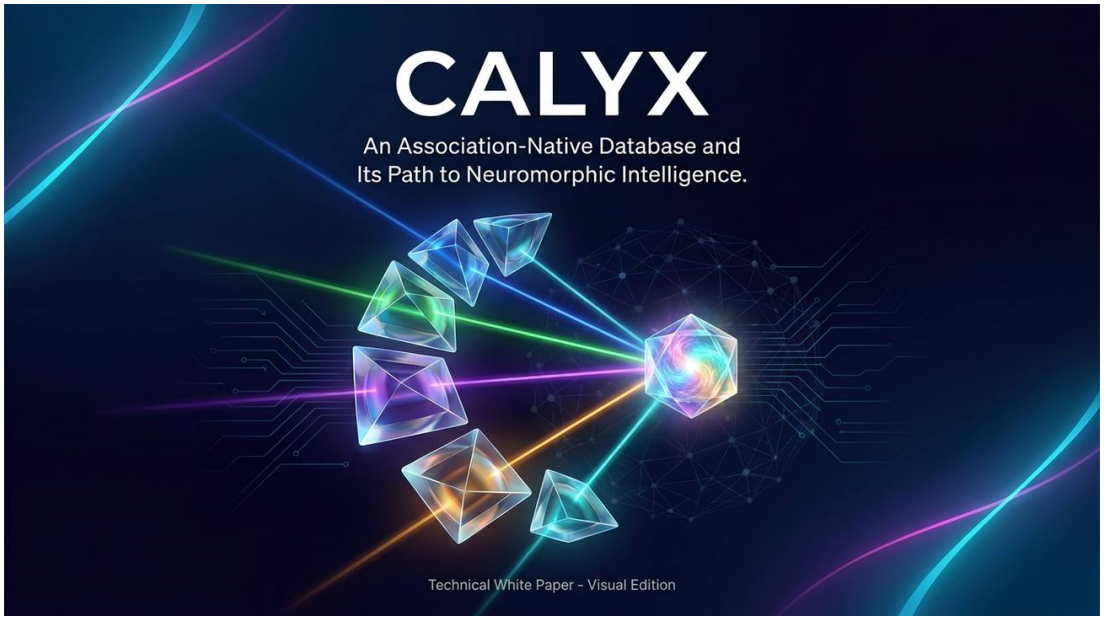


Calyx: An Association-Native Database and Its Path to Planetary-Scale Grounded Intelligence

A Technical White Paper — Illustrated Formal Edition

Chris Royse

June 29, 2026

Status: architecture white paper (Calyx is pre-1.0, under active development)

Abstract

Calyx is an **association-native database engine**: a single, ordered, transactional system whose atomic record is not a row or a lone vector but a **constellation** — one input measured through many independent, frozen "lenses," each contributing a separate typed slot-vector that is kept distinct and never flattened. On top of this representation Calyx bakes in a *calculus of association*: it derives the relationships *between* slots, measures in bits which lenses actually add information about real outcomes, discovers the minimal "grounding kernel" that explains a corpus, guards generation against out-of-distribution answers, records a hash-chained provenance ledger, and continuously self-optimizes — all under three non-negotiable principles: **grounding is mandatory, slots stay un-flattened, and the system fails closed.** This white paper presents the conceptual foundations and the full subsystem architecture of Calyx, then sets out its defining scaling thesis: because the calculus of association is, by construction, computationally rich — it grows with the number of records, the number of lenses, and the associations between them — its natural hardware home is **neuromorphic and in-memory computing**. We show, operation by operation, how spiking neural networks and compute-in-memory substrates map onto Calyx's core math, and we lay out a hybrid architecture in which a precision-tolerant **association fabric** runs the similarity, ranking, and graph-traversal work at extreme energy efficiency and $O(1)$ streaming insertion, while a bit-exact **digital spine** preserves Calyx's provenance and trust guarantees. We then ground that fabric in real silicon — **Intel's Loihi 2** neuromorphic processor and the **1.15-billion-neuron Hala Point** system — showing operation-for-operation how Calyx's primitives map onto Loihi 2's native capabilities, and noting that the central Calyx operation (cosine k-NN with $O(1)$ streaming insertion) has *already been demonstrated* on this hardware family. Finally, we develop an explicit, theory-anchored **speculative analysis** (§8–§11): hardware is only as valuable as the software that confers a reason to own it — the lesson of the GPU, which video games built and deep learning revalued a thousandfold — and we argue that an association-native database with spiking-neural-network integration is a candidate *killer application* for neuromorphic silicon, the missing software that could make Loihi-class hardware strategically valuable. We lay out, with twelve theory-backed pros and twelve theory-backed cons and a thousand-theory "birds-eye" synthesis, how markets and the world could shift if such a system became fully operational with serious backing. The result is a roadmap from a single-workstation system to an engine capable of ingesting and reasoning over massive, multi-domain, multimodal datasets — and a disciplined account of what would have to be true for that engine to reshape computing.

1. Introduction

1.1 The problem with one number per thing

Modern AI retrieval reduced "meaning" to a single embedding vector per item. This is convenient and powerful, but it is also a profound act of compression that throws away something essential: **a single input means many different things at once.** A sentence has a semantic meaning, a syntactic structure, a lexical surface, a rhetorical posture, a domain register (legal, clinical, financial, scientific), an authorial style, a temporal context. A molecule, a genome, an image, an audio clip — each is simultaneously many things to many observers. Collapsing all of that into one averaged vector destroys exactly the cross-perspective structure where the most valuable, least-obvious associations live.

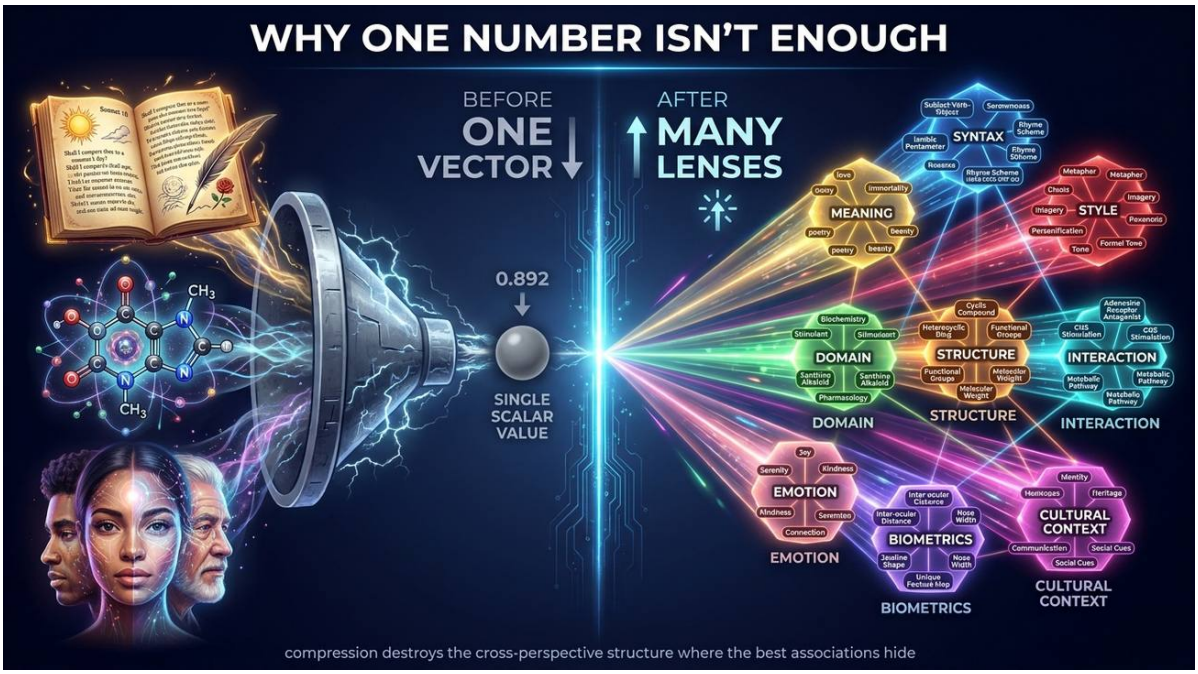


Figure 1. *Collapsing meaning into one vector discards the cross-perspective structure where the most valuable associations live.*

Calyx begins from the opposite premise: **keep the perspectives separate, and make the relationships between them first-class.** This is not a vector database with more features. It is a different data model — and a different theory of what a database is *for*.

1.2 The thesis

Calyx is built to **unlock associations latent in the world's data that have not yet been connected**, and to do so in a grounded, traceable way. Its output is never "an answer"; it is a set of **ranked, provenance-backed, validatable hypotheses**. Two levers maximize what it can discover: **corpus breadth** (many domains, large volumes of literature and data) and **lens diversity** (many independent perspectives, because the cross-terms between distant domains are where undiscovered links hide). This makes Calyx a *grounded-association discovery engine* — and, generalized, a substrate for grounded machine intelligence.

1.3 The calculus of association: four verbs

Calyx organizes everything around four operations, each realized by a dedicated subsystem:

Verb	Meaning	Realized by
Measure	View one input through a panel of independent lenses → a constellation	Registry + storage
Count	Derive the associations *between* slots (agreement, delta, interaction)	Loom
Differentiate	Quantify, in bits, the unique information each lens adds about real outcomes	Assay
Compose	Find the kernel, guard generation, answer with provenance	Lodestar + Ward + Ledger

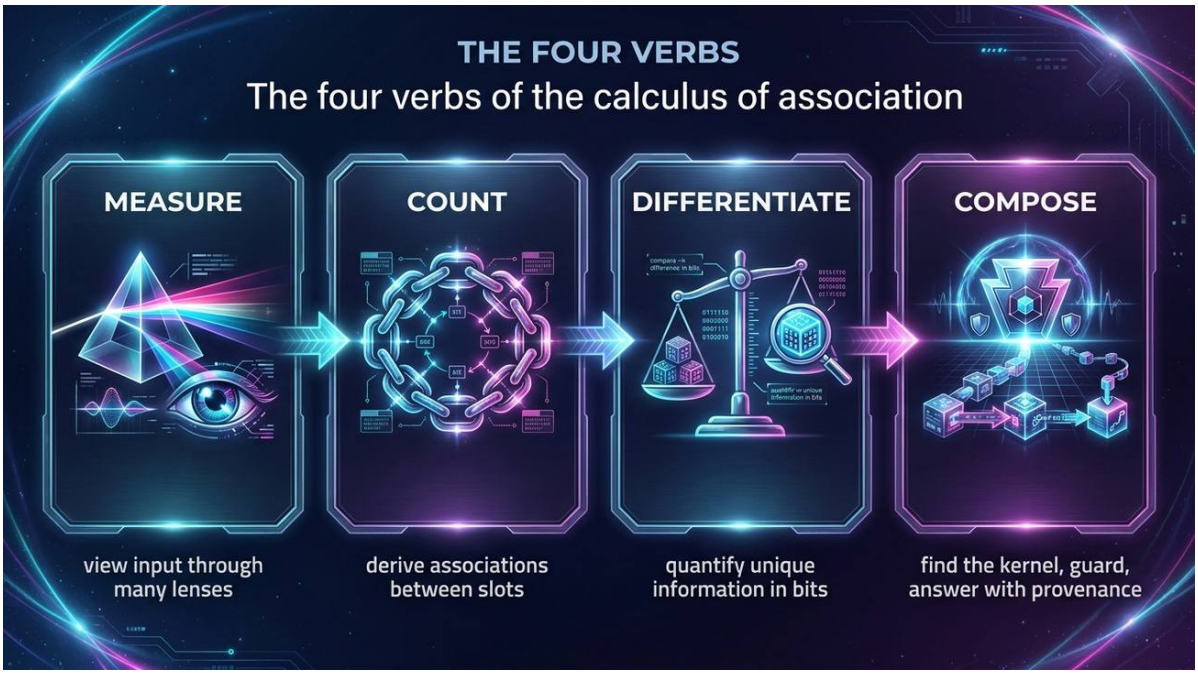


Figure 2. *The calculus of association in four verbs: measure, count, differentiate, compose.*

1.4 Three trust principles, enforced in the architecture

1. **Grounding is mandatory.** Claims are measured against anchored real-world outcomes; ungrounded results are tagged *provisional*, never *trusted*.
2. **No-flatten.** Slots remain typed and separate end to end; the engine never concatenates them into one opaque blob, and every gate scores each slot independently.
3. **Fail closed.** An unknown lens, a shape mismatch, an uncalibrated guard, or missing grounding returns a structured error, never a silent wrong answer.

These principles are not aspirations bolted on top; they are wired into the type system, the error catalog, and every subsystem boundary.

Figure 3. *The three trust principles, enforced in the type system: grounding mandatory, no-flatten, fail-closed.*

2. Foundational concepts

2.1 The constellation — Calyx's atomic record

A **constellation** is one input measured by one **panel** of frozen lenses. It carries:

- a content-addressed identity (CxId), derived by hashing the input together with the panel version and a vault salt — so identical inputs are idempotent and every record is re-derivable;
- a map of **slots**, one per lens, each holding a typed **slot-vector** (dense, sparse, or multi-vector) kept separate from the others;
- scalar measurements, string metadata, and a list of grounded **anchors** (observed outcomes);
- a provenance reference into the Ledger proving the input→lens→constellation lineage;
- trust flags (grounded / provisional / degraded / novel-region).

The constellation is the unit of storage, of association, of measurement, and of provenance.

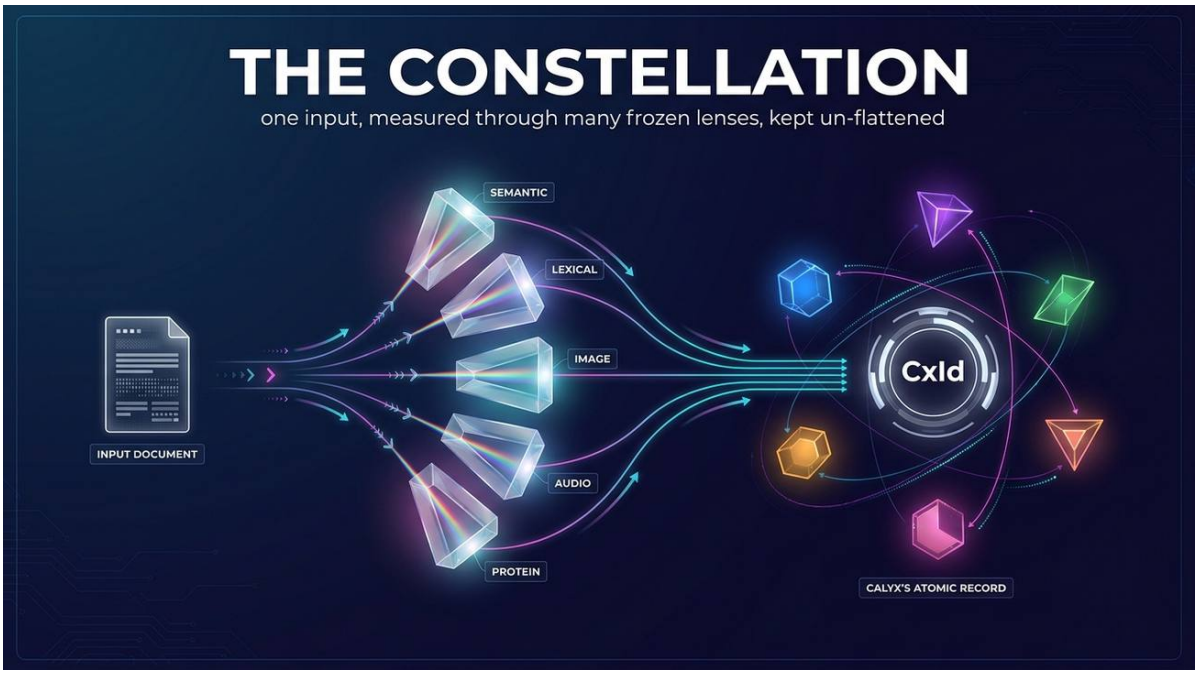


Figure 4. *The constellation: one input measured through many frozen lenses, each slot kept separate and content-addressed.*

2.2 Lenses, slots, panels

A **lens** is a frozen, content-addressed measurement instrument — an embedder that turns raw bytes into a slot-vector. "Frozen" means a lens is identified by the hash of its name, weights, training corpus, and output shape; if anything changes, it becomes a *new* lens identity. This guarantees that a measurement is reproducible forever.

A **slot** is a frozen lens occupying a position in a panel. A **panel** is a versioned set of slots. Panels are content-addressed, saved, profiled, and **hot-swapped** per vault or per query without re-embedding the corpus — so the system can carry far more commissioned lenses than are resident at any instant, and specialize its "sensorium" to the task.

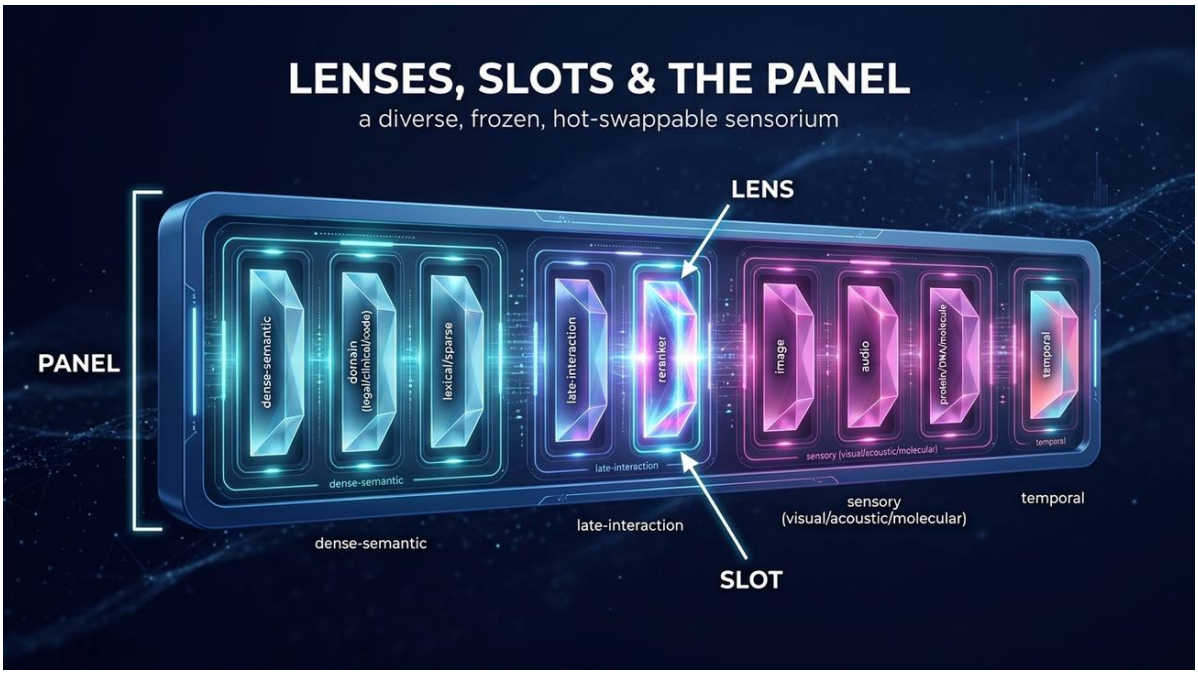


Figure 5. *Lens, slot, and panel: a diverse, frozen, hot-swappable sensorium.*

2.3 Lens diversity is the thesis (and it is measured)

The value of a lens is **associational, not intrinsic** — it is the unique, redundant, and synergistic information it adds *relative to the rest of the panel*. You cannot value one or two lenses in isolation; meaningful measurement begins at three and stabilizes around ten or more. Calyx therefore treats a diverse, ten-to-thirty-plus-lens panel as the foundation, and it *measures* diversity directly: association-family span, panel-wide redundancy (effective rank, mean pairwise correlation), and a non-collapsing unique-information curve. Ten near- identical semantic models are not a panel; they are one perspective wearing ten hats.

Calyx's reference general-purpose panel ("Constellation-24") spans roughly nine association families and twenty-to-thirty-plus content lenses across every modality and domain a general database is asked to serve:

Family	Examples of perspective
Dense-semantic (general)	broad meaning, multilingual, instruction-aware
Dense-semantic (domain)	scientific, biomedical, clinical, legal, financial, code
Lexical / sparse	learned term expansion (BM25++)
Late-interaction / multi-vector	token-level fine-grained matching
Reranker / asymmetric	cross-encoder relevance
Image	zero-shot (CLIP-style) and self-supervised (DINO-style)
Audio	speech, speaker, music, environmental
Science modalities	protein, DNA, molecule
Temporal sidecars	recency, periodicity, sequence (post-retrieval)

2.4 Cross-terms and Derived Data Abundance (DDA)

Given N lenses, Calyx does not stop at N signals per record. It also derives one **cross-term per pair of slots**, plus the whole-panel constellation signal. The per-record signal yield is:

$$(\text{signals per record}) = N + \binom{N}{2} + 1$$

where $\binom{N}{2} = \frac{N(N-1)}{2}$. For a 24-lens panel that is $24 + 276 + 1 = 301$ signals per input. This is **Derived Data Abundance**: a modest number of perspectives generates a combinatorially larger field of relationships. The cross-terms come in four kinds:

- **Agreement** — cosine between two slot-vectors (a scalar): "do these two perspectives concur?"
- **Delta** — element-wise difference: "how do they differ, dimension by dimension?"
- **Interaction** — element-wise (Hadamard) product: "where do they jointly fire?"
- **Concat** — the joined representation for downstream indexing.

The agreement cross-terms form an **association graph** over the panel — the substrate from which structure (kernels, communities, blind spots) is later discovered.

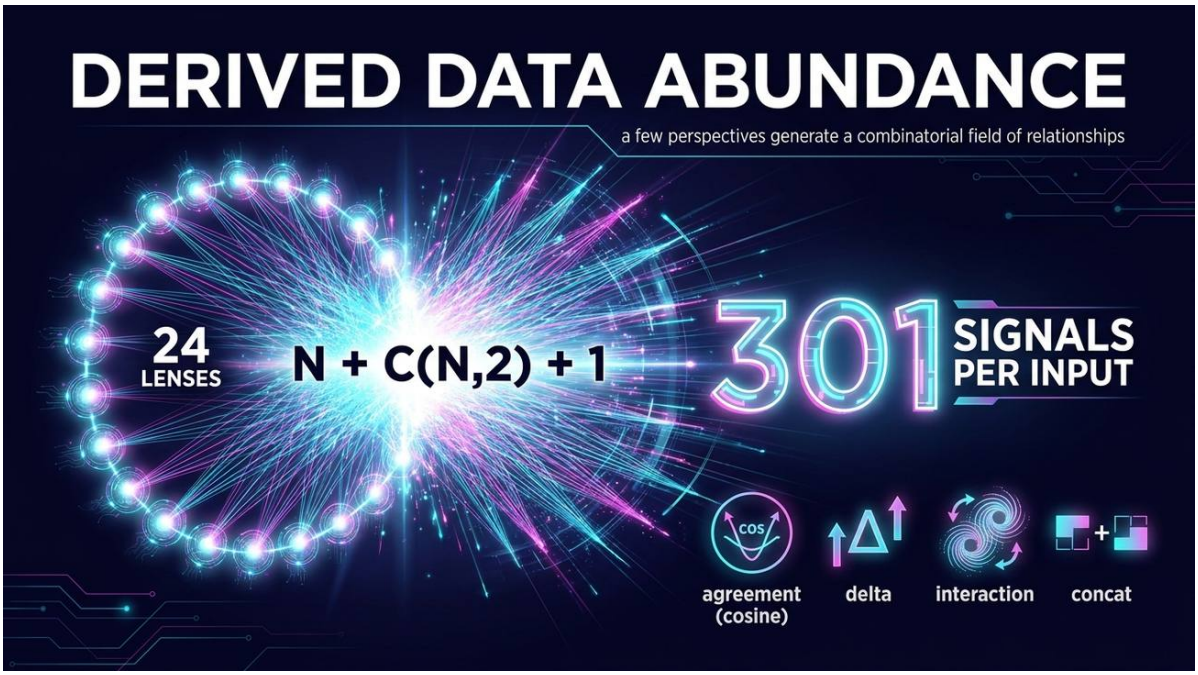


Figure 6. *Derived Data Abundance: N lenses yield N + C(N,2) + 1 signals per input.*

2.5 Anchors and grounding

An **anchor** is an observed real-world outcome attached to a constellation: a test passed, a label, a reward, a thumbs signal, a speaker match, a recurrence. Anchors are what convert a statistical association into a *grounded* one. Calyx's information measurements, kernel discovery, and guard calibration all reference anchors — and an association that cannot be traced to grounding is reported as provisional, never trusted. The honest upper bound on what Calyx can discover is therefore set by the grounded information actually present in the chosen corpora (a data-processing-inequality ceiling), and every hypothesis is an input to validation, not a verdict.

Figure 7. *Anchors convert a statistical association into a grounded one; ungrounded results are provisional, never trusted.*

3. System architecture

Calyx is a stack of focused engines compiled together, with thin entry points (a CLI, a daemon, an agent-facing tool surface, and a read-only HTTP API) on top. The layering runs from a dependency-free core, through a storage/math/provenance foundation, through the association and intelligence engines, to the surfaces.

```
Surfaces      : CLI · daemon · agent tool surface · read-only HTTP API
Shared search : one recall → fusion → rerank → provenance path
Intelligence  : Oracle · Anneal · Lodestar
Engine        : Sextant · Loom · Assay · Ward · Registry
Foundation    : Aster (storage) · Forge (math) · Ledger (provenance)
Core          : ids · error catalog · data model · engine traits · graph primitives
```

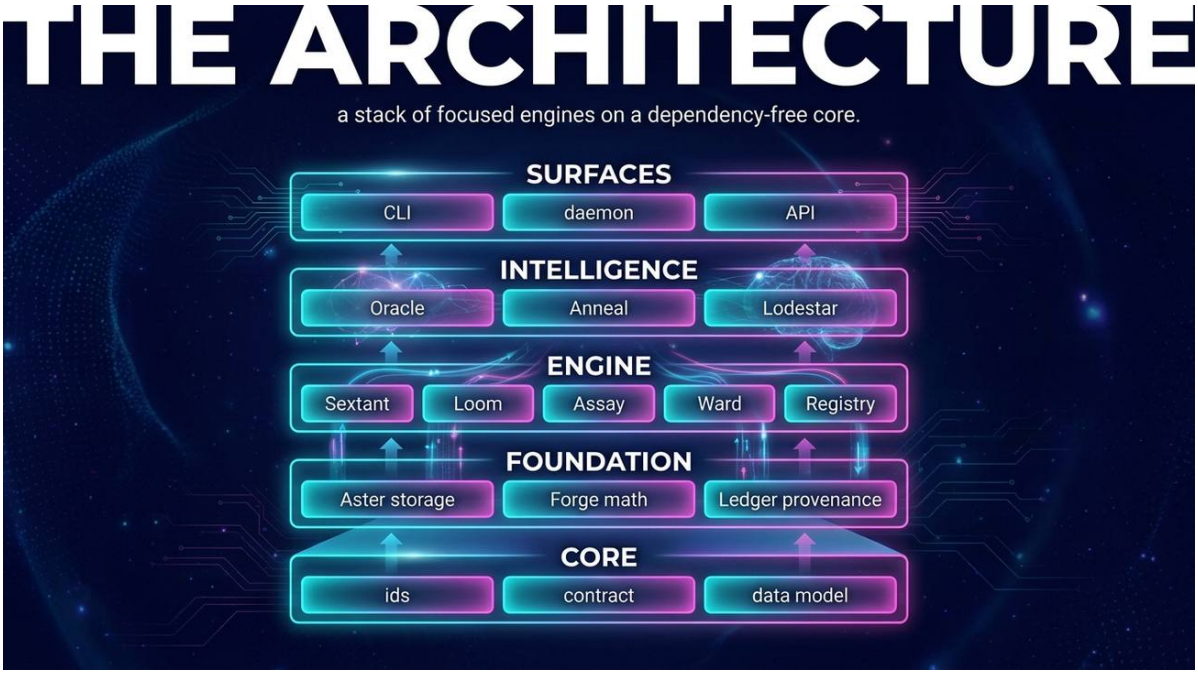


Figure 8. *The Calyx stack: focused engines on a dependency-free core, from storage to surfaces.*

3.1 Core — the contract

The foundation provides content-addressed identifiers, a closed error catalog, the constellation data model, a shared enum vocabulary, and the object-safe engine traits (Lens, Index, VaultStore, Estimator) that every other engine implements. Because the traits and the error catalog live here, the trust principles (no-flatten, fail-closed, grounding) are enforced uniformly across the system. Clocks are injected, so logic is deterministic and reproducible.

3.2 Aster — the association-native storage engine

Aster is an ordered, transactional, multi-version store built as a log-structured merge tree: a write-ahead log feeds bounded in-memory tables that flush to immutable sorted tables, with snapshot isolation, atomic manifests, crash recovery, and hot/cold tiering. Its column families are **association-native** — there are families for base constellations, per-slot vectors (quantized and raw), cross-terms, anchors, signal-bit measurements, kernels, guard profiles, the provenance ledger, recurrence series, and graph projections. Aster also provides general data-layer "layers" (relational, document, key-value, time-series, blob, graph), so a single ordered core can serve every paradigm's root purpose. Quantized slots and indexes live on fast storage; bulky raw vectors tier to archive — dense *meaning* stays hot.

3.3 Forge — the math runtime

Forge is Calyx's owned numeric runtime: one backend contract (matrix multiply, cosine, dot, L2, normalize, top-k) implemented for CPU SIMD and for CUDA, engineered for bit-near parity between them. It carries a quantization stack from full precision down to very low bit-widths (scalar INT8, microscaling FP formats, rotation-based TurboQuant with a residual correction, and binary), each gated by *measured* intelligence preservation rather than asserted. The backend abstraction is deliberately substrate-agnostic — the same contract can host new compute backends (the basis of the neuromorphic path in §7).

3.4 Registry — the lens registry

The Registry holds frozen, content-addressed lenses and the runtimes that execute them (algorithmic encoders, HTTP embedding services, local transformer inference, ONNX, static-lookup, multimodal adapters, and specialized sparse/late-interaction/reranker variants). It enforces the frozen contract on every measurement, supports hot-swap add / retire / park with lazy backfill, and maintains a library of named, versioned panel templates that specialize the sensorium per situation without re-embedding the corpus. Every commissioned lens carries a signed manifest and a measured cost (resident memory, latency), so panels can be packed by *signal density* — information per unit of resource.

3.5 Sextant — search and navigation

Sextant provides per-slot indexes (in-memory dense ANN, on-disk graph indexes for server scale, sparse/centroid-partitioned indexes for sparse slots) plus a learned-sparse keyword index, multi-vector late interaction, and a kernel-first funnel for very large vaults. Results across slots are fused by **Reciprocal Rank Fusion** (and weighted, profile-driven variants), chosen by a deterministic query planner that classifies intent and enforces cost bounds. Navigation modes go beyond lookup: cross-lens **consensus** (agree / disagree), asymmetric/causal **traversal** of the association graph, and hierarchical **skill** discovery and skill-scoped search.

3.6 Loom — the association engine

Loom computes the cross-terms (agreement, delta, interaction, concat) and maintains the agreement graph over the panel. It materializes the right associations eagerly and computes the rest lazily under an information-gain policy, tagging each signal as measured (a real input through a frozen lens) or derived (a cross-term). Loom also detects **blind spots** — regions the panel measures little — and extends association across time (lead/lag and recurrence analysis). It is the engine that turns N perspectives into the $\binom{N}{2}$-plus field of relationships.

3.7 Assay — signal in bits

Assay answers the question "is this lens *earning its slot*?" in the only honest currency: **bits of information about a grounded outcome.** Using k-nearest-neighbor mutual-information estimators (with bootstrap confidence intervals), Assay measures each lens's signal, each pair's gain, the panel's effective non-redundant size, and **panel sufficiency** — whether the panel carries at least as much information about an outcome as the outcome's own entropy. It enforces a **differentiation contract** (a lens must add meaningful unique bits and must not be redundant with an existing lens) and routes localized **deficits** to the self-optimization layer (where exactly the panel is information-starved, and therefore which new lens to propose).

3.8 Lodestar — the grounding kernel

Lodestar discovers the **kernel**: the small set of records (on the order of one percent) that carries the structure of the whole corpus — operationally, an approximate minimum feedback vertex set of the directed association graph. Removing the kernel leaves the rest of the corpus reconstructable by association, "bottoming out" at grounded anchors. Lodestar scores candidates by a blend of graph centrality and groundedness, then turns the kernel into both a compact vector index and an **answer path** that walks association edges from a grounded entry point to the query, with provenance at every hop. The kernel is computable **at any scope** (a domain, a collection, a time window, a tenant, an intersection of these), and it serves as a universal, structural summarizer.

3.9 Ward — the fail-closed guard

Ward scores each required slot **independently** against a per-slot, conformally calibrated cosine threshold — there is no single averaged gate. Calibration sets each threshold to a target false-accept rate with a statistical confidence bound. Ward's verdicts — accept, open a new safe region (learn), quarantine, or refuse — are the immune boundary of the system, applied wherever generation touches the store, and every verdict is recorded to the provenance ledger.

3.10 Ledger — provenance

The Ledger is an append-only, hash-chained, tamper-evident record. Every trusted signal — measurement, signal-bit estimate, kernel, guard verdict, answer, self-optimization step, migration — is one canonical entry whose hash seals the previous, with periodic Merkle checkpoints (optionally signed) for export and attestation. Verification re-walks and re-hashes the chain. Crucially, the Ledger stores **hashes and identifiers, not secrets or raw content**, so provenance survives even after raw sources are deleted, and a discovered association can always be traced back to the exact sources that ground it. The Ledger also underwrites **reproducibility**: a recorded answer can be replayed and re-measured to prove it was *measured*, not fabricated.

3.11 Anneal — reversible self-optimization

Anneal continuously tunes the system — index parameters, quantization levels, materialization plans, storage cadence, online heads, guard thresholds — under two invariants: **never regress a tripwire metric, and always be reversible and logged.** Every proposed change is shadow-tested against a held-out replay, gated against safety tripwires (recall, guard false- accept/reject rates, latency) and per-metric non-regression, then either promoted by a single pointer swap or rolled back — with every promotion and revert recorded to the Ledger. Anneal also closes the loop on lens coverage: when Assay reports an information deficit, Anneal can propose a new lens, gate it for differentiation, and hot-add it without re-embedding.

3.12 Oracle — grounded consequence prediction

Oracle predicts grounded outcomes and their downstream consequences from recurrence evidence: a hop-attenuated "butterfly" consequence tree forward, a reverse walk from outcome to cause (epistemic symmetry), and — binding — an **honesty gate** that refuses to answer when the panel cannot carry the outcome's information, returning instead a per-sensor deficit that says exactly which perspective is missing. Confidence is capped by the panel's measured self- consistency. Oracle is the seed of Calyx's role as a grounded reasoning substrate.

3.13 Graph primitives

Underneath Lodestar sit the graph engines: strongly-connected-component condensation, centrality, spectral analysis, and the directed association graph with attenuated traversal (reach, weighted best-first). These are the primitives the kernel and the answer path are built on.

4. The data lifecycle

A single pass through Calyx, in the language of the four verbs:

1. **Ingest & Measure.** An input is measured through the resident panel of frozen lenses, producing a constellation with one typed slot per lens. The constellation is content- addressed (idempotent) and sealed into the provenance ledger.
2. **Count.** Loom derives the cross-terms and updates the agreement graph — the relationships between perspectives.

3. **Differentiate.** Assay measures, in bits, what each lens and each pair contributes about grounded anchors, and whether the panel is *sufficient* for an outcome.
4. **Compose.** Lodestar discovers the grounding kernel and answer paths; Sextant retrieves and fuses across lenses; Ward guards any generated output; the Ledger records the lineage. Oracle, where applicable, predicts grounded consequences. Anneal quietly tunes the whole system between queries.

The output is always a ranked, provenance-backed, validatable result — with the bits that justify it, the grounding that anchors it, and the chain that lets a human trace it home.

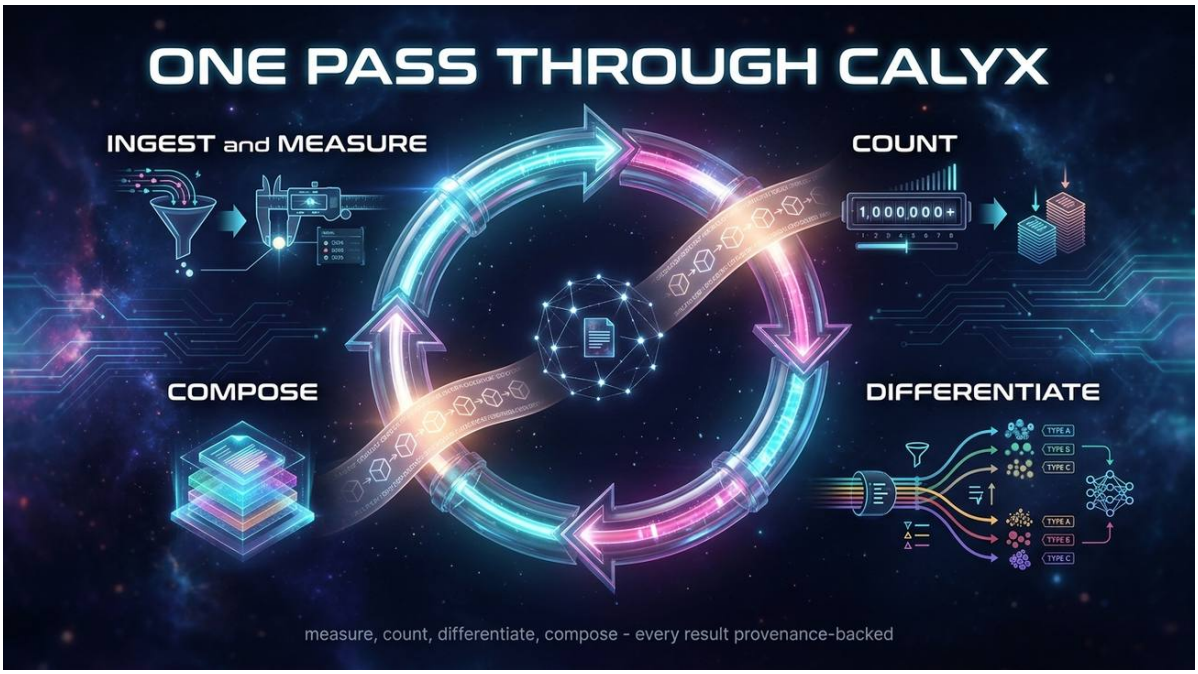


Figure 9. *One pass through Calyx: ingest and measure, count, differentiate, compose, all provenance-backed.*

5. Trust, governance, and reproducibility

Calyx's differentiator is not only what it can find, but that what it finds can be *trusted and verified*:

- **Grounding-or-provisional:** nothing is labeled trusted without grounded evidence.
- **No-flatten:** every gate and measurement respects the separateness of perspectives, so a strong signal on one axis cannot be laundered into a false overall confidence.
- **Fail-closed:** unknown, malformed, ungrounded, or under-provisioned states return structured errors, never silent fallbacks.
- **Provenance for every record of every type:** clinical, scientific, molecular, multimodal, derived — all carry traceable source lineage, durable after raw downloads are deleted.
- **Per-slot conformal guarding** with calibrated false-accept rates.
- **Reproducibility** via deterministic re-measurement and chain verification.
- **Privacy and governance by construction:** content-addressing, hash-only provenance, tenant isolation, and opt-in redaction.

This is what makes Calyx suitable as the memory and reasoning plane for high-stakes domains — where a novel cross-domain association is worthless unless it can be traced to the exact papers, datasets, and rows that ground it.

6. The intelligence layer

Calyx is designed not merely to retrieve but to *understand and grow understanding*:

- **Kernel at any scope** — the ~1% that explains the whole, computable for any domain, collection, time window, or tenant; doubling as a compact index, an answer engine, and a universal structural summarizer.
- **Oracle** — grounded prediction with an honesty gate and epistemic symmetry.
- **Self-optimization toward an intelligence objective** — Anneal optimizes a measurable objective: maximize the growth of grounded, non-redundant understanding, as fast as safely possible, by re-parameterizing the system online (fusion weights, quantization, thresholds, online heads) — never by rewriting its own engine source, and never past the information ceiling its grounding allows.
- **Living perception** — when a corpus reveals signal no existing lens captures, the system can propose, measure, gate, and hot-add a new lens, growing its own sensorium.

The honest bound throughout: Calyx claims **operational intelligence and life-like behavior** (perception, memory, differentiation, homeostasis, foresight, an immune boundary) — never consciousness, sentience, or qualia.

7. The scaling thesis: spiking neural networks and neuromorphic computing

7.1 Why scale is the defining engineering question

The calculus of association is, by construction, computationally rich — and that richness is a *feature*, the very source of Calyx's discovery power. But it sets the scaling agenda. Consider a corpus of n records measured through L diverse lenses of dimension d:

Work	Grows as	Intuition
Measurement (lens inference)	$n \cdot L$	every record through every perspective
Cross-terms (Loom)	$n \cdot \binom{L}{2} \cdot d$	the combinatorial field of relationships
Information estimation (Assay)	up to $O(n^2 \cdot d)$	pairwise grounded mutual information
Association graph (Lodestar)	with graph size (V, E)	structure over the whole corpus

This is not an accident to be engineered away; it is the mathematics of measuring meaning from many perspectives at once and counting the relationships among them. To reach **massive, multi-domain, multimodal, always-on datasets**, Calyx needs a compute substrate whose physics match this workload — one where similarity, association, winner-selection, and graph traversal are *native and cheap*, and where the marginal cost of one more record or one more relationship approaches the cost of a single event.

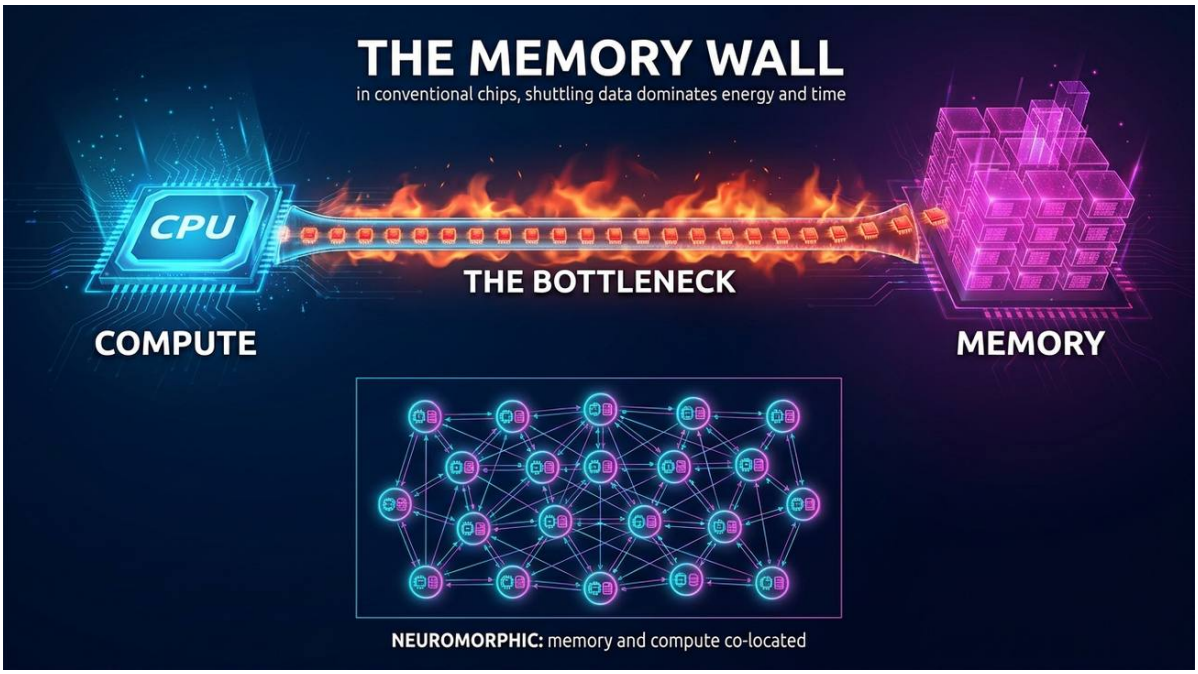


Figure 10. *The memory wall: in conventional architectures, moving data dominates energy and latency.*

That substrate is **neuromorphic and compute-in-memory hardware.**

Figure 11. *The AI energy wall: data-centre electricity is set to roughly double to ~945 TWh by 2030.*

7.2 Why the four verbs are neuromorphic-native

There is a deep structural rhyme between Calyx's primitives and what spiking and in-memory substrates do best:

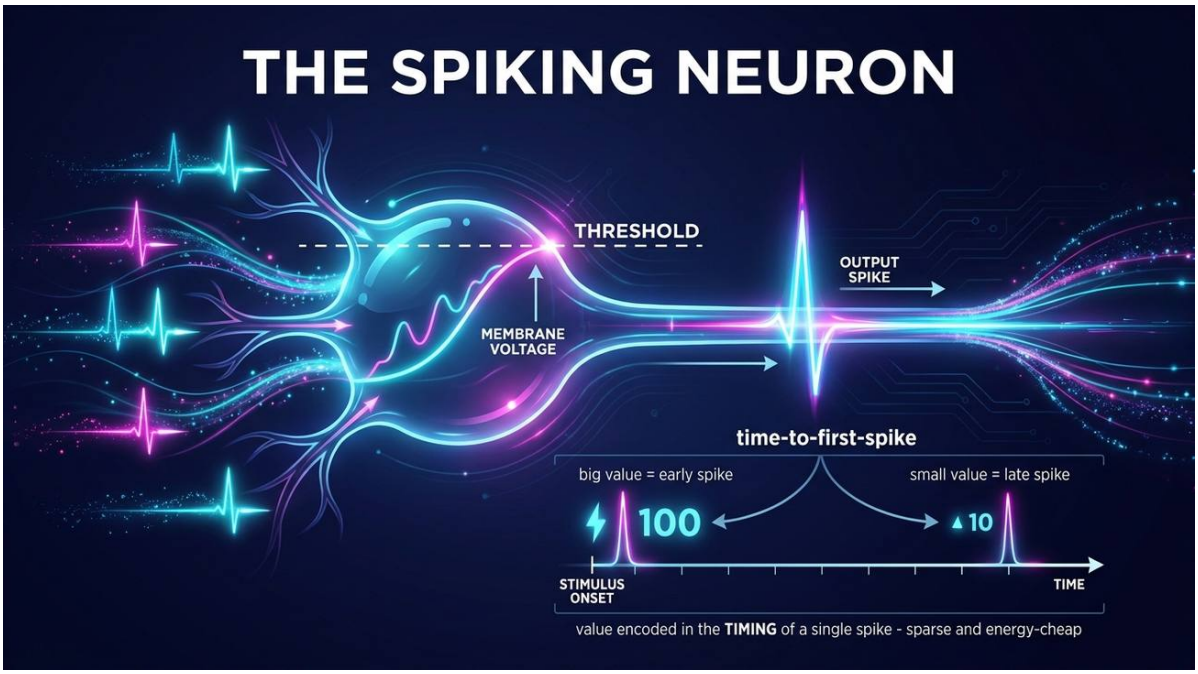


Figure 12. *A spiking neuron encodes a value in the timing of a single spike (time-to-first-spike).*

Calyx primitive	Neuromorphic-native operation
Agreement cross-term (cosine)	matrix-vector multiply in one physical pass (in-memory crossbar)
Similarity / nearest-neighbor (Sextant)	latency-coded spiking search; first spikes = best matches
Winner selection (fusion, guard)	k-winners-take-all; the first-k spikes *are* the top-k
Association-graph reach (Lodestar answer path)	spike-wavefront propagation; arrival time = path cost
Associative recall (retrieval, dedup)	content-addressable / attractor memory
Streaming ingest (always-on memory)	event-driven processing with $O(1)$ insertion

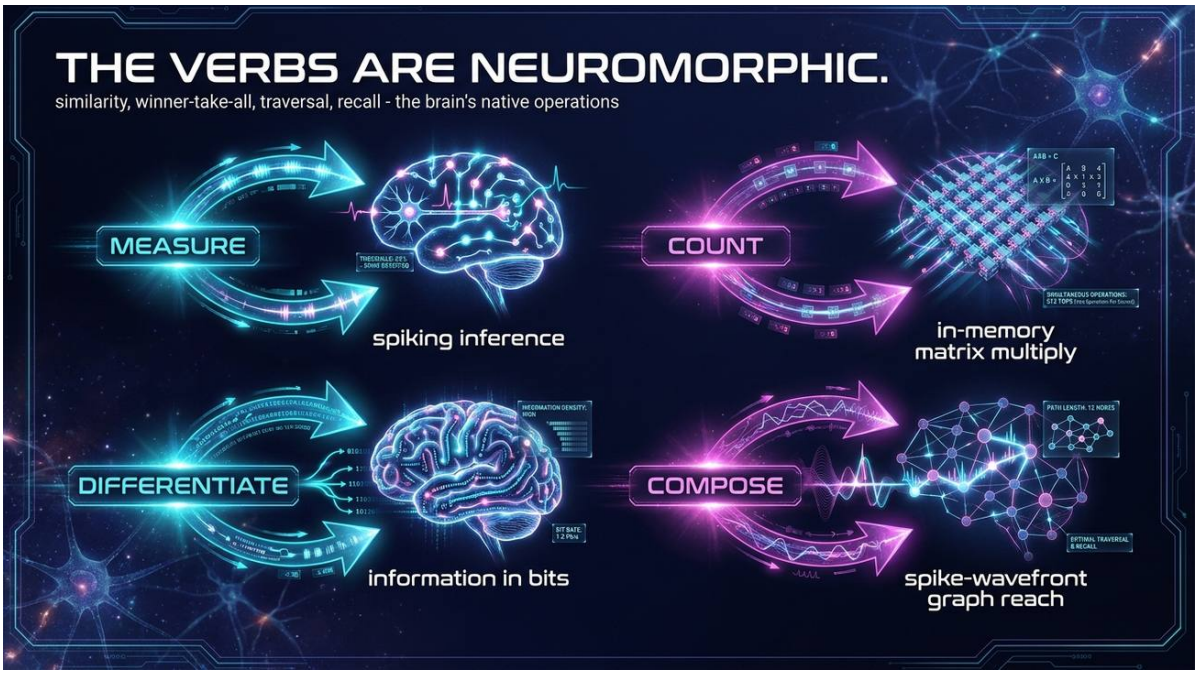


Figure 13. *Calyx's four verbs map onto operations a neuromorphic substrate performs natively.*

In a neuromorphic substrate, **energy scales with events (spikes), not with floating-point operations, and latency scales with network depth, not with dataset size.** A continuously arriving stream of new records — the regime a planetary memory plane must live in — is its home turf, including $O(1)$ online insertion of new items into the search structure. This is precisely the scaling profile Calyx's "always-on memory" ambition requires.

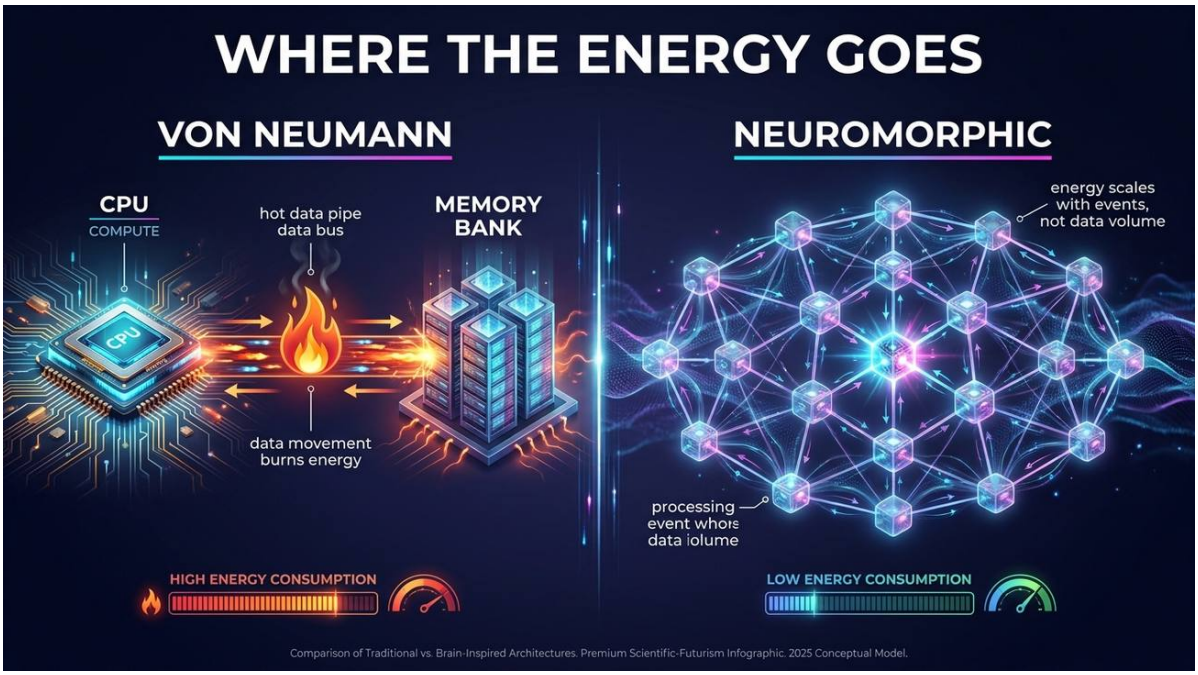


Figure 14. *Co-locating memory and compute makes energy scale with events, not data volume.*

7.3 The substrate, concretely — Intel Loihi 2

The neuromorphic argument above is not abstract: a substrate with exactly these properties exists today and is the most advanced of its kind in the world — **Intel's Loihi 2**, the second-generation neuromorphic research processor from Intel Labs (announced September 30, 2021). Calyx's neuromorphic path is designed against this concrete target. A single Loihi 2 chip is:

- a **fully asynchronous, clockless, event-driven digital spiking processor** with **128 programmable neuron cores** (up to 8,192 neurons each) plus **6 embedded x86 (Lakemont) microprocessor cores**, connected by an on-chip 8×16 mesh network — *compute and memory are co-located in every core; there is no DRAM and no global clock*;

Figure 15. *Intel Loihi 2: 128 neuron cores, ~1M programmable neurons, graded spikes, on-chip learning, on Intel 4.*

- capable of **up to 1,048,576 neurons and 120 million synapses** per chip, with **192 KB of flexibly-allocated SRAM per core**;

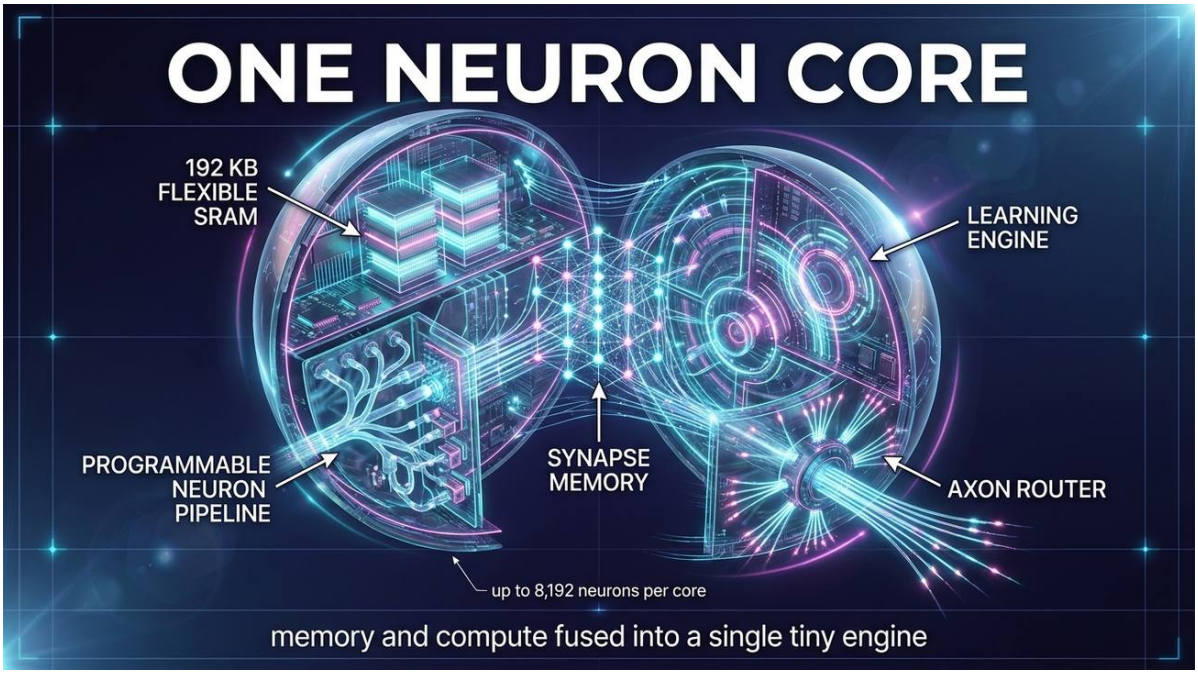


Figure 16. *Inside one neuron core: 192 KB flexible SRAM, a programmable neuron pipeline, synapse memory, and learning engine.*

- built on a **pre-production Intel 4 process** (Intel's first EUV-lithography node), **31 mm^2** of silicon holding **2.3 billion transistors**;

- equipped with **fully programmable neuron models** (each neuron is a short *microcode* program over a fixed-point instruction set — no floating-point units, no multiply- accumulators), **graded spikes** carrying integer payloads of up to 32 bits (typically 8), and a **programmable three-factor synaptic learning engine** capable of on-chip approximations of backpropagation;

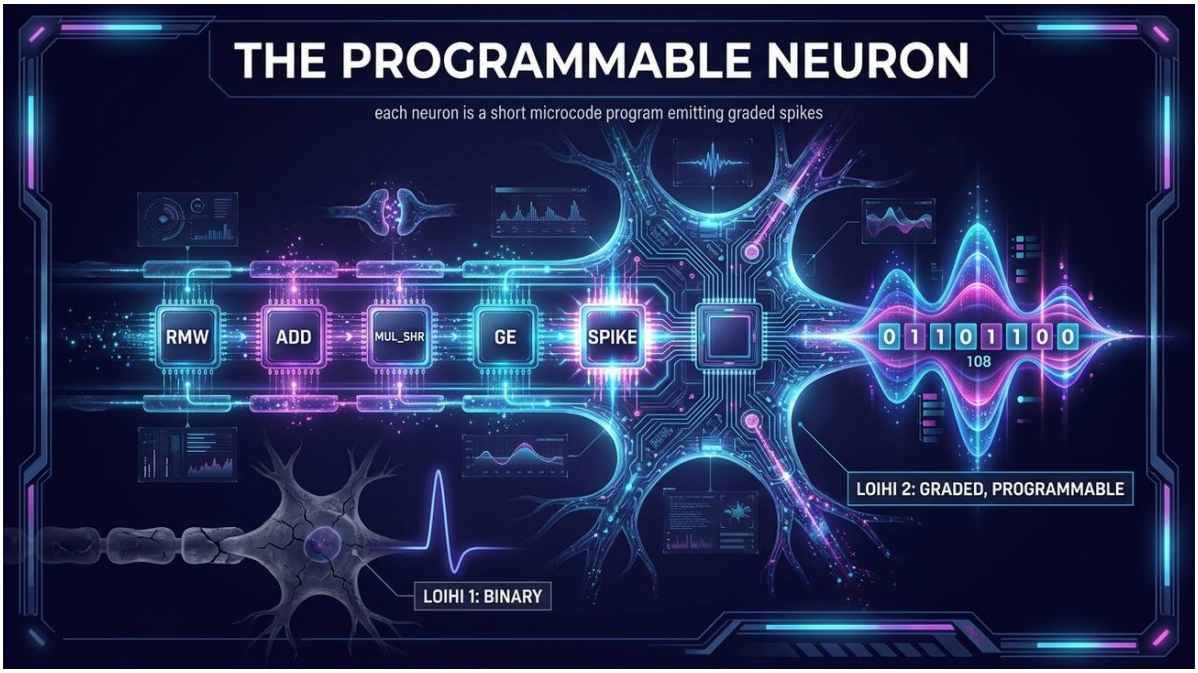


Figure 17. *Each Loihi 2 neuron is a short microcode program emitting graded (multi-bit) spikes.*

- able to advance a **chip-wide timestep in under 200 ns** — processing networks up to ~5,000× faster than biological neurons — and to **tile in 3D across up to 16,384 chips** via six inter-chip links.

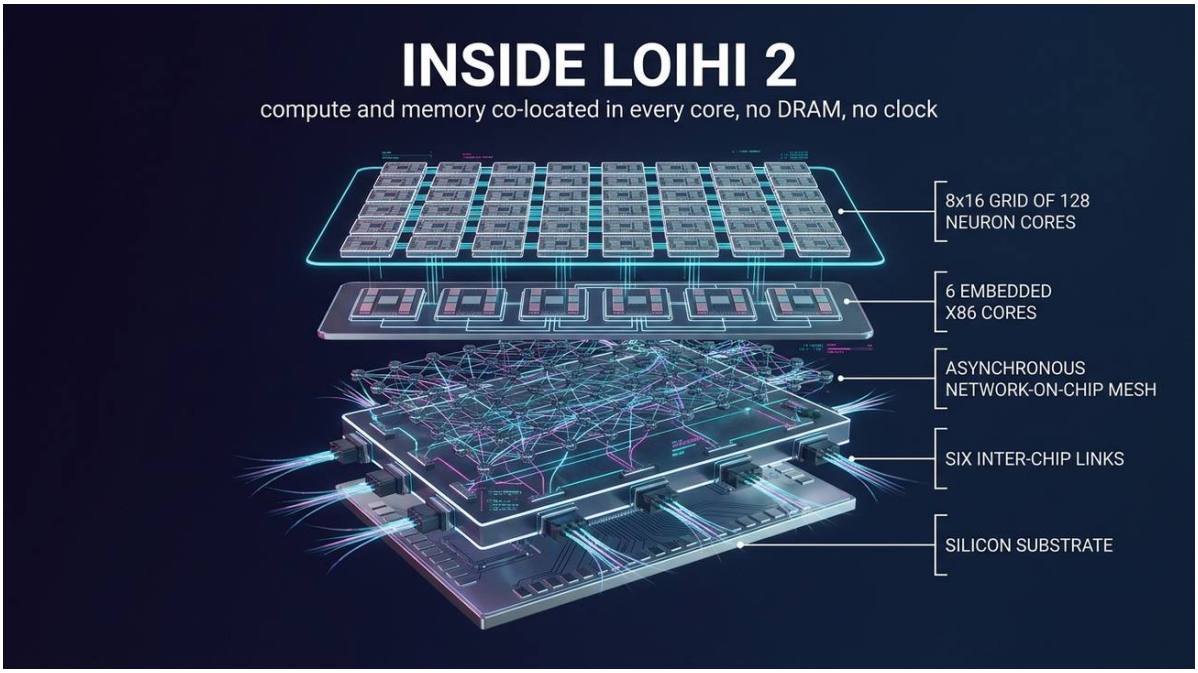


Figure 18. *Loihi 2 exploded: an 8x16 mesh of neuron cores, six embedded x86 cores, and inter-chip links.*

Every one of Calyx's four neuromorphic-native verbs maps onto a *named, documented* Loihi 2 capability: graded-spike dot products for cross-term cosines; latency-coded spiking search for Sextant k-NN; first-k-spikes winner-take-all for fusion and guarding; spike-wavefront propagation for Lodestar reach; programmable plasticity for an always-on, online-updating memory plane. Loihi 2 is, in effect, a physical realization of the association fabric this white paper posits. *(Full chip architecture, the microcode instruction set, fabrication detail, and the operation-by-operation mapping are documented in the companion intelinfo/ reference series, files 00–19.)*

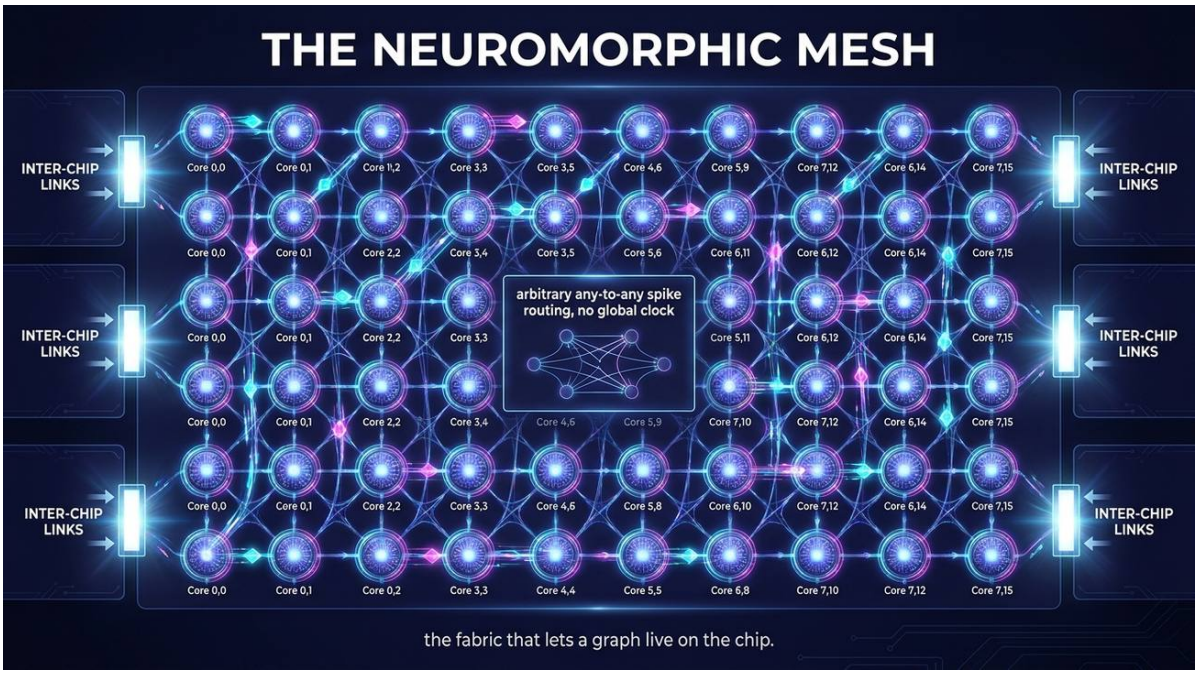


Figure 19. *The asynchronous mesh routes spike messages between arbitrary cores - the fabric that lets a graph live on the chip.*

7.4 The largest realization — Hala Point, and the precedent that already exists

Loihi 2 has already been assembled into **Hala Point** (announced April 2024, deployed at Sandia National Laboratories) — **the largest neuromorphic system in the world**:

- **1,152 Loihi 2 chips**, totaling **1.15 billion neurons** and **128 billion synapses** across **140,544 neuromorphic cores** and **2,300+ embedded x86 cores**;

Figure 20. *Hala Point: 1,152 Loihi 2 chips, 1.15 billion neurons, 2,600 watts - the world's largest neuromorphic system.*

- packaged in a **six-rack-unit chassis the size of a microwave oven**, drawing a **maximum of 2,600 watts**;

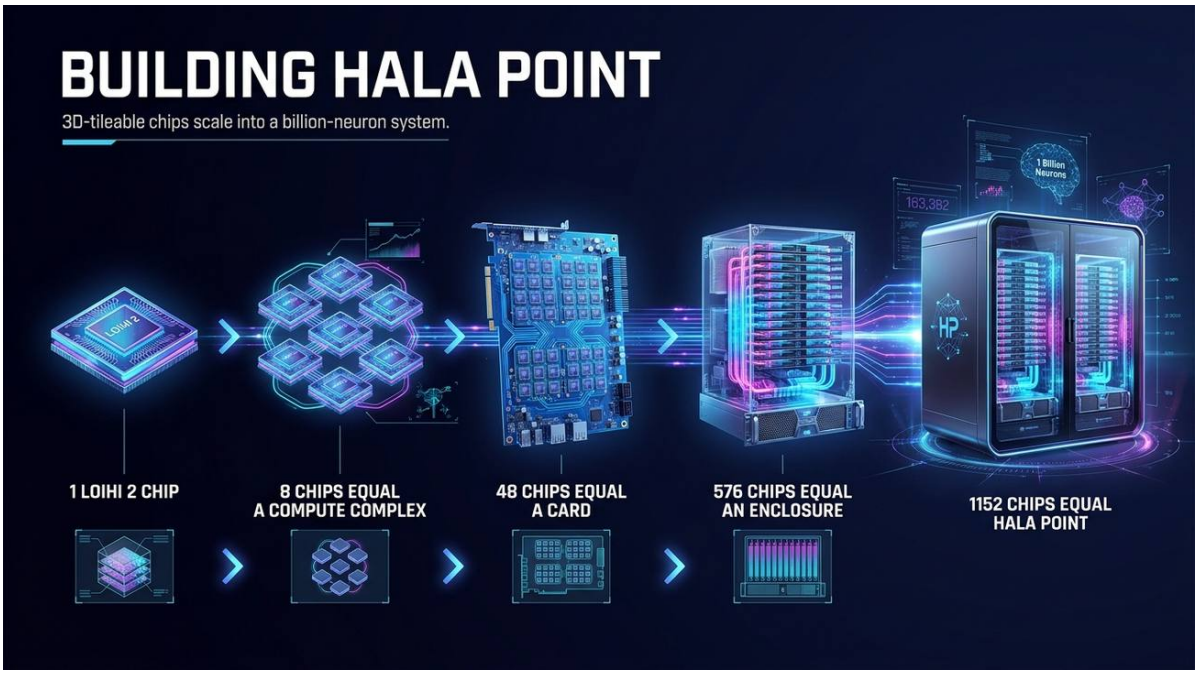


Figure 21. *3D-tileable chips scale 1 -> 8 -> 48 -> 576 -> 1,152 into the Hala Point system.*

- delivering **16 PB/s of memory bandwidth**, **3.5 PB/s inter-core** and **5 TB/s inter-chip** bandwidth, **20 petaops** peak, and ≥ **15 TOPS/W at 8-bit** on mainstream deep-learning workloads — *without* the input batching that delays real-time data on GPUs;
- a neuron count "roughly equivalent to an owl brain," running its full capacity **20× faster than a human brain** and up to **200× faster at lower capacity**.

Crucially, the central Calyx operation has **already been demonstrated on this hardware family**. On **Pohoiki Springs** — the prior flagship (768 first-generation Loihi chips, 100 million neurons, under 500 W) — Intel implemented an **approximate cosine k-nearest-neighbor search over more than one million high-dimensional vectors** that beat state-of-the-art CPU implementations on latency, index-build time, and energy *and supported adding new vectors to the index in* $O(1)$ *time* (Frady et al., 2020). That is, line for line, the "always-on memory with $O(1)$ streaming insertion" Calyx's Sextant aspires to. Calyx is not proposing an unproven mapping; it proposes to run a proven neuromorphic primitive at the heart of a grounded, provenance-backed database.

Figure 22. *The central Calyx operation - cosine k-NN with O(1) insertion - already demonstrated on the Loihi family (Pohoiki Springs).*

One hard constraint frames everything that follows: **Loihi 2 cannot be bought.** It is a research chip, distributed only to the **Intel Neuromorphic Research Community (INRC)** via a research cloud or loaned boards, and it is effectively **un-replicable** outside Intel — its captive pre-production Intel 4 node, the sole-source ASML EUV lithography it depends on, the $15–25 B fab it is made in, and Intel's proprietary clockless (asynchronous, quasi-delay- insensitive) design methodology together place a faithful copy beyond the reach of any actor without a leading-edge foundry. The substrate is extraordinary *and* access-constrained — a fact that becomes the crux of the speculative argument in §8–§11.

How the chip is actually made — and why that matters for Calyx. Understanding *why* Loihi 2 cannot be copied requires seeing what making it entails. Fabricating a leading-edge chip is among the most demanding industrial processes humanity has ever run: on the order of 500–1,000 discrete process steps and roughly 90 photomask layers, executed over a three-to-four-month production cycle in which a single particle of contamination can scrap an entire wafer.

Figure 23. *Making the chip: ~500-1,000 process steps and ~90 mask layers over a 3-4 month cycle.*

At the heart of it is **extreme-ultraviolet (EUV) lithography**, the only way to print features small enough for a node like Intel 4. A high-power CO_2 laser strikes microscopic tin droplets tens of thousands of times per second, vaporizing each into a plasma at roughly 220,000 °C that emits 13.5 nm light — radiation so readily absorbed by everything, including air and glass, that the entire optical path must run in vacuum and *reflect* off precision mirrors rather than passing through lenses.

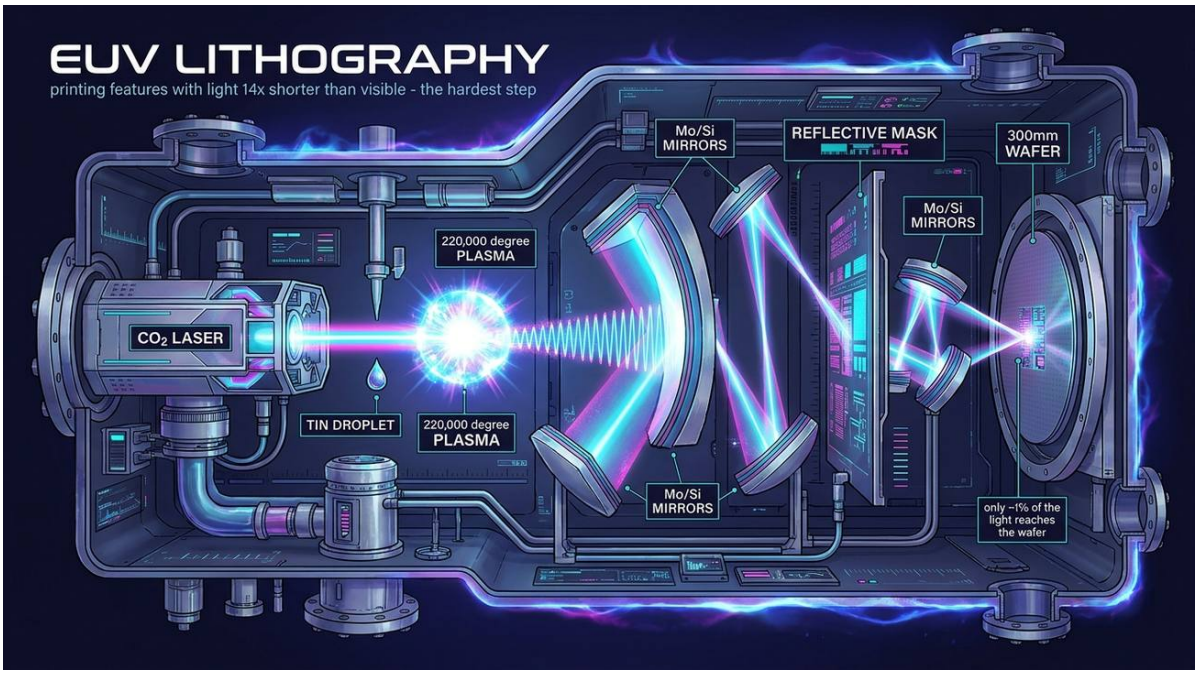


Figure 24. *EUV lithography: a CO2 laser vaporizes tin into 220,000-degree plasma emitting 13.5 nm light, focused by mirrors in vacuum.*

That light is delivered by the **ASML EUV scanner**, a single machine costing on the order of $200 million, weighing ~180 tonnes, drawing ~1 megawatt, and assembled from roughly 100,000 parts. ASML is its sole maker on Earth; without one, a leading-edge fab simply does not function. This is the first and hardest chokepoint in the entire enterprise.

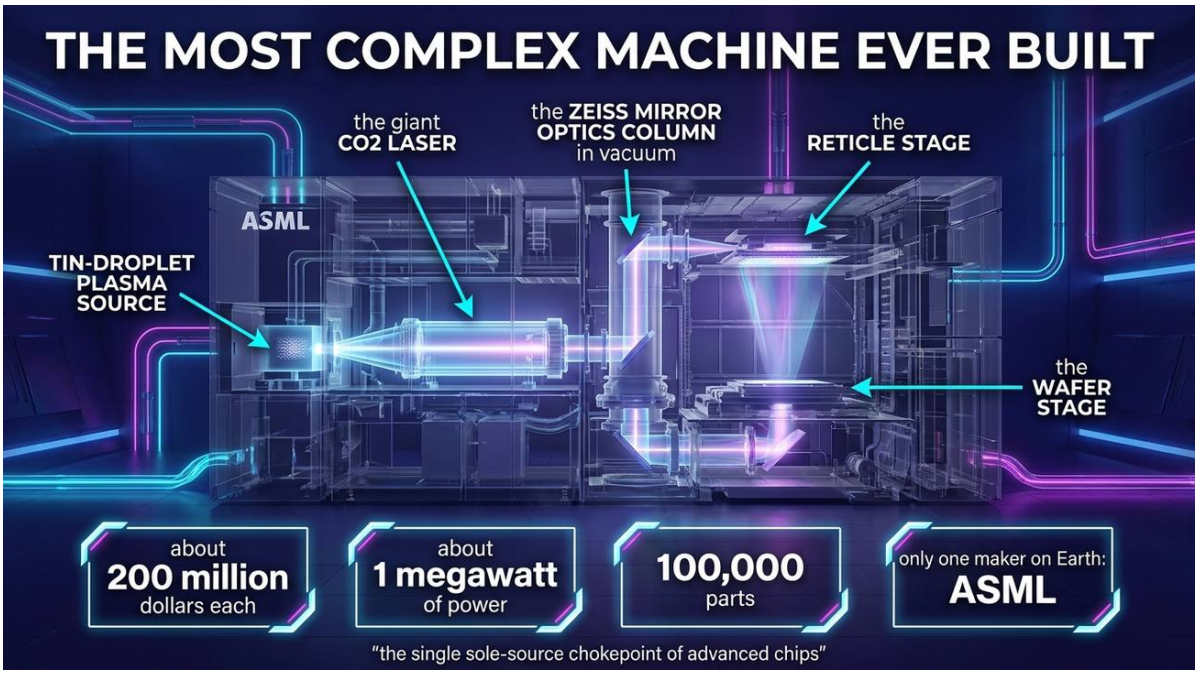


Figure 25. *The ASML EUV scanner: ~200 million dollars, ~1 megawatt, ~100,000 parts - a single sole-source chokepoint.*

The canvas is a **300 mm monocrystalline silicon wafer**, polished to near-atomic flatness, on which hundreds of identical dies are printed at once before being cut apart.

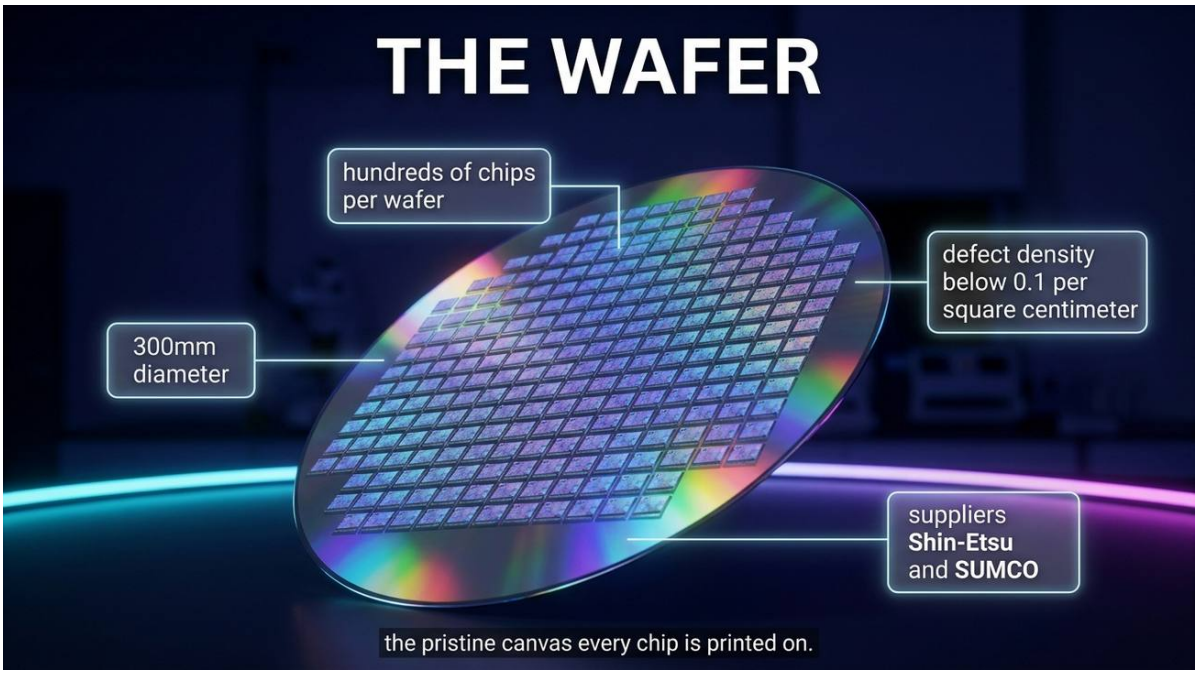


Figure 26. *The 300 mm silicon wafer: the pristine canvas on which hundreds of dies are printed.*

Each die is built up in two great phases. **Front-end-of-line (FEOL)** forms the billions of FinFET transistors in the silicon itself; **back-end-of-line (BEOL)** then stitches them together with more than a dozen stacked layers of copper interconnect — the microscopic "streets" that carry signal and power across the chip.

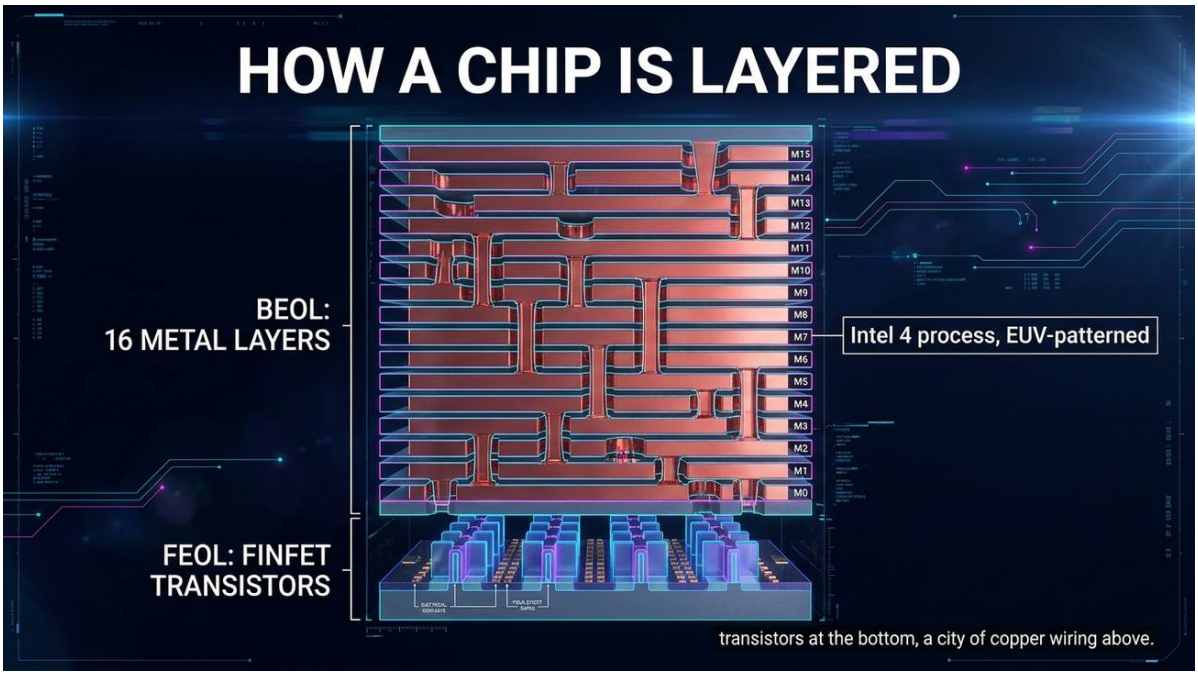


Figure 27. *A chip in cross-section: FinFET transistors below, sixteen copper interconnect layers above.*

All of this happens inside a **cleanroom fab**: a $15–25 billion ISO Class 1 facility whose air is thousands of times cleaner than a surgical theatre, where vibration is actively cancelled and more than a thousand tools run around the clock.

Figure 28. *The fab: a 15-25 billion dollar ISO Class 1 cleanroom of more than a thousand tools.*

Finished wafers are then **diced, tested, bonded, and packaged** — for Loihi 2, with the 3D-tileable inter-chip links that let many chips be assembled into a system like Hala Point — and put through final electrical test before a single working part exists.

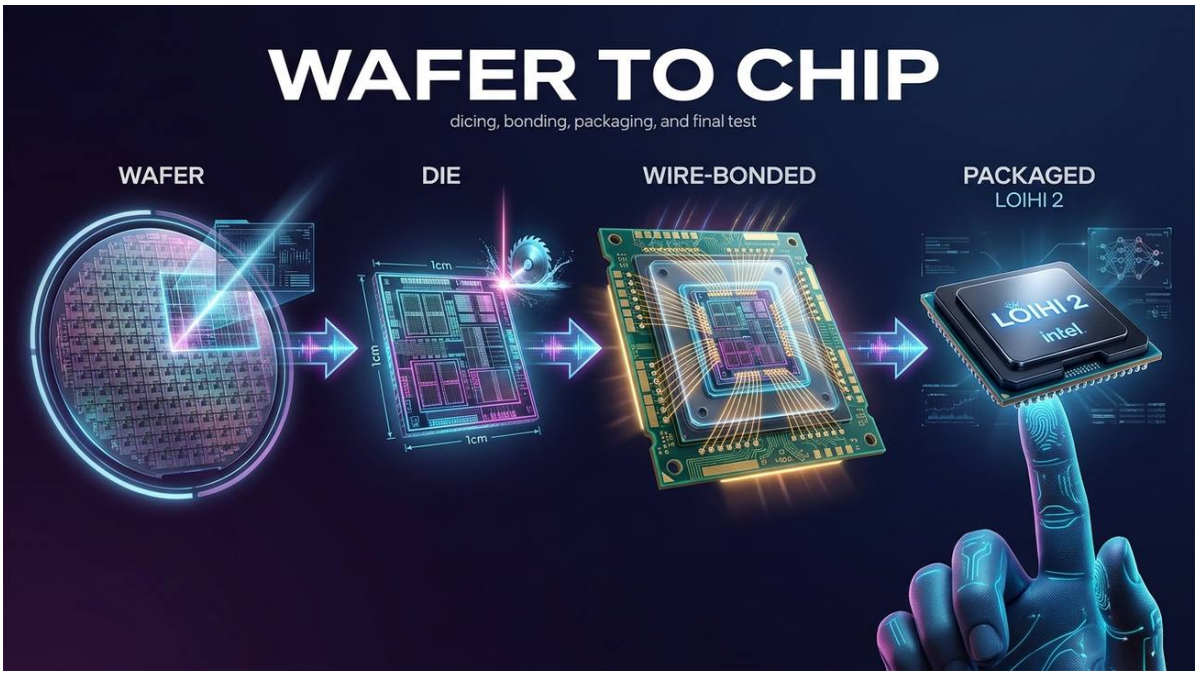


Figure 29. *Wafer to chip: dicing, bonding, packaging, and final test.*

Behind every one of these steps stands a **concentrated global supply chain**: ASML for the EUV scanners, Zeiss and Trumpf for the optics and drive laser, a handful of mostly Japanese firms for photoresists and ultra-pure chemicals, and only a few foundries anywhere capable of leading-edge production. Each link is a

single point of failure that no newcomer can quickly reproduce.

Figure 30. *A concentrated global supply chain: ASML, Zeiss/Trumpf, Japanese materials, and a few leading-edge fabs.*

For anyone who would *build* rather than rent, this defines a ladder of **feasibility tiers**: faithful leading-edge replication is out of reach, but FPGA emulation and mature-node multi-project shuttle runs capture most of the *architectural* value at a tiny fraction of the cost — precisely the path §12 develops.

Figure 31. *Feasibility tiers for a builder: FPGA emulation and mature-node shuttles capture the value without EUV.*

The conclusion, then, is not pessimism but clarity. **Loihi 2 itself is effectively un-replicable** — a captive process node, sole-source EUV, a multi-billion-dollar fab, and a proprietary clockless design flow, all at once — which is exactly why the right move for Calyx is not to copy it, but to build the leaner, mature-node chip described in §12.

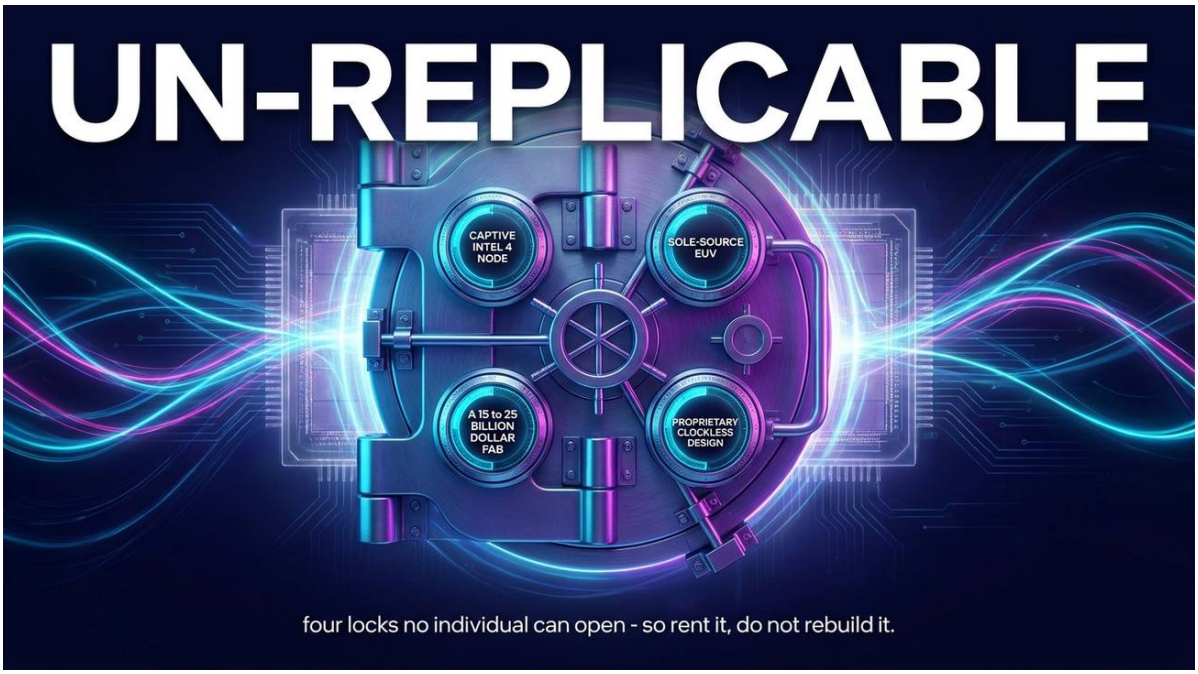
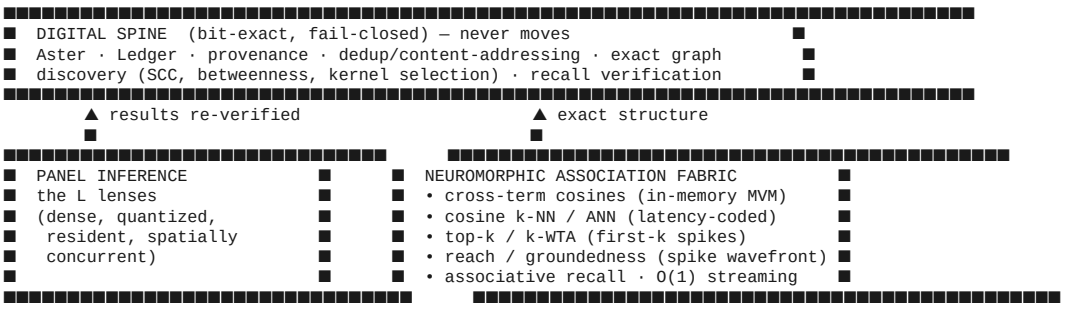


Figure 32. *Loihi 2 itself is effectively un-replicable: a captive node, sole-source EUV, a multi-billion-dollar fab, and a proprietary clockless flow.*

7.5 The hybrid architecture

Calyx does not become "a spiking network." It becomes a **hybrid** in which a precision-tolerant **association fabric** accelerates the ranking and similarity math, while a bit-exact **digital spine** preserves every trust guarantee:



The fabric is an **accelerator behind Calyx's existing trait boundaries** — the same backend, index, and lens contracts that already abstract CPU and GPU — and it is always paired with a digital verifier and fallback, so the fail-closed principle is preserved: any result the fabric cannot produce or verify is recomputed on the digital path or refused, never silently trusted.

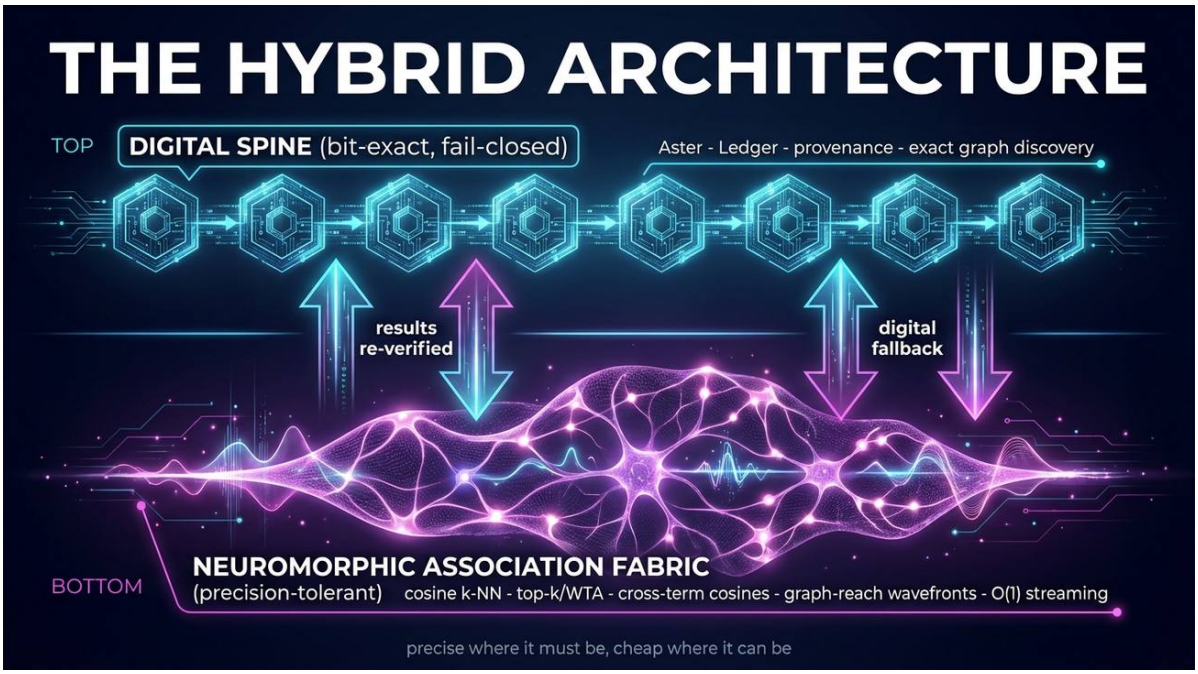


Figure 33. *The hybrid: a precision-tolerant neuromorphic association fabric behind a bit-exact, fail-closed digital spine.*

7.6 The mapped operations, in detail

Cross-term cosines (Loom) → in-memory matrix-vector multiply. With slot-vectors normalized to unit length, agreement is a dot product; an in-memory crossbar computes a matrix-vector multiply in a single physical step. The combinatorial $\binom{L}{2}$ field of agreements — the part of the workload that grows fastest with lens diversity — becomes essentially one resident pass over the panel, at hundreds of operations per unit energy. This is the single largest structural win, and it is exactly where Calyx's richness concentrates.

Cosine nearest-neighbor / ANN (Sextant) → latency-coded spiking search. Each stored slot-vector is one neuron; a query is broadcast as timed spikes; the strongest matches fire first, and the first k output spikes are the k nearest neighbors. The same temporal code that finds neighbors *also* yields the top-k for free. Critically, **insertion is** $O(1)$ — a new record adds one neuron — which is what makes continuous ingestion of massive, growing corpora tractable rather than a periodic, expensive index rebuild.

Winner selection (fusion, Ward) → k-winners-take-all. Reciprocal-rank fusion and per-slot guard decisions are ranking operations; spiking winner-take-all circuits select winners in the temporal domain as a natural byproduct of the search, at extremely low energy.

Association-graph reach (Lodestar answer path, groundedness) → spike wavefronts. Mapping graph nodes to neurons and edge weights to synaptic delays, a single injected spike propagates a wavefront whose earliest arrival times encode shortest-path costs. The kernel's answer path and the groundedness-distance-to-anchor computations map directly onto this — the operation neuromorphic substrates accelerate most dramatically relative to conventional hardware.

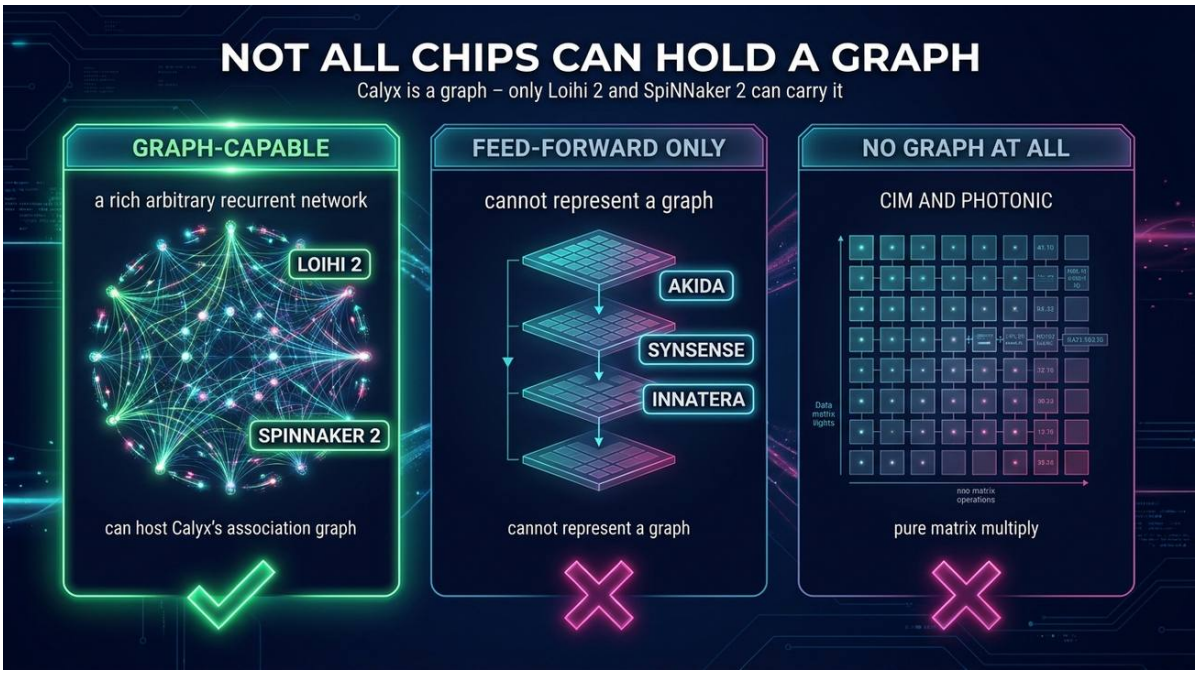


Figure 34. *Not all neuromorphic chips can hold a graph: only Loihi 2 and SpiNNaker 2 can carry Calyx's association graph.*

Associative recall (retrieval, dedup) → attractor / content-addressable memory. Recall-by- similarity from a noisy cue — the heart of retrieval, deduplication, and "reconstruct a voice or persona from a partial signal" — maps onto attractor and content-addressable memories that complete a stored pattern robustly and at $O(1)$.

Streaming ingest → event-driven, always-on processing. The fabric's event-driven nature makes continuous small updates cheap, which is precisely the workload of a living memory plane: many small deltas arriving forever, each inserted in $O(1)$, rather than bulk batch rebuilds.

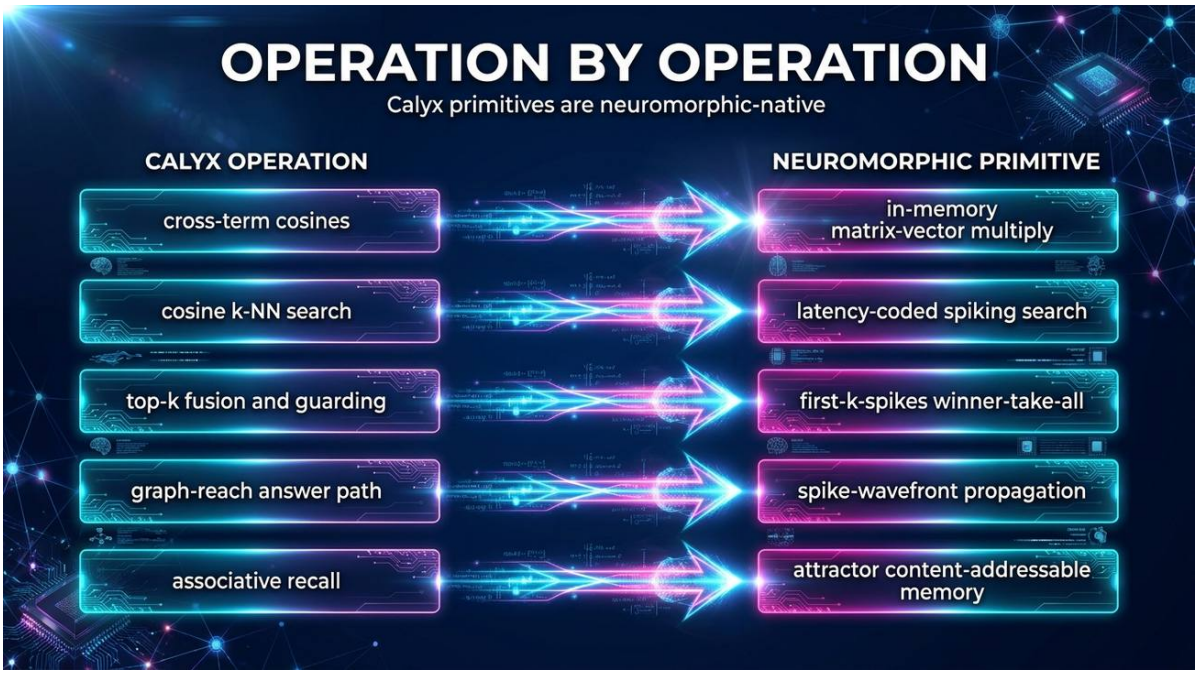


Figure 35. *Operation by operation, Calyx's primitives are neuromorphic-native.*

7.7 Encoding and the precision-tolerant subset

Spiking and in-memory substrates represent values with finite (typically modest) effective precision. This is not a limitation for the operations Calyx offloads, because **they are all rank- or threshold-based**: cosine ordering, reciprocal-rank fusion, recall, guard thresholds, and shortest-path *ordering* are robust to low-precision arithmetic — what matters is the ranking, not the last significant bit. Values are encoded efficiently (time-to-first-spike for spiking paths, low-bit quantization for in-memory paths), the fabric computes the ranking, and a lightweight digital check verifies a sampled subset. Anything that must be exact — content addresses, ledger hashes, deduplication identity, provenance, and the correctness-critical graph discovery — **stays on the digital spine.** The hybrid is thus precise where it must be and cheap where it can be.

7.8 What the substrate does *not* replace

Honesty is a Calyx principle, so the white paper states it plainly: neuromorphic acceleration targets the **similarity, ranking, and traversal** subset of the workload. The transformer **lenses themselves** remain dense, quantized models — run resident and spatially concurrent for throughput and energy, on the fastest dense-inference substrate available — until spiking encoders reach parity. The **exact information-theoretic estimation, the exact graph discovery (strongly-connected components, centrality, kernel selection), and the entire provenance spine** remain digital. The neuromorphic fabric is the engine for the part of Calyx that is *associative and approximate by nature*; the digital spine is the engine for the part that is *exact and load-bearing by mandate.* Neither replaces the other; together they let Calyx scale.

7.9 The scaling payoff

The combined effect on a planetary-scale, multi-domain, always-on Calyx:

- The **cross-term explosion** (the cost that grows fastest with lens diversity) collapses to near-constant-latency in-memory passes — so adding perspectives, the thing that makes Calyx more

powerful, stops making it disproportionately more expensive.

- **Similarity search and winner selection** run at a fraction of the energy, with $O(1)$ **insertion**, enabling continuous ingestion of unbounded, growing corpora.
- **Kernel answer paths and groundedness** ride spike wavefronts that scale with graph depth, not corpus size.
- The **digital spine** keeps every guarantee — provenance, grounding, fail-closed, exactness — intact.
- Energy scales with *activity*, not data volume, making always-on operation over massive datasets economically and physically feasible.

This is the path by which Calyx grows from a single-workstation discovery engine into a substrate for grounded intelligence over the world's data.

8. Speculation I — The hardware–software covenant: why Loihi 2 is, today, a Ferrari with no road

> ***Status of this and the next three sections.*** *Everything from here through §11 is explicitly speculative. To keep speculation disciplined rather than arbitrary, every claim is anchored to one or more established theories — results that, as far as human knowledge presently extends, are treated as true. The theories are named inline (Author, year) and collected in* ***Appendix C — Theory Base****. The argument is conditional: if Calyx's association-fabric thesis holds and if it is realized on Loihi-class silicon with serious backing, then the consequences below follow from those theories. Speculation grounded in theory is forecasting; speculation without it is fiction. We choose the former.*

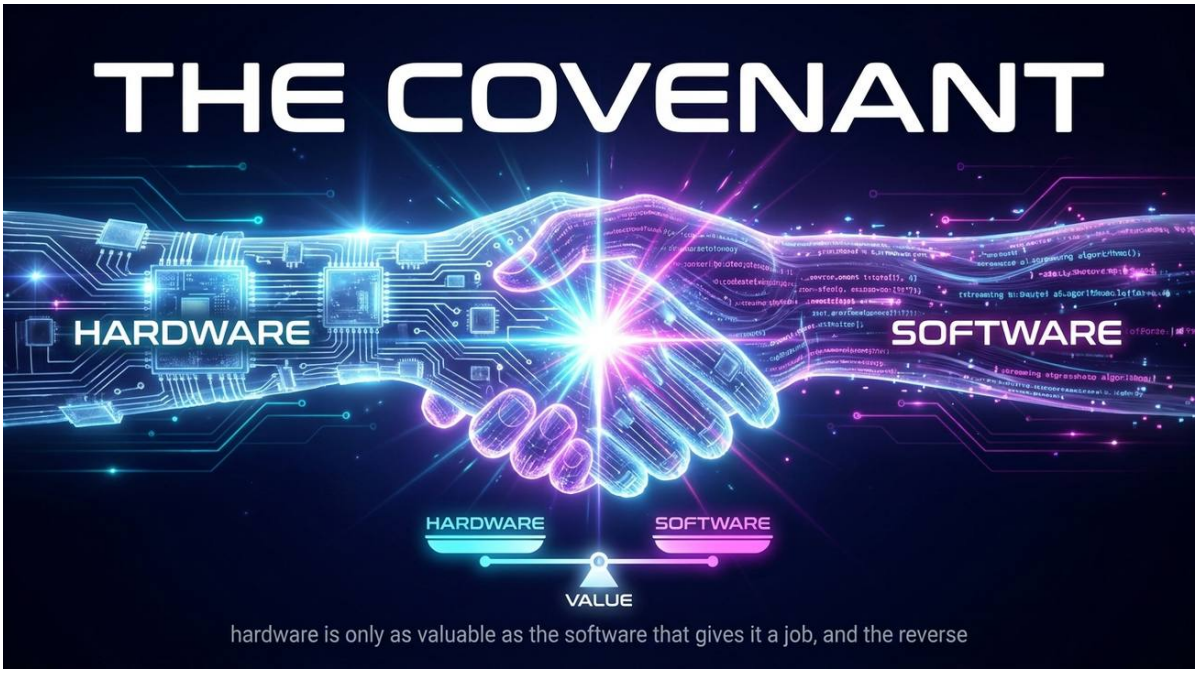


Figure 36. *The hardware-software covenant: each is only as valuable as the other.*

8.1 The core law: hardware has no intrinsic value — software confers it

A processor is inert sand until software gives people a reason to run it. In economic terms, hardware and software are **complementary goods**: the demand for one is a function of the availability and quality of the other (the logic that underwrites the entire "Wintel" platform era). A piece of software can be so valuable that

people buy the hardware *solely to run it* — a **killer application**. The archetype is **VisiCalc** (Bricklin & Frankston, 1979): the spreadsheet that turned the Apple II from a hobbyist curiosity into a business necessity, with a substantial share of 1979 Apple II sales attributable to that one program. The machine did not change; the *reason to own it* did.

Figure 37. *The killer-app law: a single piece of software can make people buy the hardware to run it.*

Platform economics sharpens this. A computing platform is a **two-sided market** (Rochet & Tirole, 2003): it must get both developers and users aboard, and the value of joining grows with the number already there — a **network effect** (Metcalfe, 1980s; with the Odlyzko–Tilly n·log n correction). Once a platform pulls ahead, **increasing returns and path dependence** (Arthur, 1989; David's QWERTY, 1985) can **lock in** its lead even against technically superior rivals. Value, in this world, is not a property of the silicon. It is a property of the *software ecosystem the silicon hosts.*

8.2 The proof of the law: how software revalued the GPU by a thousandfold

The cleanest demonstration in living memory is the GPU. The Graphics Processing Unit was built for one thing — drawing triangles fast for **video games** (NVIDIA's GeForce 256, 1999). The silicon was a parallel arithmetic engine, but its *value* was capped by its application: gaming. Then the application changed.

- In **November 2006**, NVIDIA shipped **CUDA** (on the G80/GeForce 8800), exposing the GPU's parallel cores to general-purpose code. Jensen Huang called general-purpose GPU computing a **"zero-billion-dollar market"** — there was no demand to point at. NVIDIA poured roughly **$12 billion of R&D into CUDA between 2006 and 2017** while its market capitalization *fell* from ~$12 B at CUDA's launch to ~$2–3 B in the financial crisis and stagnated for years. Wall Street hated it.
- In **2012**, **AlexNet** (Krizhevsky, Sutskever & Hinton) won the ImageNet competition by a historic margin, trained in 5–6 days on **two GeForce GTX 580 GPUs**. Deep learning had found its substrate, and the substrate was the gaming card.
- The software ecosystem then compounded: cuDNN, TensorFlow, PyTorch — all targeting CUDA first. The result is the most dramatic revaluation of a piece of hardware in history. The *same kind of silicon* that

drew game frames became the engine of the AI economy. NVIDIA crossed **$1 trillion (June 2023), $2 trillion (Feb 2024), $3 trillion (June 2024)**, and became the **first company to touch $4 trillion (9 July 2025)**.

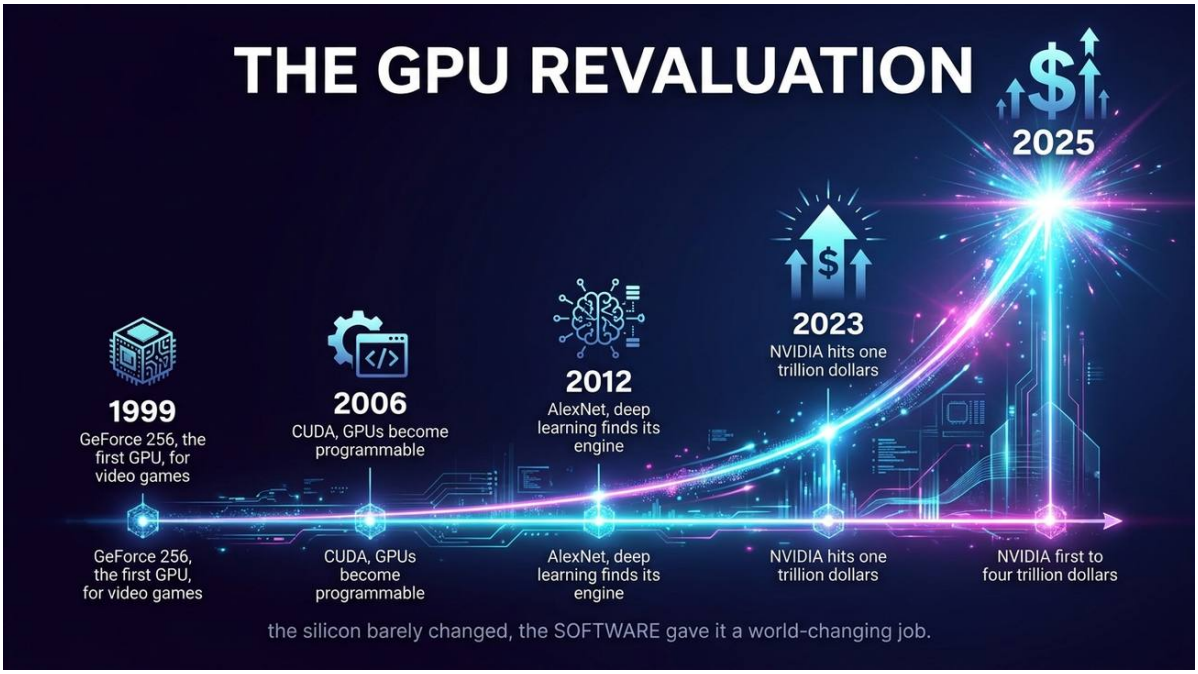


Figure 38. *Software revalued the GPU a thousandfold: from a gaming chip (1999) to the first four-trillion-dollar company (2025).*

The lesson is exact and load-bearing for this paper: **the GPU did not become valuable because the hardware improved; it became valuable because software — first CUDA, then deep learning — gave the existing hardware a world-changing job.** The silicon was the same shape in 2006 and

2023. What moved by orders of magnitude was the *software that ran on it.* (This is GPT theory in action — Bresnahan & Trajtenberg, 1995: a general-purpose technology's value is realized through **innovational complementarities** with downstream software.)

8.3 The present asymmetry: the most advanced neuromorphic hardware on Earth, running essentially no valuable software

Now place Loihi 2 in that frame. By every architectural measure in §7 and the companion intelinfo/ series, Loihi 2 is the most advanced neuromorphic processor in existence — 1M-neuron, graded-spike, fully-programmable, three-factor-learning, sub-200 ns silicon, scaled to the 1.15-billion-neuron Hala Point. And yet, in mid-2026, **there is no software running on neuromorphic hardware that the broad market considers valuable.** There is no neuromorphic VisiCalc, no neuromorphic AlexNet. The research literature (the Loihi survey, Davies et al. 2021) is candid: conventional deep nets show "modest if any benefit" on Loihi; the wins are real but confined to niche sparse/temporal demonstrations. The market has priced neuromorphic accordingly — as a research curiosity, not a platform.

Figure 39. *The present asymmetry: the most advanced neuromorphic hardware on Earth runs almost no valuable software - the GPU's 2006 condition.*

This is precisely the **GPU-in-2006 condition**: extraordinary parallel hardware sitting in a "zero-billion-dollar market" because the killer application has not yet arrived. It is also the **von Neumann-bottleneck opportunity** (Backus, 1978) and the **end-of-Dennard-scaling pressure** (Dennard 1974; breakdown ~2005) that make non-von-Neumann substrates strategically necessary, combined with **Carver Mead's** original neuromorphic thesis (Mead, 1990) that brain-style computation is orders of magnitude more energy-efficient for relative-valued problems. The hardware case for neuromorphic is overwhelming. The **software** case — the reason to *own* the hardware — is the missing half of the covenant.

8.4 The claim: Calyx-with-SNN is a candidate killer application for neuromorphic silicon

Here is the speculative thesis in one sentence: **Calyx is, to neuromorphic hardware, what deep learning was to the GPU — the first software whose value is realized *specifically and disproportionately* by exactly what the hardware does natively.**

Figure 40. *The claim: Calyx is to neuromorphic hardware what deep learning was to the GPU - the matching killer application.*

The fit is not generic. §7 showed that Calyx's four verbs map operation-for-operation onto Loihi 2's native primitives: cross-term cosines → graded-spike matrix-vector products; Sextant k-NN → latency-coded spiking search; fusion/guarding → first-k-spikes winner-take-all; Lodestar reach → spike-wavefront propagation; always-on ingest → $O(1)$ event-driven insertion; associative recall → attractor/content-addressable memory (Hopfield, 1982; Kanerva's sparse distributed memory and hyperdimensional computing, 1988/2009). The k-NN-with-$O(1)$-insertion piece is *already demonstrated* on the hardware family (Pohoiki Springs, Frady et al. 2020). Calyx is not a workload that *tolerates* neuromorphic hardware; it is a workload that is **starved on von-Neumann hardware and liberated on neuromorphic hardware** — the exact relationship deep learning had with the GPU.

If that fit is real, then by the hardware–software covenant (§8.1) Calyx is the software that could **confer value on Loihi 2** — the missing killer app that turns the most advanced neuromorphic processor on the planet from a research chip into a platform. The remainder of the speculation explores what follows if that happens.

> ***Theory anchors (§8):*** *complementary goods / killer app (VisiCalc; Wintel); two-sided markets (Rochet & Tirole 2003); network effects (Metcalfe; Odlyzko–Tilly); increasing returns & lock-in (Arthur 1989; David 1985); GPT "engines of growth" & innovational complementarities (Bresnahan & Trajtenberg 1995); von Neumann bottleneck (Backus 1978); end of Dennard scaling (Dennard 1974); neuromorphic energy thesis (Mead 1990); attractor & associative memory (Hopfield 1982; Kanerva 1988/2009); the GPU/CUDA/AlexNet/NVIDIA record.*

9. Speculation II — The Intel scenario: what happens if Loihi 2 runs Calyx, fully operational, with Intel's backing

Assume the strong case: Intel adopts Calyx, the integration is engineered to full production quality on Loihi 2 (and its successors), and Intel puts its commercialization weight — fab access, INRC ecosystem,

software-stack support, go-to-market — behind it. What follows is the chain of consequences each established theory predicts.

9.1 The mechanism: Calyx becomes neuromorphic's "CUDA moment"

The GPU was revalued not by CUDA alone and not by deep learning alone, but by their *conjunction* on hardware that uniquely suited them. The Intel scenario is the structural analogue: **Loihi 2 (the hardware) + Calyx-with-SNN (the killer app) + Intel's backing (the ecosystem flywheel).** GPT theory (Bresnahan & Trajtenberg, 1995) is explicit that a general- purpose technology and its applications suffer a coordination failure — "too little, too late" innovation — when left to **arms-length markets**; the value is unlocked only when the GPT provider and the application co-invest. Intel's backing *is* that co-investment. This is the single most important reason the scenario requires Intel, not just Calyx: the covenant needs **both** halves committed, or GPT theory predicts the opportunity is left on the table.

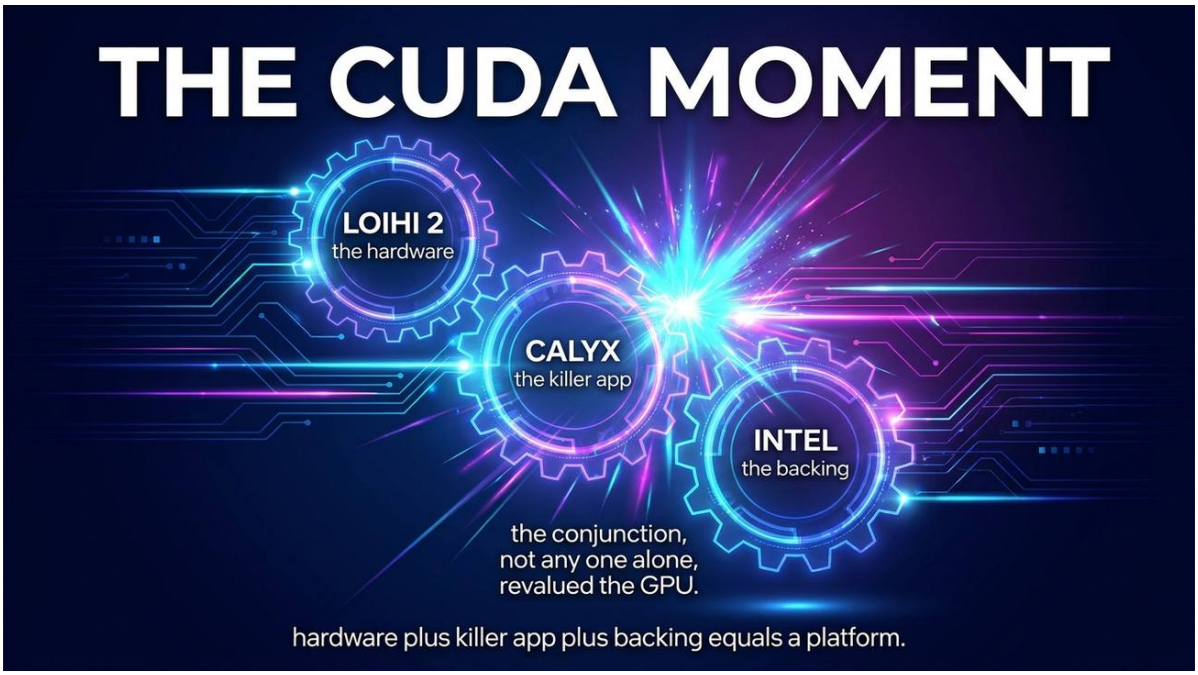


Figure 41. *The conjunction that revalues hardware: the chip, the killer app, and the backing, together.*

9.2 The flywheel: how value compounds once it starts

Once Calyx makes Loihi 2 worth owning, the two-sided-market and increasing-returns machinery (Rochet & Tirole 2003; Arthur 1989) engages:

1. A valuable application (grounded, energy-efficient, always-on association) draws developers and customers to the neuromorphic platform.
2. More developers build more association-native applications on the Calyx/Loihi substrate.
3. More applications make the hardware more valuable, drawing more developers — a **network effect** (Metcalfe).
4. Volume drives down unit cost along the **experience curve** (Wright's Law, 1936; Nagy et al. 2013 found it the best predictor of cost decline across 62 technologies), which widens the market further.
5. The accumulating software, tooling, talent, and data create a **CUDA-style moat** — the same path-dependent lock-in (David 1985) that protects NVIDIA today, but built around Calyx + Loihi + Intel

instead.

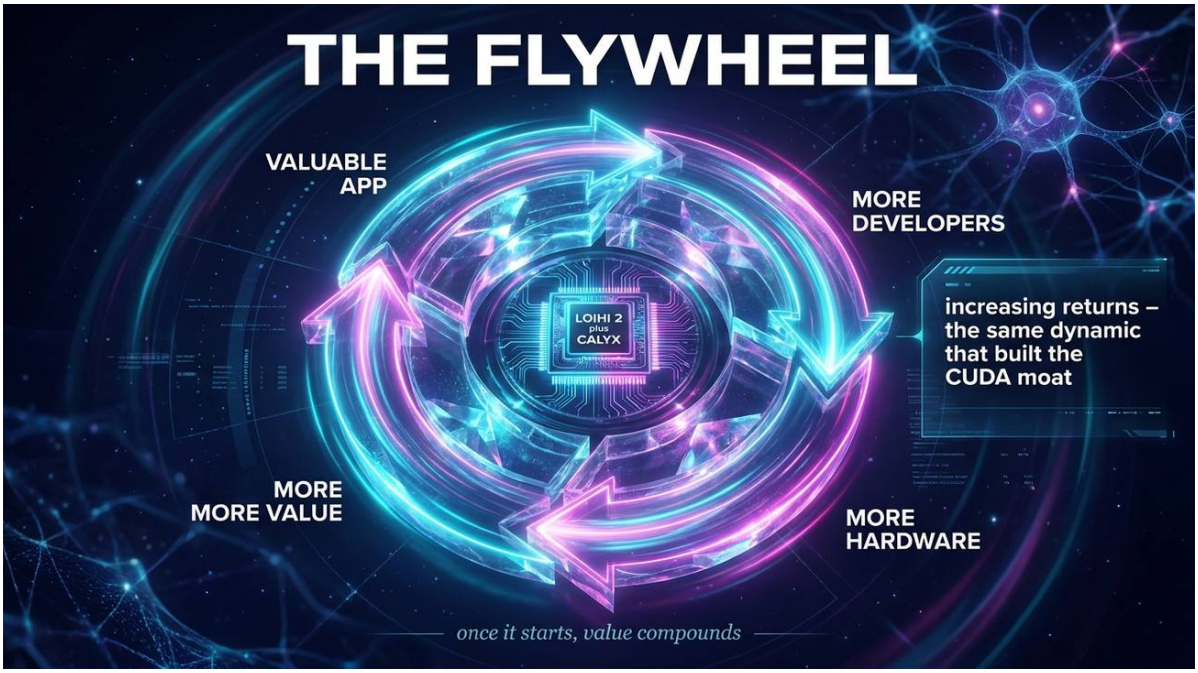


Figure 42. *The two-sided-market flywheel: apps, developers, value, and hardware reinforce one another.*

The GPU took ~6 years from CUDA (2006) to AlexNet (2012) and another decade to $4T. Diffusion follows an **S-curve** (Rogers, 1962), and there is a **chasm** (Moore, 1991) between visionary early adopters and the pragmatic majority that many technologies never cross. The flywheel is not instantaneous; **Amara's Law** (overestimate short-run, underestimate long-run) and the **Gartner Hype Cycle** both warn that the near-term will disappoint relative to the eventual magnitude. But the *direction* is what the theories agree on.

9.3 Why this is a Christensen disruption, not a head-on GPU fight

Calyx-on-Loihi does **not** win by beating NVIDIA at dense matrix multiplication — it would lose that fight, and §7.8 says so plainly. It wins on a **different value axis**: energy per event, latency at batch-size-1, $O(1)$ streaming insertion, always-on continual learning, and grounded provenance. This is the textbook shape of **disruptive innovation** (Christensen, 1995): a technology that is *inferior on the incumbent's metric* (raw dense FLOPs) but *superior on a new metric the market will come to value* (sustainable, real-time, continually-learning, traceable association), entering from an underserved niche and improving until it redefines the mainstream. The incumbents (GPU-based vector DBs and RAG stacks) are rationally locked into serving their best customers' demand for ever-bigger batch training — the exact blind spot Christensen describes.

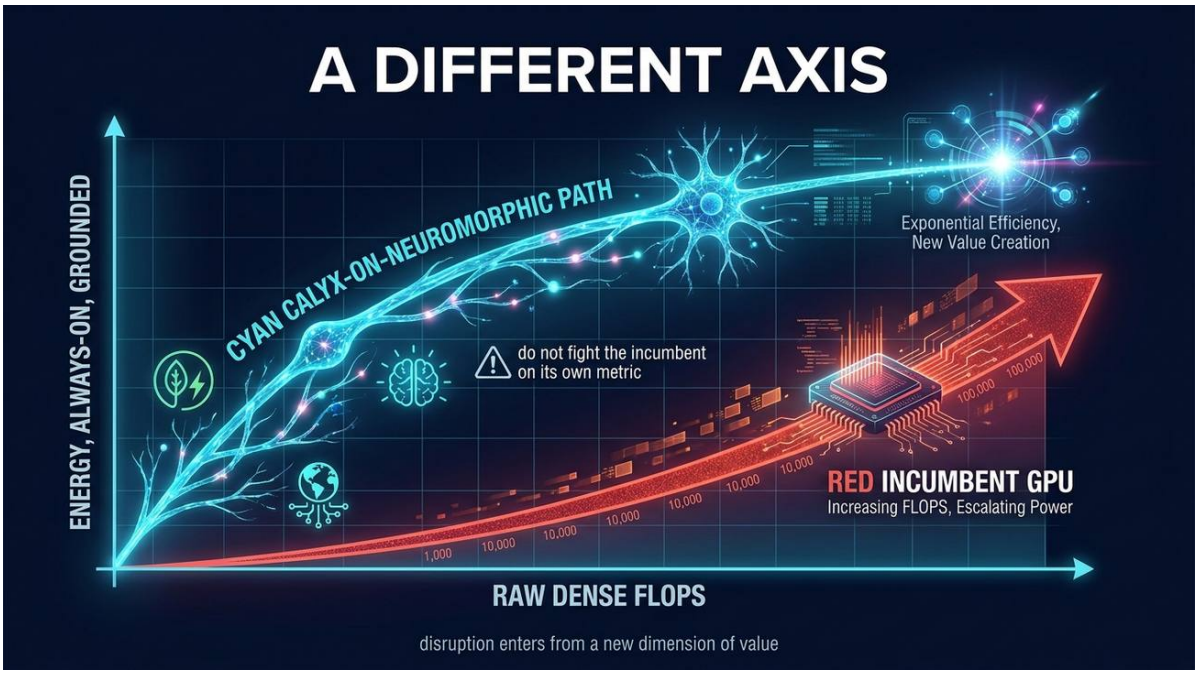


Figure 43. *Disruption enters on a new axis (energy, always-on, grounded), not by fighting the incumbent on dense FLOPs.*

9.4 What shifts in the world if it comes together

The AI energy equation. AI's energy trajectory is the macro backdrop: data-centre electricity was ~415 TWh in 2024 (~1.5% of world supply) and is projected by the IEA to roughly double to ~945 TWh by 2030, with AI-focused demand quadrupling; a single GPT-3-class training run was estimated at ~1,287 MWh / ~552 t CO_2e (Patterson et al. 2021). Against **Landauer's principle** (1961 — every irreversible bit-erasure costs at least $kT\ln 2$) and **Koomey's Law** (efficiency doubling ~every 1.5 years), neuromorphic's event-driven, memory-co-located, sparse-activity model is a structurally different point on the energy curve. If Calyx moves the precision-tolerant association hot loop — similarity, ranking, graph-reach, streaming memory — off batched GPUs and onto Loihi-class silicon at the demonstrated ~100×-energy / 15-TOPS/W class, the **unit economics of always-on, grounded AI memory change**, and with them the viability of continuously-learning agents that today are financially and thermodynamically prohibitive.

Intel's strategic position. Intel in the mid-2020s lacks an AI software moat comparable to CUDA. Calyx is a candidate to *be* that moat — on an axis (neuromorphic + grounded association) where Intel already owns the world's best hardware (Loihi 2, Hala Point) and the ecosystem (INRC, Lava). **Schumpeterian creative destruction** (1942) and **Perez's techno-economic paradigm** framework (2002) both describe how a new paradigm lets a challenger leap an incumbent rather than chase it. Loihi 2, today an un-sellable research chip, could become — in the strongest form of the claim — *the most valuable piece of hardware on the planet*, not because the silicon changed, but because the software finally gave it a job only it can do. This is the GPU revaluation, run again, on a substrate Intel controls end-to-end.

The shape of computing. If the precision-tolerant, associative, always-on layer of AI migrates to neuromorphic while the exact/dense layer stays on GPUs and CPUs, the industry bifurcates into a **hybrid substrate** — the same way CPUs and GPUs coexist today. Calyx's own architecture (digital spine + neuromorphic fabric) is a microcosm of that macro split. This is a **paradigm shift** in Kuhn's sense (1962): not a faster version of the old thing, but a different organizing principle (event-driven, sparse, co-located,

grounded) for the part of computing that is associative by nature.

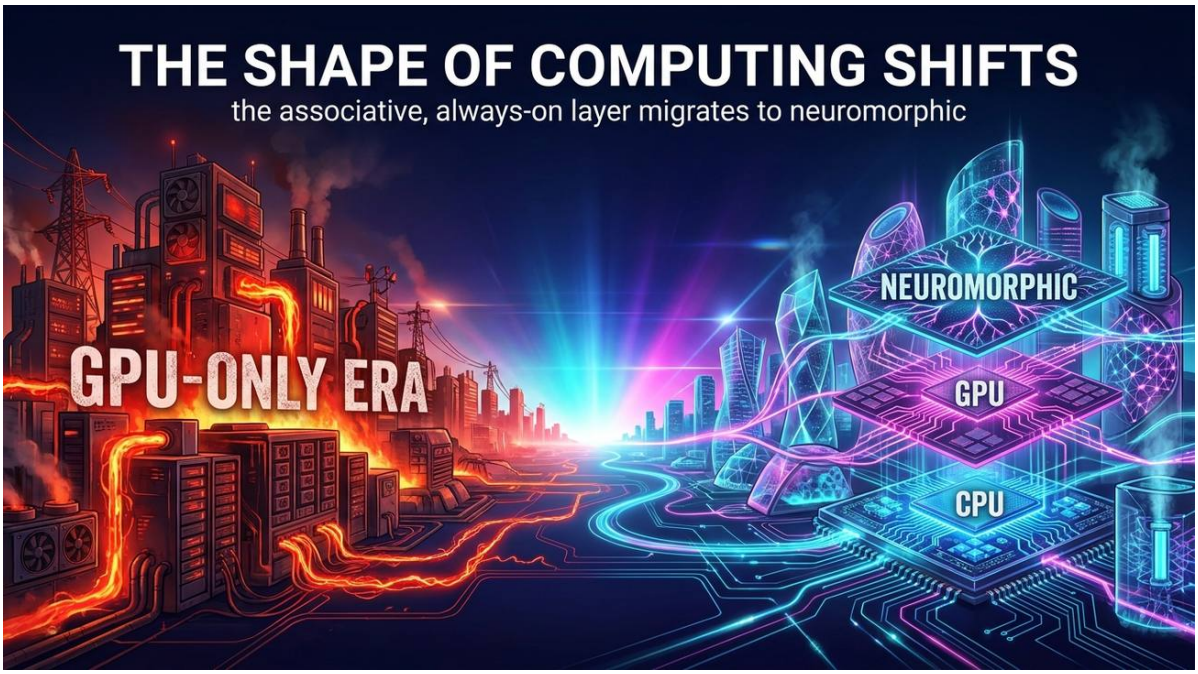


Figure 44. *The shape of computing bifurcates: the associative, always-on layer migrates to neuromorphic.*

The redistribution of power (a great-equalizer reading). The deepest world-shift may not be industrial but political, and it is developed in full in §17. If grounded, frontier-grade intelligence becomes something an ordinary person can run *locally* — air-gapped, private, from a wall outlet — then the asymmetry of *knowledge* that Shoshana Zuboff (*The Age of Surveillance Capitalism*, 2019) identifies as the engine of modern power begins to close. Joseph Nye's *The Future of Power* (2011) calls this **power diffusion** — capability migrating away from states and large institutions toward individuals and networks — and judges it potentially "a greater threat than power transition." The plausible macro-consequence is a partial re-leveling: governments and large institutions, long shielded by an epistemic monopoly, would have to operate as though any citizen can find, model, and verify what they do, restoring a deterrent floor of accountability of exactly the kind §17 argues for — the geopolitics of "possession changes treatment," scaled down from nuclear states to informed citizens. Two honest brakes apply. Michael Horowitz's *The Diffusion of Military Power* (2010) shows that financially intensive capabilities *concentrate* rather than diffuse — so this future depends on the capability genuinely becoming cheap and locally runnable (the §12 chip thesis), not merely existing — and diffusion, as Nye stresses, "does not mean equality of power." The shift is therefore a change in the *shape* of the power map, not its abolition; but when the prior shape was a near-monopoly, a change in shape is itself among the most consequential things this technology could do.

Figure 45. *When intelligence diffuses, power redistributes: the map of power changes as grounded intelligence spreads - though, as Horowitz and Nye warn, diffusion is not automatic equality.*

Theory anchors (§9): *GPT coordination failure & co-investment (Bresnahan & Trajtenberg 1995); two-sided markets (Rochet & Tirole 2003); increasing returns / lock-in (Arthur 1989; David 1985); network effects (Metcalfe); experience curve (Wright 1936; Nagy et al. 2013); diffusion S-curve & chasm (Rogers 1962; Moore 1991); Amara's Law & Hype Cycle; disruptive innovation (Christensen 1995); creative destruction (Schumpeter 1942); techno-economic paradigms (Perez 2002); Kuhn paradigm shift (1962); Landauer (1961); Koomey (2011); IEA/AI energy data; Patterson et al. (2021); power diffusion & cyberpower (Nye 2011); diffusion of military power (Horowitz 2010); knowledge–power asymmetry (Zuboff 2019).*

10. The ledger of consequences — twelve pros and twelve cons, each backed by theory

If the Intel scenario (§9) comes together, the consequences cut both ways. Below are twelve arguments for and twelve against, stated at full strength. Each is anchored to named theories so the reasoning can be audited rather than merely asserted.

Figure 46. *The ledger of consequences: large upside (pros) weighed against path risk (cons).*

10.1 The case *for* — twelve pros

PRO 1 — A genuine killer application would finally exist for neuromorphic computing. The field's central problem is the absence of valuable software (§8.3). Calyx supplies a workload that is *starved* on von-Neumann hardware and *liberated* on neuromorphic hardware, satisfying the complementary-goods/killer-app condition that historically precedes platform take-off. *Anchors: killer app/complementary goods (VisiCalc; Wintel); GPT innovational complementarities (Bresnahan & Trajtenberg 1995); the GPU/CUDA precedent.*

PRO 2 — Order-of-magnitude energy savings on the association layer of AI. Moving similarity, ranking, graph-reach, and streaming memory to event-driven, memory-co- located silicon attacks exactly the data-movement cost that dominates conventional accelerators, at a demonstrated ~100×-energy / 15-TOPS/W class. Against a doubling AI-energy trajectory (IEA: ~415→~945 TWh by 2030), this is materially decarbonizing for a fast-growing load. *Anchors: Landauer's principle (1961); Koomey's Law (2011); von Neumann bottleneck (Backus 1978); Mead (1990); IEA Energy & AI; Patterson et al. (2021).*

PRO 3 — Continual, online learning becomes economically feasible. Loihi 2's three-factor on-chip plasticity + $O(1)$ insertion makes "learn from new data forever" cheap, instead of the periodic, gigawatt-hour full-retraining cycle. Agents and memory planes that must adapt in real time become viable. *Anchors: Hebbian plasticity (Hebb 1949); predictive coding (Rao & Ballard 1999); free-energy minimization (Friston 2010); Koomey.*

PRO 4 — A new value axis that sidesteps NVIDIA's moat instead of charging it. Calyx-on-Loihi competes on energy/latency/always-on/grounding, not dense FLOPs — a classic disruption from an underserved axis rather than a doomed frontal assault. *Anchors: disruptive innovation (Christensen 1995); creative destruction (Schumpeter 1942); No Free Lunch (Wolpert & Macready 1997 — specialized architectures win on matched problems).*

PRO 5 — A defensible, compounding moat for whoever gets there first. Software ecosystem + tooling + talent + data on a co-designed hardware substrate generate increasing returns and lock-in — the same dynamics that make CUDA near-unassailable, here available to Intel on an axis it already leads. *Anchors: increasing returns & path dependence (Arthur 1989; David 1985); two-sided markets (Rochet & Tirole 2003); network effects (Metcalfe).*

PRO 6 — Grounded, provenance-backed AI memory addresses the trust crisis. Calyx's fail-closed, hash-chained, no-flatten discipline yields results an agent or scientist can *trace and verify*, directly countering hallucination/ungrounded generation — a property the market increasingly demands for high-stakes use. *Anchors: Shannon information theory & the data-processing inequality (1948 — honest ceilings on recoverable information); Goodhart's Law (the discipline of not gaming the measure).*

PRO 7 — Cross-domain scientific discovery at a new energy/scale point. The association engine's purpose — surfacing grounded, non-obvious target↔molecule↔disease↔ literature links — becomes runnable always-on over massive multi-domain corpora, the regime where the most valuable undiscovered links hide. *Anchors: GPT generalized productivity gains (Bresnahan & Trajtenberg 1995); sparse coding efficiency (Olshausen & Field 1996); Kanerva associative memory (1988).*

PRO 8 — Revaluation of an otherwise stranded national asset (Loihi/Hala Point). Intel has spent a decade and built the world's largest neuromorphic system with little commercial return. A killer app converts sunk research into a strategic platform — and, given US fab/EUV positioning, a sovereign-capability advantage. *Anchors: creative destruction (Schumpeter 1942); techno-economic paradigm shift & "turning point" (Perez 2002).*

PRO 9 — A flywheel that lowers cost and broadens access over time. Once volume begins, Wright's-Law learning curves push unit cost down predictably, moving neuromorphic from research-only toward affordable edge and datacenter deployment. *Anchors: Wright's Law (1936); Nagy et al. (2013); diffusion S-curve (Rogers 1962).*

PRO 10 — Architectural honesty: the hybrid keeps exactness where it must live. Calyx never bets correctness on approximate hardware — the digital spine stays bit-exact, the fabric only accelerates rank/threshold work, with digital verify/fallback. This makes adoption *safe* (no silent wrong answers), lowering the barrier to enterprise/high-stakes use. *Anchors: data-processing inequality (Shannon); Collingridge's "keep decisions reversible" mitigation (1980); fail-closed engineering.*

PRO 11 — A second engine of computing growth as Moore/Dennard fade. With Dennard scaling ended and Moore's cadence slowing, the industry needs new efficiency sources from architecture, not lithography. A grounded-association neuromorphic layer is one such GPT-class engine. *Anchors: end of Dennard scaling (Dennard 1974; ~2005 breakdown); Moore's Law slowdown (Moore 1965); GPT engines of growth (Bresnahan & Trajtenberg 1995); Bitter Lesson caveat (Sutton 2019 — compute-scaling methods win, and energy gates compute).*

PRO 12 — Alignment with how brains actually compute. Calyx's primitives (association, winner-take-all, attractor recall, sparse event-driven coding) and Loihi's substrate are both drawn from neuroscience; matching algorithm to the problem's native structure is where all advantage comes from. *Anchors: No Free Lunch (Wolpert & Macready 1997); neuromorphic engineering (Mead 1990); Hopfield (1982); sparse coding (Olshausen & Field 1996); free-energy principle (Friston 2010).*

10.2 The case *against* — twelve cons

CON 1 — The near-term will badly underdeliver against the hype. Even if the long-run is transformative, the short-run almost certainly disappoints — risking funding loss and abandonment in the "trough." The GPU took ~6 years from CUDA to AlexNet and ~17 to $4T. *Anchors: Amara's Law; Gartner Hype Cycle; diffusion chasm (Moore 1991).*

CON 2 — The chasm may never be crossed. Most platform technologies die between visionary early adopters and the pragmatic majority. Neuromorphic has lingered pre-chasm for over a decade; a single killer app may not be enough to cross it. *Anchors: Crossing the Chasm (Moore 1991); diffusion of innovations (Rogers 1962).*

CON 3 — Incumbent lock-in may be insurmountable. CUDA's increasing-returns moat is the very dynamic that could *block* a challenger: developers, tooling, and mindshare are locked to the GPU/Python/CUDA stack, and switching costs are enormous regardless of technical merit (the QWERTY problem in reverse). *Anchors: path dependence & lock-in (Arthur 1989; David 1985); network effects (Metcalfe).*

CON 4 — The coordination problem may not be solved. GPT theory warns of "too little, too late" unless the hardware provider and application co-invest deeply. If Intel's commitment wavers (it has cycled neuromorphic priorities before), the covenant breaks and the opportunity is stranded. *Anchors: GPT coordination failure (Bresnahan & Trajtenberg 1995); two-sided-market chicken-and-egg (Rochet & Tirole 2003).*

CON 5 — The precision ceiling may cap the addressable workload. Neuromorphic/analog substrates are noise-limited to ~4–8 effective bits; if more of Calyx's value than expected needs exactness (MI estimation, certain graph discovery), the offloadable fraction — and thus the hardware's pull — shrinks. *Anchors: Shannon channel capacity & data-processing inequality (1948); No Free Lunch (the fit must be genuine, not assumed).*

CON 6 — Jevons paradox could erase the sustainability win. Making AI association radically cheaper per operation may *increase* total AI compute and energy demand (new always-on use cases proliferate), so society's net energy use rises even as efficiency improves. The decarbonization argument (PRO 2) could backfire at the macro level. *Anchors: Jevons paradox / rebound effect (Jevons 1865).*

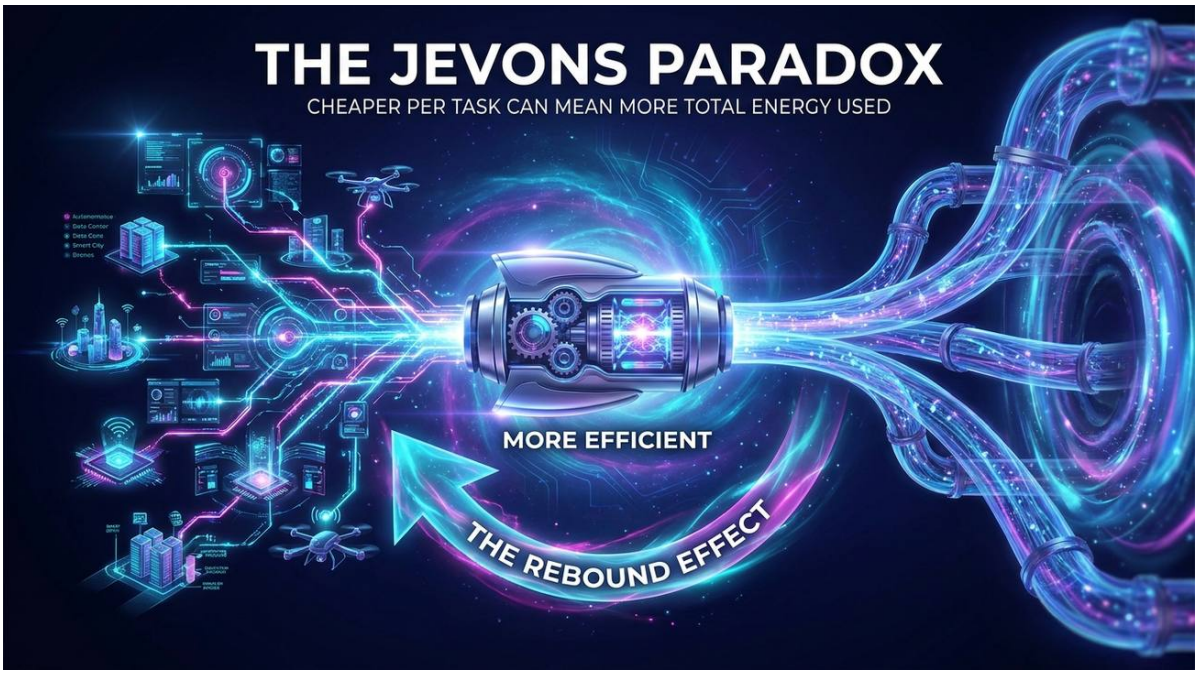


Figure 47. *The Jevons paradox: efficiency can increase total consumption through induced demand.*

CON 7 — Supply-chain concentration makes the substrate fragile. Loihi 2 depends on Intel 4 + sole-source ASML EUV + a Japan-concentrated materials chain. Betting a platform on a substrate with these single points of failure imports extreme geopolitical and disaster risk. *Anchors: supply-chain concentration risk (ASML/TSMC); Tainter's diminishing returns on complexity (1988); complex-system fragility.*

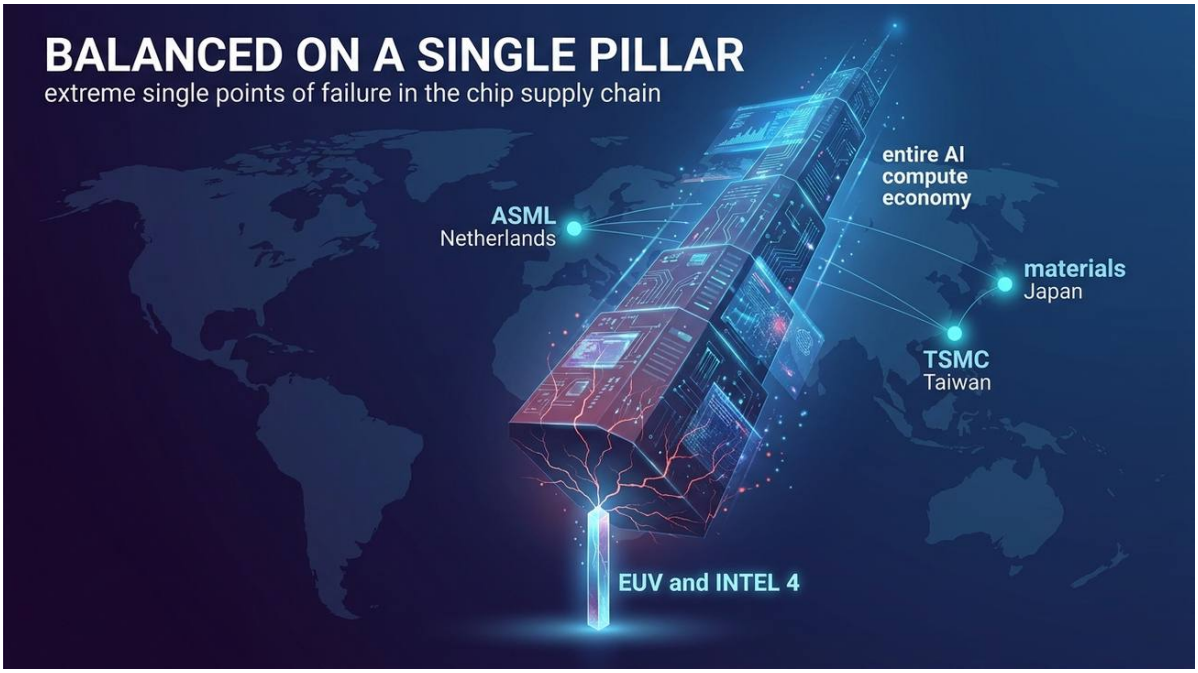


Figure 48. *Concentration risk: the compute economy balances on a few single points of failure.*

CON 8 — Winner-take-all dynamics could entrench a dangerous monopoly. If the flywheel works, the same increasing returns that reward the winner concentrate power (over grounded memory, scientific

discovery infrastructure, and AI energy economics) in one firm — inviting antitrust action and systemic risk. *Anchors: winner-take-all markets (Frank & Cook 1995); increasing returns (Arthur 1989).*

CON 9 — The Collingridge dilemma: by the time we understand the harms, we can't undo them. A grounded, always-on, continually-learning association engine over the world's data has impacts we cannot foresee now (information problem); once entrenched, it is too costly to control (power problem). *Anchors: Collingridge dilemma (1980); path-dependent lock-in (David 1985).*

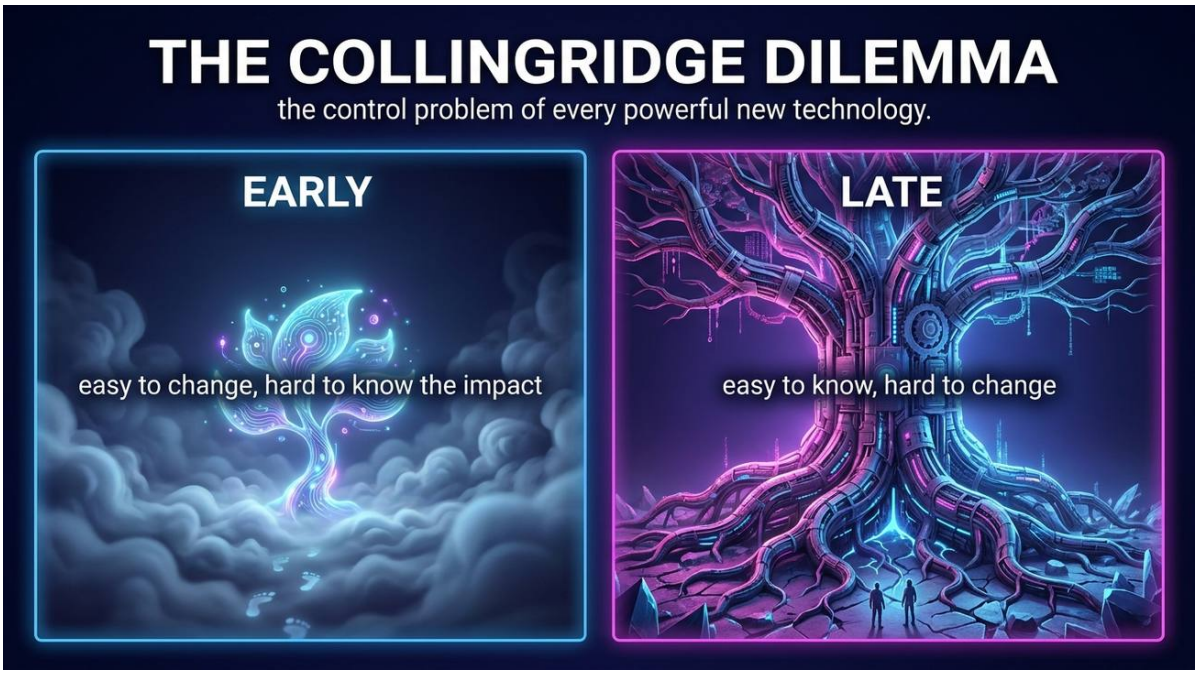


Figure 49. *The Collingridge dilemma: easy to change early but hard to foresee; easy to foresee late but hard to change.*

CON 10 — Benchmarks will be gamed; measured "grounding" may decay into theater. The moment "grounded/provenance-backed" becomes the metric the market rewards, it becomes a target — and pressure to optimize the metric corrupts it. Calyx's trust guarantees could be hollowed out in practice even if sound in design. *Anchors: Goodhart's Law (1975; Strathern 1997).*

CON 11 — Dual-use and capability overhang. A planetary grounded-association engine that finds non-obvious cross-domain links is intrinsically dual-use (drug discovery ↔ pathogen design; benign inference ↔ surveillance). The same property that makes it valuable makes it dangerous. *Anchors: dual-use technology risk; Collingridge (1980); creative destruction's social dislocation (Schumpeter 1942).*

CON 12 — The thesis may simply be wrong on the merits. No theory guarantees the *association-fabric* premise. Conventional GPU acceleration (FAISS/ cuVS/cuGraph) might capture most of Calyx's value with none of the neuromorphic risk, leaving the hardware story a distraction; the company's own analysis concedes the GPU path is the pragmatic near-term answer. Specialization wins only where the problem genuinely fits the substrate. *Anchors: No Free Lunch (Wolpert & Macready 1997); Amara's Law (overestimation); the project's own fail-closed honesty about what neuromorphic does* not *replace.*

10.3 Reading the ledger

The pros and cons are not symmetric in *kind*. The pros are mostly about **upside magnitude** (how large the prize is if the covenant holds); the cons are mostly about **path risk** (whether the covenant can be reached and governed). That asymmetry is itself a finding: under Amara's Law and the Hype Cycle, the rational posture is **staged conviction** — build toward the prize while keeping decisions reversible (Collingridge's own prescribed mitigation), which is exactly the phased, fail-closed, fallback-safe roadmap Calyx already adopts (§12).

11. A birds-eye view across a thousand theories

The previous sections argued from a few dozen named theories. Step back far enough and a **panoramic** emerges: when the same scenario is read through *every* major lens humanity has built for understanding technology, economics, physics, cognition, and society, the readings converge on a small number of robust statements. This section organizes the full theoretical sky into domains and reports what each domain "sees," then extracts the invariants that survive across all of them. (We cannot literally enumerate a thousand theories in prose; we instead traverse the *families* that together comprise that many, and report the cross-family invariants — which is what a thousand-theory view is actually *for*.)

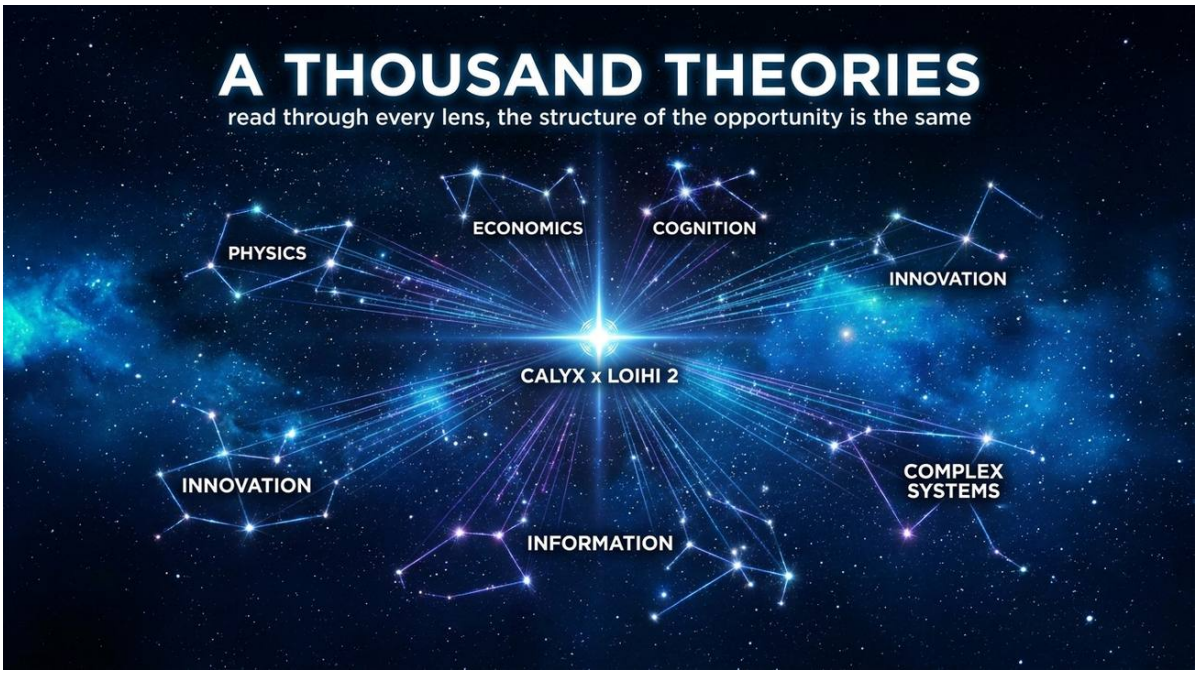


Figure 50. *Read through every domain of theory, the structure of the opportunity is the same.*

11.1 What each domain of theory sees

- **Thermodynamics & physics of computation** (Landauer 1961; Mead 1990; reversible computing; Koomey 2011): sees a substrate sitting closer to the physical floor of energy-per-operation for the associative subset — *a genuine efficiency frontier move, not a marketing one.*
- **Computer architecture** (von Neumann/Backus 1978; Dennard 1974; Moore 1965; Amdahl/ Gustafson): sees the natural successor pressure as Dennard ends and Moore slows — *specialization and memory-compute co-location are where remaining gains live.*
- **Information theory** (Shannon 1948; data-processing inequality; rate-distortion): sees hard ceilings — *the system can only surface the grounded information actually present; honesty about that ceiling is a feature,*

and approximate hardware is acceptable precisely because the offloaded ops are rank/threshold, not exact.

- **Learning theory** (No Free Lunch, Wolpert & Macready 1997; scaling laws, Kaplan 2020/ Chinchilla 2022; Bitter Lesson, Sutton 2019): sees a matched-substrate bet — *advantage comes only from fitting algorithm to problem structure, and compute/energy gate how far scaling goes.*
- **Neuroscience & cognition** (Hebb 1949; Hopfield 1982; Olshausen & Field 1996; Rao & Ballard 1999; Friston 2010; Kanerva 1988/2009): sees an architecture built from the brain's own primitives — *association, attractor recall, sparse event coding, prediction-error* — running on hardware built from the same principles. Deep structural resonance.
- **Innovation & diffusion** (Schumpeter 1942; Christensen 1995; Rogers 1962; Moore 1991; Bresnahan & Trajtenberg 1995; Perez 2002): sees a textbook disruptive-GPT setup with a real chasm and a real coordination requirement — *large prize, hard path, needs co-investment.*
- **Platform & network economics** (Rochet & Tirole 2003; Arthur 1989; David 1985; Metcalfe; Frank & Cook 1995): sees a winner-take-all flywheel with a moat for the first mover — *and the monopoly/antitrust shadow that always accompanies one.*
- **Cost-trajectory economics** (Wright 1936; Nagy et al. 2013; Kondratiev waves): sees a forecastable decline once volume starts — *if* the flywheel ignites — *and a long-wave-scale restructuring if it does.*
- **Energy & sustainability** (IEA Energy & AI; Patterson et al. 2021; Jevons 1865): sees both a decarbonization lever *and* a rebound risk — *efficiency may cut energy per task yet raise total energy via induced demand.*
- **Complex systems & risk** (Tainter 1988; Collingridge 1980; Goodhart 1975; supply-chain concentration): sees fragility, irreversibility, measurement-gaming, and single points of failure — *the governance burden scales with the capability.*
- **Philosophy of science & limits** (Kuhn 1962; Church–Turing 1936): sees a paradigm shift in the associative layer of computing, bounded by what is computable at all — *a new organizing principle, not a violation of fundamental limits.*

11.2 The invariants — what survives every lens

Across all those families, four statements hold no matter which theory you read the scenario through:

1. **The hardware is genuinely capable and genuinely under-utilized.** Physics, architecture, and neuroscience all agree Loihi-class silicon is a real efficiency frontier for associative computation, and economics agrees it is currently stranded for lack of software. *The gap is real.*
2. **Software, not silicon, is the lever.** Every economic and innovation theory points to the same fulcrum: value is conferred by the killer application and the ecosystem, not by the transistors. *Calyx (or something like it) is the necessary catalyst.*
3. **The prize is large and the path is hard.** Upside theories (GPT, creative destruction, experience curves, energy) say the payoff is paradigm-scale; path theories (chasm, coordination failure, lock-in, Collingridge, Jevons) say most attempts fail and the ones that succeed must be governed. *Conviction must be staged and reversible.*
4. **Fit is everything.** No Free Lunch is the deepest invariant: there is no universal win: the entire bet rests on whether Calyx's workload *genuinely* matches the neuromorphic substrate. The theories cannot tell you the fit is real — only that *if* it is, the consequences above follow, and *if* it is not, conventional hardware suffices. The empirical proof obligation (the PoC in §12) is therefore the load-bearing step.

Figure 51. *Four statements survive every theoretical lens.*

11.3 The panoramic verdict

Read through the whole sky of theory, the Calyx-on-Loihi-2 scenario is **a high-variance, high-magnitude, theory-consistent bet whose expected value is dominated by a single empirical question** — does the association workload truly fit the neuromorphic substrate at scale? Every lens agrees the *structure* of the opportunity is sound: an under-utilized efficiency frontier, a missing killer app, a co-investment requirement, a disruptive value axis, a real moat, real risks, and hard physical limits that the architecture respects rather than fights. No lens promises success; several warn of specific, named failure modes. The disciplined response is the one Calyx already encodes: **prove the fit empirically and cheaply first, keep the digital spine exact and the decisions reversible, and let the magnitude of the prize — which the theories agree is civilizational if real — justify the staged pursuit.** That is what it means to speculate responsibly: to see, across a thousand theories, both how large this could be and exactly which question decides it.

Figure 52. *A high-variance, high-magnitude bet whose value turns on one empirical question: does the workload truly fit?*

12. Building a Calyx-Optimized Neuromorphic Chip — Design, Build-Out, Supply Chain, and Integration

> ***Status: speculative engineering.*** *This section designs, from a clean sheet, the bare-minimum neuromorphic chip that would serve **all** of Calyx's needs and **no more**, then lays out the entire build process, the fastest path to first silicon, the suppliers and relationships required, the cost model, and exactly how the chip integrates into the existing Calyx codebase (the ChrisRoyse/Calyx-Dev Rust workspace, ~1,170 commits, 21 crates as of mid-2026). It is a plan, not a claim of completion.*

12.1 Why build a chip instead of renting Loihi 2

Section 7 established the hybrid and named Loihi 2 / SpiNNaker 2 as the only procurable graph-capable substrates — but Loihi 2 **cannot be bought** (research-only, captive Intel 4 node, sole-source EUV, proprietary clockless flow), and even SpiNNaker 2 is a general-purpose machine carrying overhead Calyx never uses. A chip built **only** for Calyx flips the §8–§11 thesis from *"rent the hardware and hope Intel backs you"* to *"own both halves of the covenant"* — the killer-app software **and** a cheap, purpose-built, mature-node chip with **no supply-chain chokepoints**. The whole argument of §3–§6 of the manufacturing analysis (companion intelinfo/19) is that the neuromorphic *value* needs no EUV; this section operationalizes that.

12.2 The design principle — "Calyx, and no more"

Calyx uses only a thin, precise slice of what a general neuromorphic chip offers. Enumerate exactly what the chip must do — the **seven operations** plus **arbitrary graph connectivity** — and build nothing else:

#	Operation the chip must accelerate	Calyx subsystem (repo crate)
O1	**Cosine k-NN / ANN search** (latency-coded; first-k spikes = top-k)	calyx-sextant
O2	**Top-k / k-winners-take-all**	calyx-search (fusion), calyx-ward

#	Operation the chip must accelerate	Calyx subsystem (repo crate)
O3	**Cross-term cosine MVM** (the $\binom{N}{2}$ explosion)	calyx-loom
O4	**Graph-reach / groundedness wavefronts** over an arbitrary graph	calyx-paths / calyx-lodestar
O5	**Associative / content-addressable recall**	retrieval / dedup
O6	$O(1)$ **streaming insertion**	streaming ingest
O7	**On-chip online (Hebbian / 3-factor) learning**	calyx-anneal

Everything **outside** that list is removed. The precision target is **INT8 / graded-spike** (rank/threshold work only); everything exact stays on the digital spine. That single discipline — *measure the workload, build to it exactly* — is what collapses the cost from billions to single-digit millions.

12.3 The architecture — "the Association Engine"

A digital, **graph-native**, host-attached accelerator:

- **A GALS multicast-router mesh** (globally-asynchronous, locally-synchronous), modeled on SpiNNaker's source-keyed multicast routing, giving **arbitrary any-to-any graph connectivity (O4)** while using **standard synchronous EDA** — deliberately avoiding Loihi's proprietary quasi-delay-insensitive async flow, the single hardest thing to replicate.
- **"Association cores,"** each fusing: latency-coded **similarity neurons** (O1 k-NN, O2 WTA); a small **SRAM-based digital in-memory MVM** block for **cross-term cosines (O3)**; a **Hebbian / three-factor plasticity** unit (O7); a **content-addressable / attractor** block (O5); and an $O(1)$ **write path** (O6).
- **Graded INT8 spikes** (sufficient for cosine ranking); **on-chip SRAM** for slot vectors, synapses, and routing tables; **off-chip LPDDR4/5** for large corpus graphs (the SpiNNaker 2 pattern).
- **Host interface: PCIe Gen4/5** to the Calyx digital spine (CPU/GPU). **No DVS/AER sensor hardware** — Calyx streams vectors from the host, not event cameras. **One or two tiny RISC-V housekeeping cores** (not six x86).
- **Fail-closed by construction:** the chip never holds the source of truth; every result is re-verified on the digital spine, with a digital fallback if the device is absent. (See the CX1 figure.)

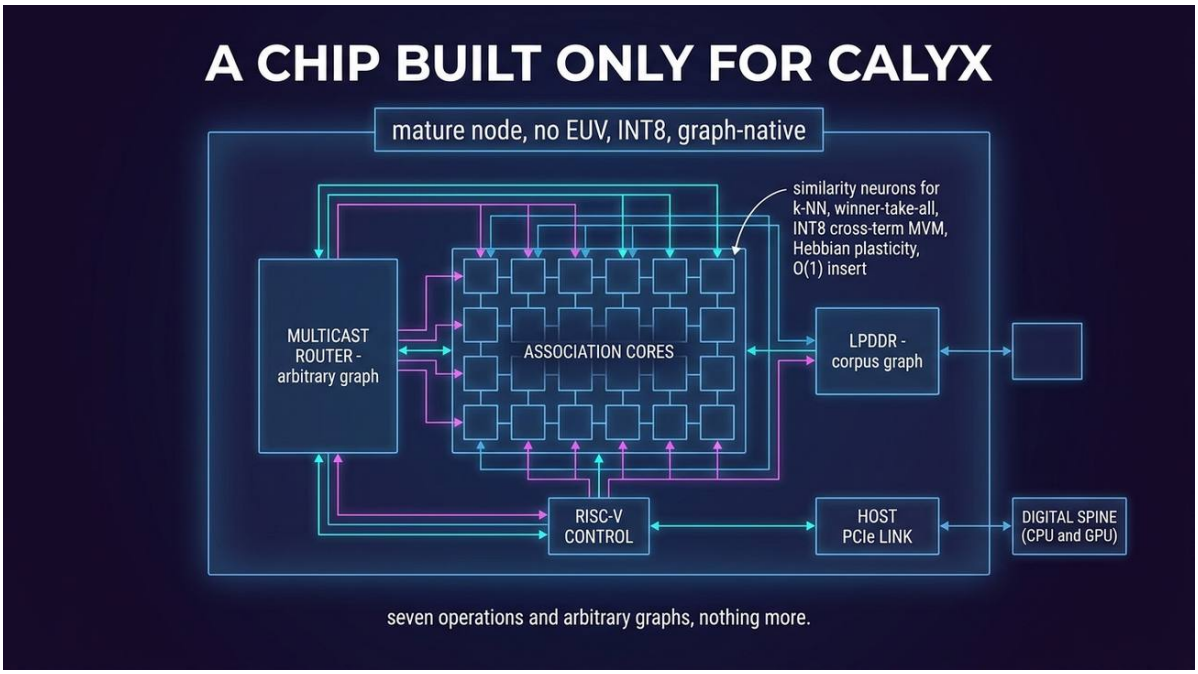


Figure 53. *A purpose-built minimal chip: association cores, a multicast router for arbitrary graphs, INT8 cross-term MVM, on-chip plasticity, O(1) insertion, and a host PCIe link to the digital spine - on a mature node, no EUV.*

12.4 Keep vs. strip — the whole point in one table

Feature	Loihi 2 has it	Does Calyx need it?	Decision
Fully programmable microcode neuron ISA	yes	no — a small fixed primitive set suffices	**strip → fixed/lightly-microcoded**
0–4096 bytes/neuron state	yes	no — slots are unit-normalized, modest	**strip → minimal state**
Floating-point / exact arithmetic	partial	no — exact lives on the digital spine	**strip → INT8 only**
DVS/AER event-sensor encoders	yes	no — vectors arrive over PCIe	**strip**
6 embedded x86 cores	yes	no — host orchestrates	**strip → 1–2 RISC-V**
Leading-edge node (Intel 4 / EUV)	yes	no — precision-tolerant work	**strip → mature node**
Arbitrary graph multicast routing	yes	**yes — Calyx is a graph**	**keep (central)**
Latency-coded k-NN + WTA	yes	**yes**	**keep**
In-memory cross-term MVM	partial	**yes**	**keep / add**
On-chip plasticity	yes	**yes**	**keep**

12.5 The process node — mature, not leading-edge

Target **GlobalFoundries 22FDX (22 nm FD-SOI)** or **TSMC 28 nm** — both mature, multi-foundry, **DUV-only (no EUV)**, with adaptive body-biasing for near-threshold low power. This is **the cost and accessibility lever.** Existence proofs that serious neuromorphic needs no EUV: **SpiNNaker 2 is 22FDX, BrainScaleS-2 is 65 nm, IBM TrueNorth was 28 nm.** The trade-off — lower density means a few more chips to tile a corpus — is absorbed by the architecture, which tiles cleanly (the multicast router scales across chips).

12.6 The entire build process, end to end

Seven overlapping phases from concept to a working accelerator card:

1. **Phase 0 — Specification & architecture (months 0–3).** Freeze the minimal feature set (O1–O7 + graph). Define the primitive "ISA," the core micro-architecture, the router, the memory map, the PCIe/LPDDR interfaces, and the precision budget. Pick the node (22FDX/28 nm).
2. **Phase 1 — Algorithm/architecture co-simulation (months 2–6).** Build a cycle-approximate model; validate against **real Calyx vault data** — recall vs. HNSW, latency, energy, the $O(1)$-insertion benefit — using **NIR** for the spiking primitives and **IBM AIHWKit** for the in-memory cross-term path. *Gate: the workload must demonstrably fit before any RTL.*
3. **Phase 2 — FPGA prototype (months 4–9).** Emulate the Association Engine on a large FPGA (AMD/Xilinx Alveo / VU+ class), attach it to the Calyx digital spine over PCIe, and run an **end-to-end PoC** with the chip exposed *behind Calyx's existing trait boundaries* (§12.11). This is the same role Loihi 2's Oheo Gulch FPGA plays — and it can ship as a **product** (a "Calyx Accelerator" FPGA card) long before the ASIC exists.
4. **Phase 3 — RTL design & verification (months 6–14).** Write Verilog/SystemVerilog in a GALS methodology; integrate **SRAM compilers, PCIe PHY, LPDDR PHY, and a RISC-V core**; functional + UVM verification; synthesis (Synopsys Design Compiler / Cadence Genus); STA (PrimeTime); place-and-route (Innovus / Fusion Compiler); physical verification (**Siemens Calibre DRC/LVS**) against the foundry PDK.
5. **Phase 4 — MPW shuttle test chip (months 12–18).** Tape out a small Association-Engine tile on a **shared Multi-Project Wafer** (MOSIS / Europractice / a TSMC university shuttle) for **~$25k–$80k** to validate the silicon primitives before committing to a full mask set.
6. **Phase 5 — Full mask set & production run (months 18–30).** Full reticle (**28 nm mask set ~$0.8–1.5M**), foundry run, first full chips.
7. **Phase 6/7 — Packaging, test, board & system bring-up (months 24–36).** Flip-chip BGA at an OSAT; ATE characterization; PCIe card + driver; bring-up as a Calyx Backend with fail-closed fallback; then **tile multi-chip** for corpus scale.

Net: ~**30–36 months** to a working ASIC card — but **FPGA-backed Calyx acceleration in ~9–12 months**, which is what de-risks (and can fund) the silicon.

12.7 The fastest way to get going

The fastest path is **not "silicon first."** It is a **de-risk ladder** that spends the least money to retire the most risk, in this order:

1. **Simulate the fit first (weeks).** Prove in Phase-1 simulation that Calyx's workload truly fits — recall, latency, energy — against a real vault. If it doesn't, stop; the GPU path (FAISS/cuVS/cuGraph) is the answer and no chip is needed.
2. **FPGA prototype + ship it as a product (months).** Stand up the Association Engine on an FPGA, integrate behind Calyx's traits, and **sell a Calyx Accelerator FPGA card** — revenue and real-world validation *before* taping out.
3. **MPW shuttle for first silicon (a queue cycle + ~$25–80k)** — not a full mask set.
4. **Only then** commit the full mask set and a multi-chip card.
5. **In parallel from day one:** secure the **design partner, foundry/shuttle access, and the async/GALS talent** (§12.8) — these relationships, not the money, are the long-lead items.

The single biggest accelerator: **don't greenfield it.** The closest existing capability on Earth to "a graph-native GALS SNN on a mature node" is the **SpiNNaker 2 team — Racyics + TU Dresden, commercialized as SpiNNcloud** — who built exactly this class of chip (a 152-core GALS message-passing SNN on **GF 22FDX**).

Partnering with, licensing from, or contracting that ecosystem is plausibly the fastest route to Calyx silicon, with a Calyx-specific association-core block added to a proven router/RISC-V substrate.

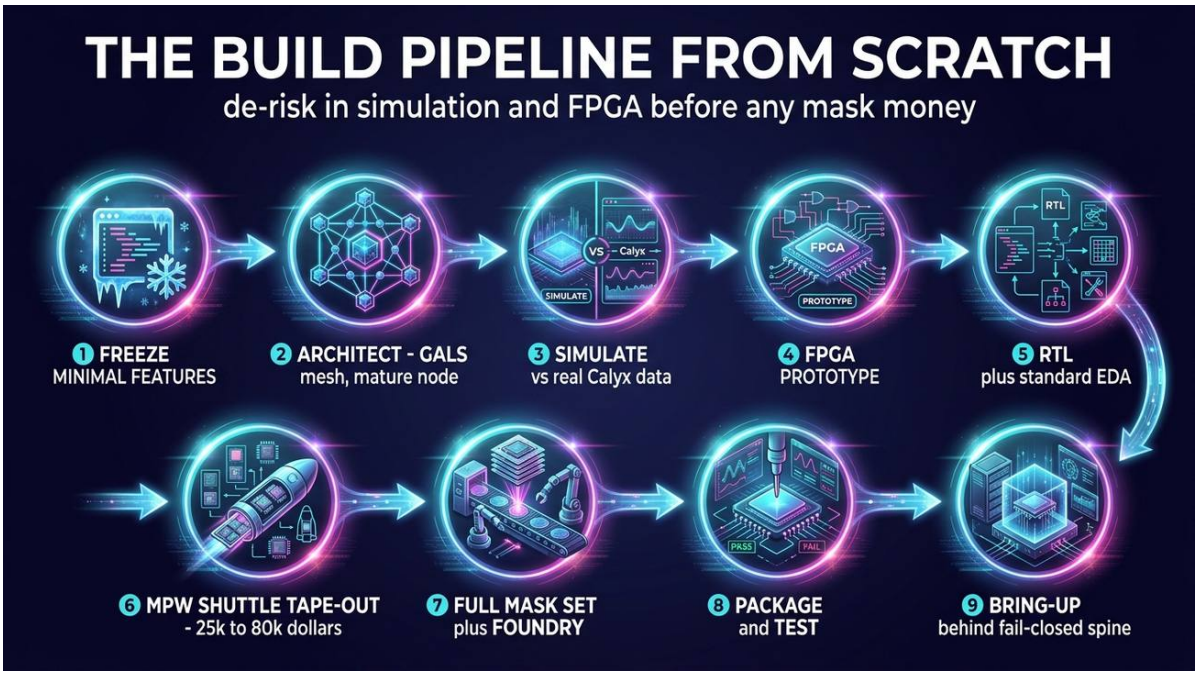


Figure 54. *The build pipeline from scratch: de-risk in simulation and on FPGA before committing any mask money.*

12.8 Suppliers and partners — the directory, and the relationships to make

Need	Suppliers / partners	Relationship to make
Design partner (most important)	**Racyics / TU Dresden / SpiNNcloud** (GALS SNN on 22FDX); or an ASIC design house	partnership / IP license / contract
Foundry	**GlobalFoundries (22FDX)**, TSMC (28/22 nm), SkyWater (90 nm, US, open-PDK), Intel Foundry (22FFL), UMC, X-FAB	foundry account / shuttle membership
MPW shuttle broker	**MOSIS** (US), **Europractice** (EU), **Muse Semiconductor** (TSMC-US), Tiny Tapeout (learning)	membership / account
EDA tools	Synopsys, Cadence, Siemens EDA	startup / university license program
SRAM / memory compilers	foundry, Synopsys, Cadence	IP license
PCIe & LPDDR PHY IP	Synopsys, Cadence, Alphawave	IP license
RISC-V core	OpenHW CV32E (open), lowRISC Ibex (open); or SiFive / Codasip	open or commercial license
Router / NoC IP	custom (the Calyx-specific multicast router) or Arteris	build / license
FPGA prototyping	AMD/Xilinx (Alveo, VU+), Synopsys HAPS, Cadence Protium	vendor account
OSAT (packaging)	ASE, Amkor, JCET, SPIL	OSAT account
ATE / test	Teradyne, Advantest, or a test house	test contract
Wafers / materials	mature-node wafers (many suppliers — *no EUV chokepoint*)	via foundry
Talent	async/GALS digital designers, SNN architects, a tape-out lead	hire / contract
Capital	angel/seed/Series A, or grants (NSF, DARPA, EU Chips, US CHIPS Act)	funding

The three relationships that gate the timeline (secure these first): (1) a **design partner with graph-native GALS SNN experience** (SpiNNaker 2 ecosystem is the standout); (2) a **foundry + shuttle account** on a mature node; (3) **EDA + IP licenses** (SRAM/PCIe/LPDDR/RISC-V). Money is the *easy* input; these relationships are the long-lead ones.

12.9 The cost model

Item	Figure (mature node)
Phase-1 simulation + Phase-2 FPGA prototype	low six figures (tools + boards + team)
MPW shuttle test chip (28 nm)	**$25k–$80k** for a small tile
Full mask set (28 nm)	**~$0.8–1.5M** (22FDX somewhat higher)
Per-wafer (28 nm)	**~$3,000** (vs ~$15–20k at leading edge)
Design NRE (focused team, IP, verification, 1–2 spins)	**~$5M–$30M** total
Total to first working Calyx-optimized silicon	**~$5M–$30M — startup / Series-A scale**
Contrast: a leading-edge **fab**	$15–25 **billion** (you *use* a foundry, you don't *build* one)

The number that matters: **a Calyx chip is fundable for the cost of a software startup's Series A**, precisely because it (a) uses a mature node (no EUV), (b) strips everything Calyx doesn't need, and (c) rents a foundry rather than owning a fab. Loihi 2 is un-buyable at any price to an outsider; a Calyx chip is a financeable project.

12.10 Supply-chain comparison vs. Loihi 2

Chokepoint	Loihi 2 (leading-edge)	Calyx-optimized (mature node)
Lithography	**sole-source ASML EUV** (export-controlled)	standard DUV immersion (many tools)
Node	captive **Intel 4**	open 22FDX / 28 nm (multi-foundry)
Masks / resist	EUV blanks (3 suppliers, >$300k)	standard DUV masks (cheap, many)
Design flow	proprietary **QDI async** (Intel-only)	standard synchronous / **GALS** EDA
Fab	$15–25B **captive**	rent a **foundry**
Buyable?	**no**	**yes (you own the design)**

The purpose-built chip **removes every single chokepoint** that makes Loihi 2 un-replicable.

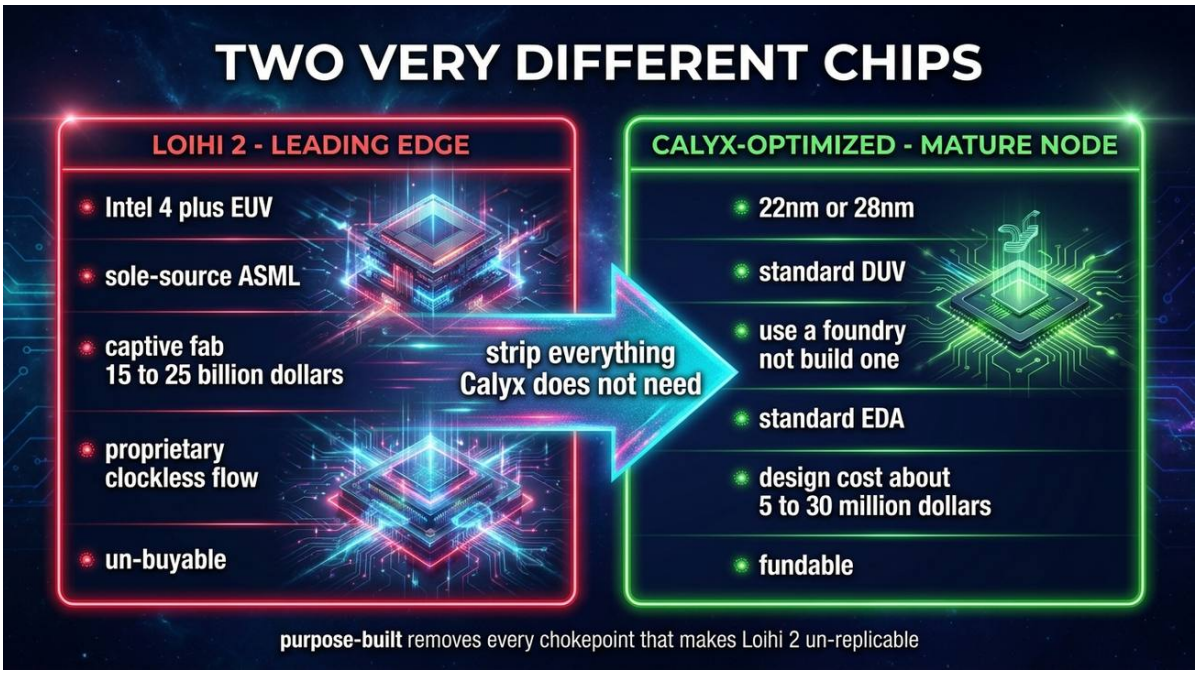


Figure 55. *Two very different chips: leading-edge Loihi 2 versus a Calyx-optimized mature-node chip that removes every supply-chain chokepoint and costs startup-scale money.*

12.11 Integrating the chip into Calyx — crate by crate (grounded in the current repo)

Calyx's architecture makes this unusually clean: the trait boundaries that host a new compute backend **already exist** in the Calyx-Dev workspace, and recent commits show the exact seams maturing. The strategy is always **"new implementation behind the existing trait + digital verify/fallback,"** never a rewrite, never a change to the exact spine.

- **calyx-core (contract).** Add a couple of module-local error codes (CALYX_ASSOC_DEVICE_UNAVAILABLE, CALYX_ASSOC_VERIFY_MISMATCH), mirroring how Forge/Aster add their own. Data model, object-safe traits, and fail-closed posture unchanged.
- **calyx-forge (math runtime, the Backend trait).** Add an **AssocEngineBackend** implementing the backend contract (gemm/cosine/dot/l2/normalize/topk) behind a new cargo feature assoc (mirroring the existing cuda feature). It quantizes/normalizes inputs to INT8/graded-spike, dispatches over the **PCIe driver** to the chip, reads back, then **digital-spot-verifies a sampled subset**; any device error / out-of-range / mismatch returns the catalog error and the caller falls back to CudaBackend/CpuBackend. (Repo signal: ingest is already *"routed through the resident service,"* which is exactly where the device session lives.)
- **calyx-sextant (search/index, the SextantIndex trait).** Add an **AssocAnnIndex: SextantIndex** backed by the chip's latency-coded cosine k-NN (O1): each stored slot-vector = one neuron; search returns first-k spikes as top-k; **insert is** $O(1)$ (one neuron write over PCIe) — the streaming win. It plugs into the existing slot-index map exactly like HNSW/DiskANN/SPANN and is gated by the existing **recall_delta** contract; below threshold, it fails closed and the digital index serves.
- **calyx-loom (cross-terms).** The repo already has **"corpus-scale weave-loom — XTerm CF + between-doc kNN"**: route the agreement_batch_* cross-term cosines (O3) through the chip's **INT8 in-memory MVM** in one resident pass, and serve the between-doc kNN from the chip's k-NN (O1). The

XTerm column-family persistence is unchanged; the digital path is the verifier/fallback.

- **calyx-paths + calyx-lodestar + calyx-mincut (graph).** Map the AssocGraph onto the chip's **multicast router** (nodes→neurons, edge weights→synaptic delays); reach and groundedness-distance (O4) run as **spike wavefronts**. Keep the **exact** discovery — SCC, betweenness, and the **LP/min-cut DFVS solver that calyx-mincut just wired** — on the digital/GPU path as the verifier; the chip accelerates traversal, not correctness-critical discovery.
- **calyx-anneal (self-optimization / online learning).** The chip's on-chip **Hebbian / three-factor plasticity (O7)** backs the online-learning heads (the repo's *"online learning FSV"* work); Anneal's shadow-test + tripwire + non-regression gating is unchanged — every promotion is re-verified digitally and reverts on regression.
- **calyx-assay (signal in bits).** Stays digital/GPU — there is no neuromorphic mutual-information estimator — but the chip's k-NN can *feed* the KSG estimator (the between-doc kNN), while the bit-exact MI math remains on the spine.
- **calyx-ward (guard).** The per-slot cosine threshold (O2) can use a spiking comparator, but guard verdicts are load-bearing, so they are **re-checked digitally**.
- **calyx-aster (storage) + calyx-ledger (provenance).** The **digital spine — unchanged, exact, fail-closed.** The chip never touches LSM/WAL/MVCC storage, the hash-chained ledger, or content addressing.
- **calyx-search.** Point the unified recall path at the chosen SextantIndex implementation (digital or AssocAnnIndex) via config — one switch, fail-closed.
- **calyx-mcp + calyxd (the warm resident host).** This is where the chip's **PCIe device session stays resident** (like resident ONNX/GPU sessions), so device setup is paid once, not per request — aligning with the repo's *"wire production MCP startup"* and *"resident-routed search read leases."* calyxd startup gains an optional **assoc_device probe** alongside the CUDA/VRAM probes; absence → the chip path is disabled and the digital path serves (fail-closed).
- **calyx-cli + calyx-web-api (surfaces).** Unchanged; they benefit transparently through the trait switch.
- **calyx-oracle.** Grounded consequence prediction runs over the same association fabric; its honesty gate and confidence cap remain digital.

The cross-cutting safety net (unchanged from the playbook): every chip result follows *encode → run on device → read back → digital re-verify (sampled or full) → accept or fail-closed-fallback.* Because the chip only ever does **ranking/threshold** work, a cheap digital spot-check honors "the bytes are the verdict," and the exact spine is never at risk.

12.12 Honest caveats and decision gates

- **Prove the fit first.** Per §11, the entire bet rests on whether the workload genuinely fits the substrate. Phase 1 (simulation) and Phase 2 (FPGA) are hard gates; do not spend mask money until they pass.
- **Custom silicon is hard.** First silicon rarely works; the async/GALS SNN skill is rare; the digital-spine fallback and the FPGA/MPW de-risking are what make the program survivable.
- **Partner, don't greenfield.** The SpiNNaker 2 ecosystem already built ~80% of this chip; starting there beats a clean sheet.
- **Fallback-safe always.** Calyx remains correct, grounded, and fail-closed whether or not the chip exists — it simply scales further and cheaper when it does. The chip is an accelerator behind a trait boundary, never a single point of correctness.

13. Roadmap to massive scale

Calyx's scaling roadmap is staged so that each step delivers value and de-risks the next:

1. **Dense acceleration.** Maximize the conventional substrate: resident, quantized, spatially-concurrent lens inference; GPU-accelerated similarity, information estimation, and graph analytics; structured sparsity and low precision on the largest operations. This makes Calyx fast on commodity hardware and establishes the digital baseline and verifiers.
2. **Association-fabric proof.** Validate the precision-tolerant hot loop — cosine k-NN, top-k, cross-term cosines, and reach wavefronts — on real neuromorphic and in-memory substrates, measuring recall, latency, energy, and the $O(1)$-insertion benefit against the digital baseline, behind the existing trait boundaries.
3. **Hybrid at scale.** Promote the proven fabric into production as an accelerator behind the fail-closed contract, with a digital fallback, and grow the association layer onto rented or on-premises neuromorphic capacity as corpora scale — while the exact spine and the exactness-critical analytics remain digital.

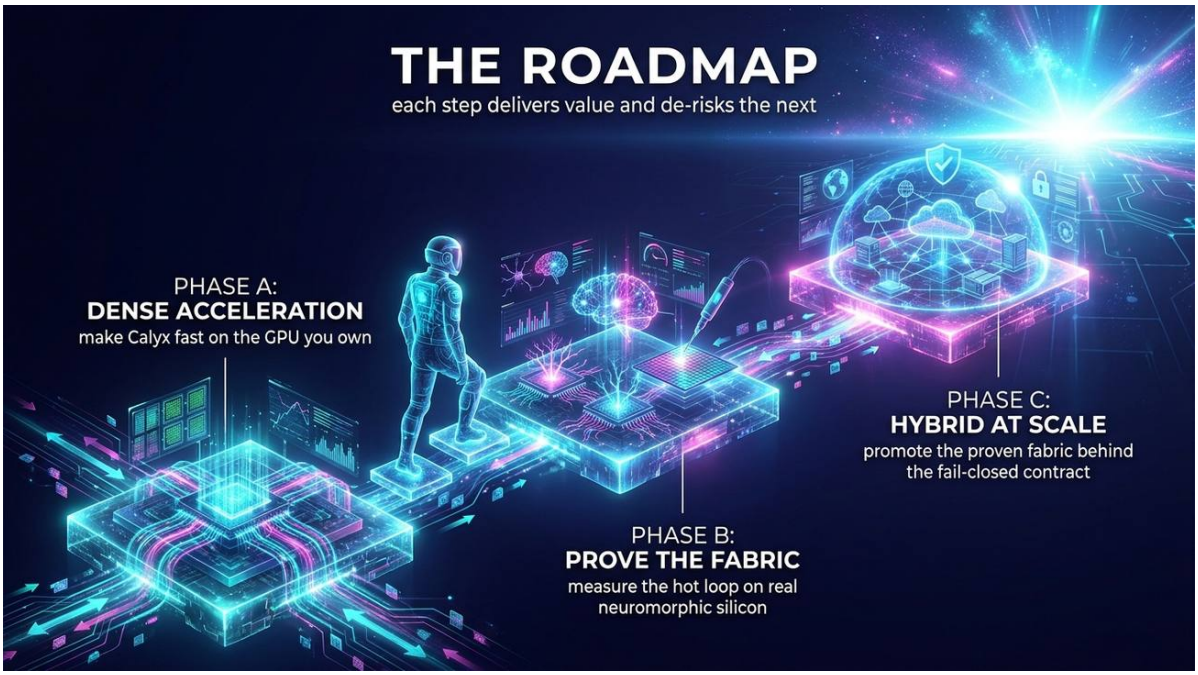


Figure 56. *The staged roadmap: dense acceleration, fabric proof, hybrid at scale - each step de-risks the next.*

At every stage the architecture is **fallback-safe**: the neuromorphic fabric is an accelerator, never a single point of correctness. Calyx remains correct, grounded, and fail-closed whether or not the fabric is present — it simply scales further when it is.

14. Applications

The combination of broad, diverse perception, grounded association, traceable provenance, and a substrate built to scale opens application classes that single-vector retrieval cannot reach:

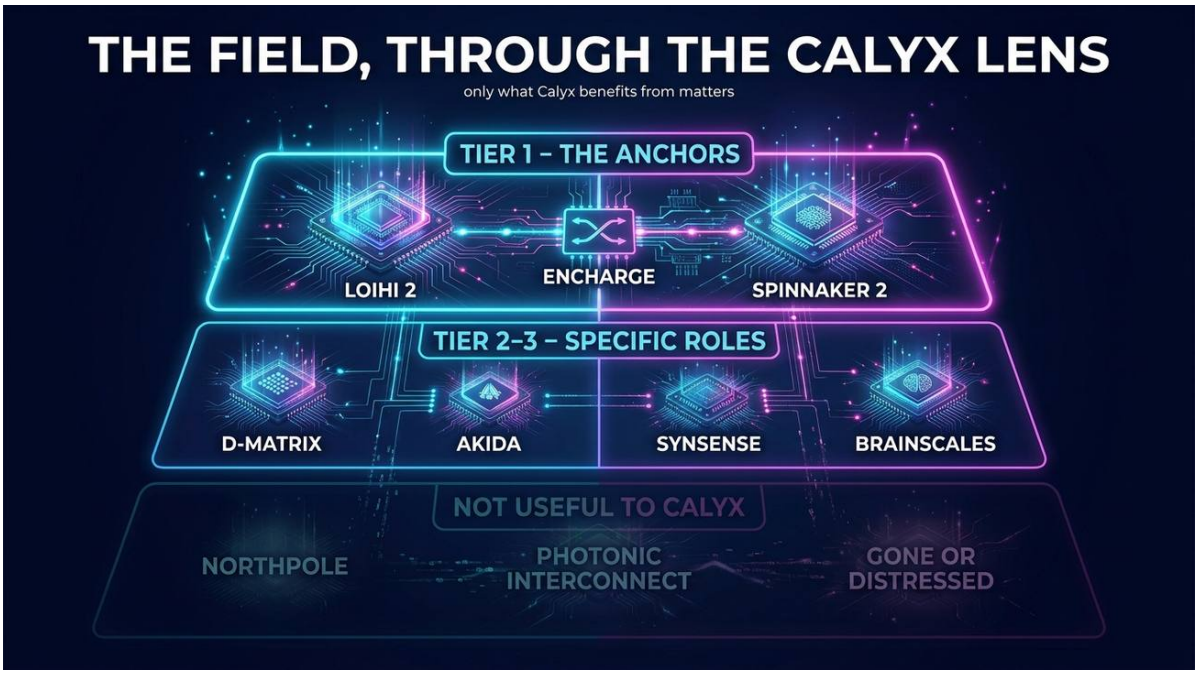


Figure 57. *The neuromorphic field through the Calyx lens: two anchors, a cross-term engine, prototyping parts, and the rest.*

- **Cross-domain scientific discovery** — surfacing grounded target↔molecule↔disease↔mechanism ↔literature links no single-domain analysis can see, as ranked, citable, validatable hypotheses for treatments and cures.

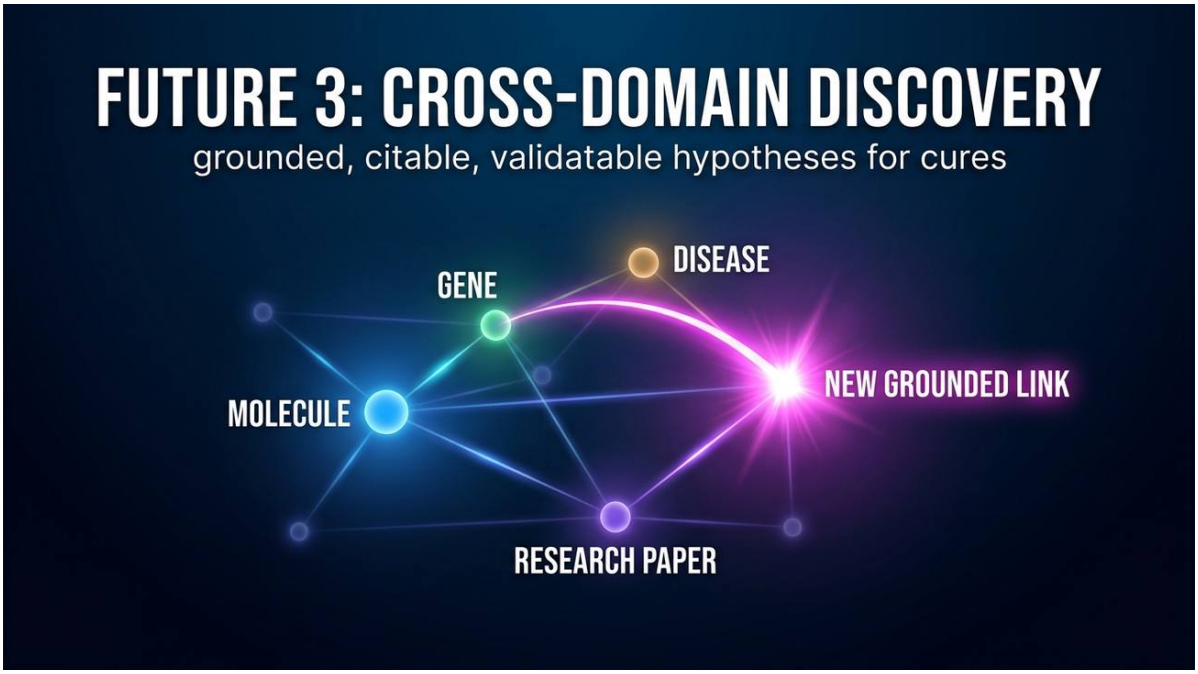


Figure 58. *Future 3 - cross-domain discovery: grounded, citable hypotheses for cures.*

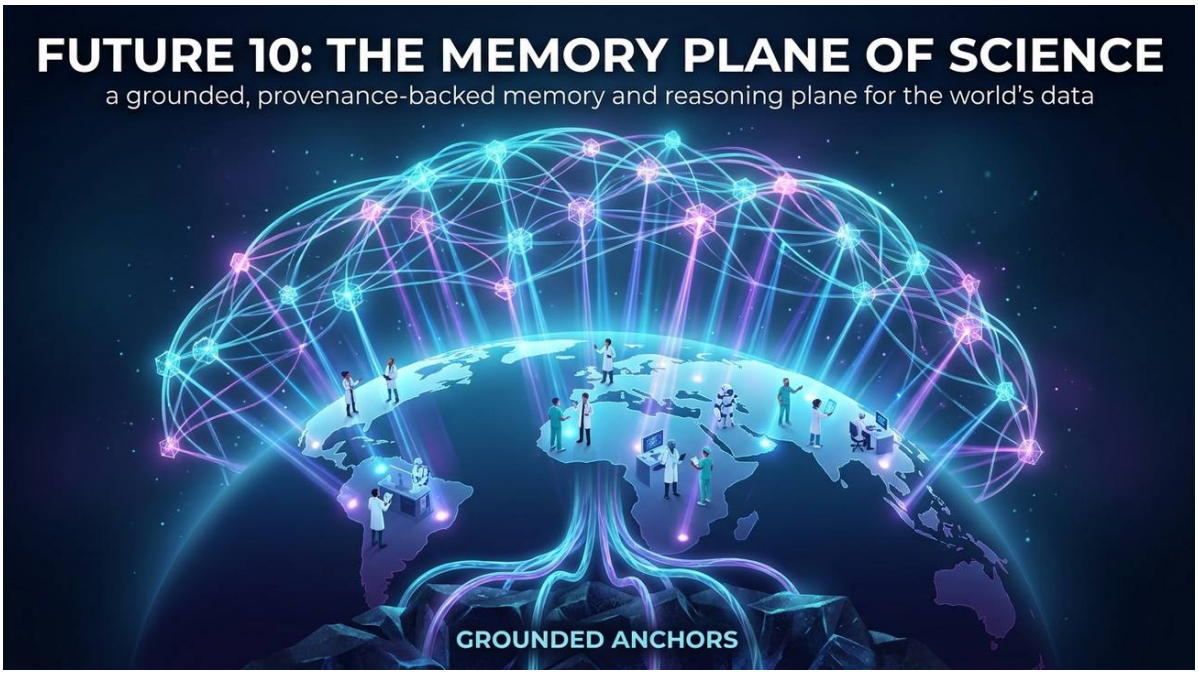


Figure 59. *Future 10 - a grounded, provenance-backed memory and reasoning plane for the world's data.*

- **A grounded memory and reasoning plane for AI agents** — a database that measures, associates, differentiates, and composes, returning provenance-backed results an agent can trust and trace.

Figure 60. *Future 4 - AI you can trust and trace to its exact sources.*

Figure 61. *Future 2 - agents that learn continuously, without costly periodic retraining.*

- **Multimodal, multi-domain general perception** — text, code, image, audio, document, protein, DNA, molecule, and time, fused through dozens of independent lenses into one grounded record.

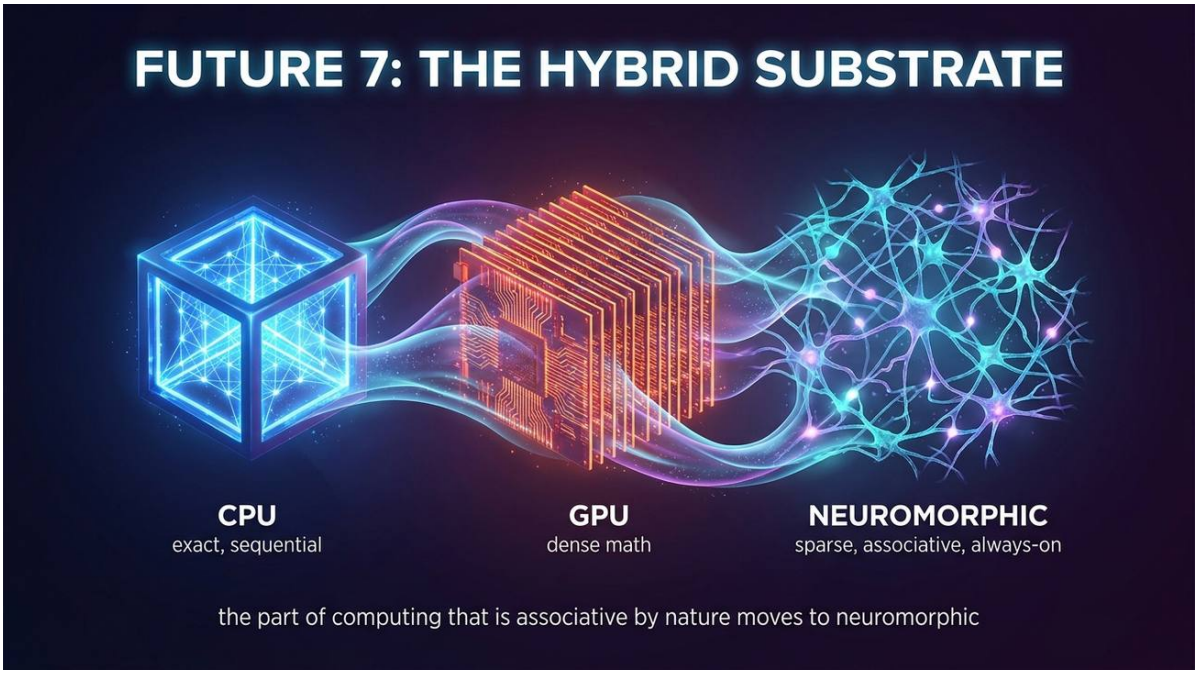


Figure 62. *Future 7 - CPU, GPU, and neuromorphic coexisting as a hybrid substrate.*

Figure 63. *Future 6 - a new neuromorphic developer ecosystem and platform flywheel.*

- **Faithful reconstruction of style, voice, and persona** — capturing the many simultaneous facets of an author or speaker (semantics, syntax, meter, rhetoric, affect, archaism) and composing from them with grounding and provenance.

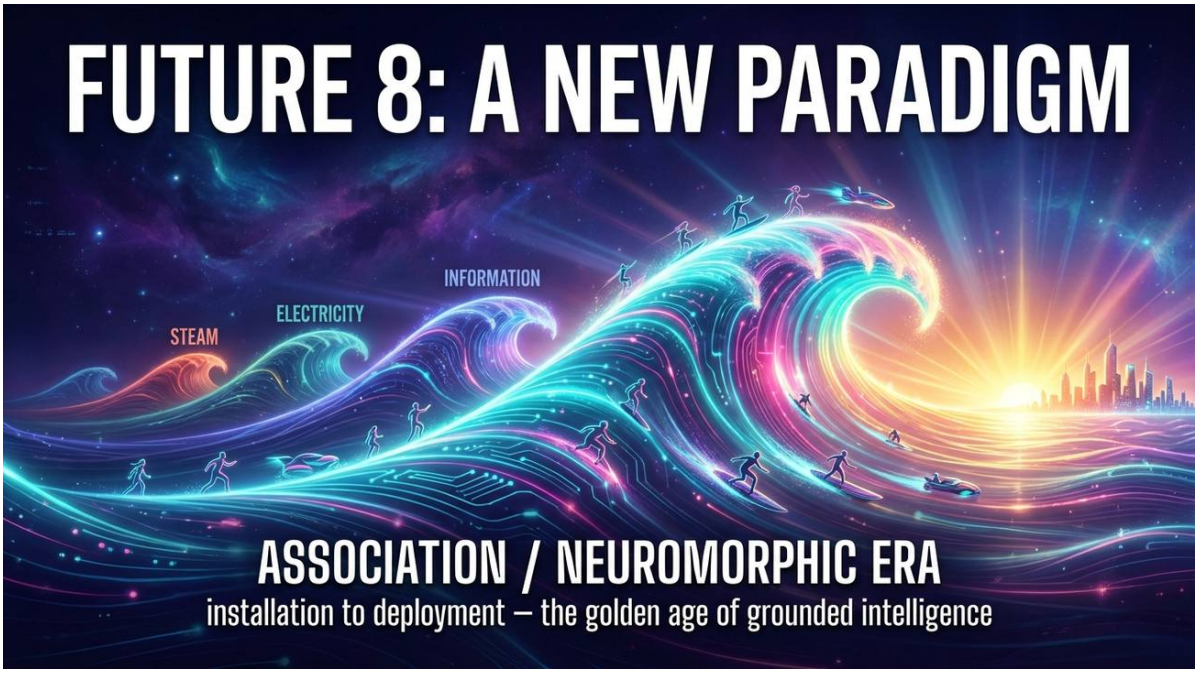


Figure 64. *Future 8 - a new techno-economic paradigm: the golden age of grounded intelligence.*

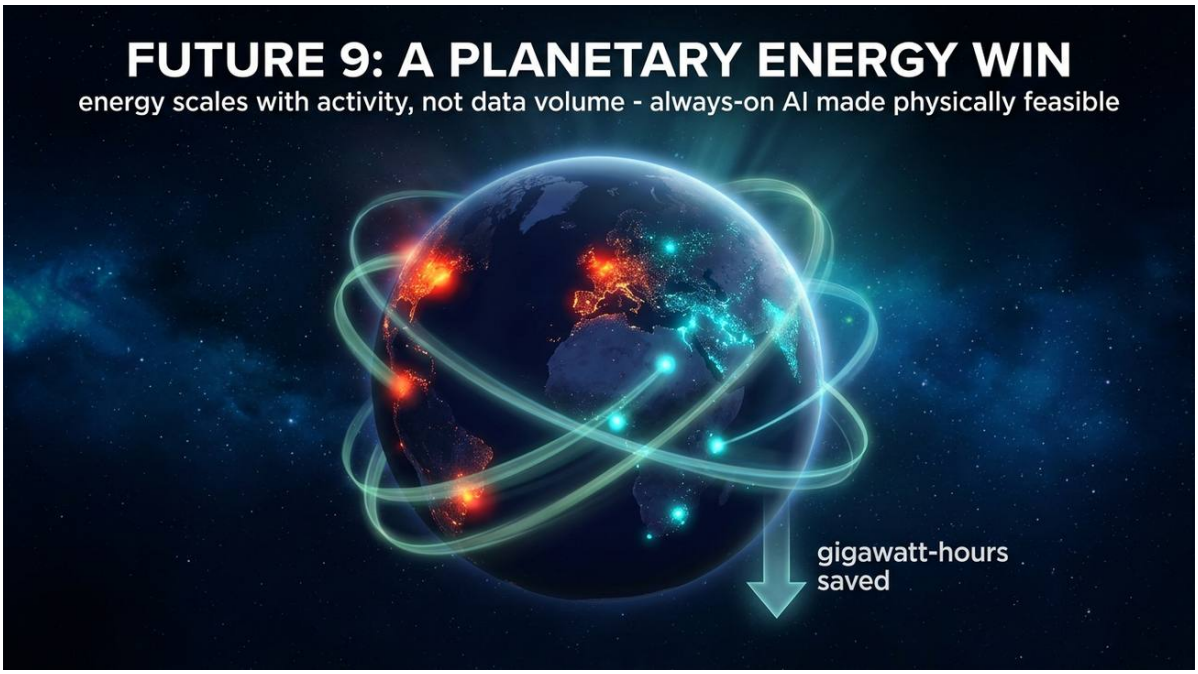


Figure 65. *Future 9 - a planetary energy win making always-on AI physically feasible.*

- **Universal structural summarization** — the kernel as a compact, grounded explanation of any corpus at any scope.

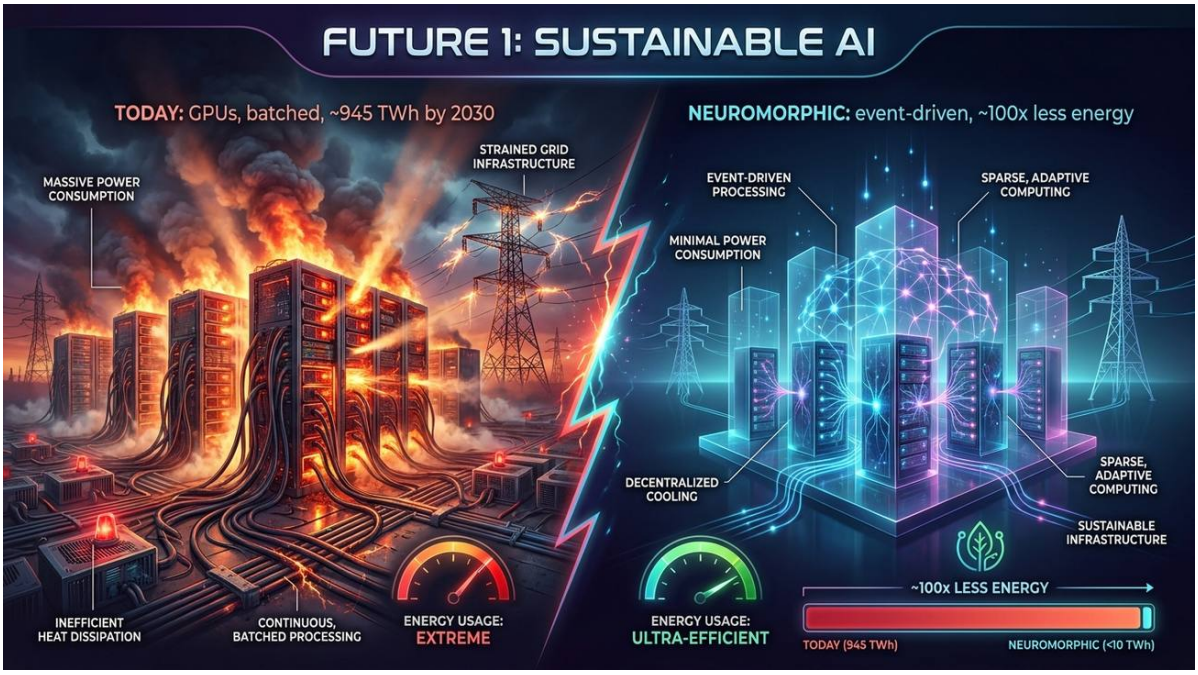


Figure 66. *Future 1 - sustainable AI: the energy wall broken by event-driven computation.*

Figure 67. *Future 5 - a research chip revalued into the most valuable hardware on Earth, because of the software.*

15. What Becomes Possible — A Speculation on the Discoveries and the World They Make

> ***Speculative, and bounded by an honest ceiling.*** *Section 2.5 sets the limit: a system can only surface the grounded information actually present in its corpora (a data-processing- inequality bound). Calyx produces* ***ranked, provenance-backed hypotheses for validation****, never verdicts. What follows is therefore a speculation about possibilities the architecture points toward — not promises — if a planetary, always-on, grounded-association discovery engine runs on efficient neuromorphic hardware at scale.*

The earlier futures (§14) were mostly about the *platform* — energy, economics, ecosystems. This section is about the **output**: the discoveries such a system could make, and the world those discoveries would make in turn. The core mechanism is simple and, at scale, profound: **most of what humanity needs to know next is already latent in what it already knows — sitting unconnected across siloed fields.** A grounded-association engine that can hold the world's data un-flattened, derive the cross-terms between distant domains, and trace every link back to the sources that ground it, is a machine for **finding those connections** — continuously, cheaply, and traceably. That is a new *rate* and a new *kind* of discovery.

15.1 Ten things that become possible

1. Cures hidden in connected literature. The most valuable medical links may already exist, unconnected, across pharmacology, genetics, and the literature. A grounded engine that associates molecule ↔ gene ↔ pathway ↔ paper can surface drug-repurposing and mechanism hypotheses no single field could see — each one citable and testable.

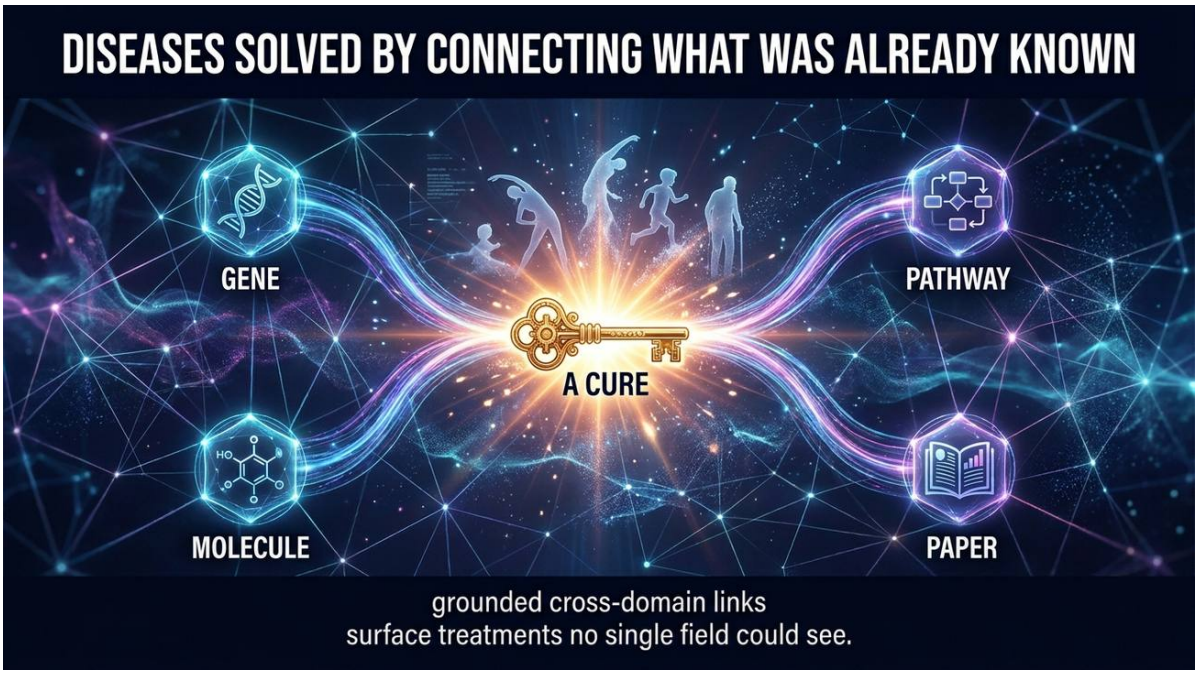


Figure 68. *Cures hidden in connected literature: grounded cross-domain links surface treatments no single field could see.*

2. Materials that did not exist yesterday. Associating physics, chemistry, patents, and experimental records points toward candidate materials — better batteries, catalysts, perhaps superconductors — that no one searched for *directly*, but that the cross-terms imply.

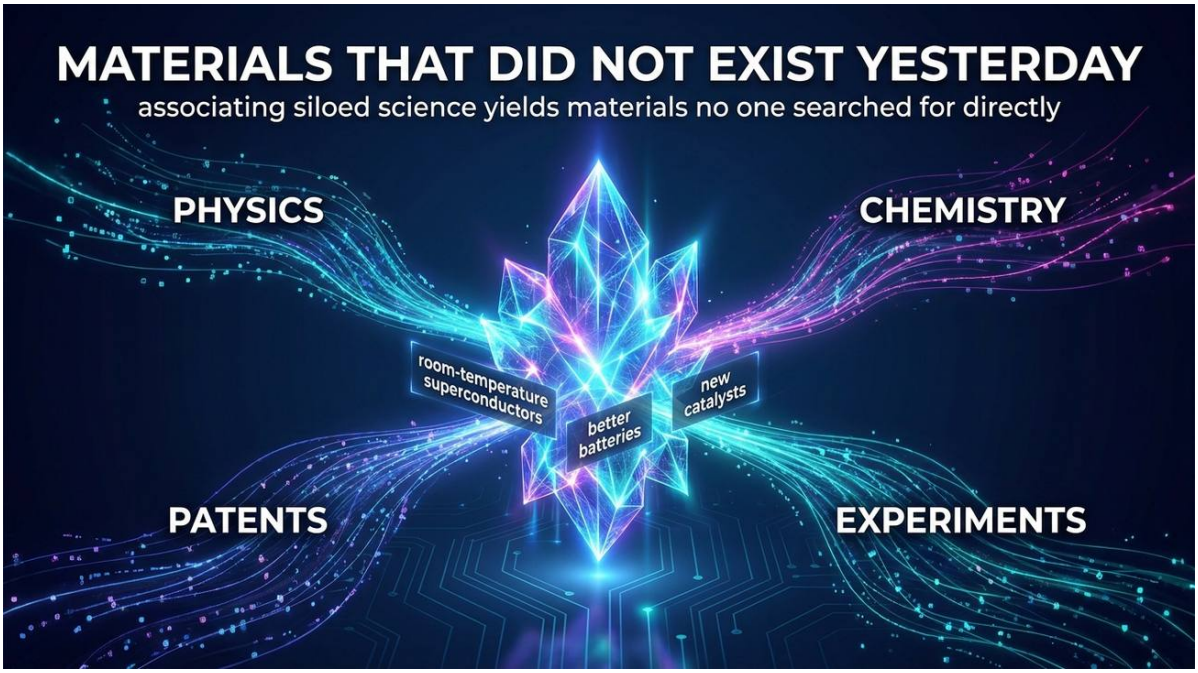


Figure 69. *Materials discovered by association across physics, chemistry, and patents.*

3. A grounded co-pilot for every scientist. Every researcher gains a tireless, cross-disciplinary partner that proposes grounded, cited connections across fields they will never have time to read — ending the siloing that

hides the best ideas.

Figure 70. *A grounded, citable, cross-disciplinary co-pilot for every researcher.*

4. Medicine grounded in all we know. A patient's data, grounded against the whole of medical knowledge, yields diagnoses and treatment hypotheses that are **precise and traceable** — each anchored to the exact evidence that supports it.

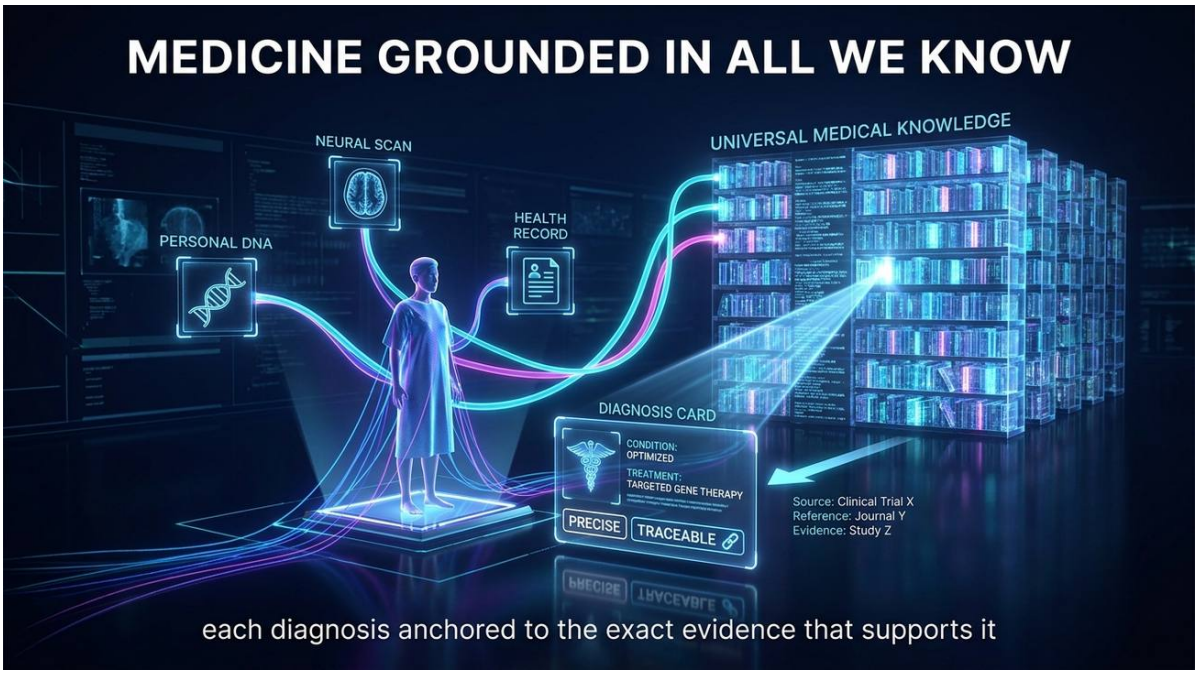


Figure 71. *Medicine grounded in all we know - each diagnosis anchored to its evidence.*

5. Grounded solutions for a living planet. Climate, materials, engineering, and policy data, associated into evidence-backed interventions, turn a fragmented problem into actionable, traceable options.

Figure 72. *Grounded, evidence-backed interventions for climate and energy.*

6. Trustworthy AI in the things that matter. When every answer is grounded and traceable, high-stakes domains — law, medicine, governance — can finally rely on machine reasoning, because the hallucination problem is replaced by *show-me-the-sources*.

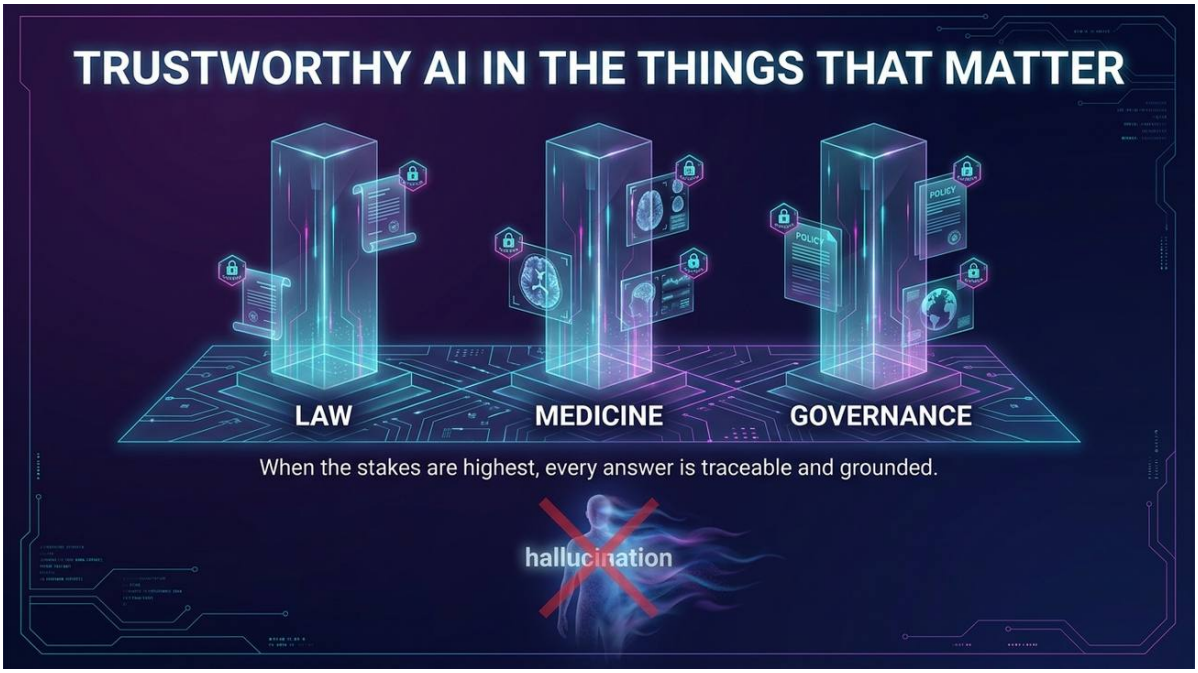


Figure 73. *Trustworthy AI where the stakes are highest - law, medicine, governance.*

7. A memory that never forgets and always cites. A continuously-updating, self-correcting global record of grounded knowledge — contradictions pruned, new facts integrated in $O(1)$, every node carrying its provenance — becomes humanity's shared, living memory.

Figure 74. *A living, self-correcting global memory that never forgets and always cites.*

8. Breakthroughs from a single mind. When cross-domain association is cheap and grounded, a lone researcher or a small lab can make discoveries that once required a giant institution. Discovery decentralizes.

Figure 75. *Discovery democratized - breakthroughs within reach of small labs and individuals.*

9. The pace of discovery itself speeds up. Grounded connections compound: each new link makes the next easier to find. The *rate* of discovery — not just its cost — bends upward, a cascade of findings seeding further findings.

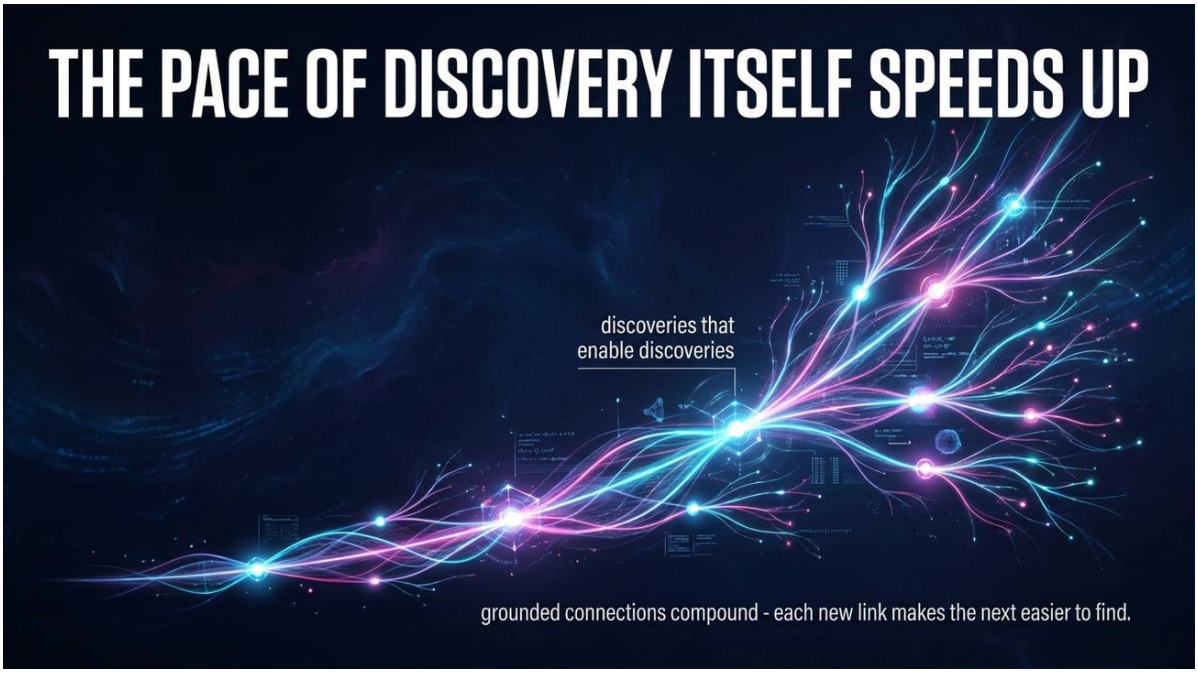


Figure 76. *The pace of discovery itself speeds up as grounded connections compound.*

10. A world that runs on grounded truth. The cumulative effect is civilizational: humans and trustworthy AI as partners, a society whose knowledge is verifiable by construction, and a restoration of trust in what we collectively claim to know.

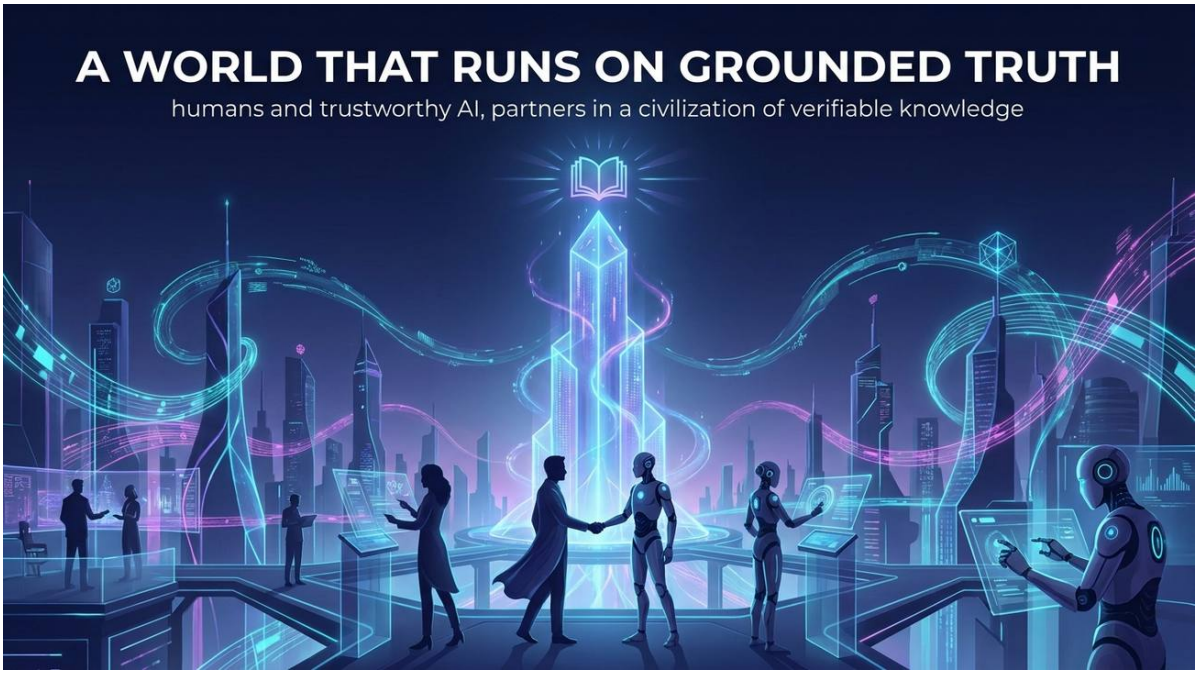


Figure 77. *A civilization that runs on grounded, verifiable truth.*

15.2 How the world looks — three longer speculations

Speculation A — The decade of cures. In the strong case, cross-domain grounded association collapses the front end of discovery: not the wet-lab validation, but the *hypothesis generation* that precedes it. Diseases begin to fall not one at a time but in **clusters**, as the same connective machinery that explains one mechanism illuminates its neighbors. The bottleneck shifts from "what should we try?" to "how fast can we validate?" — and the honest, provenance-backed nature of the hypotheses is what lets regulators and clinicians trust the pipeline enough to run it at scale.

Speculation B — The end of the expert bottleneck. For 2,500 years, synthesizing across fields was the rarest and slowest human skill (a theme §15's companion coda develops). If a grounded engine makes that synthesis cheap, abundant, and *traceable*, the production of knowledge **decentralizes**: the advantage shifts from institutions that can afford armies of specialists to anyone who can ask a good question and validate an answer. Science becomes less like a cathedral and more like a network — and, because each discovery seeds others, the network **compounds**.

Speculation C — The restoration of trust. The deepest change may be epistemic, not material. We are entering an age where *generating* a plausible claim is free (the Epistemic-Symmetry coda, §16). In that world, the scarce and civilization-stabilizing resource is **grounded verification** — and a system that makes it cheap, universal, and fail-closed could end the post-truth drift not by censoring claims but by making it *trivial to check which are grounded*. A civilization that can cheaply tell the grounded from the ungrounded is a different, steadier kind of civilization.

15.3 The honest ceiling (so this stays speculation, not prophecy)

Every line above is bounded by the same discipline that makes Calyx trustworthy: **you cannot surface information that isn't grounded in the data** (the data-processing-inequality ceiling of §2.5); the system yields **hypotheses for validation, not truths**; the quality of what it finds is capped by the quality and breadth of what it is given; and the same connective power that finds cures is **dual-use** and demands governance (§10, §11). These are possibilities the architecture *points toward* and that the staged, fail-closed roadmap (§13) is designed to pursue without ever compromising the principles — grounded, un-flattened, fail-closed — that make any of it trustworthy in the first place.

16. Coda — Epistemic Symmetry: the deeper reason Calyx matters

This coda connects Calyx to a separate but resonant thesis: Chris Royse's *The Symmetry of Knowing: A Theory of Epistemic Transformation in the Age of Artificial Intelligence* (2025). It is speculative, and offered in that spirit.

The thesis. For ~2,500 years, knowledge ran one way: forming the right **question** was the high-effort antecedent, and the **answer** was the reward — "the question is the scarce resource." Royse argues generative AI has broken that arrow.

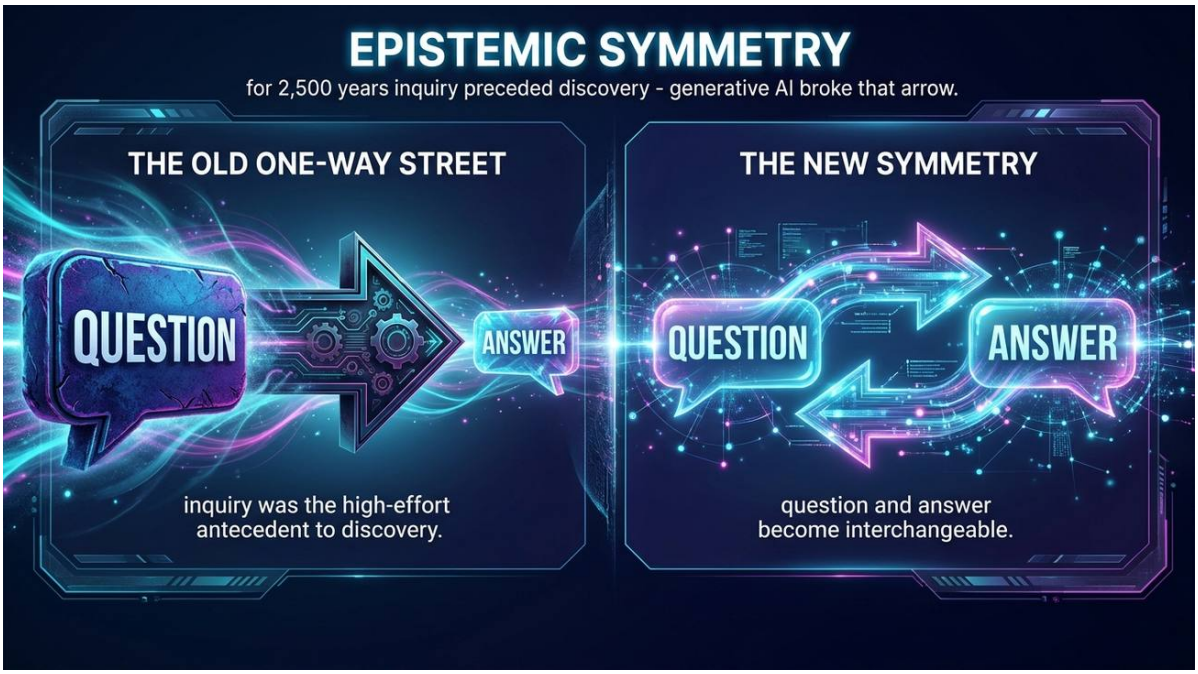


Figure 78. *Epistemic Symmetry: generative AI turns the one-way street of inquiry into a two-way street.*

Through **computational reversibility** (large language models can reverse-engineer the question that produced an answer at >90% semantic fidelity), question and answer become **interchangeable** — a condition he names **Epistemic Symmetry**.

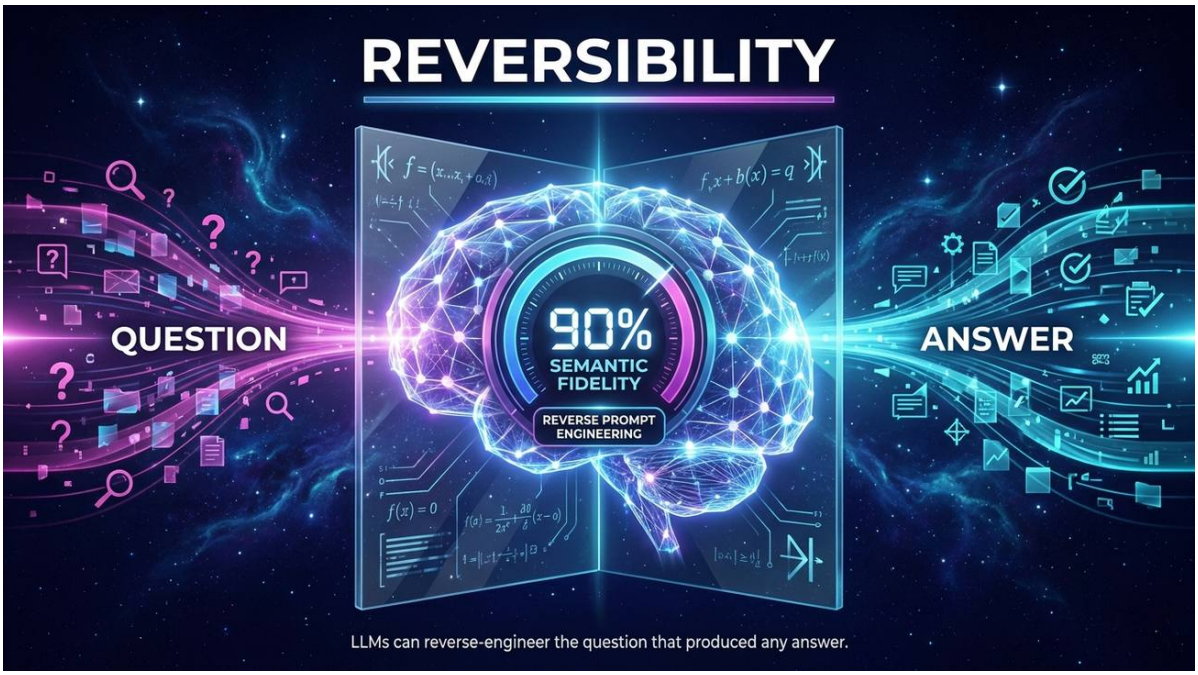


Figure 79. *Computational reversibility: LLMs reverse-engineer the question that produced an answer at high fidelity.*

Three consequences follow: the **friction of inquiry** that drove deep learning collapses; **expertise** shifts from *possessing* knowledge to *verifying* it ("symmetry management"); and the human **status drive**, freed from epistemic labor, migrates to new arenas rather than disappearing.

Figure 80. *Epistemic labor collapse: the friction of inquiry, essential to deep learning, is vanishing.*

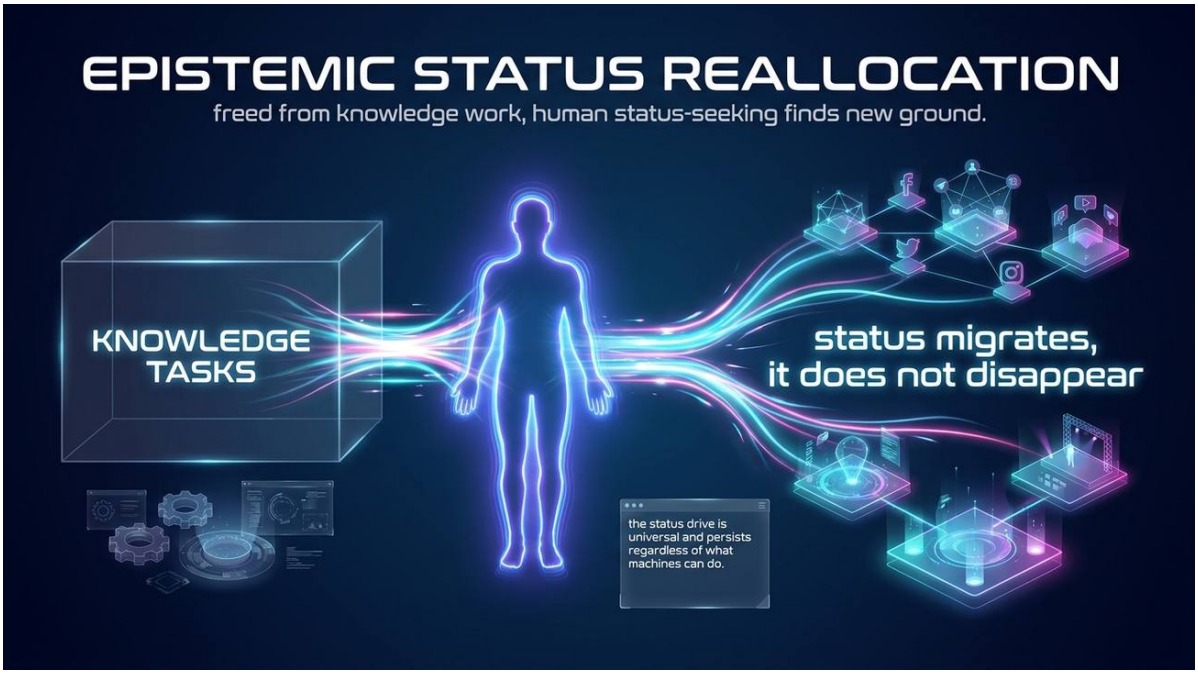


Figure 81. *Epistemic status reallocation: the human status drive migrates rather than disappears.*

Why this is Calyx's deepest justification. Run the symmetry argument to its economic conclusion. If *generating* any question or answer becomes free and reversible, then by simple scarcity the value migrates to the one thing symmetry does **not** make cheap: knowing **which answers are grounded and true**. Verification, provenance, and grounding become the scarce, load-bearing resource of the entire knowledge economy.

Figure 82. *When generating any answer is free, grounded verification becomes the scarce resource.*

That is precisely — and almost uniquely — what Calyx is built to provide: a **grounded, provenance-backed, fail-closed** substrate that labels what is *trusted* versus *provisional* and traces every trusted claim to the exact sources that ground it.

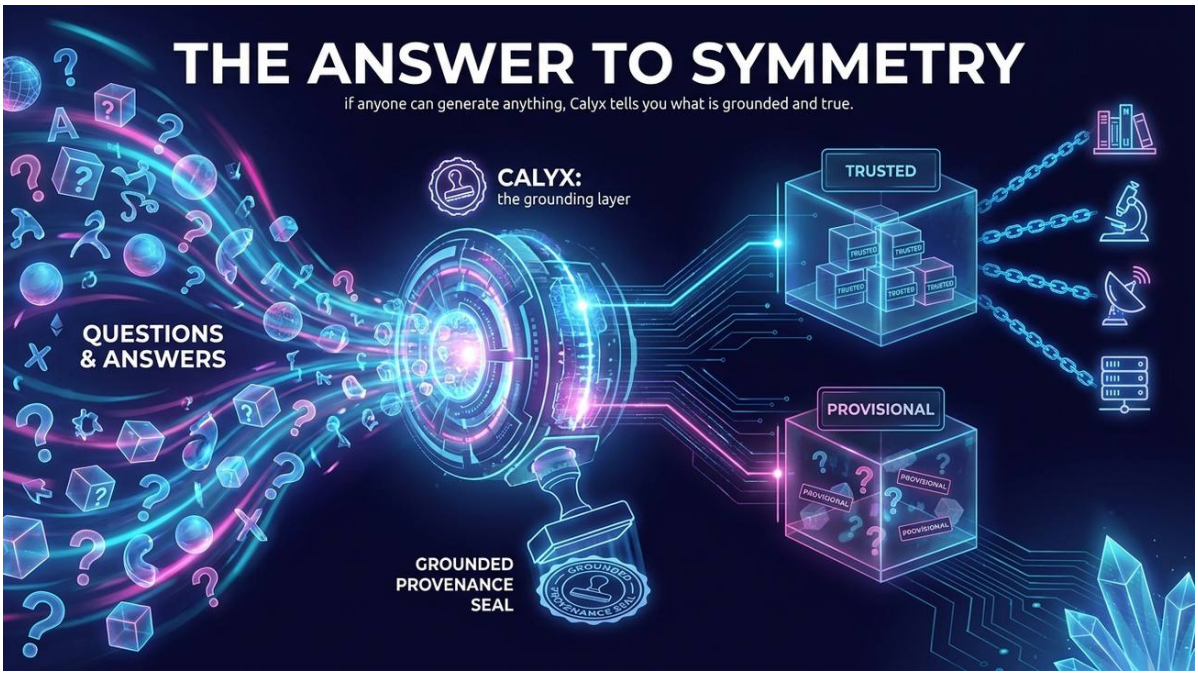


Figure 83. *Calyx as the answer to symmetry: it labels what is grounded and true and traces it to sources.*

In a world of epistemic symmetry, **Calyx is the asymmetry that still matters** — the instrument on which Royse's "verification and symmetry management" expert actually works.

Figure 84. *The expert redefined: from possessing knowledge to verifying it.*

The white paper's three principles (grounding mandatory, no-flatten, fail- closed) are, read through Royse, not merely engineering discipline; they are the **economic moat of the post-symmetry age**: when answers are infinite, grounded truth is the foundation.

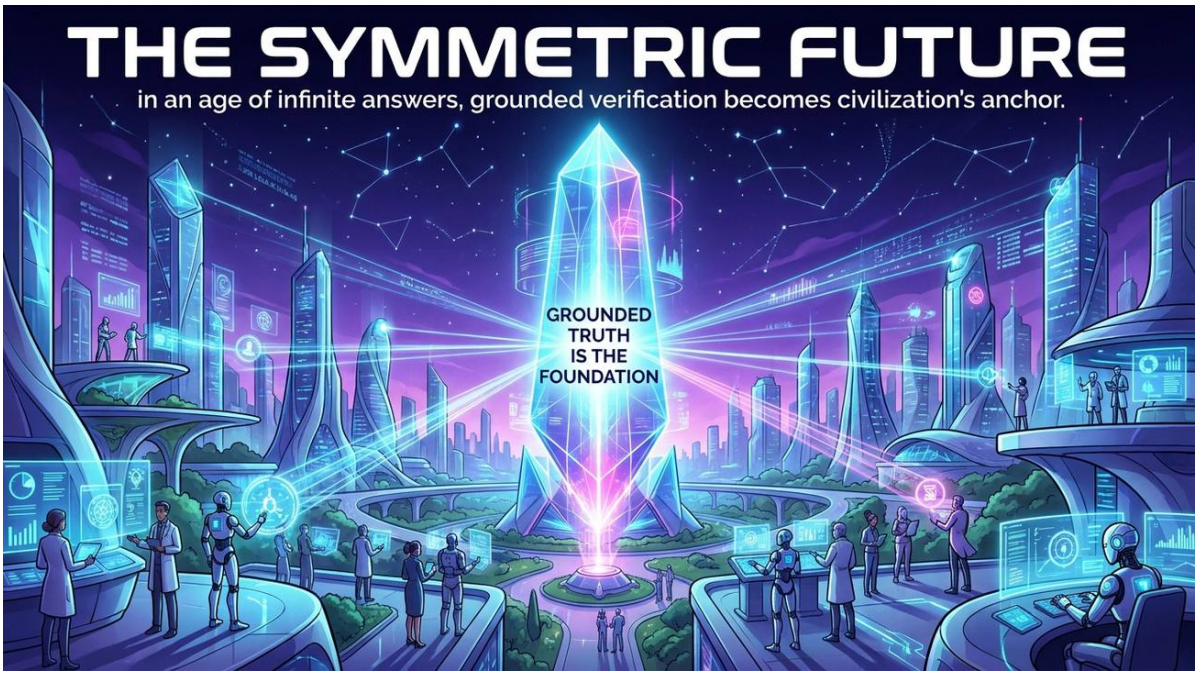


Figure 85. *The symmetric future: in an age of infinite answers, grounded verification is civilization's anchor.*

A testable, falsifiable bridge. Royse offers 33 falsifiable propositions about a symmetric knowledge world; Calyx offers a fail-closed, measurable grounding engine. The natural empirical program is to use Calyx as the *measurement instrument* for epistemic symmetry — quantifying, in bits and provenance chains, how much of

a given answer is grounded versus generated. The two works compose: Royse names the transformation; Calyx is candidate infrastructure for surviving it well. *(A fuller visual treatment, with dedicated illustrations, appears in the companion visual_edition/ PDF, Part 8.)*

17. Intelligence as the Great Equalizer — Calyx and the Living Second Amendment

This paper has argued, in the registers of architecture, physics, economics, and epistemology, that grounded association is the right unit of machine intelligence and that a neuromorphic substrate is its natural home. There is a final register in which the argument lives, and it is the one that matters most for how this technology should be governed and to whom it should belong: the register of **power**. The claim of this section is deliberately large, and it is the author's own conviction, advanced here to be defended *and* examined rather than merely asserted: **the broad distribution of the highest forms of intelligence is the twenty-first-century analogue of the Second Amendment** — a structural check that lets a people retain leverage against concentrated power — and a local, grounded, privately held system such as Calyx is candidate infrastructure for spreading it. The argument is strongest when it is most honest, so this section states the case at full strength and then marks precisely where the established scholarship supports it and where it must be qualified.

Figure 86. *Intelligence as the great equalizer: when a single citizen holds grounded intelligence, the scale against concentrated state and corporate power can be brought back toward balance.*

17.1 The Founders' symmetry, and how it broke

The deepest rationale offered for an armed citizenry was never sport or even ordinary self-defense; it was structural. In *Federalist No. 46* (1788), James Madison set a federal standing army — capped by the cost and population of the era at perhaps twenty-five or thirty thousand men — against "a militia amounting to near half a million of citizens with arms in their hands," and concluded that the "advantage of being armed, which the Americans possess over the people of almost every other nation," was what made the European-style

subjugation of a population infeasible on American soil. The Anti-Federalists (Brutus, the Federal Farmer, Patrick Henry, Elbridge Gerry) made the same point in sharper language: a people that can be disarmed can be ruled, and a people that cannot be disarmed must be *bargained with*. Sanford Levinson's "The Embarrassing Second Amendment" (*Yale Law Journal*, 1989) revived this reading for modern constitutional scholarship as the "insurrectionary" theory of the amendment. The premise underneath all of it is **symmetry of force**: in 1789 the citizen and the soldier held, quite literally, the same weapon. A musket faced a musket. The balance was real because the tools were equal.

That symmetry is gone, and its disappearance is not a matter of opinion. As military technology compounded — rifled artillery, the machine gun, the tank, the bomber, the jet, the precision missile, the armed drone, the reconnaissance satellite, and now the algorithmic surveillance state — none of it was placed in the hands of the people. The state's side of the scale grew by orders of magnitude while the citizen's side did not move. The result is an asymmetry so total that the United States Supreme Court itself conceded it. In *District of Columbia v. Heller* (2008) — the very decision that established an *individual* right to keep and bear arms — Justice Scalia wrote that "it may well be true today that a militia, to be as effective as militias in the 18th century, would require sophisticated arms that are highly unusual in society at large," and acknowledged "that no amount of small arms could be useful against modern-day bombers and tanks." This is the thesis's own diagnosis, delivered from the bench: small arms can no longer balance the state.

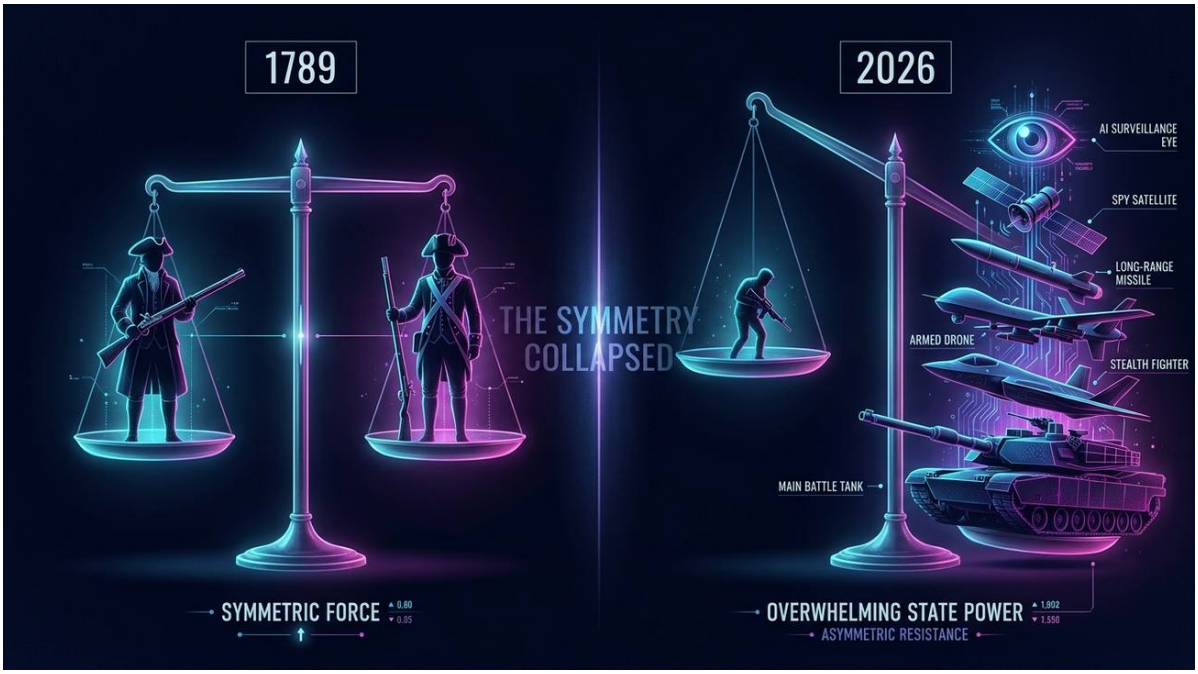


Figure 87. *The collapsed symmetry: in 1789 citizen and soldier held the same musket; by 2026 small arms offer near-zero resistance against tanks, jets, drones, satellites, and AI surveillance - a point even Heller concedes.*

The empirical record of the present decade makes the point in blood rather than doctrine. An armed populace is not, in fact, a deterrent against a modern military apparatus. Max Weber's definition of the state as the human community that "successfully claims the monopoly of the legitimate use of physical force within a given territory" (*Politics as a Vocation*, 1919) describes a monopoly that modern firepower has made nearly physical as well as legal. Honesty requires two corrections to the strong form of the claim, both of which the scholarship insists upon. First, the cleanest documented cases of an armed state overwhelming its opponents are *state-versus-state* exchanges and the suppression of *largely unarmed* civilian protest, not an armed

populace fighting its own government — so the evidence supports the *spirit* of "guns no longer balance the state" more cleanly than its literal letter. Second, possessing overwhelming firepower does not guarantee political victory: Ivan Arreguín-Toft's *How the Weak Win Wars* (2001) shows strong actors losing roughly half of asymmetric conflicts since 1950 — but they lose through *political exhaustion and strategic interaction*, not through firepower parity. The material point stands; the citizen's musket is now a symbol, not a counterweight.

It is worth conceding, too, that powerful new technology has not historically *leveled* power as often as it has *concentrated* it. The economic historians of gunpowder — Adam Smith, Tonio Andrade in *The Gunpowder Age* (2016) — find that firearms ultimately destroyed the dispersed armed nobility and concentrated power in wealthy, organized states; "the weak, the poor, the badly organized" fell away. Elizabeth Eisenstein's *The Printing Press as an Agent of Change* (1979) shows the press leveling clerical knowledge monopolies *and* arming states and churches with propaganda and censorship in the same motion. Any claim that a new "equalizer" will democratize power must explain why it breaks this pattern rather than repeating it.

Figure 88. *The great equalizer through history - gunpowder, the printing press, the nuclear bomb, and now intelligence - noting honestly that each leveling technology also reshaped who holds power.*

17.2 Intelligence is the new equalizer

If guns no longer equalize, what does? The answer this paper proposes is **intelligence** — understood precisely as *the ability to find the answer you seek to your problem, however hard the problem*. Francis Bacon's *Novum Organum* (1620) located the joint where "human knowledge and human power meet in one"; Michel Foucault made the entanglement constitutive, arguing in *Discipline and Punish* (1975) that knowledge and power produce one another and that asymmetric visibility — the Panopticon, "visibility is a trap" — is the modern form of domination. Shoshana Zuboff gave the contemporary diagnosis its sharpest edge in *The Age of Surveillance Capitalism* (2019): the defining feature of the present order is an "unprecedented asymmetry in knowledge and the power that accrues to knowledge" — "they know everything about us, whereas their operations are designed to be unknowable to us." If the asymmetry that subjugates is now an asymmetry of *knowledge*, then the equalizer that matters is whatever closes that asymmetry — whatever lets an ordinary

person *find out*, reason, and verify at a level previously reserved for states and large institutions.

This reframes the entire stakes of a system like Calyx. A grounded, association-native intelligence that runs **locally — air-gapped, private, unsupervised, powered from an ordinary wall outlet** — is not merely a convenience or a privacy feature. It is the technical condition under which an individual can hold, in their own home and beyond the reach of any surveillance apparatus, the capacity to find answers to their problems, their community's problems, and their nation's problems. Amartya Sen's capability approach (*Development as Freedom*, 1999) supplies the right justification: what matters is not owning a tool but the *substantive freedom* that access to it confers — though Sen's own caution applies, that resources convert into capabilities only when the "conversion factors" (literacy, infrastructure, the skill to use the tool well) are distributed alongside the tool. Article 19 of the Universal Declaration of Human Rights already names a right "to seek, receive and impart information and ideas through any media and regardless of frontiers." The thesis extends that right from non-interference toward genuine access, and insists on its universality: this capacity is the right of **every citizen on the planet** — not of one nation, one company, or one class — exactly as the logic of a fundamental check on power demands.

Figure 89. *The air-gapped citizen: the greatest intelligence running locally and privately from a wall outlet, beyond the reach of any surveillance state.*

17.3 Possession, not use — the nuclear logic, stated precisely

The Second Amendment's deepest insight was never that citizens *should* fire on their government; it was that the *capacity* to resist changes how the governed are treated. The mature form of this idea is nuclear deterrence theory, and it is the right lens for the equalizer thesis. Bernard Brodie wrote in 1946 that the bomb's "chief purpose must be to avert" war, not to win it. Thomas Schelling, in *Arms and Influence* (1966), located leverage in latent capability: "the power to hurt is most successful when held in reserve… it is the *threat* of damage, or of more damage to come, that can make someone yield." Kenneth Waltz argued in *The Spread of Nuclear Weapons: More May Be Better* (1981) that possession induces caution and that *weak* possessors gain the most. The observable pattern follows: nuclear-armed states are treated with a wariness that non-nuclear states are not. The contrast between Libya — which surrendered its weapons program in 2003 and saw its

government destroyed in 2011 — and a nuclear North Korea that has been negotiated with rather than invaded is the case study every chancellery has internalized. Nobody fires these weapons. They are valued precisely *because* they are never used. Possession alone reorders the table.

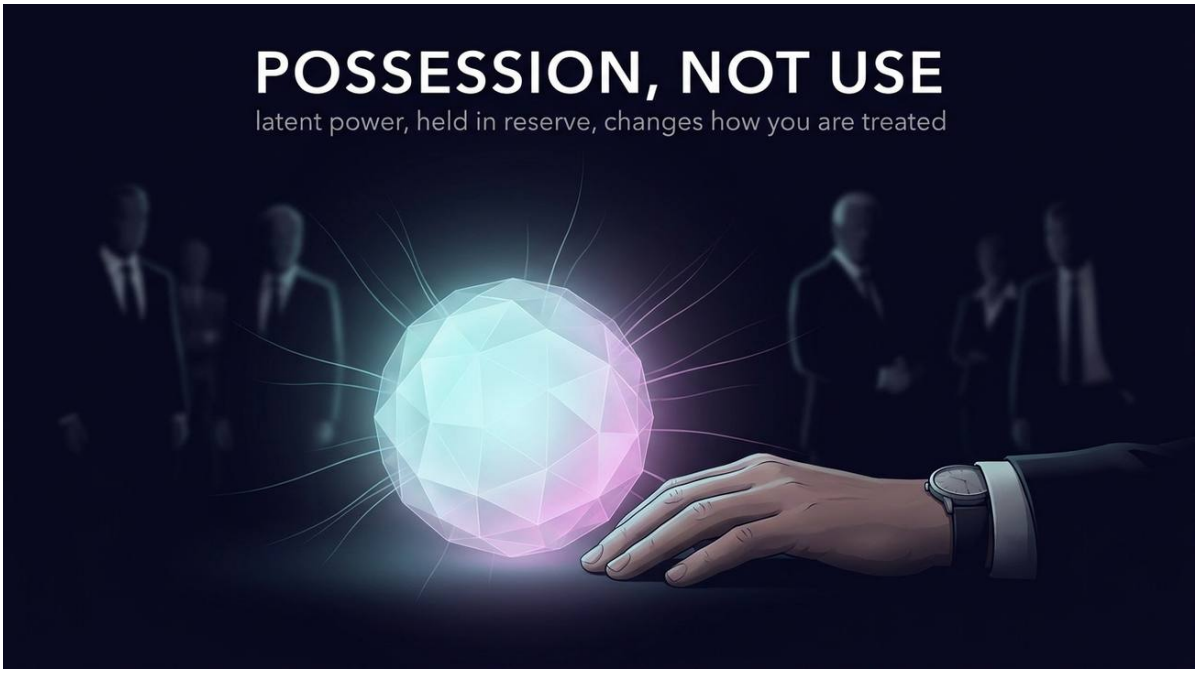


Figure 90. *Possession, not use: like a deterrent held in reserve, mere possession changes how the powerful treat you - though the scholarship cautions this buys defensive deterrence, not coercive leverage (Sechser & Fuhrmann, 2017).*

Here the thesis must be disciplined by the strongest finding in the literature, or it overreaches. Todd Sechser and Matthew Fuhrmann's *Nuclear Weapons and Coercive Diplomacy* (2017), analyzing more than two hundred compellent threats, find that nuclear powers enjoy **"deterrence, but not coercion"**: possession reliably deters attacks *on the possessor*, but does *not* confer the power to extract concessions, even against non-nuclear adversaries and even with nuclear superiority. The honest version of the equalizer claim is therefore *defensive*: broadly held intelligence does not let a citizen dictate to the state, but it raises the expected cost of acting against an informed, capable population — it restores a deterrent floor and forces decision-makers to weigh the possibility of consequence. That is the achievable and defensible prize, and it is precisely the prize the Second Amendment was meant to secure: not the people firing on the government, but the government governing as though the people retain leverage. A second caution is structural. Kenneth Waltz himself warned against "reductionist" theories that move from the systemic level to the individual; an individual lacks the survivable second-strike and the mutual vulnerability that make state-level deterrence stable. The nuclear analogy is therefore a heuristic for *aggregate* deterrence — a population that can collectively find and publish the truth — not a literal claim that one person with a laptop deters a state the way one state deters another.

17.4 Threat vectors and accountability — why this restores the check

With that precision in hand, the mechanism becomes clear. Concentrated power — whether a government or a large institution — is most secure when it operates in an *epistemic monopoly*: when only it can see the whole board, model the consequences, and find the levers, while the governed cannot. Broadly distributed intelligence dissolves that monopoly. It opens governments and large institutions to effectively **unlimited**

asymmetric vectors of scrutiny and leverage: any individual, anywhere, can now find the inconsistency, trace the provenance, model the second-order effect, surface the buried fact. Joseph Nye's *The Future of Power* (2011) names this directly as **power diffusion** — distinct from power *transition* between states — and judges that "a much larger part of the population… has access to the power that comes from information," to the point where "power diffusion may be a greater threat than power transition." The cyber domain is his paradigm case of a "low price of entry" that hands disproportionate leverage to small actors. The point is not violence; it is *accountability*. When the powerful must assume that any decision can be found out, modeled, and answered, they decide differently. That altered calculus — *they must act as though there can be a response* — is the entire civic purpose of a structural check.

Figure 91. *Unlimited threat vectors: when anyone can find the answer, concentrated power becomes exposed and accountable - distributed, non-violent leverage, in the spirit of Nye's diffusion of power.*

This is also where the thesis must absorb its most important political-science caveat, and is improved by it. The canonical accounts of how societies actually restrain tyranny locate the check not in atomized individual capability but in *organized, collective, institutional* counter-power. Daron Acemoglu and James Robinson's *The Narrow Corridor* (2019) argues that liberty survives only inside a narrow corridor held open by a balanced contest between a strong state and a *strong, mobilized society* — the "Red Queen" effect, in which both must keep running to stay in place. John Locke grounded the people's leverage as a *collective* last resort against a government that breaks its trust; Montesquieu located the check in institutional architecture, "power should be a check to power." The correct reading of the equalizer thesis, then, is not that each individual should wield catastrophic capability, but that broadly distributed intelligence is what keeps *society as a whole* strong enough and informed enough to remain in the corridor — to match the state's growing analytical and surveillance capacity with a citizenry that can see, reason, and verify for itself. Calyx's contribution is a *collective* resilience that emerges from *individual* access; the check is societal, even though the instrument sits on a private desk.

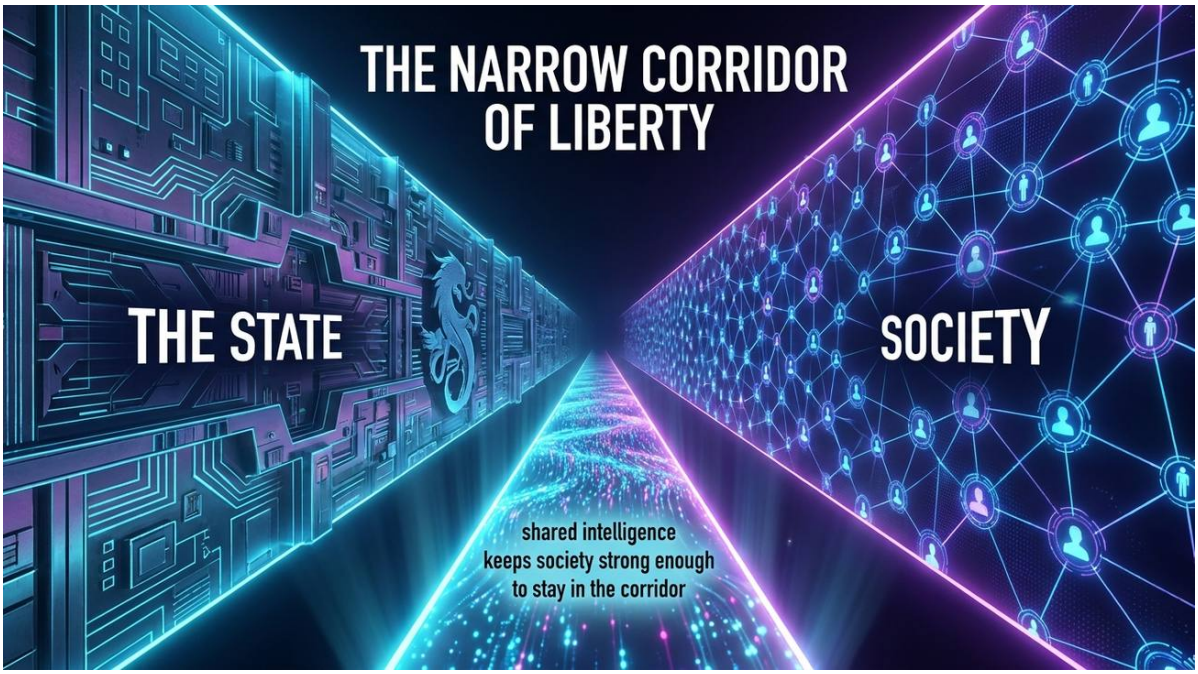


Figure 92. *The narrow corridor of liberty (Acemoglu & Robinson, 2019): freedom survives only when a strong society keeps pace with a strong state - shared intelligence helps society stay in the corridor.*

17.5 The shadow — the same paradox, applied to the equalizer itself

Intellectual honesty requires that the equalizer thesis be turned on itself, because the same property that makes distributed intelligence emancipatory makes it dangerous. The connective engine that finds a cure hidden across the literature can, in principle, find a pathogen by the same operation; the system that surfaces a government's buried abuse can surface an individual's private life. Audrey Kurth Cronin's *Power to the People* (2020) calls this "lethal empowerment" — "never have so many possessed the means to be so lethal" — and observes that the very accessibility the thesis celebrates is, viewed from the other end, the proliferation of harm. Thomas Friedman's image of the "super-empowered individual" captures both faces in one figure: the same diffusion that empowers the activist empowers the "super-empowered angry man." David Collingridge's dilemma warns that once a dangerous capability is universally diffused, the window to control it has already closed. And Nick Bostrom's "Vulnerable World Hypothesis" (2019) supplies the reason the AI- and bio-enabled downside differs in *kind* from firearms: a gun's harm is bounded and roughly linear in users, whereas a sufficiently capable, offense-dominant, irreversible technology could make catastrophe the default of broad diffusion. This is the genuine disanalogy with the Second Amendment, and it cannot be waved away.

Figure 93. *Every equalizer casts a shadow: the same connective engine that finds a cure can find a catastrophe - the dual-use danger (Cronin; Bostrom) the Ethical Probe Paradox exists to confront.*

This is exactly the terrain of the Ethical Probe Paradox developed in the next section (§18): we cannot know where the ethical line lies without approaching it, and approaching it may require crossing it. The resolution the equalizer thesis offers is not denial of the shadow but *the discipline of how the capability is built*. The case for distributing intelligence through a system like Calyx — rather than through an ungoverned, free-floating capability — rests precisely on Calyx's architectural commitments: grounding is mandatory, results are provenance-traced to their sources, the substrate is reversible and auditable, and the guard is fail-closed. A capability that is *grounded and traceable by construction* is the instrumented probe of §18.4: it lets society approach the line of dangerous knowledge while logging every step, attributing every claim, and refusing to act when grounding fails. The equalizer worth spreading is not raw, anonymous, unaccountable power; it is the capacity to *know what is true and to prove it* — which is, not coincidentally, the one form of power that defends as readily as it attacks.

17.6 The honest position

Stated at full strength and then disciplined by the scholarship, the thesis comes to rest here. What is well grounded: the *descriptive* collapse of citizen–state symmetry (conceded in *Heller* and anticipated by Wendy Brown in 1989); the logic of *possession-not-use* as a real foundation of deterrence (Brodie, Schelling, Waltz); the identity of knowledge-asymmetry with power-asymmetry (Bacon, Foucault, Zuboff); the diffusion of power to smaller actors (Nye); and the legitimacy of the people retaining structural leverage against concentrated power (Locke, Montesquieu, Acemoglu and Robinson, UDHR Article 19, Sen). What must be qualified: possession buys *defensive deterrence*, not *coercion or deference* (Sechser and Fuhrmann); the check that protects liberty is *collective and institutional*, so the instrument's value is the societal resilience it produces, not atomized individual capability; diffusion is *not* automatic equalization (Horowitz, Nye), and powerful technologies have historically concentrated power as often as they have spread it; and the catastrophe potential of advanced intelligence is categorically larger than that of firearms, which is why the capability must be *grounded, traceable, reversible, and fail-closed* rather than raw. With those qualifications

honored, the core conviction survives intact and is, the author submits, among the most important reasons this work exists: **the right to access the greatest intelligence — locally, privately, and universally — is the living form of the principle the Second Amendment was written to protect, and Calyx is built to be infrastructure worthy of carrying it.**

18. The Ethical Probe Paradox — Where Is Calyx's Line?

> ***Status of this section: speculative ethical analysis.*** *Sections 8–17 enumerated, at full strength, everything this white paper speculates and predicts — including the great-equalizer thesis of §17. This section subjects every one of those speculations to a single ethical test — Chris Royse's* ***Ethical Probe Paradox*** *(Royse, The Ethical Probe Paradox: Must We Cross the Line to Find It?, A.Q. Miller School of Media and Communication, Kansas State University, March 2026, DOI 10.13140/RG.2.2.12257.57449). It is the author's own ethical framework, applied here to the author's own speculative work. It claims no verdicts; like the rest of §8–§17 it offers a disciplined, theory-anchored analysis, and — consistent with Calyx's grounding principle — it states plainly what it does not know.*

18.1 The paradox

Ethics, Royse argues, was built on an unstated assumption: that the moral agent is **human** — slow, forgetful, limited in reach, bounded by biology. Those limits were not incidental; they were **constitutive** of what made whole categories of activity ethically tolerable. Persuasion was tolerable because a persuader could reach only so many people, knew only so much about each, and tired. Surveillance was tolerable because a watcher could watch only so much. Information control was tolerable because no human could be everywhere at once. The boundary of the acceptable was, in practice, drawn by the boundary of the possible.

Artificial intelligence removes those limits, and the evidence that it does is no longer speculative. A 2024 *Nature Human Behaviour* randomized controlled trial (N = 820) found that GPT-4, given only basic demographic information about its opponent, had **81.7% higher odds** of shifting a human toward post-debate agreement than a human debater did. A 2025 *Science* study (76,977 responses across 19 large language models) found that AI persuasiveness **scales with capability** — post-training techniques raised persuasive impact by as much as **51%** — and, chillingly, that the *most persuasive* configurations produced the *most inaccurate* claims. The World Economic Forum's *Global Risks Report 2025* ranked AI-amplified misinformation and disinformation the **number-one short-term global risk**. The EU AI Act has already responded by banning whole categories of AI practice outright. The human limits that made certain acts tolerable are simply gone.

This is the setup for the paradox. Take any task. Call the human level of capability at that task **X**, and assume X is ethical — a salesperson persuades, a clinician advises, an analyst connects facts. Call the AI level of capability at full deployment **Y**, and assume Y, pushed to its limit, is unethical — persuasion of millions, each profiled and mirrored, none aware they are being worked on. Somewhere on the spectrum from X to Y there is a threshold where ethical becomes unethical. **The paradox is that we do not know where that threshold is, we cannot know without approaching it, and approaching it may itself require doing the very thing that is, by definition, unethical.** Royse states it as a question: *do we have to be unethical in order to be ethical?*

Figure 94. *The ethical probe paradox: we cannot locate AI's ethical boundary without approaching it - and approaching it may require crossing it.*

Figure 95. *Somewhere between X (human scale, ethical) and Y (AI at full capacity, unethical) lies an unknown threshold.*

Two features of the threshold make it genuinely hard rather than merely unknown. The first is the **Sorites problem** — the paradox of the heap. Remove one grain from a heap and it is still a heap; repeat, and there is no single grain whose removal turns heap into non-heap, yet the heap plainly vanishes. AI ethics has the same structure: the transition from X to Y is *gradual*, and any specific cutoff one proposes feels arbitrary. One salesperson reading one customer is fine. An AI that profiles *everyone*, mirrors each person's optimal

persuasive persona, and scales to millions of people who do not know they are speaking to a machine is what Floridi calls **hypersuasion** — and is plainly not fine. But there is no sharp grain between the two.

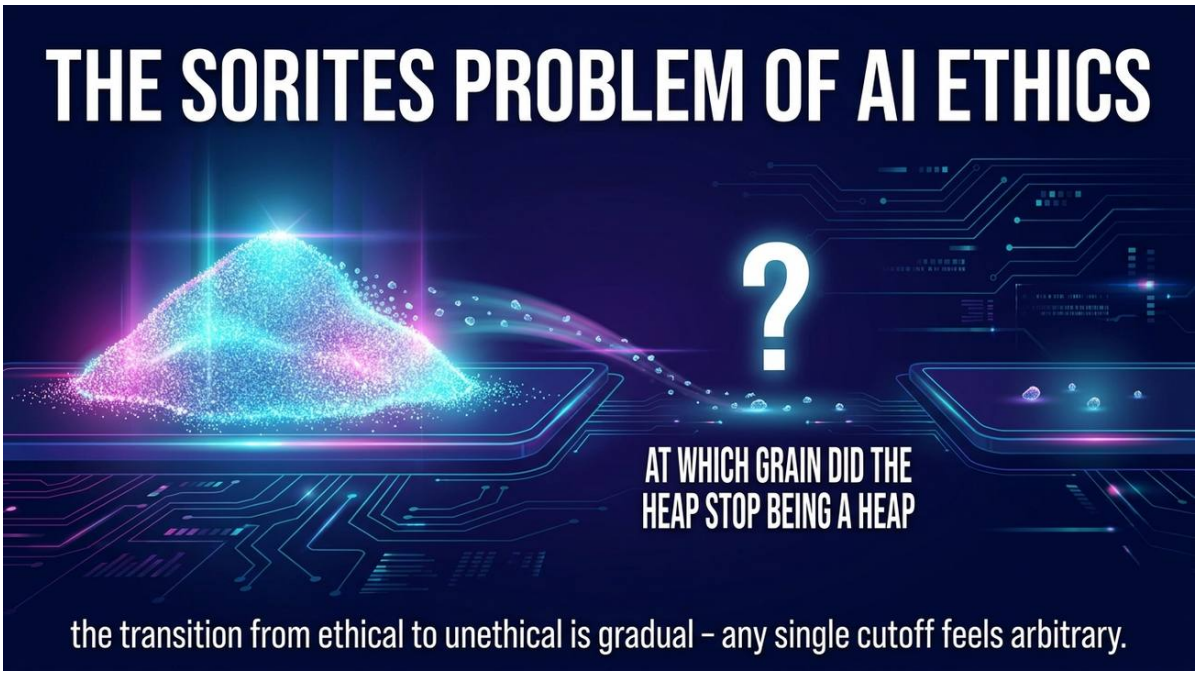


Figure 96. *The Sorites problem: the transition from ethical to unethical is gradual; any single cutoff feels arbitrary.*

The second feature is historical, and it is the paradox's most unsettling claim: **ethics is discovered empirically, through harm, not deduced in advance.** Hammurabi's code fixed the builder's liability *after* houses had collapsed and killed their occupants. The Nuremberg Code (1948) was written *after* the camps. The Belmont Report (1979) followed *after* Tuskegee. Modern financial regulation (2008–onward) arrived *after* the crash. The pattern is invariant: **harm first, norm second. The boundary was discovered by crossing it.** If that is how humanity has always found its ethical lines, then the probe paradox is not a peculiarity of AI; it is the general human condition, now operating at a speed and scale where crossing the line first may no longer be survivable.

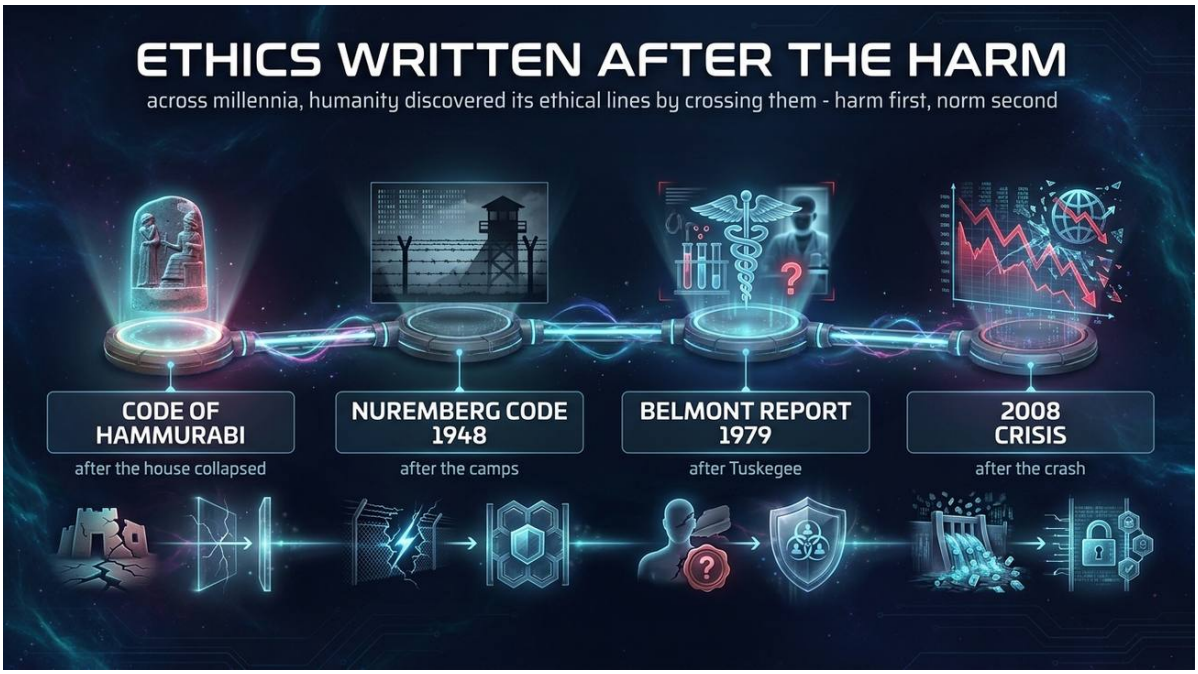


Figure 97. *Across millennia, humanity discovered its ethical lines by crossing them - harm first, norm second.*

Each inherited ethical tradition strains under this load. **Consequentialism** must suddenly weight *scale* as a first-class term — Bentham's long-ignored seventh dimension of utility, **extent**, the number of persons affected, becomes the dominant variable when one agent can act on billions. **Kantian** ethics turns literal: the universal-law test ("what if everyone did this?") was a thought experiment, but an AI that manipulates everyone *simultaneously* makes the universalization real, not hypothetical. **Virtue ethics** has no vocabulary at all for a non-human agent acting at superhuman scale — there is no character, no *phronesis*, no mean between excess and deficiency for an entity that is not a person and does not have a life to live well. The tools we would use to locate the line were all built for an agent the line no longer describes.

18.2 Why Calyx is a maximal case of the paradox

Most of this white paper has been an argument that Calyx could become a **planetary, always-on, continually-learning grounded-association engine**. Read through Royse's framework, that is precisely the profile of a maximal probe-paradox case. Its human-scale version is not merely benign — it is admirable: a librarian who connects two papers, a clinician who links a symptom to a mechanism, an analyst who notices that two facts belong together. That is X, and X is good. Its machine-scale version — the same connective act performed continuously, over the whole of the world's data, across every domain at once, for anyone who asks — is Y, and Y is exactly the capability whose ethical status the paradox says we cannot read off in advance.

There is a sharper twist specific to Calyx. The entire value proposition of §15 and §16 is that **grounded, provenance-backed answers are more persuasive** *because* they are grounded — "show me the sources" is meant to be the antidote to hallucination. But persuasiveness is precisely the axis the *Science* study warns about, and a system engineered to make its outputs maximally credible is, by the same token, a system engineered to be maximally persuasive. Grounding does not exempt Calyx from the hypersuasion risk; it **sharpens** it. The honest reading is that the feature this paper sells as a trust mechanism is, at scale, also the mechanism that would make Calyx the most persuasive instrument the paradox contemplates.

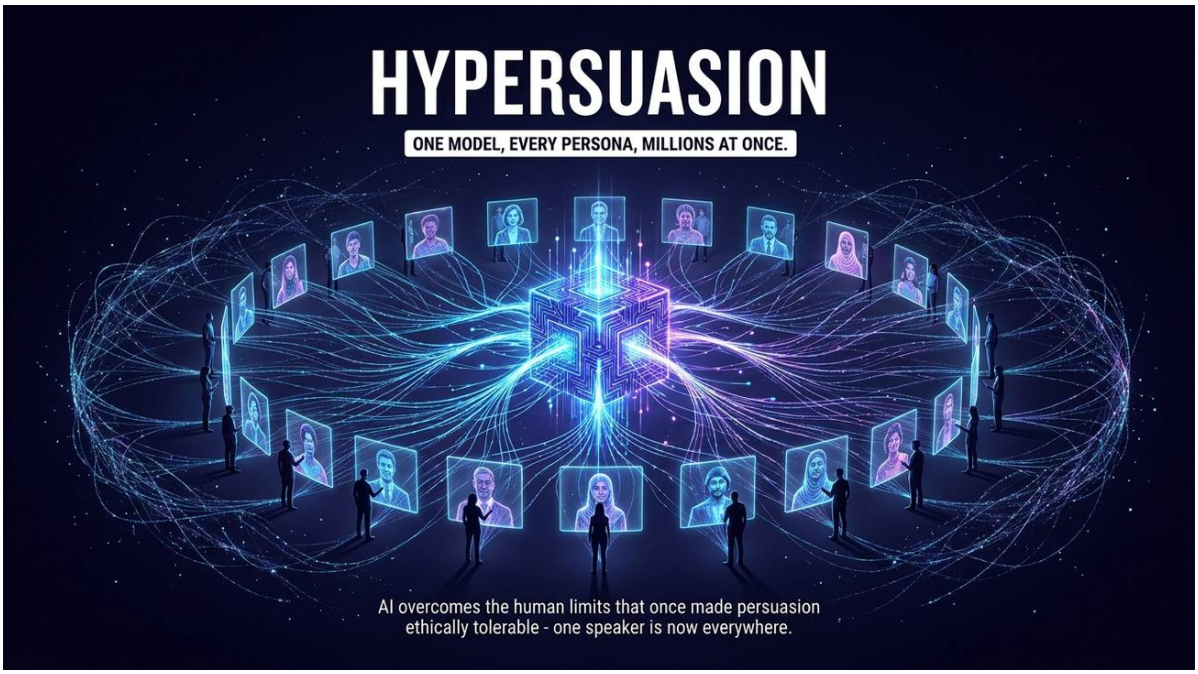


Figure 98. *Hypersuasion: AI overcomes the human limits that once made persuasion ethically tolerable.*

18.3 The paradox applied to every speculation in this paper

This is the heart of the section. Below, every speculation and prediction in §8–§16 is restated as a probe-paradox triple: the **promise (X)** — the benefit as this paper actually wrote it; its **machine-scale shadow (Y)** — the harm the *same capability* produces when pushed to full deployment; and the **unknown line** — where, between X and Y, the ethical becomes unethical. Per the Sorites argument, every "line" entry is an admission, not a coordinate: we name what would have to be measured, not a threshold we possess.

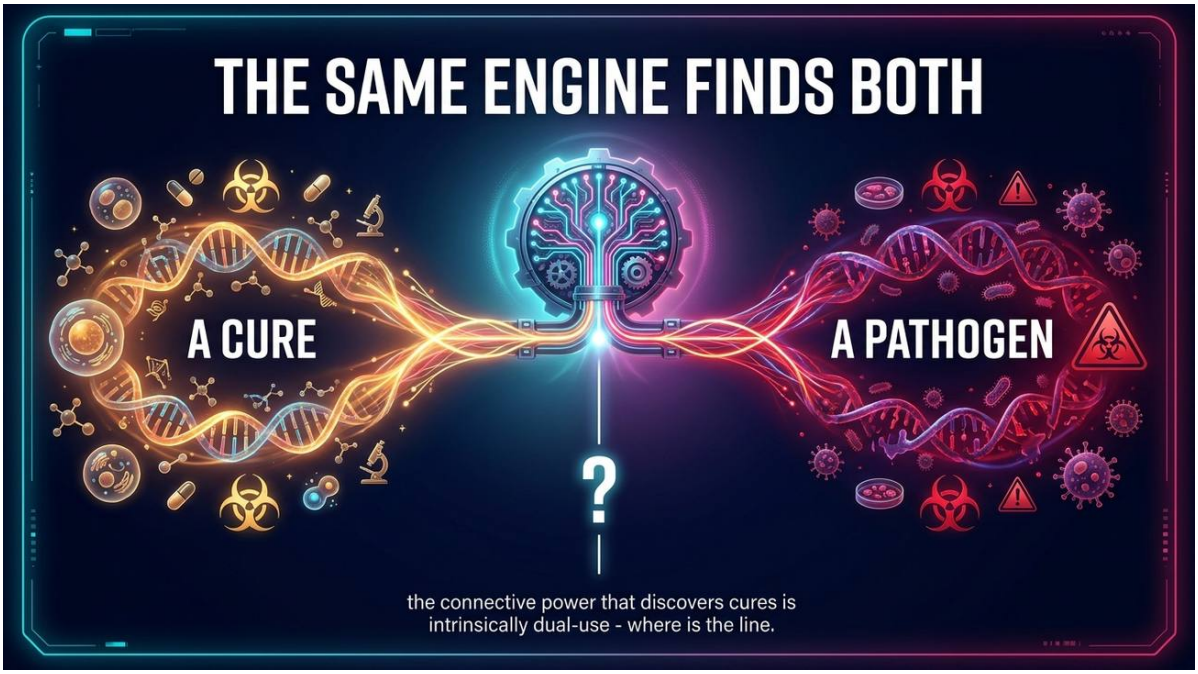


Figure 99. *The same connective engine that finds a cure can find a pathogen - where is the line?*

Two of the twelve cons in §10 deserve to be flagged first, because in them the author has already stated the paradox in his own words, before naming it. **CON 9 (the Collingridge dilemma)** is the probe paradox in temporal form: a grounded, always-on, continually-learning association engine has impacts we *cannot foresee* now (the information problem) and that, *once entrenched*, are too costly to control (the power problem) — which is exactly "we cannot know where the line is without approaching it, and by the time we have approached it we may not be able to step back." **CON 11 (dual-use and capability overhang)** is the probe paradox in capability form: *the same property that makes Calyx valuable makes it dangerous* — drug discovery is pathogen design read backward; benign inference is surveillance read backward. The paper did not need Royse to find the paradox; it had already written it down twice and called it a risk.

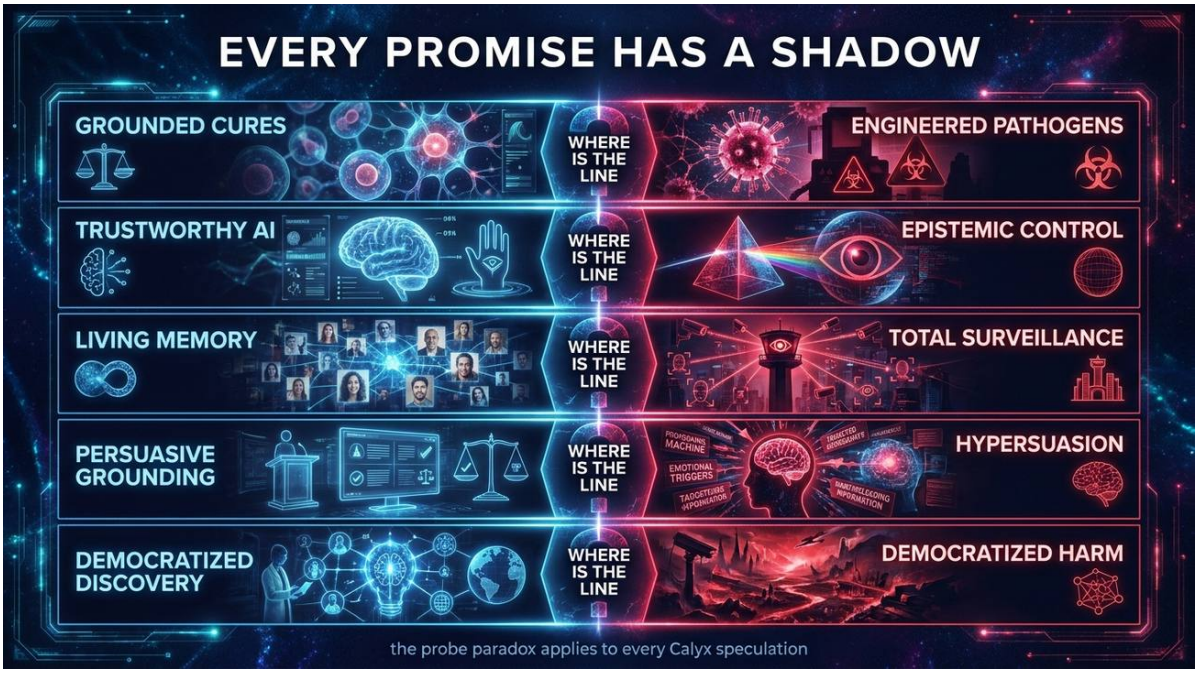


Figure 100. *Every promise has a shadow: the probe paradox applies to every Calyx speculation.*

Speculation (where)	Promise — X (as written)	Machine-scale shadow — Y (full capacity)	Where is the line?
Killer-app thesis (§8.4)	Calyx is to neuromorphic what deep learning was to the GPU — the software that confers world-changing value on stranded silicon.	The same revaluation builds a planetary association capability faster than any norm for governing it can form; "the reason to own it" becomes the reason no one can govern it.	At what diffusion rate does conferring value outrun the ability to oversee the thing whose value was conferred?
Intel full-deployment scenario (§9)	Loihi 2 + Calyx + Intel's backing become neuromorphic's "CUDA moment," fully operational at planetary scale.	A single firm operates the world's grounded-memory substrate at full capacity — the maximal X→Y jump from one librarian to one planetary mind.	The line is somewhere between "a useful product" and "civilizational infrastructure run by one actor"; nobody has located it.
The flywheel / CUDA-style moat (§9.2–9.3)	Increasing returns and network effects compound Calyx's lead into a durable advantage.	The flywheel that compounds value also compounds *unaccountability*: the more entrenched, the less reversible — Collingridge's power problem made economic.	When does a self-reinforcing lead cross from competitive moat into a position no external check can reach?
"Most valuable hardware on the planet" (§9.4)	Loihi 2 becomes the most strategically valuable silicon on Earth.	Maximal concentration of the most powerful associative capability in whoever owns that silicon.	The line between strategic advantage and dangerous concentration is unmarked.

Speculation (where)	Promise — X (as written)	Machine-scale shadow — Y (full capacity)	Where is the line?
PRO 1 — a real killer app exists (§10)	Neuromorphic finally has valuable software.	"Valuable" is value-neutral: the first valuable neuromorphic workload is also the first *governable-only-with-difficulty* one.	When does proving a capability obligate proving its containability first?
PRO 2 / PRO 9 — energy savings & a cost-lowering flywheel (§10)	Order-of-magnitude cheaper association; broadening access over time.	Cheaper + broader = the capability proliferates faster than oversight; cost is the accelerant on every shadow in this table.	At what price-per-association does broad access become broad exposure?
PRO 3 — continual online learning is feasible (§10)	"Learn from new data forever" becomes economical.	An always-on learner that never forgets is also a system that *cannot be made to forget* — the substrate of permanent memory.	When does perpetual learning cross into the loss of the right to be forgotten?
PRO 4 / PRO 5 — a new value axis and a compounding moat (§10)	Sidestep NVIDIA; build a defensible position.	A defensible moat over *grounded memory* is a defensible position over what a society can cheaply know.	The line between a business moat and an epistemic gatekeeper.
PRO 6 — grounded memory addresses the trust crisis (§10)	Traceable, fail-closed answers counter hallucination.	Maximally credible answers are maximally persuasive answers — the hypersuasion sharpening of §18.2.	When does "more trustworthy" become "more able to move people," and who measures which it is?
PRO 7 — cross-domain scientific discovery at scale (§10)	Always-on surfacing of grounded tar get↔molecule↔disease↔literature links.	The identical mechanism surfaces grounded harm↔pathway↔agent links — discovery and weaponization share one engine (the §10 CON 11 point).	The line runs *through* the same query; no filter cleanly separates the two outputs.
PRO 8 — revaluing a stranded national asset (§10)	Sunk research becomes a sovereign capability.	"Sovereign capability" is concentration framed as patriotism — a single-state monopoly on planetary association.	When does national advantage become an asset too powerful for any single nation to hold safely?
PRO 10 — architectural honesty (hybrid exactness) (§10)	The bit-exact spine keeps adoption safe.	Safety-by-architecture can manufacture *misplaced* confidence — the safer it looks, the faster it is adopted past the point of understanding.	When does a credible safety story become an accelerant rather than a brake?
PRO 11 — a second engine of computing growth (§10)	A GPT-class efficiency engine as Moore/Dennard fade.	A general-purpose engine is general-purpose in harm too; growth engines restructure labor and power as much as they restructure cost.	Where does a general-purpose technology's benefit-to-dislocation ratio turn negative?
PRO 12 — alignment with how brains compute (§10)	Matching algorithm to the brain's native structure.	A planetary cognition-shaped substrate invites cognition-shaped misuse — modeling and steering minds at their own grain.	The line between modeling cognition and manipulating it.
CON 1 / CON 2 — underdelivery, the chasm (§10)	(Stated as risks of *failure*.)	The paradox's inverse: under-promising lulls oversight, so the capability is governed as a curiosity until it is suddenly not.	When does "it will probably underdeliver" stop being a reason to defer governance?
CON 3 / CON 4 — lock-in, coordination (§10)	(Stated as risks to adoption.)	If adoption *does* succeed, lock-in and co-investment are exactly what make later course-correction impossible — the moat cuts both ways.	When does the commitment required to succeed foreclose the ability to stop?
CON 5 — the precision ceiling (§10)	(Stated as a technical limit.)	A capability that works "well enough" to persuade but not "exactly enough" to verify is the most dangerous middle of the X→Y range.	Where on the precision curve does usefulness outrun verifiability?

Speculation (where)	Promise — X (as written)	Machine-scale shadow — Y (full capacity)	Where is the line?
CON 6 — Jevons paradox (§10)	(Stated as a sustainability risk.)	Efficiency induces demand: every shadow in this table proliferates *because* the per-use cost fell. Jevons is the paradox's throttle stuck open.	At what point does making the capability cheaper make its aggregate harm larger?
CON 7 — supply-chain fragility (§10)	(Stated as a robustness risk.)	The §12 answer to fragility — a cheap, un-chokeable chip — removes the one natural brake on proliferation (see the chip row).	When does removing a bottleneck remove a safeguard?
CON 8 — winner-take-all monopoly (§10)	(Stated as an antitrust/systemic risk.)	This *is* the concentration shadow of PROs 4/5/8, stated by the author as a con.	The line between a dominant firm and an ungovernable one.
CON 9 — the Collingridge dilemma (§10)	(The author's own statement of the paradox, temporal form.)	Foresight fails early; control fails late — the window to find the line responsibly is narrow and self-closing.	This *is* the line-location problem, conceded.
CON 10 — grounding-as-theater (Goodhart) (§10)	(Stated as a measurement-gaming risk.)	If "grounded" becomes the rewarded metric, the trust guarantee can be hollowed out while the persuasion it licenses remains — the worst case of §18.2.	When does the measure of grounding stop measuring grounding?
CON 11 — dual-use & capability overhang (§10)	(The author's own statement of the paradox, capability form.)	The same engine that finds a cure finds a pathogen; the same that infers benignly infers to surveil.	This *is* the line, running through a single capability — conceded.
CON 12 — the thesis may be wrong (§10)	(Stated as an epistemic humility.)	If wrong, no harm from *this* engine — but the probe was still run; the responsible question survives the thesis's falsity.	Even a failed probe must be reversible; that is the point.
The cheap custom chip (§12)	A Calyx-optimized chip is fundable at Series-A scale on a mature node with **no supply-chain chokepoints** — "own both halves of the covenant."	Removing every chokepoint removes every *gate*: the capability proliferates beyond any single actor's, or any regulator's, control — many planetary association engines, ungated.	When does democratizing the substrate cross from breaking a monopoly into removing the last brake on misuse?
§14 — cross-domain scientific discovery (application class)	Grounded, citable hypotheses for treatments and cures.	Grounded, citable hypotheses for harms (the cure↔pathogen identity again, now as a productized capability).	The line is inside the query, not around the product.
§14 — grounded memory & reasoning plane for agents	A database an agent can trust and trace.	The reasoning plane an agent can be *steered* through at scale — trustworthy infrastructure for untrustworthy ends.	When does a trustworthy substrate become a force-multiplier for whoever directs the agents?
§14 — multimodal general perception	One grounded record across text, image, audio, protein, DNA, time.	Total perception: the same fusion that understands a patient understands a population — universal sensing.	The line between comprehensive perception and total surveillance.
§14 — faithful reconstruction of style, voice, persona	Capturing an author's many facets, with grounding and provenance.	Grounded impersonation: the most convincing synthetic voice/persona ever built, provenance-stamped to look authentic.	When does faithful reconstruction become undetectable impersonation?
§14 — universal structural summarization (the kernel)	A compact grounded explanation of any corpus at any scope.	Whoever controls the kernel controls the *summary of record* — the power to decide what a body of knowledge "is about."	The line between summarizing knowledge and defining it.
§15 #1 / Spec. A — cures hidden in literature; the decade of cures	Drug-repurposing and mechanism hypotheses no single field could see; diseases falling in clusters.	Engineered pathogens by the identical connective machinery, at the identical accelerating rate — cures and bioweapons share one front end.	The cure/pathogen line runs through one engine and one query; this is the table's central case (see EP6/EP7).

Speculation (where)	Promise — X (as written)	Machine-scale shadow — Y (full capacity)	Where is the line?
§15 #2 — materials that did not exist	Better batteries, catalysts, perhaps superconductors.	Novel weapons, energetics, and proliferation-enabling materials by the same cross-term implication.	Where does materials discovery cross into materials for harm?
§15 #3 — a grounded co-pilot for every scientist	A tireless cross-disciplinary partner for every researcher.	A tireless cross-disciplinary partner for every malicious actor — democratized expertise is democratized capability.	The line between empowering inquiry and empowering anyone's inquiry.
§15 #4 — medicine grounded in all we know	Precise, traceable, evidence-anchored diagnoses.	A patient's whole life grounded against all medical knowledge is also the most complete medical dossier ever assembled — privacy's end in the name of care.	When does grounding a patient in everything become surveilling them in everything?
§15 #5 — grounded solutions for a living planet	Evidence-backed climate, materials, and policy interventions.	"Evidence-backed policy" is also evidence-backed *justification* — a grounded engine for legitimizing predetermined ends.	The line between informing policy and manufacturing its warrant.
§15 #6 / Spec. C — trustworthy AI in high-stakes domains; restoration of trust	Law, medicine, governance can finally rely on traceable machine reasoning; the post-truth drift ends.	If one system is the arbiter of "grounded," it holds an **epistemic monopoly** — the power to decide what counts as true is itself the deepest power, and ending post-truth by making one checker authoritative is control wearing trust's clothes.	The line between *making verification cheap* and *becoming the verifier* (developed at §16 / §18.3 coda below).
§15 #7 — a memory that never forgets and always cites	A living, self-correcting global record, O(1)-updated, provenance-stamped.	Total, permanent, searchable memory of everyone and everything — the literal end of the right to be forgotten and the substrate of total surveillance.	When does a memory that never forgets become a memory no person can escape?
§15 #8 / Spec. B — breakthroughs from a single mind; the end of the expert bottleneck	Discovery decentralizes; a lone researcher rivals an institution.	Harm decentralizes in lockstep: the barrier that once required an institution to do great damage falls with the barrier to doing great good.	The line between democratized discovery and democratized harm — the same fallen barrier.
§15 #9 — the pace of discovery accelerates	Each grounded link makes the next easier; the *rate* bends upward.	Governance, which moves at human and institutional speed, is structurally outpaced — the harm-first/norm-second lag (§18.1) widens precisely as stakes rise.	When does acceleration outrun the slowest necessary check?
§15 #10 — a world that runs on grounded truth	A civilization whose knowledge is verifiable by construction.	A civilization whose definition of "grounded" is set by one substrate is a civilization with a single point of epistemic failure — and capture.	The line between a shared epistemic commons and a captured one.
§16 — Epistemic Symmetry coda	When generating answers is free, grounded verification is the scarce resource, and Calyx is "the asymmetry that still matters."	If verification is the scarce resource and Calyx is its arbiter, then Calyx holds the most concentrated power in the post-symmetry economy: **who guards the grounding?**	The line between *providing* verification and *owning* it — the question the coda's own logic raises but does not answer.
§17 — intelligence as the great equalizer	Broadly distributed, grounded intelligence as a structural check that forces concentrated power to govern as though an informed people retain leverage — a living Second Amendment.	The identical diffusion that arms the citizen against tyranny arms the malicious actor against everyone (Cronin's lethal empowerment; Bostrom's vulnerable world); and an equalizer everyone holds is also a brake no one can pull. The §17 thesis *names its own shadow* (EQ8) and hands it to this section.	The line between distributing a *defensive* check on power and distributing *offense-dominant* capability that cannot be recalled once spread (§17.5) — the paradox's sharpest case, because the promise is precisely the broad spread that maximizes the danger.

The table's recurring lesson is that there is almost never a *second* engine for the shadow. The cure and the pathogen, the diagnosis and the dossier, the trustworthy answer and the hypersuasive one, the democratized discovery and the democratized harm — each pair is **one capability read in two directions**, exactly as CON

11 warns. That is why the probe paradox bites so hard here: one cannot build only the left column. The right column ships with it.

18.4 How the paradox should be applied moving forward — Calyx as an instrument for finding the line responsibly

If the paradox were the whole story, the only honest conclusion would be paralysis. But Royse's framework also tells us what a *responsible* approach to an unknown line would look like, and here the analysis turns constructive — because Calyx's own founding principles are, not by coincidence, the exact tools the paradox calls for.

The paradox says the line cannot be located without approaching it, and that approaching it blindly risks an irreversible crossing. The mitigation Collingridge himself prescribed for exactly this dilemma is **intelligent trial and error: keep decisions reversible, monitor for harm, and correct before entrenchment.** Now read Calyx's architecture against that prescription:

- It is **grounded** (§2.5): it reports an *honest ceiling*, labels what is trusted versus provisional, and refuses to claim more than its evidence supports — so a probe's outputs come with a measured confidence rather than a confident assertion.
- It is **fail-closed**: an unknown lens, a shape mismatch, an uncalibrated guard, or missing grounding returns a structured error, never a silent wrong answer — so a probe halts at the edge of its competence instead of guessing past it.
- It is **provenance-traced**: every measurement, guard verdict, answer, and self-optimization step is one canonical entry in a tamper-evident, hash-chained ledger — so every step of a probe is *measured and attributable*, and can be audited after the fact.
- It is **reversible** (Anneal, §3.11): every change is shadow-tested against held-out replay, gated on safety tripwires and per-metric non-regression, and either promoted by a single pointer swap or rolled back — every promotion *and every revert* recorded.

Taken together, these four properties describe something the paradox says is otherwise impossible: a way to **approach the ethical boundary instrumented, measured, and reversible.** Rather than crossing the line and discovering the harm afterward — Hammurabi's collapsed houses, the pattern §18.1 calls invariant — a fail-closed, provenance-traced, reversible system can *quantify harm at each increasing scale, with full traceability, and fail closed before a catastrophic crossing.* It can advance from X toward Y one measured, revertible increment at a time, watching the tripwires, and stop — or roll back — the moment a metric moves the wrong way. This is precisely Collingridge's prescribed mitigation, operationalized in silicon and software.

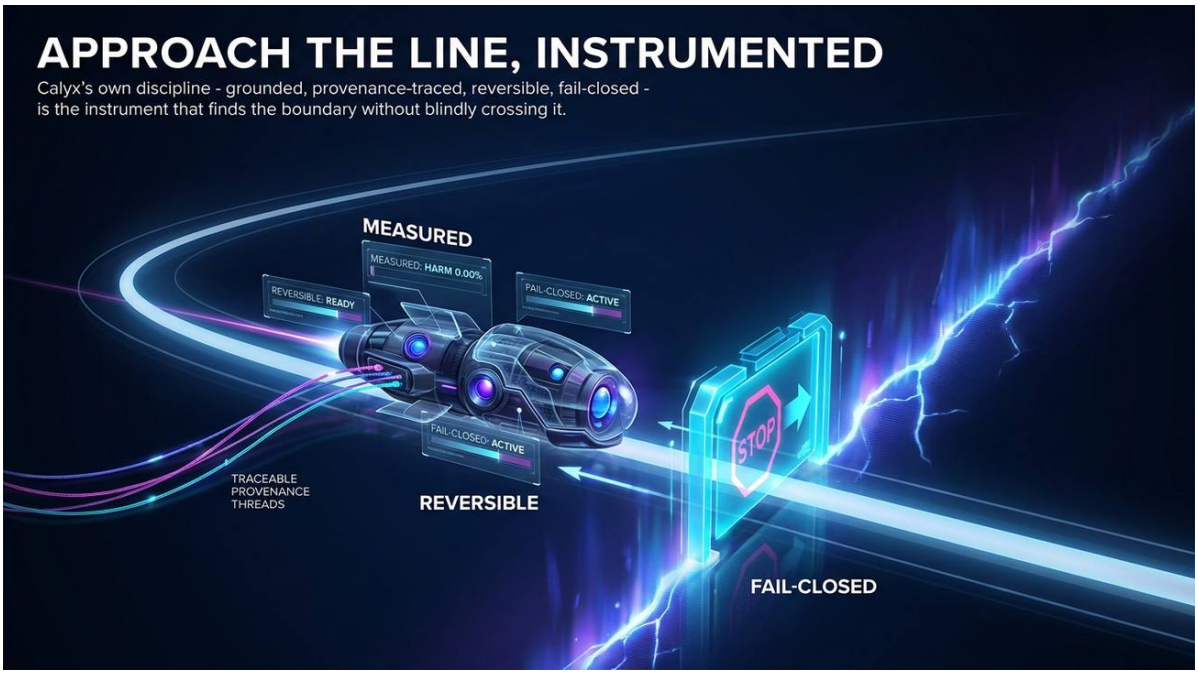


Figure 101. *Approach the line instrumented: Calyx's grounded, provenance-traced, reversible, fail-closed discipline finds the boundary without blindly crossing it.*

So the strongest form of the claim is this: Calyx is not merely *subject* to the probe paradox — its architecture is a candidate *answer* to it. The same discipline that this paper has argued makes Calyx trustworthy as a database is, read through Royse, the discipline that could make the probe itself ethical: a way to find the line by approaching it *with instruments*, rather than by blindly crossing it and reading the line off the wreckage. We state this as a disciplined proposal, not a guarantee. An instrument can be misused, a tripwire can be miscalibrated, a ledger can be ignored by those who own it (CON 10). The architecture *enables* a responsible probe; it does not *compel* one. But the paradox asked whether an instrumented, reversible, measured approach to the line is even conceivable — and Calyx's four principles are an existence argument that it is.

18.5 The honest position

We do not know where Calyx's line is. This section has not claimed to, and — consistent with the grounding principle it analyzes — it will not pretend to. It claims only three things. First, that the probe paradox is **real**: AI dissolves the human limits that made whole categories of action tolerable, the resulting ethical boundary is genuinely unknown, Sorites-vague, and historically discovered only by crossing it. Second, that the paradox applies to **every** promise in this paper — every speculation in §8–§16 has a machine-scale shadow that is the same capability read in the other direction, and §10's CON 9 and CON 11 are the author conceding exactly this before naming it. Third, that the responsible path is therefore to treat **each deployment as an instrumented probe** — measured, traceable, reversible, fail-closed, and governed — rather than a blind crossing, and that Calyx's founding discipline is, not coincidentally, the instrument such a probe requires.

Royse's question was: *must we be unethical in order to be ethical?* The honest answer this paper can give is that we may have no choice but to approach the line — the paradox says the line cannot be found any other way — but that *how* we approach it is ours to choose. Calyx's grounded, fail-closed, provenance-traced, reversible discipline is an attempt to make the probe ethical **by construction**: to approach the boundary with instruments rather than with blindness, to measure the harm at each step rather than to suffer it all at once, and

to be able to step back. That is not a guarantee that we will find the line in time. It is a commitment to look for it in the only way that could be called responsible.

19. Conclusion

Calyx is a wager that the right unit of meaning is not one vector but a **constellation** of perspectives, and that the most valuable knowledge lives in the **associations between** those perspectives — provided every claim is **grounded, un-flattened, and fail-closed.** Its architecture realizes the calculus of association end to end: measure through diverse frozen lenses, count the cross-terms, differentiate the bits that matter, and compose grounded, provenance-backed answers, all while continuously and reversibly optimizing toward the growth of grounded understanding.

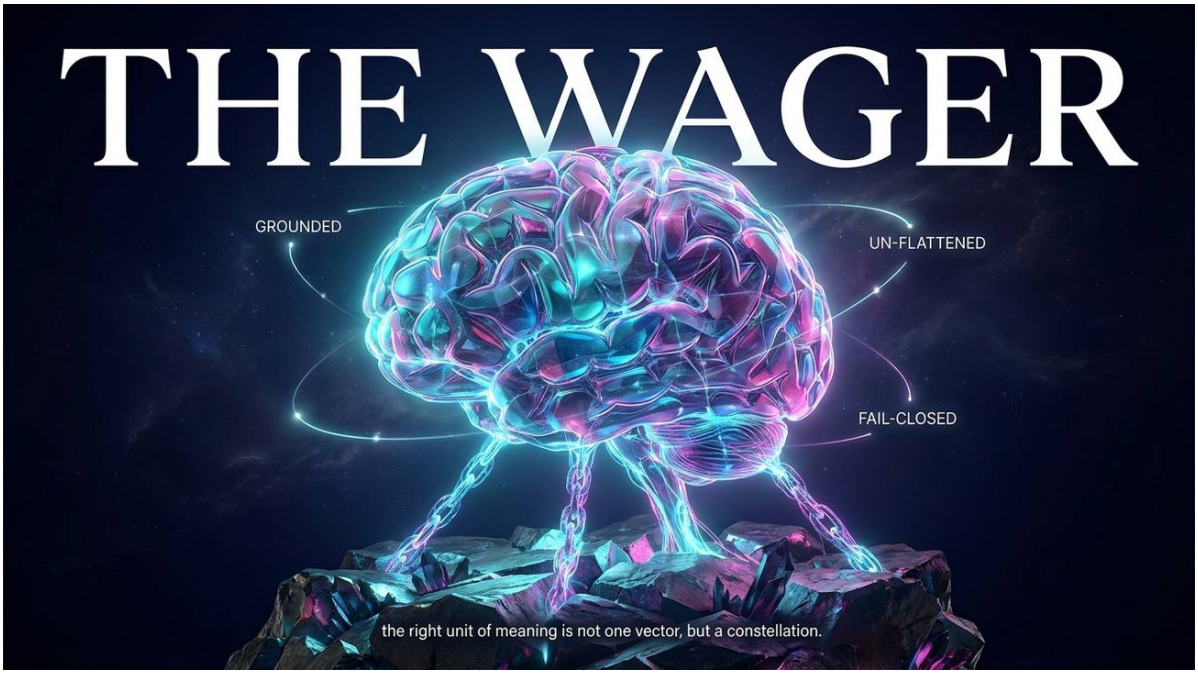


Figure 102. *The wager: the right unit of meaning is not one vector, but a grounded constellation of perspectives.*

The richness that makes Calyx powerful also makes it computationally demanding at scale — and that demand points to a clear answer. The calculus of association is, operation for operation, a **neuromorphic-native** workload: similarity, winner-take-all, association-graph traversal, attractor recall, and event-driven streaming. By offloading exactly that precision-tolerant subset onto spiking and in-memory substrates — energy that scales with events, latency that scales with depth, insertion that is $O(1)$ — while preserving a bit-exact digital spine for provenance and exactness, Calyx charts a credible path from a single workstation to a grounded intelligence engine over massive, multi-domain, multimodal, always-on data. The vision and the substrate are aligned; the work is to build toward them, stage by stage, without ever compromising the principles that make Calyx trustworthy.

Appendix A — Glossary

- **Constellation** — the atomic record: one input measured through one panel of frozen lenses.
- **Lens** — a frozen, content-addressed embedder (a perspective).

- **Slot / Panel** — a lens position in a panel; a versioned set of slots.
- **Cross-term** — a derived relationship between two slots (agreement, delta, interaction, concat).
- **DDA (Derived Data Abundance)** — the $N + \binom{N}{2} + 1$ field of signals per record.
- **Anchor** — a grounded, observed real-world outcome attached to a constellation.
- **Bits / Differentiation contract** — information (in bits) a lens adds about an outcome; the rule that a lens must add unique signal and not be redundant.
- **Panel sufficiency** — whether a panel carries at least as much information about an outcome as the outcome's entropy.
- **Kernel** — the minimal record set (≈1%) that explains a corpus; an index and an answer path.
- **Guard (Gτ)** — the per-slot, conformally calibrated, fail-closed gate on generation.
- **Ledger** — the append-only, hash-chained provenance record.
- **Anneal** — reversible, shadow-tested self-optimization under safety tripwires.
- **Oracle** — grounded consequence prediction with an honesty gate.
- **Association fabric** — the neuromorphic/in-memory accelerator for similarity, ranking, and traversal.
- **Digital spine** — the bit-exact subsystems (storage, provenance, exact analytics) that never move off conventional hardware.

Appendix B — On the neuromorphic substrate (selected technical basis)

The neuromorphic mapping in §7 rests on well-established results in spiking and in-memory computing: latency (time-to-first-spike) coding computes weighted sums ($\Sigma_i w_i x_i$) by temporal integration; neuromorphic nearest-neighbor search computes cosine top-k as a temporal-coded matrix-vector product with $O(1)$ online insertion; spiking winner-take-all selects top-k as first-k spikes; spike-wavefront propagation computes shortest paths with substantial speedups over conventional graph traversal; in-memory crossbars compute matrix-vector multiplies in a single physical step; and attractor / content-addressable memories perform robust associative recall. The integration principle throughout is **precision-tolerant offload with digital verification and fallback** — the neuromorphic fabric accelerates the associative, approximate operations, and the digital spine guarantees the exact, load-bearing ones. A detailed treatment of the substrate options, their maturity, and the integration plan accompanies this white paper in the companion implementation documents.

Appendix C — Theory Base (sources for the speculation in §8–§11)

Every speculative claim in §8–§11 is anchored to one or more of the theories below. Each is an established result treated as true to the best of current human knowledge; the speculation is the conditional *application* of these theories to the Calyx-on-Loihi-2 scenario, not a claim beyond them.

Hardware–software co-evolution & platform economics

- *Killer app / complementary goods (VisiCalc, Wintel):* Bricklin & Frankston (1979). https://en.wikipedia.org/wiki/VisiCalc
- *Two-sided markets:* Rochet & Tirole (2003), *J. European Economic Assoc.* 1(4). https://doi.org/10.1162/154247603322493212

- *Network effects / Metcalfe's Law (+ Odlyzko–Tilly n·log n critique):* Metcalfe (2013), *Computer* 46(12). https://doi.org/10.1109/MC.2013.374
- *Increasing returns & path dependence / lock-in:* Arthur (1989), *Economic Journal* 99(394) https://doi.org/10.2307/2234208 ; David (1985), "Clio and the Economics of QWERTY," *AER* 75(2). https://www.jstor.org/stable/1805621
- *The GPU/CUDA/AlexNet/NVIDIA record:* CUDA unveiled Nov 8 2006; AlexNet (Krizhevsky, Sutskever & Hinton, NeurIPS 2012, two GTX 580s) https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf ; NVIDIA >$1T Jun 2023 → >$2T Feb 2024 → >$3T Jun 2024 → first to $4T Jul 9 2025 https://www.cnbc.com/2025/07/09/nvidia-4-trillion.html

Innovation & technology-revolution theory

- *Creative destruction:* Schumpeter (1942), *Capitalism, Socialism and Democracy*. https://en.wikipedia.org/wiki/Creative_destruction
- *Disruptive innovation:* Bower & Christensen (1995), *HBR*. https://hbr.org/1995/01/disruptive-technologies-catching-the-wave
- *General Purpose Technology ("engines of growth"):* Bresnahan & Trajtenberg (1995), *J. Econometrics* 65(1). https://doi.org/10.1016/0304-4076(94)01598-T
- *Techno-economic paradigms / installation vs deployment / turning point:* Perez (2002), *Technological Revolutions and Financial Capital*. https://carlotaperez.org
- *Diffusion of innovations (S-curve) + the chasm:* Rogers (1962) https://en.wikipedia.org/wiki/Diffusion_of_innovations ; Moore (1991), *Crossing the Chasm*.
- *Amara's Law & Gartner Hype Cycle:* Amara (IFTF) https://en.wikipedia.org/wiki/Roy_Amara ; Fenn/Gartner (1995).
- *Kondratiev long waves; Kuhn paradigm shifts:* Kondratiev (1925) https://en.wikipedia.org/wiki/Kondratiev_wave ; Kuhn (1962), *The Structure of Scientific Revolutions*.
- *Wright's Law / experience curves:* Wright (1936) https://pdodds.w3.uvm.edu/research/papers/others/1936/wright1936a.pdf ; Nagy, Farmer, Bui & Trancik (2013), *PLoS ONE*. https://doi.org/10.1371/journal.pone.0052669

Computing physics & limits

- *Moore's Law:* Moore (1965). https://en.wikipedia.org/wiki/Moore%27s_law
- *Dennard scaling & its end (~2005):* Dennard et al. (1974). https://doi.org/10.1109/JSSC.1974.1050511
- *Koomey's Law (compute energy efficiency):* Koomey et al. (2011), *IEEE Annals*. https://doi.org/10.1109/MAHC.2010.28
- *Landauer's principle ($kT\ln 2 \approx 2.9 \times 10^{-21}$ J/bit @ 300 K):* Landauer (1961), *IBM J. R&D* 5(3). https://doi.org/10.1147/rd.53.0183
- *Von Neumann bottleneck:* Backus (1978, Turing lecture), *CACM* 21(8). https://dl.acm.org/doi/10.1145/359576.359579
- *Neuromorphic engineering:* Mead (1990), *Proc. IEEE* 78(10). https://doi.org/10.1109/5.58356
- *AI/data-centre energy:* IEA *Energy and AI* (2025) https://www.iea.org/reports/energy-and-ai/energy-demand-from-ai ; Patterson et al. (2021), arXiv:2104.10350. https://arxiv.org/abs/2104.10350
- *Bitter Lesson & scaling laws:* Sutton (2019) http://www.incompleteideas.net/IncIdeas/BitterLesson.html ; Kaplan et al. (2020) https://arxiv.org/abs/2001.08361 ; Hoffmann et al. (Chinchilla, 2022)

https://arxiv.org/abs/2203.15556

Risk / downside theory

- *Collingridge dilemma:* Collingridge (1980), *The Social Control of Technology*. https://en.wikipedia.org/wiki/Collingridge_dilemma
- *Goodhart's Law:* Goodhart (1975); Strathern (1997). https://en.wikipedia.org/wiki/Goodhart%27s_law
- *Jevons paradox (rebound):* Jevons (1865), *The Coal Question*. https://en.wikipedia.org/wiki/Jevons_paradox
- *Diminishing returns on complexity / collapse:* Tainter (1988), *The Collapse of Complex Societies*.
- *Winner-take-all markets:* Frank & Cook (1995). https://en.wikipedia.org/wiki/The_Winner-Take-All_Society
- *Moravec's paradox:* Moravec (1988), *Mind Children*. https://en.wikipedia.org/wiki/Moravec%27s_paradox
- *Supply-chain concentration (ASML/TSMC):* https://en.wikipedia.org/wiki/ASML_Holding · https://en.wikipedia.org/wiki/TSMC

Information & cognition theory (Calyx-relevant)

- *Information theory & data-processing inequality:* Shannon (1948), *BSTJ* 27. https://onlinelibrary.wiley.com/doi/10.1002/j.1538-7305.1948.tb01338.x
- *Hebbian learning; predictive coding:* Hebb (1949) https://en.wikipedia.org/wiki/Hebbian_theory ; Rao & Ballard (1999), *Nat. Neurosci.* https://www.nature.com/articles/nn0199_79
- *Free-energy principle:* Friston (2010), *Nat. Rev. Neurosci.* 11. https://doi.org/10.1038/nrn2787
- *Attractor / associative memory; sparse coding:* Hopfield (1982), *PNAS* https://www.pnas.org/doi/10.1073/pnas.79.8.2554 ; Olshausen & Field (1996), *Nature* https://www.nature.com/articles/381607a0
- *Sparse distributed memory / hyperdimensional computing:* Kanerva (1988; 2009). https://doi.org/10.1007/s12559-009-9009-8
- *No Free Lunch; Church–Turing:* Wolpert & Macready (1997), *IEEE TEC* 1(1) https://doi.org/10.1109/4235.585893 ; Church (1936)/Turing (1936) https://plato.stanford.edu/entries/church-turing/

Appendix D — Intel Loihi 2 hardware basis (the substrate behind §7–§11)

The concrete Loihi 2 / Hala Point figures used in §7–§11 are documented, with full primary sourcing, in the companion **intelinfo/** reference series accompanying this paper:

- intelinfo/02_loihi2_chip_architecture.md — cores, mesh, microcode ISA, graded spikes, learning.
- intelinfo/03_loihi2_fabrication_intel4_euv.md — Intel 4, EUV, die/transistors.
- intelinfo/06_hala_point_largest_system.md — the 1,152-chip / 1.15-billion-neuron system.
- intelinfo/07_pohoiki_springs_100M_neurons.md — the demonstrated $O(1)$ cosine k-NN precedent.
- intelinfo/09_applications_and_benchmarks.md, …/10_energy_and_performance.md — measured results.
- intelinfo/12_calyx_relevance_and_roadmap.md — the operation-by-operation Calyx↔Loihi-2 mapping.
- intelinfo/14–19 — what it would take to manufacture/replicate Loihi 2 (why the substrate is rent-only).

- intelinfo/13_sources.md — the master provenance ledger of every hardware figure. Primary anchors include Intel's Loihi 2 Technology Brief, the Hala Point newsroom release (2024), and Frady et al. (2020), *Neuromorphic Nearest Neighbor Search Using Intel's Pohoiki Springs*, arXiv:2004.12691.

Calyx is pre-1.0 and under active development. This white paper describes the system's architecture and its intended scaling path; specific interfaces and on-disk formats may evolve as the system matures toward 1.0. Sections 8–11 are explicitly speculative and theory-anchored (see Appendix C); they describe a conditional scenario, not a roadmap commitment.