



THE OPEN-MARKET PARADOX

How America's Campaign Against Chinese Open-Source AI
Betrays the Principles It Claims to Defend

By

Kelly Cunningham, CTO
AI Rebooted Incorporated

Executive Summary

A campaign is underway in Washington to restrict American access to open-source artificial-intelligence models developed in China. It is being sold as a matter of national security. This paper argues that the security rationale, while not entirely empty, is largely a costume worn by a much more ordinary motive: the protection of a handful of well-capitalized incumbents from a form of competition they did not expect and cannot easily beat on price.

The timing is the tell. For years, open-weight AI models were released, celebrated, and treated as a competitive advantage - so long as the leading ones bore American names. The objection to open weights materialized at almost the precise moment that Chinese open models reached rough capability parity with closed Western systems while costing, in some cases, a small fraction as much to run. A principle that survives right up until the moment it becomes inconvenient was never really a principle. It was a market position.

The argument that follows has five parts. First, that the restrictionist push is driven primarily by commercial self-interest, and that the internal contradictions of the pro-restriction coalition give the game away. Second, this represents a striking hypocrisy: a country and an industry that lecture the world about free markets and competition are reaching for protectionism the instant they begin to fall behind. Third, that the specific national-security claims about open weights are considerably weaker than advertised, and in several cases confuse a downloadable file with a hosted service. Fourth, that even the genuine residual risks are far better addressed by origin-neutral, deployment-layer governance than by provenance-based bans. And fifth, that on the present facts, China's open approach to AI is more defensible (i.e. more consistent, more economically rational, and more aligned with the stated values of the liberal international order) than the closed, protectionist turn the United States is contemplating.

None of this requires believing that the Chinese state is benign, that open models carry no risks, or that origin is irrelevant to a threat assessment. It requires only noticing the difference between a security argument and a security *pretext* and being willing to say which one is doing the work.

1. A Principle That Expired on Schedule

There is a certain kind of argument that is only ever made by the party that is losing. When a company dominates a market, it champions competition, consumer choice, and the free flow of goods and ideas. When that same company begins to lose, the language changes. Suddenly the conversation is about security, stability, standards, and the hidden dangers of the upstart. The principles do not evolve; the market share does, and the principles follow it down.

The current effort to wall off Chinese open-source AI models from American developers and companies is a textbook case. To understand why, it helps to two things. The first is that open-weight models, models whose trained parameters are published for anyone to download, inspect, run, and

modify, were, until very recently, an American calling card. Meta's release of its Llama family of open-weight models was widely praised in the same Washington and Silicon Valley circles now sounding the alarm. Openness was framed as a strength: it seeded an ecosystem, it accelerated research, it kept the United States at the center of global AI development. Nobody convened emergency briefings about the national-security implications of downloadable model weights when the weights in question were made in California.

The second fact is what has changed. Over roughly eighteen months, Chinese laboratories - DeepSeek, Alibaba's Qwen team, Moonshot AI (Kimi), Zhipu AI (GLM), MiniMax, and others - released a succession of open-weight models that closed most of the capability gap with the best closed Western systems while undercutting them dramatically on cost. Data from OpenRouter, a widely used routing layer for AI model traffic, tells the story starkly: Chinese open-weight models went from under 1.2 percent of weekly token consumption in late 2024 to a majority share of consumption among the most-used models by mid-2026, with snapshots of American startup traffic showing Chinese models handling roughly half of all tokens. Programming and agentic workloads - the serious, production-grade uses, not casual experiments - drove much of the shift.

It was at *this* point, and not before, that open weights became a national security problem. That sequence is the single most important fact in the entire debate, and no security argument can explain it. A model architecture is not more dangerous because of the flag on the passport of the researchers who trained it. What changed was not the technology's threat profile. What changed was who was winning.

2. What Is Actually on the Table

It is worth being precise about the policy landscape, because vagueness is where a lot of the persuasion happens. The restrictionist agenda is not a single bill but a cluster of measures at varying stages of seriousness.

At the concrete end, the No DeepSeek on Government Devices Act, introduced in 2025 by Representatives Josh Gottheimer and Darin LaHood, would bar the use of DeepSeek on federal government devices, citing national-security and data-privacy concerns. Beyond that, officials have floated designating Chinese models as a supply-chain risk, which would restrict how companies that do business with the federal government could use them - a designation that functions as a soft ban by making the models radioactive for any firm with government contracts. There is parallel activity on the hardware side, such as the GAIN AI Act, aimed at tightening the export of advanced AI chips, which analysts at the Center for Strategic and International Studies have warned could undermine the very U.S. chipmakers it is meant to protect. And there have been recurring reports of a coming White House move to broaden restrictions on Chinese models, reportedly reinvigorated after each new headline-grabbing Chinese release.

The stated justifications fall into a few buckets: that the models contain censorship baked in at Beijing's direction; that they might hide back doors; that hosted versions route user data to servers in China; and that American reliance on Chinese models constitutes a dangerous dependency. Each of these deserves a serious hearing, and Section 5 gives them one. But before evaluating the arguments, it is worth looking at who is making them, and what those advocates stand to gain.

3. Follow the Money

The most efficient way to understand a policy is often to ask who profits from it. The beneficiaries of a wall around Chinese open models are not difficult to identify. They are the small number of firms whose business model depends on you renting intelligence from them by the token, at prices they set, through an interface they control.

3.1 The Commercial Threat Is Real - and That Is the Point

Consider the economics that the security framing is carefully draped over. Published comparisons in mid-2026 put the cost of a leading Chinese open model's output at well under a dollar per million tokens, against twenty-five to thirty dollars per million tokens for the flagship closed models from the largest American labs - a difference of more than an order of magnitude, and by some accounts nearly thirtyfold. On many standard benchmarks, the gap in capability has narrowed to single digits; on the hardest, contamination-resistant tests, a real gap remains at the very frontier. But most commercial work does not require the frontier. It requires "good enough, cheap, and yours to control," and on that combination the open Chinese models are not just competitive with the closed American ones; they are winning outright.

When a competitor offers ninety percent of your product's capability at three percent of its price, you have two options: cut your prices toward theirs, or change the subject to national security. Only one of those options protects your margins.

This is not a conspiracy theory; it is just incentives. A firm that has raised money at a valuation predicated on selling closed-model access at premium prices faces an existential threat when a free, downloadable model does most of the same work for pennies. The rational response - rational for the *firm*, if not for the country - is to raise the cost of the alternative by any available means. If you cannot win on price, you can try to win on policy. "This competitor is a security threat" is a far more palatable thing to tell a regulator than "this competitor is cheaper than we can afford to be."

3.2 The Coalition That Cannot Agree Gives the Game Away

If the driving concern were genuinely national security, we would expect the security-minded coalition to be roughly aligned on what security requires. It is not. The fissures within the pro-restriction camp reveal that the participants are pursuing their own commercial interests and merely converging, opportunistically, on the security vocabulary.

The clearest example is the split between the frontier labs and the chipmakers. Firms like OpenAI, whose interest is in suppressing a cheaper substitute, favor tighter controls. Nvidia, whose interest is in selling as many chips into as many markets as possible, has argued the opposite - that restrictions should be *loosened*, on the theory that keeping China hooked on American hardware is the better strategic play. These two positions cannot both be what “national security” demands. They can, however, both be what corporate interest demands. When the members of a security coalition are pulling in opposite directions, and each direction happens to coincide precisely with the arguer’s revenue model, the security frame is not the cause of their positions. It is the window dressing.

3.3 “American” Companies Are Not So American

There is a further cynical sleight of hand embedded in the patriotic framing. The entities being protected are routinely described as “American companies,” as though the policy were defending the national interest as such. But the largest AI firms are multinational corporations. They incorporate, bank, and book revenue across jurisdictions; they sell into global markets including, wherever permitted, China itself; they manufacture and source across borders; and their duty runs to a global shareholder base, not to the American public. Their interests align with the national interest exactly when it is convenient and diverge from it the rest of the time.

This matters because the entire rhetorical structure of the restrictionist case depends on collapsing “what is good for these particular firms” into “what is good for America.” Once that collapse is refused, once we insist on treating a multinational’s margin and a nation’s security as the distinct things they are, the argument loses most of its moral force. Protecting a multinational’s pricing power from a cheaper foreign competitor is a perfectly ordinary act of industrial protectionism. It is not a defense of the republic and dressing it as one should not earn it a discount at the counter of public opinion.

3.4 The Conflation Trick: Weights Versus Service

The most revealing move in the whole debate is a quiet conflation that runs through nearly every alarming headline: the blurring of a *hosted service* with a *downloadable model*. When critics warn that “DeepSeek sends your data to China,” they are describing the behavior of a phone app or a hosted API - a service, running on someone else’s servers, that naturally transmits data to those servers. That is a real and legitimate concern, and it is also completely orthogonal to open weights, which are inert files that run on hardware *you* choose and transmit data to *no one* unless you wire them to. A model you download and run inside your own environment has no more ability to phone Beijing than an Excel spreadsheet does.

The security case leans heavily on this confusion because the confusion is load-bearing. The genuinely worrying property - data exfiltration - belongs to the hosted service and is trivially avoided by self-hosting. The property being restricted - the open weights - does not have that property at all. Invoking the app’s data flows to justify restricting the weights is not a security argument that happens to be mistaken. It is a rhetorical bait-and-switch, and its persistence in the discourse is itself evidence that the goal is to restrict the competitor, not to address the risk.

4. Free Markets Until the Scoreboard Turns

Step back from the mechanics and the larger hypocrisy comes into focus. The United States, and the industry lobbying it, have spent decades constructing a moral vocabulary around open competition, free trade, and the superiority of market outcomes over central planning. That vocabulary was deployed with enthusiasm when American firms were the ones benefiting from open global markets. It is being quietly abandoned now that the competition has produced a result its architects did not want.

A commitment to free markets that evaporates the moment you start losing was never a commitment to free markets. It was a commitment to winning, with free-market language as the ceremonial dress.

The pattern is old enough to be predictable. In the nineteenth century, a young, industrializing United States was one of the world's most enthusiastic pirates of British intellectual property, copying technologies and printing foreign authors without payment, precisely because doing so served a developing economy. Only after America became a net exporter of intellectual property did it discover a deep and abiding conviction that intellectual-property rights were sacrosanct and their violation a moral outrage. The principle tracked the trade balance. It is doing so again.

There is also a hard practical lesson the restrictionists are refusing to learn from their own recent history. The semiconductor export controls of the early 2020s, intended to freeze China's advanced-chip capability, are now widely assessed to have accelerated exactly the domestic Chinese development they were meant to prevent, by converting the threat of dependence into a national mobilization to build indigenous alternatives. Restriction did not create a moat; it created a competitor with a grievance and a subsidy. There is little reason to expect a wall around open models to behave differently - and open weights, unlike fabrication plants, are already published, copyable, and impossible to recall.

The deepest irony is that openness was supposed to be the *American* value in this contest. The liberal case for open science, open source, and open markets is that distributed, permissionless innovation outperforms closed, centrally gatekept systems over time. That case was the West's to make. In the specific arena of AI models, it is now China that is running the open playbook - publishing weights, seeding a global developer ecosystem, and letting the world build on its work - while the United States edges toward the closed, gatekept, state-managed posture it has spent a century criticizing in others. To watch American policymakers argue for restriction and control while Chinese labs argue, in effect, for openness and diffusion is to watch a role reversal that ought to be deeply uncomfortable for anyone who took the original principles seriously.

5. The Security Case, Weighed Honestly

It would be too easy, and not quite honest, to dismiss the security concerns entirely. Intellectual seriousness requires steelmanning the strongest version of the case and then showing precisely where it fails. The restrictionist argument, at its best, rests on four pillars: ideological manipulation, back doors, dependency, and legal exposure. Each is weaker than it first appears.

5.1 Ideological Manipulation Is the Easiest Concern to Defeat

The fear that a Chinese model has been trained to slant its outputs - to soften criticism of Beijing, to shade its treatment of sensitive topics, or in the lurid version, to teach American users to see their own country as the adversary - is the concern most often gestured at and the easiest to neutralize. Open weights are, by definition, modifiable. Overt ideological framing and refusals are surface behaviors that fine-tuning reliably scrubs; a downstream developer can retrain, steer, and re-align an open model to remove them. Whatever worldview a set of open weights arrives with is not a fixed property of the artifact but a starting point the recipient can edit. The very openness that the restrictionists treat as the danger is what makes this class of concern manageable. A closed model, by contrast, you must take or leave as its vendor tuned it, with far less ability to inspect or correct its leanings. Both open-source and closed-source AI models are black boxes. But a user can modify the weights of an open-source model and significantly augment the underlying content. There is no such mechanism for closed models.

5.2 Back Doors: A Real Category, a Manageable Risk

The more sophisticated worry is a hidden back door: a latent trigger, invisible in normal use, that causes malicious behavior only when some rare, attacker-chosen input appears. This is a legitimate category - research has shown that certain planted behaviors can survive standard safety training - and it should not be waved away. But it is also the concern for which robust, well-understood defenses already exist, and those defenses do not require knowing a model's nationality.

The key insight is that a back door's *effect* must eventually cross the boundary of the system running the model. A trigger that does nothing observable is not a threat; a trigger that does something - issues a command, exfiltrates data, ships flawed code - must produce output that leaves the sandbox, where it can be monitored, flagged, and gated. A layered deployment discipline addresses this directly:

- **Containment.** Run the model in a sandbox that denies it network egress and access to sensitive systems. A back door that cannot reach anything cannot accomplish anything, whether or not it is ever detected.
- **Boundary monitoring and human-in-the-loop.** Inspect outputs at the sandbox boundary and require human authorization for any high-consequence action. A latent trigger that tries to issue a catastrophic command is caught precisely because the command is legible where it crosses the line.
- **Detonation in a simulator.** Route a sensitive payload first into a decoy environment that faithfully mimics the real endpoint but lacks its destructive capability. The payload runs, any malicious

behavior is provoked and flagged in a harmless twin, and only vetted output proceeds. This is the same detonation-sandbox pattern that malware analysis has relied on for years.

- **Differential testing across independent models.** Run two independently sourced models and compare their outputs. A trigger present in one surfaces as an anomalous divergence from the other - a check that also catches the subtler failure mode of quietly corrupted judgment, which pure boundary monitoring can miss.
- **Randomized, extended detonation.** Because the classic evasion is a payload that detects the sandbox and plays dead, randomize the observable variables - dwell time, iteration count, synthetic inputs - so the payload cannot reliably predict when it is safe to fire.

These techniques are not speculative; they are adaptations of standard security engineering. The honest limit is that no layer is a silver bullet - a sufficiently sophisticated trigger keyed to a signal that only the genuine live endpoint can produce may still evade a simulator, and detecting subtly corrupted content remains harder than catching overt actions. But two points follow. First, the residual risk is dramatically reduced, not eliminated, and reducing risk to a low, managed level is what security engineering actually does. Second - and decisively - every one of these residual risks is *origin-neutral*. A domestic closed model with a compromised training pipeline or a malicious insider can carry exactly the same back door, and you cannot run these inspections on a closed model at all, because you never hold its weights. The defenses argue for deployment discipline and for preferring auditable open models, not for a ban keyed to a flag.

5.3 Dependency: The Argument Runs Backwards for Open Weights

The most respectable-sounding concern is dependency: that a large and growing share of American AI development now sits atop Chinese models, and that this is a structural vulnerability. Reports in 2026 that a substantial majority of U.S. AI startups build on Chinese open-weight models, and that the leading open-source models by adoption are overwhelmingly Chinese, are cited as evidence of a dangerous reliance.

But the dependency argument quietly assumes the wrong kind of dependency. A dangerous dependency is one in which a supplier can *cut you off, change the terms, or reach into your operation* - and that is a precise description of a closed model accessed through a metered API, not of open weights. Once you have downloaded a set of open weights, the vendor - Chinese or otherwise - cannot repossess them, cannot deprecate them out from under you, cannot throttle you, cannot reprice you, and cannot read your traffic. You hold the artifact in perpetuity. The genuinely high-dependency option in the AI market is the *closed* frontier model reachable only through an endpoint the provider can alter or revoke at will. The open Chinese model is the low-dependency option. The restrictionists have the risk exactly inverted: they are urging American developers away from the artifacts they can fully control and toward the services they cannot.

What survives of the dependency worry is thin: a reliance on the upstream lab's *future* release cadence for the next generation of models, plus some ecosystem and tooling lock-in. But that is a treadmill-you-do-not-control problem that applies to standardizing on *any* single model family, American ones

included, and its remedy is diversification across independently sourced models - again, an origin-neutral, deployment-layer answer rather than a ban.

5.4 Legal Exposure Cuts Both Ways

A final argument holds that licensing disputes over Chinese models would be litigated in Chinese courts under a system of “rule by law” rather than “rule of law.” There is a real distinction here worth weighing - the predictability and neutrality of a legal system is a legitimate consideration. But the argument proves less than it claims. Any cross-border commercial relationship means accepting some foreign jurisdiction; a firm dealing with an American company submits to American courts, and firms operating in Europe submit to European law. Complying with the laws of a counterparty’s jurisdiction is ordinary international business, not a national-security emergency. If a firm judges a jurisdiction’s legal risk unacceptable, the market answer is simple: do not enter that relationship. That is a business risk decision for individual firms to make, not a rationale for a government to foreclose the option for everyone. And for the many Chinese open models released under permissive licenses such as MIT, the ongoing legal entanglement is minimal to begin with - you are using a published artifact under a permissive grant, not signing an enduring service contract.

6. Govern the Behavior, Not the Birthplace

The through-line of Section 5 is that every legitimate risk associated with open models is best addressed at the layer where the model actually *acts* - its deployment - rather than at the layer of where it was *born* - its provenance. This points toward a coherent alternative to the ban: a regime of graduated, risk-proportionate governance that regulates behavior and is blind to the flag.

Such a regime would treat origin the way a competent analyst treats any risk factor - as an input that adjusts a prior, not as a categorical switch. It is entirely rational to hold that a model produced by, or capturable by, an adversarial state with both the capability and the motive to plant a back door warrants a higher prior suspicion than a model from a party with no such motive. Conceding that point does not concede the ban. It argues instead for scaling the *intensity of scrutiny* to the elevated prior:

- **Low-stakes, low-prior deployments** run under standard containment - the ordinary sandboxing any prudent operator applies to any model.
- **Elevated-prior sources** trigger heightened controls: longer and randomized detonation, stricter egress denial, mandatory differential testing against an independently sourced model, and human-in-the-loop gating of consequential actions.
- **Genuinely critical national-security systems** may reasonably exclude foreign - or any single-sourced - models outright. This is cheap insurance where the stakes are absolute, and conceding it makes the framework more consistent, not less.

This regime has every advantage the ban lacks. It is origin-neutral, so it disciplines a compromised domestic closed model exactly as it disciplines a foreign open one - closing a hole the provenance ban

leaves wide open. It is proportionate, spending scrutiny where risk is highest rather than imposing a blunt prohibition. It preserves the enormous economic benefit of cheap, controllable, auditable models for the vast majority of uses where the stakes are modest. And crucially, it is far harder to weaponize for protectionism, because it does not discriminate by nationality and therefore cannot be captured as neatly by an incumbent seeking to lock out a competitor.

The one honest caveat is that a governance regime is itself a lever that can be captured. Whoever certifies the sandboxes or writes the testing standards holds a chokepoint, and chokepoints are exactly what incumbents lobby to occupy. The design imperative, then, is to build the governance layer so that it relocates neither the security risk to the network nor the protectionism to the certification office. That is a real engineering and institutional challenge - but it is a challenge worth solving, and it is categorically better than the alternative of enshrining protectionism directly into a border.

7. Why China's Approach Is, For Now, More Defensible

It is worth stating the provocative conclusion plainly, because it follows from everything above: on the present facts, China's open approach to AI models is more defensible than the closed, restrictionist turn the United States is contemplating. This is a claim about *approach*, not about regimes. One can hold no illusions about the Chinese state and still recognize that, in this specific contest and at this specific moment, the more open strategy is the more principled and the more strategically sound one.

It is more defensible on consistency. The open strategy actually practices the values - permissionless innovation, broad diffusion, building-in-the-open - that the West has long claimed as its own and its advantage. A publicly released model that anyone on earth can download, inspect, audit, and improve is, whatever its origin, a more transparent object than a closed model whose weights are a corporate secret. The restrictionist position asks Americans to prefer the opaque thing they must trust over the transparent thing they can verify - an odd posture for the tradition that claims capitalism as a sacred virtue.

It is more defensible on economics. Diffusion tends to win. A model that becomes the cheap, ubiquitous default - the substrate that a global developer base builds tools, businesses, and skills around - accrues an ecosystem advantage that a premium, gatekept competitor struggles to match, however capable. By seeding that global default, the open strategy is playing the stronger long game, while the restrictionist strategy risks the worst of both worlds: it fails to stop diffusion abroad, since published weights cannot be recalled, and it cedes the global default to the very models it is trying to marginalize, while handicapping its own developers at home.

And it is more defensible on the security merits themselves, once those are examined rather than invoked. As Section 5 showed, the potential harms of open weights are either trivially avoidable (self-host to defeat data exfiltration), removable (retrain to defeat ideological slant), or manageable through origin-neutral deployment discipline (containment and detonation to defeat back doors). A strategy that keeps models open and governs their deployment is not the reckless option. It is the option that

keeps the artifacts inspectable, keeps the ecosystem competitive, and addresses the real risks where they actually live.

The uncomfortable summary for American policymakers is this: in the AI-model contest, the United States is preparing to abandon its own best arguments - openness, competition, verification, diffusion - at exactly the moment a rival has picked them up and started winning with them. That is not a national-security strategy. It is the reflex of compromised businesses that confuse their own market position with the national interest.

8. The Strongest Objections

Fairness requires setting out the best case on the other side and conceding what genuinely lands. Several objections have real force, and a serious reader should weigh them.

Origin does shift the probability of a threat, not just its mechanism. The strongest objection to this paper's framing is that "origin-neutral" can smuggle in a false equivalence. That a back door is *possible* regardless of origin does not mean it is *equally likely* regardless of origin. A model shaped by, or capturable by, a state with both capability and motive to compromise widely adopted software carries a higher prior probability of compromise than one from a party with no such motive. This is correct, and this paper concedes it without reservation. But this is a reason to scrutinize harder, not a reason to ban - it is a term in a cost-benefit calculation that also includes the cheapness of the mitigations and the substantial costs of exclusion. The rational response to a higher prior is more detonation and stricter gating, not prohibition. And "Chinese" remains a noisy proxy for the variable that actually matters - adversarial-state control or coercion - which is why the sharper version of the concern targets state coercion, a threat that deployment discipline addresses directly.

Subtle corruption is harder to catch than overt action. A second serious objection grants that boundary monitoring stops catastrophic actions but notes it is weaker against quietly corrupted judgment - a coding model that plants a subtle flaw on a rare trigger, or an analytical model that shades an assessment. This residual is real. But it is also speculative, thin on demonstrated instances, best addressed by the differential-testing layer described above, and, once more, origin-neutral, since a compromised domestic model presents the identical risk. It argues for multi-source verification, not for a provenance ban.

Not all open models are Chinese, and dependency concentration is real. True - and this cuts in favor of the paper's recommendation, not against it. A healthy, diversified ecosystem drawing on open models from many origins, American ones included, is precisely the hedge against concentration that a differential, multi-source deployment regime is built to encourage. The answer to over-reliance on any single source is diversification, which a ban on one major source undermines rather than advances.

Some systems really are too sensitive for any residual risk. Also true, and already conceded. For genuinely critical national-security infrastructure, excluding foreign or single-sourced models is

defensible cheap insurance. The disagreement was never about nuclear command systems. It is about whether a small business building a customer-service tool or a coding assistant should be forbidden a cheap, controllable, auditable model because of the flag on its training run. On that question, the sensitive-systems concession offers no support to a broad ban.

Taken together, the objections establish that origin is a legitimate risk input and that the residual risks are non-zero. They do not establish that a categorical, provenance-based restriction is the right instrument. Every place the objections have force, they point toward harder scrutiny and better deployment discipline - the graduated regime of Section 6 - rather than toward the ban. That is the honest limit of the security case, and it is a long way short of where the self-interested business interests would have you believe..

9. Conclusion

Strip away the vocabulary of threat and the shape of the thing is plain. A cheaper, open, controllable competitor has arrived in a market dominated by expensive, closed, incumbent products, and the incumbents - multinational firms whose interests are not identical to the nation's - are reaching for the one lever that price cannot touch: the power of the state to declare their rival dangerous. The national-security frame is not wholly empty; origin is a real risk input, and some systems warrant exclusion. But that threat is doing far less work than the money is. The tell is in the timing, in the coalition's internal contradictions, in the studied conflation of hosted services with downloadable files, and in the sudden expiration of a free-market faith that held firm only while self-interested multinational corporations were ahead.

The better path is not hard to describe. Govern models by what they are permitted to *do*, not by where they were *made*. Scale scrutiny to genuine risk. Reserve exclusion for the handful of systems where the stakes are absolute. Preserve, for everyone else, the enormous benefit of models that are cheap, controllable, and open to inspection. And recover the courage of the actual liberal conviction - that open competition, not protected incumbency, is how free societies win.

The deepest irony is the one worth ending on. In this contest, openness, verification, competition, and diffusion - the values the West has claimed as its birthright - are, for the moment, being practiced more faithfully by China's open-model strategy than by America's restrictionist one. A country confident in those values would answer a cheaper open competitor by competing, by out-building, by winning in the open. Building instead a wall to keep the competition out is not the move of a nation that believes its own principles.

Sources & Further Reading

The references below are ordered as a hierarchy of evidence — government and official analysis first, then institutional and academic research, then technical and market evidence, then policy commentary and journalism. Quantitative claims are anchored to primary benchmark and government sources rather than secondary summaries wherever possible. Adoption, pricing, and specific legislative facts continue to move quickly, and readers are encouraged to consult the originals.

1. Government and official analysis

1. National Telecommunications and Information Administration (NTIA), “Dual-Use Foundation Models with Widely Available Model Weights” (2024) — the U.S. government report finding insufficient evidence to restrict open weights and recommending monitoring: <https://www.ntia.gov/programs-and-initiatives/artificial-intelligence/open-model-weights-report>
2. U.S.–China Economic and Security Review Commission, “Two Loops: How China’s Open AI Strategy Reinforces Its Industrial Dominance” (2026): <https://www.uscc.gov/research/two-loops-how-chinas-open-ai-strategy-reinforces-its-industrial-dominance>

2. Institutional and think-tank research

3. RAND Corporation, “Securing AI Model Weights: Preventing Theft and Misuse of Frontier Models” (Nevo et al., 2024): https://www.rand.org/pubs/research_reports/RRA2849-1.html
4. RAND Corporation, “Open Models, Soft Power, and the Spectrum of U.S.–China Artificial Intelligence Competition” (PEA4686-1, 2025): <https://www.rand.org/pubs/perspectives/PEA4686-1.html>
5. Center for Security and Emerging Technology (CSET), “Open Foundation Models: Implications of Contemporary Artificial Intelligence”: <https://cset.georgetown.edu/article/open-foundation-models-implications-of-contemporary-artificial-intelligence/>

3. Academic research

6. Kapoor, Bommasani, et al., “On the Societal Impact of Open Foundation Models” (Stanford HAI / Princeton, 2024) — the marginal-risk framework (arXiv 2403.07918): <https://arxiv.org/abs/2403.07918>
7. Anderljung et al., “Open-Sourcing Highly Capable Foundation Models” (Centre for the Governance of AI, 2023) — the risk-side evaluation (arXiv 2311.09227): <https://arxiv.org/abs/2311.09227>

4. Technical safety research from leading labs

8. Hubinger et al., Anthropic (2024), “Sleepers Agents: Training Deceptive LLMs that Persist Through Safety Training” (arXiv 2401.05566): <https://arxiv.org/abs/2401.05566>
9. Anthropic (2024), “Simple Probes Can Catch Sleeper Agents” — evidence that planted behaviors are often detectable: <https://www.anthropic.com/research/probes-catch-sleeper-agents>

5. Benchmarks, market data, and ecosystem trends

10. Artificial Analysis, “DeepSeek Is Back Among the Leading Open-Weights Models with V4 Pro and V4 Flash” — independent benchmarks and pricing: <https://artificialanalysis.ai/articles/deepseek-is-back-among-the-leading-open-weights-models-with-v4-pro-and-v4-flash>

11. Artificial Analysis, open-source model comparison (intelligence, price, performance): <https://artificialanalysis.ai/models/open-source>
12. OpenRouter, “The Open-Weight Models That Matter” (June 2026) — direct token-consumption/adoption data: <https://openrouter.ai/blog/insights/the-open-weight-models-that-matter-june-2026/>
13. Stanford Institute for Human-Centered AI, “The 2025 AI Index Report” — canonical data on model trends, openness, investment, and the narrowing U.S.–China performance gap: <https://hai.stanford.edu/ai-index/2025-ai-index-report>

6. Policy analysis and commentary

14. Carnegie Endowment for International Peace, “Beyond Open vs. Closed: Emerging Consensus and Key Questions for Foundation AI Model Governance” (2024): <https://carnegieendowment.org/research/2024/07/beyond-open-vs-closed-emerging-consensus-and-key-questions-for-foundation-ai-model-governance>
15. Brookings Institution, “Market Concentration Implications of Foundation Models”: <https://www.brookings.edu/articles/market-concentration-implications-of-foundation-models-the-invisible-hand-of-chatgpt/>
16. CSIS, “China’s Localization Drive in Semiconductors Gains Impetus from Allied Chip Export Controls”: <https://www.csis.org/analysis/chinas-localization-drive-semiconductors-gains-impetus-allied-chip-export-controls>
17. CSIS, “The GAIN AI Act Will Undermine the Global Competitiveness of U.S. AI Chipmakers”: <https://www.csis.org/blogs/perspectives-innovation/gain-ai-act-will-undermine-global-competitiveness-us-ai-chipmakers>
18. Center for European Policy Analysis (CEPA), “The AI Kill Switch: Dangerous Chinese Open Source” — a leading statement of the restrictionist case: <https://cepa.org/article/the-ai-kill-switch-dangerous-chinese-open-source/>
19. TechPolicy.Press, “How the US and China Can Step Back from the Battle for AI Model Supremacy”: <https://www.techpolicy.press/how-the-us-and-china-can-step-back-from-the-battle-for-ai-model-supremacy/>

7. Journalism

-
20. Semafor, “Washington Confronts China’s Open-Source Models” (July 2026): <https://www.semafor.com/article/07/15/2026/washington-confronts-chinas-open-source-models>
 21. Tom’s Hardware, “Trump Administration Reportedly Reviving Push to Ban Chinese AI Models” (July 2026): <https://www.tomshardware.com/tech-industry/artificial-intelligence/trump-administration-reportedly-reviving-push-to-ban-chinese-ai-models-following-kimi-k3-launch-citing-cybersecurity-concerns-downloadable-open-weights-could-make-an-outright-u-s-ban-nearly-impossible-to-enforce-amid-growing-adoption>

Note on method and stance: This paper argues a position and is written as advocacy. It concedes, in Section 8, the strongest objections to its own thesis. Readers weighing AI policy should treat it as one side of a live debate and consult the restrictionist sources above - particularly the CEPA analysis - for the opposing case in its own words.