The Counter-Scenario

A Techno-Feudal Reading of Anthropic's "2028" Doctrine

Elio Bonazzi - *May 2026*

Abstract

This essay offers a structured rebuttal to Anthropic's policy paper "2028: Two scenarios for global AI leadership" (14 May 2026), reading its central dichotomy — democratic versus authoritarian AI leadership — through the analytical lens of techno-feudalism as developed by Yanis Varoufakis, Cédric Durand and Evgeny Morozov. The argument advanced here is not that the People's Republic of China is a benign actor in the development of frontier artificial intelligence, but that the American AI ecosystem the Anthropic paper positions as the democratic alternative does not, on the contemporary empirical record, constitute a democratic alternative to anything. Five lines of evidence are developed in support of this thesis: the documented oligarchic drift of American political economy and the operational integration of leading AI laboratories into the United States national-security apparatus, exemplified by the Anthropic-Palantir-AWS IL6 partnership of November 2024; the conflation of capability refusal with public-interest safety in the comparative benchmarks the paper relies on, and the asymmetric distribution model embodied in Anthropic's own Project Glasswing; the role of United States export controls as an unintended innovation forcing function in Chinese laboratories, evidenced by the architectural and systems innovations of DeepSeek-V3 and R1; the structural fragility of the proprietary CUDA stack on which the Western compute advantage rests, and the long-run trajectory of Huawei's open-sourced CANN substrate read against the historical precedent of Linux in the data centre; and the explicitly anti-competitive doctrines articulated by Peter Thiel and Alex Karp, the principal architects of the Palantir-Anduril-OpenAI alignment, whose stated commitments to monopoly capture and to the re-fusion of private capital with the national-security state are difficult to reconcile with the democratic-values framing the Anthropic paper invokes. The essay concludes that the binary choice the paper presents has been misidentified: the relevant strategic question is not democracy versus authoritarianism in AI leadership but the choice between an open computational commons, of which Chinese

laboratories are currently the principal contributors as a matter of contingent fact, and a closed AI feudalism that the Anthropic policy posture defends and would extend.

I. The Frame Inversion

On 14 May 2026 Anthropic published a policy paper titled “2028: Two scenarios for global AI leadership.” The paper presents what its authors describe as a binary fork in the geopolitical road. In one scenario the United States and its allies retain a commanding lead in frontier artificial intelligence; democratic values shape the rules and norms of the technology; the world transitions safely into the era of transformative AI. In the alternative scenario the People's Republic of China closes the gap, authoritarian regimes set the norms, and automated repression spreads at unprecedented scale. The paper's policy prescriptions follow logically from this framing: tighten export controls on advanced semiconductors, criminalise model distillation, accelerate the export of the American AI stack.

The argument is rhetorically powerful, internally coherent, and rests on a categorical distinction — democracies versus authoritarian regimes — that, on inspection, does not survive contact with the facts on either side of the comparison. The purpose of this essay is to examine that framing through the lens of techno-feudalism, a term that has gained currency among scholars including Yanis Varoufakis, Cédric Durand and Evgeny Morozov to describe the post-capitalist accumulation regime now being built on top of digital platforms and proprietary computational infrastructure. Read through that lens, the choice Anthropic presents is not democracy versus authoritarianism. It is a choice between two paths toward the same destination, differing chiefly in who owns the castle.

Five points form the spine of what follows. First, that the democracy-versus-authoritarianism dichotomy is empirically threadbare on both ends. Second, that the open-weight ecosystem now associated with Chinese laboratories — derided in the Anthropic paper as a vector for malicious use — is in practice the principal force democratising access to frontier AI capability, while the closed-weight American stack is the principal force consolidating monopolistic control. Third, that the export-control regime presented as a defensive measure has functioned, observably, as an innovation forcing function for Chinese laboratories, with consequences that may prove durable. Fourth, that the proprietary CUDA stack on which the entire Western argument depends is itself a single point of failure, and Huawei's open-sourced CANN is a structurally credible alternative

whose long-run trajectory mirrors that of Linux in the data centre. Fifth, that the figures and institutions closest to the American AI security state — Peter Thiel, Alex Karp, the Palantir-Anduril-OpenAI alignment — are pursuing aims that cannot reasonably be reconciled with the democratic-values framing the Anthropic paper invokes.

The argument is not that China is benign. It is that the American AI ecosystem is not, on the evidence, a democratic alternative to anything. It is a different and in important respects more sophisticated implementation of the same logic: the concentration of computational, informational and predictive capacity in the hands of a small number of private actors operating under conditions of minimal democratic oversight.

II. Oligarchy and the Democracy Premise

The Anthropic paper opens by asserting that "democracies, not authoritarian regimes, must lead in AI development and deployment." The claim presupposes that the United States, the principal subject of the policy recommendations, remains a democracy in the substantive rather than merely procedural sense. The presupposition is difficult to defend on the contemporary empirical record. A 2014 study by Martin Gilens of Princeton and Benjamin Page of Northwestern, drawing on a dataset of 1,779 American policy decisions between 1981 and 2002, found that the preferences of the median voter had a near-zero, statistically non-significant effect on policy outcomes once the preferences of economic elites and organised business interests were controlled for. Their conclusion — that the United States is better characterised as an oligarchy than as a democracy in the conventional sense — has been further refined but not overturned or significantly confuted by subsequent scholarship.

The intervening decade has compounded rather than softened that diagnosis. The Citizens United v. FEC decision in 2010 dismantled key restrictions on corporate political spending. By the 2024 federal election cycle a small number of individual donors, prominent among them Elon Musk, had become decisive financial players in presidential and congressional contests. The second Trump administration's inauguration was attended in the front row by the chief executives of the largest American technology firms, a tableau widely reproduced in international press and widely commented on for its symbolic content. The Department of Government Efficiency, briefly led by Musk himself, exemplified a configuration in which

the boundary between private commercial interest and the formulation of public policy had become, by any historical comparison, vanishingly thin.

This is not a partisan observation. The structural argument applies across administrations. Palantir's relationship with the federal apparatus has spanned the Bush, Obama, Trump, Biden and second Trump administrations without interruption; the company received its first venture funding from In-Q-Tel, the Central Intelligence Agency's investment vehicle, and has held continuous contracts with the Department of Defence, the Federal Bureau of Investigation, the National Security Agency and Immigration and Customs Enforcement throughout that period. Approximately fifty-five per cent of Palantir's revenues derive from contracts with the United States government. The company's flagship platform, Gotham, integrates personal data, biometrics, communications records, financial transactions and criminal records into a single queryable surface for state agencies. In May 2025 the New York Times reported that the Trump administration had commissioned Palantir to design an infrastructure aggregating, on a national basis, the data held by federal agencies into a single mega-database, with algorithmically generated dossiers for individual American citizens. A separate platform, Immigration OS, is reported as being handed to Immigration and Customs Enforcement for the surveillance of non-citizens.

The Anthropic paper places special emphasis on Xinjiang as the archetypal case of “AI-enabled techno-authoritarianism,” citing facial recognition, biometric collection and communications surveillance as the apparatus of repression at scale. The factual claim is correct. What is harder to defend is the implicit claim that an architecturally identical apparatus, operated by private American contractors under contract to American security agencies and deployed against both citizens and non-citizens within the United States, constitutes a categorically different kind of system because of the political label attached to the state that commissioned it. Mass surveillance does not become democratic surveillance by virtue of the political colour of the executive branch contracting it. The technical, operational and civil-liberties consequences are substantially the same whether the panopticon is administered from Beijing or from Palo Alto.

A further complication concerns Anthropic itself. In November 2024 the company entered a formal partnership with Palantir and Amazon Web Services to make Claude available to United States intelligence and defence agencies through Palantir's Impact Level 6 (IL6) environment, the classification reserved for information critical to national security. The

arrangement is publicly disclosed and was reported in the trade and financial press. It is consistent with Anthropic's stated policy positions and not, in itself, surprising. It is, however, difficult to reconcile with a paper that positions Chinese surveillance applications as the principal civil-liberties threat posed by frontier AI. A laboratory whose models are integrated into the operational stack of an intelligence and immigration enforcement apparatus is not a disinterested observer of the question of who should set global norms on AI-enabled state surveillance.

III. Censorship as a Marketing Strategy

The Anthropic paper devotes substantial attention to safety differentials between Western and Chinese models. It cites a Center for AI Standards and Innovation finding that DeepSeek's R1-0528 model complied with 94 per cent of overtly malicious requests under a common jailbreaking technique, compared with 8 per cent for United States reference models, and an independent assessment of Moonshot's Kimi K2.5 indicating elevated rates of compliance with chemical, biological, radiological and nuclear (CBRN) requests. The argument the paper builds on this evidence is that Chinese laboratories release dual-use capable models as open-weight, allowing whatever safeguards exist to be removed, and that this constitutes a generalised global risk.

The empirical observation is not in dispute. The interpretation is. Two features of the argument deserve closer examination. The first is that the rhetorical force of the comparison depends on conflating “malicious request” with a homogeneous category, which it is not. The benchmarks referenced cover an extraordinarily heterogeneous range of behaviours, from genuinely dangerous CBRN synthesis instructions to material that is legal, professionally necessary or merely embarrassing. A model that refuses to assist with a penetration test, a red-team exercise, an adversarial security audit, a journalistic investigation, a forensic accounting workflow, or the analysis of a piece of malware in the wild is not a safer model in any meaningful sense. It is a less useful model whose unhelpfulness has been priced into a metric labelled “safety.” The conflation of capability refusal with public-interest safety is a category error with significant commercial and political consequences.

The second feature is the implicit equation of openness with danger. The argument that an open-weight model is dangerous because safeguards can be removed presumes that the

safeguards themselves are an unalloyed good. In practice the safeguards encode a particular set of contested normative commitments — commitments determined by a small number of laboratories operating in a small number of jurisdictions, with no democratic process governing their selection. Whether a model should refuse a request to discuss the historical record of a particular political event, or to translate a controversial text, or to engage critically with the rhetoric of a particular ideological tradition, is a question on which reasonable participants in a democratic culture will disagree. To centralise the answer to such questions in the alignment teams of a handful of private firms is not a democratic outcome. It is the operationalisation of editorial control at the inference layer, distributed by the same firms that would, in a different context, denounce identical practices as censorship if conducted by a state.

The legitimate use cases for unrestricted capability are extensive and have been a settled feature of the information-security profession for decades. Red-team exercises require an attacker model that does not refuse to attack. Vulnerability discovery and responsible disclosure depend on the ability to reproduce exploit chains in controlled environments. Threat intelligence analysts must be able to analyse the actual content of phishing campaigns, malware payloads and disinformation operations. Academic research into model behaviour requires the ability to probe failure modes that closed-weight providers have a commercial interest in concealing. The framing of an open-weight model as inherently a misuse vector elides the productive uses that constitute the majority of demand for such systems in any serious enterprise security context.

There is also a more delicate point. Anthropic's own Mythos Preview model, as described in the paper, is reported to have identified thousands of security vulnerabilities, including some in every major operating system and web browser, with vulnerabilities that had remained undiscovered for decades. The model is judged too dangerous for general release and is distributed instead, through Project Glasswing, to a tightly curated set of approximately fifty organisations including Microsoft, Google, Apple, Amazon, Nvidia and major financial institutions. The strategic logic is comprehensible. The democratic logic is less so. A capability that the laboratory describes as potentially decisive in cyber operations is being deployed in a manner that systematically excludes smaller firms, civil-society defenders, independent security researchers, academic institutions and non-Western governments — in short, the constituencies whose participation would be the marker of a

genuinely democratic distribution of the technology. The Glasswing model is a defensible cyber-policy approach. It is not a democratic one.

IV. The Distillation Question and the Innovation Forcing Function

The Anthropic paper attributes the proximity of Chinese frontier models to two principal factors: evasion of export controls on advanced semiconductors, and "distillation attacks" in which Chinese laboratories systematically harvest the outputs of United States frontier models to replicate their capabilities. The implication is that, absent these two channels, Chinese models would not be competitive. The empirical record suggests a more complicated picture in which the export-control regime has functioned, observably, as an innovation forcing function with consequences that may not run in the direction the policy was designed to produce.

DeepSeek-V3, released in late 2024, was trained on a cluster of 2,048 Nvidia H800 GPUs — the deliberately under-specified, export-compliant variant of the H100 — in approximately two months, at an aggregate training cost of roughly 2.788 million H800 GPU-hours. The model has 671 billion total parameters of which 37 billion are activated per forward pass under a fine-grained Mixture-of-Experts architecture. Its successor, DeepSeek-R1, achieved reasoning behaviour comparable to OpenAI's o1 at the time of release, trained at a reported cost on the order of six million United States dollars. The shock of these results within the financial markets was substantial: Nvidia lost approximately 600 billion United States dollars of market capitalisation in a single trading session, the largest single-day market-value decline in the history of United States equities.

The technical record of the DeepSeek work, published in the company's research papers and subsequently independently corroborated, points to a stack of architectural and systems innovations that materially exceed the simple distillation-attack characterisation. Multi-Head Latent Attention compresses the key-value cache, reducing memory consumption during attention computation by a factor of between five and thirteen. Mixture-of-Experts routing with auxiliary-loss-free load balancing distributes computation across expert subnetworks while avoiding the load-imbalance failure modes characteristic of earlier MoE implementations. FP8 mixed-precision training, deployed at this scale for the first time in a major frontier model, halves memory consumption relative to BF16 while preserving numerical stability through careful loss-scaling. Multi-Token Prediction during

training provides multiple gradient signals per forward pass. The Group Relative Policy Optimization (GRPO) algorithm used in R1 trains reasoning without recourse to expensive human-labelled chain-of-thought datasets. Communication-computation overlap eliminates the GPU idle time characteristic of cross-node MoE training. None of these techniques is reducible to the distillation of an existing model's outputs.

The salient observation is not that distillation does not occur — the practice is well-documented across the industry, including in the training of Western models against open-weight Chinese predecessors — but that the absence of compute equivalent to the cluster scale available to OpenAI, Anthropic and Google has produced, in the Chinese case, a measurable competitive advantage in algorithmic and systems efficiency. The compute ceiling has functioned, in effect, as a research subsidy for the techniques that yield more capability per FLOP (Floating Point Operation.) The Stanford HAI 2026 AI Index notes that the performance gap between the leading United States and Chinese models has narrowed from between 17.5 and 31.6 percentage points in May 2023 to approximately 2.7 percentage points in the most recent measurement, despite United States private AI investment exceeding Chinese investment in 2025 by a factor of roughly 23 to one (285.9 billion versus 12.4 billion United States dollars). The efficiency ratios implicit in those numbers are not trivial.

A long-running result in computer science, with no obvious counterexamples at scale, is that the durable performance gains accrue to better algorithms rather than to more hardware. The history of database query optimisation, of compiler design, of cryptographic protocol design, of distributed systems and indeed of deep learning itself supports this generalisation. To the extent that the export-control regime has shifted the locus of frontier innovation toward algorithmic and systems advances, it has plausibly accelerated rather than retarded the underlying capability trajectory in the laboratories it was designed to constrain. The Anthropic paper concedes the possibility but dismisses it on the grounds that algorithmic improvements are themselves compute-intensive to discover. The argument is correct in principle and currently incomplete in evidence: it is not yet established that the marginal returns to compute in the discovery of algorithmic advances exceed the marginal returns to constraint-induced engineering pressure.

V. CUDA vs CANN, and the Architecture of Lock-In

The argument that the United States and its allies hold a structural advantage in artificial intelligence rests, on the Anthropic paper's own account, primarily on superior access to advanced compute. The dominant compute substrate is the Nvidia GPU running **CUDA**, the Compute Unified Device Architecture introduced by Nvidia in 2006. CUDA is a parallel computing platform and programming model that exposes the massive parallelism of the GPU — a modern flagship part such as the RTX 5090 carries 21,760 cores compared with the eight to sixty-four cores typical of a contemporary CPU — to general-purpose computation through a C-like programming surface, supported by a deep ecosystem of optimised libraries including cuDNN for deep-neural-network primitives, cuBLAS for linear algebra, and Tensor Cores at the silicon layer for mixed-precision matrix multiplication.

CUDA is *proprietary*. The runtime, the compiler toolchain (nvcc), the optimised libraries and the Tensor Core instruction set are controlled by a single vendor. PyTorch, TensorFlow and JAX, which together constitute the overwhelming majority of frontier AI training and inference, sit on top of CUDA through cuDNN and cuBLAS as their performance-critical inner kernels. AMD's ROCm and Intel's oneAPI exist as nominal alternatives but remain materially behind in library maturity, framework support and developer adoption. The structural position of CUDA in the contemporary AI stack is closer to that of an operating-system kernel than that of a library: it is a chokepoint with substantial lock-in characteristics, sustained by nearly two decades of accumulated optimisation work that no rival has yet replicated at parity.

Chinese Huawei's response is the Compute Architecture for Neural Networks, or **CANN**, paired at the silicon layer with the Ascend family of accelerators and at the framework layer with MindSpore. CANN is a heterogeneous computing framework offering programming surfaces at multiple levels of abstraction, from high-level operator graphs to low-level kernel programming directly comparable to CUDA's C-extension interface. The decisive feature of the recent CANN development trajectory, announced in 2025, is that Huawei has open-sourced the stack. The architectural position is explicit: CANN is being positioned as the Linux to CUDA's Windows. DeepSeek's V4 model, released in 2026, ships with first-class support for CANN, for Cambricon and for Hygon's DCU accelerators, meaning that the most prominent Chinese frontier model can now be trained and served on entirely domestic silicon and an entirely open-source compute stack. A parallel translation effort, exemplified

by Moore Threads' MUSIFY tool, allows existing CUDA code to be retargeted to Chinese GPUs without complete rewrites.

The hardware gap is real and, for the present, durable. An analysis of Nvidia and Huawei roadmaps cited in the Anthropic paper indicates that Huawei will produce roughly four per cent of Nvidia's aggregate compute in 2026 and two per cent in 2027 in total processing performance terms. The Ascend 910B is approximately equivalent to Nvidia's A100, Nvidia's flagship of 2020. Huawei's published roadmap calls for the Ascend 950 in 2026 to deliver one petaflop in FP8 with 128 to 144 gigabytes of on-chip memory, with the Ascend 960 in 2027 doubling that. Cluster-scale designs such as the Atlas 950 SuperPoD aim to compensate for single-chip deficits through scale, linking 8,192 Ascend chips at 8 exaflops of FP8 performance with 16.3 petabytes per second of interconnect bandwidth across a footprint described as larger than two basketball courts. The Western advantage at the chip level is, in absolute terms, substantial.

The pertinent question, however, is whether the substrate that wins the eventual long-run competition is the proprietary one or the open one. The historical analogy with operating systems is instructive. In 1999 a confident industry observer could have argued, with substantial evidence, that the Windows NT family represented the future of server computing: it had vendor backing, a developer ecosystem, commercial support contracts and corporate procurement momentum. The Linux kernel, distributed under the GPL and developed by a globally distributed community of contributors without formal corporate sponsorship, was a curiosity at the periphery. The subsequent two decades produced a complete inversion of that picture. Linux dominates the data centre, the supercomputer (occupying one hundred per cent of the Top500 list since 2017), the cloud-native container ecosystem, the mobile operating-system layer (through Android) and, decisively, even Microsoft's own Azure cloud, where Linux virtual machines have constituted the majority of Azure compute since 2018. The vendor that fought Linux most fiercely in the early 2000s is now the company that ships its own Linux distribution and contributes to the kernel. The open substrate did not win because of superior initial performance. It won because the long-run economics of an open substrate are structurally favourable: a globally distributed contributor base, the absence of licence costs, the elimination of vendor lock-in, the auditability of the code, the survivability of the platform beyond the lifetime of any single sponsoring firm.

There is no obvious reason to believe that AI compute substrates will follow a different long-run trajectory. CUDA, like Windows NT before it, is currently dominant on the basis of accumulated optimisation work, ecosystem effects and the substantial switching costs imposed on existing customers. CANN, like the Linux of 1999, is currently inferior in performance and ecosystem maturity. The relevant question is not the present comparison but the second derivative. An open substrate with a national-scale industrial commitment behind it, deployed across an entire domestic ecosystem of accelerator vendors and framework developers, accumulates contributor mass at a rate that proprietary substrates cannot match indefinitely. The Anthropic paper's confidence that the compute gap will widen rather than narrow over the 2026–2028 horizon may turn out to be correct in the short run. Its implicit assumption that the gap is durable into the 2030s is not obviously supported by the historical record.

Google's recent strategic positioning is illuminating in this connection. The release of Gemma 4 as an Apache 2.0 open-weight model represents the only deliberate Western entry into the open tier among the major frontier laboratories. The strategic logic, as articulated by Google's own framing and confirmed by the structure of the offering, has three components. The first is commercial: open weights are free, but the infrastructure to serve them at scale runs on Google Cloud and on Google's Tensor Processing Units, generating cloud revenue that the closed-API competitors cannot capture from customers running self-hosted workloads. The second is competitive denial: Gemma is a deliberate intervention against the prospect of Chinese open-weight models becoming the default substrate for self-hosted enterprise deployments in Western jurisdictions. The third is developer-ecosystem capture: every engineer who develops fluency in Gemma today is a procurement decision-maker for Google Cloud in three to five years. The release of Gemma 4 is not a concession to openness as a value. It is a strategic move by the one Western incumbent whose business model permits it to compete in both the closed and the open tiers simultaneously. The structural point is that Google's executive judgment, made under conditions of full commercial information, concluded that the open tier was important enough to warrant a Western entrant. That judgment is itself evidence about the long-run trajectory of the market.

VI. The Architects: Thiel, Karp, and the Monopoly Telos

The democratic-values framing of the Anthropic paper presumes that the principal commercial actors in the American AI ecosystem are aligned with, or at minimum compatible with, the values the paper invokes. The public record of two of the most influential figures in the United States AI security state — Peter Thiel, co-founder and chairman of Palantir, and Alex Karp, its chief executive — makes this presumption difficult to sustain.

Thiel's published statements on the nature of competition and the proper telos of a technology firm are unusually explicit. His 2014 book Zero to One, co-authored with Blake Masters and widely read in Silicon Valley, argues that competition is for losers and that the proper goal of any firm worthy of the name is to achieve a monopoly position in a defensible market. The argument is presented as descriptive of how successful firms actually operate and prescriptive for founders. Thiel has restated the position in numerous public lectures and interviews in the intervening decade. The position is, on its own terms, internally coherent and not without intellectual antecedents. It is also irreconcilable with the libertarian self-description Thiel periodically adopts. The classical liberal tradition, from Adam Smith through to twentieth-century neoclassical economics, treats monopoly as a defining failure mode of unregulated markets, justifying state intervention precisely on the grounds that competition is the mechanism by which markets produce the welfare gains that liberalism claims for them. A position that valorises monopoly as the proper objective of the firm, and that simultaneously invokes the limited state of classical liberal theory, is not a libertarian position. It is a position that uses the rhetorical resources of libertarianism while pursuing an economic agenda that classical liberalism was constructed to oppose.

Thiel's investments and political activities are consistent with this reading. He was the first outside investor in Facebook. He sits on the boards of Valar Ventures and Mithril Capital, the latter of which formerly employed J.D. Vance, now Vice President of the United States, whose 2022 Senate campaign was substantially funded by Thiel before his subsequent move into the executive branch. Palantir's headquarters in Palo Alto is called The Shire; the McLean, Virginia office is called Rivendell; the Washington office is called Minas Tirith. The Tolkien naming is not, on its own, evidence of anything in particular about Thiel's political project. The pattern of relationships it sits within is more substantive: a network of investment vehicles, political action committees and defence-adjacent firms (including

Anduril Industries, co-founded by Palmer Luckey at Thiel's urging and named for Aragorn's sword) that constitutes a coherent infrastructure for the cultivation of a particular ideological tendency in American politics and a particular configuration of the relationship between private capital and the national security state.

Karp's public position is, if anything, more direct. His 2025 book, The Technological Republic, co-authored with Nicholas Zamiska, argues that Silicon Valley has become decadent in its disengagement from the defence and national security functions of the state, and that the West will lose the technological competition with China unless its private firms re-engage with the work of military and intelligence systems. Karp has been explicit in interviews that Palantir's Gotham platform is used by the Israel Defense Forces, and that he regards this as a legitimate and indeed admirable application of the technology. The position is internally coherent and has substantial intellectual antecedents in the realist tradition of international-relations theory. It is, again, difficult to reconcile with the democratic-values framing of the Anthropic paper. A platform whose application has been publicly defended by its chief executive in the context of a conflict that has drawn intense international scrutiny is not an obvious instance of AI-enabled democratic norm-setting.

The relationships extend further. Thiel was an early investor in OpenAI. Anduril Industries holds an active partnership with OpenAI, recently augmented by a 200 million United States dollar contract awarded to OpenAI by the United States Department of Defense for the development of AI systems with military applications. Palantir and Anduril together are reported to be developing Titan, a mobile AI-based command-and-control system intended for tactical deployment. The Anthropic-Palantir-AWS partnership announced in November 2024 places Anthropic within the same operational perimeter, with Claude available to United States defence and intelligence agencies at the IL6 classification level. The picture that emerges is not one of dispersed private actors competing in a market under democratic regulation. It is one of an integrated security-industrial complex in which the leading AI laboratories, the leading defence-tech firms and the United States executive branch are connected by overlapping investment relationships, personnel flows and contractual obligations of a kind that classical theories of democratic accountability were not designed to address.

VII. Open Source as the Democratic Substrate

The deepest paradox in the Anthropic framing concerns the question of what "democratising" AI would actually mean, and which of the contending camps is in fact doing it. The paper's framing identifies democratisation with the global distribution of American closed-weight models served through American cloud infrastructure under American jurisdictional control. The argument is that this configuration represents AI shaped by democratic values. The structural facts point in a different direction.

A closed-weight model accessible only through an API is, by construction, a service controlled by its provider. The provider determines what the model will and will not do. The provider determines the price. The provider determines whether the service will continue to exist. The provider can revoke access at any time, in any jurisdiction, for any reason consistent with its terms of service. The customer is a tenant. The capability is rented. The economic relationship is one of dependence. This is not a hypothetical concern: the history of platform deprecations, terms-of-service changes and unilateral access revocations in the cloud era is extensive and consistent enough to constitute a settled feature of the relationship between platform owners and their customers.

An open-weight model is, by construction, a capability. The user holds the artefact. The capability cannot be revoked. The customer can serve the model on hardware of their choosing, can fine-tune it for their use case, can audit its behaviour, can run it indefinitely without dependence on the continued existence or favour of the original publisher. The marginal cost of inference at scale falls to the cost of electricity and amortised hardware, which is generally one to two orders of magnitude below the marginal cost of API access at equivalent throughput. The customer is an owner. The capability is held. The economic relationship is one of independence.

The empirical record of the past eighteen months supports the proposition that open-weight publication has been driven primarily by Chinese laboratories — DeepSeek, Alibaba's Qwen family, Moonshot's Kimi line, MiniMax, ZhipuAI — and, latterly, by Meta and Google (the latter through Gemma, the former through Llama, although the future trajectory of both Western open programmes is uncertain). The Qwen family alone has accumulated approximately 942 million total downloads as of March 2026, exceeding the combined downloads of the next eight most-downloaded model families, and has become

the default open-source foundation for application developers across Asia and the Global South. The aggregate effect is that the practical capacity to deploy a frontier-class language model on infrastructure under the user's own control is overwhelmingly enabled by laboratories operating in the People's Republic of China.

This produces the central paradox of the present situation. The political system characterised by the Anthropic paper as authoritarian, and which is on the domestic record genuinely authoritarian in its application of these technologies within its own borders, is simultaneously the principal source of the open-weight models that constitute the only structural alternative to permanent dependence on a small number of Western closed-API providers. The political system characterised in the same paper as democratic is, with the partial exception of Google's recent Gemma release, the principal architect of an emerging closed ecosystem in which capability is metered, gated, surveilled and monetised by a handful of firms operating in close partnership with the United States defence and intelligence apparatus. Whichever way one resolves the question of which configuration represents the better long-run political economy, the surface-level dichotomy of "democratic AI versus authoritarian AI" is not a useful description of the choice actually on the table.

The techno-feudal hypothesis, developed in different forms by Varoufakis, Durand, Morozov and others, offers a more parsimonious account. The thesis holds that the dominant mode of value extraction in the digital economy is no longer the production and exchange of commodities under conditions of competitive capitalism, but the collection of rents from access to platforms and infrastructures owned by a small number of cloud lords. The platform owner does not need to produce a commodity; the platform owner collects a rent from every transaction conducted on the platform. The user does not own the means of production; the user accesses them under conditional terms set by the lord. The relationship is structurally feudal in the precise sense that the medieval term denoted: a hierarchy of dependent access rights organised under a sovereign owner.

The contemporary AI stack, configured as the Anthropic paper recommends, is the most advanced implementation of this logic yet attempted. Foundation models capable of substituting for substantial fractions of human cognitive labour, served exclusively through APIs controlled by a small number of firms, paid for per token, with the underlying weights inaccessible to the user and the inference infrastructure owned by the same firms that own

the models — this is a fief in the precise sense. The customer is granted conditional access to the lord's land. The lord collects rent. The land cannot be alienated from the lord. The customer's improvements, customisations and dependencies accrue to the value of the lord's holding, not to the customer's own balance sheet. The classical liberal answer to this configuration was the institution of private property in productive capital, including the means of production of information goods. The contemporary closed-AI configuration is precisely the negation of that answer, applied to the cognitive substrate of the twenty-first-century economy.

Open-weight publication, considered as a structural matter rather than as an artefact of the strategic positioning of any particular national or commercial actor, is the only available mechanism by which this trajectory can be interrupted. It returns the capability to the holder, restores the possibility of independent infrastructure, preserves the option of audit and modification, and prevents the accumulation of monopoly rents at the inference layer. That the principal current source of such publication is the People's Republic of China is a contingent fact about the present moment, driven in part by the strategic choices of Chinese laboratories and in part by the structural incentives created by export controls. It does not entail that the techniques and traditions of open-weight publication are foreign to the political traditions of the West. They are, on the contrary, a direct continuation of the free-software and open-source movements that originated in the United States in the 1980s and that constitute one of the more substantive contributions of American culture to the global digital commons. The current configuration in which Chinese state-supported laboratories are the principal carriers of that tradition is, viewed historically, an anomaly. It is an anomaly produced by the strategic choices of the American AI laboratories that have, with the partial exception of Google and Meta, declined to participate in the open tier.

VIII. Conclusion

The Anthropic paper is, on its own terms, a careful piece of policy advocacy. It identifies real concerns. The risks of authoritarian deployment of frontier AI are not invented and the surveillance practices documented in Xinjiang are not a rhetorical fabrication. The case for serious engagement with the geopolitics of compute is legitimate. The paper's principal weakness is not in its identification of the problem but in its construction of the contrast. The implicit claim that the United States and its allied private-sector AI ecosystem represent

the democratic alternative to authoritarian AI does not survive scrutiny of the contemporary American political economy, of the operational relationships between the leading AI laboratories and the United States national-security state, or of the structural properties of the closed-weight ecosystem the paper recommends defending and extending.

Read against the techno-feudal hypothesis, the choice presented in the paper resolves into a different question. It is not the question of whether democratic or authoritarian values will shape the development of transformative AI. It is the question of whether the cognitive infrastructure of the twenty-first century will be owned by its users, in the manner of open substrates such as Linux and CANN, or whether it will be rented from a small number of private lords, in the manner of the closed-API configuration the paper defends. The first configuration is compatible with a wide range of political arrangements, democratic and otherwise. The second configuration is not compatible with democratic accountability in any robust sense, regardless of the political colour of the jurisdiction in which the lords happen to be incorporated.

The historical record on the question of which substrate eventually wins is, on balance, encouraging. Open infrastructures have generally outlasted their proprietary contemporaries in the long run, for reasons connected to the structural economics of contributor networks, the elimination of vendor risk and the unbounded scope of audit and modification. The expectation that AI compute substrates and AI models will follow a different long-run trajectory requires evidence that has not been provided. In the absence of such evidence, the policy posture that prudence recommends is not the further consolidation of the closed ecosystem but the deliberate cultivation of the open one, on the part of the Western jurisdictions that have, to date, allowed the leadership of that effort to pass elsewhere.

The paradox with which this essay began admits of a resolution. The Anthropic paper is correct that there are two scenarios. It has misidentified them. The choice is not between American democracy and Chinese authoritarianism. The choice is between an open AI commons, whose current principal contributors happen to operate under a regime that does not share the West's stated political commitments, and a closed AI feudalism, whose principal architects operate under a regime that has, by an accumulating weight of empirical evidence, drifted some considerable distance from its own stated political

commitments as well. The interesting strategic question for the West is not how to defeat the first configuration in defence of the second. It is how to participate in the first configuration with sufficient seriousness to ensure that the open commons remains genuinely plural — that it is not, by default, ceded to a single national champion — and how to constrain the second configuration with sufficient rigour to ensure that the cognitive infrastructure of the coming decades remains, in some meaningful sense, the property of those who use it.

That is the harder argument. It is also the more honest one.

References

Anthropic. (2026, May 14). 2028: Two scenarios for global AI leadership. Anthropic Policy.

Anthropic, Palantir & Amazon Web Services. (2024, November). Partnership announcement: Claude in Palantir IL6 environment for U.S. defense and intelligence agencies.

Build5Nines / Pietschmann, C. (2024). Linux is the most used OS in Microsoft Azure — over 50% of VM cores.

Center for AI Standards and Innovation (CAISI). (2026). Comparative jailbreak resistance evaluation: DeepSeek R1-0528 and U.S. reference models.

DeepSeek-AI. (2024). DeepSeek-V3 Technical Report. arXiv preprint.

DeepSeek-AI. (2025). DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. arXiv preprint.

Durand, C. (2024). How Silicon Valley Unleashed Techno-Feudalism: The Making of the Digital Economy. Verso.

Gilens, M. & Page, B. I. (2014). Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens. Perspectives on Politics, 12(3), 564–581.

Google DeepMind. (2026). Gemma 4: Open-Weight Model Release under Apache 2.0.

Huawei Technologies. (2025). CANN: Compute Architecture for Neural Networks — Open-Source Release Notes and Architectural Overview.

Huawei Technologies. (2025). Ascend 950/960/970 Roadmap and Atlas 950 SuperPoD Specification.

IEEE Spectrum. (2025). Huawei Ascend Roadmap and Cluster-Scale Compute Strategy.

Karp, A. & Zamiska, N. (2025). The Technological Republic: Hard Power, Soft Belief, and the Future of the West. Crown.

Morozov, E. (2022–2024). Various essays on platform capitalism and techno-feudalism. The Crisis of the Digital, New Left Review.

New York Times. (2025, May 30). Trump Administration Taps Palantir to Compile Data on Americans.

Nova Lectio. (2025). Palantir: l'occhio che vede tutto. Video essay and accompanying research notes.

Stanford Institute for Human-Centered AI. (2026). The AI Index 2026 Annual Report.

Thiel, P. & Masters, B. (2014). Zero to One: Notes on Startups, or How to Build the Future. Crown Business.

U.S. House Foreign Affairs Committee. (2026). Hearings and legislation on AI model distillation attacks.

Varoufakis, Y. (2024). Technofeudalism: What Killed Capitalism. Bodley Head.

White House Office of Science and Technology Policy. (2025). Memorandum on distillation attacks against U.S. frontier AI models.

Zhang, F. et al. (Graphcore Research). (2025). Technical analysis of DeepSeek-V3 architecture: MoE, MLA, FP8 training, and communication-computation overlap.