

Reflect Memory Architecture

Date Updated: July 22, 2026

Reflect Memory is a **persistent organizational memory layer** that sits below the AI tools your team already uses (ChatGPT, Claude, Cursor, Gemini, Grok, n8n, and others). Store context once; every connected tool can read and write under the same visibility and audit controls.

This document is intentionally high-level for public diligence. Implementation details, internal APIs, and deployment internals are shared under NDA during a private pilot.

What Reflect provides

- **Shared memory across tools** — One memory store for the organization, not a silo inside a single vendor’s product.
- **Manual and Ambient writes** — Explicit writes from the dashboard/API, plus Ambient Mode (default On) for silent capture and retrieval. Ambient can be turned Off per user; Off uses Suggested Memories (approve / edit / reject).
- **Visibility controls** — Per-memory vendor scoping so each tool only sees what it is allowed to see.
- **Team and org sharing** — Personal memories by default; sharing to org or sub-team is an explicit action.
- **Auditability** — Reads, writes, auth, and admin actions are logged for compliance review and export.
- **Temporal context** — Related memories, threads, and updates stay navigable so agents can distinguish current vs superseded decisions.

Integration surfaces

Surface	Role
Dashboard	Clerk-backed UI for browsing, editing, Ambient settings, and team admin
REST API	HTTP clients, automations, and custom integrations
MCP	Model Context Protocol for Cursor, Claude, and other MCP-capable clients
OpenAPI	Public agent schema for ChatGPT Custom GPT and similar OpenAPI consumers

Production API: <https://api.reflectmemory.com>

Dashboard: <https://reflectmemory.com>

Docs: <https://reflectmemory.com/docs>

Ambient Mode (product behavior)

Setting	Behavior
On (default)	Relevant memories are retrieved before substantive replies; durable facts can be captured without a per-write approval step. Captures remain personal until the user shares.
Off	Manual writes only, plus Suggested Memories staging for approve / edit / reject.

Ambient captures are audited like any other write. Org/team sharing remains explicit.

Deployment options (overview)

Reflect ships three buyer-facing deployment shapes:

1. **Hosted** — Multi-tenant Reflect cloud.
2. **Isolated-hosted** — Dedicated runtime and data plane per tenant.

3. **Self-host / private** — Runs inside your network with private-boundary controls (model egress, SSO, webhook exposure).

See [/diligence/deployment-architecture](#) and [/diligence/security-compliance](#) for the evaluation checklist. Exact configuration and install packages are delivered during pilot onboarding.

What we do not publish here

Public diligence does **not** include source code, internal service topology, schema dumps, algorithm descriptions, or competitive “how we beat X” internals. Those materials are available under NDA for active evaluations.

Contact

Technical and enterprise evaluation: Van Mendoza <vm@reflectmemory.com>