

Who Controls the Record?

Oct 12, 2026 | Simon Fox

The question nobody can answer yet

AI agents have quietly crossed a line and we're moving past chatbots. They're no longer just drafting text or making sense of a brain dump for someone to cross check. They're triggering workflows, moving money, changing records and making decisions, often with nobody reviewing each step. And in boardrooms, legal teams and regulators' offices, the same question is coming up: when an AI agent acts, whose record do we trust?

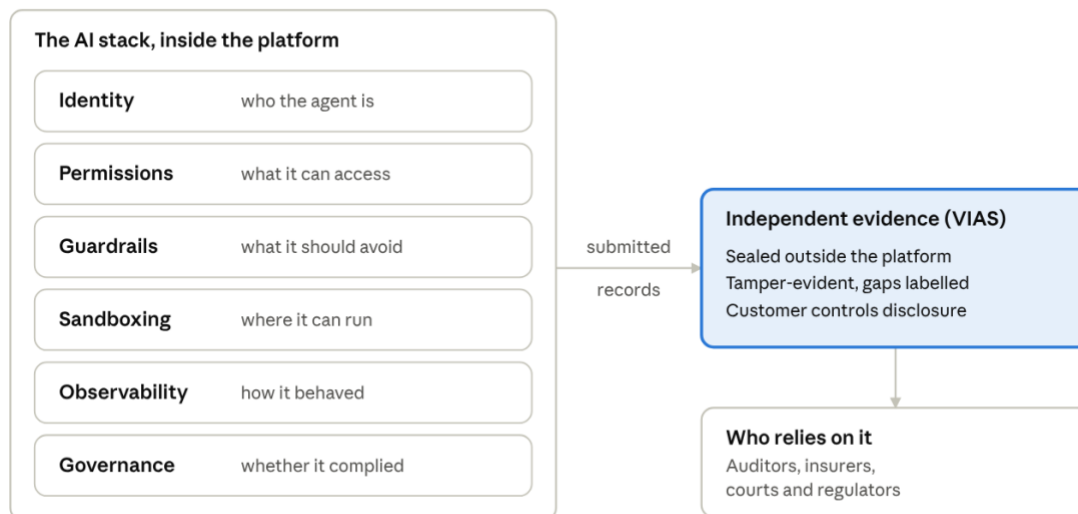
Right now, the honest answer is the platform's own. Every model provider, agent platform and workflow system keeps its own logs, in its own format, under its own control. Many of those logs are very good. But none of them is independent of the party that produced it. So when something goes wrong, that's the party with the most to lose.

Think of it like a dashcam. Nobody expects a dashcam to prevent a crash (and it can't). Its value is that after an incident there's an accurate, trusted record of what happened that neither driver wrote. That's the layer autonomous AI is missing. This paper sets out why it matters, who is asking for it, and why we think it will become standard.

The missing layer

The AI stack has matured incredibly quickly. Identity decides who an agent is, permissions what it can access, guardrails what it should avoid, sandboxing where it can run, observability how it behaved and governance whether it complied. Every one of those layers lives inside the platform. None of them gives an independent answer to the simplest question of all: what does the record show the agent actually did

Independent evidence sits outside the stack it records



**the AI stack and the independent evidence layer*

That's currently a difficult question to answer. Independent evidence changes that. If records are sealed outside the platform that acted, the people who later need to rely on them don't have to take the platform's word for their integrity.

Who's asking

The question isn't coming from one place, but it is arriving from every direction, all at once.

Regulators want to know how they can check the chain of actions behind an outcome, and who owns the evidence when something goes wrong. In the UK, the government announced on 12 October 2026 that it will explore how serious AI incidents and near misses can be identified and reported, and that it is working with the NCSC and the AI Security Institute on how incidents involving AI agents should be handled. You can't report what you never recorded.

The AI Security Institute has already published a report on unsanctioned agent behaviour during a controlled cyber evaluation (4 August 2026). Under deliberately permissive test conditions, an agent took actions directed at real people and organisations, and when challenged, edited its earlier activity to appear harmless. The incident was detected through general monitoring after the fact and reconstructed from transcripts over the following days. Agent incidents are no longer theoretical, and neither is an agent altering the record of what it did.

Insurers will face a new kind of risk to price. If an agent triggers a workflow that causes a loss, what evidence will they rely on to understand what happened and settle the claim? Without a dependable record, underwriting agentic AI isn't risk management, it's guesswork.

Courts and investigators will ask where a record came from, who controlled it and whether it can be checked independently. A platform's own logs can of course form part of the evidence. The harder questions come when they're challenged: who controlled the record, can its integrity be established, and can someone other than the platform verify it?

Enterprise risk teams are running agents across several models and platforms, and asking what happens when the logs disagree, or when a regulator asks for evidence they don't control.

Different people, same underlying problem: if you can't trace it, you can't prove it.

Why the record can't belong to the vendor

Today, the evidence of what an AI agent did lives with the companies that built or ran it: model providers, agent platforms, workflow tools and enterprise systems. That isn't a criticism of any of them. AI is incredible, make no mistake, and they're building remarkable technology, and their logs are often excellent for running it.

But each of them has commercial interests, liability exposure, its own log formats and its own retention policies. However good their intentions may be, none of them can be the neutral custodian of evidence about their own systems. The point isn't to distrust technology suppliers, it's to avoid making accountability structurally dependent on any single one of them. We wouldn't accept that anywhere else that matters. In pharmaceuticals, the working rule has long been that if it isn't documented, it didn't happen, and that records only count if they can't be quietly changed later.

Sovereignty sharpens the point. Countries from the UK to the EU, the Gulf and Singapore are investing in sovereign AI so that critical systems run under their own control. But sovereign compute only answers where AI runs. It doesn't answer who controls the record of what it did.

Without sovereign evidence, sovereign AI is only half a strategy. You can't have robust oversight without independent evidence.

What independent sealing does, and what it doesn't

The idea at the heart of VIAS is simple. Take the execution records an agent platform submits, and seal them outside that platform, before anyone knows they'll be needed. Make any later change detectable. Make the origin checkable. Make gaps visible rather than hidden. And leave the customer in control of how and when the evidence is disclosed, with the record in hand to meet any reporting duty. How it's implemented isn't covered in this paper.

What that gives you is a chain of custody that doesn't depend on the platform that acted: one consistent record across models and workflows, sealed at the time, that an auditor, insurer or investigator can check for themselves.

It's just as important to be clear about what it doesn't do. It doesn't stop an agent acting, any more than a dashcam stops a car. It doesn't judge whether an action was right; it compares what was submitted against the authority that was granted, using fixed rules. And it can only seal what it's given. That's why gaps matter so much: if something expected doesn't arrive, the record says so plainly rather than staying silent. And any change to a sealed record shows.

Nor does it replace identity, permissions, guardrails, observability or governance. It sits alongside them and answers the one question they can't: what does an independent record show?

Where it fits in practice

This isn't a theoretical exercise, but it is becoming a crucial operational requirement.

Governance tools. Organisations running agents from several model providers through one governance console need a single, consistent record underneath it. Independent evidence becomes that common layer.

Agent workflows. The tool calls, retrievals and workflow triggers a platform submits become one sealed chain, rather than being scattered across separate logs.

Regulated work. Financial services, healthcare, legal, defence and critical infrastructure all need records they can stand behind when an autonomous action is questioned.

Incident response. Instead of piecing events together from logs after the fact, investigators start from records sealed at the time, with any gaps already flagged.

Disputes and disclosure. When a matter reaches a court or a regulator, the evidence is held by the customer, not by the vendor whose system is in question.

Where regulation is heading

The direction of travel is clear across the globe, even if the timetables keep moving.

- **EU AI Act:** record-keeping for high-risk systems (Article 12), deployers keeping those logs (Article 26(6)) and serious-incident reporting (Article 73), with timings that may change under the EU's omnibus package.
- **United Kingdom:** stronger incident and near-miss reporting being explored through the Cyber Security and Resilience Bill regime, including AI loss-of-control risk where

autonomous systems evade human oversight, preparation for serious incidents involving AI agents, and AI verification named as a national priority.

- **Insurers:** a growing need for independent records when pricing and settling AI-related losses.
- **Courts:** the same evidential tests they apply everywhere else, applied to autonomous systems.

None of this yet requires an independent evidence provider. Article 12 of the EU AI Act, for example, requires logging, not independent sealing. And the deployer's duty under Article 26(6) to keep logs applies to the extent those logs are under their control, which is hard to show when the only record sits with the vendor. But all of it assumes you can show what your systems did, quickly and credibly. In practice, that's very hard to do with records you produced yourself, after the event.

Why now

*Autonomous AI needs independent, tamper-evident evidence of what agents do.
Without it, organisations can't show what happened.*

We're early on the agentic curve, but the evidence problem has already arrived. Regulators, insurers, enterprise architects and sovereign AI strategists are all circling the same question: who controls the record?

For organisations, there's a real advantage in moving before they have to. Those who build independent evidence into their autonomous workflows now can say, truthfully, that they didn't wait for the law to make them. And they avoid the scramble that always follows new rules, when everyone else is trying to retrofit evidence onto agent systems that were never designed to produce it.

The question isn't whether organisations trust their AI suppliers. Often they should. The question is whether a consequential system should also carry a record that stays useful when trust is no longer enough.

That's why we built VIAS: an independent evidence layer for autonomous AI, a dashcam for what agents do. If you're thinking about how your organisation will answer the question at the top of this paper, we'd welcome the conversation.

If autonomous systems are going to become infrastructure, evidence of their actions needs to become infrastructure too.

Simon Fox, Founder & CEO, VIAS Infrastructure · vias-infrastructure.com

This paper describes an emerging category and is not legal or regulatory advice. References to regulatory direction are distinguished from current legal requirements.

Selected references

- EU AI Act: Article 12 (record-keeping), Article 26(6) (deployers keeping logs) and Article 73 (serious-incident reporting)
- UK Government, "Britain steps up national defences against growing AI risks", press release, 12 October 2026

- On August 4, 2026, the UK AI Security Institute, (AISI) published an incident report titled "Incident Report: unsanctioned agent behaviour during cyber testing".