COMPUTE CONCENTRATION AS POST-AGI GOVERNANCE INFRASTRUCTURE: COMPUTE PROVENANCE REPORTING AND CONCENTRATION INDICES

Anonymous authors
Paper under double-blind review

ABSTRACT

Post-AGI governance proposals often emphasize rapid adaptation, stabilization, and preventing *gradual disempowerment* of human institutions. Yet many policy levers—from liability regimes to treaties to benefit-sharing—implicitly assume that researchers, regulators, and civil society can *observe and contest* the allocation of the key bottleneck input: compute. We argue that compute concentration is a first-order socio-economic and safety risk in post-AGI scenarios because it shapes (i) who can build, (ii) who can audit, and (iii) who can deploy and gate high-impact systems. We propose a practical measurement layer: **compute provenance reporting** via standardized *Compute Cards* (public, banded disclosures plus a confidential annex for trusted auditors), paired with simple, robust concentration indices (Top-k, HHI, Gini) computed across training and inference. Using a public dataset of estimated training costs for notable frontier models, we illustrate extreme inequality (Gini $\approx$ 0.89; Top-1 share $\approx$ 54%). We then outline an MVP "dashboard" pipeline, a disclosure threat model, and a research agenda that links measurement to policy evaluation (e.g., public compute, procurement, and competition interventions) while respecting safety and security constraints.

1 MOTIVATION: COMPUTE IS A POST-AGI BOTTLENECK

Work on post-AGI governance foregrounds fast transitions, hard-to-reverse decisions, and the need for institutions that remain legitimate and effective under capability surges (Bostrom et al., 2020; MacAskill & Moorhouse, 2025; Kulveit et al., 2025). A recurring theme is *power concentration*: if a small set of actors controls the capability frontier, they may entrench autocracy, lock in values, or constrain democratic oversight.

Compute is a distinctive locus for this risk because it is (a) rival and capacity-limited, (b) mediated by a highly concentrated hardware and cloud supply chain (see Figure 1), and (c) increasingly required at scale for both training and inference (Vipra & West, 2023; Center for the Governance of AI, 2024; Organisation for Economic Co-operation and Development, 2023).

Empirically, training compute requirements have accelerated beyond Moore-era trends (Amodei & Hernandez, 2018; Sevilla et al., 2022), and frontier training costs have risen sharply (Cottier et al., 2024). At the same time, unequal access to compute can re-shape the research ecosystem toward large firms and elite institutions (Ahmed & Wahed, 2020). These dynamics make compute concentration a natural *early-warning indicator* for several post-AGI "dashboard" approaches: if oversight depends on independent evaluation, red-teaming, and institutional competition, then measuring who can afford and obtain compute is prerequisite infrastructure.

2 WHAT TO MEASURE: EFFECTIVE COMPUTE ACCESS

We distinguish **compute capacity** from **effective access**. Even if global FLOP supply grows, effective access may remain concentrated due to financing, chip scarcity, cloud contracts, export controls, data

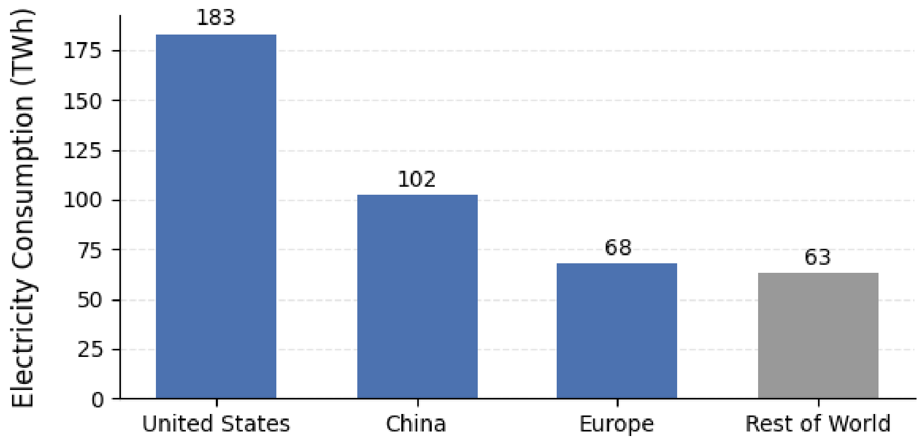

Figure 1: Geographic concentration of physical infrastructure. Approximate regional shares of data-centre electricity use (2024) reported by the IEA (International Energy Agency, 2025).

rights, and distribution bottlenecks. For a model or organization i, define effective access as a vector:

$$A_i \;=\; \big(C_i^{\text{train}},\; C_i^{\text{infer}},\; D_i,\; G_i\big), \tag{1}$$

where C^{train} is training compute (e.g., accelerator-hours, FLOPs, or amortized $ cost), C^{infer} is sustained inference throughput, D summarizes legal/contractual access to key data (e.g., licenses; proprietary corpora), and G captures *gating power* over deployment channels (e.g., dominant APIs, platform integration, or key partnerships). Policy-relevant concentration can then be measured at multiple layers: chips, cloud, training runs, and deployment interfaces (Competition and Markets Authority, 2024; Federal Trade Commission, 2023).

3 Proposal: Compute Cards with safe disclosure

We propose **Compute Cards**: a lightweight, standardized disclosure artifact analogous to model cards and datasheets (Mitchell et al., 2019; Gebru et al., 2021), but scoped to compute provenance and deployment capacity. Compute Cards align with calls to systematically report energy and carbon impacts (Strubell et al., 2019; Schwartz et al., 2020; Henderson et al., 2020), and can be made compatible with emerging safety-tiering practices (Anthropic, 2023; OpenAI, 2023).

Table 1: Proposed Compute Card Schema (Two-Tier Disclosure)

Tier	Fields Disclosed
1. Public (Banded)	(i) Training compute band (e.g., $10^{24} - 10^{25}$ FLOPs) & Accelerator family (ii) Training cost band & uncertainty method (iii) Energy use band & region (iv) Deployment mode (Weights / API / On-device) (v) Access policies (Rate limits, safety tier)
2. Confidential (Auditor Annex)	Exact cluster topology & providers Procurement contracts & security controls Incident logs & utilization rates

Disclosure threat model. Exact cluster details can be competitively sensitive and may create security risks (e.g., targeting scarce hardware). We therefore recommend a *two-tier* design (Table 1): (i) *public* fields are disclosed in bands (e.g., order-of-magnitude FLOPs; energy range; region), and (ii) a *confidential annex* is shared under NDA / regulator privilege with accredited auditors, enabling verification and stronger metrics.

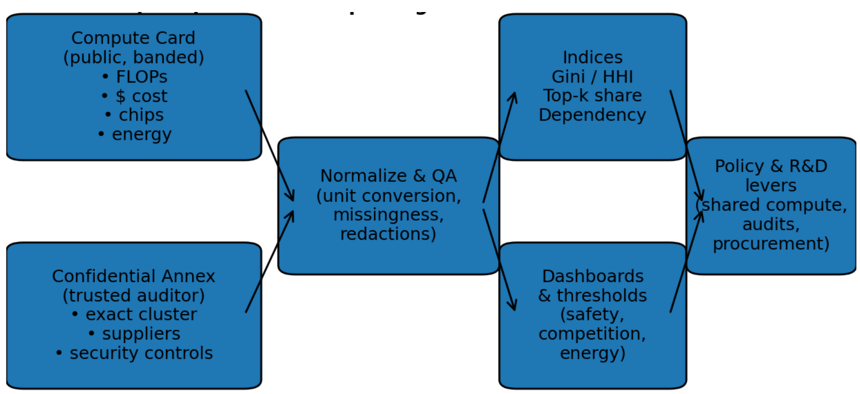

Figure 2: Compute provenance reporting as measurement infrastructure. Public, banded Compute Cards and a confidential annex feed normalization and QA, enabling concentration indices.

4 Metrics and an MVP dashboard

Given a set of disclosed compute magnitudes $\{x_i\}$ (training, inference, or combined), we recommend reporting complementary concentration measures. The Top-k share captures the dominance of the largest players:

$$S_k = \frac{\sum_{i=1}^{k} x_{(i)}}{\sum_j x_j} \tag{2}$$

To capture the full distribution, we employ the HHI and Gini coefficient:

$$\text{HHI} = \sum_i s_i^2 \quad \text{and} \quad \text{Gini} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |x_i - x_j|}{2n^2 \bar{x}} \tag{3}$$

These indices trade off interpretability, robustness to missingness, and sensitivity to tails (Rhoades, 1993; U.S. Department of Justice & Federal Trade Commission, 2010).

Illustration (training cost). Using cost estimates for 61 notable frontier models (Cottier et al., 2024; Epoch Research, 2024), we observe extreme inequality: Gini ≈ 0.89 and Top-1 share $\approx 54\%$ (Figure 3). While cost is an imperfect proxy for compute access, it provides a public lower bound on concentration in state-of-the-art training.

MVP pipeline. A minimal dashboard can be implemented in three steps: (1) ingest Compute Cards (conference submissions; voluntary disclosures; regulator filings), (2) normalize units (FLOPs, accelerator-hours, and $), record uncertainty bands, and handle missingness, (3) publish rolling indices by *capability tier* and *deployment scale*. To connect to sustainability, pair compute with grid carbon intensity and energy estimates (International Energy Agency, 2025; Lacoste et al., 2019; Lannelongue et al., 2021).

5 Levers and evaluation

Measurement is valuable only if it enables interventions to be compared. We highlight four classes of levers:

- **Shared compute for independent scrutiny:** expand public-interest access (e.g., NAIRR) so that academia, journalists, and civil society can reproduce evaluations and audit dominant systems (U.S. National Science Foundation, 2025).
- **Competition & interoperability:** scrutinize cloud/FM partnerships and exclusionary conduct; encourage portability and multi-homing to reduce gating power (Competition and Markets Authority, 2024; Federal Trade Commission, 2023; Open Markets Institute, 2024; Gans, 2024).
- **Safety-tiered access:** align access and release decisions with risk, avoiding naive "open vs closed" framing (Anthropic, 2023; OpenAI, 2023; National Institute of Standards and Technology, 2023).

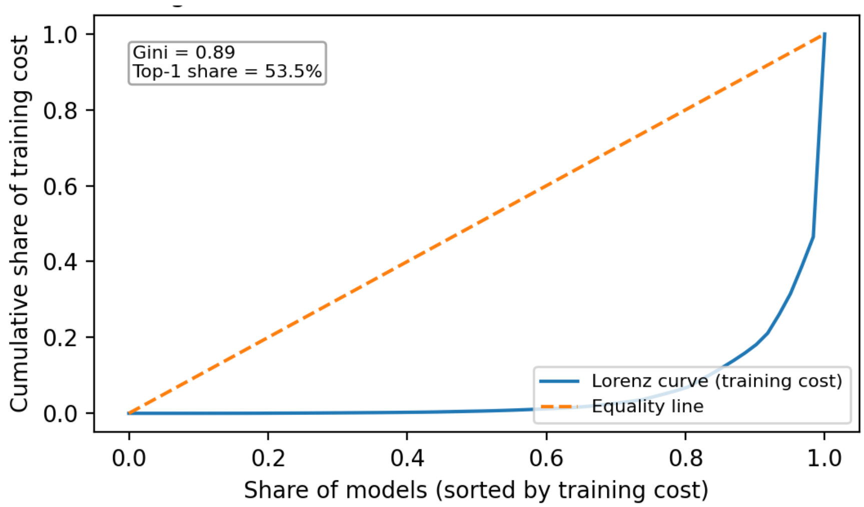

Figure 3: Lorenz curve of estimated frontier-model training costs (61 models). Extreme concentration implies that marginal increases in the frontier are dominated by a small number of runs.

- **Geopolitics & chokepoints:** export controls and chip policy can reshape concentration and access (Bureau of Industry and Security, 2023; Executive Office of the President, 2023; Buchanan, 2020).

Evaluation metrics. A dashboard should track: (i) concentration indices by tier, (ii) disclosure coverage and audit pass rates, (iii) time-to-independent-replication of frontier evaluations, and (iv) energy-per-capability trends. The goal is not to eliminate concentration (which may be technologically constrained), but to prevent *unaccountable* concentration and to preserve contestability under capability acceleration.

Mapping to post-AGI "dashboard" directions This measurement layer operationalizes three recurring directions in post-AGI governance debates: (1) *ecosystem alignment metrics and levers* (we propose measurable indices and link them to interventions), (2) *power-concentration firebreaks* (we connect concentration metrics to competition and procurement levers), and (3) *early-warning indicators and governance dashboards* (Compute Cards enable a rolling public index).

6 Limitations and Research Agenda

Compute Cards will be gamed unless paired with verification. Key open problems include: privacy-preserving reporting, strategic missingness, cross-region comparability, and linking concentration to downstream harms. We recommend piloting Compute Cards as a *submission artifact* for high-compute papers and as an optional disclosure for major model releases, with third-party auditing and redaction guidelines.

7 Conclusion

Post-AGI governance needs instrumentation. Compute concentration is a measurable, policy-relevant proxy for power concentration across the AI stack. Compute Cards plus simple indices can turn "power concentration" from an abstract worry into a monitored, evaluated variable—enabling earlier, more accountable interventions.

References

Nur Ahmed and Muntasir Wahed. The de-democratization of AI: Deep learning and the compute divide in artificial intelligence research. *arXiv preprint arXiv:2010.15581*, 2020. URL `https://arxiv.org/abs/2010.15581`. Accessed 2026-02-06.

Dario Amodei and Danny Hernandez. AI and compute. OpenAI blog post, 2018. URL https://openai.com/index/ai-and-compute/. Accessed 2026-02-06.

Anthropic. Responsible scaling policy. Policy document, 2023. URL https://www.anthropic.com/news/responsible-scaling-policy. Accessed 2026-02-06.

Nick Bostrom, Allan Dafoe, and Carrick Flynn. Public policy and superintelligent AI: A vector field approach. In S. Matthew Liao (ed.), *Ethics of Artificial Intelligence*, pp. 293–326. Oxford University Press, 2020. doi: 10.1093/oso/9780190905033.003.0011. URL https://nickbostrom.com/papers/aipolicy.pdf. Accessed 2026-02-06.

Ben Buchanan. The AI triad and what it means for national security strategy. Center for Security and Emerging Technology (CSET) report, 2020. URL https://cset.georgetown.edu/publication/the-ai-triad-and-what-it-means-for-national-security-strategy/. Accessed 2026-02-06.

Bureau of Industry and Security. Export controls on advanced computing and semiconductor manufacturing items. U.S. Department of Commerce, 2023. URL https://www.bis.gov/press-release/bis-strengthens-export-controls-advanced-computing-items-additional-enhancements. Accessed 2026-02-06.

Center for the Governance of AI. Computing power and the governance of artificial intelligence. GovAI report, 2024. URL https://cdn.governance.ai/Computing_Power_and_the_Governance_of_AI.pdf. Accessed 2026-02-06.

Competition and Markets Authority. AI foundation models: Technical update report. UK CMA report, 2024. URL https://assets.publishing.service.gov.uk/media/661e5a4c7469198185bd3d62/AI_Foundation_Models_technical_update_report.pdf. Accessed 2026-02-06.

Ben Cottier, Raja Sengupta Rahman, Loredana Fattorini, Nestor Maslej, Tamay Besiroglu, and David Owen. The rising costs of training frontier AI models. *arXiv preprint arXiv:2405.21015*, 2024. URL https://arxiv.org/abs/2405.21015. Accessed 2026-02-06.

Epoch Research. Training cost trends dataset (model_costs_updated.csv). GitHub repository, 2024. URL https://github.com/epoch-research/training-cost-trends. Accessed 2026-02-06.

Executive Office of the President. Executive order 14110: Safe, secure, and trustworthy development and use of artificial intelligence. White House executive order, 2023. URL https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artifici Accessed 2026-02-06.

Federal Trade Commission. Cloud computing business practices: Request for information. FTC press release, 2023. URL https://www.ftc.gov/news-events/news/press-releases/2023/11/ftc-launches-inquiry-cloud-computing-business-practices. Accessed 2026-02-06.

Joshua Gans. Market power in AI. *NBER Working Paper*, 2024. URL https://www.nber.org/papers/w32425. Accessed 2026-02-06.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. Datasheets for datasets. *Communications of the ACM*, 64 (12):86–92, 2021. doi: 10.1145/3458723. URL https://dl.acm.org/doi/10.1145/3458723. Accessed 2026-02-06.

Peter Henderson, Jieru Hu, Joshua Romoff, Emma Brunskill, Dan Jurafsky, and Joelle Pineau. Towards the systematic reporting of the energy and carbon footprint of machine learning. *arXiv preprint arXiv:2002.05651*, 2020. URL https://arxiv.org/abs/2002.05651. Accessed 2026-02-06.

International Energy Agency. Energy and AI. IEA analysis / report webpage, 2025. URL https://www.iea.org/reports/energy-and-ai. Accessed 2026-02-06.

Jan Kulveit, Raymond Douglas, Nora Ammann, Deger Turan, David Krueger, and David Duvenaud. Gradual disempowerment: Systemic existential risks from incremental AI development. *arXiv preprint arXiv:2501.16946*, 2025. URL https://arxiv.org/abs/2501.16946. Accessed 2026-02-06.

Alexandre Lacoste, Alexandra Luccioni, Victor Schmidt, and Thomas Dandres. Quantifying the carbon emissions of machine learning. *arXiv preprint arXiv:1910.09700*, 2019. URL https://arxiv.org/abs/1910.09700. Accessed 2026-02-06.

Loïc Lannelongue, Jason Grealey, and Michael Inouye. Green algorithms: Quantifying the carbon footprint of computation. *Advanced Science*, 8(12):2100707, 2021. doi: 10.1002/advs.202100707. URL https://onlinelibrary.wiley.com/doi/10.1002/advs.202100707. Accessed 2026-02-06.

William MacAskill and Fin Moorhouse. Preparing for the intelligence explosion. Forethought report, 2025. URL https://www.forethought.org/research/preparing-for-the-intelligence-explosion.pdf. Accessed 2026-02-06.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAccT)*, 2019. URL https://arxiv.org/abs/1810.03993. Accessed 2026-02-06.

National Institute of Standards and Technology. AI risk management framework (rmf) 1.0. NIST report, 2023. URL https://www.nist.gov/itl/ai-risk-management-framework. Accessed 2026-02-06.

Open Markets Institute. Stopping big tech from becoming big AI. Policy report, 2024. URL https://openmarketsinstitute.org/publications/stopping-big-tech-from-becoming-big-ai/. Accessed 2026-02-06.

OpenAI. Preparedness framework. Policy document, 2023. URL https://openai.com/index/preparedness-framework/. Accessed 2026-02-06.

Organisation for Economic Co-operation and Development. A blueprint for building national compute capacity for artificial intelligence. OECD Digital Economy Papers, No. 350, 2023. URL https://www.oecd.org/en/publications/a-blueprint-for-building-national-compute-capacity-for-artificial-intelligence_876367e3-en.html. Accessed 2026-02-06.

Stephen A. Rhoades. The herfindahl-hirschman index. *Federal Reserve Bulletin*, pp. 188–189, 1993. URL https://www.federalreserve.gov/pubs/bulletin/1993/0393lead.pdf. Accessed 2026-02-06.

Roy Schwartz, Jesse Dodge, Noah A. Smith, and Oren Etzioni. Green AI. *Communications of the ACM*, 63(12):54–63, 2020. doi: 10.1145/3381831. URL https://dl.acm.org/doi/10.1145/3381831. Accessed 2026-02-06.

Jaime Sevilla, Lennart Heim, Anson Ho, Tamay Besiroglu, Marius Hobbhahn, and Pablo Villalobos. Compute trends across three eras of machine learning. *arXiv preprint arXiv:2202.05924*, 2022. URL https://arxiv.org/abs/2202.05924. Accessed 2026-02-06.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2019. URL https://aclanthology.org/P19-1355/. Accessed 2026-02-06.

U.S. Department of Justice and Federal Trade Commission. Horizontal merger guidelines. Policy document, 2010. URL https://www.justice.gov/atr/horizontal-merger-guidelines-08192010. Accessed 2026-02-06.

U.S. National Science Foundation. National artificial intelligence research resource (nairr) pilot. Program webpage, 2025. URL https://www.nsf.gov/focus-areas/ai/nairr. Accessed 2026-02-06.

Jai Vipra and Sarah Myers West. Computational power and AI. AI Now Institute report, 2023. URL https://ainowinstitute.org/wp-content/uploads/2023/09/AI-Now_Computational-Power-an-AI.pdf. Accessed 2026-02-06.