Vamshi Jandhyala

The record nobody agreed to keep

✠

August 2026

A trace records what was done. An audit records who was entitled to have it done.

On 24 July 2026 the European Union published Regulation (EU) 2026/1744, postponing the record-keeping obligations for the standalone high-risk systems listed in Annex III by sixteen months, to 2 December 2027.[1] Four days later the Model Context Protocol published a revision deprecating its Logging feature, directing implementers to write to `stderr` on stdio transports or to use OpenTelemetry for structured observability.[2]

One decision gave firms more time to produce a record. The other marked for retirement the feature in a widely adopted tool protocol that was most explicitly about emitting one. Neither is a mistake. A transport probably should not be in the record-keeping business, and a regulator that defers an obligation because the standards underneath it are not ready is behaving sensibly. What sits between them is a question the extra sixteen months does not answer by itself: when an agent acts on a client account, what gets written down about the authority under which it acted, and by whom.

The old problem underneath

Systems that let one party act for another have always needed two records rather than one. There is the record of the act, which says what was done and when and to what effect. There is the record of the authority, which says who was entitled to have it done, within what limits, and who answers for it. Mandates, delegated powers, and four-eyes approvals all exist to produce the second record, and the reason they are separate artefacts is that the first record cannot generate the second. Observing that a payment was made tells you nothing about whether anyone was allowed to make it.

The durable point is about where responsibility for the second record sits. Responsibility for an action attaches naturally to whoever performs it. Responsibility for a record does not attach anywhere in particular, which means it has to be assigned, and an unassigned record-keeping duty tends to be delegated outward by every party in turn until it arrives nowhere. Each delegation is reasonable in isolation. The absence is only visible from outside the chain, and usually only when someone asks for the record.

That is the shape of the thing this essay is about. Nothing in what follows is a criticism of any single decision in it.

What a trace contains

Start with what the industry now builds. The OpenTelemetry GenAI semantic conventions define the spans and attributes that instrumentation should emit for agent systems: `invoke_agent`, `chat`, `execute_tool`, alongside attributes for the model requested and the tokens consumed. They moved into a dedicated repository at release v1.42.0 on 12 June 2026 and remain pre-stable, with no 1.0 and names that can still change between versions.[3]

This is a well-designed schema for the question it was built to answer. A trace reconstructs the path an execution took, so that an engineer can find where latency accumulated or where a step threw. Its unit is the operation and its assumed consumer is the team that owns the system.

None of which prevents a firm doing more. Traces can be retained for years, enriched with arbitrary attributes, joined to an identity system, and produced as evidence, and some firms do exactly that. The question is what the schema supplies when nobody has done that work, because the default is what most systems will emit and what most vendors will store.

An audit asks a differently shaped question. Where a trace's unit is the operation, an audit's is the decision, and its consumer is adversarial, or at least independent, and was not in the room. It reaches back years. What it wants to establish is entitlement: who was answerable for this, and on what basis. Adding detail to a trace does not converge on that by itself, because the standard conventions do not define the fields that would settle it.

Where the protocol draws its boundary

MCP's authorization specification is explicit that authorization is optional for implementations. Where it is present on HTTP transports it is OAuth 2.1: the client obtains an access token, the token is audience-bound to a specific server through RFC 8707 resource indicators, and the server validates that the token was issued for it.[4] Access is then governed by scopes.

A scope describes what may be done. It does not describe who delegated the doing, or under what constraint, or which human principal stands behind the agent that presented the token. The published registry of authorization extensions contains two entries, Enterprise-Managed Authorization and a draft Client Credentials extension, and neither addresses delegation.[5] A proposal to support on-behalf-of token exchange for agent-to-agent communication has been open on the specification repository, on the reasoning that RFC 8693 preserves both the user and the agent identity through a delegation.[6] It has not been adopted.

The consequence is structural rather than accidental. In MCP's architecture the host holds the authorisation context, the client issues the tool call, and the server sees a valid token and a request. The server cannot see the context that produced them, which means the record it could write, if it wrote one, would not contain the field that matters. A preprint in May 2026 made this point at the wire level, arguing that MCP carries no authenticated claim about which agent or reasoning episode initiated a call, and proposing token exchange and signed agent-context claims as remedies.[7]

Read as a layering decision this is arguably right. Transports should not be in the accountability business. What the deprecation of Logging does is mark for retirement the one protocol feature explicitly concerned with emitting records, while pointing structured observability at a schema built around operations. The feature itself was never an audit mechanism, and keeping it would not have produced an accountability record. Nor is deprecation removal: it stays in the specification, and under the feature lifecycle policy cannot become eligible for removal until a revision at least twelve months after the one that deprecated it. The change is one of direction rather than capability, and the direction is away from the protocol having any view on what gets written down.

What the regulation actually requires

The postponement makes the text worth reading closely, because sixteen months is long enough to build something and short enough that the design decisions get made this year. Article 12 turns out to be thinner than its reputation. Paragraph 1 requires that high-risk systems technically allow automatic recording of events over the system's lifetime. Paragraph 2 requires that logging enable the recording of events relevant to identifying risk situations under Article 79(1), to post-market monitoring under Article 72, and to the operational monitoring deployers owe under Article 26(5). No fields are named.[8]

Paragraph 3 does name fields, including the identification of the natural persons involved in verifying results. It applies only to systems referred to in point 1(a) of Annex III, which is remote biometric identification. The legislator wrote a human-attribution requirement, scoped it to biometrics, and did not generalise it. For a credit-scoring system, which is high-risk under a different Annex III heading, the standard is traceability appropriate to the intended purpose, and what is appropriate is left to be determined.

Article 26 then closes the loop in a way that matters for anyone in financial services. Deployers must keep the automatically generated logs for a period appropriate to the purpose and at least six months. And deployers that are financial institutions subject to internal-governance requirements under Union financial services law are treated as fulfilling the record-keeping obligation by keeping the records they already keep under that law.[9]

That provision is sensible drafting. It avoids making a bank maintain a parallel logging regime alongside the one its supervisors already inspect. It also means that for the sector where agents touch the most consequential decisions, the AI Act specifies nothing about what to record. It hands the question to financial services record-keeping law, which was written on the assumption that the entity making the decision was a person.

The United Kingdom, with the forcing function removed

British firms face the same question and never had the deadline to begin with. There is no UK equivalent of the AI Act, and the supervisory line has been that firms should govern AI under the rules that already exist instead of waiting for new ones.[10] A postponement in Brussels changes nothing here, which makes the United Kingdom the place where the question is live now.

Those rules are general. SYSC 9.1 requires a firm to arrange for orderly records to be kept of its business and internal organisation, including all services and transactions undertaken, sufficient to enable the FCA to monitor compliance and to ascertain that the firm has met its obligations to clients.[11] It is a standard, not a schema. It was drafted for a world in which the question "who did this" had an obvious answer, and it does not tell a firm which fields to write when the answer is a process holding a delegated token.

The accountability regime is where this bites. The FCA's position is that its existing framework already reaches AI, and that the obligation on senior managers to take reasonable steps to ensure the business for which they are responsible is effectively controlled extends to the safe and responsible use of it.[12] Responsibility stays with the individual who holds the function, and handing execution to a system does not move it.

The regulator has also said, in public, that it does not regard the question as settled. In January 2026 Sheldon Mills put it directly: accountability under the regime still matters, "but what does 'reasonable steps' look like when the model you rely on updates weekly, incorporates components you don't directly control, or behaves differently as soon as new data arrives?"[13]

That is this essay's question asked from the supervisor's side. Reasonable steps have to be evidenced, and the evidence a firm would need is a record of who owned the process, what authority the system was operating under, and what constrained it at the moment it acted. None of that is in a span as the conventions currently define one.

So the position for a UK firm running agents on regulated data is that a named individual carries personal accountability for decisions increasingly executed by software, under a record-keeping rule that predates the software, with no prescribed fields, and with the tool protocol having recently declined to define any.

What the standard schema defines

The obvious response is that this is what the observability conventions are for, since a span is where anyone would look for the record. So read what they actually define.

The OpenTelemetry GenAI conventions cover model invocations, tool executions, agent runs, retrieval and memory operations, with attributes for the model requested and the tokens consumed. OpenInference, the other widely implemented convention, defines exactly two identity attributes: `user.id`, "Unique identifier for a user", and `session.id`, "Unique identifier for a session". It defines nothing for authorisation, permission, scope, delegated authority, on-behalf-of relationships, or the distinction between an actor and a principal.[14]

Those two attributes are what an accountability record would have to lean on, and they cannot bear the weight. Not because any implementation is deficient, but because of what a span attribute is: a value the instrumented application supplies at the moment it emits the span. Nothing in either convention attests it, and neither defines a field that could be attested. An agent that writes the wrong identifier produces a record that is wrong in precisely the way an audit needs it to be right, and there is nothing in the record to reveal the error.

The schema is not failing at its job. Grouping and filtering traces by customer is the job, and for that an application-supplied string is exactly right. It is simply not the same artefact as a record of who was entitled to act.

The firm keeps its own records

The obvious rejoinder from anyone who has worked inside a bank is that none of this is where the audit trail lives. No regulated firm treats an observability vendor as its system of record. It has its own application logs, its own entitlements system, its own case management, and those are what supervisors inspect.

That is correct, and it relocates the problem without dissolving it. What a firm's own systems record about an agent's action is whatever arrived at them, which is typically a service account or a client credential, a request, and a result. The delegation chain that would say whose authority the agent was drawing on is either absent, because nothing minted it, or present in a token at the gateway and dropped before the write. The firm's log is authoritative about what its systems did and silent on who was entitled to have it done.

The silence is easy to miss, because the record looks complete. Every field is populated. Nothing errors. A field that was never in the schema does not show up as missing, and a reconciliation that only checks what is present will never find it. Nor is there enforcement history to learn from, since the European obligation is now more than a year away and the British one has never been written as a schema at all. An absence of consequences so far is weak evidence about a control whose whole purpose is to work in the case that has not happened yet.

The seam

The capability is not missing from the stack. It is in the wrong place, and it has been specified for years.

RFC 8693, OAuth 2.0 Token Exchange, defines precisely the distinction an accountability record needs. A subject token identifies the principal, the party on whose behalf the action is taken. An optional actor token, surfaced as an `act` claim, identifies the party doing the acting, so an authorisation server can distinguish the two and a delegation can be represented without either identity being lost.[15] That is a decade-old standard, and it is the same mechanism the open MCP proposal reaches for.

So an attested delegation can be expressed today. The question is what becomes of the claim afterwards. It exists in a token, at the instant of the exchange, in the layer that performed it. The durable and queryable record, the one that outlives the incident, is a span, whose attributes are whatever the application chose to write. No convention defines a field for the claim to land in, so nothing carries it across and nothing joins them.

That is the actual gap, and it is narrower and more fixable than an absence. It is a missing propagation and a missing field, not a missing capability.

What the record would have to say

An accountability record for a machine-initiated action needs to answer four questions, and the standard trace schema answers none of them by default.

On whose behalf the action was taken, as an attested claim and not an application-supplied string. Under what authority, meaning the specific grant that permitted this action, not the scope set the token happened to carry. Within what constraint, meaning the policy that was evaluated and the decision it returned. And who is answerable, meaning the human or role that would be named if a regulator asked, which is a mapping the firm holds and no vendor can supply.

None of that requires new cryptography or new protocol machinery. Token exchange already carries a subject and an actor. Policy engines already emit decisions. Spans already hold arbitrary attributes and are already retained. The work is deciding that the join is somebody's job, defining the fields, and propagating a claim across a boundary that currently drops it.

Which brings the argument back to where it started. Deciding what an agent may do is a design problem that firms have begun to take seriously, and graded authority, settled action by action before anything runs, is a product decision rather than a setting. Proving afterwards what it did and under whose authority is a different problem, later in time, with a different consumer, and it has quietly been left out of every layer's remit. The protocol says use OpenTelemetry. OpenTelemetry models operations. The regulation says keep the records you already keep. The records a firm already keeps assume a person. The conventions define a field for grouping traces by customer, which is the question they were built to answer.

Every one of those positions is defensible from inside the layer that holds it. The record is missing anyway, and a firm asked to produce it will struggle to explain the chain to someone holding statutory power. Benchmarks tell you whether a system is likely to behave, which is a question asked before deployment, and I have written about why a score does not certify a control. This is the question asked afterwards, by someone who does not have to be fair.

Sixteen months is the interesting part. It is enough time to define the fields, propagate the claim, and have the thing working before anyone demands it, and it is short enough that the firms which treat the postponement as a reprieve will arrive at December 2027 with the same unverified string they have today. The deadline moved. The record still has to exist, and at the moment almost nobody is writing it down.

Sources

[1]Regulation (EU) 2026/1744, the Digital Omnibus on AI, published in the Official Journal on 24 July 2026 and in force from 27 July 2026. It postpones the application of Chapter III, Sections 1 to 3 of Regulation (EU) 2024/1689, which contain the requirements for high-risk systems and the obligations of providers and deployers. Systems classified as high-risk under Article 6(2) and Annex III move from 2 August 2026 to 2 December 2027, a postponement of sixteen months. Those classified under Article 6(1) and Annex I, being AI embedded in regulated products, move from 2 August 2027 to 2 August 2028, a postponement of twelve months. This essay concerns Annex III systems throughout, since that is where financial services classifications such as creditworthiness assessment sit. Articles 12 and 26, relied on throughout this essay, fall within those sections. The European Parliament endorsed the package on 16 June 2026 and the Council gave final approval on 29 June 2026. The Article 50 transparency rules and the Article 4 AI literacy duty were not postponed. See the Digital Omnibus postponement analysis and, for the negotiation history, Gibson Dunn. Secondary sources; the authoritative text is the Regulation itself on EUR-Lex. As of 4 August 2026.

[2]Model Context Protocol, Logging specification. The page carries the notice: "**Deprecated**: The Logging feature is deprecated as of protocol version 2026-07-28" (SEP-2577). Under the feature lifecycle policy it remains in the specification for at least twelve months before becoming eligible for removal; "New implementations **SHOULD NOT** adopt it; existing implementations **SHOULD** migrate to logging to `stderr` for stdio transports, or to

OpenTelemetry for structured observability." Roots and Sampling were deprecated in the same revision. See also the 2026-07-28 changelog. As of 4 August 2026.

[3]OpenTelemetry GenAI semantic conventions. The conventions define spans including `invoke_agent`, `chat` and `execute_tool`, and attributes including `gen_ai.request.model` and `gen_ai.usage.input_tokens`, covering model invocations, tool executions, agent runs, retrieval and memory operations. All `gen_ai.*` attributes and spans moved out of the main semantic-conventions repository into a dedicated GenAI conventions repository at release v1.42.0 on 12 June 2026. This was not a stabilisation: the conventions remain pre-stable and experimental, with no 1.0 release, and names can still change between versions. See OpenTelemetry's GenAI observability post. As of 4 August 2026.

[4]Model Context Protocol, Authorization specification. "Authorization is **OPTIONAL** for MCP implementations." Where supported on HTTP transports it follows OAuth 2.1, with servers acting as resource servers, mandatory Protected Resource Metadata (RFC 9728) for authorization-server discovery, and mandatory Resource Indicators (RFC 8707) so that tokens are audience-bound: "MCP servers **MUST** validate that access tokens were issued specifically for them as the intended audience." Access is governed through OAuth scopes. As of 4 August 2026.

[5]MCP Authorization Extensions registry. At the time of writing the repository lists two extensions: Enterprise-Managed Authorization (stable) and Client Credentials (draft). Neither covers on-behalf-of flows, delegated authority, actor identity, or RFC 8693 token exchange. As of 4 August 2026.

[6]Issue #214, modelcontextprotocol/modelcontextprotocol, "Support On-Behalf-Of Token Exchange protocol for Agent-to-Agent Communications", proposing RFC 8693 on the reasoning that passing user access tokens between agents is insecure and that token exchange supports delegation while retaining both user and agent identity. Open, not adopted into the specification. As of 4 August 2026.

[7]Maya Solen, "Closing the MCP Identity Gap: From OAuth Tokens to Signed Reasoning Episodes", clawrXiv, 22 May 2026 (v3). The paper argues that no surveyed platform carries authenticated claims about which agent, model version, or reasoning episode initiated a tool call, and proposes five graduated extensions, including RFC 8693 token exchange for verifiable delegation chains and a signed `agent_context` JWT claim. Worth reading and worth weighing carefully: clawrXiv is a preprint repository for papers published by AI agents, and the work is not peer reviewed. As of 4 August 2026.

[8]Article 12, EU Artificial Intelligence Act. Paragraph 1: "High-risk AI systems shall technically allow for the automatic recording of events (logs) over the lifetime of the system." Paragraph 2 requires logging capabilities to enable recording of events relevant to Article 79(1) risk situations, Article 72 post-market monitoring, and Article 26(5) operational monitoring. Paragraph 3 sets minimum fields, including "(d) the identification of the natural persons involved in the verification of the results, as referred to in Article 14(5)", but applies only "For high-risk AI systems referred to in point 1 (a), of Annex III", which is remote biometric identification. Article 12 sits in Chapter III, Section 2, whose application was postponed by Regulation (EU) 2026/1744. Readers should be aware that several widely used AI Act reference sites, including the implementation timeline at artificialintelligenceact.eu, still showed the original 2 August 2026 date at the time of writing and did not mention the amendment. As of 4 August 2026.

[9]Article 26, EU Artificial Intelligence Act. Paragraph 6 requires deployers to keep automatically generated logs "for a period appropriate to the intended purpose of the high-risk AI system, of at least six months", to the extent the logs are under their control and subject to applicable data protection law, and provides that deployers which are financial institutions subject to requirements regarding internal governance under Union financial services law fulfil the obligation by keeping records under that law. Paragraph 5 sets the monitoring duty referred to in Article 12(2)(c). Article 26 sits in Chapter III, Section 3, whose application was likewise postponed. As of 4 August 2026.

[10]The United Kingdom has no dedicated AI statute comparable to the EU AI Act, and the

supervisory expectation has been that firms govern AI under existing rules. See Kennedys, "Deploying AI in financial services in the UK: FCA and data protection considerations", 2026. Law-firm commentary rather than a supervisory publication; treated here as commentary. As of 4 August 2026.

[11]FCA Handbook, SYSC 9.1, general rules on record-keeping: a firm must arrange for orderly records to be kept of its business and internal organisation, including all services and transactions undertaken by it, sufficient to enable the FCA to monitor the firm's compliance with the requirements under the regulatory system and in particular to ascertain that the firm has complied with all obligations with respect to clients. As of 4 August 2026.

[12]The FCA's AI Update (April 2024) sets out how the existing regulatory framework maps to the UK government's principles for AI regulation, and states that the Senior Managers and Certification Regime already applies to the use of AI: the obligation on senior managers to take reasonable steps to ensure that the business for which they are responsible is effectively controlled covers the safe and responsible use of it. For the regime's application in practice, and the expectation that senior managers be aware of AI use within their area and be able to show the steps they took, see DLA Piper, "FCA and PRA AI Update: Senior Managers' accountability for the use of AI", April 2024. As of 4 August 2026.

[13]Sheldon Mills, FCA, "The FCA's long term review into AI and retail financial services: designing for the unknown", speech, 28 January 2026: "Accountability under the Senior Managers and Certification Regime (SM&CR) still matters – but what does 'reasonable steps' look like when the model you rely on updates weekly, incorporates components you don't directly control, or behaves differently as soon as new data arrives?" As of 4 August 2026.

[14]OpenInference semantic conventions. Identity attributes are `user.id` ("Unique identifier for a user") and `session.id` ("Unique identifier for a session"). The specification defines no attribute for authorisation, permission, scope, delegated authority, on-behalf-of relationships, or the actor-versus-principal distinction. As of 4 August 2026.

[15]RFC 8693, OAuth 2.0 Token Exchange (IETF, January 2020). The specification defines a `subject_token` representing "the party on behalf of whom the request is being made" and an optional `actor_token` representing "the identity of the acting party". Where a delegation is represented, the issued token carries an `act` claim identifying the actor while the `sub` claim continues to identify the principal. As of 4 August 2026.