SMART VERIFY

AI DATA SECURITY & COMPLIANCE · 2026

Your AI agents are running. Who's watching the data?

Every AI regulation and every data-security framework asks the same question: what did your systems do with regulated data, and how do you know? That answer is created, or lost, at the data layer.

PER-QUERY EVIDENCE | FRAMEWORK-MAPPED | TAMPER-EVIDENT | INLINE ENFORCEMENT

10
Compliance frameworks mapped end to end

39
Distinct evidence types, each confidence-graded

<50ms
P90 inline overhead. Zero code changes

2
Normalisation hubs: AI and data security

"I was at AWS when the agentic shift happened. I watched us try to adapt cloud-era compliance tooling to autonomous AI agents, and I understood exactly why it couldn't work. Security, compliance, and data governance had always been separate silos that customers stitched together by hand. With AI agents, you cannot stitch fast enough."

Kajal Deepak · CEO & Founder, SmartVerify · Former GM, AWS Audit Manager

EXECUTIVE SUMMARY

What this paper says, in one page

AI agents operate inside your data perimeter with credentials you issued. Once inside, almost nothing observes what they actually do. That silence is now a compliance liability across two separate bodies of regulation at once.

1 The evidence does not exist by default

When an authorised AI agent queries a customer database, a clinical record store, or a payments ledger, it produces no compliance evidence. Identity tools recorded that it was allowed in. Nothing recorded what it asked for, what came back, or whether the access was appropriate. A log that was never written cannot be produced later.

2 Two sets of obligations converge on the same missing record

New AI regulation (TRAIGA, Colorado ADAI, the EU AI Act, California DROP) creates liability that traces back to data access events. Established frameworks you are already audited against (SOC 2, HIPAA, PCI DSS, CMMC, GDPR) have always required audit controls and access enforcement over regulated data. AI agents are a new class of actor those controls were not built to observe. Both sets of obligations are answered by the same evidence.

3 SmartVerify produces that evidence inline, per query

SmartVerify deploys between AI agents and your data stores with no application code changes. Every query is intercepted, classified, evaluated against policy, and written to a cryptographically signed record. Enforcement happens at wire speed. Latency overhead is under 50 ms at P90.

4 Every record maps to the frameworks your teams already manage

Evidence is normalised through two hubs, NIST AI RMF for AI regulation and NIST SP 800-53 for data security, and from there to the specific control language of ten frameworks. One observed access lands on an AI-regulation control and a data-security control simultaneously.

5 Every claim is graded, including the ones we cannot make

Deterministic evidence is marked FULL. Probabilistic machine-learning evidence is marked PARTIAL and described as corroborating. Controls requiring governance or application-layer action are marked GAP and documented as your obligation. The coverage arithmetic is published rather than asserted.

THE BOTTOM LINE FOR A CISO

SmartVerify is not another monitoring tool and it is not a governance platform. It is the evidentiary substrate underneath both: the verified, per-query record of what your AI systems did with regulated data, produced automatically because the system sits on the path the data travels. Without it, your governance documentation describes controls nobody can demonstrate were operating.

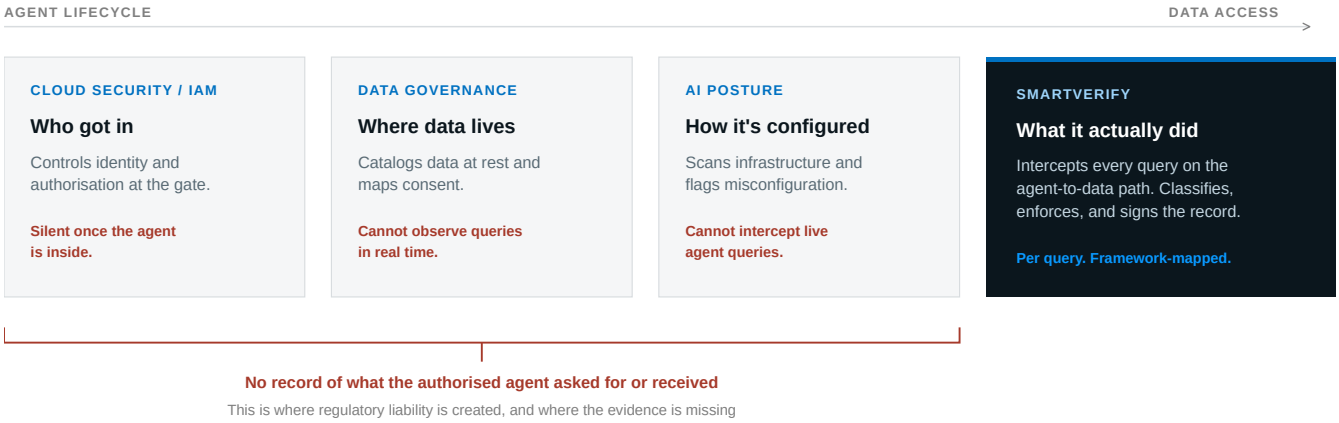
01 · THE GAP

Every tool in your stack stops before the question that matters

Your perimeter knows who was authorised. Your catalog knows where sensitive data lives. Your posture tool knows how the environment is configured. None of them know what your AI agent actually did once it was inside.

EXHIBIT 1

Where existing tooling goes silent



Identity, catalog, and posture tooling remain necessary. None of them observes the live interaction between an agent and the data it queries.

Why this gap did not exist before agents

Human users and static applications operate on predictable access patterns. You could define rules in advance, sample evidence periodically, and produce a defensible record. AI agents construct novel queries at runtime. They reach fields nobody scoped. They operate at volumes that make sampling meaningless. The compliance model built for cloud infrastructure assumed a predictable actor, and that assumption no longer holds.

An absence of records is not a neutral fact in an investigation. It reads as an absence of reasonable care.

Both TRAIGA and Colorado ADAI provide cure periods of 60 and 90 days respectively before penalties accrue following written notice. That protection is only usable if you have the monitoring in place to detect the problem, diagnose it, and demonstrate correction inside the window. An organisation without data-layer monitoring cannot act on a cure period it cannot see.

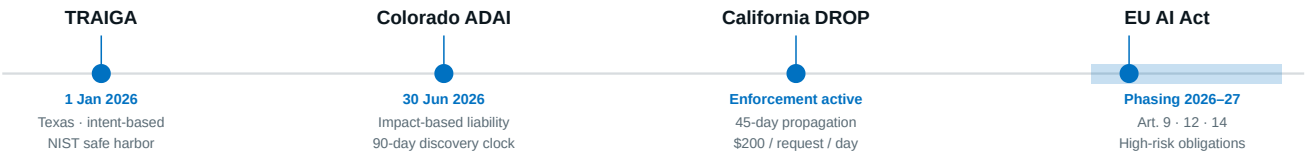
02 · REGULATORY LANDSCAPE

The clock started without you

Four AI-specific regimes create active obligations for enterprise AI in 2026. They differ in mechanism but share a common trigger: liability traces back to what data an AI system accessed and whether you can prove what happened.

EXHIBIT 2

AI regulation in force through 2026



TRAIGA TEXAS · IN FORCE

The first US AI statute with meaningful enforcement. It prohibits AI systems that unlawfully discriminate against protected classes, and it turns on intent rather than outcome. The provision that matters commercially is the safe harbor: substantial compliance with the NIST AI Risk Management Framework is an affirmative defence. That defence requires documented evidence that your systems were monitored against NIST sub-categories in production, not a policy asserting alignment.

Up to $200,000 per uncurable violation · 60-day cure period

Colorado ADAI SB 24-205 · IN FORCE

The most demanding US framework. Liability is impact-based, meaning an AI system can cause unlawful discrimination without anyone intending it, if outputs produce disparate outcomes. It requires continuous behavioural monitoring and notification within 90 days of discovering that a deployed system has caused or is likely to cause algorithmic discrimination. If your agents inform decisions on employment, housing, credit, insurance, or healthcare, it applies.

90-day discovery and notification obligation

California DROP ENFORCEMENT ACTIVE

A unified deletion and opt-out mechanism that must propagate across the entire data estate. The difficulty is structural: personal data proliferates into vector stores, RAG caches, fine-tuning corpora, and agent conversation logs that conventional deletion workflows cannot enumerate, let alone purge. Proving propagation requires enforcing deletion semantics at the point of access rather than at rest.

$200 per request per day, automated

EU AI Act PHASING THROUGH 2026–27

The most technically prescriptive regime globally. Article 12 requires logging sufficient to reconstruct decision inputs after the fact. Article 9 mandates a continuous risk management system across the lifecycle. Article 10 sets data governance requirements. Organisations with EU exposure accumulate an evidence deficit with every unmonitored query.

02 · REGULATORY LANDSCAPE, CONTINUED

The audits you are already failing

AI regulation is the newer problem. It is not the nearer one. Most organisations have a SOC 2, HIPAA, or PCI examination already on the calendar, and AI agents are generating findings in those examinations now.

These frameworks predate AI by decades, and they have always required audit controls, access enforcement, and integrity protection over regulated data. Nothing in those requirements exempts an AI agent. An agent querying a patient record is a system accessing electronic protected health information, and HIPAA §164.312(b) applies to it exactly as it applies to anything else. What has changed is that the actor is new, and existing audit tooling was not built to observe it. Three assumptions break at once.

Attribution

Audit controls assume privileged actions belong to individual accountable users. An agent typically runs on a service account. The log shows the service account acted; it does not show which agent, which query, or what was returned.

Volume

A single agent generates thousands of queries an hour, joining tables and constructing cohorts no human would. Existing logging records that access occurred. It does not classify whether the pattern fits the agent's approved purpose.

The response

Access logging captures the request. It does not capture what sensitive data actually came back: whether PHI was returned, whether a card number appeared in the output. That is the difference between logging access and proving exposure.

The frameworks in scope

FRAMEWORK	TYPE	WHAT IT REQUIRES OF AI AGENT ACTIVITY
SOC 2	Data security	Privileged actions attributable to accountable identities; system monitoring; protection of audit information from alteration.
HIPAA	Data security	Audit controls over systems containing ePHI (§164.312(b)), access control, and integrity protection (§164.312(c)(1)).
PCI DSS v4	Data security	Audit logs for all access to cardholder data (Req. 10.2), audit record content (10.3), and log protection (10.5).
CMMC 2.0	Cybersecurity	NIST SP 800-171 practices for access control and audit accountability where agents touch controlled unclassified information.
GDPR	Data protection	Records of processing (Art. 30) and security of processing (Art. 32). Privacy-governance obligations sit outside the data layer.
ISO/IEC 42001	AI management system	Demonstrated operational control over the AI system lifecycle and evidence that AI systems are used as intended. Addressed in Section 07.

WHY THIS MATTERS STRATEGICALLY

The same per-query record answers both bodies of regulation. A record showing which agent accessed which tagged field, what policy applied, and what came back is simultaneously NIST AI RMF safe harbor evidence under TRAIGA and audit-control evidence under HIPAA §164.312(b). **You are not buying two compliance programmes. You are producing one evidence set and mapping it twice.**

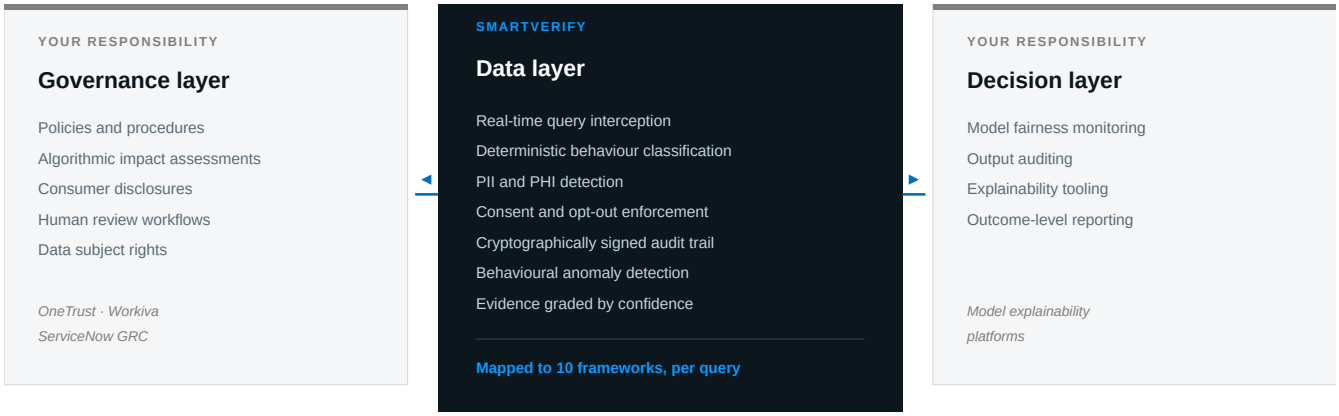
03 · ARCHITECTURE

AI compliance is a three-layer problem. SmartVerify owns the one that proves the others.

The most common strategic error is attempting to satisfy every regulatory requirement with a single platform. Compliance for AI is structurally three layers, each requiring different capabilities, different owners, and different tooling.

EXHIBIT 3

The three-layer model and where responsibility sits



The data layer supplies the operational evidence both other layers depend on

A policy nobody can demonstrate was enforced, and a fairness claim nobody can trace to an access record, are both undefendable

Why the provable facts live here

Every enforcement action against an AI system ultimately traces to a data access event. A discriminatory lending decision traces to the features the model queried. A DROP violation traces to a query that returned an opted-out user's record. A TRAIGA inquiry traces to the pattern of queries an agent made and the intent that pattern reveals.

The data layer is also the only place real-time intervention is possible. By the time a governance review or an outcome audit runs, the access has already happened.

YOUR IAM SAYS

"This agent was authorised."

Necessary, and not the same as knowing what it did. Both records are needed in an investigation.

SMARTVERIFY ADDS

"Here is exactly what it accessed, under what policy, and what was returned."

Per query. Signed. Framework-mapped.

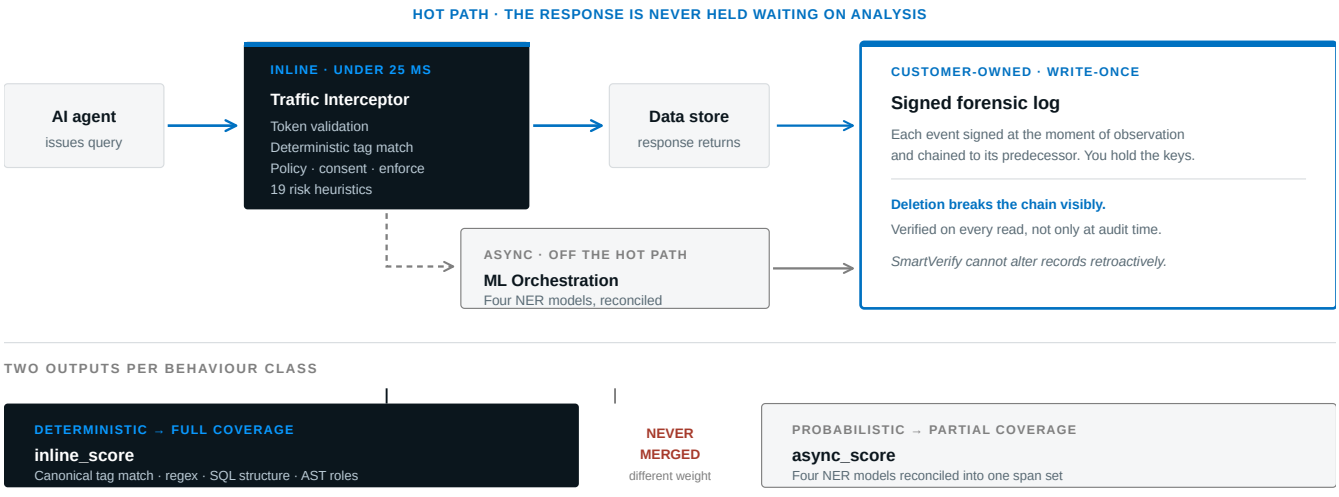
04 · HOW IT WORKS

Two inspection paths. One never blocks. Neither is ever merged.

SmartVerify deploys inline between AI agents and your data stores as a proxy: a connection-string change, with no application code modified. Every query is inspected twice, along two independent paths that produce two different kinds of evidence.

EXHIBIT 4

The query path and the two inspection paths



Raw query and response content never leaves the customer environment. Only derived labels, counts, and scores cross the data-minimisation boundary.

Why the two paths are kept apart

The inline path is deterministic. Field names are matched against a canonical tag vocabulary and return a definite yes or no. Regular expressions match structured identifiers or they do not. Structural analysis of the query either finds an unbounded projection or it does not. Nothing about it is an estimate.

The asynchronous path is machine learning. Four named-entity models run over the request and response text and their overlapping outputs are reconciled into a single span set. It catches what tags cannot: personal data in free text, and entities in an unstructured response. It is probabilistic.

These are different kinds of evidence and they carry different weight in front of an auditor. Merging them into a single score would hide which method fired. SmartVerify keeps them as separate fields, and the compliance mapping grades them differently, which is the subject of the next section.

05 · EVIDENCE ARCHITECTURE

We grade our evidence. Most compliance vendors do not.

SmartVerify produces 39 distinct evidence types, from deterministic audit records to probabilistic behavioural scores. They are not equally strong, and a mapping that treats them as equally strong is overstating most of them.

EXHIBIT 5

How every evidence record is graded

FULL COVERAGE

Deterministic

High-confidence records produced by direct observation, not inference. These prove a control was satisfied.

Per-query audit log
Policy rule fired · enforcement outcome
Consent state · opt-out redaction
Event signature and chain verification
Structured identifier detection

PARTIAL COVERAGE

Corroborating

Medium-confidence machine-learning evidence. Contributes materially. Does not alone complete a control.

Entity detection in free text
Protected-characteristic scoring
Behavioural anomaly baselines
Clinical PHI recognition

GAP · YOUR LAYER

Not ours to produce

Controls requiring governance or application action. Documented with the tooling that addresses each.

Algorithmic impact assessments
Consumer-facing AI disclosure
Human review and appeal workflow
Workforce training records
Third-party agreements

Coverage % = (FULL + 0.5 × PARTIAL) ÷ Applicable controls · published, not asserted

Evidence integrity is itself a control

Every producer signs its own event at the moment of observation. Each event is signed over its content plus the hash of its session predecessor, so the log forms a per-session chain. Altering a record breaks its signature. Deleting one breaks the chain exactly where the deletion happened, a visible break rather than a silent gap. Keys are held in your key management system, which means SmartVerify structurally cannot alter records after the fact.

Verification runs continuously rather than on demand. The service that assembles per-request records verifies every signature and chain hash as it reads, and alarms on any break. Tampering surfaces in normal operation, not during an audit six months later.

AN HONEST FAILURE MODE, HANDLED

When a field is encrypted by you, or contains an image rather than text, pattern matching finds nothing. The naive result is an evidence record reading "no sensitive data detected", a confident false negative. SmartVerify instead records which of three bases produced each classification: content-based, schema-based because the content could not be inspected, or uninspected. An entropy check catches encrypted fields you forgot to declare and falls back to the highest sensitivity tier rather than treating silence as safety.

Every vendor claims broad coverage. We publish the arithmetic, and we publish what falls outside it.

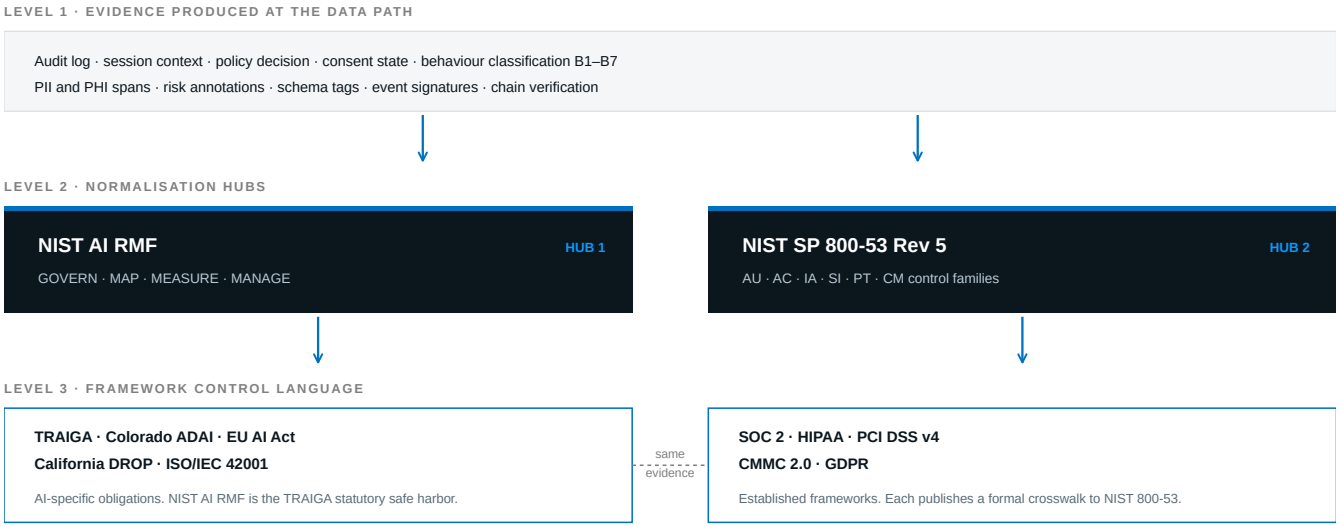
06 · FRAMEWORK MAPPING

One observed access. Two regulatory worlds. Simultaneously.

Evidence is only useful if it speaks the language your auditor already works in. A HIPAA assessor does not read NIST AI RMF. A TRAIGA safe-harbor argument does not run on SOC 2 Common Criteria. SmartVerify normalises every record through two hubs and maps outward from each.

EXHIBIT 6

The dual-hub mapping architecture



Frameworks never map directly to evidence. They map to a hub, and the hub maps to evidence. Adding a framework is one crosswalk, not a re-mapping exercise.

What this buys you

A single query in which an agent reads a tagged clinical field produces one record. Through Hub 1 that record is NIST AI RMF *MEASURE 2.3* evidence, which is TRAIGA safe-harbor material and EU AI Act Article 10 material. Through Hub 2 the same record is NIST 800-53 *AU-2* and *AU-13* evidence, which is HIPAA §164.312(b) and SOC 2 CC7.2 material. You collected once.

Your compliance team can enter any control reference (SOC 2 CC7.2, HIPAA §164.312(b), TRAIGA §552.055) and trace it back through the hub to the specific evidence records that satisfy it, each with its confidence grade and a FULL, PARTIAL, or GAP verdict.

07 · COVERAGE

What SmartVerify covers, and what it does not

SmartVerify does not claim end-to-end AI regulatory compliance. No product can. What follows is the honest mapping: where we directly satisfy a requirement, where we contribute but need other inputs, and where the obligation belongs to your governance or application layer.

EXHIBIT 7

Requirement coverage by framework

COMPLIANCE REQUIREMENT	TRAIGA	ADAI	DROP	EU AI	ISO 42001	SOC 2	HIPAA	PCI	CMMC
Per-query audit trail of agent data access	●	●	●	●	●	●	●	●	●
Tamper-evident evidence integrity	●	●	n/a	●	●	●	●	●	●
Real-time enforcement (block, redact, allow)	●	●	●	●	●	●	●	●	●
NIST AI RMF safe-harbor evidence	●	●	n/a	●	●	n/a	n/a	n/a	n/a
Discriminatory access pattern detection	●	●	n/a	●	●	n/a	n/a	n/a	n/a
PII / PHI detection in AI responses	●	●	●	●	●	●	PART*	●	●
Opt-out enforcement and 45-day propagation proof	n/a	n/a	●	n/a	n/a	n/a	n/a	n/a	n/a
Access control and enforcement evidence	n/a	n/a	n/a	n/a	●	●	●	●	●
Configuration and version control evidence	●	●	n/a	●	●	●	●	●	●
AI system lifecycle governance evidence	●	●	n/a	●	●	n/a	n/a	n/a	n/a
Accountable-human attribution of agent activity	PART	PART	n/a	PART	PART	PART	PART	PART	PART
Behavioural anomaly and 90-day discovery support	PART	PART	n/a	PART	PART	PART	PART	n/a	PART
Bulk extraction and exfiltration monitoring	●	●	●	●	●	●	●	●	●
Algorithmic impact assessment	PART	PART	n/a	PART	PART	n/a	n/a	n/a	n/a
Consumer-facing AI disclosure	o	o	n/a	o	o	n/a	n/a	n/a	n/a
Human review and appeal workflow	o	o	n/a	o	o	n/a	n/a	n/a	n/a
Workforce training and policy attestation	o	o	n/a	o	o	o	o	o	o

● Directly satisfied · **PART** Contributes materially, requires additional inputs · o Governance or application-layer obligation · n/a Not applicable under this framework

* Clinical PHI recognition is conditional on a pending data use agreement and is marked PARTIAL until resolved rather than claimed in advance.

ON ACCOUNTABLE-HUMAN ATTRIBUTION

Every access is attributed to an authenticated, registered application identity. Where an application resolves user authorisation internally and calls downstream with its own service credential, the human was stripped before the data path saw anything. Naming the application narrows the generic-service-account finding your auditor is raising. It does not close it, and we mark that PARTIAL rather than claim otherwise.

08 · ISO/IEC 42001

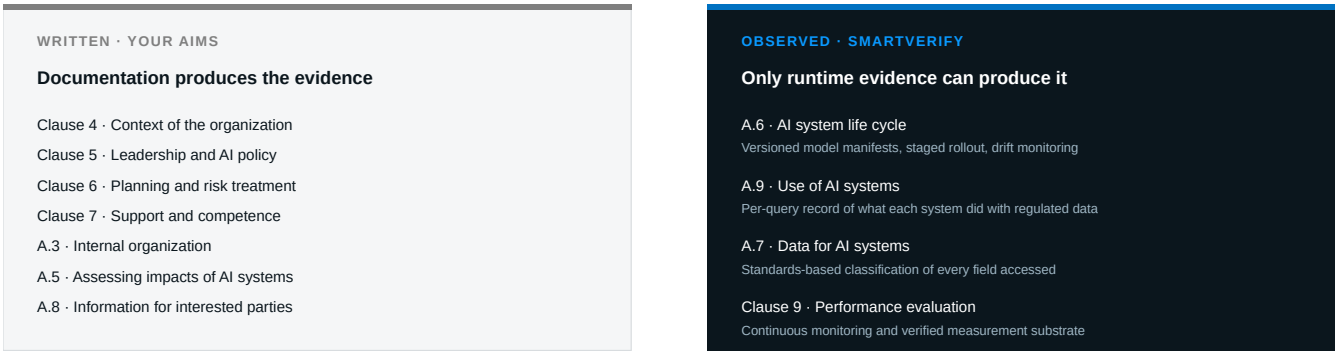
You can write the management system. You cannot write the evidence it claims.

ISO/IEC 42001 is becoming the framework enterprises choose when they want a certifiable AI governance posture rather than a jurisdiction-specific one. It is also the framework where the gap between documentation and demonstration is widest.

An AI management system is largely a documentation exercise, and organisations are generally good at it. You can define scope, appoint owners, write an AI policy, and record a risk treatment plan. What you cannot produce from a document is evidence that your AI systems actually behaved the way the management system says they do. **That evidence is what a certification auditor asks for, and it is the part most organisations cannot answer.**

EXHIBIT 8

Where ISO 42001 needs written policy, and where it needs observed evidence



A.10 · Third-party relationships · SmartVerify's own model governance is inheritable supplier evidence

The questions a certification auditor will ask

- *"Your AIMS says these AI systems access only approved data domains. Show me."*
- *"Demonstrate that model versions in production match your approved manifest."*
- *"Evidence that monitoring for unintended use is operating, not just defined."*

None of these are answered by a policy document. All three are answered by a per-query record showing which system accessed which classified field, under which model and policy version, with the result signed and chained at the moment of observation.

THE SUPPLIER ASSURANCE ANGLE

ISO 42001 asks you to manage AI risk arising from suppliers. SmartVerify runs its own models under a documented eight-stage lifecycle with semantic versioning, per-tenant version manifests, staged rollout, and drift monitoring. **That is vendor-side AI management system evidence you can cite directly in your own assessment**, which most AI vendors in your estate cannot provide.

SmartVerify is not an AI management system and does not replace one. It is the operational evidence layer that makes yours auditable.

09 · CLAUDE ACCESS ATTESTATION

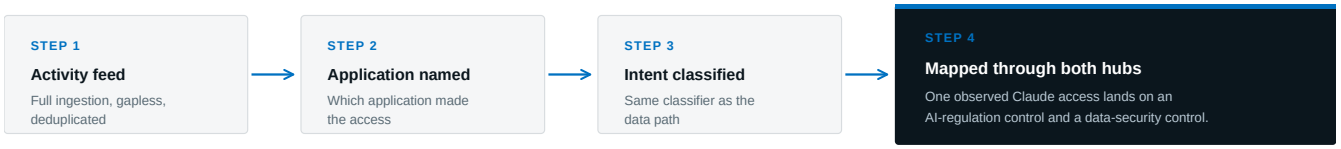
When your auditor asks what Claude accessed last quarter, what will you show them?

Most organisations cannot answer that question today. Claude Enterprise activity is one of the largest unmonitored AI surfaces in the regulated enterprise, and it sits outside every data-path control you have deployed.

SmartVerify ingests Claude Enterprise telemetry through the Claude Compliance API and routes it through the same dual-hub mapping as data-path evidence. Claude becomes a second evidence source, not a second product.

EXHIBIT 9

The Claude attestation chain



What this produces: an application-level attestation of Claude data access, mapped to your frameworks, from an authoritative source

What it does not: name the accountable human behind each access where the application called with its own service credential

The boundary, stated plainly

The attestation names the application behind every Claude access and maps it to your controls. Where an application passes a delegated user token, the human is visible. Where it calls with a service credential, the human was stripped upstream and the record resolves to the application only.

We surface that split rather than hiding it, because the proportion of your Claude activity that resolves only to a service account *is itself the finding*. It is the number your auditor will ask for, and most organisations have never measured it.

6 yrs	<1 min	1
Compliance API retention. Evidence you cannot recreate later	Activity queryable after the event occurs	Mapping framework. Claude reuses the existing hubs

Requires Claude Enterprise with the Compliance API enabled by your organisation.

10 · STACK FIT

Complementary to everything you have. Covering what none of it reaches.

DIMENSION	CLOUD SECURITY / IAM	DATA GOVERNANCE	AI POSTURE	SMARTVERIFY
What it secures	Identity and perimeter	Data at rest	Infrastructure config	**Data in motion, agent to data**
Core question	Who was authorised?	Where does data live?	Is it configured correctly?	**What did the agent actually do, and why?**
Regulatory evidence	Access grants and denials	Data maps, consent records	Posture snapshots	**Per-query trail mapped to AI RMF and 800-53**
Handles novel queries	**No, static rules only**	**No**	**No**	**Yes, evaluated at runtime**
Real-time enforcement	Access gates only	**No**	Alerts, limited blocking	**Yes, inline redact and block**
Tamper-evident records	**No**	**No**	**No**	**Yes, signed and hash-chained**
Covers AI data sprawl	**No**	Partially, known stores	**No**	**Yes, enforced at access**

What SmartVerify does not replace

We are direct about this, because the posture that survives an investigation is the honest one. For each of these we document the obligation and the tooling category that addresses it.

- Algorithmic impact assessments: legal and governance work no data-layer product replaces
- Consumer-facing AI disclosure: an application and product obligation
- Human review and appeal workflows for adverse decisions
- Model weights, training data composition, and explainability analysis
- Data subject rights intake, consent programme management, and DPIAs
- Workforce training records and third-party contractual agreements

WHERE THE GOVERNANCE PLATFORMS FIT

Platforms like OneTrust and Workiva map to regulation as comprehensively as we do, and they usually arrive with the compliance buyer already in place. What they lack is runtime evidence. Their control attestations rest on inventories, assessments, and approval workflows, which is to say on what people recorded rather than on what systems did. That is an evidence-supply problem, and it is precisely what SmartVerify produces. We treat that category as a partner more often than a competitor.

NEXT STEPS

Thirty days to a defensible baseline

SmartVerify deploys in audit-only mode: full visibility, no enforcement, no application code changes. Nothing in your production path changes behaviour. Everything becomes visible.

1 Week one · deploy and observe

A connection-string change places SmartVerify in front of your existing data stores. Works across AWS, Azure, GCP, Snowflake, and on-premises. Sub-50 ms P90 overhead. Your AI applications require no modification.

2 Weeks two and three · behavioural baseline

Per-agent profiles accumulate: which data domains each agent touches, at what volume, with what intent classification. Most organisations discover at least one agent operating well outside what anyone believed its scope was.

3 Week four · the evidence package

A framework-mapped compliance evidence package, graded FULL, PARTIAL, and GAP against your specific obligations, with the gaps documented and routed. This is the artefact you take to your auditor, your board, or your regulator.

THREE QUESTIONS WORTH ASKING YOUR OWN TEAM FIRST

1. If a regulator asked what our AI agents accessed last quarter, what would we produce?
2. What proportion of our AI data access resolves to an accountable identity rather than a shared service account?
3. Our next SOC 2 or HIPAA examination: do our AI agents appear in the audit scope at all, and can we evidence what they did?

SMART VERIFY

See it running against your own data

Request a technical briefing or a proof-of-concept deployment. We will scope it against your actual framework obligations, not a generic demo.

smartverify.ai/contact

Bellevue, Washington

Deployment	Inline proxy, no code changes
Latency	Under 50 ms P90 overhead
Data residency	Raw content never leaves your account
Frameworks	10 mapped end to end
Evidence types	39, each confidence-graded