Shadow AI: the risk you cannot see

light

40% of healthcare professionals have encountered an unauthorized ("shadow") AI tool in their organization, and nearly one in five admit to using one. Source: Wolters Kluwer Health, January 2026 (survey conducted by CITE Research, December 2025, n=518 U.S. healthcare professionals).

$670K added to the average breach cost when shadow AI was involved. Source: IBM Cost of a Data Breach Report 2025.

$4.44M global average total cost of a data breach; the U.S. average reached a record $10.22 million, and healthcare remained the costliest industry at $7.42 million. Source: IBM Cost of a Data Breach Report 2025.

What shadow AI actually is

It is not one tool. It is every path to a model that was never sanctioned, reviewed, or even known to exist. The defining trait is not the tool, it is that no one in your institution chose it, so no control was ever placed in front of it.

- An employee pasting a record into a public chat model from a personal account your institution does not administer.
- A coding assistant or command-line client reaching a provider that was never reviewed.
- A browser session signed into an AI tool the institution does not know exists.
- A new agent wired up by one team, invisible to everyone else.

Why most controls miss it

Most controls assume traffic takes a sanctioned path. A network gateway sees only what was routed to it. A closed assistant governs only itself. Provider-side settings apply only after the data has already arrived. A VPN or cloud proxy cannot see inside an HTTPS session, so it cannot decide what an AI tool may do. Each one assumes away the exact traffic that defines shadow AI, so it can report clean while the real risk runs beside it, unseen.

Why it is no longer about pasted text

Modern AI tools do not just chat. They run shell commands, read and write files, query production databases, call internal APIs, and browse the web. Point one of those at a system of patient records, criminal histories, student files,

or controlled information, with no control in the path, and the exposure is no longer measured in a leaked paragraph. It is measured in **actions taken on your most sensitive systems**, with no record of who did what. In healthcare alone, breaches that occurred in 2024 exposed the protected health information of roughly 243 million individuals, driven largely by a single incident (HHS Office for Civil Rights reports to Congress).

What closes the gap

THE EXPOSURE YOU OWN TODAY

- **Personal accounts:** staff reach providers from accounts your institution does not administer.
- **Unreviewed clients:** coding assistants and command-line tools call out from the machine with no policy in the path.
- **Browser sessions:** tools signed into in a tab that no gateway or closed assistant can see.
- **No record:** when an action does reach a sensitive system, nothing captures who did what.

WHAT CHANGES WITH VERILLIAN

- **One position:** a control on the device, where traffic leaves the machine, that sees the sanctioned and the unsanctioned alike, whoever owns the account.
- **Decided before egress:** allow, redact, or block before content reaches a provider. Sensitive-data detection and redaction are best-effort, not a guarantee that every value is caught.
- **Across the tools on the device:** coverage spans the AI tools that reach a provider, not only the ones wired to a gateway. Audit is fully parsed for Anthropic, Claude, and Claude Code today, with best-effort detection across other providers.
- **Evidence you hold:** every captured interaction is signed on the device and hash-chained into a tamper-evident record, encrypted under a key only you hold.

This is runtime governance and evidence for regulated AI, placed at the one position that can see traffic before it leaves the machine. The record is tamper-evident, not tamper-proof: any change to a captured interaction is detectable, and the content is verifiable mathematically, independent of vendor attestation. **The tools you approved are not the problem. The ones you did not are.**

Seeing the AI traffic you never approved is the whole job.

If your staff are already using AI you cannot see, a short pilot in your own environment, on your own infrastructure, under your own keys, will show you exactly what is leaving your machines. Name a sponsor and a technical contact and we will scope the first week.

hello@verillian.ai