Winner: Mem0 Self-Hosted

```
## §6 — Recommended Winner + Rationale (D4 / AC#4)

**Pre-condition gate:** /tmp/forged-99124-evidence.jsonl contains 42 records (≥40 required).
Verified: `wc -l /tmp/forged-99124-evidence.jsonl → 42` at 2026-05-04T21:20Z.

### winner: Mem0 self-hosted

### runner-up: claude-mem

### decision_matrix_score

Weights (CEO-locked):

| Factor | Weight | Mem0 | claude-mem | LightRAG-resurrect |
|--------|--------|------|------------|--------------------|
| Pillar #9 compatibility (hard gate) | GATE | PASS | PASS | PASS |
| $30 combined ceiling (hard gate) | GATE | PASS ($0 L3 incr.) | PASS ($0) | PASS (~$1) |
| OAuth-only auth (hard gate) | GATE | PASS (local Ollama) | PASS (no LLM client) | PASS (Ollama
| Semantic recall capability | 30% | 9/10 (vector search, 865 facts, 80% baseline) | 2/10 (BM25
| Current deployment state | 25% | 10/10 (running, wired) | 7/10 (installed, not primary) | 4/10
| Multi-client SVE isolation | 20% | 6/10 (user_id field exists; needs schema extension) | 1/10
| Integration risk | 15% | 9/10 (lowest risk, already passing Phase 1) | 7/10 (zero infra risk,
| Recall@10 ≥80% (chip-huyen SC-1) | 10% | 10/10 (80% confirmed) | 1/10 (no baseline, BM25 limit

**Weighted scores:**
- Mem0: (0.30×9 + 0.25×10 + 0.20×6 + 0.15×9 + 0.10×10) = 2.7+2.5+1.2+1.35+1.0 = **8.75**
- claude-mem: (0.30×2 + 0.25×7 + 0.20×1 + 0.15×7 + 0.10×1) = 0.6+1.75+0.2+1.05+0.1 = **3.70**
- LightRAG-resurrect: (0.30×7 + 0.25×4 + 0.20×3 + 0.15×2 + 0.10×3) = 2.1+1.0+0.6+0.3+0.3 = **4.3

### defend_stop-hook-l3-memory-spec

The pre-commitment in `stop-hook-l3-memory-spec.md` (MC #99071) is **DEFENDED**.

Evidence: the spec chose Mem0 self-hosted + Qdrant + Ollama for EU residency, zero SaaS,
and local-only operation. All three constraints remain valid in 2026-05-04 context. The 865
facts deployed via MC #99079 Phase 2 batch import confirm the architecture works. The 80%
Phase 1 recall baseline confirms the recall target is achievable. Nothing in the MC #99124
research overrides this choice.

### why_not_others

**claude-mem:** BM25 keyword search cannot replace semantic vector recall. When John asks
"what was the root cause of the Drop outage?" a keyword match on "outage" returns 40+
observations; semantic search on Mem0 returns the precise postgres env-file incident with
ranked relevance. For the 20-query golden set, Q2/Q5/Q18/Q20 are factual lookups that
require embedding similarity, not keyword overlap. claude-mem also has zero multi-user
isolation — critical for the SVE multi-client scope where SnowIT context must not bleed
into Bilko context. AGPL-3.0 license creates commercial-use risk for client-facing
```

deployments. Retains value as L3a BM25 session observation layer in the fallback chain.

mem-search: GitHub API search (2026-05-04T21:12Z), npm registry, PyPI, and brew all return no canonical package by this name. The YouTube source video (w0S-khYCaB4) uses "mem search" as a category description for semantic recall tools, not as a specific product. No installation path, no version, no maintainer. Cannot be evaluated or deployed.

Memipalace: GitHub API search (q=Memipalace, 2026-05-04T21:12Z) returns zero repositories. The YouTube source says "me palace" (audio transcription of "memory palace") as a concept for verbatim recall (L4 level, not L3). No software package exists under this name. Cannot be evaluated or deployed.

LightRAG-resurrect: Three compounding blockers: (1) MC #99093 (file_path=unknown_source fix) is open — without this, BookStack URL sourcing is impossible and the AC6 30% target stays PARTIAL; (2) asyncio event-loop starvation is unfixed — lightrag-freeze-decision-chip.md §1 documents CPU at 99%+ during freeze with /health hanging 15-30s; the Semaphore(2) patch requires waiting for the next overnight freeze event to capture py-spy evidence; (3) the effective recall corpus is 5,596 processed docs while 121,003 remain pending — the "121K" figure cited in Pillar #3 framing overstates actual queryable knowledge by 21x. Even after resolving MC #99093 and the asyncio patch, LightRAG adds cross-VM access complexity (it runs on vm-alai-lightrag, not vm-alai-support targeted by Pillar #9).

kill_criteria

Conditions that would invalidate the Mem0 winner choice within 6 months:
1. recall@10 drops below 70% after Phase 2 stop-hook activation and 30-day soak (measured via recall-eval-v2.sh Q1-Q20 baseline comparison)
2. Ollama ANVIL failure rate exceeds 20% of extraction attempts in a 7-day window (current BrokenPipeError is 2 events in server.log — acceptable; >20% is not)
3. Multi-client SVE schema cannot be extended beyond user_id='john' without a full collection-per-client migration costing >40h (§8 must clarify this by Phase 3)

tradeoffs_accepted

- HNSW index not built at 865 points (full scan latency ~200ms acceptable at this scale; index will build automatically when points_count exceeds 10,000)
- No graph-style entity relationships (LightRAG strength abandoned); Mem0 recall is semantic similarity, not graph traversal — acceptable for L3 operation facts
- AGPL-3.0 claude-mem in fallback chain creates license dependency; mitigated by it being a read-only search tool, not a deployed service

dissent_log

anthropic-architecture concern: AC6 of MC #99079 returned PARTIAL because LightRAG ingestion lacks file_path source URLs. Do not assume 121K docs are usable — the effective corpus is 5,596. INCORPORATED: §2.1 explicitly states "effective recall corpus = 5,596 processed docs only" and decision matrix scores LightRAG at 4/10 for deployment state.

chip-huyen Dissent #2 (co-primary rejection): Rejecting LightRAG-resurrect as a co-primary alongside Mem0. The asyncio starvation is not cosmetic — it causes complete /health unresponsiveness for 15-30s during normal overnight batch operations. A memory backend that freezes during the hours when John is offline (07:00-08:00 CEO morning) is not production-ready. Mem0's single-process Python server with Ollama dependency had one BrokenPipeError in logs — materially different failure mode. INCORPORATED: singular winner,

```
no co-primary.

---
```

Revision #2
Created 2026-05-07 10:24:44 UTC by John
Updated 2026-06-07 20:01:26 UTC by John