

AI Safety and Security Enable Innovation in Global Majority Countries

Joanna Wiaterek¹, Jared Perlo²

¹Centre for AI Security and Access

²Centre pour la Sécurité de l'IA (Center for AI Safety)

joanna@casa-ai.org, jared@securite-ia.fr

Extended Abstract

The rapid advancement of Artificial Intelligence (AI) capabilities and accelerating AI adoption has thrust the technology to the forefront of policy agendas worldwide. However, AI is a dual-use technology; it simultaneously holds promise of significant benefits and the potential for widespread disruptions and harms. In this context, governments face an apparent zero-sum choice between two high-stakes priorities: limit regulation and encourage AI progress and innovation or regulate AI risks and prioritize safety. While Global Majority governments do not uniquely face this apparent dilemma, Global Majority governments do face particular and pressing challenges in navigating this purported antagonism due to the scarcity of resources and competing socioeconomic issues.

This paper argues that depicting AI safety and security (S&S) and innovation as trade-offs presents a false dichotomy and ultimately undermines emerging economies' development prospects. Instead, this paper argues that investments in AI S&S can render long-term development benefits and enable beneficial AI ecosystems, particularly for Global Majority countries.

We begin by offering context on AI S&S efforts and their current global governance. We then argue that AI S&S enable and complement development goals through a case study analysis. Section II offers specific recommendations for the Global Majority actors on building AI S&S capacity and engaging in global AI S&S governance, especially relevant ahead of the upcoming Global AI Summit in India. Lastly, Section III addresses alternative views to the argument that AI S&S are prerequisites for development and should be an increased focus for Global Majority countries.

There is still little cross-disciplinary conversation tying AI S&S to existing global development practices. AI may act as a polycrisis intensifier, exacerbating rather than solving the wide array of existing global challenges. For example, systemic AI risks may compromise or even

undo global progress on the Sustainable Development Goals (SDGs). Focusing on either AI safety or AI security in isolation is insufficient. As a result, for the purposes of this paper, we refer to AI S&S efforts together.

Both AI S&S are intrinsically linked to key factors affecting a country's socioeconomic growth. For example, AI S&S could dramatically affect institutional stability, financial independence, or technologies that could significantly boost productivity and real income. Societal safeguards from AI risks emerge as critical, rather than competing, policy considerations for Global Majority decision-makers. To name a few, non-exhaustive examples: AI S&S capacity (e.g. laws, safety culture and agencies, trained personnel) helps foster preparedness to develop an AI industry effectively and sustainably; AI's adoption and utility depend on users' trust, which is contingent on contextualized safety guardrails and value alignment; and defined, local AI S&S priorities can enable robust digital sovereignty and minimize digital colonialism.

Given the rapid pace of AI progress, countries must urgently embrace AI S&S. Investing in AI S&S now could unlock significant (and in extreme scenarios, perhaps exponential) economic growth. We propose several concrete actions at national and international levels. For example: national AI strategies in Global Majority countries should incorporate a much stronger focus on S&S with an added focus on informational harms. Regional bodies (like the African AI Council) can pool scientific talent and provide representative voices on higher-level, potentially-catastrophic risks. AI S&S research in Global Majority countries should not strive to replicate existing S&S efforts but instead focus on tailored, locally-relevant issues. This could include establishing better metrics to measure AI's impact on criteria more pressing for Global Majority countries (e.g. AI's impact on and uptake within the informal economy), building culturally responsive and appropriate tools to measure AI capabilities, and measuring the safety of models in local languages, amongst other key areas. AI S&S are not trade-offs for development. Rather, AI S&S can help secure long-term benefits in the age of AI.

References

- Adan, S. N.; Trager, R.; Blomquist, K.; Dennis, C.; Edom, G.; Velasco, L.; Abungu, C.; Garfinkel, B.; Jacobs, J.; Okolo, C. T.; Wu, B.; and Vipra, J. 2024. Voice and Access in AI: Global AI Majority Participation in Artificial Intelligence Development and Governance. Technical report, Oxford Martin School.
- Araujo, R.; Fort, K.; and Guest, O. 2024. Understanding the First Wave of AI Safety Institutes Characteristics, Functions, and Challenge. Technical report, Institute for AI and Policy Strategy. <https://www.iaps.ai/research/understanding-aisis>, Accessed 2025-05-02.
- Bengio, Y.; Mindermann, S.; Privitera, D., et al. 2025. International AI Safety Report. ArXiv:2501.17805.
- Boakye, B.; Singh, S.; Drake, G.; Harriet, C.; Britton, L.; Adams, R.; Medon, F.; Eitel-Porter, R.; and Mo'kander, J. 2025. How Leaders in the Global South Can Devise AI Regulation That Enables Innovation. Technical report, Tony Blair Institute for Global Change. www.institute.global/insights/tech-and-digitalisation/how-leaders-in-the-global-south-can-devise-ai-regulation-that-enables-innovation. Accessed: 2025-05-20.
- Brundage, M.; Avin, S.; Clark, J.; Toner, H.; Eckersley, P.; Garfinkel, B.; Dafoe, A.; Scharre, P.; Zeitoff, T.; Filar, B.; Anderson, H.; Roff, H.; Allen, G. C.; Steinhardt, J.; Flynn, C.; hE' igeartaigh, S.; Beard, S. J.; Belfield, H.; Farquhar, S.; Lyle, C.; Crootof, R.; Evans, O.; Page, M.; Bryson, J.; Yam- polskiy, R.; and Amodei, D. 2024. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. ArXiv:1802.07228.
- Caroli, L. 2025. France's AI Action Summit. Technical report, Center for Strategic and International Studies. <https://www.csis.org/analysis/frances-ai-action-summit>. Accessed 2025-05-18
- Chaudhuri, R. 2023. Decoding the G20 Consensus on Digital Public Infrastructure: A Key Outcome of India's Presidency. Technical report, Carnegie Endowment for International Peace. www.carnegieendowment.org/research/2023/09/decoding-the-g20-consensus-on-digital-public-infrastructure-a-key-outcome-of-indias-presidency?lang=en. Accessed 2025-05-14.
- Chawla, U.; Mohnot, R.; Singh, H.V.; Banerjee, A. 2023. "The Mediating Effect of Perceived Trust in the Adoption of Cutting-Edge Financial Technology among Digital Natives in the Post-COVID-19 Era." *Economies* 11 (12): 286. <https://doi.org/10.3390/economies11120286>.
- Ferrara, E. 2023. "Fairness And Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, And Mitigation Strategies." *Sci* 6 (1): 3. <https://doi.org/10.3390/sci6010003>.
- Guembe, B.; Azeta, A.; Misra, S.; Osamor, V.C.; Fernandez-Sanz, L.; and Pospelova, V. 2022. "The Emerging Threat of Ai-Driven Cyber Attacks: A Review." *Applied Artificial Intelligence* 36 (1): 2037254. <https://doi.org/10.1080/08839514.2022.2037254>.
- Korinek, A. 2024. "Economic Policy Challenges for the Age of AI." Version 1. Preprint, arXiv. <https://doi.org/10.48550/ARXIV.2409.13168>.
- Kuewor, R. 2025. *How AI Is Reshaping the Future of Informal Work in the Global South*. World Economic Forum. <https://www.weforum.org/stories/2025/05/ai-reshaping-informal-work-global-south/>
- Laue, H.J. forthcoming. "Nuclear Power and Technology Transfer." *IAEA Bulletin* 24 (4). <https://www.iaea.org/sites/default/files/24404781923.pdf>.
- Naughton, J. 2016. "The Evolution of the Internet: From Military Experiment to General Purpose Technology." *Journal of Cyber Policy* 1 (1): 5–28. <https://doi.org/10.1080/23738871.2016.1157619>.
- Powell, R.; Stockwell, S.; Sharadjaya, N.; Boyes, H. 2024. *Towards Secure AI: How Far Can International Standards Take Us*. Alan Turing Institute. https://cetas.turing.ac.uk/sites/default/files/2024-03/cetas_research_report_-_towards_secure_ai_0.pdf.
- Reuters. 2025. "Emerging Economies Lead the Way in AI Trust, Survey Shows." April 29. <https://www.reuters.com/business/emerging-economies-lead-way-ai-trust-survey-shows-2025-04-28/>. Accessed 2025-05-10.
- Schwachula, A.; Seoane, M.V.; Hornidge, A-K. 2014. *Science, Technology and Innovation in the Context of Development: An Overview of Concepts and Corresponding Policies Recommended by International Organisations*. ZEF Working Paper Series No. 132. <https://www.econstor.eu/handle/10419/99990>.
- Sinha, A. 2025. "India Will Host the Next Global AI Summit. What Mark Will Modi Make?" *Tech Policy Press*, March 7. <https://www.techpolicy.press/india-will-host-the-next-global-ai-summit-what-mark-will-modi-make/>. Accessed 2025-05-10.
- Stadler, C. 2024. *How M-PESA Cornered the Market in Kenya*. Warwick Business School. <https://www.wbs.ac.uk/news/how-mpesa-cornered-the-market/>. Accessed 2025-05-10.
- Suleyman, M. 2023. *The Coming Wave: Technology, Power, and the Twenty-First Century's Greatest Dilemma*. First edition. Crown.
- The Singapore Consensus on Global AI Safety Research Priorities*. 2025. Singapore Conference on AI. <https://www.scai.gov.sg/2025/scai2025-report>.
- The Universal Digital Public Infrastructure Safeguards Framework: A Guide to Building Safe and Inclusive DPI for Societies*. 2024. United Nations Office of the Secretary-General's Envoy on technology. <https://dpi-safeguards-framework.org/frameworkpdf>.
- Trammell, P.; Korinek, A. 2023. *Economic Growth under Transformative AI*. National Bureau of Economic Research. <https://doi.org/10.3386/w31815>.
- Vergara Cobos, E.; Cakir, S. 2024. *A Review of the Economic Costs of Cyber Incidents*. World Bank. <https://documents1.worldbank.org/curated/en/099092324164536687/pdf/P17876919ffee4079180e81701969ad0a18.pdf>. Accessed 2025-05-11.
- Vinuesa, R.; Azizpour, H.; Leite, I.; Balaam, M.; Dignum, V.; Domisch, S.; Felländer, A.; Langhans, S.D.; Tegmark, M.; Nerini, F.F. 2020. "The Role of Artificial Intelligence in Achieving the Sustainable Development Goals." *Nature Communications* 11 (1): 233. <https://doi.org/10.1038/s41467-019-14108-y>.
- Wiaterek, J.; Abungu, C.; Okolo, C.T.. 2025. *Building Regional Capacity for AI Safety and Security in Africa*. Brookings Institute. <https://www.brookings.edu/articles/building-regional-capacity-for-ai-safety-and-security-in-africa/>.
- Zhou, T. 2012. "Examining Mobile Banking User Adoption from the Perspectives of Trust and Flow Experience." *Information Technology and Management* 13 (1): 27–37. <https://doi.org/10.1007/s10799-011-0111-8>.