FAIR LABS FRONTIER SERIES

VOL 07 · 2026

The Compute Compact

Compute is the one input to advanced AI that is physical, measurable, and hard to hide — which makes it either the sharpest oversight lever or the deepest source of concentration

PREPARED FOR POLICYMAKERS · CHIPMAKERS · CLOUD PROVIDERS · RESEARCH FUNDERS

FAIR Labs
FAIR ARTIFICIAL INTELLIGENCE RESEARCH LABS

February 2026
FL-2026-07

FAIR LABS FRONTIER SERIES · VOLUME 07

The Compute Compact: Compute is the one input to advanced AI that is physical, measurable, and hard to hide — which makes it either the sharpest oversight lever or the deepest source of concentration

The seventh volume of the Frontier Series treats compute — chips, clusters, energy, interconnect — as AI's most governable input and its most concentrated one. The report builds the case for compute-indexed thresholds as the backbone of Volume 04's governance ladder, confronts the security logic and limits of export controls, and makes the affirmative case for a publicly funded compute commons so that oversight capacity does not live only inside the firms being overseen. It closes with a decoded map of the compute stack for readers who have never had to think about a wafer fab before.

REPORT CODE	FL-2026-07
PUBLICATION DATE	February 2026
PRIMARY AUDIENCES	Policymakers · Chipmakers · Cloud Providers · Research Funders
SERIES	FAIR Labs Frontier Series, Fair Artificial Intelligence Research Labs
SUGGESTED CITATION	*The Compute Compact: Compute is the one input to advanced AI that is physical, measurable, and hard to hide — which makes it either the sharpest oversight lever or the deepest source of concentration.* FAIR Labs Frontier Series, Vol. 07. Fair Artificial Intelligence Research Labs, 2026.
KEYWORDS	compute governance, chips, export controls, concentration, compute commons, interoperability, thresholds

IN BRIEF · EXECUTIVE SUMMARY

You cannot hide a gigawatt. That makes compute the best lever we have — if the access fight doesn't decide everything first.

Every other input to advanced AI — data, algorithms, talent — is easy to move, copy, or hide. Compute is not: chips are manufactured in countable facilities, clusters draw metered power, and training runs leave a physical footprint. That makes compute the most promising substrate for governance in Volume 04's toolkit. It is also the most concentrated layer of the entire AI stack, which means the fight over who gets access to it will shape who gets a say in everything downstream.

01 **Compute is measurable where everything else is not.**

You can meter a cluster's power draw; you cannot meter an algorithm. That asymmetry is why compute-indexed thresholds anchor Volume 04's obligation ladder rather than any capability self-report.

02 **Concentration compounds fastest at the physical layer.**

Chip design and advanced fabrication are more concentrated than model training itself — the chokepoint sits upstream of the part everyone argues about.

03 **A threshold set once ages badly.**

Training compute has grown far faster than any legislative calendar. A number written into a 2022 statute is a rounding error by the time it is enforced without review.

04 **Export controls are a security tool with real costs, not a free action.**

Restricting advanced-chip access to specific actors is a defensible security measure and a standing tax on alliance management and innovation diffusion. Both are true at once.

05 **Oversight capacity needs its own compute.**

Independent safety and public-interest research cannot check what it cannot run. A funded compute commons is the precondition for outside verification meaning anything.

06 **Lock-in hardens fast; interoperability has to arrive before it does.**

Switching costs at the model, cloud, and chip layer compound yearly. The remedy is cheap now and expensive later.

CONTENTS

Inside this report

01

The measurable input

Data can be copied at zero cost. Algorithms leak. Talent moves between employers overnight. Compute is the one input to advanced AI that stays where you can see it.

A frontier training run requires thousands of accelerator chips, drawing megawatts of metered power, housed in a countable number of facilities, for weeks at a time. None of that can be hidden the way a weight file or a training recipe can be copied onto a drive and walked out the door. This is not a minor administrative convenience; it is the single structural fact that makes AI governance more tractable than it would otherwise be. Volume 04 builds an obligation ladder keyed to "a measurable capability or training-compute index" precisely because compute is the rare input that can actually play that role honestly.

The public groundwork already exists to build on. Export-control regimes restricting advanced-chip access to specific destinations are operating today; that is a qualitative fact about the current landscape, not a FAIR Labs proposal. What this volume adds is the governance architecture around that fact: how to turn "compute is measurable" into thresholds that work, remain current, and do not simply entrench whoever already has the most of it.

> **WHY THIS VOLUME PAIRS WITH VOLUME 04**
>
> Volume 04 designs the obligation ladder — transparency register, external evaluation, pre-deployment review, licensing. This volume builds the one input precise enough to key that ladder to: compute, not calendars.

The chapters ahead move from the physical stack to the policy stack built on top of it. Chapter 2 maps where concentration actually sits — not always where public attention goes. Chapter 3 details the threshold mechanics. Chapter 4 takes the security logic of export controls seriously and pairs it with the interoperability remedies that keep the market from calcifying around today's incumbents. Chapter 5 makes the case for a com-

pute commons. Chapter 6 turns this into an eight-point agenda; Chapter 7 is a plain-language map of the stack for readers encountering it for the first time.

> *You cannot hide a gigawatt. Everything this volume proposes follows from that one fact.*
>
> THE PREMISE OF THIS VOLUME

02

The concentration problem

Public attention fixes on which lab trained the biggest model. The tighter chokepoints sit further upstream, where fewer firms compete and switching is harder.

2.1 The stack has more layers than the headlines suggest

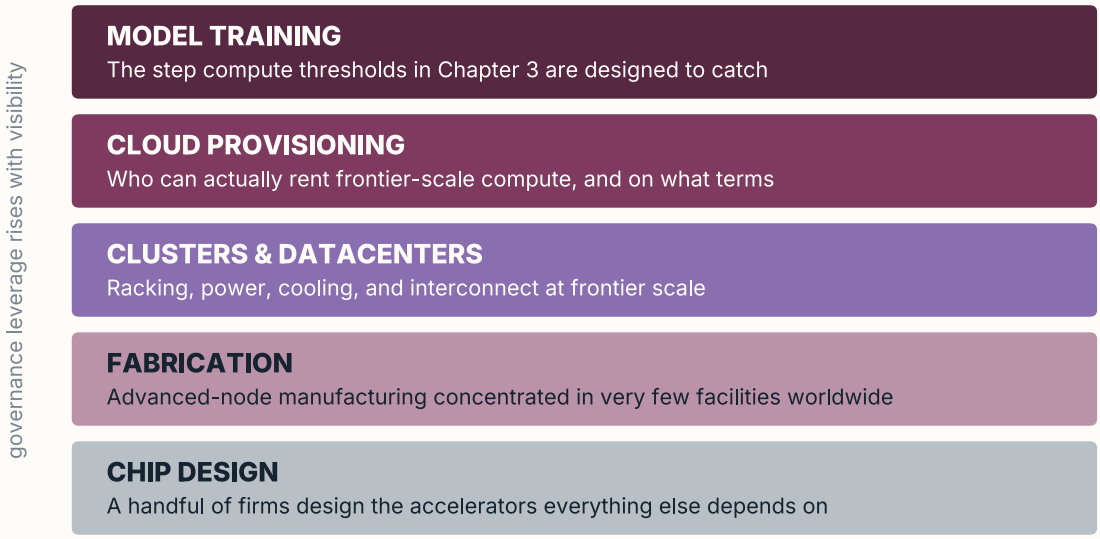
Advanced-node chip fabrication is concentrated in a small number of facilities capable of the relevant process nodes; accelerator chip design sits with a handful of firms whose architectures frontier training depends on; cluster construction and cloud provisioning add a further layer of capital intensity that only a few organizations can sustain. By the time a model is trained, the concentration in the physical layers beneath it is often tighter than the concentration among the labs doing the training — a fact policy debates focused only on "which company has the biggest model" routinely miss.

2.2 Capital intensity is a moat that compounds

Each successive generation of frontier-scale compute costs more than the last, in dollars and in lead time. That is an ordinary capital-intensive-industry dynamic — the same logic that concentrated aircraft manufacturing and semiconductor fabrication generations ago — but it interacts badly with AI's winner-take-most dynamics in adjacent markets. Volume 03 files this under structural drift and Volume 02 asks who ends up collecting the resulting economic dividend; this volume is where the remedy actually has to be engineered, because competition policy applied only at the model layer will not reach a chokepoint that sits at the fabrication layer.

EXHIBIT 7.1 · CONCEPTUAL FRAMEWORK

The compute stack: five layers, uneven visibility, uneven concentration



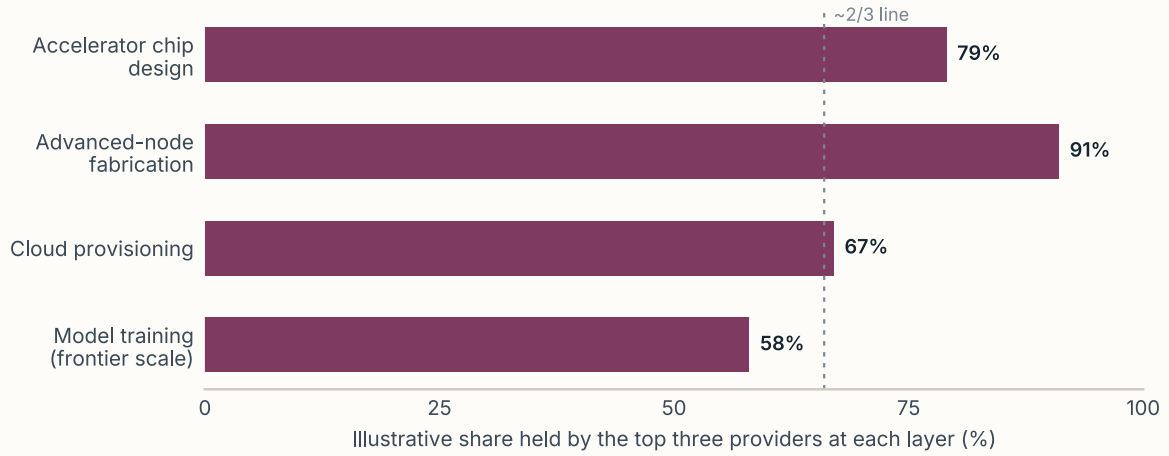
Illustrative construct organizing the physical and commercial layers beneath a trained model; not a map of any specific company's position.

2.3 Where the concentration actually sits

Ranking layers by how tightly held they are produces a counter-intuitive result: the layers with the least public debate are often the most concentrated ones. This matters for where remedies belong — competition enforcement and interoperability mandates aimed only at the most visible layer will leave the tighter chokepoints untouched.

EXHIBIT 7.2 · ILLUSTRATIVE

Concentration by layer: the least-discussed layers are often the tightest



Illustrative index for orientation; not a measured market-share study of any named firm or product.

03

Thresholds as governance

A compute threshold is only as good as its ability to stay current. This chapter details the mechanics Volume 04's obligation ladder depends on.

3.1 What makes a good threshold

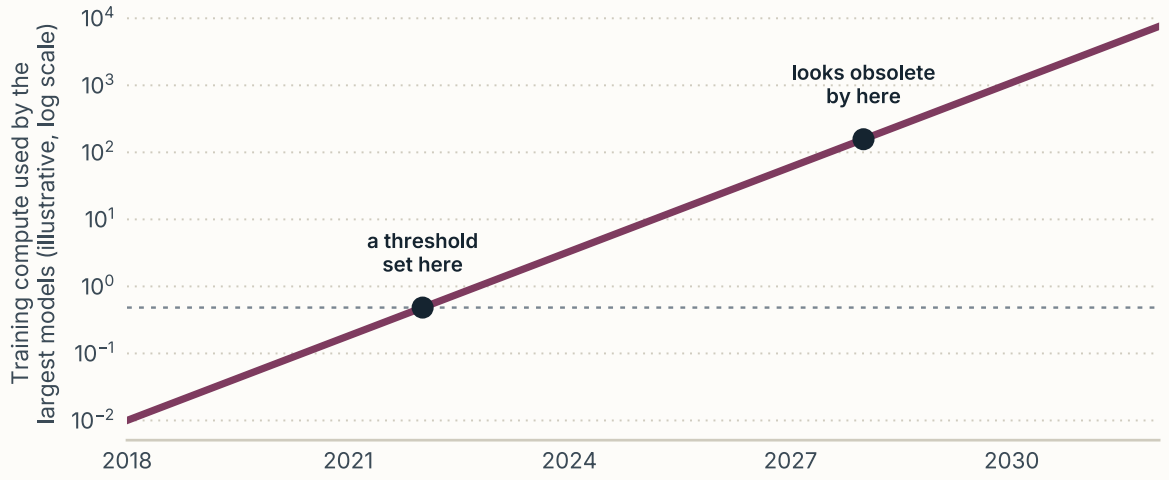
A workable compute threshold has three properties: it is measurable by a third party without relying on the trained party's own report; it correlates plausibly with the capability the regime actually cares about, acknowledging the correlation is imperfect and loosens over time as algorithms grow more efficient; and it is reviewed on a fixed schedule rather than left as a fixed number in statute. Training-compute order of magnitude is the most defensible current proxy — not because it is perfect, but because it is the hardest of the available options to fake.

3.2 Thresholds age; that is not a flaw, it is a feature to design for

The compute required to reach a given capability level falls over time as algorithms and hardware improve. A threshold calibrated correctly today will, left alone, eventually classify yesterday's frontier system as routine and next year's mid-tier deployment as exempt from scrutiny it may still warrant. This is not a reason to abandon compute thresholds; it is the specific argument for Volume 04's recommendation that thresholds be reviewed annually rather than fixed once in statute.

EXHIBIT 7.3 · STYLIZED

Why a static threshold ages out of relevance



Stylized illustration of training-compute growth outpacing a fixed threshold; axes are relative and not a specific numeric forecast.

3.3 Know-your-customer at the provisioning layer

Because cloud provisioning is itself a chokepoint, it doubles as a natural enforcement point: providers of frontier-scale compute can be required to know who they are renting to and at what scale, the same way financial institutions know their customers before moving large sums. This reaches actors who would never file a voluntary transparency report, without requiring anyone to inspect a training run directly.

3 tests

third-party measurable, capability-correlated, and hard to fake — what a good threshold needs

1 layer

cloud provisioning, the natural enforcement point KYC rules can reach

annual

the review cadence Volume 04 recommends for every threshold this chapter proposes

04

Export controls and interoperability

Two policy tools sit at opposite ends of the same problem: one restricts who can reach frontier compute, the other keeps the market for it from calcifying around whoever already has it.

4.1 The security logic, honestly stated

Restricting export of the most advanced chips and manufacturing equipment to specific destinations is a real, operating policy today, grounded in a straightforward security logic: the same physical measurability that makes compute governable also makes it interdictable. FAIR Labs does not dispute the logic. It also does not pretend the tool is free: export controls tax legitimate diffusion of the technology to allied and developing economies, accelerate incentives for restricted actors to build independent capacity, and complicate alliance management when partners weigh the controls differently. A security policy that presents itself as costless is not being honest about the trade-off it is making.

4.2 Interoperability before lock-in hardens

The opposite failure mode is lock-in at the model, cloud, or chip layer — switching costs that compound every year a customer stays put, foreclosing exactly the competitive entry that would discipline the concentration Chapter 2 describes. Volume 02 names the remedy in brief: interoperability, portability, and access terms for essential inputs, enforced before today's advantages harden into the next decade's tollbooths. This volume adds the specific compute-layer version — data and model portability between cloud providers, non-discriminatory access terms for compute capacity, and standardized interfaces that keep switching a decision rather than a multi-year migration project.

HOLDING BOTH TRUTHS AT ONCE

Export controls and interoperability mandates look like they pull in opposite directions — one restricts access, the other guarantees it. They target different populations: controls restrict a narrow set of destinations on security grounds; interoperability protects everyone else from concentration among the firms freely permitted to compete.

05

The compute commons

Every instrument in this series that depends on independent verification — Volume 04's accredited evaluators, Volume 05's alignment research — depends on this chapter existing first.

5.1 Oversight capacity needs its own compute

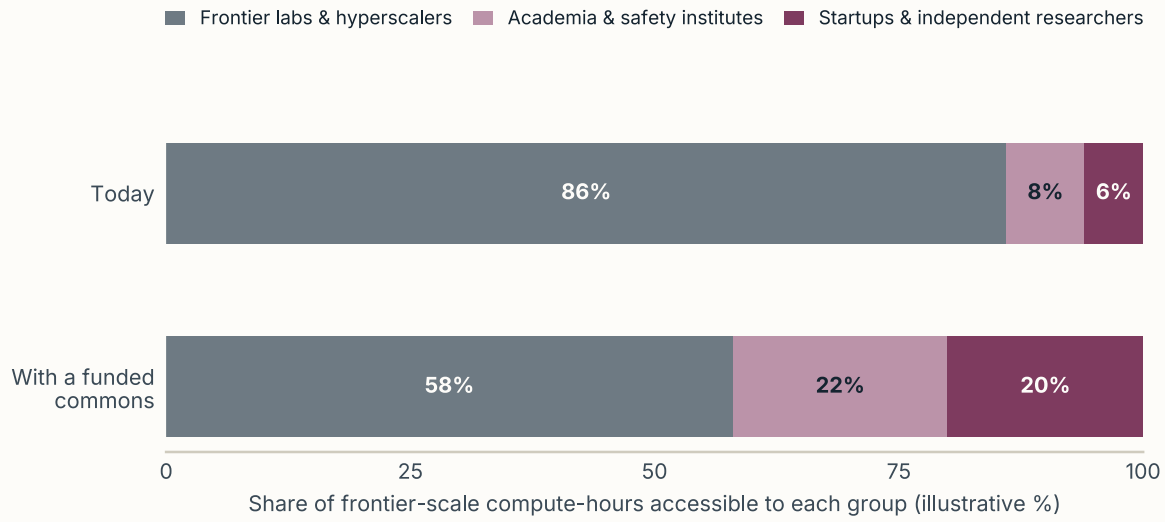
Volume 05 names "guaranteed structured access plus commons compute" as a precondition for independent alignment research to become a stable career rather than a sabbatical from industry. The same logic extends to every other independent verification function this series proposes: an accredited evaluator, an academic auditor, or a public-interest researcher who must ask a frontier lab for compute to check that lab's own claims is not independent in any way that matters. A compute commons — publicly funded access to frontier-scale compute, allocated to safety research, academic work, and public-interest verification — is the precondition, not a nice-to-have alongside it.

5.2 Design options, not a single blueprint

A commons can take several forms, and jurisdictions will reasonably choose differently: a national compute grant program allocating cycles competitively; subsidized cloud credits for accredited safety and academic institutions; or a shared, publicly governed cluster operated for public-interest research specifically. What matters more than the exact mechanism is the allocation principle — access should track public-interest verification need, not just academic prestige or institutional connections, or the commons will simply reproduce the concentration it exists to counteract.

EXHIBIT 7.4 · ILLUSTRATIVE

Who gets to run frontier-scale compute, today and under a funded commons



Illustrative allocation shares for orientation, not a measurement of any specific program or jurisdiction's current compute budget.

5.3 Paying for it

A compute commons is not charity; it is the same logic that funds public research infrastructure in every other capital-intensive science — telescopes, particle accelerators, sequencing capacity — extended to the infrastructure independent AI oversight now requires. The realistic funding base is public research budgets supplemented by contributions from the frontier labs and cloud providers whose own long-term license to operate depends on credible outside verification existing.

06

Recommendations

Eight moves, each with an owner, that turn a measurable input into an actual governance and access architecture.

CC-01 **Tier obligation to a compute index, reviewed annually**

The physical anchor for Volume 04's four-tier obligation ladder; a number reviewed on a schedule, not fixed once in statute.

OWNERS: **GOVERNMENTS · CLOUD PROVIDERS · CHIPMAKERS** · DEEP DIVE: VOL 04

CC-02 **Fund the compute commons as core research infrastructure**

Budgeted like a telescope or an accelerator, not treated as a discretionary grant program cut in the first lean year.

OWNERS: **RESEARCH FUNDERS · GOVERNMENTS** · SEE ALSO: VOL 05

CC-03 **Require know-your-customer rules at the provisioning layer**

Cloud providers renting frontier-scale compute should know who they serve and at what scale — the natural enforcement point in the stack.

OWNERS: **CLOUD PROVIDERS · GOVERNMENTS**

CC-04 **Keep export controls narrowly targeted and reviewed, not a blanket wall**

State the security logic honestly, scope it to the actual risk, and reassess as the technology and the landscape shift.

OWNERS: **GOVERNMENTS · STANDARDS BODIES**

CC-05 **Mandate interoperability and portability before lock-in hardens**

Switching rights at the model, cloud, and chip layer, enforced now while switching is still a choice rather than a migration project.

OWNERS: **REGULATORS · CLOUD PROVIDERS** · DEEP DIVE: VOL 02

CC-06 **Publish a public compute-concentration index annually**

Track concentration at every layer of the stack, not just the model layer the headlines cover.

OWNERS: **STATISTICS AGENCIES · REGULATORS**

CC-07

Ring-fence a share of public compute for safety and alignment research

A protected allocation that survives budget cycles, sized to make outside verification of frontier claims genuinely possible.

OWNERS: **GOVERNMENTS · SAFETY INSTITUTES**

CC-08

Re-threshold on a fixed schedule, not by amendment fights

Build the review into the original statute so recalibration is routine administration, not a fresh legislative battle every time.

OWNERS: **REGULATORS · LEGISLATURES**

07

The compute stack, decoded

A plain-language map of the stack for readers who have never had to think about a wafer fab, keyed to the governance lever that reaches each layer.

STACK LAYER	WHO TYPICALLY SITS HERE	GOVERNANCE LEVER THAT REACHES IT	DEEP DIVE
Chip design	A handful of accelerator-architecture firms	Export controls; design-level reporting	Ch. 4
Fabrication	Very few advanced-node manufacturing facilities	Export controls on equipment and destination	Ch. 4
Clusters & datacenters	Hyperscalers and specialized operators	Energy/siting disclosure; KYC on capacity	Ch. 3
Cloud provisioning	A small number of frontier-scale providers	Know-your-customer duties; interoperability rules	Ch. 3–4
Model training	Frontier labs, a growing set of challengers	Compute-indexed reporting and licensing thresholds	Vol 04
Independent verification	Academics, safety institutes, accredited evaluators	Compute commons access	Ch. 5

FAIR Labs reference map, intended as an orientation tool for non-specialist policymakers; not exhaustive of every firm or role at each layer.

WHERE TO GO NEXT

Regulators tying this volume's thresholds into a full obligation ladder should read Volume 04 next. Researchers and funders designing commons allocation: Volume 05. Economists tracing who collects the resulting dividend: Volume 02. Boards weighing dependency on concentrated compute suppliers: Volume 10. Every threshold in this volume assumes the reporting and verification institutions Volume 04 builds.

Fair Artificial Intelligence Research Labs (FAIR Labs) is an independent research institute working to ensure that advanced artificial intelligence is developed safely, governed wisely, and shared fairly. Through rigorous technical research, policy analysis, and public engagement, FAIR Labs works with scientists, governments, companies, and communities to widen the circle of people who benefit from — and have a say in — the most consequential technology of our time.

The **Frontier Series** is FAIR Labs' flagship collection of stakeholder briefings on advanced AI — one shared evidence base, ten vantage points. Each volume stands alone; together they form a common map for scientists, executives, policymakers, workers, and citizens navigating the same transition.

FL-2026-07 · FAIR Labs Frontier Series · 2026