# How Verillian is designed for trust

light

## Why this document exists

Two things are true at once in regulated sectors. AI is already in use by your staff, and the cost of getting the controls wrong is measured in millions and headlines. The figures below are the backdrop every IT and security leader is working against, and they are why a governance control that you can prove, not just deploy, is now a procurement question rather than a research one.

| | |
|---|---|
| **The breach is expensive** | The global average total cost of a data breach is $4.44 million, and in the United States it reached a record $10.22 million (IBM Cost of a Data Breach Report 2025). Healthcare is the costliest industry at $7.42 million, and financial services is second at $5.56 million (same report). |
| **Public-sector targeting is rising** | Ransomware attacks on government organizations rose 41 percent in the first nine months of 2025 over the same period in 2024, 276 attacks versus 196, with 443,522 records breached in the confirmed attacks (Comparitech, October 2025). The FBI's 2024 Internet Crime Report logged 220 ransomware complaints from the Government Facilities sector, the third-most of any critical-infrastructure sector. |
| **The data at stake is enormous** | Roughly 243 million individuals had protected health information exposed in U.S. healthcare breaches that occurred in 2024, per HHS Office for Civil Rights reports to Congress, driven largely by a single incident. In education, a threat actor claimed to have stolen data on about 62.4 million students in the PowerSchool breach, a figure PowerSchool did not confirm (BleepingComputer, January 2025). |
| **And the tools are already inside** | In a December 2025 survey of 518 U.S. healthcare professionals, 40 percent had encountered an unauthorized, or shadow, AI tool in their organization and nearly one in five admitted to using one (Wolters Kluwer Health, January 2026). The exposure that is hardest to govern is the use no one approved. |

## Where your data lives

Verillian is runtime governance and evidence for regulated AI, and it is built so the sensitive parts never leave your control. The design choice underneath everything below is that the company you buy from is never in custody of your prompts, your outputs, or your record.

| | |
|---|---|
| **On your infrastructure** | The admin service, ingest service, and database run in your environment. Verillian does not host them. |
| **At the device** | The sentinel sits on your staff's devices, before content reaches a provider. The decision is made locally. |
| **Nothing flows back** | Verillian delivers signed software and a license key. Nothing about your traffic returns to us, and the license verifies offline with no phone-home. |

## How your content is protected

Customer-key encryption is shipped today. The content Verillian handles is unreadable to anyone but you, including to us.

| | |
|---|---|
| **Encrypted before it is stored** | Content is encrypted with your organization's key using X25519 envelope encryption. Your server holds metadata and opaque ciphertext, not readable content. |
| **You hold the key** | Decryption happens in your authorized users' browsers. Verillian has no access to your key and cannot decrypt your data. |
| **Tamper-evident audit** | Every captured interaction is SHA-256 hash-chained and Ed25519-signed on the device. Any modification to a captured record breaks the chain and is detectable, verifiable mathematically and independent of vendor attestation. |

## How enforcement works

Policy is applied at execution, on the device, before any content reaches a provider. The control sits where the action happens, not downstream where the traffic has already left.

| | |
|---|---|
| **Decided at execution** | The decision is made on the device before content reaches a provider: allow, redact, or block. A blocked request never leaves. |
| **Fail-closed** | When a decision is uncertain or no valid policy is present, Verillian blocks. It never defaults to allow, and unaudited AI is a stop condition, not a degraded mode. |
| **Signed policy** | Policies are Ed25519-signed by your admin service and verified by each sentinel before enforcement. Unsigned or revoked policies are rejected. |
| **Best-effort redaction** | Detection and redaction of regulated data run before egress. This is best-effort and does not guarantee that every sensitive value is caught, so it complements your own data-handling controls rather than replacing them. |

## What this maps to in your sector

The control sits at the device and writes a signed record, which is the shape most regulated-sector audit requirements take. Verillian is designed to satisfy these controls and is aligned to their requirements, not independently certified.

| | |
|---|---|
| **Criminal justice information** | Built to CJIS Security Policy v6.0: tamper-evident logging mapped to the NIST 800-53 AU (audit and accountability) family, one-year retention, and fail-closed on loss of audit. CJIS compliance is validated through your agency's state-agency audit. |
| **Protected health information** | Built for HIPAA-regulated environments: best-effort redaction before the boundary, customer-held keys, and six-year documentation retention. |
| **Controlled unclassified information** | Keys on the device, localized deployment, and enforcement at the moment of execution, including air-gapped, supporting CMMC 2.0 and NIST 800-171 environments. |

## Questions security teams ask first

| QUESTION | ANSWER |
|---|---|
| Where is our data stored? | On your infrastructure, encrypted with your key. Verillian stores none of it. |
| Can Verillian see our prompts or outputs? | No. We never receive them, and content is encrypted with a key only you hold. |
| Do you train on our data? | No. We never receive it. |
| Is the audit log tamper-evident? | Yes. Every captured interaction is hash-chained and signed, so any modification to a captured record is detectable. |
| Which AI tools and providers are covered? | Any AI tool or agent that reaches a provider over standard secure web traffic. Audit coverage is fully parsed for Anthropic, Claude and Claude Code today, with best-effort detection across other providers that expands to the ones your institution uses. |
| How is the software authenticated? | Server components build from reviewed, tagged source. The sentinel is code-signed and notarized. Policies and licenses are Ed25519-signed. |
| Does it work air-gapped? | Yes. License verification and core operation require no Verillian connectivity. |
| What does Verillian hold about us? | Business-contact and billing information only. |

## Verillian's internal security posture

Verillian maintains a documented security program: enforced MFA on all critical systems, a password-manager standard, least-privilege access control, a documented incident-response plan, backup and recovery procedures, and code-repository controls including branch protection, secret scanning, and signed releases. Our offline license-signing keys are protected separately. The product is distributed under the Elastic License 2.0 (source-available) with an Ed25519 license-key commercial model. Summaries are available to customers under NDA.

## Your responsibilities

Because Verillian runs in your environment, you are responsible for your infrastructure and its security, backing up the database, safeguarding your encryption key, managing your users, and applying the updates we provide. Losing the encryption key means losing the ability to decrypt your own data, by design.

> Your keys, your record, your environment. Verillian provides the governance and the proof, and never holds your data.

This overview is intended to be forwarded to the people who decide. If your security team has questions before a pilot, send them over and we will answer in writing.

hello@verillian.ai