability. This involves a common understanding of key legal definitions and concepts as well as the meaning, intent, nuances, and context of data and actions.

Foster Cross-Border AI Testing and Simulation Environments
Create safe, secure, controlled environments where countries and organizations can collaboratively test AI systems across borders. These testing environments would enable the simulation of AI deployments under varying regulatory frameworks and technical standards, allowing stakeholders to identify interoperability issues before real-world implementation. A cross-border AI testing infrastructure can help establish best practices, increase confidence in AI systems, and ensure that AI technologies perform safely and ethically across diverse jurisdictions.

Recommendations on Legal Interoperability

Leverage global and international regulatory interoperability principles. Policymakers should promote the use of global and international regulatory principles in bilateral, regional, and multilateral agreements. Local regulations need to be able to adapt to cross-border challenges and opportunities, take into account international solutions, and ensure alignment with global standards.

Increase international regulatory cooperation. Strengthening international regulatory cooperation can help regulators address cross-border policy challenges at the right level of governance, limit unnecessary frictions and divergences among regulatory frameworks, and broaden the evidence base for regulation¹⁵⁹. National regulators should strengthen cross-border and pan-industry cooperation. Unnecessary costs and barriers due to different regional requirements should be avoided. This could create an impetus to strengthen regulatory quality and coherence.

Develop global standards for categorizing AI risks. Develop a unified and widely accepted risk categorization framework across jurisdictions to jointly define risk levels for different types

¹⁵⁹(OECD, 2013[11]. Long-standing of such efforts to address transboundary air pollution provide a good example of this (OECD, 2020[9])

of AI systems¹⁶⁰. This involves creating a consensus on precise definitions of AI risks¹⁶¹, and establishing a widely accepted framework for categorizing risk.

Recommendations on Technical Interoperability

Promote global alignment on AI standards. These alignments need to be scientifically grounded and respect international law. Internationally interoperable technical tools, standards or practices need to be developed and deployed through joint international agreements or treaties.

Use AI technologies in initiatives to increase interoperability. Use AI technologies to standardize, clean, and structure data to significantly improve interoperability. AI can facilitate better data integration and sharing, making it easier for different systems to communicate effectively. Develop interoperable platforms that allow different AI systems to work together seamlessly to reduce siloed data and incompatible technologies.

International collaboration in AI technology R&D and deployment: Incentivize joint AI research projects.

Resilient and interconnected Internet infrastructure. In the realm of AI and data governance, ensuring a robust, trustworthy and interconnected internet infrastructure is a cornerstone for successful AI interoperability. A resilient Internet infrastructure is critical to foster a conducive environment for AI innovation and ethical data management. Policymakers should prioritize investments in internet infrastructure that enhance its security, resilience, and capacity to support AI technologies, ensuring trust and interoperability.

Recommendations on Data and Privacy Interoperability

Global Data Framework and International Data Sharing. Develop a global data framework, building on existing international and regional data and privacy protection guidelines. This will facilitate the sharing of AI training data while ensuring robust protections for personal data and privacy. Create international an data commons for AI research, where countries agree to share anonymized, sector-specific datasets (such as in healthcare and transportation) under secure conditions. Mechanisms like data trusts, trusted research environments, and multi-party computation can enable the secure sharing of training data across jurisdictions. These efforts

¹⁶⁰DFRLab, [User in the Middle: An Interoperability and Security Guide for Policymakers](#), 2024

¹⁶¹Risk has several acceptable definitions, such as impact of uncertainty on objectives ISO31000 vs the probability of a negative outcome affecting people, systems, or assets- UNDRR. The UNDRR definition (probability of a negative outcome affecting people, systems, or assets) emphasises direct impact on stakeholders and systems. The ISO 31000 definition (impact of uncertainty on objectives) may complement this, especially for organizational resilience.

align with the Global Digital Compact's objective of achieving interoperable data governance¹⁶².

Interoperability between national data protection legal frameworks and AI governance:

There are several steps that can be taken to strengthen support to all countries to develop effective and interoperable national data governance frameworks. First, Develop consistency and interoperability between national data protection legal frameworks and AI governance efforts. Through mandating transparency obligations of AI system developers and deployers, and data protection impact assessments. Respect data subjects' rights, enable data to flow with trust and mutual benefit. Respect lawful grounds for processing personal data as training data for AI systems.

International organizations' role in data protection regulation. International organizations could lead in developing data protection laws that countries can adopt or adapt, coupled with technical and financial support for implementation. Alternatively, regional or multilateral organizations could pool resources to create cohesive data governance strategies.¹⁶³

Contextualize solutions for data privacy. Current data protection frameworks often fail to consider the unique needs and contexts of different regions and industries. More flexible and adaptive approaches are needed to ensure that data protection does not hinder innovation, particularly in sectors that are vital for development, for example in AI for SDG initiatives. A common understanding of privacy in data is the foundation of data being globally anonymised and shared.

4. Conclusions

While various regulatory frameworks and technical standards exist, significant discrepancies in their requirements, adoption and implementation continue to create challenges. To ensure effective AI interoperability, a set of mechanisms for international compatibility, alignment and coordination is essential. This includes developing universal guidelines that can be reviewed, updated, and endorsed by international organisations. Encouraging contextualised regional collaboration, aligning global, international, regional and national standards, creating compatible instruments, and strengthening multistakeholder engagement and capacity building.

Global multistakeholder cooperation and input are crucial for promoting inclusive governance frameworks and coordinating and AI interoperability efforts across different regions and parts of the world. This discussion paper emphasizes the importance of multistakeholder cooperation in a number of ways. Open and accessible global initiatives like the IGF Policy Network on AI can help identify regulatory and standards gaps, provide inclusive policy recommendations and best

¹⁶² Global Digital Compact, https://www.un.org/global-digital-compact/sites/default/files/2024-09/Global%20Digital%20Compact%20-%20English_0.pdf

¹⁶³For example, the dedicated working group on data governance at the UN's Commission on Science and Technology for Development proposed in the GDC to provide recommendations towards equitable and interoperable data governance arrangements.

practices, and support responsible AI development. Development that prioritizes innovation, interoperability and human rights. Strengthening international cooperation and focusing on shared goals will be vital as we build an interoperable, safe, and sustainable global AI ecosystem.

Authors

PNAI Sub-group on AI governance, Interoperability, and Good practices

Team leaders: Dr. Yik Chan Chin (Beijing Normal University, China); Dr. Chafic Chaya (RIPE NCC, United Arab Emirates); Sergio Mayo (Aragon Institute of Technology, Spain); Amado Espinosa (Medisist, Mexico)

Penholders: Debora Irene Christine (Tifa Foundation, Indonesia); Dooshima Dapo-Oyewole (College of Engineering and College of Business and Economics, Boise State University, USA); Dr. Allison Wylde (Glasgow Caledonian University, UK); Dr. Jin Cui (ITU, China); Antara Vats (National Law School of India University, India); Dr. Patrick M. Bell (National Defense University, USA); Heramb Podar (Center for AI and Digital Policy, India); Jochen Michels (Kaspersky, Germany); Lufuno T Tshikalange (Orizur Consulting Enterprise, South Africa); Muriel Alapini (Benin IGF, Benin); Poncelet Ileleji (Jokkolabs Banjul/NRI, Gambia, Gambia); Xitao Jiang (CNNIC, China); Concettina Cassa (AgID, Italian Government, Italy)

We sincerely thank the following individuals for their valuable comments and contributions: Isabelle Lois, Neha Mirsha, Raymond Forbes, Huw Roberts, Titti Cassa, Dr. Fadi Salem (MBR School of Government, UAE), Azeem Sajjad (Ministry of IT & Telecom, Pakistan), Saba Tiku Beyene (Independent AI Researcher, Ethiopia), as well as all participants in the Sub-group meetings and those who provided feedback during PNAI meetings and workshops.

Appendix

Key Concepts

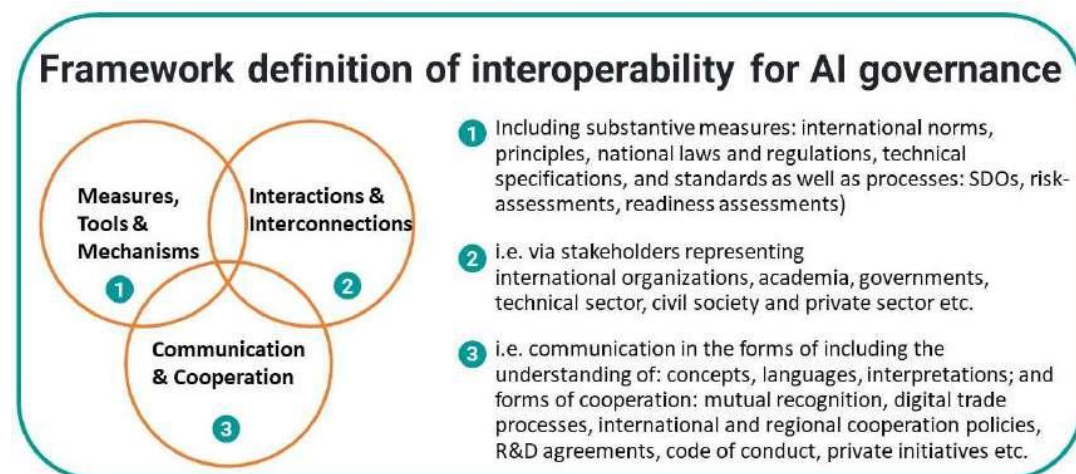
AI Governance: Processes, policies, regulations, and standards that govern the development, deployment, and operation of AI technologies to ensure their ethical, secure, and effective use.

Global AI governance: The process through which diverse interests that transcend borders are accommodated, without a single sovereign authority, so that cooperative action may be taken in maximizing the benefits and mitigating the risks of AI.

Good Practices: Practices that ensure AI systems are developed and used in ways that are ethical, responsible, and beneficial to society. For example: guidelines and strategies that mitigate risks.

Interoperability: The ability of both different AI systems to operate together as well as ability of AI governance frameworks to work together. For example, alignment and coordination of standards, policies and regulations across various jurisdictions. A key factor in ensuring seamless collaboration and data sharing between AI systems, platforms, and components.

Our group's definition of interoperability in AI governance brings together three key aspects: (1) the foundational tools, resources, measures and mechanisms involved in developing and implementing AI, (2) multistakeholder interactions and interconnections and (3) defining a consensus about the mechanisms to communicate and cooperate. All three are necessary to support a common understanding, interpretation and implementation of transborder governance of AI.



AI Governance and Interoperability: Highlights in 2024

The United Nations (UN) General Assembly adopted two resolutions on AI in 2024. The resolution on international cooperation on AI capacity building¹⁶⁴ emphasises that AI should benefit humanity. The resolution encourages international cooperation in strengthening AI capacity building in developing countries. Another landmark resolution¹⁶⁵ promotes development of a regulatory and governance framework to ensure safe, secure and trustworthy AI. A symbiotic relationship between innovation and regulation is emphasised: AI development and application should be safe, reliable, serve the collective interest and protect human rights. Governance measures must be interoperable, flexible, adaptable, inclusive, and based on international law, meet the needs and capabilities of different countries, and ensure fair benefits worldwide.

United Nations' Global Digital Compact (GDC)¹⁶⁶ interoperability related measures and proposals include: collaboration between standards development organizations in interoperable AI standards; cooperation in developing representative high quality data sets, affordable compute resources, and local solutions; increasing access to open AI models and systems, opening training data and compute; facilitating AI model training and development; promoting interoperability between national, regional and international data policy frameworks. GDC proposes establishing a dedicated working group on data governance under the Commission on Science and Technology for Development, a multidisciplinary Independent International Scientific Panel on AI, and a Global Dialogue on AI Governance.

UNESCO has mapped Emerging Regulatory Approaches for AI across the world.¹⁶⁷

United Nations High Level Advisory Body on AI has emphasised inclusivity, public interest, and alignment with established international norms and framework in global AI governance¹⁶⁸. It proposes to enhance “Common Understanding” of AI capabilities and risks, “Common Ground” to establish interoperable governance approaches and “Common Benefits” referring for example to AI’s contribution in reaching the Sustainable Development Goals (SDGs). The High-Level Advisory Body proposes for example setting up a light and agile AI Office in the UN Secretariat to work as “glue” to unite AI initiatives as well as establishing an International Scientific panel on AI¹⁶⁹.

¹⁶⁴ The 78th session on [Enhancing international cooperation on capacity-building of artificial intelligence](#), July 2024

¹⁶⁵ [A/78/L.49 General Assembly](#)

¹⁶⁶ United Nations, [Pact for the Future, Global Digital Compact, and Declaration on Future Generations](#), September 2024

¹⁶⁷ UNESCO, [UNESCO launches open consultation to inform AI governance](#) news article, August 2024

¹⁶⁸ UN, [AI Advisory Body](#) information website, Accessed in September 2024

¹⁶⁹ UN AI Advisory Body, [Governing AI for Humanity](#), September 2024

The African Union. The Continental AI Strategy and the African Digital Compact¹⁷⁰ were endorsed in 2024, final approval is expected in early 2025. The Strategy emphasizes ethical AI use, minimizing risks, and leveraging opportunities for digital advancement. Key components of the AU's AI regulatory landscape include: AU Convention on Cybersecurity and Data Protection¹⁷¹; AfCFTA Digital Trade Protocol adopted in 2024; Collaborative frameworks through the Network of African Data Protection Authorities (NADPA) and other initiatives to harmonise data protection and build public trust in AI. National AI Frameworks (including Tanzania, Ghana, Egypt, Rwanda, Kenya, Nigeria, South Africa, and Mauritius) align with each nation's social and economic contexts and ethical standards. AU Digital ID Framework¹⁷² aims to establish a unified and secure digital identity for African citizens to facilitate access to services and enhance socio-economic development.¹⁷³ Introducing AI technologies in low-resource environments could perpetuate current inequalities and further entrench the already skewed power from global socio-technical systems. The Continental AI Strategy highlights that effective and robust governance is crucial for ensuring that AI technologies serve the interests and development needs of African societies¹⁷⁴. A robust governance regime for Africa will align with existing relevant national legislation and continental framework¹⁷⁵.

ASEAN. The Association of Southeast Asian Nations (ASEAN) Guide on AI Governance and Ethics was published in 2024 and focuses on comprehensive alignment within ASEAN and fostering interoperability of AI frameworks across jurisdictions. The key components of alignment include Internal governance structures and measures; Determining the level of human involvement in AI-augmented decision-making; Operations management; and; Stakeholder interaction and communication.¹⁷⁶ A template for AI Risk Impact Assessment (AI RIA) is recommended to promote interoperability between ASEAN Member States in conducting AI RIA. An ASEAN Working Group on AI Governance will drive and oversee the alignment and interoperability in the region. Guides will be produced by it to address the governance of generative AI on developing a shared responsibility framework. One goal is to gather use cases that demonstrate practical implementation of the Guide.

¹⁷⁰AU, [African Ministers Adopt Landmark Continental Artificial Intelligence Strategy, African Digital Compact to drive Africa's Development and Inclusive Growth](#) press release, June 2024

¹⁷¹Malabo Convention. Less than 20 countries in the African continent have ratified it.

¹⁷²AU, [AU Interoperability Framework for Digital ID](#)

¹⁷³AU Interoperability Framework for Digital ID provides standards and protocols for different digital identity systems to communicate and work seamlessly together. It enables exchanging data securely and integration of ID systems across borders and sectors.

¹⁷⁴AU, [Continental Artificial Intelligence Strategy](#), August 2024

¹⁷⁵Ibid.

¹⁷⁶ASEAN, [ASEAN Guide on AI Governance and Ethics](#)

The Middle East (Arab States)¹⁷⁷. The League of Arab States is developing the Arab AI strategy to coordinate AI initiatives across Member States and to promote knowledge sharing and resources to boost AI development in the region. Both the United Arab Emirates (UAE) and Saudi Arabia have adopted a soft law approach to AI, with focus on guidelines and principles that reflect best practices and interoperability across regions. The UAE developed its AI Strategy for 2031 and established the UAE Council for AI and Blockchain, issued AI Ethics and Principles and Generative AI Guidelines. Dubai created "Digital Dubai" for policy oversight in IT. The Kingdom of Saudi Arabia formed the Saudi Data & AI Authority (SDAIA) and the National Strategy for Data & AI, and aims to be a leading AI economy by 2030. Other Middle Eastern countries are also advancing their AI capabilities: Qatar focuses on AI applications in education and smart city development, while Egypt leverages AI for agricultural advancements. Bahrain and Oman are enhancing their financial services and government efficiency through AI. These initiatives, combined with significant investments in AI education and training aim to build a robust AI talent pipeline and drive economic diversification across the region.

Latin America. The Santiago Declaration¹⁷⁸, forged during a crucial AI summit of high-level authorities from across Latin America and the Caribbean (LAC) in October 2023, underscores a commitment to not only participate in, but to also actively influence the global dialogue on AI. The Declaration highlights a concerted effort from LAC countries to develop governance and regulatory frameworks based on interoperability standards. Columbia chairs an UNESCO committee to implement UNESCO AI Ethics in Latin America. The region's integration into the international technical landscape, coupled with its dependence on foreign investment and technologies, highlights the need for a regulatory approach that is adaptable to both global standards and local realities. Most countries in Latin America are drawing inspiration for their AI bills from the EU AI Act. However, Latin America must consider adapting and refining these ideas to fit its own regulatory, economic and technological landscape. International standards¹⁷⁹ play a pivotal role by providing well-established guidelines and benchmarks to help ensure Latin American AI technologies are globally compatible.

The European Union (EU). The European AI Office¹⁸⁰ was established to oversee AI development across the EU and implementation of the EU AI Act regulation that entered into force in August 2024. The AI Office has engaged stakeholders to help prepare the

¹⁷⁷University of York, [AI regulation and policy landscape in the Middle East](#) news item, March 2024

¹⁷⁸Cumbre Ministerial y de Altas Autoridades de América Latina y el Caribe, [Declaración de Santiago](#). Babl, [Unpacking the Declaración de Santiago: A New Dawn for AI Ethics in Latin America and the Caribbean](#)

¹⁷⁹Set by organizations such as the International Organization for Standardization, see: [ISO/IEC JTC 1/SC 42 Artificial intelligence](#)

¹⁸⁰European Commission, [Commission establishes AI Office to strengthen EU leadership in safe and trustworthy Artificial Intelligence](#) news item, May 2024

first General-Purpose AI Code of Practice.¹⁸¹ It promotes the EU's AI approach internationally, fosters international cooperation, and supports the development of international agreements. Interoperability discussions include technical standards, transparency, and compliance.¹⁸²

Council of Europe (CoE) AI Treaty¹⁸³. Interoperability discussions include technical standards, transparency and accountability of AI systems and compliance. Other proposed efforts include international cooperation in exchanging relevant and useful information and strengthening cooperation to prevent and mitigate risks and adverse impacts on human rights, democracy and the rule of law.

EU, UK & USA have set up joint efforts to promote common understanding of competition risks and principles in generative AI foundation models and AI products.¹⁸⁴

The USA's Executive Order on AI, published in October 2023, mandates increased AI engagement, accelerated AI standards development, and safe, responsible AI deployment. The USA aims to lead global conversations and collaborate on critical infrastructure standards. National Institute of Standards and Technology (NIST) has developed *A Plan for Global Engagement on AI Standards*¹⁸⁵ that focus on terminology, developing metrics and measurements, digital content origins, risk management, security, privacy as well as incident response. The US Federal Trade Commission has conducted a series investigation on AI claims and provided guidelines regarding "Deceptive AI Claims".¹⁸⁶

China. In 2024, China set up two AI Safety and Governance Institutes and Chinese AI Safety Network¹⁸⁷ as platforms for dialogue, mapping, interoperability, and collaborations. Newly published AI Safety Governance Framework¹⁸⁸ promotes broad consensus on a global AI governance system. It unveiled the AI Capacity-Building Action Plan for the Benefit of All¹⁸⁹. China's AI domestic interoperability approach emphasizes technical standardization, open platforms and data sharing, and cross-domain

¹⁸¹European Commission, [The kick-off Plenary for the General-Purpose AI Code of Practice took place online](#), September 2024

¹⁸²European Commission, [European AI Office](#) information web page (Accessed in September 2024)

¹⁸³CoE, [The Framework Convention on Artificial Intelligence](#)

¹⁸⁴CMA, [Joint statement on competition in generative AI foundation models and AI products](#), July 2024

¹⁸⁵NIST, [A Plan for Global Engagement on AI Standards](#), July 2024

¹⁸⁶FTC, [FTC Announces Crackdown on Deceptive AI Claims and Schemes](#), September 2024

¹⁸⁷Chinese AI Safety Network, [zChinese AI Safety Network](#) information website

¹⁸⁸<https://www.tc260.org.cn/upload/2024-09-09/172584919284100989.pdf>

¹⁸⁹Ministry of Foreign Affairs of China, [AI Capacity-Building Action Plan for Good and for All](#), September 2024

application demonstrations in fields that often require interoperability between different systems and platforms, for example healthcare, education, and transportation.¹⁹⁰ It

International interoperability focuses on AI R&D and application; establishing open-source and inclusive AI communities to share best practices and knowledge; AI capacity-building programs tailored for developing countries; diverse AI language and data resources; developing data Infrastructure to fair and inclusive use of global data. AI policy synergy and joint risk management, shared mechanism for AI testing, evaluation, certification, and regulation¹⁹¹.

PNAI Approach to AI Interoperability

Interoperability is often understood as the ability of different systems to communicate and work seamlessly together, this may require there are clear agreements about how to deal with differences across borders. An interoperability framework enables various regulatory systems to coexist and communicate, a critical requirement for cross-border AI applications. This concept is vital in balancing global integration with domestic regulatory autonomy. The development of international agreements, such as the Global Digital Compact highlights ongoing efforts to establish a common framework while accommodating diverse domestic approaches. Such communication includes different levels of integration (technical, conceptual, data format and structure, functionality, etc). We argue that more emphasis should be placed in analysing if and how the different initiatives to regulate and govern AI across the world could collaborate and through that become more impactful.

Key Gaps of AI Interoperability

The rapid development of AI technologies has already begun to exert considerable influence across sectors, including, healthcare, justice, education, cyber-physical systems, autonomous vehicles, employment, and personal privacy. The need for AI integration across the ethical, legal, technical and public policy issues necessitate an examination of existing policies and mechanisms required to address common challenges on a global scale. Without effective

¹⁹⁰四部门关于印发国家人工智能产业综合标准化体系建设指南（2024版）的通知; Article 15,

https://www.gov.cn/zhengce/zhengceku/202407/content_6960720.htm;

科技部等六部门关于印发《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》的通知_国务院部门文件_中国政府网; https://www.gov.cn/zhengce/zhengceku/2022-08/12/content_5705154.htm

¹⁹¹ [Joint Statement between the People's Republic of China and the French Republic on Artificial Intelligence and Global Governance](https://www.linking-ai-principles.org/term/198); <https://www.linking-ai-principles.org/term/198> Article 4, [中华人民共和国和法兰西共和国关于人工智能和全球治理的联合声明](https://www.mfa.gov.cn/eng/xw/zyxw/202407/t20240704_11448351.html)

https://www.mfa.gov.cn/web/ziliao_674904/zt_674979/dnzt_674981/xjpdfsxjxgsfw/zxxx/202405/t20240507_11293821.shtml; [Full text: Shanghai Declaration on Global AI Governance Ministry of Foreign Affairs of the People's Republic of China](https://www.mfa.gov.cn/eng/xw/zyxw/202407/t20240704_11448351.html); https://www.mfa.gov.cn/eng/xw/zyxw/202407/t20240704_11448351.html

governance, the societal implications of AI are likely to intensify as the technology evolves. Despite current developments, significant gaps remain in achieving effective AI governance interoperability.

One major challenge is the absence of a globally accepted mechanism framework that can coordinate regional and multilateral efforts. While individual states and global organizations have developed regional and multilateral frameworks, there is a lack of coordination and consensus on values, principles and objectives for regulating AI. This inconsistency is compounded by many interoperability policies often lacking clear definitions, frameworks, and measures essential for practical implementation. The challenge of interoperability is further complicated since many governance proposals originate from industrial, intergovernmental organisations and regional bodies (the UN, the EU, the US, China, and ASEAN governments) while lacking input from the global south. These initiatives frequently overlook the unique realities and challenges faced by the Global South. Added "unique" for emphasis and adjusted the tone for clarity and specificity. The results may increase disparity in how AI technologies and governance may develop globally, leading to an uneven distribution of benefits.

Additionally, the lack of coordination among regulatory approaches creates further obstacles. Global solidarity and resource-sharing mechanisms are not being adequately leveraged to ensure that AI's benefits are inclusive. Thus, regions that may lack the infrastructure or resources to fully engage in AI development and governance, may risk further marginalization in the global digital economy.

Effective AI interoperability requires the active collaboration of multiple stakeholders, including governments, the private sector, technical community and civil society. However, current initiatives often fall short in terms of comprehensive stakeholder involvement, particularly from underrepresented and marginalized groups. Increased engagement from these groups, supported by initiatives from global organizations like the UN, could help bridge these gaps through creating more inclusive and effective governance structures.

Significant risks associated with a lack of technical incompatibilities as AI systems develop based on different regional standards, platforms, and protocols. This divergence may inhibit cross-border data flows for algorithm training or technical collaboration, resulting in difficulties for international companies to navigate these varying standards. AI systems are being developed according to regional standards, platforms, and protocols that may not be compatible with each other. This incompatibility can prevent cross-border data sharing and hinder international AI collaboration. A model for addressing such challenges through providing a more unified approach to AI governance is provided by the Interoperable Europe Act¹⁹²

Ethical inconsistencies may emerge due to the lack of a shared understanding of AI's societal functions and implications. The lack of a shared ethical framework for AI leads to varying interpretations of principles like fairness, transparency, and accountability. These

¹⁹²[Interoperable Europe act: Council adopts new law for more efficient digital public services across the EU - Consilium](#)

discrepancies mean that what is acceptable in one region might be prohibited in another, creating compliance challenges for multinational AI systems. These inconsistencies lead to fragmented approaches that may erode public trust. Similarly, the lack of semantic interoperability, which is essential for ensuring that different systems can consistently interpret and use data, poses a significant barrier. The development and adoption of common taxonomies will be crucial in creating a shared language for AI applications and ensuring that systems can effectively communicate across borders.

While interoperability is necessary for fostering regulations on transparency, explicability, and accountability, there is also a risk that efforts to achieve consensus may result in watered-down standards. This could compromise critical elements such as human rights if such considerations are not carefully integrated into the regulatory process. Moving towards a singular global set of interoperable standards on AI can also lead to the stifling of innovation and erode public trust in AI systems. Over regulation and standardisation will limit the deployment of new innovations. For Example AI learning models and AI algorithms will constantly develop as computing power becomes accessible and research is achieved.

1. Methodology: framework for comparing AI interoperability Initiatives

The key patterns for comparison include:

- 1. Objectives of interoperability:** This refers to the intended goals of the interoperability framework, such as promoting cross-border data flows, enhancing regulatory coordination, or ensuring the ethical alignment of AI systems.
- 2. Principles and values of interoperability:** This pattern focuses on the foundational principles and values underpinning each interoperability initiative. These may include transparency, accountability, inclusivity, fairness, and respect for human rights, which shape the design and implementation of the interoperability framework.
- 3. Top-down vs. bottom-up approaches:** Interoperability can emerge through different pathways. A bottom-up approach may develop organically, as countries learn from each other and replicate best practices, often through multistakeholder collaborations. Conversely, a top-down approach may involve deliberate decisions by governments or international institutions, which establish a "meta-framework" to coordinate and support domestic frameworks.
- 4. Binding nature:** Interoperability frameworks vary in their legal force. Some manifest as non-binding declarations, taxonomies, or mutual recognition agreements, while others take the form of binding treaties or standards.
- 5. Level of integration:** Interoperability models differ in the degree of specificity they provide. Some frameworks, such as the Internet & Jurisdiction toolkits, offer highly detailed guidelines on how interoperability can be implemented. Others are more flexible and general, aiming for compatibility rather than strict alignment across jurisdictions.

6. Components of interoperability frameworks: Interoperability is not limited to technical standards. A comprehensive framework may include legal, organizational, semantic, and technical dimensions. Addressing all these components is essential to ensure the continued functionality of AI systems in a globally interconnected environment.

2. Primary Objectives States Want to Achieve for Data and Privacy Interoperability

Five primary objectives have been identified to address the challenges of global data privacy and interoperability:

- **Prevent Data Protection Disparities and Legal Arbitrage:** Establish uniform standards to ensure consistent protection of personal data across all jurisdictions, eliminating vulnerabilities caused by regional differences.
- **Harmonize Regulatory Environments:** Reduce fragmentation in global data privacy regulations by fostering alignment between regulatory bodies and promoting common standards, thereby simplifying compliance and enhancing protection.
- **Enhance Transparency:** Ensure clear and accessible information about data collection, usage, and protection practices, empowering individuals to make informed decisions and hold organizations accountable.
- **Improve Consumer Redress Mechanisms:** Implement and communicate clear processes for consumers to file complaints and seek resolutions when their data is mishandled, while also reporting on these issues to identify areas needing stronger protections.
- **Cross-Border Interoperability for AI Training Data Sharing:** Create mechanisms that allow secure cross-border sharing of training data, particularly in high-risk AI systems (e.g., healthcare, financial systems), while respecting national data protection laws.

Promoting Multistakeholder Dialogue on Artificial Intelligence Related Labour Issues

Policy Network on Artificial Intelligence (PNAI)
Sub-group on Labour issues throughout AI's life cycle

1. Introduction

Artificial intelligence (AI) is increasingly becoming a fundamental part of modern society, permeating sectors, such as healthcare, finance, manufacturing, and the service industry. AI is transforming the workforce landscape, for example automating tasks and creating new job roles that demand advanced technical skills. As AI evolves, its impact on labour and employment is of critical concern. As other major technological innovations in the past, AI holds the potential to both enhance and disrupt labour markets all over the world. The transformative capabilities of AI are reshaping industries, leading to both opportunities and challenges for the workforce.

On the one hand, AI's capacity to both complement or substitute tasks previously handled by humans¹⁹³, might be particularly beneficial for workers whose skills are complemented, as they could see a substantial increase in their productivity and income. On the other hand, while it is too early to claim how many jobs have been directly impacted,¹⁹⁴ there are concerns on job substitution (where tasks are entirely taken by AI systems) and displacement (where tasks are replaced with new ones, because they are partially taken by AI systems), as well as unemployment and other labour issues.

Additionally, from a development perspective, AI holds potential to accelerate the achievement of Sustainable Development Goals (SDGs), as well as create jobs in new and emerging sectors such as renewable energy. Examples can be found in waste management, and recycling, mining, and manufacturing industries. Also, expanding use of AI could help close inequalities by integrating traditionally excluded populations into the workforce. Examples of AI supporting digital inclusion of differently abled persons are deaf people using speech to text AI to participate in the labour market more easily, and neurodiverse persons who are provided with content in Easy Read format (in multiple languages including visuals) through generative AI.

¹⁹³ ILO. Information web page: [Artificial Intelligence](#). Accessed in September 2024

¹⁹⁴ There is research that estimates that 84% of tasks in UK central government bureaucratic decision-making processes can be automated to some degree. See: Vincent J. Straub, Youmna Hashem, Jonathan Bright et al. [AI for bureaucratic productivity: Measuring the potential of AI to help automate 143 million UK government transactions](#) (March 2024)

We note that the accelerating development and uptake of AI systems across sectors has translated into new roles and new career paths, for example AI scientists, AI trainers, AI UX developers, AI assisted health and disability care workers, and AI governance specialists. Because of this, vocational and technical training is required, as well as reskilling or upskilling large parts of the workforce around the world, especially in the Global South. Also, some of the new positions, for example those related to data labelling for AI training, are mainly carried out in the Global South and raise new challenges regarding fair working conditions and protection of workers' rights¹⁹⁵.

Besides this, as AI is being used to perform managerial tasks (such as hiring, monitoring, supervising, and training workers) to optimize human resources (HR) processes, issues regarding the role of AI oversight over workers and guaranteeing workers' rights in this new employer-employee relation emerge. A pertinent example would be the increasing use of AI-powered Applicants Tracking Systems (ATS) in resume screening, which may sometimes overlook candidates in the initial screening process.¹⁹⁶

Taking this into account, in the following pages we analyse examples that show how labour issues related to AI are being tackled, propose best practices and review International Labour Organization (ILO), Organisation for Economic Co-operation and Development (OECD), and The United Nations Educational, Scientific and Cultural Organization (UNESCO) recommendations. With this scope in mind, we advise on how the labour market stakeholders (employers, employees, unions, and the government) might carry out valuable dialogue. Our goals are:

- Provide an overview of key concerns, challenges, and opportunities arising from the impact of AI on labour and the workforce
- Review laws, policy, and other global initiatives in relation to AI and labour
- Explore policy recommendations and strategies that can help mitigate some of the labour challenges and appropriately deploy the benefits of AI

Our multi-stakeholder drafting team will provide recommendations on how to better assess the labour related issues pointed out, and how to promote meaningful policy discussions that defend a human-centred and human rights-based development and deployment of AI systems in the labour market.

¹⁹⁵ Adrienne Williams and Milagros Miceli, Essay [Data Work and its Layers of \(In\)visibility](#). (Accessed in September)2024 and (<https://dl.acm.org/doi/10.1145/3415186>, Milagros Miceli and Julian Posada, [The Data-Production Dispositif](#), Proceedings of the ACM on Human-Computer Interaction (November 2022) and Milagros Miceli, Martin Schuessler and Tianling Yang, [Between Subjectivity and Imposition: Power Dynamics in Data Annotation for Computer Vision](#), [Proceedings of the ACM on Human-Computer Interaction](#) (October 20202)

¹⁹⁶ [Claire Cain Miller and Josh Katz \(April 2024\) What Researchers Discovered When They Sent 80,000 Fake Resumes to U.S. Jobs](#), [New York Times](#). Katarina Drucker (2016) [Avoiding Discrimination and Filtering of Qualified Candidates by ATS Software](#). [Glassdoor. Is your ATS discriminatory? blog post \(January 2023\)](#). [Dave Zielinski \(March 2022\) Is Your Applicant Tracking System Hurting Your Recruiting Efforts?](#), [HR Magazine](#). Alex Rosenblat, Tamara Kneese, and Danah Boyd (2014) [Networked Employment Discrimination](#), Data & Society Working Paper

2. Opportunities and Challenges of AI in the Labour Market: A State of the Art

AI has many converging interests of workers, however, without an international labour standard or instrument that addresses emerging technologies, governing AI remains both an opportunity and a challenge. It is important to consider how both arise on the labour market with the use of AI. We highlight the following examples of opportunities and challenges of AI in the labour market:

2.1. Opportunities of AI in the Labour Market

AI's positive impact on worker productivity and competitiveness. Given AI's complementarity with human work, it has the potential to boost productivity and competitiveness of workers and companies. For example, generative AI can boost the performance of highly skilled workers almost by 40% compared with workers who don't use it¹⁹⁷. Additionally, in a recent OECD survey most employees who use AI in their work reported improved performance, improved job enjoyment as well as better mental and physical health¹⁹⁸.

AI-based new occupations and new fields of work. The advent and general deployment of AI around the world will require new roles and occupations to be carried out by humans. This will translate into new jobs in the labour markets. The World Economic Forum (WEF) uses the concept of creative destruction to point out how AI might lead to some jobs disappearing, while at the same time creating new roles and positions. WEF¹⁹⁹ clusters the roles under three categories: 1) Trainers, people involved in developing AI²⁰⁰, 2) Explainers, people making AI easy to use for members of the public²⁰¹, and 3) Sustainers, people guaranteeing AI systems are used as good as possible²⁰².

AI governance empowers workers and addresses current challenges and inequalities in the workplace. AI can be used to reach broader talent pools, reduce biases, and promote diversity in hiring processes²⁰³, additionally AI might empower Diversity, Equity and Inclusion (DEI) processes in HR departments. This is due to reduction of harmful biases in hiring processes, as well as promotion of more neutral analysis of workers, both translating into more inclusive

¹⁹⁷ Dell'Acqua, F., et al (2023) Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. Harvard Business School Draft Paper 24-013. Available at: <https://mitsloan.mit.edu/sites/default/files/2023-10/SSRN-id4573321.pdf>

¹⁹⁸ Lane, M., Williams, M., & Broecke, S. (2023) [The impact of AI on the workplace: Main findings from the OECD Ai survey of employers and workers](#), *OECD Social, Employment and Migration Working Papers*, OECD.

¹⁹⁹ World Economic Forum (2023) Jobs of tomorrow: [Large Language Models and Jobs](#)

²⁰⁰ Roles such as engineers and scientists working on AI.

²⁰¹ Roles such as AI user experience designers, personalized AI assistants, tutors or coaches.

²⁰² Roles such as content creators, data curators, and ethics and governance specialists.

²⁰³ PWC (2024) [How AI is being adopted to accelerate gender equity in the workplace](#)

hiring of groups that have been often excluded from the workforce. In addition, AI could help to include traditionally excluded groups through digital inclusion, such as differently abled people.

AI to empower education and reskilling of workers. AI has the potential to transform teaching and learning practices across different levels and innovate new kinds of teaching and developing skills that are required for life and work in the AI era²⁰⁴. On that note, AI will play a key role in empowering workers and providing them with traditional and new skills to improve their livelihoods and participation in the workplace.

AI supports more capacity in strained sectors such as healthcare, other critical public services and utilities. Increasing demand in the provision of critical public services and utilities requires an ever-growing number of workers to take care of basic needs in countries all over the world. In key sectors, such as healthcare²⁰⁵, there are major deficits of workers that affect the quality and coverage of the provision of fundamental services. On that note, strengthening current workers with AI would improve their capacity and productivity, reducing pressure in these sectors.

2.2. Challenges of AI in the Labour Market

Risk to job quality. Many more jobs will be transformed by AI technology rather than displaced. However, how technology is designed and integrated into the workplace could have consequences for job quality. Algorithmic systems are also deployed to manage workers in their daily tasks at work through tracking devices on workers' computers, phones, work vehicles or other machinery used, or through wearable technology, such as badges. The objective of these tools, sometimes referred to as "HR analytics" or "people analytics", is to collect continuous and real-time data on employees' work in order to inform decisions about task distribution as well as evaluate worker performance. At their worst, these systems could reduce the scope for workers' professional judgement, potentially creating situations where individuals are pressured to work in unsafe or unsustainable ways. Additionally, many of the jobs at risk of displacement in developing countries are in formal waged employment, thus increasing the risk of informality.²⁰⁶

Data workers' conditions. Workers in countries of the Global South are increasingly engaged in data work such as data labelling, cleaning, moderation, tending to the ever-increasing demand for training data for AI systems. In some instances, they have been forced to work long hours in repetitive, monotonous, and precarious conditions²⁰⁷.

Job loss and decrease of income due to automation. As task automation substitutes and replaces workers' activities, this might translate into job losses especially in sectors particularly

²⁰⁴ UNESCO (2019) [Beijing consensus on Artificial Intelligence and education](#)

²⁰⁵ WEF estimates a shortfall of 10 million healthcare workers worldwide by 2030.

²⁰⁶ Gmyrek, P., Winkler, H. & Graganta, S. (2024) [Buffer or bottleneck? Exposure to Generative AI and the Digital Divide in Latin America](#). ILO Working Papers

²⁰⁷ Billy Perrigo (January 2023) [OpenAI Used Kenyan Workers on Less Than \\$2 Per Hour to Make ChatGPT Less Toxic](#), Time magazine

exposed to automation²⁰⁸. Additionally, task automation might lead to less human workforce needed in certain roles, which could translate in less wages and income for workers.

Skills gap and upskilling requirements. AI's impact has been seen with the need to upskill, and adequately train the workforce for new jobs or tasks required as AI proliferates the workspace. While upskilling and retraining is a significant challenge for workers of all ages, there is also a growing risk of deskilling of some workers²⁰⁹.

Mental health and job insecurity. Recent strikes by artists and creative unions in the USA underscore insecurity and fear around AI replacing jobs in creative industries. Additionally, psycho-social effects related to widening use of AI might affect workers. The most frequently noted effects on the psyche of the workforce have negative connotations: Replacing one's own labour force through AI might lead to fear of losing one's job, and stress from reskilling as AI is increasingly capable of tasks that traditionally required human work. However, at the same time, replacing or supplementing the workforce through AI, might have positive psycho-social consequences by giving the individual more space and time for self-fulfilment.

Wage polarization. IMF report²¹⁰ on AI's impact on wages and jobs highlights how professionals in high income jobs are more likely to be both exposed to the impact of AI, and gain higher than proportionate income through it. Additionally, most AI models and Intellectual Property, as well as advanced model development are situated and carried out in the Global North. This situation might lead to AI capacities not being distributed equitably geographically. Both these cases risk polarization of wage and income. In time, this could aggravate the productivity differences between Global North and Global South²¹¹.

Worker surveillance and privacy concerns. Companies are increasingly using AI for worker tracking and oversight tools to monitor their actions and performance. This might lead to unlawful tracking of rests, bathroom, and food breaks especially in warehouse and logistics operations²¹². This might leave workers vulnerable to privacy and surveillance risks in some sectors.

Rise in insecure and irregular work. AI has the potential to alter industrial relations, potentially leading to more insecure contracts such as 'just-in-time' worker contracts, which, as US white house reports²¹³, lead to less inclination of companies to engage in worker training and stable work relationships.

Discrimination and bias in AI systems. There are concerns that some AI systems might have discriminatory outcomes, particularly in recruitment. It is possible for bias to be introduced at

²⁰⁸ Holzer, H. (2022) [Understanding the impact of automation on workers, jobs, and wages](#)

²⁰⁹ Nithya Sambasivan, Rajesh Veeraraghavan (2022) [The Deskilling of Domain Expertise in AI Development](#), CHI 2022

²¹⁰ IMF (2024) [Gen-AI: Artificial Intelligence and the Future of Work](#), staff discussion note

²¹¹ United Nations & International Labour Organization (2024) [Mind the AI Divide: Shaping a Global Perspective on the Future of Work](#)

²¹² [European Commission](#) (2024) [Algorithmic management practices in regular workplaces](#), study by European Commission Joint Research Center

²¹³ Rebecca Stropoli (2023) [AI is going to disrupt the labour market. It doesn't have to destroy it](#), Chicago Booth Review

different stages of the recruitment process, including: sourcing (algorithm makes decision about which candidates are shown online ads); screening (questions posed to potential candidates, sometimes involving games or puzzles that candidates must solve; software trained on past hiring practices to filter potential candidates); interviewing (software conducts video interview and uses sentiment analysis to analyse responses), or background checks (automatically making decisions on candidate's fit based on their social media postings or other online presence)²¹⁴.

2.3. The Role of Unions on Labour Issues Related to AI systems

Unions play a crucial role in integrating the digital dimension into collective bargaining and advocating for regulatory reforms that protect workers' rights in the age of AI. They play a key role in mitigating risks AI brings by ensuring equitable treatment for all workers, regardless of gender, race, political beliefs, or other personal characteristics.

Unions have the potential to play a pivotal role in retraining workers by leveraging their bargaining power and involvement in the processes of technological adaptation across various economic sectors, as well as in the internal management of personnel within organizations. Trade unions themselves need to follow the AI development landscape closely, build an understanding of AI plan proactively, and ensure they have resources and internal AI expertise in their organizations. Particularly, unions have the potential to play a role in the design, implementation, use and integration of technologies at the workplace. For example, German works councils negotiate over the technology that is being adopted, how it is designed and how it is used, and in the process, ensure that the technology benefits work quality in addition to productivity²¹⁵.

Trade unions could actively engage in the uptake, regulation and use of AI and other new technologies in the workplace. They can contribute to technological innovations being adopted in a manner that safeguards fair and equitable working conditions for workers in emerging digital environments.

Unions can lead in key actions such as: forming strategic alliances with social actors who advocate for workers' rights, promoting continuous training to update skills, integrating new competencies for managing AI and other emerging technologies, and ensuring mechanisms for oversight to prevent misuse and ensure ethical, responsible technology use.

As an example, African labour unions have voiced significant concerns regarding the impact of AI on Labour and employment security. Their stance typically addresses several key issues: job displacement, wage pressure, workers' rights, inclusive policy development, social safety nets, and the regulation of AI. Unions in South Africa and across Africa advocate for proactive measures to tackle the challenges posed by AI, emphasizing worker protections, equitable transitions, and

²¹⁴ International Labour Organization (2024) [Challenges and opportunities of digitalization](#)

²¹⁵ Krzywdzinski, M., Gerst, D., & Butollo, F. (2023) [Promoting human-centred AI in the workplace. Trade unions and their strategies for regulating the use of AI in Germany](#). Transfer: European Review of Labour and Research, 29(1), 53-70.

the necessity for inclusive policymaking. Some examples can be found in the South African Mining Sector, Kenya's Digital Economy, the Nigerian Health Sector and South African Transport Sector.

2.4. Current regulation of AI in the labour market

The norm-setting of AI governance to realize justice and equity for the workers must be done in a comprehensive manner that promotes workers' rights as well as innovation. Therefore, the available global Internet governance fora is an ideal scenario for guiding recommendations, discussing common consensus, and exploring possible pathways and modalities to establish AI norm-setting.

Taking the opportunities and challenges into account, and even though there is not a globally binding agreement on AI in the workplace, we recognize valuable advances made in some countries and regions around the world that can provide valuable insights for recommendations to address labour issues related to AI. Namely, we bring forward the following:

- i) The EU AI Act²¹⁶ laid out protections for workers' rights by establishing selected uses of AI systems in the workplace as high-risk. Particularly, the regulation states that certain AI systems used should have special oversight considering the impact they might have on future career prospects of individuals, livelihoods of those persons and workers' rights.
- ii) The US Department of Labor established the AI Principles for developers and employers, which lay out guidelines for the use of AI in the labour market²¹⁷, that promote empowering workers employment in the design, development, testing, training, use, and oversight of AI systems for use in the workplace and ethically developing AI systems to protect workers, that consider AI governance and human oversight.
- iii) In response to Chinese governmental concerns about algorithms controlling the dissemination of news and online content, the Cyberspace Administration of China enacted the Provisions on the Management of Algorithmic Recommendations in Internet Information Services in 2021. The regulation includes extensive provisions for content control and provides protections for workers impacted by algorithms, among other measures. It also establishes the "algorithm registry" for use in future regulatory frameworks. Article 20 of the regulation establishes protections for workers whose schedules and salaries are determined by algorithmic systems.²¹⁸

²¹⁶ [Regulation \(EU\) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence](#))

²¹⁷ US Department of Labor (2024) [Department of Labor's Artificial Intelligence and Worker Well-being: Principles for Developers and Employers](#).

²¹⁸ Cyberspace Administration of China (2021) [Provisions on the Management of Algorithmic Recommendations in Internet Information Services](#)

On that note, we recognize relevant efforts being carried out to develop regulatory frameworks for AI in the labour market, that establish different levels of worker protection in different regions of the world such as the USA, the EU and China. We also recognize that there is existing omnibus regulation that protects workers such as ILO fundamental rights on discrimination, freedom of association and occupational safety and health, which are recognized as human rights and thus relevant to the deployment of AI.

3. Recommendations - promoting workers-led AI governance

In general terms, we underline the importance of promoting workers-led AI governance, that considers the promotion of workers' rights in the AI era as well as innovation and productivity. On that sense, we lay out a framework of recommendations based on the principles of empowerment and participation:

- **Establish frameworks that enable workers and trade unions to actively engage in AI decision-making processes** at the national, regional, and multilateral levels, and promote appropriate opportunities to engage in consultation on AI implementation at company or organization level.
- **Ensure that AI serves as a tool for productivity enhancement** rather than job replacement. Safeguard workers' rights and job security amidst AI proliferation in helping people, businesses, and communities to unlock their potential. Therefore, redeployment should be encouraged when possible. Also, the productivity benefits should be shared with workers and societies.
- **Ethical Frameworks:** Organizations should create codes of conduct that outlines responsibilities and accountability for both workers and management in AI usage, considering transparency, training and support, participatory governance, and protection of rights, laid out in international ethics and legal instruments
- **Incorporate Worker Feedback in designing AI systems:** Involve workers in the design and testing phases of AI systems that they will interact in their work. Incorporating this feedback would also help to improve the efficiency and usability of AI systems developed for the workplace.
- **Establish Joint Committees:** Form committees with equal representation from workers and management to oversee AI integration in organizations, addressing concerns related to labour issues.
- **Encourage Sectoral Open Dialogue:** Organize regular forums for workers to voice concerns and suggestions about AI use in different sectors, building trust and open communication, while promoting workers' rights, productivity, and innovation.

- **Develop both safeguard policy and AI systems** that promote respect to worker autonomy and well-being, gender-mainstreaming, respecting cultural sensitivity, ensuring the worker's freedom of belief and personal cultural and/or religious practices. The AI systems should be interoperable and respond to sectoral specificity. Encourage international organizations and development partners to provide an enabling environment, financial support. Design key pilot projects partnering with relevant government, workers' unions, and employer institutions.
- **Strengthen governance frameworks:** Develop comprehensive, human-centred, international AI governance standards that include clear ethical guidelines and labour protections to apply to algorithmic management and the protection of workers' personal data. Promote AI transparency and accountability to ensure that workers are aware and understand how AI affects their employment, and their performance evaluation and the expected impacts towards their career pathways.
- **Support reskilling and upskilling programs:** Promote the creation of funds and capacity centres for reskilling and upskilling initiatives, particularly in sectors that are specially exposed to job displacement. On that note, stakeholders should work together to provide accessible training on AI-related skills and roles such as data analytics, machine learning, digital tools and even prompt engineering. Partnering with educational institutions, governments, and private sector companies to develop tailored training programs that empower the workforce to adapt to the changing job market, focusing on skills in AI, data management, and technology operation.
- **Mainstream AI use in safeguarding the workers' rights:** Develop guidelines for AI use to address the intersectional issues in workplaces (gender, religious, and cultural context) to be governed by equitable standards.
- **Promote the development of global and uniform standards for monitoring AI's impact on the labour market:** There are multiple heterogeneous methodologies for monitoring AI's impact on labour, leading to inconsistencies on how to measure and tackle this issue. Without standardized metrics and methodologies, it is challenging to measure AI's effects on the workforce accurately and establish policy based on precise data.
- **Support cross-discipline research on labour issues related to AI:** Investigate the key labour challenges and opportunities that arise throughout the various stages of AI's lifecycle, including its creation, implementation, ongoing upkeep, and regulatory oversight.

4. Conclusions

Considering the impact that AI has in different tasks including those managerial ones, it is relevant to recognize the importance of governance frameworks throughout AI's lifecycle to guarantee that all relevant stakeholders are considered, to protect workers' rights while also promoting technological innovation and productivity. Tackling the labour-related issues brought on by AI calls for an all-encompassing, multistakeholder strategy that balances innovation and worker rights protection, following the example of the Internet Governance Forum. To ensure AI benefits the workforce without aggravating current disparities, it is imperative to implement tiered governance for AI, conduct thorough bias testing, develop national policies and norms, and foster international collaboration. To guarantee that AI develops in a way that upholds ethical principles and promotes an inclusive workplace, the PNAI should concentrate on expanding its interactions with global organizations and regional stakeholders as we go forward.

Authors

Policy Network on AI Sub-group on Labour issues throughout AI's life cycle

Team leaders

Olga Cavalli, University of the Defense of Argentina, Argentina

Germán Lopez Ardila, Javeriana University, Colombia

Penholders

Muhammad Shahzad Aslam, Assistant Professor, Xiamen University Malaysia

Azeem Sajjad, Ministry of IT & Telecom, Pakistan

Mary Rose Ofianga, Program Manager, Wadhvani Foundation

Samuel Jose, Children and Youth Major Group to UNEP as individual contributor, Indonesia

Lufuno T Tshikalange, Orizur Consulting Enterprise, South Africa

Shantanu Prabhat, Individual contributor, India

Shobhit S., Individual contributor, India

Christian Gmelin, GIZ, Germany

Julian Theseira, Center for AI and Digital Policy (CAIDP), Malaysia

Mondli Mungwe, Mungwe Industries International, South Africa, Saba Tiku Beyene Independent AI Researcher, Ethiopia

Editing

Maikki Sipinen, PNAI, Finland