Bridging the International AI Governance Divide:

Key Strategies for Including the Global South

Heramb Podar[1]*, Joseph Awuah Baffour[2], Oluwakorede Ajibona[3], Adrian Klaits[4], Omer Alaiashy[5], K. Surya Kailash[6], Shreya Sampath[7], Piyal Uddin8, Elina Haber[9], Safina Soataliyeva[10], Clay Gitobu[11]

[1]Encode India,[2]Encode Ghana,[3]Encode Nigeria,[4]Encode United States of America,[5]Encode Saudi Arabia,[6]Encode India,[7]Encode United States of America,[8]Encode Bangladesh,[9]Encode Lebanon,[10]Encode Uzbekistan,[11]Encode Kenya

Preface: Our Collective Call to Action

As youth representatives and advocates from the Global South, we emphasise that true global AI governance cannot be achieved without our active involvement. Youth leaders from the Global South represent the largest demographic group in the regions most impacted by AI-driven transformations. The risks and opportunities presented by AI are shared across borders, and in accordance with this reality, our collective voices must shape the frameworks that govern this technology. If voices like ours are left out, the frameworks will miss what matters most to the people they're meant to protect.

We call on leaders at the AI Safety Summit and beyond to prioritise understanding bottlenecks across contexts, building partnerships, and co-creating solutions that ensure AI serves all of humanity—responsibly, equitably, and sustainably. The future of AI must be a shared endeavour grounded in mutual respect and a commitment to global solidarity. To this end, we have written a report outlining actionable steps for global leaders to address the risks faced by the Global South and work towards distributing the benefits from AI systems.

Together, we can bridge the digital divide and ensure that AI technologies uplift and protect every community, leaving no one behind.

*Primary Author, Corresponding Email: podar_hd@cy.iitr.ac.in

Signatories:

Heramb Heramb Podar Encode India	Joseph Awuah Baffour, Encode Ghana
Oluwakorede Ajibona, Encode Nigeria	Omer Alaiashy, Encode Saudi Arabia
Piyal Uddin Piyal Uddin, Encode Bangladesh	Elina H. Elina Haber Encode Lebanon
Safina Soataliyeva Safina Soataliyeva, Encode Uzbekistan	Clay Gitobu, Encode Kenya

Executive Summary

With the AI Safety Summit approaching, it is imperative to address the global digital divide and ensure that the voices of the Global South are not just heard but actively incorporated into AI governance. As AI reshapes societies across continents, the stakes are high for all, particularly for the Global South, where the impact of AI-driven inequalities could be devastating.

Including the Global South in AI governance strengthens the effectiveness of global AI regulations, preventing regulatory arbitrage that companies could exploit. It also enhances the legitimacy of international agreements, ensuring they are seen as fair and universally applicable. Moreover, the unique insights from the Global South—such as managing AI risks in informal economies and addressing cultural and linguistic biases—equip the Global North with strategies to make AI systems safer and more adaptable across diverse contexts.

In this report, we address why the Global North should take special care to include the Global South in international AI Governance and what the Global North can do to facilitate this process[1]. We identify the following 5 key objectives pertaining to AI, which are of global importance, with the Global South being a particularly relevant stakeholder.

Objective 1: Establish AI Safety Institutes to Build State Capacity in the Global South
The Global South lacks the infrastructure and regulatory systems needed to manage AI risks. By initiating partnerships and developing localised adaptation blueprints, the Global North can support the establishment of AI Safety Institutes. This will enable proactive governance and safeguard against unregulated AI deployments.
Call to Action: Launch a feasibility study identifying bottlenecks for AI Safety Institutes in the Global South within six months, with plans for implementation within two years.

Objective 2: Coordinate a Global Moratorium on Lethal Autonomous Weapons Systems (LAWS)
LAWS present unique threats to the Global South, where fragile institutions and conflict prevalence make these regions vulnerable to destabilisation. Technical workshops and regional risk assessments can build a strong case for international moratoriums.

[1] Notably, this report does not go into what Global South countries should do regarding these objectives. This will be addressed in an upcoming report by the authors.

Call to Action: Convene technical workshops within four months to inform future negotiations on LAWS governance.

Objective 3: Leverage AI Responsibly for Achieving the Sustainable Development Goals (SDGs)

AI can accelerate progress toward the SDGs but must be deployed ethically. Establishing guidelines and safety audits for AI projects will ensure that development efforts are effective and fair.

Call to Action: Form a working group to draft contextual AI guidelines within three months, with pilot testing by the end of the year.

Objective 4: Safeguard Human Rights, Democracy, and the Rule of Law in AI Governance

Regulatory lag in the Global South puts populations at risk of digital exploitation. Ex-ante Human Rights Impact Assessments and transparency mandates can close this gap.

Call to Action: Establish a task force to explore Human Rights Impact Assessments, with findings presented within 12 months.

Objective 5: Mitigate Language and Cultural Bias in AI Systems

Current AI models often fail to represent the linguistic and cultural diversity of the Global South, leading to systemic inequities. Conducting regulatory stress tests can address these biases and make AI systems globally robust.

Call to Action: Begin stress test simulations within three months, with a comprehensive report within one year.

Summary of Our Recommendations
Objective 1: **Establish AI Safety Institutes to Build State Capacity in the Global South**
● Develop Localised Adaptation Blueprints ● Coordinate Joint Emergency Response Mechanisms ● Form Institutional Partnerships and Fellowships
Objective 2: **Coordinate a Global Moratorium on Lethal Autonomous Weapons Systems (LAWS)**

● Perform Red Teaming Simulations ● Delineate AI Safety Institutes’ Role in Risk Mapping ● Provide Technical Assistance and Information for Weapon Verification Systems
Objective 3: Leverage AI Responsibly for Achieving the Sustainable Development Goals (SDGs)
● Establish Public-Private AI Deployment Guidelines ● Carry Out AI Safety Auditing for Development Projects ● Develop Ethical AI Assessment Frameworks
Objective 4: Safeguard Human Rights, Democracy, and the Rule of Law in AI Governance
● Mandate Interoperable Standards for Global Tech Firms ● Work on Pre-Deployment Safety Reports ● Facilitate the Implementation of ex-ante Human Rights Impact Assessments(HRIAs)
Objective 5: Mitigate Language and Cultural Bias in AI Systems
● Perform Regulatory Stress Testing ● Deploy Monitoring and Evaluation Toolkits ● Conduct Cross-Continental Joint Safety Evaluations

INDEX

Introduction:

The technology of AI is penetrating all sectors across continents, transcending borders and reshaping our shared future. It follows, then, that our dialogue on AI Safety and governance must mirror this boundlessness[2].

From healthcare and education to finance and security, AI's transformative power offers immense potential for human advancement but also poses significant risks if not governed responsibly. As AI continues to evolve, the need for comprehensive, inclusive governance becomes more urgent.

However, the global conversation around AI safety and ethics has often been dominated by a few powerful voices, primarily from the Global North[3]. This lack of inclusivity risks reinforcing existing inequities and exacerbating the digital divide between those with the infrastructure, regulatory frameworks, and economic stability to manage AI risks and those without. For many in the Global South, the challenges are compounded by underdeveloped regulatory systems, data privacy concerns, and a lack of representation in international policy discussions.

The upcoming AI Safety Summit on 10-11 February 2025 in France provides a pivotal moment to address these imbalances. It is an opportunity for stakeholders to recognise that AI's challenges—and its solutions—are global. The Seoul AI Summit has already acknowledged the need for international collaboration and the need for interoperability across frameworks[4]. The safety and ethical standards developed today will shape the lives

[2] "'Irrefutable' Need for Global Regulation of AI: UN Experts | UN News," September 19, 2024, https://news.un.org/en/story/2024/09/1154541.

[3] UNCTAD, ed., *Forging the Path beyond Borders: The Global South* (New York Geneva: United Nations, 2018).

[4] "Seoul Declaration for Safe, Innovative and Inclusive AI by Participants Attending the Leaders' Session: AI Seoul Summit, 21 May 2024," GOV.UK,

of billions, and they must be robust enough to protect vulnerable populations while promoting the equitable distribution of AI's benefits. Bridging the digital divide[5] and ensuring the Global South is not left behind requires cooperation, mutual understanding, and a shared commitment to responsible AI governance.

Why Must The Global North Prioritise Including The Global South In AI Governance?

1. Preventing Regulatory Arbitrage and Ensuring Policy Effectiveness

Fragmented AI regulations open opportunities for regulatory arbitrage, where companies exploit differences in regulatory environments to minimise compliance costs. This undermines global safety standards and creates uneven enforcement landscapes.

- **Case in Point**: In the 1980s and 1990s, insufficient global cooperation allowed waste-exporting practices to flourish, where developed nations shipped toxic waste to developing countries with weaker environmental laws. These countries bore the environmental and health consequences, highlighting the dangers of regulatory gaps[6]. This crisis eventually led to the Basel Convention[7], a coordinated international effort to manage hazardous waste effectively.

  The lesson for AI governance is clear: without global coordination, companies could similarly exploit regions with weak AI regulations, transferring risks and harms to the most vulnerable communities.

2. Mitigating Borderless AI Risks Through Comprehensive Safeguards

AI risks transcend national borders, including data breaches, algorithmic biases, and cyber-attacks. Weak regulations in one region can have widespread repercussions, much like a leaky bucket—where even one gap compromises the whole system.

https://www.gov.uk/government/publications/seoul-declaration-for-safe-innovative-and-inclusive-ai-ai-seoul-summit-2024/seoul-declaration-for-safe-innovative-and-inclusive-ai-by-participants-attending-the-leaders-session-ai-seoul-summit-21-may-2024.

[5] "Widening Digital Gap between Developed, Developing States Threatening to Exclude World's Poorest from Next Industrial Revolution, Speakers Tell Second Committee | Meetings Coverage and Press Releases," https://press.un.org/en/2023/gaef3587.doc.htm.

[6] Jennifer Clapp, "The Toxic Waste Trade with Less-Industrialised Countries: Economic Linkages and Political Alliances," *Third World Quarterly* 15, no. 3 (1994): 505–18, https://www.jstor.org/stable/3993297.

[7] U. N. Environment, "Basel Convention on the Control of Transboundary Movements of Hazardous Wastes | UNEP - UN Environment Programme," December 9, 2011, https://www.unep.org/resources/report/basel-convention-control-transboundary-movements-hazardous-wastes.

- **Case in Point**: The global spread of the 2008 Chinese milk scandal[8] exemplifies the leaky bucket effect. Due to lax regulatory oversight, milk and infant formula contaminated with melamine, a toxic chemical, were initially produced in China. These products were exported and distributed worldwide, affecting more than 300,000 children's health and prompting 24 countries across Africa, Latin America and Asia to place bans[9]. The incident highlighted how weak regulations in one country can have devastating global repercussions, especially when it comes to public health and safety.

3. Strengthening the Legitimacy of International Agreements

International AI governance frameworks that exclude the Global South risk being dismissed as Western-centric or neo-colonial, reducing compliance and hindering global cooperation.

- **Case in Point**: The Global Compact for Migration faced challenges in gaining legitimacy, illustrating that international agreements require buy-in from a diverse array of nations to succeed[10]. Without inclusive participation, especially from both the Global South and the Global North, such frameworks risk being perceived as one-sided or being hit by withdrawals, undermining their effectiveness and global cooperation[11]. AI governance efforts must take note of this. Without the Global South's perspectives, agreements risk being ineffective and resisted, weakening global regulatory efforts.

4. Leveraging Unique Insights to Address Overlooked Risks

The Global South brings critical and often unique insights into AI's dual nature. Regions in Africa, Asia, and Latin America have experienced both the transformative benefits and the new risks technology can bring. It also cannot be ignored that it is the Global South that has the comparative advantage of and context regarding the unique risks and challenges faced by the collective Global South.

- **Case in Point**: Mobile banking in sub-Saharan Africa has significantly improved financial inclusion[12]. Access to M-PESA increased per capita consumption levels

[8] "Melamine-Contaminated Powdered Infant Formula in China - Update 2," https://www.who.int/emergencies/disease-outbreak-news/item/2008_09_29a-en.

[9] Jane Parry, "China's Tainted Milk Scandal Spreads around World," *BMJ* 337 (October 1, 2008): a1890, https://doi.org/10.1136/bmj.a1890.

[10] Carolina Gottardo and Nishadh Rego, "The Global Compact for Migration (GCM), International Solidarity and Civil Society Participation: A Stakeholder's Perspective," *Human Rights Review* 22, no. 4 (December 1, 2021): 425–56, https://doi.org/10.1007/s12142-020-00611-z.

[11] "UN to Adopt Migrant Pact Hit by Withdrawals," France 24, December 10, 2018, https://www.france24.com/en/20181210-united-nations-adopt-migrant-pact-hit-withdrawals-usa-belgium.

[12] Omwansa, T. (2009). "M-Pesa: Progress and Prospects" innovations / Mobile World Congress 2009. Pg 107-123.

and lifted 194,000 households, or 2 % of Kenyan households, out of poverty[13]. However, it has also exposed populations to new financial fraud risks, such as opaque pricing and concentrating power in the hands of a few companies. Policymakers from these regions have developed innovative responses, providing lessons that are unknown or overlooked. Overlooking these experiences is not just an oversight but a strategic error. The Global South can identify risks and solutions that the Global North may not have anticipated, such as impacts on informal labour markets and political disinformation.

5. Alignment of Standards as a Global Public Good

Harmonizing AI standards is a public good that benefits everyone. Coordinated and interoperable standards ensure that AI technologies are developed and deployed in ways that are safe, ethical, and fair across borders[14]. This alignment reduces confusion, promotes collaboration, and provides a stable regulatory environment that fosters innovation while safeguarding against misuse.

When standards are misaligned, it creates inefficiencies and barriers, fragmenting the global AI ecosystem[15]. A unified approach enables seamless cross-border cooperation, enhances trust among stakeholders, and sets a consistent baseline for safety and ethical considerations. The Global North and Global South both stand to gain from this stability, as it minimises the risk of AI-related harms and maximises the technology's benefits for global economic and social development.

- **Case in Point:** For instance, the aviation industry operates on aligned international safety standards, benefiting every country involved by ensuring the safety of global air travel. Organisations such as the International Civil Aviation Organization (ICAO) and the International Air Transport Association (IATA)[16] have established comprehensive frameworks such as the Global Aviation Safety Plan (GASP)[17] that provide a strategic direction for safety management to member states and airlines, ensuring uniform safety measures and operational consistency worldwide. Similarly, AI governance must be approached as a collective endeavour, where the alignment of standards ensures security, trust, and equitable progress for all.

6. Upholding Ethical Responsibility and Global Fairness

The principle of "nothing about us without us" underlines the ethical imperative that decisions affecting communities should not be made without their active involvement. It

[13] "The Long-Run Poverty and Gender Impacts of Mobile Money | Science," https://www.science.org/doi/10.1126/science.aah5309.

[14] "PNAI Report | Internet Governance Forum," https://intgovforum.org/en/filedepot_download/282/28491

[15] Ibid

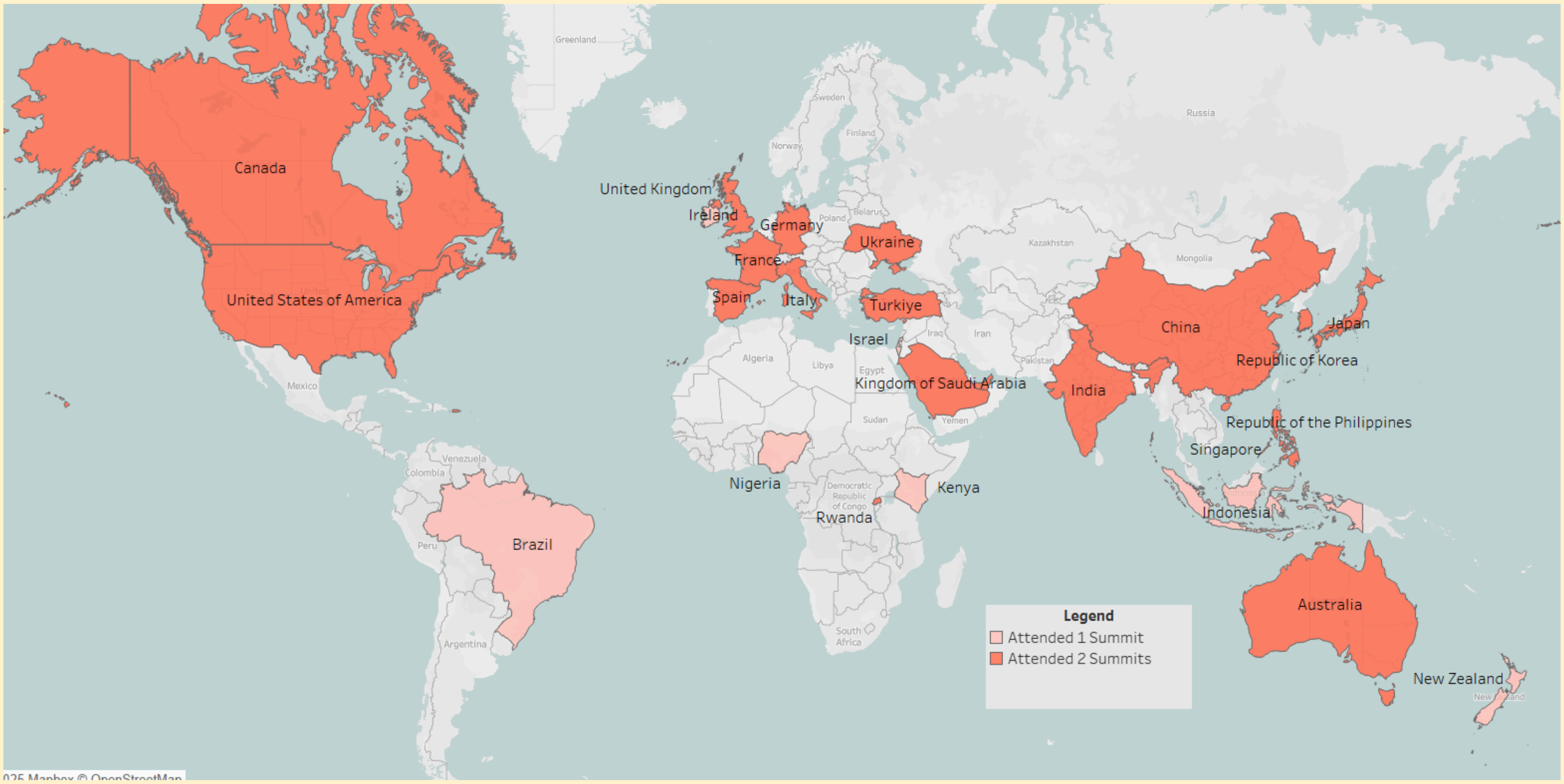
[16] Wragg, David W. (1973). A Dictionary of Aviation (first ed.). Osprey. p. 164.

[17] "Pages - Safety Management," https://www.icao.int/safety/safetymanagement/Pages/default.aspx.

has been shown that the ideological stance of an LLM often reflects the worldview of its creators[18]. Excluding the Global South—which comprises 88% of the world's population—from AI governance frameworks introduces biases that favour the Global North. Such frameworks fail to address the diverse needs and realities of the majority of humanity.

- **Case in Point:** Facial recognition systems have been shown to have inaccuracies in identifying individuals with darker skin tones, resulting in discriminatory outcomes for Global Majority individuals[19].

Global South Inclusion in Future Safety Summits:



Safety Summit Attendees[20],[21]

[18] "[2410.18417] Large Language Models Reflect the Ideology of Their Creators," https://arxiv.org/abs/2410.18417.

[19] "Unmasking the Bias in Facial Recognition Algorithms | MIT Sloan," https://mitsloan.mit.edu/ideas-made-to-matter/unmasking-bias-facial-recognition-algorithms.

[20] "AI Safety Summit: Confirmed Attendees (Governments and Organisations)," GOV.UK, https://www.gov.uk/government/publications/ai-safety-summit-introduction/ai-safety-summit-confirmed-governments-and-organisations.

[21] "AI Seoul Summit: Participants List (Governments and Organisations)," GOV.UK,https://www.gov.uk/government/publications/ai-seoul-summit-programme/ai-seoul-summit-participants-list-governments-and-organisations.

Global South countries might not be participating in forums such as the Safety Summit because of the following reasons:

- The countries in question were not extended invitations to the event.
- These countries lacked the necessary state capacity and resources to contribute meaningfully.
- The proposed agenda did not adequately reflect or prioritise concerns specific to the Global South.

This absence of participation contributes to a siloed approach in global AI governance, creating trust deficits and exacerbating the Digital Divide. A post-event delegatory consultation could be conducted to bridge these gaps, inviting non-attendee governments to provide feedback on the resolutions discussed and suggest agenda items for future Safety Summits. Furthermore, engagement could be strengthened through the AISI network[22], which could organise parallel consultations at forums like the International Telecommunication Union (ITU) to ensure a more inclusive and comprehensive dialogue.

Recommendations:

We identify the following 5 key objectives pertaining to AI, which are of global importance, with the Global South being a particularly relevant stakeholder. The recommendations were designed to create practical steps for the Global North governments to work towards the Global South's integration into global AI safety efforts via the Safety Summit or otherwise. This will require both the Global North and South to work together and is expected to be a gradual process with a requirement for a lot of dialogue and mutual understanding.

Objective 1: Establish AI Safety Institutes to Build State Capacity in the Global South

Problem: *Preventing a Widening AI Divide Before It Takes Root*

The Global South often lacks the foundational infrastructure, regulatory frameworks, and institutional capacity required to manage and mitigate AI risks effectively[23]. Without

[22] Kristina Fort, "The Role of AI Safety Institutes in Contributing to International Standards for Frontier AI Safety" (arXiv, September 17, 2024), https://doi.org/10.48550/arXiv.2409.11314.
[23] "2024 Government AI Readiness Index," 2024.

dedicated AI Safety Institutes[24], these nations remain vulnerable to unregulated AI deployments, data misuse, and biased algorithms. As global AI standards are developed, the Global South risks becoming a passive recipient of frameworks that may not align with or address its socio-economic realities. This lack of proactive governance structures exacerbates inequality, leaving these regions susceptible to exploitation and harmful technological consequences[25].

Call to Action: Initiate a collaborative effort to support the development of AI Safety Institutes in the Global South over the next two years, starting with a feasibility study and resource assessment to ensure these institutes address local needs effectively and sustainably.

Timeline: Launch the feasibility study within six months, followed by a detailed plan and funding proposals within one year, aiming for the initial groundwork to begin by the two-year mark.

<u>Recommendations for the Global North:</u>

- **Coordinate Joint Emergency Response Mechanisms**: AI Safety Institutes in the Global North should set up collaborative emergency response protocols to treat any emergent AI risks with the most vulnerable Global South governments for AI-related crises (e.g., algorithmic failures or AI-driven security threats). These protocols could detail how expertise and resources can be rapidly deployed in a crisis scenario.

- **Develop Localised Adaptation Blueprints**: Require AI Safety Institutes in the Global North to develop “localised adaptation blueprints[26]” of their AI safety protocols, which are designed to the context of and in partnership with Global South institutions. These could include modular safety frameworks that account for differing levels of regulatory and infrastructural maturity, risk mapping for regional vulnerabilities, cultural and linguistic customisation of AI systems, and phased implementation strategies aligned with domestic AI strategies.

[24] “Understanding the First Wave of AI Safety Institutes: Characteristics, Functions, and Challenges,” Institute for AI Policy and Strategy https://www.iaps.ai/research/understanding-aisis.

[25] “AI Safety Institutes: Can Countries Meet the Challenge?,” https://oecd.ai/en/wonk/ai-safety-institutes-challenge.

[26] Localized Adaptation Blueprints are modular frameworks that customize AI safety protocols for the Global South, addressing region-specific risks like data misuse or biased algorithms. These blueprints prioritize local linguistic and cultural integration, regulatory alignment, and scalable infrastructure to make AI systems contextually effective.

- **Form Institutional Partnerships and Fellowships**: Develop long-term partnerships between established AISIs in the Global North and new or potential AISIs in the Global South. This could involve knowledge exchange programs, technology transfer agreements, joint research projects, and hosting fellowships for Global South researchers and policymakers.

Analogy:
The Global Health Security Agenda (GHSA)[27] facilitates knowledge exchange, joint projects, and capacity building to enhance global health preparedness. It could be looked at as a model for partnerships between developed and developing countries.

Objective 2: Coordinate a Global Moratorium on Lethal Autonomous Weapons Systems (LAWS)

Problem: *Preventing Tech-Fueled Conflict Hotspots*

The development and deployment of Lethal Autonomous Weapon Systems (LAWS) represent an urgent, high-stakes risk, disproportionately affecting the Global South[28]. These autonomous systems, which can select and engage targets without human intervention, are poised to reshape warfare in ways that exacerbate existing vulnerabilities. The Global South, often characterised by higher conflict prevalence[29], fragile institutions[30], and the proliferation of non-state actors[31], stands at the frontline of these emerging threats. Without the means to manage or influence the governance of LAWS, these regions could become testing grounds for technologies that are unreliable and unpredictable[32] which destabilise fragile societies and escalate violence.

Call to Action: Facilitate technical workshops and research forums over the next year to generate data and insights that inform global discussions on a LAWS moratorium, with the ultimate goal of building a technical case for future international agreements.

[27] “The Global Health Security Agenda (GHSA): 2020-2024,” n.d.
[28] “The Global South and Autonomous Weapons Controls | Arms Control Association,” https://www.armscontrol.org/act/2024-11/features/global-south-and-autonomous-weapons-controls.
[29] “ACLED Conflict Index - ACLED,” https://acleddata.com/conflict-index/.
[30] John Idriss Lahai and Helen Ware, eds., *Governance and Societal Adaptation in Fragile States* (Cham: Springer International Publishing, 2020), https://doi.org/10.1007/978-3-030-40134-4.
[31] International Law Editorial, “Exploring the Impact of Non-State Actors on Global Governance - World Jurisprudence,” December 1, 2024, https://worldjurisprudence.com/impact-of-non-state-actors/.
[32] Whitepaper Alycia Colijn and Heramb Podar, “Technical Risks of (Lethal) Autonomous Weapons Systems,” n.d.

Timeline: Organize the first technical workshop within four months and establish a research consortium to release a comprehensive technical report within one year.

Recommendations for the Global North:

- **Perform Red Teaming Simulations**: Organize Red Teaming exercises[33], where Global North military and AI safety experts simulate potential scenarios involving LAWS in Global South contexts to better understand the threats and design robust mitigation strategies. The findings should be shared with policymakers to inform regulations.

- **Delineate AI Safety Institutes' Role in Risk Mapping**: Global North AI Safety Institutes could perform "regional risk assessments" focused on how the spread of autonomous weapon systems could destabilise Global South regions. These risk assessments[34] could inform policy recommendations and build geopolitical cases against the spread of LAWS.

- **Provide Technical Assistance and Information for Weapon Verification Systems**: While this still involves technical support, AI Safety Institutes can lead the development of AI-powered verification tools that are designed to detect and monitor LAWS activities in Global South regions. These tools should be deployed with assistance from Global North defence ministries to enhance global monitoring and cover all global conflict hotspots.

Note:

Collaborating with defence ministries, these institutes can deploy such tools in regions susceptible to LAWS proliferation, enhancing global monitoring and mitigating potential conflicts. This approach mirrors existing verification mechanisms in arms control, such as the use of remote sensing and open-source data for monitoring compliance with Weapons of Mass Destruction (WMD) treaties[35]. For example, these tools could analyse data from various sources to detect unauthorised deployment of Lethal Autonomous Weapon Systems (LAWS), ensuring adherence to established norms.

[33] Tessa Baker, "What Does AI Red-Teaming Actually Mean?," *Center for Security and Emerging Technology* (blog), October 24, 2023, https://cset.georgetown.edu/article/what-does-ai-red-teaming-actually-mean/.

[34] "ISO - ISO 31000 — Risk Management," ISO, December 10, 2021, 900, https://www.iso.org/iso-31000-risk-management.html.

[35] Veronica Borrett et al., "Science and Technology for WMD Compliance Monitoring and Investigations," November 12, 2020, https://unidir.org/publication/science-and-technology-for-wmd-compliance-monitoring-and-investigations/.

Objective 3: Leverage AI Responsibly for Achieving the Sustainable Development Goals (SDGs)

Problem: *Unlocking AI's Potential for Global Good*

Artificial Intelligence holds immense potential to accelerate progress toward the Sustainable Development Goals (SDGs) in the Global South[36]. However, realising this potential requires responsible, well-governed AI deployments that address regional challenges rather than exacerbate them and not merely focus on including underrepresented groups but also make it work for all[37]. The Global South faces deeply embedded structural limitations like economic inequality, limited social inclusion, and lack of technical infrastructure[38].

Call to Action: Form a working group to draft practical, contextual guidelines for AI projects to achieve SDGs in the Global South, with a timeline for publishing these guidelines and testing them in pilot projects within one year.

Timeline: Establish the working group within three months, with draft guidelines ready for feedback within six months, and begin pilot testing in one year.

<u>**Recommendations for the Global North:**</u>

- **Establish Public-Private AI Deployment Guidelines**: Draft practical guidelines[39] for AI deployments, focusing on minimising negative impacts and maximising SDG benefits. These guidelines should emphasise the use of transparent, explainable AI systems with clear red lines[40], especially in critical sectors like healthcare and agriculture.

- **Carry Out AI Safety Auditing for Development Projects**: Establish auditing mandates for any AI project by development agencies (like USAID[41] or the UK's

[36] Brigitte Hoyer Gosselink et al., "AI in Action: Accelerating Progress Towards the Sustainable Development Goals," n.d.

[37] Alan Chan et al., "The Limits of Global Inclusion in AI Development" (arXiv, February 2, 2021), https://doi.org/10.48550/arXiv.2102.01265.

[38] "The 'AI Divide' between the Global North and Global South," World Economic Forum, January 16, 2023, https://www.weforum.org/stories/2023/01/davos23-ai-divide-global-north-global-south/.

[39] Such guidelines must prioritize data sovereignty, ensuring local control and ethical data use while fostering regional capacity-building to reduce reliance on external actors. They should integrate context-specific safeguards to address systemic inequities and mandate transparent accountability mechanisms to prevent safety-washing

[40] "The AI Red Line Challenge | TechPolicy.Press," https://www.techpolicy.press/the-ai-red-line-challenge/.

[41] For their part, USAID acknowledges the potential risks associated with deploying AI in developing-country contexts in its "Artificial Intelligence in Global Development" report and noting

FCDO) to undergo a pre-deployment safety and ethics audit. AI Safety Institutes should develop these auditing guidelines to minimise the risk of harm in Global South deployments.

Analogy:

The Millennium Villages Project, which aimed to alleviate poverty in Sub-Saharan Africa through targeted interventions, encouraged farmers to plant maize (corn) to boost food security. However, this approach faced contextual challenges: farmers encountered difficulties selling surplus maize due to distant markets, leading to post-harvest losses and limited income generation; additionally, maize cultivation required substantial water, which was often inaccessible, especially in regions lacking adequate irrigation infrastructure[42].

- **Develop Ethical AI Assessment Frameworks**: Work with Global South governments to develop an "AI for SDGs Ethics Checklist" that ensures AI projects are designed with safety, fairness, and protection of human rights in mind. Provide training for the concepts and tools needed to implement this framework[43]. Best practices and case studies could be shared between actors in a multi-stakeholder format.

Objective 4: Safeguard Human Rights, Democracy, and the Rule of Law in AI Governance

Problem: *Closing the Regulatory Lag*

AI technology is advancing at an unprecedented rate, outpacing the development of regulatory frameworks in the Global South[44]. This regulatory lag leaves countries ill-equipped to confront the ethical, social, and economic challenges AI introduces. Populations are increasingly at risk of digital rights violations, unchecked surveillance, and economic exploitation. Principles like openness and explainability are often designed with the Global North in mind, assuming access and agency that may not exist in other contexts. When applied in the Global South, these principles can be impractical or

that as of May 2023, only two AI systems were in use within its foreign assistance programs according to an AI inventory submitted to the Office of Management and Budget(OMB).

[42] "The Idealist by Nina Munk: 9780767929424 | PenguinRandomHouse.Com: Books," https://www.penguinrandomhouse.com/books/118598/the-idealist-by-nina-munk/.

[43] D. Leslie, C. Rincón, M. Briggs, A. Perini, S. Jayadeva, A. Borda, S. J. Bennett, C. Burr, M. Aitken, M. Katell, J. Fischer Wong, and I. Kherroubi Garcia. *AI Sustainability in Practice. Part One: Foundations for Sustainable AI Projects.* London: The Alan Turing Institute, 2023. https://www.turing.ac.uk/news/publications/ai-ethics-and-governance-practice-ai-sustainability-practice-part-one-foundations.

[44] Collingridge, David. *The Social Control of Technology.* New York: St. Martin's Press; London: Pinter, 1980. ISBN 0-312-73168-X.

counterproductive, highlighting the urgent need for tailored regulatory approaches that fit local realities.

Call to Action: Set up an exploratory task force to evaluate the implementation of ex-ante **Human Rights Impact Assessments** for AI technologies, focusing on aligning these mechanisms with local governance readiness levels in the Global South, with findings to be presented within 12 months. Inspiration could be taken from the learnings shared by the team implementing the Chilean Readiness Assessment Methodology[45].

Timeline: Form the task force within three months, conduct evaluations over nine months, and publish a report with recommendations by the end of the 12-month period.

<u>**Recommendations for the Global North:**</u>

- **Mandate Interoperable Standards for Global Tech Firms**: Require tech companies operating in the Global North to publicly disclose how their algorithms are trained, audited, and evaluated, especially when deployed in the Global South. Work towards incorporating Global South concerns and establishing international transparency benchmarks to hold firms accountable. Global North and Global South Actors must work together to combine soft and hard-law approaches in a way that establishes meaningful international guidelines while also being cognizant of regional or local regulations[46].

- **Facilitate the Implementation of ex-ante Human Rights Impact Assessments(HRIAs)**: Encourage Global North countries to adopt ex-ante HRIAs[47] for AI systems, similar to impact assessments used for environmental projects, to evaluate the human rights implications of AI deployments. Share best practices with the Global South and promote the integration of these assessments into local governance[48].

- **Work on Pre-Deployment Safety Reports**: Mandate that Global North AI Safety Institutes prepare and share comprehensive safety evaluation reports with Global South governments before deploying AI technologies. This ensures that potential

[45] “7 Lessons from Implementing the RAM in Chile | UNESCO,” https://www.unesco.org/en/articles/7-lessons-implementing-ram-chile.
[46] “PNAI Report | Internet Governance Forum,” , https://intgovforum.org/en/filedepot_download/282/28491.
[47] Hickok, Merve, Marc Rotenberg, Christabel Randolph, and Sneha Revanur. "Artificial Intelligence and Human Rights." Statement to the U.S. Senate Judiciary Committee, Subcommittee on Human Rights and the Law, June 13, 2023. Center for AI and Digital Policy. https://www.judiciary.senate.gov/committee-activity/hearings/artificial-intelligence-and-human-rights.
[48] Yoshua Bengio, “International Scientific Report on the Safety of Advanced AI - Interim Report,” n.d.

risks are flagged and mitigated collaboratively[49]. Facilitate early access for Global South policymakers to AI safety evaluation reports by Global North institutes[50]. This allows for timely feedback and adaptation, ensuring AI models are better suited to diverse linguistic and cultural contexts.

Objective 5: Mitigate Language and Cultural Bias in AI Systems

Problem: *Ensuring AI Speaks Every Language*

Current AI models are heavily biased toward the cultural and linguistic norms of the Global North[51]. This creates a significant risk and disadvantage for the Global South, where diverse languages and cultural practices are often misrepresented or wholly excluded. Such biases perpetuate digital inequality, systemic racism, and a lack of access to fair AI-driven services. Additionally, the opacity of these AI algorithms makes it difficult to hold developers accountable for the harm caused by biased systems[52]. The Global South needs AI that recognises and respects its cultural diversity, yet it currently faces systemic barriers that deepen existing inequalities and marginalise already vulnerable populations.

Call to Action: Commission a series of regulatory stress tests over the next year to identify and address language and cultural biases in AI models, ensuring these findings are incorporated into global AI safety standards.

Timeline: Begin designing stress tests within three months, conduct simulations over the following six months, and present results and recommendations within one year.

Recommendations for the Global North:

- **Deploy Monitoring and Evaluation Toolkits**: Provide open-source toolkits for monitoring AI deployments, including metrics for safety, bias, and performance[53]. Offer training programs to empower local regulators and institutions to use these tools effectively.

[49] Markus Anderljung et al., “Frontier AI Regulation: Managing Emerging Risks to Public Safety” (arXiv, November 7, 2023), https://doi.org/10.48550/arXiv.2307.03718.
[50] “AI Safety Institute Releases New AI Safety Evaluations Platform - GOV.UK,” https://www.gov.uk/government/news/ai-safety-institute-releases-new-ai-safety-evaluations-platform.
[51] “[2410.18417] Large Language Models Reflect the Ideology of Their Creators,” https://arxiv.org/abs/2410.18417.
[52] “The ‘Missed Opportunity’ with AI’s Linguistic Diversity Gap | World Economic Forum,” https://www.weforum.org/stories/2024/09/ai-linguistic-diversity-gap-missed-opportunity/
[53] Rachel K. E. Bellamy et al., “AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias” (arXiv, October 3, 2018), https://doi.org/10.48550/arXiv.1810.01943.

- **Perform Regulatory Stress Testing**: Encourage Global North regulatory bodies to conduct “stress tests”[54] of their AI governance models using simulated Global South scenarios. This would expose potential blind spots and allow for the creation of more resilient frameworks that account for varied socioeconomic realities.

- **Conduct Cross-Continental Joint Safety Evaluations**: Global North AI Safety Institutes should carry out safety evaluations of AI systems designed for projects in the Global South. This collaboration helps both sides: it highlights vulnerabilities and biases that may not be obvious in a Western context (benefiting the Global South) while also refining the safety mechanisms of these AI systems to make them globally robust (benefiting the Global North). For example, a healthcare AI model might perform well in Europe but fail to consider unique disease prevalence or healthcare delivery methods in Africa. Addressing these disparities strengthens the system overall.

Conclusion:

The path forward for AI governance must embrace the spirit of working together, for we have far to go- whether it be in building bridges or understanding bottlenecks. The efforts today will define the landscape of tomorrow, making it imperative that AI governance is shaped by comprehensive, culturally sensitive collaborations that reflect our global diversity.

The Global South brings critical insights and unique experiences that can strengthen international AI standards, making them more robust and globally applicable. For the Global North, engaging meaningfully with the Global South is not just an act of fairness but a strategic investment in a safer, more balanced, and resilient future. We must build governance frameworks that are not only ethical and effective but also representative of our diverse world. This requires rigorous risk assessments, culturally attuned rights protections, equitable capacity building, inclusive standards, and preparedness for emergent crises—all woven into collaborations that reflect the unique strengths and challenges of both the Global North and South.

Acknowledgements:
We acknowledge the use of the Large Language Model GPT-4 in copyediting this paper.

[54] These could be sandboxes with different environments, scarce resources or simulations of how SDG metrics perform.