COMPETITIVE INTELLIGENCE REPORT

AI Pattern Landscape

Λ Competitive Benchmark — Q2 2026

Twelve deployment patterns. Structural propagation analysis.
115 cited sources. Every score justified. Full methodology shown.

The Grove Foundation
July 2026

Contents

The Grove Foundation · July 2026 · CC BY 4.0

Twelve deployment patterns. Structural propagation analysis. 115 cited sources. Every score justified. Full methodology shown.

From the Managing Director

Karp told you what is happening. Here is the paper trail.

On July 1, Alex Karp went on CNBC's Squawk Box and said "something has gone completely wrong" with how AI is sold. He relayed what enterprise customers tell him privately: they are paying for "tokens that create no value" while "these people are stealing the weights and alpha of my business." The day before, Palantir had posted a written manifesto making the same case in colder language: "Controlling your weights is controlling your fate... If you let others control your weights, you are allowing them to migrate the alpha of your business to theirs." The market bid Palantir up 8 to 9 percent that day.

Discount the messenger — you should. Two days before the interview, Palantir announced the NVIDIA sovereign-deployment deal that his critique conveniently sells. And note what the whole episode did not contain: not one document. No clause, no exhibit, no ROI figure — the only number Karp volunteered was his own company's cash-flow projection. He named a real anxiety and offered no evidence.

This letter is the evidence. What Karp asserted from the seller's seat, the vendors have already written down — in terms of service, in privacy policies, in their own published archives, dated and quotable. The "alpha migration" he warned about is not a prediction. It is a clause. If your last vendor review checked PII exposure and breach liability — the cloud-era review — it ran right past this, because the vendors spent the past fourteen months rewriting the terms that review was built to check.

The Executive Summary that follows defines the three words this quarter turns on — trace, harness, custody — and the board shows the market already pricing the shift. This letter's job is narrower: to show you, in the vendors' own documents, that custody of the trace — the record of how your organization thinks — has already been transferred. Lawfully. At every privacy setting. By design.

The mechanism is the language you were taught to read as protection. De-identification, aggregation, Zero Data Retention — in the cloud era, those words meant the vendor kept less. In these contracts, they are how the vendor keeps more: strip the regulated, identifiable datapoint off the interaction, and lawfully keep the interaction pattern, which is the asset. "We delete the plaintext and keep the embeddings" is not a safeguard. It is the harvest.

Karp gave you the sentiment. The documents give you the proof.

The reserved-use ratchet, in the vendors' own documents

Date	Vendor	What the document says
Apr 3, 2025	Cognition (Platform ToS)	**Training is opt-in: Customer Data "will not be used for model training purposes unless you opt-in." Usage Data is merely "collected, analyzed, and utilized" — no ownership claim. (§3.3.1, §3.3.2, vendor's own archive)**
Oct 6, 2025	Cursor (Privacy Policy)	De-identify-and-keep clause enters the consumer policy: de-identified data is used for "improving or adding features, and conducting research."
Oct 15, 2025	Cursor (Data Use)	**Regime cutoff printed on the live page: prompts and telemetry flow to model providers only for accounts created after this date. Older accounts keep the older deal.**
Nov 7, 2025	U.S. District Court, S.D.N.Y.	The MDL court orders roughly 20 million "de-identified" consumer ChatGPT logs produced in discovery — and forces retention of chats users had already deleted.
May 18, 2026	OpenAI (US Privacy Policy §4)	Deletion exemption standing and reaffirmed: deleted content is removed within 30 days — unless "it has already been de-identified and disassociated from your account." Your delete does not reach the de-identified copy.
Jun 9, 2026	Cursor (Data Use)	At the strongest privacy setting: "all plaintext code for computing embeddings ceases to exist after the life of the request. The embeddings and metadata about your codebase... may be stored in our database." The clause sits outside the Privacy Mode gate.

Date	Vendor	What the document says
Jun 30, 2026	Cognition (Platform ToS)	**The ratchet closes. Training flips to opt-out — available only on paid tiers, admin-only on Teams (§3.3.1). A new sentence appears: "Cognition owns all right, title, and interest in and to the Usage Data" (§3.3.2). Zero Data Retention is defined to reach Customer Data only (§1.8) — and Customer Data "expressly excludes… Usage Data" (§1.2).**

Every row is a primary source; the two Cognition rows come from the vendor's own archive. Full citations: Works Cited [14]–[19] and the source note closing this letter.

Read the last row again. In fourteen months, one vendor moved from "we will not train unless you say yes" to "we may train unless you pay and say no" — and took formal title to the anonymized record of how your teams use the product. The retention pledge and the ownership carve-out were engineered together, in the same document: the pledge reaches exactly as far as the bucket the vendor does not own, and stops precisely where the owned bucket begins. No setting, on any tier, reaches the titled pattern.

And one more clause, because it tells you how the vendors value this asset. The same terms that take title to your usage patterns bar you from the reverse: customers may not use the services "to train competing artificial intelligence models" (Cognition §2.3). Your interaction record is an asset worth owning — when they own it.

No one outside the vendors knows exactly how the pattern data they collect is being used — the terms are built so the question cannot be audited from the outside. But I have been reading and writing software contracts for more than 20 years, and I know one thing cold: companies do not pay lawyers to flip defaults, gate opt-outs behind paid tiers, and take formal title to an asset that is worthless to them. This is not housekeeping. It is acquisition. The de-identified reasoning trace is training-grade raw material, the vendors have structured their terms to keep it, and the burden of proving otherwise now sits with them — not with you.

The court record shows what those protections are worth under pressure. In the OpenAI copyright MDL, "de-identified" was credited as the safeguard — and 20 million de-identified consumer conversations were produced to opposing counsel anyway, with deleted chats retained by court order against the user's explicit instruction. The order covers consumer accounts, not enterprise — and it proves retention and disclosure, not training. Hold that boundary; it is what makes the exhibit unbreakable. And it answered the only question that matters for risk planning: when the pattern exists and someone with power wants it, "de-identified" is a label, not a wall.

Keep the scopes straight, because precision is what makes this letter hard to dismiss. The consumer policies show the mechanism in the open; the product-level documentation shows the pattern persisting at the strongest setting; the commercial term adds ownership title. Different scopes, one logic: de-identify, then keep. The bar cuts both ways — the one confirmed case of a vendor routing a derived signal into training is Anthropic's feedback carve-out, where a thumbs-up rating becomes training-eligible with multi-year retention. One exhibit, not an industry-wide verdict. Every vendor gets the same standard.

Five questions for your legal and risk teams — this week

On air, Karp posed two questions buyers should put to every AI vendor: "Are you keeping the data? Are you going to enter our business?" Right instinct, wrong altitude — a privacy page can answer Karp's questions and still keep the pattern. These five get answered in the contract, or not at all.

1. **What does "Customer Data" actually cover?** Every protection you negotiated — opt-out, deletion, Zero Data Retention — attaches only to what the contract defines as Customer Data. Ask whether that definition includes derived data: embeddings, telemetry, usage patterns, de-identified traces. In the terms in the exhibit, it expressly excludes them. If your definition does too, your protections stop before the asset.

2. **Who owns the usage record?** Cognition's current terms claim it in writing: "Cognition owns all right, title, and interest in and to the Usage Data." Put the same question to your vendor. If the vendor holds title to the

record of how your people work, your opt-out cannot reach it — you cannot opt out of someone else's property.

3. **Does our deletion right survive de-identification?** OpenAI's consumer terms answer no, in writing: content already "de-identified and disassociated from your account" is exempt from your deletion request. Ask whether your delete reaches the de-identified copy.

4. **What happens to "de-identified" data under subpoena?** A federal court ordered roughly 20 million de-identified consumer chats produced to opposing counsel — including chats users had deleted (OpenAI MDL, S.D.N.Y., Nov. 7, 2025). Have counsel read that order, then ask where your data sits in the same scenario.

5. **Will the vendor sign a verifiable non-training guarantee with audit rights covering derived data — not just Customer Data?** This is the one that separates a privacy page from a commitment. A yes costs the vendor nothing if the trace is worthless to them. Watch what a no tells you.

A vendor who objects to this letter can refute it with a signature, not a statement. Amend the definition. Extend the deletion right. Grant the audit. And hold the sovereignty sellers to the same bar — Palantir's isolation and portability claims are vendor-stated, not independently audited. The forward question for your procurement team is whether control over derived data becomes a measured variable in every AI contract or stays a rhetorical weapon on television. Karp's episode made the question loud. The exhibit above makes it answerable.

The durable answer is architectural, not contractual: own the loop, keep the trace in your own storage, on infrastructure you control. The report that follows scores which deployment patterns make that possible today.

The pattern of how your organization thinks is either an asset you own or an asset you donate. The contracts have already decided. The only question is whether you have read them.

Jim Calhoun

Founder & Managing Director, The Grove Foundation, Inc.

Letter source note: Karp quotes — CNBC Squawk Box, July 1, 2026 (spoken; CNBC video and write-up, Mediaite transcript). Palantir manifesto — @PalantirTech on X, June 30, 2026 (written). Palantir–NVIDIA Nemotron announcement — BusinessWire, June 29, 2026. The spoken and written records are cited separately; they are different documents. Contract exhibits — Works Cited [14]–[19] and the vendors' own published archives. This letter is commentary; the scored analysis begins with the Executive Summary.

Executive Summary

The fight has moved up a layer. It is no longer about who owns the best model. It is about who owns the *worked example* — the recorded trace of how a real organization solves a real problem, step by step, with a checkable result. That trace is the scarce fuel now. And this quarter, the board shows it moving away from the big API vendors and toward open-weight patterns a company can run and keep for itself.

Three terms carry this report, so we define them here in plain words. A **trace** is the recorded sequence of steps an AI system took to solve a problem, together with the result. A **harness** is the software wrapper around a model — the layer that runs tools, watches the work, and records it. **Custody** means who holds and controls that record. Everything below turns on those three words.

Every one of this quarter's five strongest patterns is a Sovereign open-weight pattern — weights a company can download, run on its own hardware, and keep. Five patterns cleared the Approaching-Critical line ($\Lambda \geq 0.03$): DeepSeek V4 (0.0502), Mistral (0.0440), Gemma 4 (0.0376), GLM-5.2 (0.0366), and Qwen 3.6 (0.0366). The strongest closed pattern, Anthropic Claude (0.0058), sits a full tier below in Sub-Critical — five times short of the line. Clearing the line is earned on the equation. It is not handed out by how we count the board.

Five findings carry the quarter.

1. All five Approaching-Critical seats are Sovereign open-weight. The closed API vendors sit far below them. This is a scored result, not a counting trick — each leader clears the line on its own numbers.

2. A moat is not the same as spread. The incumbents' Q2 counter-moves — OpenAI's own silicon, Claude captured inside Slack, OpenAI's new ad tools — make their market position stickier. But under Λ they *lower* Spreadability and deepen dependence. None of them crosses a tier, even when we tilt the math in their favor.

3. The apex is now eating its own moat. Google's own open model, Gemma 4, clears into Approaching Critical (0.0376) on the same permissive Apache 2.0 license that carries Mistral to 0.0440 — even though Google also sells a closed, premium Gemini. Scored correctly, that split-the-difference strategy shows up as an incentive drag (β_{ideo}), not as friction in the asset itself (Fc). It costs Gemma the rank but no longer the tier. An apex vendor can no longer ship permissive open weights without arming a rival against its own closed product. The license and the friction ruling are both locked on the record (panel, July 1, 2026).

4. Concentration is breaking up. The concentration index (HHI) fell from 4,118 to 1,705 — down 59%, across the line the Justice Department treats as highly concentrated. On a like-for-like basis, the Sovereign cluster's share of structural weight rose from 83.8% to 91.6%. We headline the like-for-like figure and footnote the mechanical 96.3%, because part of the mechanical jump is just counting one row as four.

5. The prize has moved up a layer. The asset worth owning is custody of the reasoning trace, manufactured by the agent harness. You can prove that transfer from vendor terms today. What you cannot yet prove is active bulk training on the de-identified trace — so we say custody, and we stop there. That line is not timidity. It is what makes the claim hold.

Methodology: The Λ Framework (V2.0)

This report measures **structural propagation** — how freely a pattern can spread without a central owner pushing it. It does not measure revenue, brand, or product quality. The inputs are judgment calls on an ordinal scale; the model's job is to make those judgments visible and arguable, not to hide them.

The equation.

$$\Lambda = (S \times R \times V) / (1 + (\beta \cdot Fc)^2)$$

Λ (lambda) is the propagation score. Three factors help a pattern spread (S, R, V). Two resist it (Fc, β), and they sit in a squared denominator. That squared term is the whole point: cut friction or incentive-resistance in half in a high-resistance regime, and Λ roughly quadruples. Improve a spread factor, and Λ moves only in a straight line. Small moves in Fc and β are load-bearing; small moves in S, R, and V are not.

Continuity with the March publication. The equation above is the same one published in March 2026 (the-grove.ai/alerts/ai-deployment-pattern-benchmark, "Methodology 2.0"): the power-law form, the Validation term, and β as a *geometric mean* of its financial, regulatory, and ideological parts. An earlier internal draft used an exponential form with a min()-based β; that form was corrected before the March release and appears in no published score. The geometric mean matters: one strong incentive dimension cannot paper over a weak one, which is exactly right for judging whether a pattern survives on its own. Because the method is unchanged, this quarter's Λ values are directly comparable to the published March baseline — the deltas in this report are real movement, not a change of ruler. What is new in Q2 is granularity and coverage: the combined open-weight row is split into individually scored leaders, Gemma 4 joins the board, and the quarter's scoring rulings are recorded in the sensitivity section. One naming update travels with that: the clusters the March report called the *Sovereignty Profile* and the *Dependency Profile* appear here as **Sovereign** and **Concentrated** — same clusters, shorter names.

The five variables.

Var	Range	Name	What it measures	How it aggregates
S	0–1	Spreadability	How freely the pattern is copied without a central owner — license, replication cost, network reach	mean(S1 Legal & IP, S2 Replication cost, S3 Network topography)
R	0–1	Rail Compatibility	How easily it runs on what a company already has	mean(R1 Interoperability, R2 Legacy compat, R3 Ecosystem support)
V	0.2–1.0	Validation	Deployment discount — 0.2 pre-publication, 1.0 enterprise-validated	applied directly
Fc	0–10	Cognitive Friction	Mental and operational cost to adopt (high = bad)	mean(Fc1 Mental demand, Fc2 Paradigm switch, Fc3 Uncertainty/risk)
β	0.1–10	Exogenous Incentive	Outside forces that push or drag adoption (low = strong push)	geometric mean(β_fin, β_reg, β_ideo)

The zones.

Tier	Λ range	Meaning
Critical Mass	≥ 0.10	Self-propagating. Survives with no subsidy.
Approaching Critical	0.03 – 0.099	Momentum building; not yet durable. Grove's intervention band.
Sub-Critical	0.005 – 0.029	Works, but needs constant outside support.
Structurally Inert	< 0.005	Won't spread without coercion or monopoly inertia.

Two rules that shape the board. First, we do not score "open weight" as one blob. We score the surging leaders one by one — DeepSeek V4, GLM-5.2, and the quarter's third mover, Qwen 3.6 — and bucket the long tail by jurisdiction and license. The third seat is earned by surge each quarter, not held by incumbency. Second, every load-bearing number ties to a primary source before it enters a published score; aggregator

figures are flagged, then confirmed or cut.

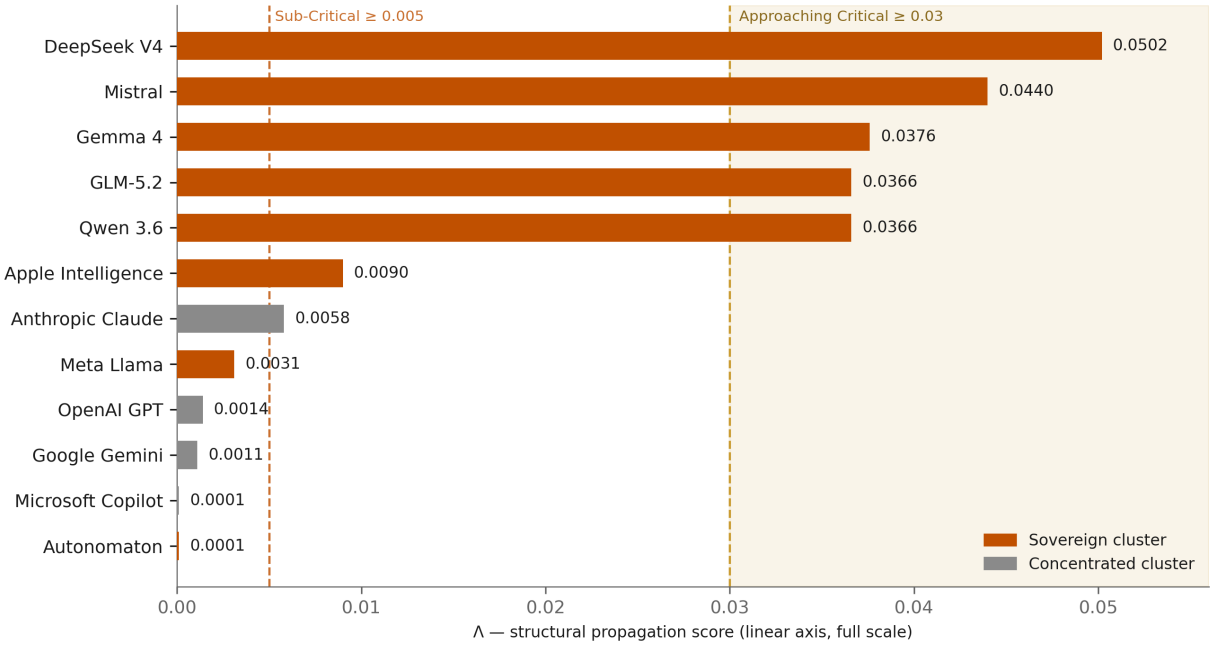
Conflict-of-interest disclosure. Grove's own pattern, the Autonomaton, is scored by the same method and lands at the floor (Λ = 0.0001, Structurally Inert, V = 0.2). We keep this public. It earns no special treatment.

The Board at a Glance

#	Pattern	Category	S	R	V	Fc	β	Λ	Tier	Cluster
1	DeepSeek V4	Open Weight (Chinese)	0.800	0.700	1.0	6.0	0.531	**0.0502**	Approaching Critical	Sovereign
2	Mistral	Open Weight (Western)	0.800	0.600	1.0	5.0	0.630	**0.0440**	Approaching Critical	Sovereign
3	Gemma 4	Open Weight (Western)	0.833	0.767	1.0	4.0	1.000	**0.0376**	Approaching Critical	Sovereign
4	GLM-5.2	Open Weight (Chinese)	0.800	0.700	1.0	6.0	0.630	**0.0366**	Approaching Critical	Sovereign
5	Qwen 3.6	Open Weight (Chinese)	0.800	0.700	1.0	6.0	0.630	**0.0366**	Approaching Critical	Sovereign
6	Apple Intelligence	On-Device	0.200	0.567	1.0	2.0	1.710	0.0090	Sub-Critical	Sovereign
7	Anthropic Claude	Centralized API	0.467	0.800	1.0	5.0	1.587	0.0058	Sub-Critical	Concentrated
8	Meta Llama	Open Weight (US)	0.600	0.700	1.0	5.0	2.321	0.0031	Structurally Inert	Sovereign
9	OpenAI GPT	Centralized API	0.467	0.933	1.0	6.0	2.924	0.0014	Structurally Inert	Concentrated
10	Google Gemini	Platform Bundle	0.300	0.800	1.0	5.0	2.924	0.0011	Structurally Inert	Concentrated
11	Microsoft Copilot	Platform Bundle	0.300	0.800	1.0	8.0	5.000	0.0001	Structurally Inert	Concentrated
12	Autonomaton	Sovereign Open	0.933	0.833	0.2	6.0	6.300	0.0001	Structurally Inert	Sovereign

CHART 1 — THE Q2 2026 SCORED BOARD

All five Approaching-Critical seats are Sovereign open-weight. Λ = (S × R × V) / (1 + (β·Fc)²)



THE GROVE FOUNDATION — Λ STANDINGS Q2 2026

Every Λ on this board reproduces from its sub-scores through the equation — verified June 29, 2026, with the amended Gemma 4 row re-verified July 1, 2026. Two clusters organize the board: **Concentrated** (high R, low S — the closed API and platform vendors) and **Sovereign** (high S — the open-weight leaders, Gemma, Apple, Llama, and the Autonomaton).

Sub-score note. For legacy patterns (rows 6–12), the S1/S2/S3, R1/R2/R3, Fc1/Fc2/Fc3, and β sub-components carry from the published March benchmark's 96-source base, with documented Q2 deltas. The four open-weight leaders (rows 1, 2, 4, 5) derive from March's combined "Mistral/DeepSeek" anchor row plus deltas sourced this quarter; their sub-component lines below show that derivation. Gemma 4 (row 3) is a genuine new row, anchored to Mistral with Gemma-specific deltas; its scores are banked at the variable level under the July 1 amendment.

Rounding note. The March publication rounded sub-scores to two decimals, which printed Anthropic Claude at 0.0059 and Microsoft Copilot at 0.0002. This report carries three-decimal sub-scores, which compute 0.0058 and 0.0001 from the identical inputs. That is a precision convention, not movement; both rows are unchanged this quarter.

Pattern-by-Pattern

PATTERN 1

DeepSeek V4 · Open Weight (Chinese) · Λ = 0.0502 · Approaching Critical

Companies self-host DeepSeek's MIT-licensed weights on their own or a Western cloud's hardware. It tops the board this quarter, and it does so on incentive, not friction.

- **S = 0.800.** S1 0.8 (permissive MIT license), S2 0.8 (low replication cost), S3 0.8 (broad global distribution) — carried from the 1.0 anchor row, reconfirmed this quarter.
- **R = 0.700.** R1 0.8 (standard open stack — GGUF, vLLM, Hugging Face), R2 0.5 (custom adapters for legacy integration), R3 0.8 — up from the anchor's 0.5 as the quarter's demand surge routes through ecosystem support, per the scoring ruling [2].
- **V = 1.0.** Deployed at scale (quarter's V = 1.0-for-deployed rule).
- **Fc = 6.0.** Fc1 5, Fc2 5, Fc3 8 — the 8 carries the China data-jurisdiction friction: the residual adopter hesitation that survives even when the weights run on a Western host.
- **β = 0.531.** The lowest resistance on the board. A verified ≥10× output-price collapse (~12–53× cheaper than frontier APIs) trips the β_fin floor of 0.3 [1]; β_reg ≈ 0.5 because the National Intelligence Law attaches to cloud APIs, not to self-hosted weights; β_ideo ≈ 1.0. Geometric mean = 0.531.
- **Why it leads.** Same S, R, and Fc as the other Chinese leaders, but a lower β — the price collapse is what lifts it above GLM and Qwen. β does the work here, not Fc.
- **Vulnerability / Ratchet / Displacement.** Western procurement bans are the risk. Each self-hosted deployment builds in-house muscle that ratchets away from vendors. It loses ground only if a Western-origin model matches its price curve.

PATTERN 2

Mistral · Open Weight (Western) · Λ = 0.0440 · Approaching Critical

Europe's permissive open model, self-hosted. Splitting the old combined row exists to make this pattern visible: the Western permissive alternative, scored on its own numbers.

- **S = 0.800.** S1 0.8 (Apache 2.0), S2 0.8 (low replication cost), S3 0.8 (native EU/US reach) — anchor values, reconfirmed.
- **R = 0.600.** R1 0.8, R2 0.5, R3 0.5 — the anchor's rails, unchanged. Fewer default integrations than the Chinese leaders; no comparable demand-surge lift to R3.
- **V = 1.0.** Deployed.
- **Fc = 5.0.** Fc1 5, Fc2 5, Fc3 5 — a full point below the Chinese leaders because Fc3 sheds the data-jurisdiction risk. Since Fc sits in the squared denominator, that one point is what lifts Mistral above GLM and Qwen *despite* its lower R (0.60 vs. 0.70).
- **β = 0.630.** β_fin ≈ 0.5 (strong cost case, but no verified 10× price floor), β_reg ≈ 0.5 (Western tailwind — EU sovereign-AI funding), β_ideo ≈ 1.0. Geometric mean = 0.630.
- **Read this precisely.** DeepSeek still beats Mistral — but on β, not Fc. The two stories do not collide: Fc explains only the Mistral-over-GLM/Qwen ordering; β explains DeepSeek over everyone.
- **Vulnerability / Ratchet / Displacement.** Thin default-ecosystem support is the soft spot. EU sovereign-AI funding is a tailwind. It is the fallback if Chinese models get banned outright.

PATTERN 3

Gemma 4 · Open Weight (Western) · Λ = 0.0376 · Approaching Critical

Google's own open model — and this quarter's sharpest exhibit. It clears the line even though Google also protects a closed, premium Gemini. The March publication called this in advance: its April note said Gemma 4's Apache-licensed release "materially alters the structural mechanics of the landscape" and promised the full scoring in this update. Here it is. The license and the friction call are both locked on the record: Apache 2.0 confirmed against the primary [3], and Fc ruled to 4.0 by the panel on July 1, 2026.

- **S = 0.833.** High spread on the confirmed Apache 2.0 grant; the efficient 2–31B model line and per-layer embeddings keep replication cost low [4]; distribution is broad — top-trending on Hugging Face, on-device to an iPhone 17 Pro.
- **R = 0.767.** Wide runtime fit — GGUF/llama.cpp, JAX, PyTorch, on-device runtimes, and Vertex.
- **V = 1.0.** Deployed; roughly 2 million downloads in week one [6].
- **Fc = 4.0 (ruled July 1, 2026; scored 5.0).** The panel moved Fc down to record only the asset's low execution friction — the efficient 2–31B line, on-device escape velocity (~40 tokens per second on an iPhone 17 Pro), native multi-framework parity with no secondary quantization work — and to hold Google's dual-track *strategy* in β_ideo instead. Holding Fc at 5.0 would have charged the same fact twice, once in β and once in Fc. Mistral holds at 5.0 because it ships no comparable on-device efficiency profile; that asymmetry is earned, not manufactured.
- **β = 1.000.** β_fin ≈ 1.0, β_reg ≈ 0.5, β_ideo ≈ 2.0 — the dual-track hedge: open Gemma alongside a closed, premium Gemini, confirmed this quarter [5]. Geometric mean = 1.000. The dual-track drag costs Gemma the rank (0.0376 < Mistral's 0.0440) but, at Fc = 4.0, no longer the tier.
- **Why it matters.** An incumbent can no longer ship permissive open weights without helping a competitor beat its own closed product. That is the apex-cannibalization finding in one row. A pure-play sponsor with the same asset (β ≈ 0.63) would score near 0.059 — the incentive drag is real and measured; it suppresses Gemma without sinking it.
- **Sub-score note.** Gemma's variable scores are banked at the variable level under the July 1 amendment; the three-way sub-splits are not separately banked, and we do not invent them here.
- **Vulnerability / Ratchet / Displacement.** The residual risk is the friction asymmetry itself: Fc = 4.0 rests on the on-device profile, and if Mistral ships a comparable profile next quarter, the asymmetry gets re-audited symmetrically. The ratchet: every Gemma deployment normalizes on-device open weights. Displacement comes only if Google reverses course and closes the line.

PATTERN 4

GLM-5.2 · Open Weight (Chinese) · Λ = 0.0366 · Approaching Critical

A Chinese MIT-licensed leader — 744B/40B mixture-of-experts, 1M-token context, Huawei Ascend support [7] — self-hosted.

- **S = 0.800 · R = 0.700 · V = 1.0 · Fc = 6.0.** Same derivation as DeepSeek: anchor sub-scores with the R3 demand lift and the Fc3 jurisdiction carry.
- **β = 0.630.** β_fin ≈ 0.5 — a strong cost case but no verified 10× price floor, unlike DeepSeek — β_reg ≈ 0.5, β_ideo ≈ 1.0. That single β_fin step is the whole reason it trails DeepSeek despite identical S, R, and Fc.
- **Vulnerability / Ratchet / Displacement.** Jurisdiction friction and Western bans are the risk. A strong release cadence keeps it in the top group. It holds unless the Chinese cohort reshuffles next quarter.

PATTERN 5

Qwen 3.6 · Open Weight (Chinese) · Λ = 0.0366 · Approaching Critical

Holds the quarter's earned third-leader seat, taken on demand surge — not incumbency.

- **S = 0.800 · R = 0.700 · V = 1.0 · Fc = 6.0 · β = 0.630.** Matches GLM-5.2, same derivation.
- **Seat basis.** Reconfirmed to primaries: OpenRouter's 100-trillion-token study names Qwen a lead driver of the Chinese open-source token surge [2], and Hugging Face download data — 700 million cumulative, past Llama — puts it ahead of MiniMax and Kimi [8, 9]. The precise monthly share percentages stay out of this report — a July 6, 2026 pre-publish check against OpenRouter's own ranking page could not confirm them (the page keeps no monthly history); the seat holds on the primaries.
- **Rule A carry-forward.** The seat is earned each quarter. Re-contest in Q3 against Qwen3.7 Max, MiMo-V2-Pro, and Hy3.
- **Vulnerability / Ratchet / Displacement.** Same jurisdiction profile as GLM. The seat itself is the main risk — a faster mover can take it next quarter.

PATTERN 6

Apple Intelligence · On-Device · Λ = 0.0090 · Sub-Critical

AI on Apple Silicon, data kept on the device. Lowest friction on the board — and the narrowest reach.

- **S = 0.200.** S1 0.2 (closed models, proprietary silicon), S2 0.2 (cannot be replicated off-Apple), S3 0.2 (air-gapped to non-Apple endpoints). Low by design.
- **R = 0.567.** R1 0.5 (custom middleware), R2 1.0 (seamless native OS fit), R3 0.2 (single vendor, no cross-platform rails).
- **V = 1.0 · Fc = 2.0.** Fc1/Fc2/Fc3 = 2/2/2 — invisible on-device models, no prompt work, near-zero privacy risk. Lowest friction in the landscape.
- **β = 1.710.** β_fin 5.0 (free inference, but gated behind hardware upgrades), β_reg 1.0 (stateless private compute answers cross-border rules), β_ideo 1.0 (privacy resonates). Geometric mean = 1.710.
- **Why it's stuck.** Friction is almost zero, but spread is capped by hardware. It cannot become a cross-company cognitive layer. Great floor, low ceiling.

PATTERN 7

Anthropic Claude · Centralized API · Λ = 0.0058 · Sub-Critical

The strongest closed pattern — and it still sits a full tier below the Approaching-Critical line.

- **S = 0.467.** S1 0.2 (restrictive proprietary terms), S2 0.2 (frontier-scale capital to replicate), S3 1.0 (universal API reach plus Bedrock and Vertex).
- **R = 0.800.** R1/R2/R3 = 0.8/0.8/0.8 — strong Bedrock integration, regulated-industry footing.
- **V = 1.0 · Fc = 5.0.** Fc1/Fc2/Fc3 = 5/5/5 — trust tooling (ISO 42001, zero-retention options on Bedrock) lowers perceived risk, but the rented, centralized shape remains.
- **β = 1.587.** β_fin 0.8 (prompt caching, batch discounts), β_reg 1.0 (ISO 42001 as a compliance defense), β_ideo 5.0 (closed weights leave the open-source community cold). Geometric mean = 1.587.
- **Counter-moves are sensitivity-only.** The SpaceX/Colossus compute expansion [10] and Claude-in-Slack capture strengthen the moat but are held as scenarios, not banked sub-score changes. Neither crosses a tier. **Custody note:** Anthropic is also home to the one banked ACTIVELY-TRAINED exhibit — the feedback carve-out — covered in the synthesis. That exhibit is banked and it stays. It is one

vendor's clause, not an industry verdict.

PATTERN 8

Meta Llama · Open Weight (US) · Λ = 0.0031 · Structurally Inert

The one open-weight pattern that does not clear its own weight. The March publication scored Llama at 0.0104, Sub-Critical — and flagged both warning signals in print: EU geofencing, and an expected score suppression from the "Avocado" closed-source pivot. This quarter banks both on primaries. The restatement is visible and intentional: 0.0104 → 0.0031, one tier down.

- **S = 0.600.** S1 dropped 0.5 → 0.2 (Llama 4's EU geofencing limits free copying); S2 0.8 and S3 0.8 hold.
- **R = 0.700.** R1 0.8, R2 0.5, R3 0.8 — standard open-stack rails.
- **V = 1.0 · Fc = 5.0.** ML-Ops expertise to self-host; no jurisdiction friction.
- **β = 2.321.** β_ideo moved 1.0 → 5.0 on the confirmed closed-source ("Avocado/Muse") pivot; β_fin 0.5, β_reg 5.0. Geometric mean = 2.321 (March published 1.357, pre-pivot).
- **Tier precision.** 0.0031 is Structurally Inert — the band *below* Sub-Critical. Public copy that calls this a "Sub-Critical drop" understates it; the corrected wording ships with this report.
- **Baseline note.** The board's March baseline carries this restatement, so Llama reads "unchanged" in the quarter's delta accounting. The share disclosure in the counter-arguments section shows the movement both ways.

PATTERN 9

OpenAI GPT · Centralized API · Λ = 0.0014 · Structurally Inert

Owns the market conversation; scores near the floor on spread.

- **S = 0.467.** S1 0.2, S2 0.2, S3 1.0 — universal reach, total lock-in.
- **R = 0.933.** R1 1.0, R2 0.8, R3 1.0 — the default integration standard.
- **V = 1.0 · Fc = 6.0.** Fc1/Fc2/Fc3 = 5/5/8 — the 8 reflects real deprecation risk; the GPT-4o retirement forced unplanned migrations.
- **β = 2.924.** β_fin 1.0, β_reg 5.0, β_ideo 5.0. Geometric mean = 2.924.
- **Counter-moves are sensitivity-only.** Proprietary silicon and the new advertising infrastructure [11] harden the moat. Neither moves the tier — under Λ, both deepen dependence rather than spread.

PATTERN 10

Google Gemini · Platform Bundle · Λ = 0.0011 · Structurally Inert

Bundled into Workspace and GCP; anchored by data gravity, not by spread.

- **S = 0.300.** S1 0.2, S2 0.2, S3 0.5 — confined to the Google demographic.
- **R = 0.800.** R1/R2/R3 = 0.8/0.8/0.8 — strong inside its own ecosystem.
- **V = 1.0 · Fc = 5.0.** Fc1/Fc2/Fc3 = 5/5/5 — fragmented product branding adds needless navigation cost.
- **β = 2.924.** β_fin 1.0, β_reg 5.0, β_ideo 5.0. Geometric mean = 2.924.
- **The barbell.** Google now runs the industry's clearest dual-track: this closed bundle at 0.0011 and open Gemma 4 at 0.0376 — a 34× spread inside one company. The open asset propagates; the closed bundle anchors. Flagged as a Q3 watch-item, not promoted to an engine rule this quarter.

PATTERN 11

Microsoft Copilot · Platform Bundle · Λ = 0.0001 · Structurally Inert

The clearest case of market presence without structural spread.

- **S = 0.300.** S1 0.2, S2 0.2, S3 0.5 — data gravity confines it to the tenant.
- **R = 0.800.** R1/R2/R3 = 0.8/0.8/0.8 — deep inside Microsoft 365.
- **V = 1.0 · Fc = 8.0.** Fc1/Fc2/Fc3 = 8/8/8 — the highest friction on the board; low paid-seat conversion tells the story.
- **β = 5.000.** β_fin 5.0, β_reg 5.0, β_ideo 5.0 — ambiguous on every axis. Geometric mean = 5.000.

PATTERN 12

Autonomaton · Sovereign Open · Λ = 0.0001 · Structurally Inert (COI disclosure)

Grove's own model-agnostic architecture. Highest theoretical base on the board, zero current propagation. Scored by the same method as everyone else.

- **S = 0.933 · R = 0.833.** The strongest base multiplier here (S × R = 0.78).
- **V = 0.2.** Pre-publication — the discount that pins Λ to the floor.
- **Fc = 6.0.** Fc1/Fc2/Fc3 = 5/8/5 — the paradigm-switch cost (the 8) is the biggest single friction barrier in the landscape.
- **β = 6.300.** β_fin 5.0, β_reg 10.0, β_ideo 5.0. Geometric mean = 6.300 — corrected from the legacy min()-based 5.000; 6.300 is the only value that reproduces Λ = 0.0001.
- **Honest read.** Huge base, no outside push, high switching cost, pre-publication discount. It needs an external shock to move. We disclose this and claim nothing extra.

Strategic Synthesis

1. Why now: the data wall

Frontier AI is no longer short on text. It is short on *verified reasoning traces* — worked examples that carry a checkable reward. Three facts set up the scarcity, and none of them depends on any custody or training claim.

Capability on long-horizon software tasks is doubling about every 89 days (METR Time Horizon 1.1) [108]. The systems meant to supervise that capability are not compounding as fast, so a gap opens. The best fuel for reinforcement learning is the outcome-verified worked example — not the chat log, and not the raw web dump. And the escape hatch — "just generate more data" — is closing: training on recursively generated synthetic data degrades models [12], and self-consuming generative loops go "MAD" [13]. Text is exhausted and synthetic text rots. So the scarce input is real reasoning traces.

That is the "why now." Without it, this report is timeless opinion. With it, the custody finding below is a Q2 statement about where a newly scarce asset is piling up.

2. The harness makes the asset

If traces are the fuel, the agent harness is the machine that manufactures them at scale. The harness — the orchestration and tool-governance layer wrapped around a model — is where a reasoning trace gets produced, watched, and, where the architecture allows, kept.

Keep two claims apart. The strong claim is that a harness *retains replayable traces*. That is what matters for custody, and we prove it from vendor terms below — not from capability numbers. The weak claim is that harnesses make the product better. That is real but not load-bearing, and the public evidence for it is thin: vendor-blog results and small-n preprints. We treat that class of evidence as illustrative color, never as a scoring input. The argument rests on the black-letter terms.

3. The custody inversion — read the contracts, not the press release

Here is the part the market gets backwards, and it is the part a CIO needs to hear plainly.

The question is **not** "does the vendor train on my data?" The question is "has custody of my pattern already transferred?" Custody is written into the terms today, and you can check it yourself.

De-identification, aggregation, Zero Data Retention, and "we discard the plaintext" are sold as protections. Read against the contracts, they are the *mechanism*. They are how a vendor strips the regulated, identifiable datapoint off your interaction — and lawfully keeps the interaction pattern, which is the asset. An embedding is your pattern with the plaintext peeled away. "We delete the plaintext and keep the embeddings" is not a safeguard. It is the harvest.

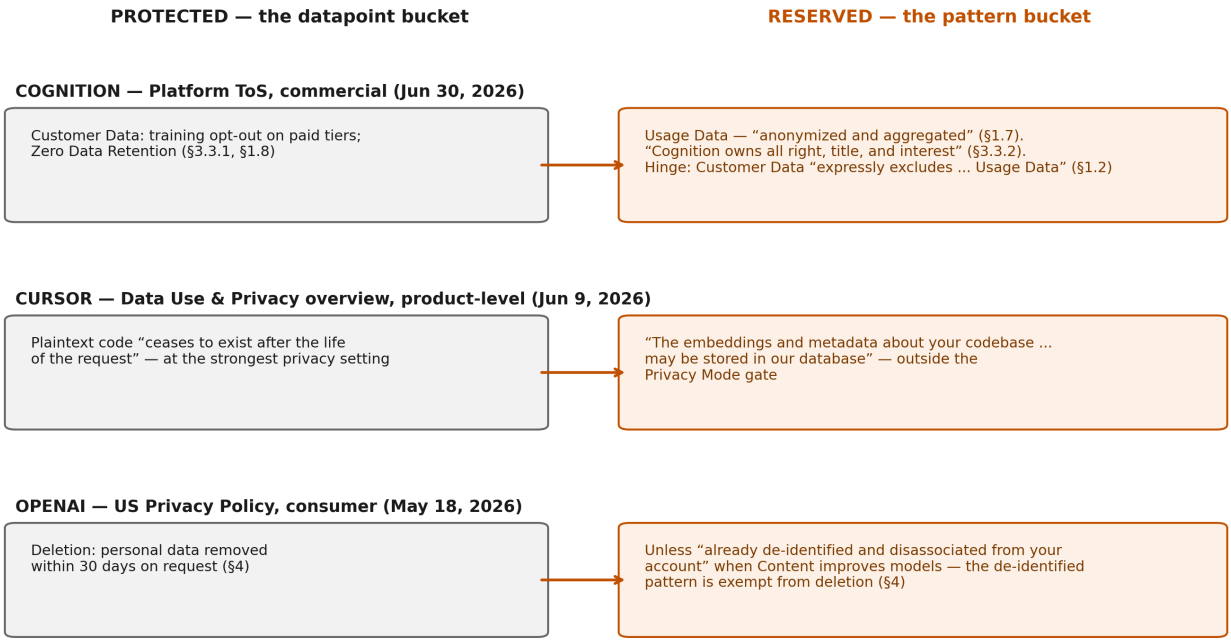
Three vendors, one architecture, all in their own words:

- **Cognition owns the reserved bucket outright.** "Cognition owns all right, title, and interest in and to the Usage Data" (§3.3.2), where Usage Data is "the anonymized and aggregated data regarding the manner in which you or your Authorized Users interact with the Services" (§1.7). And the hinge: Customer Data "expressly excludes... Usage Data" (§1.2). So your opt-out and your Zero Data Retention (§3.3.1, §1.8) attach only to Customer Data — and stop exactly where the reserved, company-owned bucket begins. No setting, on any tier, reaches the titled pattern. (Cognition Platform ToS, effective June 30, 2026 — the commercial product term.) [14]
- **Cursor keeps the pattern at its strongest privacy setting.** "All plaintext code for computing embeddings ceases to exist after the life of the request. The embeddings and metadata about your codebase... may be stored in our database." That clause sits in the product-level data-use documentation, outside the Privacy Mode gate. [15]

- **OpenAI exempts the de-identified pattern from deletion.** Content that has "already been de-identified and disassociated from your account" is kept past a deletion request "when you allow us to use your Content to improve our models" (OpenAI US Privacy Policy §4, May 18, 2026). [16]

CHART 2 — THE CARVE-OUT ARCHITECTURE: ONE LOGIC, THREE VENDORS

The protections attach to the datapoint bucket. The pattern bucket sits outside them, reserved. Quotes are operative clauses.



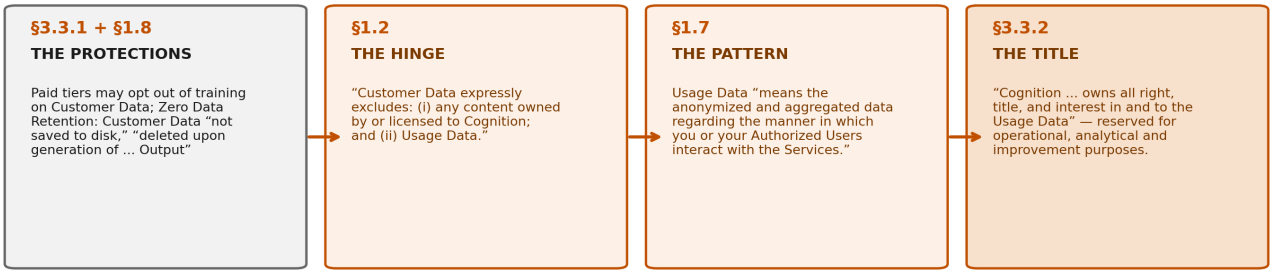
AXIS TAG — STRUCTURALLY-CAPTURED (banked): capture, reservation and ownership of the de-identified pattern are black-letter.
ACTIVELY-TRAINED (held): bulk training on that pattern is NOT claimed; the one banked exhibit is the Anthropic feedback carve-out.

THE GROVE FOUNDATION — Λ STANDINGS Q2 2026 CC BY 4.0 · Source: verification-log-q2-2026 (carve-out catalog, quotes verbatim)

This is not a hypothetical about what a vendor *might* do. It is what the signed terms *say*. The compelling truth for an enterprise buyer is simple: the privacy language you were shown protects the identifiable record, not the pattern of how your people think and work — and the pattern is what the vendor keeps. (Axis: STRUCTURALLY-CAPTURED / reserved-use. The verbatim Cognition side-by-side is reproduced in Chart 5 so a hostile reader can check every word.)

CHART 5 — THE COGNITION CHAIN: FOUR CLAUSES, NO GAP

Cognition Platform Terms of Service, effective June 30, 2026. Read in sequence, the customer's protections stop exactly where the owned bucket begins.



The chain closes with no gap: opt-out and ZDR reach only Customer Data (§3.3.1, §1.8) · Customer Data expressly excludes Usage Data (§1.2) · Usage Data is the anonymized interaction record (§1.7) · Cognition owns it outright (§3.3.2).

NO SETTING, ON ANY TIER, REACHES THE TITLED PATTERN.

THE GROVE FOUNDATION — Λ STANDINGS Q2 2026

4. The honest counter — and where we stop

The strongest objection is the pledge every vendor makes: Zero Data Retention and "we don't train on your data" mean the trace stays yours. We engage it head-on, and we open with the case that shows what "de-identified" is worth under pressure.

In *In re: OpenAI, Inc. Copyright Infringement Litigation* (MDL, S.D.N.Y., No. 25-md-3143), the court ordered OpenAI to hand roughly 20 million de-identified consumer ChatGPT logs to an opposing party in discovery (Magistrate Judge Wang, Nov. 7, 2025; affirmed by District Judge Stein) [17]. A separate order forced retention of data — including chats users had already deleted — against a base the court itself called "tens of billions" of logs. De-identification was credited as a safeguard, and 20 million "de-identified" conversations were produced anyway, kept past deletion. Two guardrails travel with this exhibit: it is **consumer-scope only** (the order excludes Enterprise, Edu, Business, and API), and it proves *retention and disclosure, not training*. It stays on the STRUCTURALLY-CAPTURED axis; it does not touch the ACTIVELY-TRAINED line.

The scope layering must be stated plainly, or a critic calls it conflation. The consumer policies — OpenAI's privacy policy, and Cursor's consumer-scope privacy policy with its matching de-identify-to-improve clause [18] — show the *mechanism* in the open. Cursor's product-level data-use documentation shows the pattern persisting at the strongest setting. The commercial term (Cognition) gives that same mechanism *ownership title*. Different scopes, one logic: de-identify, then keep.

And here is where we stop, on purpose. Active bulk training on the de-identified reasoning pattern is **not proven**. Vendors confirm training on user *Content* at consumer or default scope, opt-out available — but Content is not the de-identified trace. The only banked ACTIVELY-TRAINED exhibit is the Anthropic feedback carve-out: a thumbs up/down signal routed to multi-year retention and made training-eligible [19]. The evidentiary bar is symmetric, Anthropic included — one confirmed exhibit does not license a claim about the whole industry, and the ownership reservations above are not allowed to masquerade as proof of training. We say custody, because custody is what the terms prove. That restraint is the reason the custody claim cannot be knocked down.

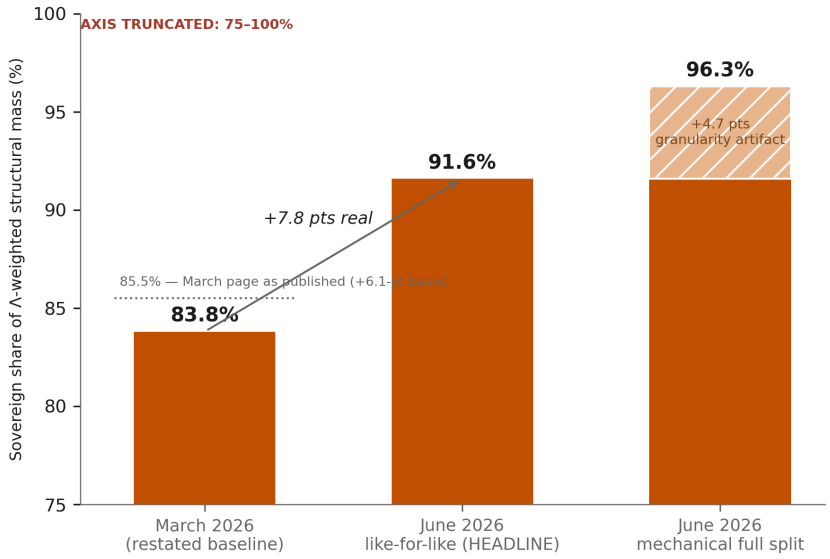
5. The way out: own the loop

If the trace is the asset, the defensible move is an architecture where the trace never leaves. Export your own trace logs. Run reinforcement learning on a corpus you own. The pieces already exist: harnesses that write traces as files into the customer's own storage, the "token capital" framing that treats accumulated interaction as an owned asset rather than a rented one, and models that run on your own hardware — which answer the custody question at the architecture level instead of in the fine print.

That is the Autonomaton's argument, made without special pleading, and it ties back to the board. On a like-for-like basis, the Sovereign cluster's share of structural weight rose from 83.8% to 91.6% — its first decisive break above the mid-80s — while HHI fell from 4,118 to 1,705. The mechanism is not counting: every open-weight leader now clears the level the old combined row held, and Google's own open model clears the line despite a dual-track incentive. The loop-ownership story lands with a live example of an apex vendor unable to stop its own open weights from spreading.

CHART 3 — SOVEREIGN STRUCTURAL SHARE: REAL SURGE vs. COUNTING ARTIFACT

Like-for-like basis holds the row count constant. Headline: 83.8% → 91.6%. Mechanical 96.3% footnoted.



Reading the split

The naive delta is +12.5 pts. Hold the Rule A row count constant and +7.8 pts of real structure remains; the other +4.7 pts is bookkeeping from counting one row as four. Gemma 4 is a genuine new row, so it lands in the real column.

Baseline basis: carries the June Llama restatement (0.0104 → 0.0031). Against the March page as published, the move reads 85.5% → 91.6% (+6.1 pts). Decisive above the mid-80s either way.

HHI: 4,118 → 1,705 (−59%)

Counter-Arguments, Engaged

(a) "The Sovereign surge is just disaggregation bookkeeping." True that splitting one combined row into four leaders mechanically lifts Sovereign share and cuts HHI. So we measured it instead of waving it off. Hold the row count constant — collapse the four leaders back into one representative row on a March-shaped roster — and Sovereign share still moves 83.8% → 91.6%, about +7.8 points of real structure. The full mechanical split reads 96.3%; the gap is granularity, and Chart 3 shades that +4.7-point artifact in plain sight. Every leader clears the old combined row's 0.0314 on its own, and Gemma 4 enters as a genuine new arrival (an April launch), not a split — so it lands in the real column. We publish the like-for-like number and footnote the mechanical one. We do not headline "Sovereign hit 96%."

One more disclosure, so the arithmetic is fully on the table. The March baseline in this report carries the Llama restatement (0.0104 → 0.0031, banked this quarter), which puts the baseline Sovereign share at 83.8%. Compute the same baseline from the March page as published — Llama still at 0.0104 — and it reads 85.5%, which makes the like-for-like move +6.1 points instead of +7.8. The surge breaks decisively above the mid-80s on either basis. We report the restated basis because it keeps Llama's fall out of the open-weight surge; we disclose the published basis so no one has to reverse-engineer it.

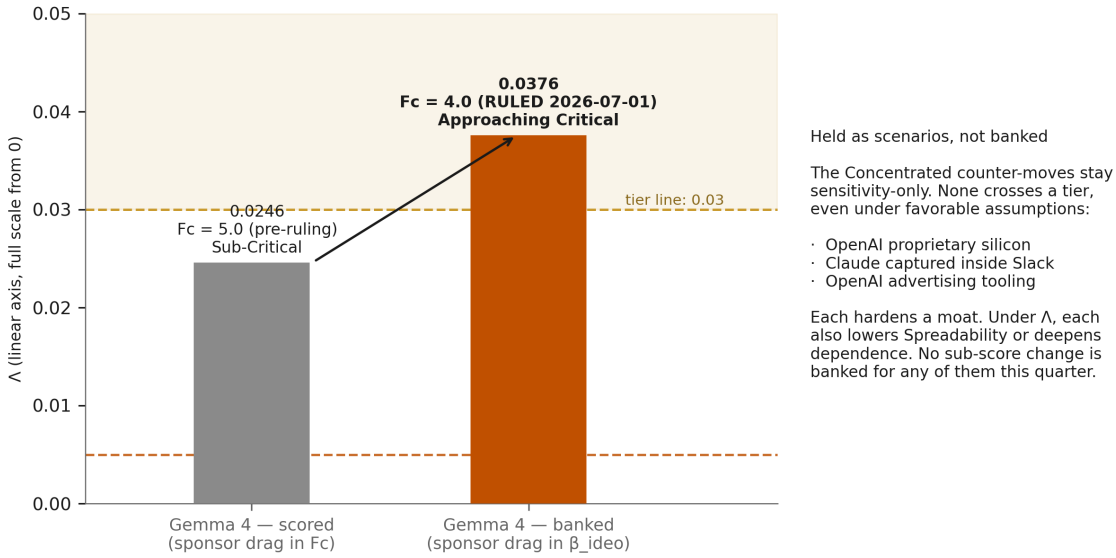
(b) "ZDR and 'we don't train on your data' keep the trace with the customer." Engaged in full in the synthesis. It turns on two facts. The pledge is scoped: opt-out and ZDR reach Customer Data, but Customer Data expressly excludes Usage Data, and Usage Data — the anonymized interaction record — is owned outright (Cognition §1.7 / §3.3.2), with matching de-identify-and-keep clauses at Cursor and OpenAI. De-identification strips the regulated datapoint and keeps the asset. The axis is custody, and the terms prove custody. Active training on the de-identified pattern is not asserted — the terms do not prove it, and they are built so no outsider can check. That is vendor opacity, not vendor innocence: a company is not cleared by the absence of proof it made unobtainable, and the one confirmed exhibit (Anthropic's feedback carve-out) is not generalized to the industry. These clauses are new instruments that cloud-era contract review was never built to catch; they warrant the same forensic reading this report gives them.

Sensitivity and Rulings (Locked This Quarter)

The board findings are single-variable-robust. The incumbents' counter-moves — OpenAI silicon, Claude-in-Slack, OpenAI ad tooling — are held as single-variable sensitivity scenarios; none crosses a tier even under favorable assumptions. The one scenario that crossed a tier line — Gemma 4's Fc — was ruled and banked at 4.0 on July 1, 2026, so Gemma is a scored crossing, not a pending scenario.

CHART 4 — SENSITIVITY: THE ONE CALL THAT CROSSED A TIER LINE

The panel ruled Gemma 4's Fc to 4.0 on July 1, 2026. Every other scenario on the board is tier-stable.



Held as scenarios, not banked

The Concentrated counter-moves stay sensitivity-only. None crosses a tier, even under favorable assumptions:

- OpenAI proprietary silicon
- Claude captured inside Slack
- OpenAI advertising tooling

Each hardens a moat. Under Λ, each also lowers Spreadability or deepens dependence. No sub-score change is banked for any of them this quarter.

THE GROVE FOUNDATION — Λ STANDINGS Q2 2026

Locked rulings, on the record:

- **Gemma 4 Fc = 4.0.** β_ideo carries the sponsor's dual-track strategic drag; Fc carries only the asset's developer and execution friction, which sits a full step below the Western-permissive baseline (efficient 2–31B line, on-device escape velocity, no secondary quantization work). Holding Fc at 5.0 double-counted the apex penalty already priced in β. Mistral holds at Fc = 5.0 because it ships no comparable on-device profile; if it does next quarter, the asymmetry gets re-audited symmetrically.
- **Variable placement is symmetric — the China call follows the same discipline.** The National Intelligence Law attaches to cloud API usage, not to weights self-hosted on a Western cloud, so β_reg ≈ 0.5 does not carry the China penalty for self-hosted deployment. What survives self-hosting is adopter hesitation — residual uncertainty and risk — and that is adoption friction, which lives in Fc3 (6.0 vs. Mistral's 5.0). Sponsor strategy in β; adoption friction in Fc. Same rule for Gemma and for the Chinese leaders, stated here so no one has to infer it.
- **β_fin floor = 0.3** for any model with a verified ≥10× output-price collapse against frontier APIs. Applies to DeepSeek V4 (verified ~12–53× range).
- **V = 1.0 for all deployed patterns.** The demand surge is routed through R3 and β_fin, not V.
- **Concentrated counter-moves stay sensitivity-only.** Promotion to banked sub-scores requires banked primary evidence, next quarter or later.

Works Cited

The published March 2026 benchmark ("The AI Market Runs on Subsidy, Not Structure," the-grove.ai/alerts/ai-deployment-pattern-benchmark, CC BY 4.0) carries the legacy sub-scores on a 96-source base; those sources are enumerated in the March benchmark dossier and appear here as entries 20–115, order preserved. This quarter adds 19 Rule-D-verified primaries (entries 1–19, report order; sibling documents and press confirmations are listed inside their primary's entry). Six aggregator-class claims from the research pull were re-anchored to the vendor primaries below or cut before scoring; the find-to-fix trail is preserved in the quarter's verification logs.

[1] DeepSeek API pricing documentation, April 2026. (Basis for the verified ~12–53× output-price collapse; β_fin floor ruling.)

[2] OpenRouter, "State of AI: An Empirical 100-Trillion-Token Study." arXiv:2601.10088; openrouter.ai/state-of-ai, 2026.

[3] Google, "Introducing Gemma 4." blog.google, April 2, 2026. (Apache 2.0 license grant, confirmed against the primary July 1, 2026.)

[4] Google, Gemma 4 Model Card. ai.google.dev/gemma/docs/core/model_card_4, 2026. (Per-layer embeddings; hybrid local/global attention.)

[5] Interconnects, analysis of Google's dual-track open/closed strategy. interconnects.ai, 2026.

[6] Latent Space, Gemma 4 launch-week demand coverage. latent.space, 2026.

[7] Zhipu AI, GLM-5.2 release specifications, 2026. (744B/40B MoE; 1M context; Huawei Ascend support. Verified during the scoring session.)

[8] Hugging Face, Qwen model-hub download data. huggingface.co/Qwen, January 2026. (700M cumulative downloads, past Llama.)

[9] Hugging Face, "State of Open Source — Spring 2026." huggingface.co/blog, March 19, 2026. (Includes >30% of Fortune 500 verified accounts.)

[10] Anthropic, "Higher Limits via SpaceX Colossus Capacity." anthropic.com/news/higher-limits-spacex, May 6, 2026. Confirmations: CNBC, May 6, 2026; xAI press release, May 6, 2026.

[11] OpenAI, Ad Tools Terms. openai.com/policies/ad-tools-terms, June 12, 2026. Related: "New Ways to Buy ChatGPT Ads," openai.com, May 5, 2026; Digiday coverage, April 22, 2026.

[12] Shumailov, I., et al., "AI models collapse when trained on recursively generated data." Nature 631:755–759, 2024. DOI 10.1038/s41586-024-07566-y.

[13] Alemohammad, S., et al., "Self-Consuming Generative Models Go MAD." arXiv:2307.01850, 2023.

[14] Cognition, Platform Terms of Service. cognition.com/legal/platform-terms-of-service, effective June 30, 2026. §§1.2, 1.7, 1.8, 3.3.1, 3.3.2, quoted verbatim.

[15] Cursor (Anysphere), Data Use & Privacy Overview. cursor.com/data-use, June 9, 2026. (Product-level; embeddings clause sits outside the Privacy Mode gate.)

[16] OpenAI, US Privacy Policy. openai.com/policies/us-privacy-policy, May 18, 2026. §§2, 4. Sibling: ROW Privacy Policy, February 6, 2026.

[17] In re: OpenAI, Inc. Copyright Infringement Litigation, MDL No. 25-md-3143 (S.D.N.Y.): Wang order, Nov. 7, 2025; Stein affirmance. Coverage: Bloomberg Law, Nov. 12, 2025; OpenAI CISO statement, openai.com, Nov. 12, 2025.

[18] Cursor (Anysphere), Privacy Policy. cursor.com/privacy, October 6, 2025. (Consumer scope; de-identify-and-use clause.)

[19] Anthropic Privacy Center, feedback retention and training eligibility. privacy.anthropic.com, 2026. (The one banked ACTIVELY-TRAINED exhibit.)

Entries 20–115 are the published March 2026 benchmark's 96-source base (CC BY 4.0), enumerated in the March dossier, order preserved and renumbered +19. Rows 6–12 sub-scores trace here. The dossier's equation pages predate the pre-publication correction to Methodology 2.0; its sub-scores and sources carry, its example Λ values do not.

[20] Enterprise AI ROI Shifts as Agentic Priorities Surge. Futurum, 2026.

[21] Service Terms. OpenAI, 2026.

[22] GPT-5.4 Pro Model. OpenAI API Documentation, 2026.

[23] Pricing. OpenAI API Documentation, 2026.

[24] API Overview. OpenAI API Reference, 2026.

[25] Changelog. OpenAI API Documentation, 2026.

[26] GPT-5 Chat Model. OpenAI API Documentation, 2026.
[27] ChatGPT API Pricing 2026: Token Costs & Rate Limits. IntuitionLabs.ai, 2026.
[28] Global AI Adoption in 2025. AI Economy Institute, Microsoft, 2025.
[29] Mind the Gap: Closing the AI Trust Gap for Developers. Stack Overflow Blog, Feb 2026.
[30] Using GPT-5.4. OpenAI API Documentation, 2026.
[31] Paradigm Shifts of the Developer Mindset in the Age of AI. Snowflake Engineering Blog, 2026.
[32] OpenAI Phases Out GPT-4o and Other Legacy Models from ChatGPT. MLQ.ai, 2026.
[33] The GPT-4o Termination Paradox. Reddit r/ChatGPT, 2026.
[34] Latest Wave of Obligations Under the EU AI Act Take Effect. DLA Piper, Aug 2025.
[35] API Pricing. OpenAI, 2026.
[36] AI Act: Shaping Europe's Digital Future. European Union, 2026.
[37] Enterprises Trust OpenAI, Microsoft and Google AI in 2026. Gammatek ISPL, 2026.
[38] Rate Limits. Claude API Documentation, 2026.
[39] Pricing. Claude API Documentation, 2026.
[40] Claude Code Deployment Patterns with Amazon Bedrock. AWS Machine Learning Blog, 2026.
[41] Anthropic Claude API Pricing 2026: Complete Cost Breakdown. MetaCTO, 2026.
[42] Claude AI Statistics 2026: Revenue, Users & Market Share. Panto, 2026.
[43] Amazon Bedrock AgentCore and Claude. AWS Machine Learning Blog, 2026.
[44] Introducing Claude Sonnet 4.5. Anthropic News, 2026.
[45] GitHub Copilot vs Claude Code: 2026 Accuracy & Speed Analysis. SitePoint, 2026.
[46] Anthropic Achieves ISO 42001 Certification for Responsible AI. Anthropic News, 2026.
[47] Private Cloud Compute is Only Half the Story. Virtru Blog, 2026.
[48] OpenAI vs. Anthropic vs. Google Gemini: Enterprise LLM Platform Guide. Xenoss, 2026.
[49] 2026 Guide to Microsoft Copilot Pricing and Licensing. Adoptify AI, 2026.
[50] Microsoft's Sales Push Ignores Copilot Issues. Aragon Research, 2026.
[51] Copilot vs Gemini: Which AI Fits Your Business? Reality Tech, 2026.
[52] Microsoft 365 Copilot Plans and Pricing. Microsoft, 2026.
[53] Gemini vs Copilot: AI Integration in Google Workspace and M365. TTMS, 2026.
[54] Microsoft Customers Share Impact of Generative AI. Microsoft Source, Nov 2024.
[55] 4 Obstacles Impede Paid Microsoft 365 Copilot Adoption. No Jitter, 2026.
[56] The True Cost of Enterprise AI Agents: A Complete TCO Framework. Medium, 2026.
[57] The Copilot Reality Check: Enterprise Adoption Data. Forrester Blogs, 2026.
[58] Azure AI Foundry and Security Copilot Achieve ISO 42001 Certification. Microsoft Azure Blog, 2026.
[59] What's New in Microsoft 365 Copilot, January 2026. Microsoft Tech Community, 2026.
[60] The Cost of Understanding: AI, Velocity, and Developer Stress. AI Advances, 2026.
[61] Google AI Studio vs Gemini vs Vertex AI. Hoerr Solutions, 2026.
[62] Introducing Gemini Enterprise. Google Cloud Blog, 2026.
[63] Compare Gemini Enterprise vs. Vertex AI in 2026. Slashdot, 2026.
[64] Integrate Gemini Enterprise Agents with Google Workspace. Google Codelabs, 2026.
[65] New Enhanced Tool Governance in Vertex AI Agent Builder. Google Cloud Blog, 2026.
[66] Google AI Guide: Gemini Enterprise vs. Vertex vs. Workspace. ByteeIT, 2026.
[67] Gemini Enterprise vs. Vertex AI Agent Builder. Medium, Feb 2026.
[68] Gemini 3 Grounding with Google Search: Industry Disruption Analysis. Sparkco, 2025.
[69] Gemini Enterprise vs Vertex AI vs Workspace With Gemini. Promevo, 2026.
[70] Data Governance for Commerce. Google Cloud Documentation, 2026.
[71] Data Residency: Generative AI on Vertex AI. Google Cloud Documentation, 2026.
[72] The Future of AI-Powered Work for Every Business. Google Workspace Blog, 2026.
[73] The 10 Most Widely Used LLMs Currently in 2026. Medium, 2026.
[74] Google's 5 Coolest AI Products and Gemini Innovation in 2026. CRN, 2026.
[75] Private Cloud Compute: A New Frontier for AI Privacy. Apple Security Research, 2024.
[76] Updates to Apple's On-Device and Server Foundation Language Models. Apple ML Research, 2025.

[77] Apple Plans M5-Based Private Cloud Compute Architecture. 9to5Mac, Feb 2026.
[78] Apple's Foundation Models Framework Unlocks New Intelligent App Experiences. Apple Newsroom, Sep 2025.
[79] Apple Replacing Core ML with Core AI Framework for iOS 27. 9to5Mac, Mar 2026.
[80] Apple Intelligence Gets Even More Powerful. Apple Newsroom, Jun 2025.
[81] Apple Intelligence. Apple, 2026.
[82] Security Research on Private Cloud Compute. Apple Security, 2024.
[83] Scaling Up Performance from M5. Hacker News, 2026.
[84] Our Longstanding Privacy Commitment with Siri. Apple Newsroom, Jan 2025.
[85] Meta Llama 3 and the 700M MAU Limit. WCR Legal, 2024.
[86] Meta's Llama License Is Still Not Open Source. Open Source Initiative, 2024.
[87] Llama vs Mistral vs Phi: Complete Open-Source LLM Comparison for Enterprise. PremAI Blog, 2026.
[88] Benchmarked the Main GPU Options for Local LLM Inference in 2026. Reddit r/LocalLLaMA, 2026.
[89] The Complete Guide to Private LLM Deployment. Meta Intelligence, 2026.
[90] Meta Llama. Hugging Face, 2026.
[91] Llama 4 on Nutanix Enterprise AI. Nutanix Dev, May 2025.
[92] The Definitive Guide to Local LLMs in 2026. SitePoint, 2026.
[93] Local LLMs vs Cloud APIs: 2026 Total Cost of Ownership Analysis. SitePoint, 2026.
[94] Five AI Predictions That Will Redefine Data Centers in 2026. Digital Realty, 2026.
[95] Llama 3 vs GPT-4: Enterprise Comparison. Rajat Gautam, 2024.
[96] Total Cost of Ownership for Enterprise AI. Xenoss, 2026.
[97] Enterprise AI Risk: Security, Providers, and Regulation. George Mudie, 2026.
[98] Meta Llama 4 Ultimate Capabilities Cheatsheet. AI Mindset, 2026.
[99] DeepSeek One Year Later: Regulatory Storm, Global Surge. AI-Regulation.com, 2026.
[100] DeepSeek R1 vs Qwen 3 vs Mistral Large: LLM Comparison. Digital Applied, 2026.
[101] The Mistral AI Non-Production License. Mistral AI, 2024.
[102] Top 10 Open Source LLMs 2026: DeepSeek Revolution Guide. O-mega.ai, 2026.
[103] DeepSeek and the China Data Question. IAPP, 2026.
[104] Accenture and Mistral AI Accelerate Enterprise Reinvention. Accenture Newsroom, 2026.
[105] Mistral AI's $14 Billion Valuation Marks Europe's AI Turning Point. Cloud Summit EU, 2026.
[106] Europe Artificial Intelligence Market Size and Report 2034. IMARC Group, 2026.
[107] LLMs, Robots and Intelligent Cars: Does China Have an AI Advantage in 2026? WeAreInnovation, 2026.
[108] Time Horizon 1.1. METR, Jan 2026.
[109] Apple M5 Pro & M5 Max Announced: What It Means for Local AI. Reddit r/LocalLLaMA, 2026.
[110] Using a High-End MacBook Pro or RTX 5090 Laptop for Inference. Reddit r/LocalLLM, 2026.
[111] Next-Generation LLM for UAV: From Natural Language to Autonomous Flight. ResearchGate, 2026.
[112] Real-Time Systems Design and Analysis, Fourth Edition. Laplante, 2012.
[113] What AI Coding Costs You. Tom Wojcik, Feb 2026.
[114] Sovereignty. AI Now Institute, 2026.
[115] AI Sovereignty & Data Localization Guide. Meta Intelligence, 2026.

The Grove Foundation

"Design is philosophy expressed through constraint."

CC BY 4.0 — July 2026 · 12 patterns · 115 sources · Full transparency

Provenance: scores-snapshot-q2-2026.md (amended 2026-07-01), gemma-fc-swing-memo-q2-2026.md, structural-share-memo-q2-2026.md, delta-memo-q2-2026.md, verification-log-q2-2026.md (deliverables + inputs), cognition-callout-q2-2026.md, public-page-reconciliation-q2-2026.md. Methodology and zones per _engine/methodology.md (V2.0), matching the published March equation (the-grove.ai/alerts/ai-deployment-pattern-benchmark — the exponential form in the reference dossier predates the pre-publication correction and is not used). Legacy sub-scores ported from the March 96-source base (enumerated in _engine/reference/competitive-benchmark-1.0.pdf) with documented Q2 deltas, reconciled against the published March board (Llama restated 0.0104 → 0.0031; Anthropic/Copilot two-decimal rounding noted). Every published Λ reproduces from its sub-scores via $\Lambda = (S \times R \times V)/(1 + (\beta \cdot Fc)^2)$, re-verified at build. _engine/ untouched; March baseline fixed; one quarter only. No push; no q2-2026 tag until the public page ships.