Shadow — Security & Architecture Brief

Independent, open-source verifier for automated credit decisions — adverse-action evidence tracing under Reg B / ECOA and GDPR Article 22.

Version 2026-07 · MIT-licensed · github.com/alex-jb/shadow-mentor

---

1 · What Shadow is

Shadow verifies automated credit decisions **independently**: it did not build your model, which is why a regulator can trust it to check the record. Every adverse-action decision is traced to its evidence bundle and reason codes, the first step that will not hold up under examination is flagged, and the exam-ready record preserves a human's sign-off.

2 · Architecture: local-first, offline-verifiable

- **Capture:** decision events are recorded in a `shadow-evidence/v1` bundle — each event hash-chained to its predecessor, the batch root signed with Ed25519 at capture time.
- **Verification:** the open-source CLI (`npm run verify:bundle -- bundle.json --public-key public.pem`) recomputes the chain and batch root and checks the signature — offline, with CI-friendly exit codes (0 verified, 1 failed, 3 I/O).
- **Primitives:** `shadow-attest-core` (npm) — zero-LLM-dependency cryptographic evidence primitives, independently verifiable offline.
- **This site:** a static build. All case processing happens in the visitor's browser; there is no server-side compute, no database, no telemetry. The public demo case is entirely synthetic.

3 · Threat model

Shadow assumes the record-keeper may be the adversary: a lender's own logs can be edited after the fact. Bundles are therefore hash-chained and signed at capture, so tampering after issuance is detectable by anyone holding the public key — without trusting the party that produced the record. The public site's attack surface is limited to static asset hosting, served with `nosniff`, a strict referrer policy, restricted permissions, and a report-only Content Security Policy. A malformed or tampered case fails validation loudly rather than rendering a misleading audit state.

4 · Key management

Bundles are signed with Ed25519. Keypairs are generated by the open-source CLI on the operator's own machine; the private key never leaves that machine and is never held by this site or any hosted service. Verification requires only the public key (SPKI/PEM), which is safe to distribute to examiners, auditors, and counterparties. Key rotation and custody are the operator's responsibility today; HSM/KMS integration is on the roadmap, not claimed now.

5 · What verification does — and does not — prove

Verification proves that the supplied evidence matches the sealed record: the hash chain is unbroken and the signature is valid for the stated public key, so nothing was added, removed, or reordered after sealing.

> **Verification ≠ correctness.** It does not prove analytical correctness, business correctness, or regulatory compliance — and it never substitutes for the accountable human reviewer whose sign-off the record preserves.

6 · Data handling

Local-first by default: cases, packages, and preferences are processed and stored in the user's browser (IndexedDB / localStorage). Nothing is uploaded — there is no Shadow backend to receive it. Optional features that touch a network are off by default, disclosed at the point of use, and act only on explicit user action:

- AI command understanding (bring your own key): sends command text and view names to api.anthropic.com only after the user supplies their own key, stored only in their browser. Case content is never included.
- Voice commands: audio is processed by the browser vendor's speech service; the application never receives it.
- Internet sync (experimental): shares only the tour step and focused node via a public signaling server (signaling.yjs.dev) — never case content.

7 · Compliance roadmap (dated, honest)

Nothing below is a certification Shadow already holds.

- **2026 Q3 — shipped:** MIT-licensed verifier CLI with CI-friendly exit codes; automated privacy/secrets scan on every commit; this brief.
- **2026 Q4 — planned:** third-party penetration test, scoped with the first guided pilot.
- **2027 H1 — planned:** SOC 2 Type I examination, contingent on a production pilot. Not started, not held.
- **Under evaluation, no committed date:** ISO/IEC 27001; HSM/KMS-backed key custody for managed deployments.
