Evaluating an AI governance control

measured, plain, procurement-grade

Your staff are already using AI. The board wants it in play and the regulator wants the controls, so the decision is no longer whether to govern AI but which control to trust with the proof. These questions test the two things a regulated buyer is accountable for: runtime governance, what the control sees and decides before content leaves the device, and evidence, whether the record survives a challenge to its integrity. Each question is paired with what a weak answer tells you.

$4.44M

Global average cost of a data breach in 2025, and a record $10.22 million in the U.S. Healthcare stays the costliest sector at $7.42 million. Use these to calibrate how hard to press on the evidence questions. Source: IBM Cost of a Data Breach Report 2025.

The questions

Score each vendor's answer pass, partial, or fail. A partial is a qualifier, and the qualifier is the gap. The categories are ordered the way risk accumulates: a control that fails coverage cannot be saved by strong evidence, because it never saw the interaction to record it.

1 · COVERAGE: CAN IT SEE THE AI YOU DID NOT APPROVE

- [ ] Does it see unsanctioned tools, personal accounts, and new providers, or only what we configured it to watch? If coverage depends on per-tool integration, shadow AI is invisible to it by design, and shadow AI is where the exposure lives.
- [ ] Does it govern the AI tools and agents that reach a provider over standard secure web traffic, without changes to those tools? Per-tool or per-provider coverage leaves gaps that grow every time staff adopt something new, and you inherit a new project each time.

2 · THE DECISION: DOES IT ACT BEFORE THE DATA LEAVES

- [ ] Is the decision made on the device before content reaches a provider, or only observed after it has already gone? After-the-fact detection cannot prevent the leak; it only documents it. Ask exactly where in the path the decision is made.
- [ ] Can it redact or block sensitive content before egress, not just flag it afterward? Confirm whether detection and redaction are best-effort or claimed as complete. A vendor that claims it catches every sensitive value is overclaiming; the honest answer is that detection is best-effort and complements your own data-handling controls.
- [ ] When a decision is uncertain, or the audit pipeline fails, does it fail closed? A control that defaults to allow under ambiguity is not a control in the cases that matter most. Unaudited AI should be a stop condition, not a degraded mode.

3 · DATA AND KEYS: WHO CAN READ WHAT FLOWS THROUGH

- [ ] Who holds the encryption keys, and can the vendor read our content? If content is encrypted under a key only the institution holds, the vendor cannot read it, even the vendor. If a vendor can read it, that vendor becomes the custodian of your most sensitive data and a new breach surface, the one the figures above are made of.
- [ ] Does any plaintext leave our boundary to the vendor, ever, including for support or troubleshooting? A single plaintext path defeats the custody answer. The defensible posture is that the server stores ciphertext it cannot read and the vendor never receives your AI content.

4 · EVIDENCE: WILL THE RECORD SURVIVE A CHALLENGE

- [ ] Is each captured interaction signed at the source and hash-chained, or is it an ordinary log the keeper could quietly edit? Only a signed, append-only record where each entry depends on the one before it is tamper-evident, so any change is detectable. Note the wording: tamper-evident, which is provable, not tamper-proof, which no one can honestly claim.
- [ ] Do we hold the evidence, and can we verify it mathematically, independent of vendor attestation? Evidence you cannot independently verify is the vendor's word, not yours. In an audit, a lawsuit, or a criminal case, the vendor's assurance is not in the room; the record is.
- [ ] Does the record cover the captured interaction end to end: what was asked, what the decision was, what was redacted, and where it went? Ask what is captured and what is not. A record of the captured interactions is honest; a claim of complete or non-omissible coverage of every interaction is not.

5 · DEPLOYMENT: CAN IT RUN WHERE THE NEED IS GREATEST

- [ ] Does it run on our own infrastructure, including air-gapped environments, and verify its license without phoning home? Cloud-only controls cannot serve the most regulated environments, where the need is greatest and outside connectivity is restricted by rule.
- [ ] Does it deploy through the device management we already run, without changing how staff work? If enrollment is a custom integration rather than a software push through Jamf, Intune, or Group Policy, the timeline and the risk both grow.

6 · COMPLIANCE FIT: DOES THE RECORD MAP TO THE FRAMEWORK WE ANSWER TO

- [ ] Does the vendor map controls to the specific framework you report against, and state plainly whether it is aligned or independently certified? Aligned to a framework's requirements is an honest claim; certified is a different and verifiable one. Press on which they mean, and ask for the control-by-control mapping in writing.
- [ ] Does the retention and audit behavior match your obligations, by name and version? For criminal justice data, look for CJIS Security Policy v6.0 tamper-evident logging, one-year retention, and validation through your agency's state-agency audit; for protected health information, redaction before the boundary and six-year HIPAA documentation retention; for controlled unclassified information, enforcement at the moment of execution with keys on the device. The NIST 800-53 audit (AU) family is the common spine.

7 · COMMERCIAL: CAN YOUR TEAM ACTUALLY BUY IT

- [ ] Is pricing predictable and on a vehicle your finance team has approved before? Per-device, per-year pricing on cooperative purchasing routes shortens the path from decision to deployment, because the buying mechanism is already cleared.
- [ ] Is the software source-available so your team can read what runs in your environment? Code your security team can inspect under a license like Elastic License 2.0 is code they can stand behind in a review. Source-available is not the same as open-source; ask which it is and under what license.

> Many vendors answer the first three categories cleanly and start hedging at evidence. That hedge is where a control that reports cleanly parts ways with one that holds up under a regulator, an auditor, or a court.

How Verillian answers the same questions

We hold ourselves to this checklist. Here is the short version, on the record, so you can score us alongside everyone else.

THE QUESTION	VERILLIAN'S ANSWER
Coverage of unsanctioned AI	A sentinel on each device governs any AI tool or agent that reaches a provider over standard secure web traffic, with no plugins or per-tool integration. Detection across providers is best-effort; audit coverage is fully parsed for Anthropic, Claude, and Claude Code today and expands to the providers our first institutions use.
The decision point	On the device, before content reaches a provider: allow, redact, or block. Redaction before egress is best-effort and complements your own controls. When a decision is uncertain or the audit pipeline fails, it fails closed.
Data and keys	Audit content is encrypted under a key only you hold, on a server that stores ciphertext it cannot read. Verillian never receives your AI content, and even Verillian cannot read it.
Evidence	Every captured interaction is Ed25519-signed on the device and SHA-256 hash-chained. Any change to a captured entry breaks the chain and is detectable. You hold the record and can verify it mathematically, independent of vendor attestation.
Deployment	Self-hosted on your infrastructure, air-gapped capable, license verified offline with no phone-home. Sentinels deploy through Jamf, Intune, Kandji, Configuration Manager, or Group Policy.
Compliance fit	Built to CJIS Security Policy v6.0, for HIPAA, and to NIST 800-53 audit controls; aligned to these frameworks, not independently certified. CJIS is validated through your agency's state-agency audit. Control-by-control mappings on request.
Commercial	Per device, per year, available on cooperative purchasing vehicles. Source-available under Elastic License 2.0, so your team can read what runs in your environment.

Watch where the answers turn into qualifiers. The qualifier is the gap, and in a regulated sector the gap is the finding.

Bring this checklist to your evaluation of Verillian. We will answer every question on the record, in your own environment, under your own keys.

hello@verillian.ai