

# Insurance-grade evidence

*What underwriters actually want from an AI agentic audit trail.*

TNQ RESEARCH PERSPECTIVE · 2026-05-29 · Distribution: public, unrestricted

## **An audit trail built to satisfy your engineers is not the same as one built to satisfy an underwriter.**

The two readers ask different questions. The engineer wants to know whether the system worked. The underwriter wants to know something colder: when it fails, can the loss be attributed, bounded, and proven — and therefore priced. That is a higher bar, and it is quietly becoming the one that matters most, because the insurance market has already begun refusing to cover what it cannot see. This piece is about what insurance-grade evidence actually consists of, why underwriters want each part of it, and how to think about an AI agent's audit trail as the testimony it will one day be asked to give.

## **01 The audit trail has a reader you have not met**

Most audit trails are built by the people who build the system, for the people who operate it. Their purpose is debugging, observability, and — at the more disciplined end — compliance. These are real readers with real needs, and a trail that serves them is already better than most. But there is a third reader, rarely considered at design time, whose question differs from either the engineer's or the auditor's, and whose verdict carries a price tag. That reader is the underwriter, and standing behind the underwriter are the claims examiner, the forensic investigator, and eventually the opposing lawyer.

The engineer asks whether the system worked. The auditor asks whether the controls were designed correctly. The underwriter asks a different question entirely: when something goes wrong, can the resulting loss be attributed to the right party, bounded to a knowable size, and proven to a hostile audience. An audit trail can pass the first two examinations comfortably and fail the third completely — and it is the third that decides whether a loss is covered, what the cover costs, and who ultimately pays for it.

## **02 What the underwriter is actually buying**

It helps to be precise about what underwriting is. An underwriter prices a distribution of possible losses — their frequency and their severity — and accepts that distribution in exchange for a premium. The most expensive thing in that calculation is not risk. It is uncertainty. A well-understood risk, even a large one, can be priced and transferred efficiently. A risk that cannot be seen, sized, or attributed cannot be priced at all, and the only rational responses to an unpriceable risk are to charge punitively for it or to refuse it outright.

This is not hypothetical, and it is not years away. In January 2026 the industry's standard-setting body, the Insurance Services Office, introduced new general-liability exclusions allowing policies to carve out artificial-intelligence exposures — the CG 40 47, CG 40 48 and CG 35 08 endorsements — and carriers began adopting them at once; several major insurers have started trimming AI-related cover across their management- and professional-liability lines. The stated reason is precisely the one above. By one 2026 industry survey, more than forty per cent of insurers track no AI loss data at all, and a risk with no loss history is a risk an underwriter cannot price. When they

cannot price it, they exclude it. AI losses compound the problem by sitting in the worst quadrant for an underwriter: low in frequency but high in severity, the classic shape of a tail risk, where the rare event is the one that matters and the data to anticipate it barely exists.

The implication for anyone deploying AI agents is direct. The underwriter is not, in the end, buying your controls. They are buying their own ability to see and bound those controls. Evidence is the instrument that moves a risk from the unpriceable column into the priceable one — and the quality of that evidence is, quite literally, a term in the premium.

*Underwriters do not buy your controls. They buy their ability to see and bound them.*

03 The properties that make evidence insurance-grade

Not every record is evidence. A log is a convenience for the people who wrote it; evidence is a record that survives an adversary's attempt to discredit it. The gap between the two is where most audit trails fail the underwriter's test. The properties below are what close that gap. None is exotic. Each exists because a specific kind of doubt, raised after a loss when the incentives have turned adversarial, can otherwise defeat a claim.

| Property                  | What it means in practice                                                                        | Why the underwriter wants it                                                                                              |
|---------------------------|--------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------|
| Attributability           | Every action ties to a specific agent, the authority it acted under, and the party it acted for. | Coverage and subrogation both turn on who did what; a loss that cannot be attributed cannot be apportioned or recovered.  |
| Tamper-evidence           | The record is provably unaltered since the moment it was written.                                | A log the insured could have edited carries little weight; opposing counsel will assume it was edited.                    |
| Completeness              | The whole action set is captured, not only the paths that went well.                             | Gaps invite the inference that the missing record is the damaging one. Selective logging reads as concealment.            |
| Contemporaneity           | Records are made at the time of the action, in the ordinary course of operation.                 | Contemporaneous records are trusted; reconstructions assembled for the claim are not.                                     |
| Decision provenance       | The inputs, the policy version, and the authority behind each action are preserved.              | Establishes whether the agent acted inside its authorised envelope — the line between a covered loss and an excluded one. |
| Boundedness               | The record proves the agent could only ever reach what it was scoped to reach.                   | A demonstrable cap on the maximum loss lets the underwriter price a bounded tail instead of an open-ended one.            |
| Reproducibility           | The decision can be replayed from the same inputs and the same policy version.                   | Replay converts a one-off, contested incident into an analysable, defensible event.                                       |
| Independent verifiability | A third party can validate the record without taking the insured's word for it.                  | Self-attested evidence is discounted; independently verifiable evidence is credited.                                      |

The unifying idea is that every one of these properties answers a question someone will ask after the loss, once the incentives have turned adversarial. A record that cannot answer those questions is

not worthless — it may still help the engineers find the bug — but it is not insurance-grade, and the underwriter will price the risk as though the record were not there at all.

## 04 Why agents make the trail the only witness

For most of the history of operational risk, the audit trail was never the whole story. When a human process failed, there were other witnesses: the people involved remembered what happened, the emails survived, the paper existed, and an investigation could reconstruct events from many partial records. The system log was one source among several, and rarely the most important. With an autonomous agent, that redundancy disappears. The agent acts on its own, frequently faster than anyone is watching and at a scale no individual could match; by the time a human notices, the event is already complete. There is no colleague to interview, no recollection to take a statement from. The agent is the only witness to its own actions, and the trail it kept is the only testimony it will ever be able to give. If that testimony is thin, there is nothing behind it.

This is also why the agentic-specific property in the table — decision provenance — matters more than it ever did for conventional systems. The decisive question after an agent-driven loss is rarely what the agent did; the action itself is usually plain to see. The decisive question is whether it was entitled to. Did the agent act within the authority it was granted, under the policy in force at the time, on inputs it was permitted to use. That is the question separating a covered operational loss from an excluded act outside the agent's mandate, and it cannot be answered by a trail that records outcomes alone. The very features that make agents valuable — their autonomy, their scale, their capacity to act on ambiguous inputs — are precisely the ones that make fault hardest to attribute after the fact. The audit trail is where that difficulty is either solved in advance or left to be litigated.

*When an agent fails, it is the only witness to its own actions — and the trail it kept is the only testimony it can give.*

## 05 How to think about it: write for the hostile reader

If there is a single shift in mindset that produces insurance-grade evidence, it is this: design the audit trail for the reader who is trying to use it against you, not the one who is trying to keep the system running. Four consequences follow.

- **The friendly reader is not the one who shows up.** Engineers read logs to fix things, and they read charitably, filling gaps with context they already hold in their heads. The reader who matters after a loss — a claimant's lawyer, a subrogating carrier, a regulator — reads adversarially, treating every gap as concealment and every ambiguity as fault. Build for the second reader and the first is served automatically. The reverse is not true.
- **The audit trail is a financial instrument, not a technical artifact.** It determines whether a loss is covered, how the risk is priced, and who ultimately pays. A record with those consequences deserves the seriousness an organisation gives its financial controls — not the afterthought it usually receives as a byproduct of logging.
- **Evidence is generated, not assembled.** The two properties that carry the most weight after a loss — contemporaneity and completeness — are precisely the two that cannot be added later. A trail assembled in response to a claim is, by definition, neither made at the time nor demonstrably complete. The evidential value of an audit trail is fixed at design time, before anyone knows it will be needed.
- **The control and the proof of the control are different assets.** An organisation can have flawless least-privilege scoping and a perfect human-approval boundary and still lose a claim, if

it cannot prove those controls were in force at the moment of the incident. Underwriters do not credit controls; they credit demonstrable controls. The evidence is the asset they pay for.

## 06 The market is already on this path

None of this is speculative, because the insurance industry has run the experiment once already, in cyber. Cyber cover followed a recognisable arc: first the losses were quietly absorbed, then carved out by silent exclusion, then priced as a bolt-on, and finally underwritten on the basis of demonstrated controls — at which point carriers stopped accepting attestations and began demanding proof. The firm that could show multi-factor authentication actually in force, with evidence, was insurable on good terms. The firm that could only assert it was repriced or refused.

Agentic AI is moving down the same path, and faster, because the playbook already exists and the carriers are more nervous about a risk they understand less well. The exclusions of early 2026 are the silent-carve-out phase. Control-based underwriting is what comes next, and it will arrive sooner than the cyber version did. In an immature market, uncertainty is always priced punitively, because the underwriter has no loss history to lean on and must assume the worst. That is exactly the moment when good evidence is worth the most.

An organisation that can show an underwriter a bounded, attributable, tamper-evident, reproducible account of what its agents can and cannot do is offering something most applicants cannot: a risk the underwriter can actually see. That is the difference between being quoted and being excluded — and, among those who are quoted, a material difference in the number on the page. The audit trail, described in the plainest financial terms, is how an organisation lowers its own cost of transferring agentic risk.

## 07 Build the testimony before you need it

The properties an underwriter wants are not, on inspection, insurance-specific. They are the properties that make any record survive adversarial use — whether the adversary is a claimant's lawyer, a regulator, a court, or an internal investigation. Insurance is simply the most honest forcing function, because it is the one that puts an explicit, recurring price on the quality of an organisation's evidence and charges it every year.

This is, as it happens, the shape of how a Digital Employee is built at TrueNorth Quantum: a distinct and scoped identity for every agent, so each action is attributable; an immutable audit ledger that records actions, decisions, and escalations as they occur, so the record is contemporaneous and tamper-evident; decisions pinned to the policy version in force at the time, so the authorised envelope is provable; a human approval boundary on consequential actions, recorded whenever it is exercised; and a deliberately bounded, reversible action set, so the maximum loss is a known quantity rather than an open question. We did not design these to satisfy underwriters. We designed them to govern the AI workforce — and found, as the insurance market began to specify what it wanted, that a well-governed agent and an insurable one are very nearly the same thing. The underwriter's checklist turns out to be the governance checklist, written in the language of *money*.

***An audit trail is testimony written in advance — by the only witness present when it matters, for a reader who only shows up after the loss.***

The question worth asking of any agent now being deployed is not whether it keeps a record. Almost all of them do. The question is whether that record was written for the friendly reader or the hostile

one — because only one of them ever arrives after something goes wrong, and by then the trail is already whatever it was always going to be.

---

## COLOPHON

POV-2026-002. A TNQ Research Perspective, dated 2026-05-29. Perspectives are point-of-view essays published under TrueNorth Quantum's standing publication cadence; they argue a position and are not customer-specific. Companion documents: POV-2026-001 The Governance Lesson We Already Learned; RPT-2026-001 Governance for the AI Workforce; RPT-2026-002 The Risk Tier Framework; MEM-2026-001 Multi-track Commission Administration. Distribution: public, unrestricted. Comments and corrections: [research@truenorthquantum.com](mailto:research@truenorthquantum.com). © 2026 TrueNorth Quantum Inc.