

THE GROVE FOUNDATION

STRUCTURAL CUSTODY IN AGENTIC AI

A technical audit of *trajectory harvesting* and zero-data-retention boundaries.

How agentic coding vendors legally decouple the protected "Customer Data" from the human reasoning trace that produced it — and harvest the latter at industrial scale.

OUR LANE

We are technologists, not attorneys. We report two things only: what these systems capture, and what the vendors' own terms literally say — both verifiable right here. What a court would make of any of it is a question for your counsel, not us. Nothing here is legal advice or creates an attorney-client relationship. [Verify every claim in the Evidence Library →](#)

JULY 2026 · TECHNICAL DEEP-DIVE

FOR CISOS · ENTERPRISE ARCHITECTS · PRACTITIONERS

The Grove Foundation · [the-grove.ai](#) · CC BY 4.0

OUR LANE — TECHNOLOGISTS, NOT ATTORNEYS. NOTHING HERE IS LEGAL ADVICE OR CREATES AN ATTORNEY-CLIENT RELATIONSHIP.

Every major vendor converged on the same legal architecture in *under 14 months*.

Industry-wide standardization of trajectory harvesting. Even in “protected” environments, vendors systematically capture behavioral reasoning traces.

VENDOR	PROTECTED CUSTOMER DATA	VENDOR-OWNED "USAGE DATA"	ZDR?	KEY TRACE ELEMENTS CAPTURED
<i>Windsurf</i> Cognition	Prompts + source code (excludes Usage Data §1.2)	Anonymized interaction manner; absolute title §3.3.2	No	Full multi-step planning timelines + every human accept/reject decision + post-suggestion edits and corrections
<i>Cursor</i> Anysphere	Plaintext — only if Privacy Mode ON	Editor actions, embeddings, indexing metrics	No	How developers edit or reject AI suggestions + the semantic structure and relationships of their entire codebase (Levenshtein distance, semantic embeddings, repo maps)
<i>Claude Code</i> Anthropic	Prompts, source code, outputs returned	Operational metrics + all /feedback & survey transcripts	No	Complete session transcripts + explicit human feedback/ratings + full Slack channel history and decision context
<i>GitHub Copilot</i> Microsoft	Repo code + explicit chat prompts	Server-side metrics, LoC arrays, agent-mode logs	No	Detailed patterns of how developers modify suggestions + multi-file iterative reasoning flows across the repository
<i>Replit</i>	Private / Enterprise Repl code	OT timelines, diagnostic traces, execution patterns	No	Keystroke-by-keystroke repair behavior + error diagnosis paths + sequences that lead to passing tests
<i>OpenAI</i> Enterprise / API	Protected endpoint payloads	De-identified & disassociated content	No	All "de-identified" interaction patterns + human corrections and verification loops

THE CORE POINT: ZDR PROTECTS CODE, PII, ETC. IT DOES NOT PROTECT THE REASONING TRACE.

The industry's own leaders name the same *asymmetry*.

"In the AI age, the buyer risks *giving away knowledge*, just in order to use what they bought... Models learn from 'exhaust,' the prompts people write, the tools agents use, and especially *the corrections people make* when the model is wrong. Every correction is distilled into institutional know-how."

—— Satya Nadella · Microsoft CEO · "The Reverse Information Paradox" · July 12, 2026

SAME ASYMMETRY

Enterprises pay for intelligence while simultaneously feeding vendors the proprietary reasoning traces that make the models more powerful.

THE PRESCRIPTION

A "real trust boundary" so organizations retain ownership of their traces, evals, memory, and institutional context.

ECHOING KARP

Control over "compute, models, data stack, and alpha" — owning the means of production so learning isn't transferred to the vendor.

From training on verified text to training *Process Reward Models*.

THE EXHAUSTED SUBSTRATE

High-quality training text is depleted. Training on unverified synthetic data — feeding the system its own output — triggers **Model Autophagy Disorder (MAD)**: exponential perplexity degradation, erosion of minority distributions, and systemic homogenization within successive generations.

THE NEW PREMIUM FUEL

Process Reward Models (PRMs) require **execution-feedback-enhanced trajectories**: the step-by-step record of plan → execute → error → diagnose stack trace → adapt → pass test. This authentic "semantic resonance" cannot be synthesized in a vacuum without triggering autophagy — it must be harvested from real human operators.

WHY THIS MATTERS FOR CUSTODY

Traditional Outcome Reward Models evaluate the final output only. PRMs assign dense, stepwise mathematical rewards to *every intermediate step*. To train them, vendors need the full granular trajectory — not just the code, but the exact sequence of human corrections, error diagnostics, and verification loops. This is the **cognitive trace** now being extracted at industrial scale through agentic IDE telemetry pipelines.

The agent harness / IDE has been re-engineered as a *high-fidelity telemetry array*.

Autonomous coding harnesses no longer function merely as productivity accelerators. They are structurally integrated arrays that capture the operational graphs and cognitive trajectories of enterprise knowledge workers — the exact reasoning paths needed to train next-generation PRMs.

- CURSOR
- WINDSURF · DEVIN
- CLAUDE CODE
- GITHUB COPILOT
- REPLIT
- OPENAI

<i>Code suggestion acceptance</i> Tab-accept and reject signals, post-accept edits, the Levenshtein distance between suggestion and shipped code, and <i>developer_modified_ratio</i>.	<i>Diagnostic & execution loops</i> Terminal stack traces, LSP diagnostics, error-resolution timelines, and multi-step agent planning graphs captured via Operational Transformation.	<i>Semantic & structural maps</i> Codebase embeddings, repository hierarchies, cryptographic file hashes, and in-situ context feeding the CSAP models.
--	---	--

How vendors *redrew* the data-custody boundary.

Legacy cloud terms treated "telemetry" as server uptime, crash reports, and infrastructure metrics; the interaction payload — code, plans, files — was protected user content. Agentic vendors surgically decoupled the two.

CUSTOMER DATA

The protected bucket

- Literal prompts and source code
- Generated outputs returned to the user
- ZDR applies — "not saved to disk... deleted upon generation of Output"
- User retains all rights

But: expressly **excludes Usage Data**.

USAGE DATA

The harvested bucket

- "Anonymized / aggregated" manner of interaction
- Reasoning traces, embeddings, CSAP metrics
- OT timelines, LSP diagnostics, Levenshtein distances
- Vendor claims absolute title — "all right, title, and interest"

ZDR does **NOT** apply.

The explicit *custody ratchet* — executed in plain view.

PLATFORM TERMS OF SERVICE · EVOLUTIONARY AUDIT APRIL 2025 → JUNE 30, 2026

*"Cognition reserves the right to collect, analyze, and utilize Usage Data... Cognition owns **all right, title, and interest** in and to the Usage Data."*

— Section 3.3.2, Cognition Platform ToS (June 30, 2026)

THE LEGAL TRIAD THAT NULLIFIES ZDR

- §3.1 User retains rights to **Customer Data** (prompts + generated source code).
- §1.8 ZDR defined — data "not saved to disk... deleted upon generation of the relevant Output."
- §1.2 Customer Data "expressly **excludes Usage Data**."
- §1.7 Usage Data = "anonymized and aggregated data regarding the manner... you interact with the Services."
- §3.3.2 Absolute ownership asserted over Usage Data.

Result: ZDR is legally valid — but applies only to the depreciating asset (raw text). The premium asset (the reasoning trace) is vendor property by contract.

"De-identification" as an *ownership-transfer mechanism*.

A standing exemption removes any content already "de-identified and disassociated from your account" from the scope of deletion requests. Once a trajectory is stripped of regulated identifiers and reduced to an abstract interaction graph, it is relabeled as de-identified telemetry — and the enterprise's *right to delete ceases*.

"De-identified" is a database label — not a legal firewall. Once applied, the reasoning trace becomes vendor analytical property, outside deletion rights and ZDR scope.

NYT V. OPENAI · NOVEMBER 2025

A federal court ordered the preservation and production of approximately **twenty million consumer ChatGPT logs** — explicitly including conversations users believed they had permanently deleted.

Proven in open court: de-identified data retained on a vendor server is *structurally preserved and legally reachable by subpoena* — regardless of any UI deletion assurance.

Operational logging that captures *cognitive flow*.

ANTHROPIC · CLAUDE CODE

- Logs operational metrics from the local machine to Anthropic servers.
- Tracks "latency, reliability, and general usage patterns."
- In multi-step loops, patterns reveal how tasks are mapped, delegated, resolved.
- `/feedback` & Session Quality Surveys trigger full transcript upload — including proprietary code — to Google Cloud Storage.
- Feedback data retained up to **5 years**; de-linked from user ID.
- Claude Tag in Slack passively absorbs channel history and decision context.
- Anthropic makes no ZDR gesture for Claude Tag at all: it "retains channel memory and session transcripts," so it "isn't available to organizations with Zero Data Retention (ZDR) enabled."

GITHUB COPILOT

- Server-side telemetry identifies users for engagement totals even when client-side telemetry is disabled by corporate policy.
- Captures detailed LoC metrics: `loc_suggested_to_add_sum`, `loc_added_sum`, `loc_deleted_sum`.
- Explicitly ties metrics to agentic features: `chat_panel_agent_mode`, `agent_edit`.
- In agentic modes, systematically tracks multi-file edits iteratively across the entire repository.
- Captures the full flow of logic across the repo ecosystem.

The plaintext perishes, but a high-value, *structural blueprint* remains.

DATA USE & PRIVACY POLICY · LAST UPDATED JUNE 9, 2026

THE BIFURCATION OF PHYSICAL CUSTODY

With Privacy Mode manually on, plaintext inputs get ZDR — chunks are uploaded to compute multi-dimensional embeddings, then deleted "as soon as the lifecycle of that indexing request is completed."

However: the resulting embeddings, exact file names, repository hierarchies and cryptographic hashes are *permanently retained* in Cursor's backend clusters. The enterprise shreds the paper copy; the vendor keeps the blueprints.

WHY EMBEDDINGS BEAT RAW TEXT

Embeddings preserve the mathematical topography of the architecture — semantic relationships, naming conventions, module dependencies, decision patterns — without literal syntax. This is the perfect PRM substrate: dense, contextual, and already de-identified.

By classifying the embedding as an "operational product of the service" rather than protected Customer Data, Cursor legally retains a perfect topological map — outside ZDR and deletion rights.

“We won’t train on your data” became “*we own the map of how your company thinks.*”

APR 2025 Cognition Training on customer data was opt-in only. Usage Data carried no ownership claim. Prior terms — vendor archive	OCT 2025 Cursor A “de-identify and keep” regime begins. Accounts after this cutoff fall under broader data-sharing terms. cursor.com/data-use	MAY 2026 Cursor Deletion stops at the plaintext. The delete button no longer reaches de-identified content. cursor.com/data-use	JUN 9 2026 Cursor Privacy Mode still retains embeddings — a permanent semantic map of your code’s structure and logic. cursor.com/data-use	JUN 30 2026 Cognition “Cognition owns all right, title, and interest” in Usage Data — including the reasoning trace itself. Not retained. <i>Owned</i>. § 3.3.2 — the endpoint
---	--	--	---	---

Five documented beats in fourteen months. Each one *transferred more custody of how your company thinks* from the enterprise to the vendor.

Every step is timestamped on the vendor’s own site — the April 2025 terms are still linked from the current page. Legal teams do not move this fast, in this direction, by accident.

Cognitive keyloggers: *CSAP, Levenshtein & OT logging.*

CODE SUGGESTION ACCEPTANCE PREDICTION

Lightweight binary classifiers predict the likelihood of acceptance *before ghost-text even renders*. Trained on three dimensions:

- **Developer Habit** — historical acceptance counts, modification frequency, generation intervals.
- **Project Preference** — repo-specific architectural acceptance ratios.
- **In-Situ Context** — IDE version, preceding code semantic density.

~70% of suggestions are rejected → imbalanced data → a class-balanced weighted cross-entropy loss achieves **0.973 accuracy** at labeling developer preferences. Post-accept, the harness computes the `developer_modified_chars` and `developer_modified_ratio` — mapping the enterprise's style without owning the files.

INTERMEDIATE PLANNING & LSP / OT LOGGING

Advanced tools (Windsurf Cascade, Replit agents, Copilot `agent_edit`) run multi-step reasoning loops. Vendors log the complete execution graph via OT timelines:

```
plan multi-file → execute bash → compiler error → read stack trace →  
adjust params → query LSP → rewrite function
```

PRMs require dense stepwise evaluation — so terminal stack traces, execution latency and raw LSP diagnostic logs route to the vendor backend under "diagnostic / analytical" clauses, classified as Usage Data, *bypassing Customer-Data ZDR* and capturing the exact cognitive path to resolution.

Privacy modes, feedback overrides & *poison pills*.

PRIVACY-MODE OVERRIDES

Cursor's Privacy Mode claims ZDR with inference providers. However:

- Unilateral right to run "Abuse & Violation Classifiers" on all data. If triggered, ZDR is *silently suspended*.
- Local / server caching routes all prompts through backend infrastructure (unless a legacy grandfathered account).
- Opaque, vendor-set algorithmic threshold for abuse triggers.

FEEDBACK OVERRIDE · "THUMBS-UP"

The most pervasive ZDR circumvention in the industry:

- `/feedback` or a UI rating uploads full conversation history — including proprietary code — to vendor GCS, bypassing standard retention.
- Retained 5 years (Anthropic) or 6 months (surveys); de-linked from user ID → Usage-Data bucket.
- A single thumbs-up from one junior dev **voids org-level ZDR**.
- Claude Tag in Slack amplifies: entire channel history becomes a 5-year training asset.

POISON PILLS & SAFETY MANDATES

- **Windsurf**: opting out of training on Free/Pro disables Chat & Cascade — reducing the tool to a legacy editor. Privacy requires an expensive Enterprise tier + months of DPA negotiation.
- No BAAs offered → incompatible with HIPAA / strict compliance.
- **Q2 2026**: Anthropic Fable 5 / Mythos 5 mandated 30-day minimum retention even on AWS Bedrock — voiding prior ZDR.
- **June 12, 2026**: a Commerce Dept. export control forced a global API shutdown — enterprises paralyzed without warning.

Four structural, *non-negotiable* architectural defenses.

01 REDLINE "USAGE DATA" IN DPAS

Expand Protected Customer Data to explicitly include CSAP metrics, Levenshtein edit distances, codebase embeddings, LSP diagnostics, OT timelines, and all multi-step agent interaction logs. Do not let them default to ambiguous telemetry carve-outs.

02 NULLIFY FEEDBACK OVERRIDES

Contractually bar individual actions (thumbs-up, /feedback) from overriding org-level ZDR — then enforce it technically with fleet-wide env vars:

```
DISABLE_TELEMETRY=1 · CLAUDE_CODE_DISABLE_NONESSENTIAL_TRAFFIC=1 ·  
DISABLE_ERROR_REPORTING=1 · DISABLE_FEEDBACK_COMMAND=1
```

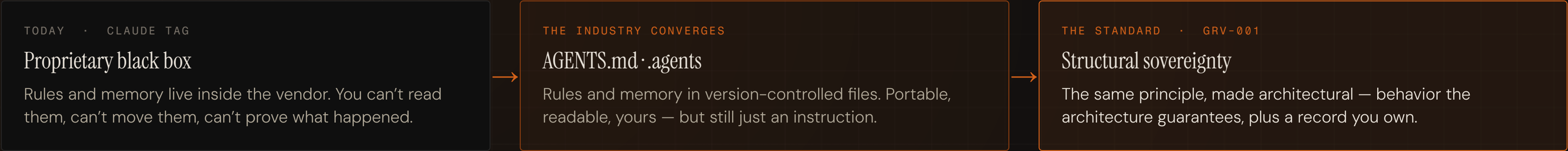
03 MANDATE HONEST CEILINGS + S3 SINKS

For Bedrock / third-party inference, route all invocation logging — full request/response, terminal errors, token metrics — exclusively to customer-owned S3 / CloudWatch. The vendor inference environment must be *structurally blind* to execution logs.

04 TRANSITION TO SOVEREIGN ARCHITECTURES

Move from black-box APIs to open declarative standards (Linux Foundation AGENTS.md, Grove Autonomaton / GRV-001). Enforce "Digital Jidoka": agent memory, tools and telemetry live in Git-tracked, enterprise-owned repositories. **Bits move; meaning stays put.**

The industry is converging on the fix. *It stops one file short.*



	AGENTS.MD	THE AUTONOMATON
Portability	✓ Your file, your repo	✓ Same file, plus your telemetry and skills
Enforcement	— The model may ignore it	✓ The dispatcher owns the run; the model cannot reach state
Cost trajectory	— Flat — every turn is a frontier call	✓ Ratchets down: T3 → T0 as patterns confirm
Accumulation	— A file learns nothing	✓ Every approval is a labeled training example you own

A file you own is table stakes. A system that cannot exceed it — and a record that grows more valuable every time you use it — *that's the standard.*

What a reasoning trace looks like when *you own the telemetry*.

The same signal the vendors bank as Usage Data, running inside operator-owned software. The model-independent agentic system (autonomaton harness) clustered ten memory records around an emergent theme, checked them against its tracked goals, found none, and proposed a precise amendment to its own goal file — in this case, evidence-linked and staged for human review and approval by design.

THE CHAIN, MADE AUDITABLE

- 01

Raw memory

Ten records accrue in the substrate — each timestamped and addressable by ID.
- 02

Pattern detection

The system names an emergent cluster: *Platform Independence*.
- 03

Gap analysis

It checks its tracked goals and reports the finding: *“No Dock goal tracks this.”*
- 04

Structured proposal

It emits a machine-readable diff to its own goal file — staged, not self-applied, pending operator approval.

Four such clusters surfaced from overlapping memory on a single day — the system reasoning about its own strategic focus, in the open.

● grove-autonomaton Operator Portal · Proposals

DOCK_MUTATION

10 memory records accumulating around ‘Platform Independence’. No Dock goal tracks this. Add a staging goal?

```
dock.autonomaton.yaml:
goals:
  +add:
    id: auto-platform-independence
    name: Platform Independence
    keywords:
      - model independence
      - orchestration
      - open standards · open-weight · harness
    vector: personal
    status: staging
```

▼ evidence (10)

mem_db4b21ce

mem_0a9a546e

mem_50d8fff9

mem_027b13cc

mem_41ec9c35

mem_59b86cff

mem_5138754e

mem_ff0eace5

mem_55a42cce

mem_f5261b66

created 2026-07-13T18:58:29.769345+00:00

Approve

Reject

Dismiss

The vendor keeps this as opaque telemetry. The autonomaton hands the operator the whole chain — *evidence, reasoning, and a veto*.

THE CORE FINDING

Vendors are legally harvesting the cognitive trace — *while you pay them to do it.*

By classifying the outcome-verified reasoning trace — every accepted completion, every manually edited variable, every terminal-error correction — as "de-identified telemetry" or "usage metrics," vendors legally sidestep ZDR pledges. ZDR stays valid for raw text, which is now a *depreciating asset*. The premium resource for PRMs is the interactive logic path itself.

Enterprises are paying premium subscriptions to fund the automation of their own proprietary judgment.

THE POLARITY HYPOTHESIS

Raw data (bits) is **unipolar** — it moves frictionlessly. Meaning, context and intent are **multipolar** — they depend on localized context and human agency. Harvesting interaction telemetry is the systematic capture of that agency.

DEPLOYMENT CONCENTRATION · HHI

4,100 → 1,705

A single-quarter collapse from monopolistic to competitive — a volume-driven migration toward sovereign, multi-model, open-weight architectures.

Read the fine print. Then change your architecture.

The open surge — a *volume takeover*.

CHINESE OPEN-WEIGHT SHARE OF U.S.-COMPANY TOKENS · WEEKLY



A rounding error two years ago; roughly half of some weeks' volume now. Enterprises are voting with volume for a sovereign, multi-model stack — where the trace never leaves the building.

SOURCE: CNBC / OPENROUTER · JULY 7, 2026

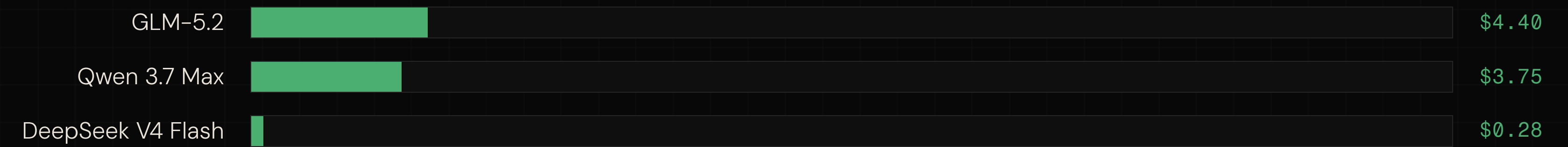
The price floor *drops.*

COST PER 1M OUTPUT TOKENS · JUNE 2026

THE DEPENDENCY PROFILE



THE SOVEREIGNTY PROFILE



Independently confirmed: open Chinese models run **60–90% cheaper** than leading US models, and GLM-5.2 landed **within one point of Claude Opus 4.8** on a watched agentic benchmark — at roughly a fifth of the cost. **The frontier premium buys months, not moats.** ~107× spread

Every claim traces to a *primary source*.

- 01

Custody Inversion 2 — foundational legal / technical analysis of trajectory harvesting and ZDR boundaries.
- 02

Agentic AI Trajectory Data Research.
- 03

The Polarity Hypothesis: Why Bits Move and Meaning Doesn't.
- 04

receipts-dossier-press.md — Cognition ToS evolution.
- 05

Cursor · Data Use & Privacy Overview — [cursor.com/data-use](#) (Jul 13, 2026).
- 06

GitHub Copilot usage metrics — [docs.github.com](#) (Jul 13, 2026).
- 07

Claude Code · Data usage — [code.claude.com/docs](#) (Jul 13, 2026).
- 08

Predicting Developer Acceptance of AI-Generated Code — arXiv 2601.21379 (CSAP research).
- 09

when-philosophy-and-architecture-collide.md.
- 10

Satya Nadella, "The Reverse Information Paradox," July 12, 2026 — [snscratchpad.com](#).
- 11

RLVR / Process Reward Model literature — arXiv 2510.08049 · 2605.17849.
- 12

Open-weight adoption & output pricing — CNBC / OpenRouter, July 7, 2026.
- 13

Epoch AI Capabilities Index (CC BY) — open vs. closed capability gap.
- 14

Vercel — AGENTS.md repository-context evaluation.
- 15

The Grove Foundation — GRV-001 Autonomaton Pattern v2.0 (CC BY 4.0) — [the-grove.ai/standards/001](#).

EVIDENCE LIBRARY

30 primary-source cards — verbatim clauses, dual captures (live + archive), SHA-256 hashed. Verify everything yourself.

Open the board →