POLICY BRIEF

T20 SOUTH AFRICA 2025

G20 SOUTH AFRICA 2025

Beyond Transparency Theatre: Open-Source as a Path to Responsible AI Development

2025

Vineet John Samuel, Doctoral Candidate in Governance, Hertie School (Berlin)
Chhavi Banswal, Independent Researcher (New Delhi)

02

Digital Transformation

Abstract

The release of DeepSeek's V3 and R1 models with open weights could be a turning point in artificial intelligence (AI) development, representing meaningful progress in making it more accessible and transparent. However, the Open Source Initiative's comprehensive definition of Open Source AI (OSAI) offers a compelling vision of what even greater transparency could achieve. Drawing from DeepSeek's recent innovations in model optimisation and efficiency, we present three interconnected arguments supporting the evolution from current open-weight practices toward full OSAI development.

First, we examine how DeepSeek's success with publicly available weights demonstrates the viability and value of increased transparency. We highlight the additional benefits of embracing the entire OSAI framework, which includes training data provenance and characteristics, complete source code for training and inference, and full model parameters. Second, we argue that pivoting toward genuine open-source standards can potentially de-escalate the emerging AI "arms race" by transforming a seemingly zero-sum game into a collective endeavour for advancement. Third, we contend that open-weight initiatives could naturally evolve toward the Open Source Initiative's standards of transparency and accountability in AI systems – a critical consideration as we approach potentially superintelligent systems where value alignment becomes paramount. The point of convergence for the recommendations will be global cooperation to establish a framework for responsible AI governance – one that effectively balances innovation with accountability.

As AI systems increasingly train successive generations of models, this policy brief argues that building upon current open-weight practices toward comprehensive open-source standards is necessary for the responsible advancement of the technology – aligning with the G20's AI governance and global safety standards. By advocating for stronger open-source AI practices, the G20 can play a crucial role in tech diplomacy, de-escalating the emerging AI arms race. With AI becoming increasingly significant in shaping economies and societies, fostering collaboration over proprietary competition also resonates with the G20's broader vision of inclusive and sustainable technological development.

Diagnosis

AI stands at a critical inflection point. The race to develop increasingly powerful AI systems has intensified, yet the predominant closed-source approach threatens both innovation and safety. Recent open-weight models offer a compelling alternative path that aligns with responsible governance and sustainable progress.

AI has transitioned from research laboratories to becoming a cornerstone of economic and geopolitical strategy. The OECD defines an AI system as "a machine-based system that infers from inputs how to generate outputs that can influence physical or virtual environments".[1] This transformative technology promises to revolutionise productivity across all economic sectors – from healthcare and manufacturing to education and governance.

This potential has ignited intense global competition backed by unprecedented resources. Private investment in AI exceeded $91.9 billion in 2023 alone, with the US and China emerging as dominant players.[2] Former US president Joe Biden called AI "humanity's most powerful tool",[3] while Chinese President Xi Jinping designated it "the strategic technology that has profoundly changed the way humans produce and live".[4] Indian Prime Minister Narendra Modi similarly

[1] OECD. (2019). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 ↩

[2] Stanford Institute for Human-Centered Artificial Intelligence (Stanford HAI). (2024). Artificial Intelligence Index Report 2024. Stanford University. https://aiindex.stanford.edu/report/ ↩

[3] The White House. (2023, October 30). Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence ↩

[4] Chinese Peoples Political Consultative Conference. (2025, April 27). Xi Jinping's statement at a study session of the Political Bureau of the CPC Central Committee. http://en.cppcc.gov.cn/2025-04/27/c_1088876.htm#. ↩

positioned AI as "a catalyst for India's development journey".[5] This rhetoric has cascaded through policy circles, with security bodies warning about competitive readiness for the "AI era".[6]

The resulting "AI arms race" has led to a concentration of computational resources. Nations compete for limited semiconductor resources, with advanced graphics processing units (GPUs) becoming strategic assets. This marks a shift from previous technological revolutions, when public research played a more substantial role.

This zero-sum framing has hindered our ability to examine what we are racing toward. Critics argue that large language models are merely "stochastic parrots" – systems that mimic language without understanding.[7] Artists and publishers allege intellectual property (IP) theft,[8] researchers document algorithmic bias, and ethicists question pursuing increasingly powerful AI without adequate governance.[9] In this competitive context, critical ethical questions receive insufficient attention, exacerbated by the dominance of closed-source development. Despite OpenAI's initial commitment to openness, it pivoted toward closed models in 2019.[10] Similarly, Anthropic, Google, and xAI maintain proprietary systems with limited scrutiny.

---

[5] Ministry of External Affairs, India (MEA). (2023, September 10). PM's address at the Global AI Summit 2023. https://www.pib.gov.in/newsite/PrintRelease.aspx?relid=238867/ ↩

[6] National Security Commission on Artificial Intelligence (NSCAI). (2021). Final report. https://reports.nscai.gov/final-report// ↩

[7] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922 ↩

[8] Samuelson, P. (2023). Generative AI meets copyright. Communications of the ACM, 66(10), 32–34. https://cacm.acm.org/opinion/u-s-copyright-offices-questions-about-generative-ai/ ↩

[9] Gabriel, I. (2022). Artificial intelligence, values, and alignment. Minds and Machines, 32(3), 457–489. https://link.springer.com/article/10.1007/s11023-020-09539-2 ↩

[10] Altman, S. (2019, March 11). OpenAI LP. OpenAI. https://openai.com/blog/openai-lp/ ↩

A notable exception is Meta's LLaMA family, which embraced open weights in 2023. These models have become a foundation for thousands of derivative works, with over 100 000 developers building upon them.[11] Yet the closed paradigm remains the dominant norm among major AI labs, restricting our collective ability to address fundamental questions surrounding AI development.

In this high-stakes environment, open-source AI offers not just technical advantages but also regulatory footholds – transparency, auditability, and oversight necessary for responsible governance. While LLaMA demonstrated the viability of open weights, the January 2024 release of DeepSeek's models reinforced this approach.

DeepSeek's open-weight models rival top commercial systems at a fraction of the cost. While GPT-4 reportedly cost between $41 million and $78 million to train,[12] DeepSeek achieved this for allegedly under $10 million (though unverified). The model represents a pivotal development – competitive AI capabilities can emerge outside the dominant closed-source paradigm, even under hardware constraints imposed by export restrictions.

This illuminates an alternative path forward. By opening weights and architecture details, such models transform AI systems from opaque black boxes to potentially more transparent ones where critical aspects of their functioning can be examined. This offers a pragmatic middle ground that balances competitive innovation with necessary oversight.

[11] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., & others. (2023). LLaMA: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971. ↩

[12] Buchholz, K. (2024, August 23). The Extreme Cost Of Training AI Models. Forbes. https://www.forbes.com/sites/katharinabuchholz/2024/08/23/the-extreme-cost-of-training-ai-models// ↩

The implications are profound and two-fold. First, it challenges the assumed necessity of secrecy in creating competitive AI capabilities. Open models facilitate knowledge sharing that makes development more efficient and less resource-intensive – seen in technological advancement from software to biotechnology.[13] Second, they enable transparency, allowing researchers and regulators to examine how systems reason and make decisions. This visibility is essential for addressing questions of liability, copyright infringement, and other ethical concerns.[14]

While open weights represent significant progress, the Open Source Initiative's comprehensive definition of OSAI offers a more complete vision for AI transparency. The OSAI framework expands beyond mere model weights to include (1) training data provenance and characteristics; (2) complete source code for training and inference; (3) full model parameters and hyperparameters; and (4) transparent documentation of evaluation methods. This comprehensive approach addresses critical gaps in current open-weight models and provides the foundation necessary for truly auditable, ethical AI development.

As AI systems grow increasingly sophisticated and integrated into critical infrastructure, closed development becomes problematic. Open-weight models offer a middle path enabling accelerated yet responsible AI development. The question is not whether we can afford to develop AI openly but whether, in a world increasingly shaped by these systems, we can afford not to.

[13] Chesbrough, H. (2020). Open innovation results: Going beyond the hype and getting down to business. Oxford University Press. ↩

[14] Buterin, V., & Weyl, E. G. (2022). Decentralized society: Finding web3's soul. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4105763 ↩

Recommendations

The geopolitical dynamics around AI are more nuanced than simply who builds the most powerful model. From access to data, investment in R&D and an enabling regulatory landscape, several factors can tip the scales. Differing national strategies, data formats, and regulatory frameworks drive proprietary competition over collaboration. But how the world views this oscillating powerplay can be redefined by the G20. What is touted as the global AI arms race is perhaps an unprecedented opportunity for collective effort.

Despite concerns around data sovereignty and IP risks, comprehensive open-source standards have the potential to solve their own issues through global alignment and are thus essential for the responsible advancement of the technology. Therefore, the following recommendations are made.

1. Aligning national policies

A shared approach to open-weight practices is a necessity for equitable AI adoption. How AI is perceived varies globally – emerging economies hold a more positive outlook while people in developed countries show low confidence in government and commercial entities in terms of fair development and regulation of AI.[15] Open-source AI systems by their nature earn higher levels of public trust, leading to increased adoption,[16] thus meriting prioritisation.

The G20 AI principles of human-centric values, transparency, and robustness[17] may have been founded in shared understanding, but a common approach to

[15] KPMG. (2023). Trust in artificial intelligence. https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/09/trust-in-ai-country-insight.pdf.coredownload.inline.pdf

[16] The Alan Turing Institute, Ada Lovelace Institute. (2025). How do people feel about AI? https://attitudestoai.uk/

[17] G20 (2019). G20 Ministerial Statement on Trade & Digital Economy. https://www.g20.utoronto.ca/2019/2019-g20-trade.html

these principles remains absent, leading to fragmented adoption and limited inputs from emerging economies. A working group for framing open-source standards under the G20 can nudge nations into aligning their national policies and catalyse global efforts to address fundamental questions surrounding AI development.

2. Interoperability for inclusive, equitable, and sustainable AI

The greatest potential and limitation of global AI is rooted in data. Data sharing between most countries today is primarily at G2G level (government to government) and the G20's focus on interoperability can create more robust and inclusive AI systems. As already demonstrated by India through its digital public infrastructure, interoperability can make technology inclusive and scalable, with equitable access.[18] Moreover, the joint communiqué by the G20 Troika (India, Brazil, South Africa) calls for fair and equitable principles for data governance and acknowledges the need for a technologically neutral approach to create a level-playing field for a seamless transition.[19]

In the current scenario of differing national standards, data formats, and legal frameworks, interoperability may seem incongruous. Even as there are some challenges posed by interoperability itself, such as ethical inconsistencies across borders, regulatory divergence, and varying data protection standards, the benefits are far more. Interoperability in AI systems leads to efficient processing of big data, precision of outcomes, and greater timeliness.[20] It can further increase

[18] Kant, A., Mishra, S. (2023, August 23). The international significance of India's Digital Public Infrastructure. World Economic Forum. https://www.weforum.org/stories/2023/08/the-international-significance-of-indias-digital-public-infrastructure/

[19] Ministry of External Affairs, Government of India (2024). Declaration on Digital Public Infrastructure, AI and Data for Governance - Joint Communique by the G20 Troika (India, Brazil, South Africa). https://www.mea.gov.in/bilateral-documents.htm?dtl/38551/

[20] India AI, Government of India (2025). Interoperability. https://indiaai.gov.in/ai-standards/interoperability

transparency and promote ethical use of AI,[21] which aligns with the G20 AI principles. Moreover, open-source AI can facilitate interoperability in training processes. For instance, Flower enables different devices or organisations to collaboratively train models using open-source frameworks, while preserving privacy through federated learning.[22] Interestingly, public databases can also reduce the amount of web crawling and scraping, making open models more environmentally sustainable.[23]

The G20's "data free flow with trust" principle[24] marks a significant advancement but needs a more directional implementation. A dedicated working group on Interoperability can initiate harmonisation of technical and legal standards to deliver more tangible results.

3. Facilitating global innovation

Open-source practices are not unique to AI development. Historically, they have laid the foundation for breakthroughs in software innovation, like Linux.[25] As AI systems increasingly train successive generations of models, open-source practice can spur innovation and efficiency. Each country's approach to AI remains highly contextualised. For instance, India's AI innovation focuses on productivity gains, while Japan is exploring AI to address its ageing demography and dwindling workforce. The US, on the other hand, has been the front runner in

[21] OECD (2024). AI Principles. https://www.oecd.org/en/topics/sub-issues/ai-principles.html#:~:text=Fostering%20an%20inclusive%20AI%2Denabling,digital%20ecosystem%20for%20trustworthy%20AI.

[22] Flower Intelligence. (2025, March 12). https://flower.ai/blog/2025-03-12-flower-intelligence/

[23] Tarkowski, A., Open Source Initiative. (2025). https://opensource.org/wp-content/uploads/2025/02/2025-OSI-DataGovernanceOSAI-final-v5.pdf

[24] G20 (2019). G20 Ministerial Statement on Trade & Digital Economy. https://www.g20.utoronto.ca/2019/2019-g20-trade.html

[25] Cogent University. (2024, December 3). How open source changed software development! https://www.cogentuniversity.com/post/how-open-source-changed-software-development

market-led innovation[26] ethics, and privacy and rights have been central to the EU's AI regulatory landscape.[27]

A centralised innovation lab under the G20, where multi-stakeholders can converge, would create optimum resources for regulatory sandboxing to promote open-source practices. Such an institution would not only grant the Global South equitable access to cutting-edge technologies but also concretise the G20's commitment to pro-innovation regulations and governance approaches that promote responsible AI, as in the Rio de Janeiro Leaders' Declaration.[28]

Conclusion

The world has already witnessed the power of collective action in the Montreal Protocol (1987), which healed the ozone layer. Instead of a reactionary stand, here is an opportunity for the G20 to be proactive. Even though open weights alone may not equate to full transparency, they can potentially transform a seemingly zero-sum game into a collective endeavour for advancement. The G20 nations must decide whether AI will fuel an arms race or a rising tide that lifts all.

[26] Chopra, C., Kasare, A., Gupta, P. (2024, May 24). How venture capital is investing in AI in the top five global economies — and shaping the AI ecosystem. World Economic Forum. https://www.weforum.org/stories/2024/05/these-5-countries-are-leading-the-global-ai-race-heres-how-theyre-doing-it/

[27] European Commission. European approach to artificial intelligence. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

[28] G20 (2024). G20 Rio de Janeiro Leaders' Declaration https://g20.org/wp-content/uploads/2024/11/G20-Rio-de-Janeiro-Leaders-Declaration-EN.pdf

T20 South Africa Convenors

The Institute for Global Dialogue (IGD)

The South African Institute of International Affairs (SAIIA)

The Institute for Pan-African Thought and Conversation (IPATC)

For publication enquiries, please contact t20@t20southafrica.org

Website: www.t20southafrica.org