From the Managing Director

Revision 2 — July 8, 2026. Changes from r1: opt-out scoping corrected in the ratchet table (opt-out reaches Customer Data only; no opt-out exists for Usage Data); Claude Tag exhibit added; market paragraph anchored to CNBC/OpenRouter (Jul 7); "technically true" framing; renewal timing on the five questions; operator line edits to the opening. New primaries banked in the verification log addendum (r2), Items A–C.

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 you're reading now 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.

Step back one layer, because the Karp episode is a symptom of the larger shift. For years we measured artificial intelligence by model capability. That yardstick broke in Q2 2026. While harness engineering and loops hold our attention, the real fight has shifted to a different layer: today, apex vendors are desperate to own 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 of the AI industry now.

Three words carry everything that follows, so here they are in plain terms. 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 the model — the layer that runs tools, watches the work, and records it. **Custody** means who holds and controls that record. Multiply one trace by every engineer, every day, and it is the most valuable byproduct your company produces — the pattern of how your organization thinks. The terms now in force transfer custody of that pattern to the vendor. Lawfully. At every privacy setting. By design.

The market has started pricing this in. This quarter's board shows momentum moving away from the big API vendors and toward open-weight patterns a company can run and keep for itself. Every one of the quarter's five strongest patterns is a Sovereign open-weight pattern — weights you can download, run on your 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, lands a full tier below at 0.0058 — five times short of the line. The demand data now says the same thing with a byline: per CNBC's July 7 report of OpenRouter's own figures, Chinese open-weight models' share of U.S.-company tokens averaged 4.5 percent in the first half of 2025, has run above 30 percent every week since February 8, 2026, and peaked at 46 percent — with GLM-5.2 landing within a point of the strongest closed model on a watched agentic benchmark at roughly a fifth of the cost. Enterprises are voting with volume for weights they can keep.

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. When a vendor tells you it does not train on "your data," the statement is technically true — under definitions the vendor wrote. Your data is protected. The record of how your people think is, by definition, not your data anymore. "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.

EXHIBIT: 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

DATE	VENDOR	WHAT THE DOCUMENT SAYS
		codebase… may be stored in our database.” The clause sits outside the Privacy Mode gate.
Jun 30, 2026	Cognition (Platform ToS)	The ratchet closes. A new sentence appears: “Cognition owns all right, title, and interest in and to the Usage Data” (§3.3.2). The opt-out the terms offer reaches **Customer Data only** — paid tiers, admin-only on Teams (§3.3.1). **For Usage Data there is no opt-out clause, on any tier, at any setting.** 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 — current terms and prior version, both linked from cognition.com. Full citations in the Q2 Standings report, sources [14]–[19]; Claude Tag and CNBC/OpenRouter citations in the verification-log addendum.

Read the last row again. In fourteen months, one vendor moved from “we will not train unless you say yes” to a two-bucket architecture with no exit: an opt-out that reaches only the bucket you already owned, and formal title to the anonymized record of how your teams use the product — a bucket no setting reaches, because you cannot opt out of someone else’s property. 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.

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.

The harvest has moved up the stack

The terms above govern the code editor. In June, the same architecture arrived in the place your whole company thinks out loud. Anthropic’s Claude Tag embeds a proprietary agent natively inside Slack — its own service accounts, channel-scoped memory that accumulates your teams’ terminology, patterns, and decision history over months. It is genuinely useful, and that is the point: at launch, Anthropic disclosed that “65% of our product team’s code is created by our internal version of Claude Tag.” That is the vendor’s own number for what this class of tool is worth when the loop is closed and kept. Two more launch facts complete the picture: introductory launch credits for Enterprise and Team organizations to accelerate adoption, and no documented export path for the accumulated channel memory — leaving means inducing corporate amnesia. This is not one vendor’s move; OpenAI’s Codex Remote runs the same play at the execution layer. The pattern is the strategy: the harness, not the model, is the moat, and every exchange it watches is another trace in the bucket the terms reserve.

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' own internal-use disclosures show what it is worth — they 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; the Claude Tag launch shows the same architecture arriving in the collaboration layer, on the vendor's own disclosures. 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, Anthropic included — which is why its own launch disclosures appear in this letter next to everyone else's terms.

Five questions for your legal and risk teams — before the next renewal

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. And the timing matters: this transfer happens at signature. Every renewal your procurement team signs this year is a fresh acceptance of whatever these definitions now say.

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 — and is there any opt-out that reaches it?** Cognition's current terms claim it in writing: "Cognition owns all right, title, and interest in and to the Usage Data" —

and offer no opt-out for that bucket, on any tier. Put the same two-part 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 full Q2 Standings report 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 — before you renew.

Jim Calhoun Founder & Managing Director, The Grove Foundation, Inc.

*The Λ Standings and the Q2 2026 benchmark report are published openly (CC BY 4.0) at the-grove.ai/lambda. Every claim in this letter traces to a primary source enumerated in the report's Works Cited [14]–[19], to the vendors' own published archives, or to the public record. 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. Claude Tag: Anthropic launch announcement and support documentation, June 2026 (vendor's own disclosures). Token-share figures: CNBC, "Chinese AI models are gaining ground with U.S. companies," July 7, 2026, reporting OpenRouter data. The spoken and written records are cited separately throughout; they are different documen