

Fault Lines

The dual-use AI governance vacuum in Asia: a seven-layer structural analysis

Vibha Amarnath

University of Cambridge

With Apart Research · Global South AI Safety Hackathon 2026 · Asia Track (Governance and Geopolitics)

Abstract

On 13 May 2025, the US Bureau of Industry and Security rescinded the AI Diffusion Rule, the only mechanism imposing governance obligations on AI compute access for developing states in Asia. No replacement was announced. The Council of Europe Framework Convention on AI (CETS No. 225), which entered into force on 1 November 2025, exempts national security applications entirely under Article 3.2, excluding the highest-risk dual-use domains from treaty scope. The result is a codified asymmetry: Article 3.2 legally shields the most advanced state-sponsored AI from compliance obligations while leaving developing states without the governance architecture to secure their own infrastructure. This paper introduces a seven-layer diagnostic framework derived from WMD governance precedents — the NPT, CWC, BWC, IAEA safeguards, and ICAO Annex 19 — and applies it across nine countries in Asia and the West. Three findings emerge: of seven required governance layers, AI has fragments of one, rescinded before enforcement; unilateral export controls accelerated the integration of unmonitored AI into Asian critical infrastructure rather than preventing it; and every institution with a dual-use enforcement mandate has the wrong jurisdiction for AI, while every institution with potential AI jurisdiction has no enforcement mandate. This structural mismatch explains why the policy vacuum has persisted despite widespread recognition of the problem, and why it cannot be closed by existing actors working harder within current mandates. The paper offers mechanism-level treaty design specifications for each of the six absent layers, as a resource for the 2027 New York AI governance negotiations. An accompanying interactive tool, Fault Lines - <https://fault-lines.netlify.app> makes the analysis navigable for treaty practitioners and researchers.

1. Introduction

On 13 January 2025, the United States published the Framework for Artificial Intelligence Diffusion (Federal Register 90 FR 4544), establishing the first systematic governance architecture for AI compute access across a three-tier country classification. On 13 May 2025, two days before its compliance deadline, the rule was rescinded, with no replacement announced. The Council of Europe Framework Convention on AI (CETS No. 225) had opened for signature in September 2024 and entered into force on 1 November 2025. Its Article 3.2 exempts all activities related to national security from treaty obligations excluding precisely the applications where dual-use risk is highest: military AI, biosecurity, and critical infrastructure. India, Vietnam, Indonesia, and Singapore are not party to the Convention. Together, these developments left Asian developing states with no binding mechanism governing AI compute access and no binding treaty with meaningful dual-use provisions.

This paper examines what those two events produced and why it matters. It addresses the **Governance and geopolitics** sub-track of the Asia track, covering compute diffusion and access, AI sovereignty, and national data security.

The structural reality this has produced is not a passive gap in an otherwise functioning system. It is a codified asymmetry. The mechanism of codification is Article 3.2 of the CoE Convention, which exempts “all activities within the lifecycle of AI systems related to the protection of national interests” from treaty obligations, a carve-out that legally shields the most advanced Western military and intelligence AI from any compliance requirement while applying equally to any developing state that invokes the same clause. In practice, only states with the institutional capacity to control external scrutiny can exploit that protection. International frameworks now legally shield state-sponsored, high-consequence military AI in the West while systematically denying developing nations the governance architecture required to evaluate and safeguard the AI operating in their own infrastructure. Working dual-use governance regimes like nuclear, chemical, and aviation safety, have historically required seven distinct functional layers to maintain stability. The biological weapons regime, which lacks verification (Layer 6), is the

counter-example that proves the point: it is the least effective of the four, and AI-enabled bioweapon design is now the highest-risk ungoverned dual-use application as a direct consequence.

One analytical distinction is critical. The Diffusion Rule never came into force, so its rescission did not dismantle an operating regime. It was an active, deliberate policy choice to leave the compute diffusion vector unmonitored — made at the precise moment that frontier open-weight models from geopolitical rivals began entering Asian telecoms, banking, and civil infrastructure without any mandatory pre-deployment safety evaluation.

The implications for AI safety are concrete. On biosecurity: while frontier LLMs currently provide modest uplift for biological attack planning (RAND 2024; OpenAI 2024; Anthropic 2024), GovAI (2026) documents that coding agents are materially altering this risk landscape, and open-weight biological foundation models such as Evo 2 and ESM3 are deployed internationally without the AISI pre-deployment evaluations mandated for Western closed-source equivalents. On electoral interference: AI-generated synthetic media deployed at cross-border electoral scale operates with no standardized attribution mechanism and no accountability framework. On critical infrastructure: the four confirmed cases in Section 4 document AI deployments in telecoms, banking, and sovereign systems with no incident reporting obligation and no international body empowered to audit cross-border failures.

Contributions of this submission

The following are original to this submission and distinct from the prior work addressed in Section 2: (1) **The Seven-Layer Structural Matrix** — a diagnostic framework derived from primary WMD treaty texts and applied to the June 2026 AI governance landscape for the first time; (2) **Empirical Mapping of Regional Redirection** — four confirmed cases where unilateral export controls accelerated integration of unmonitored AI into sovereign infrastructure rather than preventing it; (3) **The Structural Mismatch Analysis** — a systematic mapping of eight international institutions demonstrating that enforcement authority and AI jurisdiction are architecturally separated, not merely politically misaligned; (4) **Fault Lines** - an interactive policy briefing tool (<https://fault-lines.netlify.app>) providing clause-level treaty specifications for each missing governance layer, designed for use by treaty practitioners preparing for the 2027 New York negotiations. None of the prior work cited in Section 2 engages the specific post-May 2025 regulatory vacuum.

2. Related Work

The closest prior work is Sastry et al. (2024) on compute governance, which provides the most rigorous treatment of Layer 1 of the framework developed here. This paper builds on that foundation but demonstrates that Layer 1 alone is structurally insufficient — its political vulnerability, demonstrated by the May 2025 rescission, reveals the urgent necessity of six additional institutional layers that compute tracking was never designed to address. Heim et al. (2024) on training compute thresholds addresses related territory but does not engage with the post-rescission environment.

Hadfield and Clark (2023) and Dafoe (2018) provide the normative context for AI governance design. This paper departs from their approach: rather than specifying what an idealized AI regulatory body should look like, it examines what functional architecture dual-use technology regimes have empirically required to operate — using the WMD parallel as a structural benchmark, not a rhetorical comparison.

On institutional fragmentation specifically: Roberts et al. (2024) map the nascent landscape of international AI governance regimes and conclude that a governance deficit persists due to inadequacy of existing initiatives and difficulties reaching agreement on appropriate mechanisms. Czerniawski (2020) examines competing architectures for international AI governance and concludes that fragmentation will likely persist. Both papers document the *existence* of fragmentation. This paper goes further by identifying the *structural mechanism* that makes fragmentation architecturally stable rather than merely politically contingent: every institution with enforcement authority was built for a different threat domain, and every institution with AI scope was deliberately designed without enforcement authority. That pattern cannot be corrected by coordination — it requires new institutional architecture. This is the contribution the existing fragmentation literature has not made.

Birhane et al. (2022) and Mohamed et al. (2020) document how AI research concentrates power in Western institutions and encodes their valuations into global systems. This paper operates within that decolonial critique but focuses on a downstream operational failure: the structural asymmetry that leaves developing states — who inherit the collateral risks of frontier deployments — without any multilateral mechanism to evaluate or contest the AI integrated into their national infrastructure. The two problems are related: the structural mismatch documented in Section 4.3 is partly a consequence of the same institutional concentration that Birhane et al. describe.

3. Methods

3.1 The seven-layer framework

The framework derives its seven layers from the functional requirements of working WMD governance regimes, sourced from primary treaty texts. Each layer corresponds to a specific institutional failure mode in governance history that required a specific institutional solution. The layers are not derived from first principles but from what was actually built in response to dual-use governance failures in nuclear, chemical, biological, and aviation domains.

Layer	Function	WMD precedent	AI status — June 2026
L1	Physical precursor / compute control	NSG trigger lists; EAR chemical precursor bounds	Rescinded May 2025 — never enforced
L2	Capability identification	CWC Schedules 1–3; IAEA fissionable material categories	Partial — Western frontier labs only, voluntary, unverified
L3	Use-case differentiation	CWC Article VI facility sizing; NPT civilian enrichment thresholds	Entirely absent — no standard differentiates safe from dangerous use at point of access
L4	National governance capacity	CWC Article VII mandatory national authorities	Fragmented — isolated domestic laws (EU, UK, Vietnam) without international parity or interoperability
L5	Incident attribution	OPCW Technical Secretariat; international nuclear forensics	Entirely absent — no cross-border model tracking or forensic tracing framework
L6	International verification	IAEA safeguards (INFCIRC/153); OPCW challenge inspections	Entirely absent — no international right to inspect training data, weights, or compute clusters
L7	Consent and representation	NPT Article IV technology transfer; CWC multilateral negotiation	Largely absent — first inclusive forum opens 2026; Global South excluded from prior rule design

Table 1. Seven-layer dual-use governance framework mapped against AI status as of June 2026.

3.2 Country selection and urgency scoring

Nine countries were selected to represent the full governance spectrum: rule-designers (US, UK), a rule-maker (EU), Tier 1 integrated peripheral states (Japan, South Korea), and states navigating the vacuum (India, Vietnam, Indonesia, Singapore). Including Western comparators makes the structural contrast visible rather than implied.

Each country receives an urgency score weighted as: 40% national governance capacity gap (inverse of Oxford Insights Government AI Readiness Index 2024 — 188 countries, 40 indicators, G20 and UNESCO benchmark); 40% dual-use risk exposure across documented biosecurity and cyber-vulnerability profiles; and 20% export control exposure derived from BIS tier classification post-rescission (Federal Register 90 FR 4544). Scores are directional estimates, not precise measurements. Critically, India, Vietnam, and Indonesia are not OECD members; their complete absence from OECD AI Policy Observatory quantitative data streams is itself an empirical confirmation of the Layer 7 representation gap this paper identifies.

Country	Oxford AI Readiness 2024	OECD Member	Export control status	Layers covered (of 7)	Urgency score
Indonesia	65.85 / rank ~55	No	Vacuum post-rescission	0 / 7	9.4
India	~64.0 / rank 46	No	Vacuum post-rescission	0 / 7	8.1
Vietnam	61.42 / rank ~65	No	Vacuum post-rescission	1 / 7 (L4 partial)	7.8

Country	Oxford AI Readiness 2024	OECD Member	Export control status	Layers covered (of 7)	Urgency score
Singapore	84.25 / rank 2	No	Vacuum — same tier as Indonesia	2 / 7 (L2, L4 partial)	6.2
Japan	75.75 / rank 11	Yes	Tier 1 via semiconductors	3 / 7	4.8
South Korea	79.98 / rank 3	Yes	Tier 1 via Samsung/SK Hynix	3 / 7	4.2
United States	87.03 / rank 1	Yes	Rule designer	5 / 7	Comparator
United Kingdom	78.88 / rank 5	Yes	Tier 1 co-designer	5 / 7	Comparator
European Union	~76 avg / Top 10	Yes (members)	Tier 1, AI Act binding	5 / 7	Comparator

Table 3. Country urgency scores and governance stack coverage. Oxford Insights score source: Government AI Readiness Index 2024 (188 countries). Layers covered reflects presence (full or partial) of governance mechanisms in Table 1. Urgency scores are directional composites, not precise measurements.

3.3 Two analytical boundaries

On open-weight models: DeepSeek and Meta Llama are freely downloadable. A reasonable objection is that compute controls would not have prevented their adoption regardless of the Diffusion Rule. This is likely true — and is precisely the point. The rescission matters because they represented the only mechanism creating any governance obligation over AI access at all. Open-weight models make Layer 1 governance permanently incomplete. This is why Layers 2–7 are not supplementary to a working system — they are the only architecture that can reach open-weight AI after release. The governance vacuum is structural in both directions.

On the WMD comparison: The comparison is functional, not a claim of risk equivalence. Dual-use AI shares the structural properties that made WMD governance necessary: identical technology for beneficial and harmful purposes, inability to distinguish intent at the point of access, and asymmetric harm potential. The governance principles the CWC and IAEA instantiated are not chemistry-specific but they are responses to that structural problem, and they apply wherever the problem recurs.

4. Results

4.1 Finding 1: of seven required governance layers, AI has one — rescinded before it came into force

Applying the framework to AI as of June 2026: nuclear governance maintains all seven layers; chemical, six; biological, three — the missing verification layer (L6) is precisely why AI-enabled bioweapon design is the highest-risk ungoverned dual-use application today. AI has fragments of Layer 4 in a handful of Western jurisdictions, and a rescinded Layer 1.

The Layer 1 collapse is precisely documented. The Diffusion Rule (90 FR 4544) was the first attempt to apply systematic governance logic to AI compute access. Its three-tier country classification and controls on model weights represented a genuine architectural step. It was rescinded on 13 May 2025 with BIS citing that it “would have stifled American innovation.” The current US chip control regime is the pre-2022 EAR baseline — a trade-war architecture with no AI safety mandate and no capacity to enforce alignment or safety protocols.

Layer 3 — use-case differentiation — presents an additional structural challenge. An advanced open-weight model uses identical mathematical weights whether executing a pathogen gene sequence analysis or an epidemiological defence script. Deepfake detection and deepfake generation are the same architectural problem run in reverse. No access-control mechanism can distinguish intent at the point of delivery. This is why Layers 5 and 6 (attribution and verification) are not optional additions — they are the only mechanisms that can address the dual-use gap after it has been exploited, once the access-layer distinction has been shown to be insufficient.

4.2 Finding 2: export controls accelerated unmonitored AI integration

Rather than constraining technology development, Western export restrictions drove systemic integration of unmonitored AI into Asian critical infrastructure. Four confirmed cases document this redirection, with a fifth (Thailand) noted in the table caption pending primary source confirmation:

Country / Entity	What happened	Infrastructure	Operational governance
Indonesia — Indosat Ooredoo Hutchison	MoU signed Jan 2025 with AlonOS to develop solutions on DeepSeek for food security and tourism	DeepSeek (open-weight)	None
Malaysia — Skyvast Cloud	Deputy Minister announced sovereign AI on Huawei Ascend GPUs (May 2025); retracted next day under US diplomatic pressure. Huawei–Skyvast MoU for Ascend-powered sovereign cloud confirmed separately via Ministry of Digital.	Huawei Ascend (announced, disputed)	None — decision made under geopolitical pressure, not safety evaluation
Singapore — OCBC Bank	30+ GenAI applications deployed (Annual Report 2024); DeepSeek reasoning models cited as capability being adopted. MAS-designated systemically important bank.	DeepSeek (cited); multiple providers	None for Chinese-origin models
Vietnam — enterprise sector	DeepSeek and Alibaba Cloud deployments across enterprise sector. Law 134/2025 binding from March 2026 but lacks jurisdiction over Chinese-origin models.	Alibaba Cloud; DeepSeek	Law 134/2025 (domestic scope only)

Table 2. Four confirmed AI infrastructure cases in Asian sovereign networks. Sources: Business Standard (2025); Developing Telecoms (2025a, 2025b); OCBC Annual Report 2024; Pertama Partners (2026). A fifth case — simultaneous Microsoft and ByteDance datacentre investment in Thailand — is documented in regional trade press but excluded here pending primary source confirmation.

The Malaysia episode is particularly instructive. A government minister announced a sovereign AI infrastructure decision at a state event; informal US diplomatic pressure followed; the statement was publicly retracted the next day with no official explanation. No multilateral body had any jurisdiction over the decision. No international framework required a safety evaluation of either the Huawei hardware or the DeepSeek model. A developing state was forced to make a high-consequence national infrastructure choice on the basis of geopolitical pressure rather than verifiable safety standards. That is the governance vacuum operating in real time.

Vietnam exposes a different dimension of the failure. Law No. 134/2025/QH15 — the first binding, standalone AI law in Southeast Asia, in force from 1 March 2026 — establishes risk-based classification, mandatory conformity assessments for high-risk systems, and penalties for non-compliance. At the federal level, the United States has no equivalent statute. Yet Vietnam sits in the same compute access vacuum as Indonesia, which has no AI governance at all. The current architecture is entirely blind to governance quality. This creates a direct disincentive for Global South legislative investment — a country that builds rigorous domestic AI law receives the same international treatment as one that builds nothing.

4.3 Finding 3: the structural mismatch

The most analytically original finding of this paper is not the absence of AI governance mechanisms — that is well-documented — but the reason they have not been built. Mapping eight international institutions against their mandates produces a consistent pattern:

The structural mismatch: every institution with a dual-use enforcement mandate has the wrong technical jurisdiction for AI. Every institution with potential AI scope has no enforcement mandate. This is not primarily a political failure — it is an architectural one. Institutions with enforcement

capacity (OPCW, IAEA, BIS) were built for legacy threat domains. Institutions created specifically for AI (Frontier Model Forum, AISI network) were designed without enforcement authority.

BIS: WMD containment mandate; rescinded its own AI governance rule. **OPCW:** bound tightly to CWC text; cannot audit cloud-based generative biochemical design without state-level protocol extension.

CCW GGE: mandate concludes at the Seventh Review Conference (November 2026); ten years of discussion without binding output; any High Contracting Party can block by consensus. **CoE**

Convention: Article 3.2 structurally immunizes highest-consequence AI from compliance tracking; entered into force November 2025; no Asian developing state is party. **IAEA:** gold standard for physical material verification; charter lacks jurisdiction over non-nuclear compute clusters. **ITU:** technical standards only; no enforcement authority; well-positioned for AI incident reporting standards. **WHO:** has guidance on AI in health contexts; no dual-use mandate and no enforcement authority; the biosecurity-AI intersection has no institutional home with both expertise and binding authority. **UN GGE on Cybersecurity:** non-binding norms; fragmented by parallel OEWG process.

Because this mismatch is architectural rather than political, it cannot be resolved by existing institutions working harder within current mandates. Incremental expansion is a dead end. Closing the structural asymmetry requires a purpose-built multilateral treaty architecture that creates new institutional capacity, rather than extending the mandates of bodies that were never designed for this domain.

5. Discussion and Limitations

5.1 Implications for the 2027 treaty

The three findings converge on a single implication: the 2027 New York sessions must move beyond non-binding normative declarations and commit to specific, multi-layered institutional design. Layer 4 exists in fragments in a small number of Western jurisdictions; the remaining six layers are structurally absent for all states in the governance vacuum. The Fault Lines tool provides mechanism-level specifications for each of those six absent layers, with draft treaty language modelled on WMD precedents. Two mechanisms warrant specific argument here because they are the most politically consequential.

The first is a sovereign governance compute carve-out modelled on NPT Article IV. NPT Article IV stabilised global nuclear posture through a structural bargain: peripheral states surrendered weapons-development pathways in exchange for legally guaranteed access to civilian nuclear technology. The AI equivalent must mirror this architecture — developing states accept mandatory governance compliance and international verification in exchange for unthrottled access to frontier compute and verified foundation model weights for sovereign economic development. Vietnam is the clearest evidence for why this matters: Law 134/2025 makes it more rigorously governed than the United States at the federal level, yet it faces the same compute access vacuum as Indonesia, which has enacted nothing. Without this structural inducement, Global South states have no incentive to accept the sovereignty costs a treaty would impose, and the governance architecture remains permanently fractured.

The BWC is the relevant historical warning. Verification negotiations ran for seven years before the US withdrew in 2001, citing trade secret concerns and the technical difficulty of inspecting dual-use biological facilities. Both arguments have already re-emerged from frontier AI developers in 2025–26 — model-weight inspections are framed as compromising proprietary IP, and the technical difficulty of auditing training runs is real. The BWC outcome is a treaty that cannot detect the violations it was designed to prevent. Verification must be built into the foundational text of the 2027 treaty from day one, not delegated to a separate protocol that can be blocked or delayed indefinitely.

A simulation of five intervention scenarios against the seven-layer framework produces one clear finding: no single intervention is sufficient. A 2027 treaty with binding verification addresses Layers 5–7 but leaves Layers 1–3 unaddressed. A governance compute carve-out addresses Layer 4 for qualifying states. Closing Article 3.2 of the CoE Convention partially addresses Layer 3. The implication is that treaty design must be multi-layered from the outset, not sequenced through future protocols that history shows can be blocked.

5.2 Dual-use considerations

One structural concern warrants explicit acknowledgement. Detailing the specific jurisdictional blindspots of the OPCW, BIS, and CCW GGE carries an inherent dual-use risk: it makes visible exactly where non-aligned actors can

navigate structural gaps to deploy unverified AI within sensitive sectors. This paper judges that this informational risk is outweighed by the necessity of providing specific structural blueprints to international negotiators before the 2027 window closes — the same judgement that made the publication of CWC verification protocols defensible despite their dual-use implications. Transparency about governance failures is a precondition for fixing them.

Limitations

- **Urgency score methodology:** the 40/40/20 weighting reflects this paper's focus on dual-use structural exposure. Alternative weightings would produce different state rankings.
- **Layer assessment:** classifying each layer as present, partial, or absent involves judgement calls, particularly for Layer 4 where governance capacity varies significantly within Tier 1 states.
- **Deployment data:** the five cases in Table 2 represent a documented floor. Unmapped private enterprise integration of Chinese-origin models across Southeast Asia is likely significantly higher.
- **Layer 3 may be partly irreducible:** use-case differentiation faces a genuine technical constraint. Governance can address the attribution and accountability dimensions post-deployment; it cannot make the access-layer distinction at the point of delivery.

Future Work

The framework should be applied to Latin American and African regional infrastructure to map analogous structural gaps. Future iterations should integrate ASEAN-level governance structures to test whether mid-tier multilateral bodies can serve as interim verification buffers while global treaty mechanics are under construction. Empirical validation of urgency scores against incident data would strengthen the methodology.

6. Conclusion

The dangerous structural asymmetry defining contemporary AI safety is the direct result of two compounding failures: the deliberate dismantling of compute-access governance at the moment Chinese open-weight models were entering Asian critical infrastructure, and an international treaty framework that structurally immunizes high-risk national security AI from global accountability — operating on top of an institutional landscape where every body with enforcement capacity lacks AI jurisdiction, and every body with AI scope lacks enforcement capacity.

The seven-layer WMD framework introduced here is not the final word on AI governance architecture. It is an attempt to shift the policy discourse from abstract calls for better governance to the mechanical design specificity that treaty negotiation actually requires. The physical properties of silicon chips and neural network weights differ from enriched isotopes and nerve agents in important ways. But the underlying institutional challenges are structurally identical: the verification of dual-use technologies across sovereign boundaries, the attribution of harms that cross jurisdictions, and the design of compliance systems that can detect the violations they were built to prevent.

The BWC record makes the stakes concrete: the verification protocol that failed was a general call for accountability; the CWC verification regime that succeeded was a specific institutional design with defined inspection rights, confidentiality protections, and enforcement referral paths. The 2027 New York sessions represent the first structured opportunity to apply that level of design specificity to AI — with genuine Global South participation and a treaty mandate. The positions advanced in the dialogue phase directly shape what the 2027 session is mandated to address. What treaty practitioners do with this specification is the next step.

Code and Data

- **Fault Lines (interactive tool):** <https://fault-lines.netlify.app> — a six-tab self-contained HTML policy briefing instrument mapping the seven-layer stack, WMD precedent analysis, accountability scenarios, and mechanism-level treaty design specifications.
- **Primary data sources:** US Federal Register 90 FR 4544; BIS May 2025 rescission statement; CoE CETS No. 225; Vietnam Law 134/2025/QH15; OCBC Annual Report 2024; Oxford Insights Government AI Readiness Index 2024; IAEA INFCIRC/153.

References

- Birhane, A., et al. (2022). The values encoded in machine learning research. *Patterns*, 3(6). <https://doi.org/10.1016/j.patter.2022.100490>

- Business Standard. (2025). AlonOS, Indosat collaborate on DeepSeek. https://www.business-standard.com/world-news/aionos-indosat-collaborate-on-deepseek-to-drive-indonesia-s-ai-journey-125012700796_1.html
- Council of Europe. (2024). CETS No. 225. Entered into force 1 November 2025. <https://rm.coe.int/1680afae3c>
- Czerniawski, M. (2020). Fragmentation and the future: Investigating architectures for international AI governance. *Global Policy*, 12(S6). <https://doi.org/10.1111/1758-5899.12890>
- Dafoe, A. (2018). AI governance: A research agenda. Future of Humanity Institute. <https://arxiv.org/abs/1901.00854>
- Developing Telecoms. (2025a). Malaysia launches sovereign AI stack with DeepSeek and Huawei GPUs. <https://developingtelecoms.com/telecom-technology/data-centres-networks/18504>
- Developing Telecoms. (2025b). Malaysia ministry backtracks on use of Huawei GPUs. <https://developingtelecoms.com/telecom-technology/data-centres-networks/18512>
- GovAI. (2026). Coding agents are changing the biosecurity risk landscape. <https://www.governance.ai/analysis/coding-agents-are-changing-the-biosecurity-risk-landscape>
- Hadfield, G. K., & Clark, J. (2023). *Regulatory Markets: The Future of AI Governance*. MIT Press.
- Heim, L., et al. (2024). Training compute thresholds: Features and functions in AI governance. arXiv:2405.10799.
- IAEA. (1972). INFCIRC/153. <https://www.iaea.org/sites/default/files/publications/documents/infcircs/1972/infcirc153.pdf>
- ICAO. (2013). Annex 19 — Safety Management. <https://www.icao.int/safety/SafetyManagement/Pages/Annex-19.aspx>
- Mohamed, S., et al. (2020). Decolonial AI. *Philosophy & Technology*, 33, 659–684.
- Mouton, C., et al. (2024). The operational risks of AI in large-scale biological attacks. RAND Corporation. https://www.rand.org/pubs/research_reports/RRA2977-2.html
- OCBC Bank. (2025). Annual Report 2024. <https://www.ocbc.com/iwov-resources/sg/ocbc/gbc/pdf/investors/annual-reports/2024/ocbc-annual-report-2024.pdf>
- Oxford Insights. (2024). Government AI Readiness Index 2024. <https://oxfordinsights.com/wp-content/uploads/2024/12/2024-Government-AI-Readiness-Index-2.pdf>
- Pertama Partners. (2026). Vietnam AI Law 2025. <https://www.pertamapartners.com/insights/vietnam-ai-law-134-2025-compliance-guide>
- Roberts, H., et al. (2024). Global AI governance: barriers and pathways forward. *International Affairs*, 100(3), 1275–1286. <https://doi.org/10.1093/ia/iaae073>
- Sastry, G., et al. (2024). Computing power and the governance of AI. arXiv:2402.08797.
- UN General Assembly. (2025). A/RES/79/325. <https://undocs.org/A/RES/79/325>
- US BIS. (2025a). Federal Register 90 FR 4544. <https://www.federalregister.gov/documents/2025/01/15/2025-00636/framework-for-artificial-intelligence-diffusion>
- US BIS. (2025b). Rescission of AI Diffusion Rule. <https://www.bis.gov/press-release/department-commerce-announces-rescission-biden-era-artificial-intelligence-diffusion-rule-strengthens>
- Vietnam National Assembly. (2025). Law 134/2025/QH15. <https://english.luatvietnam.vn/law-no-134-2025-qh15-dated-december-10-2025-of-the-national-assembly-on-artificial-intelligence-422299-doc1.html>

LLM Usage Statement

Claude (Anthropic) was used throughout this project to stress-test the seven-layer framework, draft and iterate on sections of this paper, build the Fault Lines interactive tool, and verify empirical claims against primary sources. All factual assertions, legislative citations, and treaty references were independently cross-checked by the author against primary legal texts. The structural mismatch argument, case deployment analysis, framework derivation, and policy conclusions represent original analytical work conceived and executed by the author during the hackathon.