The Threat Silicon Valley Did Not See Coming

Structural decline of frontier-AI developers in a regulatory environment degraded by competitive authoritarianism and science denialism

AUTHOR
Miguel Moreno

AFFILIATION
University of Granada

PUBLISHED
27 June 2026

ABSTRACT +

1 The underestimated threat: from unmodelled variable to structural symptom

Accounts of the risk facing developers of frontier large language models are almost always assembled from legible threats: the scarcity of compute, the contest for talent, the pressure of open-weight models, the ceiling of the scaling laws, the competition of the Chinese laboratories. These are risks the industry knows how to name because it knows how to price them; they admit hedging, diversification, a schedule. Against that inventory, the governance environment of the developer's own jurisdiction figured as a tacit assumption rather than a variable: a predictable rule of law, reasoned administrative procedure, regulatory decisions anchored in the best available technical evidence. The thesis of this essay is that this assumption was in fact the most consequential variable of all, and that its deterioration is not a passing accident but a process with direction.

The standing of the two forces that drive it must be specified at the outset, because everything else depends on it. Competitive authoritarianism and science denialism are not two further hazards that certain firms neglected to enter into their strategic plans; they are *aggregating symptoms* of a deeper erosion of the background against which social cooperation in science, health, education and the environment operates at its best. The standing of the two forces that drive it must be specified at the outset, because everything else depends on it. Competitive authoritarianism and science denialism are not two further hazards that certain firms neglected to enter into their strategic plans; they are *aggregating symptoms* of a deeper erosion of the background against which social cooperation in science, health, education and the environment operates at its best.[1] That background — equitable opportunities of access to knowledge and culture for the majority, independent research institutions, the open circulation of talent on merit — is

precisely what allows a territory to consolidate the industrial ecosystems capable of attracting and retaining highly exacting professionals, who are sensitive to parameters of wellbeing and quality of life that no salary or corporate prestige compensates for on its own. The arbitrariness of a political management that is hyper-ideologised and refractory to expert knowledge does not, therefore, degrade an isolated parameter: it degrades the very conditions of possibility of talent-intensive activity.

It does so, moreover, with international reach, dragging towards the precipice firms and institutions in jurisdictions far removed from the locus of the decision. Two recent episodes, external to the domain of AI, fix the structural signature of the phenomenon before it appears applied to frontier models. The first, the "Liberation Day" tariff policy of April 2025, invoked economic-emergency powers to set tariffs at will on nearly every trading partner; the measure triggered a market crash and was ultimately struck down by the Supreme Court in February 2026 as an unlawful use of emergency authority, with an estimated 166 billion dollars collected under it left exposed to refund claims [1,2]. The second, the joint US–Israeli campaign against Iran — which escalated from the twelve-day war of June 2025 [3–5] into the full operation launched on 28 February 2026 — displayed the other half of the pattern. That operation was begun without congressional authorisation and on only a cursory public justification [6]; its declared aim of halting an Iranian nuclear-weapons programme stood against the International Atomic Energy Agency's finding of no evidence that such a programme existed, and against the United States' own earlier intelligence assessment that Iran was not building a weapon [7]. Its effects did not stay within Iran: retaliation reached the Arab Gulf states the campaign's own authors had courted, the Strait of Hormuz was closed to global shipping, energy prices rose, and a wider regional war followed whose costs no party could bound in advance [7]. What is most disquieting in the episode is less any single proclamation than the fragility of the decision scheme itself — its arbitrariness, its urgency, its appeal to a national-security risk whose factual basis the relevant experts could not verify — together with the high probability that the actor wielding such discretion, and the allies it draws in, are themselves harmed to degrees difficult to anticipate in advance. In both cases the same structure recurs: an authority of exception — emergency, national security — exercised at discretion, on a factual basis that is opaque or in dispute, with an effect that overruns the borders of the actor exercising it. Decisions of this kind, taken without the minimal expert-grounded weighing of consequences, do not merely raise the probability of an unmanageable chaos that degrades the very conditions of growth and stability they presuppose; in industries acutely sensitive to their operating environment, and amid ferocious geopolitical competition, they suffice on their own to set off waves of effects impossible to forecast. The hypothesis of this essay is that this structure, transposed to

the cognitive infrastructure of the frontier, defines the dominant risk for the laboratories that produce it.[2]

The combination of the two forces is not additive but complementary, and it is worth anticipating now. Competitive authoritarianism — in the precise sense in which Levitsky, Way and Ziblatt characterise the United States' turn in 2025: a regime that preserves the forms of democracy while hollowing out their substance through the selective use of formally legitimate instruments[9,10] — supplies the instrument: the discretionary application of rules conceived for another purpose. Science denialism withdraws the brake that would limit its use, by decoupling the criterion of the decision from the technical evidence that ought to constrain it. Together they produce a governance that is at once arbitrary, because it ceases to be predictable, and immune to fact, because it ceases to be correctable by proof.[3] To that object — the degradation not of the regulator but of the very criterion by which one regulates — I give the name *governance-criteria capture*, and I argue that it constitutes a failure mode distinct from both the over-regulation and the under-regulation with which the literature usually operates.

The firms' miscalculation lies exactly here. The photograph of the January 2025 inauguration, with the sector's leading executives — Musk, Bezos, Zuckerberg, Pichai, Cook — seated ahead of the administration's own cabinet nominees[11,12], is not an anecdote of protocol: it is the emblem of a strategic wager on the unconditional favour of a discretionary state[13–15]. That the sector's leadership should have placed itself there at all — at the investiture of a president whose authoritarian excesses and arbitrariness were already well known, and had already been suffered, under an earlier term — betrays less a calculated wager than a deficit of corporate vision and judgement, a servility legible in the very gestures and greetings of the occasion. The wager systematically overestimates the durability of that advantage. A quasi-monopoly position built on affinity with an executive — and not on the verifiable quality of the product — inherits the volatility of its pattern: the same power that grants preferential access can revoke it without notice, as the suspension of the most capable models of one of those very laboratories would demonstrate eighteen months later under an export-control directive. And it carries, in addition, an accumulated reputational cost.

The business models of these firms prosper, disproportionately, in literate, high-income societies, with critical consumers and a free press; these are precisely the societies in which explicit affinity with elites of an elitist, anti-immigration or openly autocratic disposition is entered as a liability. In the medium term, that liability can contaminate every niche — defence, aerospace, logistics, cloud, cybersecurity — in which the firm has an identifiable presence[16]. What such an alignment endorses, if only indirectly, is a style

of authoritarian and avowedly irrational management prepared to impose anti-social schemes by force on the most sensitive matters — border control, the treatment of vulnerable groups, healthcare coverage, food assistance, education — and its reputational reach is not easily contained. It erodes, above all, the confidence on which any collaboration with United States firms as reliable partners depends, in sectors from defence and AI to renewable energy, health and aeronautics: the favour of radically ideologised, denialist executives confirms only conditional support, contingent on submission and deference, with no guarantee of long-term stability and no commitment beyond the convenience of the immediate strategic moment[17].

> **The seating order, read closely: 20 January 2025**

Three questions follow, which the essay will develop and which it is worth leaving posed. The first is a question of declared coherence: when a company adopts competitive authoritarianism as the environment most favourable to its corporate success, what social objectives and values does it claim to serve, and how does it justify them in public? The second is a question of structural self-knowledge: if the science denialism of the actors with the greatest political leadership is judged propitious for firms that depend on a constantly developing scientific-technical base, what have their managers ceased to see about the conditions of their own ascent? It is best framed without rhetoric. The position those leaders now occupy was made possible by a system of higher education and research that consolidated its prestige over decades of autonomy from executives of every ideological stripe; an environment with academic institutions weakened, economically extorted or ideologically subjugated would hardly have been a seedbed for great technology firms — which is exactly the environment their wager helps to produce, and whose first measurable indicator is already visible: the number of AI researchers and developers relocating to the United States has fallen by 89 per cent since 2017, with 80 per cent of that decline concentrated in the last year alone[22]. Nor does the erosion stop at the universities: the same discretionary logic that purges and packs the administrative state turns its pressure on the independent regulators and scientific agencies — among them the securities regulator (SEC), the ocean-and-atmosphere service (NOAA) and the environmental-protection agency (EPA) — whose statutory autonomy exists precisely so that decisions demanding prospective analysis, which is to say almost all of them, rest on expert knowledge. That pressure installs discretion as the ordinary criterion of action and anticipates a horizon of dysfunction and unmanageable damage, born of an elementary failure of foresight[9]. The third is a question of public trust: can citizens, domestic or foreign, place durable confidence in firms whose class of governance has been structurally disabled and biased from the political environment, and which appear in the

public sphere as instances subordinate to a network of influence and conflicts of interest between regulator and regulated?

The pivot of the argument is the following. Firms developing frontier models underestimate the foreseeable effect of substituting governance with discretion, however much that substitution may temporarily swell their results. What the logic of capitalist competition tends to render invisible is that the decisive advantage of an R&D democracy does not reside in any affirmative executive act, but in a *distributed property*: a plurality of domestic and foreign actors, cross-verification, the open circulation of talent on merit, and the possibility of resorting to other jurisdictions in the face of unforeseen problems. It is that enabling framework — and not the favour of a cabinet — which generates and sustains the advantage, and it is that property which discretion erodes, because it displaces long-horizon investment and talent towards environments of predictable rule. The continuity of access to the tool with which one has learned to think then ceases to be guaranteed, no longer for the individual user — the question of second-order epistemic autonomy — but for the very R&D base on which the laboratories depend. The price of a technocratic elite so out of phase with social, economic and political reality is paid first by its own employees and, in proportion to the size of the firm, by society at large — as the September 2008 collapse of Lehman Brothers, the fourth-largest investment bank in the United States and, at some 639 billion dollars in assets, the largest bankruptcy filing in the country's history, made plain: its failure helped to freeze global credit markets and tip the world into a recession that impoverished millions[23]. That a technocratic elite should have failed to value this asymmetry admits an uncomfortable but analytically sober reading: the mechanism that selected it rewarded proximity to power over managerial judgement, and a leadership so selected optimises the immediate advantage precisely because it is not equipped to appraise the distributed condition that made it possible — an army of ambitious sycophants rather than a corps of qualified managers[24].

The sections that follow develop this argument in order: the anatomy of the two corrosions of the criterion (§2), the dated case that exemplifies them in the domain of AI (§3), the evidentiary asymmetry that lays bare the epistemic deficit of the framework (§4), the channels through which discretion propagates into decline (§5), the displacement of the frontier beyond the reach of the switch (§6), the prospective exercise that quantifies the resulting trajectories (§7), the governance that would restore a floor (§8), and the principal objections and limits (§9).

2 Two corrosions of the governance criterion

2.1 Competitive authoritarianism as an instrument

The concept of competitive authoritarianism is not a synonym for dictatorship, and the analytic force of the essay depends on not conflating the two. In Levitsky and Way's formulation, competitive authoritarian regimes retain the formal architecture of democracy — elections, courts, a press — while emptying it of substance through the selective and discretionary use of instruments that remain, taken individually, legal[25,26]. The diagnostic feature is not illegality but the selective application of rules that exist for another purpose; the discretion is the product, not the defect. To weaponise the state, elected autocrats purge and then pack it, following a blueprint legible in Orbán's Hungary, Erdoğan's Turkey, Chávez's Venezuela and Modi's India long before it was legible in Washington[9]. Levitsky, Way and Ziblatt argue that the United States crossed into this category in 2025; they insist, with equal emphasis, on its reversibility, pointing to electoral contestation and intra-coalitional fracture as the channels through which such regimes are characteristically undone[10]. Both halves of that claim matter here: the diagnosis grounds the risk variable, and the reversibility grounds one of the four futures developed in §7.

The same conjunction has been theorised independently from within the philosophy of technology: Coeckelbergh's account of *technofascism* identifies the structural complicity between platform power and the new authoritarianism, a diagnosis whose concrete instantiation this essay traces in §3[27].

Two implications bear directly on the governance of artificial intelligence. The first is that AI governance rests on instruments — export control, national-security authority, product and emergency powers — whose *selective application* is precisely the modus operandi the concept names. Frontier AI does not need to be prohibited for the regime to capture it; it suffices that its access be rendered discretionary. The second is that the phenomenon is not parochially American. The same analytic lens that diagnoses Washington also identified, until April 2026, a competitive-authoritarian member state inside the European Union: Hungary under the sixteen-year Fidesz government of Viktor Orbán. The parliamentary elections of 12 April 2026, which returned a two-thirds supermajority to the opposition Tisza party and ended the Orbán era, illustrate precisely the reversal mechanism Levitsky, Way and Ziblatt identify as the characteristic mode through which such regimes are undone[10]. That reversal does not dissolve the point; it extends it. Either type of governance — codified and contestable, or discretionary and unappealable — is available to any jurisdiction, and Hungary's trajectory, approach and withdrawal, confirms that the contrast drawn in §2.3 is empirical rather than ideological:

the EU's codified, contestable framework was the very condition that made the reversal possible.

2.2 Science denialism as the removal of the brake

If competitive authoritarianism supplies the instrument, science denialism removes the epistemic constraint that would otherwise limit it. The decisive analytic tool here is not the proposition that publics "do not believe science", but Oreskes and Conway's distinction between *impact science* and *production science*. Conservative hostility, they argue, is not directed at science as such but at the sciences that identify harms implying a need for regulation; the sciences that underwrite industry are left untouched. The corollary is the sharp one: any science can become an impact science the moment it discovers something that points to the need for government action[28]. The mechanism is older than the climate case in which it was first anatomised — the organised manufacture of doubt documented in *Merchants of Doubt*[29] is, in Proctor and Schiebinger's terms, a deliberate production of ignorance, an agnotology rather than a mere absence of knowledge[30].

The application to AI is exact and, once stated, hard to unsee. *AI-safety science is the impact science par excellence*: its central epistemic product is the proposition "this is dangerous, regulate it." A denialist posture need not deny that AI works — that is production science, and it is welcomed and funded; it need only discount, selectively, the evidence of harm. The result is a governance whose criteria become *anti-correlated* with the technical facts: the safety evidence that a responsible developer produces as its distinctive contribution loses traction in the decision, while the discretionary lever is applied on non-technical grounds. This is not a hypothetical inference. The same political environment that discounts impact science has, in its first eighteen months, cut federal research funding, cast public doubt on scientific findings, and prioritised investment outside the university system — the defunding of the very science–society contract on which the R&D democracy depends[31,32]. The contemporary scholarship that catalogues denialism as a coordinated institutional practice, rather than a private failing, reinforces the point[33]. (That the concept of "science denial" is itself contested — some hold it to be ambiguous and inflated — is addressed in §9, where the precise Oreskes–Conway version is adopted precisely because it is falsifiable and does not require positing a wholesale rejection of science.)

2.3 Governance-criteria capture, and the comparative test

The conjunction of the two forces yields an object that the standard vocabulary of regulatory analysis does not name. The classical concern, in the Stiglerian tradition, is *regulatory capture*: the regulator is taken by the regulated, and the rules come to serve the industry. What is at issue here is different and, I will argue, more corrosive. It is not the capture of the regulator but the degradation of the criterion itself — *governance-criteria capture* — such that the decision ceases to be (a) predictable and (b) correctable by evidence. Competitive authoritarianism destroys (a); science denialism destroys (b). The output is *a governance simultaneously arbitrary and fact-immune*. This is a third failure mode, orthogonal to the familiar axis along which policy debate usually runs — too much regulation versus too little. A regime can be minimal in volume and still catastrophic in kind, if the little it does is discretionary and unmoored from proof. And it produces a terminal paradox for the safety-conscious firm, developed in §3: the transparent, fair, fact-grounded statutory process that the responsible developer demands is exactly the kind of governance that an anti-empirical, competitive-authoritarian environment is structurally unable to supply[34].

The comparative test sharpens, rather than softens, the diagnosis. The European Union enacted, in Regulation (EU) 2024/1689, the world's first comprehensive horizontal framework for AI: codified, risk-tiered, with obligations rolled out on a published schedule[35]. It is the structural opposite of discretionary override, and it would be convenient for the argument to treat it as the unblemished foil. Honesty forbids this. The EU is, at the time of writing, under its own deregulatory pressure: the Digital Omnibus on AI, proposed in November 2025 and adopted by the European Parliament in June 2026, postpones the most consequential high-risk obligations to December 2027 and August 2028, and civil-society analysts have called the process a weakening conducted without proper impact assessment — "a defeat for the rule of law"[36–38]. The comparison, properly drawn, is therefore not between a regulated Europe and a deregulated America. It is between two *modes* of changing the rules. The European weakening proceeds through the ordinary legislative process — deliberative, recorded, amendable, justiciable, and reversible at the next election — whereas the discretionary override proceeds through a letter dispatched at 5:21 p.m. invoking an authority of exception, on evidence withheld, with no notice and no avenue of appeal. The first is governance one may contest; the second is the *absence of governance wearing its clothes*. That distinction is the whole of the matter, and the case to which the essay now turns is its cleanest instance.

3 The anatomy of a switch

On 12 June 2026, at 5:21 p.m. Eastern Time, the United States Secretary of Commerce sent the chief executive of a frontier developer a letter invoking national-security export-

control authority and requiring the suspension of all access to the firm's two most capable models — Fable 5 and Mythos 5 — by any foreign national, whether inside or outside the United States, including the firm's own non-citizen employees[39,40]. The directive's scope was nominally one of nationality; its practical effect was a total global shutdown, because the models ran across dozens of simultaneous cloud environments and the firm had no mechanism to verify user nationality in real time across that infrastructure. A measure formally addressed to foreigners therefore produced a universal recall. It was, by contemporaneous expert account, the first time the United States invoked export-control authority against the API access of a commercially deployed AI model — the conversion of a question of product safety into a question of sovereignty over who may touch a capability[40,41].

Two events converged to produce the directive, and each illuminates a different face of the pattern. The first was a matter of provenance: a telecommunications firm that had obtained early access to the restricted model through the developer's trusted-access programme was found to have historical corporate ties to China, and the administration ordered its access revoked[40,42]. The second was a matter of capability: security researchers identified a method of eliciting the model's cybersecurity functions through a prompt so banal it has become shorthand for the episode — an instruction, in effect, to read a codebase and fix its flaws. The model had been built to decline exactly such requests; the guardrail was the developer's public proof of responsibility. That the two events arrived within days of each other, and that the response was a recall rather than a remediation, is the first datum. The second is that the developer's own account — that the demonstrated technique recovered a small number of previously known, minor vulnerabilities, of a kind that other publicly available models, including a competitor's, discover without any bypass — could not be independently weighed, because the government's technical reasoning was not made public[39,43]. The reader should hold both facts at once: the characterisation of the episode comes largely from an interested party, and the inability to check it is itself the phenomenon under analysis.

It is here that the essay's sharpest irony becomes visible, and it is not the irony of an external accident but of a self-authored trap. The developer had, for two years, marketed its most capable systems under a discourse of danger: *this model is so capable in cybersecurity that we must restrict it; we have built it as, in effect, a controlled munition.* That branding was the firm's bid for legitimacy and its differentiator in a crowded market. It was also, precisely, the legal predicate of its own vulnerability. A product described in every release as a munition invites, sooner or later, an instrument designed for munitions; a cybersecurity researcher's verdict on the episode put it with unimprovable economy — the firm wrote the legal predicate itself and called it a brand[44]. The

discourse of responsibility had become the lever of capture. This is the *brand-as-munition trap*, and it is general: under a governance whose criterion has been decoupled from proportionate, fact-grounded assessment, the very act of advertising one's product as dangerous supplies the discretionary state with a ready-made hook, on which the timing of the pull is left entirely to its discretion.

The deeper consequence is the one the broader project has named *second-order epistemic autonomy*. The conventional question in the philosophy of AI asks whether a subject can reason well *while using* an AI system — whether the tool degrades or enhances first-order epistemic agency. The case shows that question to rest on an unexamined premise: that the relationship between user and system is stable enough to make the first-order question well posed. A measurable and growing share of research, teaching and civic reasoning across dozens of jurisdictions is now routed through a small number of providers headquartered in a single country, and none of those users holds an enforceable claim to continuity of access, a contractual right to notice, or an avenue of appeal. What the switch threatens, then, is not merely a product but a condition: the guaranteed availability of the instrument with which one has learned to think. And the case generalises the threat one tier further than the individual user. The same discretionary power that can disconnect a foreign student can disconnect a foreign-national employee of the developer itself — which is to say it can reach into the R&D base, the distributed property, on which the frontier depends [45]. A jurisdiction that can switch off its own foreign researchers overnight becomes a risky place in which to build, and capital and talent migrate towards where access is predictable. That this occurred in the same month in which the developer was reported to have filed confidentially for a public listing only underscores the point: the asset a frontier firm brings to market is its forward capability, and a forward capability hostage to a 5:21 p.m. letter is an asset of newly uncertain value [44].

> **Scenario B, observed: the Seoul launch of 17 June 2026**

4 The standard that every object satisfies

The directive's most revealing feature is not its severity but its logic, and the logic does not survive inspection. The developer's defence turned on a distinction between two kinds of jailbreak. A *universal* jailbreak would be a method capable of broadly defeating a model's safeguards, unlocking a wide range of dangerous capabilities; the firm maintained that no tester had yet found one. A *non-universal* jailbreak can elicit some capability in specific circumstances, and the firm's stated position — shared, it argued,

across the industry — is that perfect resistance to such jailbreaks is not currently possible for any provider, and that universal jailbreaks will eventually be found[39,47]. On those two premises the developer built a defence-in-depth posture: make jailbreaks either narrow or very expensive, and monitor to detect and close them. But the same two premises, once stated, dismantle the directive's standard.

Consider the operating rule the directive instantiates: *the demonstration of a non-universal jailbreak justifies suspension.* If that is the rule, and if it is simultaneously held that perfect resistance is impossible and that every deployed model is vulnerable to *some* non-universal jailbreak, then the suspension condition is satisfiable against any model at any moment. A criterion that every object of its class necessarily satisfies does not discriminate between safe and unsafe systems; it discriminates not at all. It is not, in the logician's sense, a standard — it is a lever. Whoever holds it may pull it selectively — against one provider, against one market moment, against one actor — without ever having to justify why this case and not another, because on the stated criterion every case qualifies. The developer drew the implication itself: applied across the industry, the standard would halt essentially all new frontier deployment[39]. The point for this essay is not the disruption to one firm but the demonstration that the governing criterion had been emptied of discriminating content — the very signature of governance-criteria capture, now visible at the level of a single decision.

The evidentiary asymmetry compounds the deficit into something one may fairly call radical. On one side stood a directive with the force of a total recall, affecting models deployed to hundreds of millions of users. On the other stood verbal evidence of a narrow technique whose resulting capability the developer judged to be equivalent to defensive tools in everyday use, and replicable in a competitor's publicly available model not subject to the same control[39]. The outside expert who appears to have read the underlying paper reached the same conclusion from the opposite direction: the behaviour described cannot meaningfully be removed without making the model worse at the defensive work of finding and fixing bugs, and the same capability is, or shortly will be, available in foreign and open-weight systems that export controls cannot reach[43]. A governance environment that can impose its gravest instrument on the thinnest and least auditable evidence is not one that an enterprise built on talent and expert knowledge can plan within; the deficit is not of degree but of kind, because the relation between the weight of the action and the weight of the proof has been severed.

There is a final observation that belongs to the register of the essay rather than to the case, and it is best stated plainly because it is analytically, not rhetorically, motivated. The firms most exposed to this deficit included those that had most conspicuously aligned

themselves with the administration whose discretion produced it. A leadership able to read a balance sheet but unable to read an institutional risk had wagered on the durability of executive favour and received, in return, the demonstration that favour granted at discretion is favour revocable at discretion. That the wager was made at all is the symptom; that its makers appear not to have anticipated the most predictable consequence of empowering an unconstrained and ill-advised regulator is the diagnosis. The asymmetry of proof, in other words, was not only a property of the directive. It was also a property of the strategic reasoning that had helped to install the conditions for it.

5 The channels of decline

The cost of wilful managerial ignorance has been chronicled most crisply in Lewis's account of the fifth risk [48]. If governance-criteria capture is the disease, decline is the prognosis, and the prognosis must be earned mechanism by mechanism rather than asserted. The thesis is not that discretionary favour never pays; it is that it pays on a horizon shorter than the one on which a frontier firm's value is realised, and that the same discretion which delivers the near-term gain erodes the conditions of the long-term advantage. Four channels carry the erosion. They are presented here in prose and formalised, as an interactive model, in §7.

The first channel is the brand-as-munition trap already anatomised: under a decoupled criterion, the firm's own danger-branding raises the probability of a discretionary intervention, so that the marketing that bought legitimacy also bought exposure. The second is *flight*. A large body of evidence establishes that policy uncertainty depresses investment and consumption independently of the policy's content; the mere unpredictability of the rule is sufficient to displace long-horizon commitment [49]. For a frontier laboratory the displaced commodity is not only capital but talent, and talent is the more mobile and the more consequential. The leading indicator is already in the record: the number of AI researchers and developers relocating to the United States has fallen by 89 per cent since 2017, four-fifths of that decline in the most recent year [22]; and prominent scholars of authoritarianism have themselves moved their work to Canadian universities as the political pressure on American institutions has intensified [10]. The third channel is *substitution*, treated in full in §6: as domestic frontier access becomes discretionary, adoption migrates to open-weight and foreign systems, and the strategic value of the off-switch decays as the frontier moves beyond its reach. The fourth is *legitimacy*: a discretionary intervention unsupported by public technical reasoning erodes the trust on which a firm can credibly demand fact-grounded governance, and the erosion is self-reinforcing, because a delegitimised demand for evidence-based process leaves the criterion still more decoupled from fact than before.

The four channels share a single underlying object, and naming it precisely is the analytic core of the section. The advantage of an advanced R&D economy was never an affirmative grant of the state; it was a *distributed property* of the ecosystem — a plurality of domestic and foreign actors, cross-verification across institutions, the open circulation of talent on merit, and the standing possibility of recourse to other jurisdictions when a problem or an opportunity exceeded the local frame[50]. This is the property that discretion dismantles, not by confiscating any single asset but by withdrawing the predictability that made the distributed arrangement worth investing in. And the dismantling has a structural homology with the political phenomenon that produces it: just as competitive authoritarianism preserves the *form* of democracy while hollowing its substance, a jurisdiction that retains its laboratories but loses the predictability of access to them retains the *form* of its scientific advantage while forfeiting the substance. The instrument conceived to preserve a national strategic advantage degrades the very property that generated it. This is not a paradox of execution but of kind: discretion and distributed advantage are, on a long enough horizon, incompatible.

The historical record supplies the mirror, and it is a mirror the United States of all countries should find legible, because it is the country the original asymmetry built. American scientific supremacy in the twentieth century was assembled, in substantial part, from talent that fled regimes which had ideologically purged their own institutions — the physicists and mathematicians who left fascist Europe in the 1930s and made the Anglo-American laboratory the centre of the world. The cautionary case from the other side of the ledger is Lysenkoism: a scientific base subordinated to ideological criteria, its impact sciences suppressed because their findings were politically unwelcome, with a degradation of capability that took the Soviet life sciences a generation to repair. The essay's claim is that the contemporary configuration runs the first process in reverse and rhymes with the second: a jurisdiction degrading the distributed property that attracted the world's talent, while subordinating its impact sciences — safety science foremost among them — to a discretionary, anti-empirical criterion. The 89-per-cent figure is not yet an exodus. It is the first reading on an instrument that has measured exoduses before.

6 The frontier moves beyond the switch

A switch is only a source of power for as long as what it controls is scarce and irreplaceable, and the central wager of an export-control regime over AI is precisely that frontier capability is both. The wager is poor. The capability elicited in the case at the origin of this essay — reading a codebase to find and fix its flaws — is, on expert assessment, defensive work that cannot be excised without degrading the model, and it is already, or will shortly be, available in foreign and open-weight systems beyond the

reach of any American control[43]. The history of export controls over software is a history of this miscalculation. The attempt to control strong encryption as a munition in the 1990s failed against the diffusion of mathematics; the addition of "intrusion software" to the Wassenaar Arrangement in 2013 was drawn so broadly that it threatened the defenders — vulnerability disclosure, incident response, coordinated defence — more than the attackers, and had to be renegotiated[43]. Capability encoded in weights and methods behaves like capability encoded in algorithms: it travels, and the control golden-plates the domestic incumbent's costs while the frontier reconstitutes itself elsewhere.

"Elsewhere" is no longer hypothetical. The two leading AI ecosystems have organised themselves around divergent strategies, and the divergence is precisely along the axis this essay has been tracing. The American frontier has concentrated investment in compute-intensive, closed models; China, constrained on advanced semiconductors but backed by sustained state support, has organised around open development and rapid deployment, integrating AI across its economy under a rubric of general capability[51]. Open-weight systems from Chinese developers have moved from a marginal share of global usage to a substantial one within a year, at a fraction of the training cost of their American counterparts[51]. In that configuration, a discretionary American switch does not protect a scarce asset; it accelerates the migration to an abundant substitute. The off-switch caduces — it loses its charge — exactly as the capability it gates becomes a commodity, and the jurisdiction that pulled it is left having taxed its own incumbent without having denied the capability to anyone who wanted it.

The picture is complicated, however, by a parallel architecture. The June 2026 Pax Silica framework — a political declaration signed by the European Union alongside Germany, Greece, Finland, Sweden and the Netherlands — commits adherents to what its principal architect, Jacob Helberg, US Under Secretary for Economic Affairs, described as a 'shared and trusted ecosystem' of American AI developers, in explicit opposition to what he termed the 'retrograde and counterproductive' goal of autonomous national AI infrastructure[52]. The arrangement is structurally asymmetric: signatory states obtain access to frontier AI capabilities whilst accepting American firms as the default providers of their digital-services infrastructure, reinforcing data flows and dependency channels that are difficult to reverse once embedded.[4] A framework of this kind does not accelerate the switch's caducity; it works to arrest it, by converting third-jurisdiction access to frontier AI into a quid pro quo for supply-chain loyalty — the EU–US trade agreement's parallel commitment to purchase at least 40 billion dollars in American AI semiconductors being the balance-sheet expression of the arrangement[52]. The switch therefore caduces only for unaligned jurisdictions; the Pax Silica is the strategy for minimising that set.

The precedents must be diversified, because the keys to geopolitical competitiveness are more complex than any single sector, and the essay's claim is not confined to software. The tariff episode is the non-software instance of the same structure: an authority of exception invoked at discretion, producing not the intended leverage over trading partners but a year of uncertainty that depressed investment, a market crash, and ultimately a judicial finding of illegality and exposure to large-scale refund claims — the discretionary instrument damaging the actor that wielded it[1,2,49]. The Lysenko case is the non-software instance from the science side: ideology degrading a national capability that took decades to rebuild. The through-line is constant across the sectors — defence, aerospace, logistics, cloud, cybersecurity, AI — that a firm might hope to dominate through affirmative executive favour: without the background ecosystem that guarantees a renewable supply of talent and the predictability that retains it, affirmative authoritarian executive action confers no durable benefit. It confers a quarter, perhaps a year, of advantage, purchased against the slow forfeiture of the conditions that make advantage renewable. The strategic error is to mistake the form of dominance — being the firm the state favours — for its substance, which is being the firm the world's talent and capital choose on predictable terms.

7 Four futures

Prospective claims about a fast-moving industry invite two failure modes — false precision and vacuous hedging — and the discipline of this section is to avoid both by reasoning in scenarios rather than point forecasts[53]. The first methodological move is to separate trend from regime. The extrapolation of capability is production science: by the task-horizon method, the length of task a model completes reliably has been doubling on a months-scale cadence, with wide and explicitly modelled uncertainty[54]. That trend tells us that capability is *not* the binding constraint on the frontier's future. The thesis of this essay is orthogonal to the capability curve: what governs decline is not the slope of capability but the *variance of regime*. The model below is therefore a model of regime dynamics and their propagation into corporate viability, and it is a heuristic instrument for disciplining intuition — a device for making the feedback loops, their signs, and their tipping conditions explicit — not a calibrated predictor. Every coefficient in it is a declared assumption.

The model tracks the viability of a frontier laboratory through seven state variables on the unit interval: capability lead relative to the open and foreign frontier (C), regulatory predictability (R), the epistemic coupling of governance to technical fact (E), talent stock (T), capital confidence (K), market access (M), and legitimacy (L), together with an exogenous safety-branding intensity (S). Four feedback loops connect them. *Loop 1*, the

brand-as-munition trap: high S, under low E and R, raises the probability of a discretionary intervention, which depresses M and K. *Loop 2*, flight: falling R drives capital and talent out, which erodes C. *Loop 3*, substitution: falling domestic market access raises adoption of substitutes, which decays the switch's leverage over C. *Loop 4*, legitimacy: discretionary intervention without technical basis erodes L and, with it, the capacity to demand fact-grounded governance, which further decouples E. Two axes organise the outcomes — regime type (consolidating versus reversible) and epistemic posture (denialist versus evidence-restoring) — yielding four futures.

	Denialist (E → 0)	**Evidence-restoring (E → 1)**
Consolidated competitive authoritarianism (R → 0)	**A · Sovereign capture / managed decline.** The firm survives as a national champion stripped of the distributed property that made it valuable; capability migrates abroad; "brains as brand" fully realised.	**C · Unstable combination.** A discretionary regime that nonetheless binds itself to evidence; transient and improbable, useful mainly for showing why R and E tend to co-move.
Reversible / contested (R → mid)	**B · Jurisdictional arbitrage / hollowing.** The firm hedges by offshoring (the Seoul office is the contemporary token) but faces compliance fragmentation; partial, jurisdiction-dependent decline.	**D · Recovery / re-coupling.** Levitsky's reversibility obtains; predictable, fact-grounded process is restored; safety leadership becomes an asset rather than a liability.

The scenario model below runs as a self-contained interactive figure; its controls, read-out and usage instructions travel inside it. Drag the sliders or use the scenario buttons, and read the *lower tail* of the band rather than the central line.

FIGURE · INTERACTIVE SCENARIO MODEL

Regime variance and the decline of a frontier laboratory

A heuristic model of *regime* dynamics — not a forecast. Capability is treated as a rising trend; what governs decline is the variance of the governance regime. Every coefficient is a declared assumption.

Safety-branding intensity (S) **0.70**

Initial regulatory predictability (R_0) **0.50**

Initial epistemic coupling (E_0) **0.50**

Monte-Carlo trials **600**

☑ Regime consolidates (R decays over time)

SCENARIOS A · Sovereign capture | B · Hollowing | D · Recovery | Reset

MEDIAN TERMINAL C
–

P(DECLINE · C<0.5)
–

band = 10th–90th pct · faint lines = individual runs

The line is the **median** capability lead *C* across the ensemble; the shaded band is the **10th–90th percentile**; the dashed line marks the **decline threshold** (C = 0.5). Read the *lower tail* and which loop drives it, not the central value: push *S* high while R_0 and E_0 are low to see the brand-as-munition and flight loops collapse the tail. Drag the sliders or pick a scenario.

Scenario coordinates — **A** Sovereign capture / managed decline: $R_0 \approx 0.30$, $E_0 \approx 0.30$, consolidates . **B** Jurisdictional arbitrage / hollowing: $R_0 \approx 0.50$, $E_0 \approx 0.40$, non-consolidating . **D** Recovery / re-coupling: $R_0 \approx 0.70$, $E_0 \approx 0.70$, non-consolidating . (Scenario C — discretionary regime that nonetheless binds itself to evidence — is transient and is discussed in the text rather than presented as a preset.)

Figure 2. Interactive scenario model (HTML only). The line is the median capability lead C; the shaded band is the 10th–90th-percentile Monte-Carlo envelope; faint lines are individual runs; the dashed line is the decline threshold (C = 0.5). Controls: safety-branding intensity S, initial regulatory predictability R_0, initial epistemic coupling E_0, a regime-consolidation toggle, and the ensemble size. Presets A / B / D set the scenario coordinates discussed in the text.

Note on formats

The model is to be read for its lower tail and its loop attributions, not for its central value. The decisive empirical questions are therefore framed as leading indicators that would confirm or refute the trajectory. The diagnosis is *confirmed* by repetition of the pattern — a second discretionary switch — by an acceleration of talent and capital flight towards predictable jurisdictions, by defensive offshoring of laboratories, and by the explicit discounting of safety evidence in regulatory decisions. It is *refuted* by the institutionalisation of a transparent, reasoned statutory process, by restoration with an effective right of appeal, and by a reversal of the regime — through electoral outcomes or intra-coalitional fracture — without lasting cost to the frontier. A foresight exercise earns its keep not by predicting which branch obtains but by specifying, in advance, the observations that would tell us.

8 What governance would restore the floor

The remedy follows from the diagnosis, and its structure is dictated by the fact that two distinct corrosions must be addressed at once. Three instruments, drawn from the regulatory horizon of the broader project, do the work. The first is a *minimum-continuity obligation*, analogous to universal-service requirements in telecommunications: a floor of guaranteed access below which a provider's service to dependent third-jurisdiction users may not fall without due process. The second is a *mandatory notice period* before any unilateral suspension, which converts an instantaneous recall into a reviewable act. The third is a *multilateral appeal mechanism*, before which a suspension's technical basis can be contested by parties other than the suspending state.

The non-obvious claim is that these instruments are not merely protections for the user; each re-couples the criterion to fact, and so attacks the denialism corrosion as directly as it attacks the discretion. A mandatory notice that must be *reasoned* reintroduces the obligation to state a technical basis; an appeal with *technical review* reintroduces the evidentiary check that science denialism had removed; and a continuity floor with *due process* reintroduces the predictability that competitive authoritarianism had withdrawn.

Where the disease is governance-criteria capture — a criterion rendered arbitrary and fact-immune — the cure must restore both predictability and corrigibility, and these instruments are designed to restore each.

Under a consolidated competitive-authoritarian regime, who institutes the appeal? A discretionary state will not readily bind its own discretion, and an anti-empirical one will not readily reinstate the evidentiary check it has discarded. This is precisely why the appeal must be *multilateral*: no single discretionary jurisdiction can be the guarantor of a floor against its own discretion, and the commons of cognitive infrastructure — used across dozens of jurisdictions — has the standing to demand governance no single member can supply [50], a standard against which the institutional adversary had already declared itself by June 2026. The principal architect of the Pax Silica framework publicly dismissed autonomous third-jurisdiction AI infrastructure as counterproductive fragmentation — the systematic position against which a continuity floor would have to be built, advanced from the office that holds the switch [52]. That structural difficulty is also why the reversibility thesis is load-bearing for the whole argument: the remedy presupposes a re-coupling of criterion and fact that only a political correction can deliver. The instruments describe the destination; whether the polity can travel to it is the question the objections must now confront.

9 Objections and limits

A critical essay of this kind lives or dies by the strength of the case against it, and five objections deserve a steelman rather than a dismissal.

First, the benign reading. The anchor case was framed by the provider itself as a misunderstanding to be resolved, and perhaps it was: bureaucratic improvisation rather than competitive-authoritarian strategy. The response is that the framework does not require intention; it requires structure — a legal instrument of exception, an opaque procedure, and an effect disproportionate to the alleged risk. But the concession must be made honestly: the distinction between misunderstanding and strategy is empirically open, and the essay's claim is about the *structural pattern that both readings share* — that a switch of this kind exists and is pullable at discretion. The diagnosis does not depend on attributing motive, and it should not pretend to establish one.

Second, endogenous decline. There is serious evidence that the training-and-application layer is in an economically unsustainable dynamic — capital expenditure outrunning demonstrable utility, a roughly 30-per-cent rise in compute prices, and cross-lab dependencies that look more like mutual hostage-taking than competition [55]. Perhaps the decline is a bubble, not a governance failure. The correct response is not a contest of

hypotheses but a recognition of composition: financial fragility *amplifies* sensitivity to regime risk, because a sector burning capital cannot absorb discretionary shocks. The essay should model governance and economics as multiplicative, not rival, factors, state explicitly which portion of the decline it attributes to each, and concede the under-identification that follows from their interaction.

Third, that "science denial" is an inflated concept. Recent scholarship holds that the term conflates disparate phenomena and that the alarm around it is overstated[56]. The essay disarms this by adopting only the *precise* Oreskes–Conway version — the selective discounting of impact science — which is falsifiable and requires no claim of wholesale rejection of science. The pillar is built on the version of the concept that survives the strongest objection to it.

Fourth, the national-security steelman. There is a legitimate argument that frontier offensive-cyber capability may warrant restriction, and the essay does not deny the state's authority to block genuinely unsafe deployments. What it attacks is the *form* — the absence of transparent, reasoned, technically grounded process — and it shows, through the universal/non-universal distinction and the wide availability of the capability, why this particular case fails the standard that the authority itself presupposes. To avoid a straw man, the contrary expert position must be represented: the view, argued by practitioners, that the controls in fact *harm* cyber-defence by denying defenders the tools attackers already possess[43]. The essay's claim is the narrower and more defensible one — not that the power should not exist, but that it was exercised in a manner that the power's own justification cannot license.

Fifth, reversibility. If the diagnosed regime is reversible, as Levitsky, Way and Ziblatt maintain[10], then scenario D is available and the decline may be averted. Far from undercutting the argument, this is its point: the model's value is to display the conditional trajectories, and the leading indicators of §7 are what tell us which branch the polity is on. A diagnosis that is reversible is not thereby false; it is, precisely, actionable.

A limit of method must close the section, because it is also a datum of the analysis. Much of the characterisation of the anchor case derives from an interested party, and the government's technical reasoning was never made public. The essay treats that asymmetry not as a licence to assert motive but as the phenomenon itself — the second-order opacity in which neither the public nor the sanctioned provider can audit the basis of a decision that reaches across jurisdictions. The inability to know is not a gap in the argument. It is the argument's subject.

10 Conclusion

The threat Silicon Valley did not see coming was not a competitor, a capability ceiling, or a compute wall. It was the degradation of the governance criterion on which the entire enterprise tacitly depended, by two forces that the industry treated as background noise and that are, in fact, complementary engines of a single failure: competitive authoritarianism, which makes the rule discretionary, and science denialism, which makes it fact-immune. Their conjunction — governance-criteria capture — converts the firms' own safety branding into the legal predicate of their capture, and converts the predictability that retains the world's talent into a discretionary favour revocable by letter. The firms most exposed were those that had wagered on that favour, mistaking the form of dominance for its substance; the substance was always the distributed property of an R&D democracy, and discretion dissolves it. The remedy is a governance that restores both predictability and corrigibility — a continuity floor, a reasoned notice, a multilateral appeal — and the diagnosis that motivates it is reversible but not self-correcting. The off-switch exists whether or not it is governed. The question this essay has tried to make unavoidable is whether the jurisdictions that hold it will learn, before the talent does, that a switch one can pull at will is a switch one has already begun to lose.

Acknowledgements and funding

This work was developed within the AUTAI research project *Artificial Intelligence and Human Autonomy* (PID2022-137953OB-I00), which investigates the conditions of human autonomy in interaction with generative and agentic AI systems. That remit spans the full arc of the technology: the design and training of large language models, the infrastructure through which they are used, and the governance framework — the regulatory order — by which safe and reliable services are placed within reach of the public, without unjustifiable gaps or asymmetries. The present essay belongs to the project's governance strand.

On the use of AI tools. In the spirit of the transparency this essay argues for, the author records that large language models developed by Anthropic — Claude Sonnet 4.6 and Claude Opus 4.8— were used as instruments at several stages of preparation: in articulating and structuring successive drafts of the argument; in cross-checking bibliographic references and source data against primary materials; and in constructing the interactive scenario simulation and its static fallback in §7. The author set the questions, directed each task, reviewed and verified the output, and retains sole responsibility for the analysis, the judgements, and any error that survives. That the instruments employed belong to the very class of systems the essay examines is not incidental to it: the argument is offered not from outside the dependency it describes but from within it.

References

1. Supreme Court of the United States. (2026). *Learning resources, inc. V. trump*. 607 U.S. ___ (2026), No. 24–1287. https://www.supremecourt.gov/opinions/25pdf/24-1287_4gcj.pdf

2. Manak, I. et al. (2026). *A year after 'liberation day,' experts review the costs of trump's tariffs*. Council on Foreign Relations. https://www.cfr.org/articles/a-year-after-liberation-day-experts-review-the-costs-of-trumps-tariffs
3. CNN. (2025). *Early US intel assessment suggests strikes on iran did not destroy nuclear sites*. cnn.com. https://edition.cnn.com/2025/06/24/politics/intel-assessment-us-strikes-iran-nuclear-sites
4. CSIS Nuclear Network. (2025). *Disruption or dismantlement: Diverging assessments of iran nuclear strikes*. nuclearnetwork.csis.org. https://nuclearnetwork.csis.org/disruption-or-dismantlement-diverging-assessments-of-iran-nuclear-strikes/
5. Congressional Research Service. (2025). *U.s. Strikes on nuclear sites in iran* (IN12571). Library of Congress. https://www.congress.gov/crs-product/IN12571
6. Anderson, S. R. (2026). *After the strike: The danger of war in Iran*. Brookings Institution commentary. https://www.brookings.edu/articles/after-the-strike-the-danger-of-war-in-iran/
7. House of Commons Library. (2026). *Israel/US–Iran conflict 2026: Background and UK response* (CBP-10521). Research Briefing, House of Commons Library. https://commonslibrary.parliament.uk/research-briefings/cbp-10521/
8. Ülgen, S. (2026). *From trade dependence to geopolitical leverage: The EU in an era of weaponized interdependence*. Carnegie Endowment for International Peace, Europe Program. https://carnegieendowment.org/europe/research/2026/06/from-trade-dependence-to-geopolitical-leverage-the-eu-in-an-era-of-weaponized-interdependence
9. Levitsky, S., & Way, L. A. (2025). The path to american authoritarianism: What comes after democratic breakdown. *Foreign Affairs*, *104*(2). https://www.foreignaffairs.com/united-states/path-american-authoritarianism-trump
10. Levitsky, S., Way, L. A., & Ziblatt, D. (2026). The price of american authoritarianism. What can reverse democratic decline? *Foreign Affairs*. https://www.foreignaffairs.com/united-states/american-authoritarianism-levitsky-way-ziblatt
11. Brooks, E. (2025). *Lawmakers side-eye billionaire CEOs getting prime trump inauguration seats*. thehill.com. https://thehill.com/homenews/house/5096085-trump-inauguration-tech-ceos/
12. Helmore, E. (2025). Trump inauguration: Zuckerberg, bezos and musk seated in front of cabinet picks. *The Guardian*. https://www.theguardian.com/us-news/2025/jan/20/trump-inauguration-tech-executives
13. Schaake, M. (2024). *The tech coup: How to save democracy from silicon valley*. Princeton University Press.
14. Taskale, A. R. (2026). Materialized science fiction: The tech oligarchy's blueprint for a new global order. *Global Studies Quarterly*, *6*(1), ksag002. https://doi.org/10.1093/isagsq/ksag002
15. Massanari, A. L. (2024). *Gaming democracy: How silicon valley leveled up the far right*. The MIT Press. https://doi.org/10.7551/mitpress/14531.001.0001
16. Fung, A., & Lessig, L. (2025). *Oligarchy in the open: What happens now as the u.s. Is forced to confront its plutocracy problem?* Harvard Kennedy School. https://www.hks.harvard.edu/faculty-research/policycast/oligarchy-open-what-happens-now-us-forced-confront-its-plutocracy
17. Montgomery, B. (2025). Silicon valley's tycoons are bending the knee to trump. *The Guardian*. https://www.theguardian.com/technology/2025/jan/11/trump-silicon-valley
18. Faguy, A. (2025). Washington post cartoonist quits after bezos satire is rejected. *BBC News*. https://www.bbc.com/news/articles/cn4xk4y2emro

19. Jiménez, M. (2025). Dimite una dibujante del 'washington post' por el veto a una viñeta en que bezos se arrodillaba ante trump. *El País*. https://elpais.com/comunicacion/2025-01-05/dimite-una-dibujante-del-washington-post-por-el-veto-a-una-vineta-en-que-bezos-se-arrodillaba-ante-trump.html
20. Mathur-Ashton, A. (2025). The notable attendees – and no-shows – at trump's inauguration. *U.S. News & World Report*. https://www.usnews.com/news/national-news/articles/2025-01-20/who-attended-trumps-inauguration-the-billionaires-and-far-right-leaders-who-gathered-at-the-swearing-in
21. Preskey, N. (2025). Influencers, tech bros, MMA fighters: Trump's inauguration guests. *BBC News*. https://www.bbc.com/news/articles/cgkjgmkn10ko
22. Stanford Institute for Human-Centered AI. (2026). *The 2026 AI index report*. Stanford University. https://hai.stanford.edu/assets/files/ai_index_report_2026.pdf
23. Wiggins, R. Z., Piontek, T., & Metrick, A. (2014). *The Lehman Brothers bankruptcy A: overview*. Yale Program on Financial Stability (YPFS) Case Study. https://ypfsresourcelibrary.blob.core.windows.net/fcic/YPFS/001-2014-3A-V1-LehmanBrothers-A-REVA.pdf
24. Matten, D. (2026). Fascism as a management philosophy. *Philosophy of Management*, *25*, 21–49. https://doi.org/10.1007/s40926-025-00355-1
25. Levitsky, S., & Way, L. A. (2010). *Competitive authoritarianism: Hybrid regimes after the cold war*. Cambridge University Press. https://doi.org/10.1017/CBO9780511781353
26. Levitsky, S., & Way, L. A. (2002). Elections without democracy: The rise of competitive authoritarianism. *Journal of Democracy*, *13*(2), 51–65. https://doi.org/10.1353/jod.2002.0026
27. Coeckelbergh, M. (2026). Technofascism: AI, big tech, and the new authoritarianism. *AI & Society*. https://doi.org/10.1007/s00146-026-02862-9
28. Oreskes, N., & Conway, E. M. (2022). From anti-government to anti-science: Why conservatives have turned against science. *Daedalus*, *151*(4), 98–123. https://doi.org/10.1162/daed_a_01946
29. Oreskes, N., & Conway, E. M. (2010). *Merchants of doubt: How a handful of scientists obscured the truth on issues from tobacco smoke to global warming*. Bloomsbury Press.
30. Proctor, R. N., & Schiebinger, L. (Eds.). (2008). *Agnotology: The making and unmaking of ignorance*. Stanford University Press.
31. Mills, M. A., & St. Clair, P. (2025). The strange new politics of science. *Issues in Science and Technology*, *41*(3), 40–48. https://doi.org/10.58875/NDTQ1755
32. Olesko, K. M., Eames, A., Mody, C. C. M., Löwy, I., Zeller, T., Walker, M., Ash, M. G., & Haraway, D. (2025). The crisis in american science. *History of Science*, *63*(2), 125–165. https://doi.org/10.1177/00732753251343655
33. Bruni, E., & Lefsrud, L. M. (2026). The collective organization of science denial: Toward a framework for collective response. *Open Access Government*, 184–185. https://www.openaccessgovernment.org/article/the-collective-organization-of-science-denial-toward-a-framework-for-collective-response/202441/
34. Oreskes, N. (2019). *Why trust science?* Princeton University Press.
35. European Union. (2024). *Regulation (EU) 2024/1689 of the european parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act)*. Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
36. Council of the European Union and European Parliament. (2026). *Digital omnibus on AI: Provisional agreement to simplify and streamline AI rules*. consilium.europa.eu; European

Parliament. https://www.europarl.europa.eu/RegData/etudes/ATAG/2026/789329/EPRS_ATA(2026)789329_EN.pdf

37. Liberties (Civil Liberties Union for Europe). (2026). *AI omnibus: Fast-tracking deregulation, weakening digital rights*. liberties.eu. https://www.liberties.eu/en/stories/ai-omnibus-ep-vote/45715
38. Tech Policy Press. (2026). *EU's AI act delays let high-risk systems dodge oversight*. techpolicy.press. https://www.techpolicy.press/eus-ai-act-delays-let-highrisk-systems-dodge-oversight/
39. Anthropic. (2026). *Statement on the US government directive to suspend access to fable 5 and mythos 5*. Anthropic. https://www.anthropic.com/news/fable-mythos-access
40. Tech Times. (2026). *Fable 5 export ban day six: Anthropic opens seoul office, vows models back in days*. techtimes.com. https://www.techtimes.com/articles/318668/20260618/fable-5-export-ban-day-six-anthropic-opens-seoul-office-vows-models-back-days.htm
41. Bombach, K. M., Dwyer, K., Roberson, J. E., et al. (2026). *AI company anthropic suspends access to claude fable 5, claude mythos 5 following US export control directive*. National Law Review (Greenberg Traurig LLP alert). https://natlawreview.com/article/ai-company-anthropic-suspends-access-claude-fable-5-claude-mythos-5-following-us
42. Korea JoongAng Daily. (2026). *Anthropic confident of re-enabling mythos, fable 5 access in coming days, executive says*. koreajoongangdaily.com. https://www.koreajoongangdaily.com/business/anthropic-confident-of-reenabling-mythos-fable-5-access-in-coming-days-executive/12727522
43. Moussouris, K. (2026). *The fable 5 export controls harm US cyber defense*. Luta Security. https://www.lutasecurity.com/post/the-fable-5-export-controls-harm-us-cyber-defense
44. Kahn, J. (2026). Anthropic disables fable and mythos AI models after u.s. Government bars it from giving foreigners access. In *Fortune*. https://fortune.com/2026/06/13/anthropic-fable-mythos-ai-models-disabled-us-government-foreign-access/
45. Coeckelbergh, M. (2020). *AI ethics*. MIT Press.
46. Tech Jacks Solutions. (2026, June 18). *Anthropic opens seoul office with six enterprise deals while its top models remain export-blocked*. https://techjacksolutions.com/ai-brief/anthropic-opens-seoul-office-with-six-enterprise-deals-while/
47. Anthropic. (2026). *Claude fable 5 and claude mythos 5*. Anthropic. https://www.anthropic.com/news/claude-fable-5-mythos-5
48. Lewis, M. (2018). *The fifth risk: Undoing democracy*. W. W. Norton & Company.
49. Basco, S. (2026). How trump's renewed tariff chaos will stifle investment around the world. In *The Conversation*. https://theconversation.com/how-trumps-renewed-tariff-chaos-will-stifle-investment-around-the-world-276789
50. Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.
51. U.S.-China Economic and Security Review Commission. (2026). *Two loops: How china's open AI strategy reinforces its industrial dominance*. USCC. https://www.uscc.gov/sites/default/files/2026-03/Two_Loops--How_Chinas_Open_AI_Strategy_Reinforces_Its_Industrial_Dominance.pdf
52. Ponce de León, R. (2026). *Qué es la pax silica y por qué puede ser una trampa para europa. eldiario.es*. https://www.eldiario.es/internacional/pax-silica-trampa-europa_1_13334330.html

53. Tetlock, P. E., & Gardner, D. (2015). *Superforecasting: The art and science of prediction*. Crown Publishers.
54. Kwa, T., West, B., Becker, J., et al. (2025). Measuring AI ability to complete long tasks. *arXiv Preprint*, *arXiv:2503.14499*. https://doi.org/10.48550/arXiv.2503.14499
55. Epoch AI. (2026). *How much AI compute do frontier labs use?* epoch.ai/gradient-updates. https://epoch.ai/gradient-updates/frontier-labs-dont-use-most-ai-compute
56. Contessa, G. (2025). *Science denial: Post-truth or post-trust?* Cambridge University Press.

Footnotes

[1] A caveat is in order. Nothing here presupposes that political management can be stripped of its partisan connotations and constraints where the priorities of corporate or economic governance are concerned. The point is rather the difficulty of disaggregating, within any forward business strategy, the elements that exhaustive critical analysis finds obscured beneath isolated symptoms — anti-intellectualism, an aversion to academic elites and to the agencies and institutions that steward expert knowledge — which are habitually swept up into broader categories such as science denialism *qua* avowed position of a party with presidential ambitions. Given the volume of contingency and uncertainty that burdens all political management, lacing it with an irrationality and complexity that are often unmanageable, the very act of attributing these further traits already betrays a deficit of reliable, operative and effective instruments for redirecting or confronting unforeseen risks and scenarios. ↩

[2] A useful independent calibration of this structural pattern: in June 2026, Sinan Ülgen of Carnegie Europe described the United States as having become paradoxically one of the greatest systemic risk multipliers for European supply chains — applying the vocabulary of systemic risk, from within the transatlantic partnership, to the partnership's own anchor state[8]. ↩

[3] The reach of this is easy to understate. Once the ultimate decider operates by decoupling its criteria from the evidence that ought to bind them — in vaccination policy, or in the regulation of air quality and pollution, as readily as in artificial intelligence — the pertinent question is no longer which procedure will be damaged but which will not: what part of the unavoidable interactions between corporations and political actors escapes the arbitrariness of whoever decides in the last instance, and what procedure, vital or trivial, remains intact under a scheme that admits no error correctable by proof? ↩

[4] The explicit target of Helberg's critique is the United Nations Global Digital Compact, which calls on each state to develop its own complete AI infrastructure — hardware, data, models, and national computational capacity — as the basis of digital sovereignty. Helberg characterises this vision as a form of 'synchronised mediocrity', on the grounds that it leads every country to reconstruct capabilities that already exist at greater scale elsewhere. Under the Pax Silica logic, that aspiration is replaced by a division of labour in which American frontier developers supply AI capabilities in exchange for what Helberg frames as a 'reliable

and indispensable supply chain' of critical raw materials — the arrangement presented publicly as 'mutual economic security'. ↩