Security & Compliance

Questions this doc answers

- How does Reflect approach SOC 2 / HIPAA-oriented controls?
- What is audited?
- How does Ambient Mode interact with consent and sharing?
- What changes in a private deployment?

Controls & coverage

- **SOC 2 Type II alignment** is in progress for the hosted environment. The same control themes (access control, encrypted transport, rate limits, auditability) apply to isolated and self-host pilots.
- **HIPAA-oriented boundary** for regulated workloads is achieved via private deployment, egress controls, and customer-held keys — not by sending PHI to a public multi-tenant demo. A formal BAA is not available in V1; encryption at rest for memory storage is an operator responsibility in self-host.
- **SSO / OIDC** is available for enterprise identity; missing required identity config fails closed at startup in private deployments.
- **Agent credentials** are scoped so MCP/agent traffic cannot casually reach admin surfaces.

Ambient Mode & consent

- **Default On, user-toggleable** in Settings. Off switches to Suggested Memories (approval-required).
- **Phase 1 scope:** Ambient captures stay in the user's personal pool. Org/team visibility requires an explicit share action.
- **Provenance:** Ambient writes are tagged and attributed to the writing tool/vendor.
- Regulated buyers can disable Ambient per user while keeping manual MCP/REST writes.

Audit trail

Every read, write, auth, and admin action is recorded — including Ambient captures and in-place updates. Exports support compliance review and SIEM ingestion. Logs include who/what/when style fields suitable for regulated review (user, memory, operation, timestamp, origin).

Telemetry & residency

- Cloud collects minimal service-health telemetry.
- Self-host defaults telemetry off.
- Private deployments keep data inside your boundary unless you explicitly open approved integrations.

Guardrails buyers should verify in a pilot

- Vendor visibility scoping on every read path
- Soft delete / recovery behavior
- Org vs sub-team sharing scopes
- Model-host allowlisting when running air-gapped
- Admin and rate-limit posture for your threat model

Detailed control matrices, penetration-test summaries, and questionnaire answers are shared under NDA.