

LatticeRunner

Memory infrastructure for AI products that need real recall.

THE PROBLEM

Every AI memory architecture trades against your **inference budget**.

LLM-MEDIATED MEMORY

MEM0 · LETTA · ZEP ·
GRAPHITI

Mem0 reports **7–8 s per memory recall**; Zep ~4 s. Every operation costs that latency *and* the LLM tokens consumed by extraction, summarization, and routing — on top of any tokens the consuming agent injects downstream.¹

STANDARD VECTOR-DB RAG

PINECONE · QDRANT · MILVUS ·
PGVECTOR

Avoids the memory-layer LLM cost, but production midsize implementations still inject **8,000–12,000 tokens** of retrieved context per query. At 100K queries/month: **\$3K–\$5K** on Standard tier; **\$30K–\$180K** on Premium.²

LONG CONTEXT WINDOWS

1M-TOKEN MODELS

Inflate inference cost, lose information past ~16K tokens (lost-in-the-middle³), and **vanish at session end**.

All three share a structural problem: memory is a database lookup paid for at inference time. **There is no architectural escape within the database paradigm.**

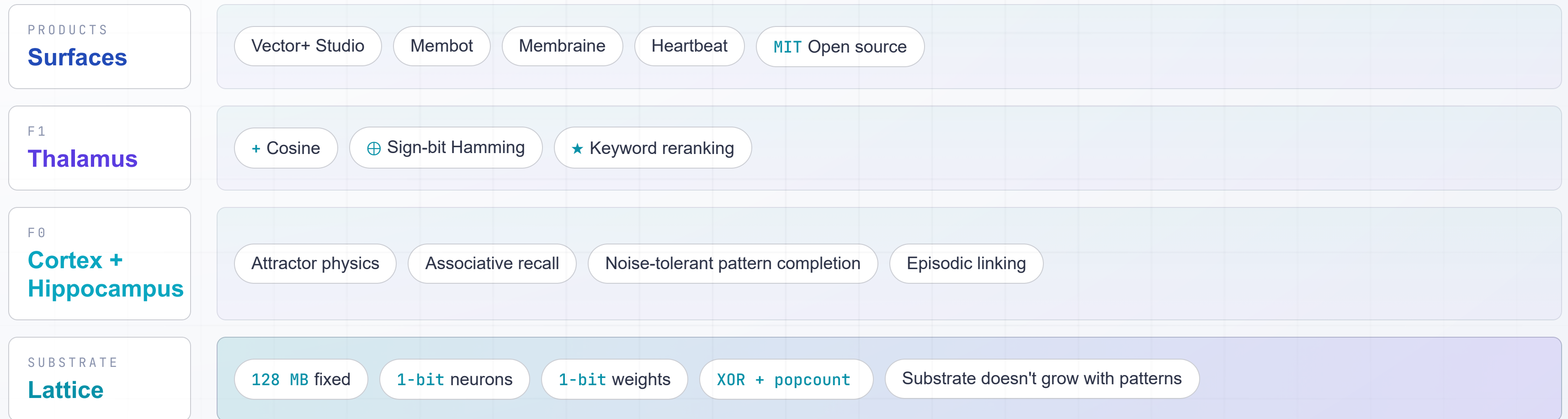
¹ Vectorize.io 2026, *Mem0 vs Letta (MemGPT)*; DEV Community 2026, *5 AI Agent Memory Systems Compared*.

² Redis 2026, *LLM Token Optimization*; Towards AI 2026, *LLM API Token Caching*; current Anthropic + OpenAI public per-million-token pricing.

³ Hugging Face 2025, *Advanced RAG cookbook*.

THE SOLUTION

A physics-based memory substrate. Patterns are stored as **attractors** and recalled by **settling** — not searched in an index.



- **Hebbian learning during imprint.** No backpropagation, no gradient descent — continuous accumulation without catastrophic forgetting.
- **Open source MIT** (Membot, Membrane). Substrate is free; products built on it monetize.
- **Hippocampus episodic linking** woven into the same substrate — associative *and* sequential recall.
- **Encoder-agnostic.** Anything you can build an encoder for — text, image, audio, video — drops onto the same lattice.

THE NUMBERS

Validated. Two live demos.

LIVE · MEMBOT

Semantic search · 2.4M arxiv passages

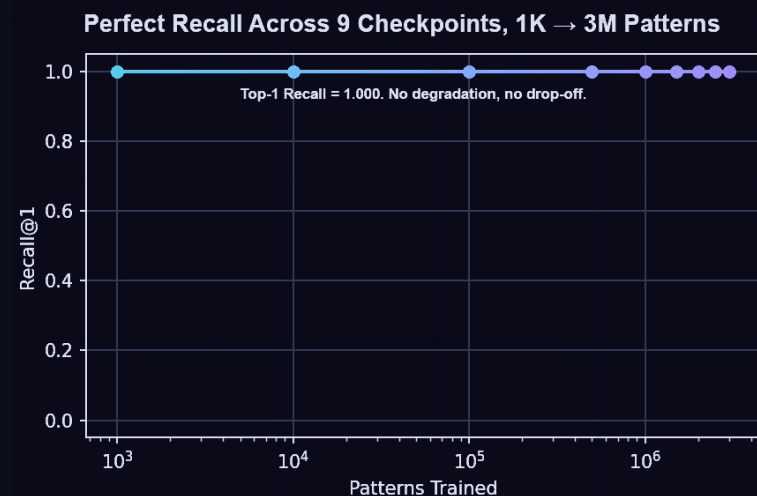
project-you.app/membot/app



LIVE · VECTOR+ STUDIO

Drag-drop cart builder · RAG+ provenance

project-you.app/vps/app



ONE CONTINUOUS RUN UP TO 3 MILLION PATTERNS.
AND WE STILL HAVEN'T FOUND THE CEILING...

No degradation visible at the tested upper bound.

METRIC	RESULT
Patterns stored, single lattice	3,000,000 · 9-checkpoint validated · no upper bound found
Recall fidelity (R@1)	1.000 at every checkpoint, 1K → 3M
Tokens per memory op	0 LLM tokens · sub-200ms retrieval
Search speed at 3M patterns	2.1ms (BALANCED) · constant with scale
LongMemEval R@5	95.0% · no LLM in retrieval pipeline
ARC-AGI-3	84.9% · 23/25 games · ~\$250 total cost
Sign-bit arithmetic empirical floor	24,926× signal vs random · 370K vocab ^a
Histopathology pilot (multimodal)	100% tumor sensitivity · 5,915 PathMNIST patches · 4 vision encoders
BEIR retrieval	Sign-zero blend matches cosine within 1pp at 32× compression
Production deployment	\$28/mo droplet · multi-agent serving

MARKET

Vector retrieval is our wedge. The substrate's reach is broader.

VECTOR DATABASE MARKET • 2025 → 2030 • CAGR 27.5%¹

Standard sign-zero retrieval at 32× compression matches Pinecone, Weaviate, Qdrant, Milvus, Zilliz — with **no reindex-on-write** and **zero LLM tokens** in the memory operation.

ADJACENT • CLOUD DBAAS

\$57.5B by 2028

CAGR 22%² • Vector DB sits inside this. Slice of a fast-growing slice.

ADJACENT • GRAPH DB

\$2.14B by 2030

FK-as-edges + episodic hippocampus map onto the same substrate.¹

The substrate is encoder-agnostic. What we ship is shaped like a vector DB; what we built is a binary-pattern memory layer that takes anything you can build an encoder for. Each new encoder unlocks an adjacent market.

TEXT

Production

Nomic embeddings • live at project-you.app

IMAGES

Pilot

4 vision encoders • 100% tumor sensitivity

AUDIO

Demo

Mel spectrograms • paired-pattern recall

VIDEO

Demo

Cross-modal cue → modality recall

¹ MarketsandMarkets, *Vector Database Market – Global Forecast to 2030* (12-25); *Graph Database Market* (01-25).

² MarketsandMarkets, *Cloud Database & DBaaS Market*.

SAVINGS

No LLM in the memory layer. Watch what that does to your bill **at scale.**

Mem0/Letta/Zep-style memory systems pay LLM tokens **twice per query**: once for the memory operation itself (extraction, summarization, routing), and again when retrieved content is injected into the consuming model's prompt. **We eliminate the first cost entirely.**

USAGE VOLUME	EFFICIENT (HAIKU / MINI)	STANDARD (SONNET / GPT-40)	PREMIUM (OPUS / PRO)
1 query	\$0.002 - \$0.05	\$0.03 - \$0.22	\$0.30 - \$1.80
1,000 queries	\$2 - \$50	\$30 - \$220	\$300 - \$1,800
100,000 queries / mo	\$200 - \$5,000	\$3,000 - \$22,000	\$30,000 - \$180,000

Savings range applies most directly to teams using LLM-mediated memory (Mem0, Letta). For teams using standard vector-DB RAG, savings come from higher-precision retrieval allowing fewer chunks per query — typically 3–5× reduction at the lower end. Cost math compounds either way; magnitude depends on what you're switching from.

ONE SUBSTRATE, THREE SURFACES

Same substrate, three monetization surfaces.

PRIMARY

DEVELOPER / DBMS SURFACE

Vector+ Studio

Hybrid DBMS · cart-as-primary-object

- Free hybrid DBMS — semantic-first, SQL as power tool
- Cart Marketplace as agent-commerce primitive · ownership-key auth doubles as DRM
- Drag-drop document ingestion → branded cartridge → portable artifact
- Tier model: free hosted demo · per-cart · hosted enterprise

live: project-you.app/vps/app

ENTERPRISE INFRASTRUCTURE

Membot

Self-hostable substrate

- Single-user · federated · multi-user (Membox) modes
- The first neuromorphic DBMS, to our knowledge
- Customer-support KB · internal company database · agentic memory layer
- Tiered pricing per deployment

live: project-you.app/membot/app

CONSUMER

Heartbeat

Chrome extension · 7 LLM platforms

- Captures chats from ChatGPT, Claude, Gemini, Mistral, Grok, Copilot, Heart
- Cross-device sync via Supabase
- "Your LLM chats stop being goldfish-brained"
- Free tier + paid sync tiers

live:

VECTOR+ STUDIO free hosted demo / **per-cart** / hosted enterprise

MEMBOT self-hosted free / **hosted enterprise**

HEARTBEAT free tier / **~\$5/mo sync**

Three surfaces, three revenue motions: **marketplace take-rate** · **enterprise-hosted tiers** · **consumer freemium**. One substrate, one cart format underneath.

HARDWARE PATH

BURIED LEDE - SEE SLIDE 09

Binary operations (1-bit neurons, 1-bit weights, XOR + popcount) **map directly to FPGA** → **ASIC silicon**. **Hardware partnerships in discussion**. A memory chip, not a software library.

FPGA → ASIC
ROYALTY + OEM
LICENSING

Open source substrate is the **distribution**. Monetization layers sit on top.

01 Running today

Open source flywheel

LIVE NOW

- MIT-licensed substrate at github.com/project-you-apps (Membot, Membrane, Looper)
- Two live public demos: Membot (2.4M arxiv) + Vector+ Studio
- Sign-bit-arithmetic + pathology results posted to LinkedIn

COMPARABLE

[Hugging Face open-model hub flywheel](#)

02 Marketplace

Cart Marketplace

V1 → V1.2 · Q3 2026

- Sellers package carts: USPTO patents, SEC EDGAR filings, public health datasets, vertical KBs
- Buyers download with ownership-key auth (DRM via authentication, not encryption)
- Take rate on transactions; recurring revenue from cart-as-product
- Network effects: more carts → more buyers → more cart-makers

COMPARABLE

[Hugging Face model hub + Zapier integration marketplace](#)

03 Enterprise

Hosted enterprise

V1.2 · Q4 2026

VALIDATED DEMAND · 04-24-26

- Self-hosted Membot is free (MIT). Hosted managed Membot is paid.
- Tiered: per-cart-month · per-query · per-seat
- **First external pitch:** consulting shop with mid-market book asked to deploy *both internally AND for a specific client engagement* before the consumer demo finished. Technical deep-dive scheduled.

COMPARABLE

[GitLab – free self-hosted + paid SaaS](#)

04 Consumer · parallel

Heartbeat consumer

Q4 2026

- Browser extension free tier — capture from 7 LLM platforms, local search
- Paid tier: cross-device sync via Supabase, semantic search across every LLM chat ever had
- Freemium · ~\$5/mo for sync

COMPARABLE

[Notion / Obsidian Sync](#)

V1 today (80% shipped). V1.2 next quarter. V2 by year-end. Silicon as the long horizon.

V1 Q3-2026

80% complete · live at project-you.app/vps/app

- **Vector+ Studio**
Hybrid DBMS surface: drag-drop cart builder, RAG+ provenance, sandboxed upload-your-own-cart, three-layer cart-format RWX
- **Cart Builder V2**
Browser-side WebGPU build pipeline. Substrate smoke-tested 05-06-26 · server-parity at 6th-decimal drift. Your data never leaves your machine.
- **SQL editor v1**
Toggle alongside semantic search · ~10 commands wired to existing Membot API
- **OAuth**
Supabase Auth + Google/GitHub/Apple · platform-wide identity
- **Cart Marketplace beta**
Sellers list · buyers download with ownership-key auth

V1.2 Q4-2026

Cognition + first enterprise customer

- **Cognition Engine substrate**
Free-wake, daydream, predictive coding, concept-cluster Hebbian (CCs filed)
- **First paying enterprise customer**
Pipeline from 04-24-26 consulting team · deep-dive scheduled
- **Pathology consortium expansion**
Phase 1: 100% sensitivity in hand · Phase 2 adds NCT-CRC + PANDA stratified, cross-tissue
- **Lattice-LM experiments**
Bigram/trigram prediction floor · Spec 06 pre-registered

V2

Marketplace at scale + silicon prototype

- **Cart Marketplace at scale**
Payment rails · take-rate revenue · agent-commerce primitive
- **SQL ingestion**
.sql files as cart source. Rows-as-passages, schema-as-manifest, FK-as-hippo-edges
- **Cloud storage backend**
Cloudflare R2 default · pluggable CartStorage interface (S3, GCS, user-hosted opt-in)
- **Silicon prototype**
FPGA build begins (partner TBD)

HARDWARE FOOTER • THE BURIED LEDE

Our binary substrate (1-bit neurons, 1-bit weights, XOR + popcount) **maps directly to FPGA → ASIC silicon**. Target: **sub-100mW per ASIC** — below the mainstream 1–3W edge segment. **Memory chip thesis**: not a software library; a piece of silicon that lives on every edge device. FPGA partnership begins V2.

\$25.65B → \$165B

EDGE AI · 2025 → 2035 · CAGR 20.5%
Precedence Research³

< 100 mW

PER-ASIC TARGET DRAW
vs. 1-3W mainstream edge

³ Precedence Research, *Edge AI Market Size to Attain USD 165.05 Billion by 2035* (current).

THE ASK

\$25–50K bridges us to the **pre-seed round**.

FOUNDER · THIS IS A FOUNDER-BET AT THIS STAGE

40 years of personal iteration on this architecture. Original concept hit paper in 1985. Production runs today on a \$28/month droplet, serving multi-agent traffic with 2.4M passages at sub-200ms.

Solo technical founder + Claude as co-developer. Early instance of an LLM listed intentionally as co-author on an academic submission (ARC-SAGE: Palatov, Grossberg, et al., *in preparation*). Claude Code uses LatticeRunner's substrate as its own memory layer — eating our own dogfood at the daily-development level.

WHAT \$50K SPECIFICALLY BUYS · 6-MONTH BRIDGE

ALLOCATION	AMOUNT	DELIVERABLE
Andy runway	\$25K	6 months focused build time at bootstrapped pace
Hardware engineer (contractor)	\$10K	2 months: GPU optimizations on the lattice runtime + breadboard XOR rough demo (64×64 L2 or 256×256 L3 scale). Physical silicon-shaped artifact in hand — turns "interesting concept" into "show me the silicon, here's the silicon."
Hardware partnership talks	\$5K	Travel + prototype hardware deposits
Hosting + marketing	\$5K	V1 launch (Reddit / HN / Product Hunt) + deeper droplet capacity
Legal	\$5K	Contracts, provisional patents, etc.

END-OF-BRIDGE MILESTONES · THE ANGEL'S "I BRIDGED THEM TO THE REAL ROUND" MOMENT

M1

V1 shipped
with Cart Marketplace beta live

M2

First paying enterprise customer landed
(pipeline from 04-24-26)

M3

Cognition Engine V1.2 prototypes running (CCs filed; OAuth + cron foundation)

M4

Heartbeat early growth — ~1K users

M5

Breadboard XOR demo of the lattice in hand — physical-substrate artifact ready for hardware-partner conversation

TRY IT RIGHT NOW

project-you.app/membot/app

project-you.app/vps/app

Pre-seed round bridging to: \$500K–\$1M · 12–18 mo runway → V2 + Series A
 Andy Grossberg · andy.grossberg@gmail.com · wavingcat.dev · github.com/project-you-apps · Discord: andygrossberg