Decision Accounting

A Framework for AI-Agent Decision Accountability in the Enterprise

Five hundred years ago, double-entry bookkeeping made capital legible: every movement of money recorded twice, every balance provable, every steward answerable. The enterprises now delegating decisions to autonomous AI systems have no equivalent discipline for cognition. This paper proposes one — a vendor-neutral framework in which every consequential action taken by an AI agent is recorded, expressed in the language of the business, authorized by an explicit mandate, and closed against its outcome. We call the discipline decision accounting. This is a framework and a vocabulary, offered so that practitioners, standards bodies and regulators can describe the same problem in the same terms. It is deliberately not an implementation specification: it does not prescribe data schemas, interfaces or conformance tests, and an organization could not build to it without further work. That specification is the necessary next step, and it should be written by more hands than these.

Dev Gandhi · Actee · September 2026 · v1.4

Revised following review by senior wholesale-banking, risk and board practitioners. See Notes on Revisions at the end.

1. The Accountability Gap

Every large enterprise now operates two workforces. The first is human: hired, trained, supervised, appraised, and answerable through a management chain refined over a century of corporate governance. The second is synthetic: AI agents that read, decide, write, transact, and coordinate — deployed in months, multiplying in weeks, and governed, in most organizations, through infrastructure controls designed for software rather than for actors exercising judgement.

The asymmetry is not an absence of control. It is a mismatch of vocabulary. For the human workforce, an enterprise can answer four questions in business terms: who acted, under whose authority, on what information, and with what result. For the agent workforce, most enterprises can answer versions of the first and third in system terms — which service identity called which API, against which records — and cannot answer the second and fourth at all. The bank knows which agent touched which system. It does not know what the agent decided about a client, whether that decision was within the authority anyone intended to grant, or what it cost.

The prevalence data supports this reading rather than a story of neglect. In a 2026 survey of 1,900 IT leaders, only 12% said they can actually govern the agents they deploy — in organizations that, almost without exception, run mature identity and entitlement stacks.[1] The U.S. National Institute of Standards and Technology has framed the specific defect: agents are commonly treated as generic service accounts, without dedicated identity, authorization, or accountability controls. The problem is not that no one is watching. It is that what is being watched is not denominated in anything a board, an auditor, or a supervisor can act on.

And the regulatory perimeter has just moved in a way that makes the gap official. In April 2026, U.S. banking regulators replaced the fifteen-year-old SR 11-7 model risk guidance with SR 26-2 — and explicitly excluded generative and agentic AI from its formal scope. The supervisory message is unambiguous: the model-risk toolkit was not built for this, and institutions must govern agentic systems through their broader risk and control environment. At this moment, that broader environment has no standard instrument for the job. Singapore's Model AI Governance Framework for Agentic AI (January 2026) — the first of its kind — points at the answer, requiring verifiable agent identity and an audit trail of who acted under whose authorization. NIST's planned AI Agent Interoperability Profile and its critical-infrastructure profile will point at it again. What is missing is the operating model, and that is what this paper proposes.

2. Why Model-Centric Governance Cannot Close the Accountability Gap

The instinct of every risk function confronting AI agents is to reach for model risk management: inventory the models, validate them, monitor drift, document limitations. That instinct fails for a structural reason, not a procedural one. Model risk management governs an artifact at rest. Agentic risk lives in actions in flight.

A validated model can still take an unauthorized action, because authorization is a property of the action's context, not of the model's weights. The same foundation model, on the same day, may draft a routine email, reprice a product, and initiate a payment — three acts with three different materiality levels, three different approval requirements, and three different evidentiary obligations, none of which is visible at the level of the model. Validation frequency compounds the mismatch: models are

revalidated in cycles measured in months, while agent fleets act thousands of times an hour and change behaviour with every prompt revision, tool grant, and context update.

The unit of governance must therefore shift, from the model to the decision. This is not a rejection of model risk management — it is its completion for the agentic era. Banks did not stop auditing employees when they started validating models; they will not stop validating models when they start accounting for decisions.

2.1 You cannot rely on preventive controls alone — but you must start with them

A necessary correction to how this problem is usually framed. Institutions with mature control environments do not govern their people by instruction alone, and they should not govern agents that way either. A trader who must not deal in a product or beyond a size is not asked to refrain; the capability is removed at the system boundary. Material non-public information is not protected by asking the uncleared not to read it; it is protected by entitlements. The same logic applies, without qualification, to agents. Every agent should carry a distinct identity rather than a shared service account, and should inherit the entitlement model that governs an equivalent human role. Where an action can be prevented, prevent it. A ledger is not a substitute for a permission that should have been withheld.

Preventive controls, however, have three boundaries that no amount of entitlement engineering removes, and it is precisely in those three gaps that accountability is lost.

- **They constrain the possible, not the chosen —** Entitlements define the space of permitted actions. Within that space, an agent still exercises judgement — which client to prioritize, which remediation to trigger, which price to offer, which exception to grant. Almost all consequential agent behaviour is permitted behaviour. Permissions say nothing about it.
- **They are silent on reason —** Supervision of people has always been limited by the impossibility of recording intent — the meeting without minutes, the judgement never articulated. Agents invert this constraint. For the first time, the reasoning behind a delegated decision can be captured contemporaneously, at the moment of action, at negligible marginal cost. This is not merely a compliance artifact; it is the input that lets the next mandate be written better than the last.
- **They do not connect to outcome —** An entitlement model can confirm that an action was allowed. It cannot say whether the action was any good. Nothing in the permission stack joins what an agent did to what subsequently happened to revenue, loss, conduct, or client outcome — which is the question boards and supervisors ultimately ask.

> The relationship, stated plainly. Permissions constrain what an agent can do. Mandates express what it is authorized to decide, in business terms, under a named owner. The ledger records what it chose within that authority and why, and closes that choice against its outcome. The first is infrastructure and must come first. The second and third are accounting, and no institution yet has them.

3. The Central Idea: Decision Accounting

> *Double-entry bookkeeping made capital legible. Decision accounting makes cognition legible.*

When Luca Pacioli codified double-entry bookkeeping in 1494, the innovation was not arithmetic. It was accountability architecture: every economic event recorded as a transaction, every transaction touching two accounts, every book reconcilable and every custodian of capital thereby answerable. That architecture scaled from Venetian merchants to global capital markets essentially unchanged, because it rested on principles rather than technologies. Financial accounting eventually acquired shared standards — generally accepted principles — because capital that crossed firm boundaries needed a common language of trust.

Enterprises are now custodians of a second scarce resource: delegated decision-making authority. They hand it to synthetic actors at scale, across vendors, without a ledger. Decision accounting applies the Pacioli move to cognition.

> One term, defined once. A Decision Event is the record of a consequential agent action — the entry, not the act. As in double-entry bookkeeping the transaction is the event and the journal entry is the record, here the agent's action is the act and the Decision Event is what the ledger holds: who acted, under which mandate, on what basis, what it touched, what it was worth, who approved it, and what followed.

Seven principles follow.

I. **I. Conservation of accountability —** Delegation transfers execution, never accountability. Every autonomous action must trace to a mandate owned by a named human. There is no orphan authority in the enterprise.

II. **II. Every consequential action is accountable; not every action is an alert —** Any agent action with business consequence — a query, a write, a transaction, a communication, an adopted recommendation — is recorded as a Decision Event in an append-only, tamper-evident ledger. Recording is not the same as reviewing. Attention is scarce and must be spent proportionately: breaches and attempted out-of-mandate actions demand immediate scrutiny, while in-mandate activity is reviewed on a cycle in aggregate. A record that generates no attention still discharges its purpose, because it is the substrate for review, attribution and evidence.

III. **III. Meaning over telemetry —** Raw telemetry is not accountability. Each Decision Event must be reconciled to the enterprise's own semantics — the entities, metrics, and data lineage it touches — so that a board can read the ledger in the language of the business, not the language of the stack.

IV. **IV. Mandate precedes action, and permission precedes mandate —** An agent's decision rights are expressed as a machine-evaluable Mandate: permitted actions, scope, thresholds, escalation rules, required approvers. Mandates sit above the entitlement model and never substitute for it — where an action can be prevented technically, it should be. Mandates govern the exercise of judgement inside what permissions already allow. They are evaluated before action where interception is possible, and immediately after where it is not, but they are always evaluated.

V. **V. Materiality is denominated in business terms —** Anomalies and breaches are scored in the enterprise's own units of consequence — revenue, margin, customers, regulatory exposure — not in tokens, traces, or latencies. We borrow the term deliberately from financial reporting, where

information is material if omitting or misstating it could reasonably influence the decisions of those relying on it, and we retain that logic while changing its object: the question is not whether a figure would influence a reader of the accounts, but whether an agent's action would influence a metric the enterprise manages. As in financial reporting, thresholds are set by the organization against its own scale, and qualitative factors — regulatory exposure, contractual commitment, reputational sensitivity — can render an action material at magnitudes a purely quantitative threshold would pass over. Concentration is one such factor: an exposure that is immaterial for a single agent may be material once it is understood that a hundred agents share the same dependency.

VI. **VI. Decisions close against outcomes —** Every Decision Event is eventually joined to its outcome. A delegation whose results are never measured is not governed; it is merely tolerated. The outcome loop is how the enterprise learns which authorities to widen and which to withdraw.

VII. **VII. Evidence is verifiable by construction —** The ledger must support export of cryptographically verifiable evidence — to auditors, regulators, and cross-sector bodies — without exposing the underlying data. Trust between institutions is built on proofs, not promises.

Taken together, these principles describe what the industry does not yet have: a common discipline by which any enterprise, in any sector, on any vendor stack, can state with evidence that it is in control of its AI operations. Whether such a discipline ever becomes generally accepted is not ours to declare. Generally accepted accounting principles were not proclaimed by their authors; they were conferred over decades by a profession, tested by crisis, and ratified through standard-setting institutions with the authority to bind. We offer these principles in that spirit — as a candidate framework put before the professions that would have to adopt, amend, or reject it, and improved considerably if they do the second.

3.1 Proportionality: what supervision actually looks like

The most common objection to a decision ledger is that recording everything is disproportionate. The objection is right about attention and wrong about records, and the distinction matters enough to state as an operating model. Wholesale banking already solved this problem for human risk-takers, and the pattern transfers directly.

Trigger	Response	Cadence
Attempted action outside mandate	Immediate scrutiny, with the attempted action, the mandate clause breached, the materiality score, and the agent's stated reason presented together to the mandate owner	Real time
Action inside mandate	No alert. Aggregated into periodic review of the agent class against its risk and outcome profile, in the same way a desk is reviewed on its risk and P&L	Daily to monthly, by materiality
Pattern shift within mandate	Flagged for review where behaviour diverges materially from the agent class baseline, without asserting a breach	Cycle-based
Everything else	Recorded, not surfaced. Available for attribution, audit and evidence; not competing for executive attention	On demand

Two questions sit behind this model, and an institution should be able to answer both: did things operate as designed, and separately, was the design as intended? The first is supervision and is answered by the exception path. The second is governance and is answered by the outcome loop — the accumulated evidence that a mandate, faithfully executed, produced the results the institution wanted. Most agent governance discussion addresses only the first. The second is where the value of an accounting discipline actually lies.

This proportionality also disposes of the objection that a decision ledger is a second data estate. It is not. The ledger stores references, mandates, verdicts, reasons and outcome links — not copies of the records an agent touched.

4. Where Your Organization Stands

Organizations do not adopt accountability architecture in one step. Five levels serve as a self-assessment and, eventually, as a cross-sector benchmark. The levels describe the vocabulary in which an institution can account for its agents, not the presence or absence of controls: a firm with excellent entitlements and no business-meaning reconciliation sits at L1, which is not an accusation of negligence but a statement about what it can currently evidence.

Level	Name	The organization can say
L0	Unobserved	"Agents act; we cannot reconstruct what they did." Rare in regulated institutions, common where agents have been adopted at the edge of the estate.
L1	Logged	"Vendor and platform logs exist, in someone's format, in system vocabulary. We can trace calls; we cannot read decisions."
L2	Reconciled	"We know what our agents did, in the language of our business — clients, products, exposures, metrics."
L3	Governed	"Every agent acts under an explicit mandate above its entitlements; breaches surface immediately with materiality attached, and in-mandate activity is reviewed on a cycle."
L4	Attested	"We can evidence to our board, our auditors and our regulator what our agents decided, under whose authority, and with what consequence."

Most enterprises today sit between L1 and L2 — well controlled at the infrastructure layer, unable to render agent activity in business terms. Supervisory expectations, insurance underwriting, and coalition membership criteria will, we expect, converge on L3 as the professional floor within two to three years.

One clarification about the top of the ladder is worth stating plainly, because the distinction is easy to lose. L4 is an evidentiary claim, not a reliability one. It does not assert that agents will not fail or deviate; failure modes are specific to each agent and to the context it operates in, and no attestation can generalize across them. What L4 asserts is narrower and achievable: that the organization can show what its agents decided, under whose authority, and with what consequence. This is the same logic as a financial attestation, which never claims that no error occurred, only that the records are complete and the controls operated as described.

5. What It Looks Like in Practice

Two worked examples, from different sectors, with the same shape.

5.1 A bank: the discount nobody approved

Gross margin in one region lands 180 basis points below forecast. The cause is roughly 4,000 orders shipped over nine days carrying discounts averaging 22%, applied by an agent deployed to speed up quote turnaround. The agent held legitimate access to the CRM. Every action fell inside its permissions. Its operational dashboard was green throughout: latency healthy, error rate near zero, throughput up 40%. Nothing malfunctioned. What the institution could not produce was the record that mattered: that these discounts moved margin by 180 basis points, that their cumulative effect crossed a threshold someone should have approved, and that no human ever did. Under decision accounting, the per-action limit is irrelevant — it is the aggregate mandate ceiling, expressed in basis points of regional margin, that surfaces the problem on day three.

5.2 A consumer goods manufacturer: the immaterial variance that wasn't

A European manufacturer deploys agents in shared services to resolve supplier invoice discrepancies below a threshold. Each resolution is individually trivial and individually permitted. Over a quarter, accepted price variances accumulate into a seven-figure supplier overpayment concentrated in one category, and a definition of "immaterial" that nobody in finance set becomes, in effect, company policy. The agent behaved exactly as designed. The failure is that the design was never expressed in terms anyone in finance could review, and the accumulation was never scored against a metric anyone owned. This is the case that makes the point most clearly: the discipline is not a financial-services discipline. It applies wherever delegated authority is exercised at a speed and volume that outruns human review.

5.3 A risk the ledger alone does not catch: concentration

Decision accounting records and evaluates decisions one at a time. That is the right unit, and it leaves one class of exposure invisible without a deliberate addition. If every agent across an enterprise depends on the same foundation model, the same vendor runtime, or the same retrieval configuration, then a single model revision, deprecation, outage or behavioural shift propagates through every delegated decision simultaneously. Individually each agent remains inside its mandate. Collectively the institution has a correlated failure across its entire agent workforce.

This is the agentic analogue of concentration risk in a loan book, and it is not hypothetical: the operational-risk literature has begun to treat AI dependency as a distinct exposure, on the principle that accountability cannot be outsourced to a supplier. Two consequences follow for this framework. The fleet registry at Layer 1 must record each agent's model and runtime dependencies, not merely its owner and entitlements, so that the question "how much of our decision-making rests on one supplier?" is answerable at all. And materiality, under Principle V, must treat concentration as a qualitative factor — an exposure that is immaterial for one agent can be material once a hundred agents share its dependency.

We flag this as an addition prompted by review rather than as a solved problem. A single institution can measure its own concentration. Whether an entire sector has converged on the same few runtimes is a

question no institution can answer alone, and it is one of the strongest arguments for the federated exchange described in the companion paper.

6. The Framework: Five Layers

The principles resolve into a layered architecture. Each layer is independently implementable and independently auditable; together they form the accountability perimeter around an enterprise's agent fleet. Layers 1 and 2 are largely infrastructure, and many institutions have parts of them already. Layers 3 to 5 are where the discipline is new.

Layer	Function	Core artifact
1. Identity	Every agent enrolled with owner, vendor, cost centre, entitlement profile, model and runtime dependencies, and mandate binding; agent activity outside the register detected and surfaced (no shadow agents)	Fleet registry entry
2. Capture	Activity records from different agent platforms — which arrive in different formats and different vocabularies — normalized into one common event structure held in an append-only, tamper-evident store	Decision Event; Decision Ledger
3. Meaning	Events reconciled to the enterprise's own entities, metrics, lineage and organizational structure, with human correction of low-confidence mappings that compounds accuracy over time	Reconciled event graph
4. Control	Mandates compiled from policy and evaluated at runtime above the entitlement model; pre-action gates where runtimes permit interception, post-action flags with materiality scoring where they do not; approval workflows with tracked human decisions	Mandate; approval record
5. Evidence	Verifiable exports mapped to external frameworks — regulator templates, auditor evidence packs and board attestations	Evidence pack; attestation

Two design constraints keep the framework deployable in regulated environments. First, the ledger stores references and aggregates, not copies of raw enterprise data — accountability without a second data estate. Second, events are content-addressed and hash-chained, so any auditor can verify integrity without trusting the operator.

7. Alignment with the Emerging Framework Landscape

Decision accounting is not a competing framework. It is the operational layer the current instruments presuppose but do not specify:

Framework / instrument	Where decision accounting connects
SR 26-2 (U.S. interagency model risk guidance, Apr 2026)	Explicitly excludes generative and agentic AI; institutions must govern agents through broader controls. The Decision Ledger is that broader control, purpose-built.
CRI Financial Services AI RMF (Feb 2026, 230 control objectives, Treasury-announced)	Ledger events, mandates and evidence packs supply the "effective evidence" the Control Objective Reference Guide asks for — continuously generated rather than assembled at examination.
COSO, Internal Control over Generative AI (Feb 2026)	Observes that organizations can control outputs but cannot evidence the decision path. Decision accounting is the architecture that discharges that expectation.
NIST AI RMF and forthcoming agent profiles	Supplies the identity, authorization and accountability controls NIST has flagged as missing where agents run as generic service accounts.
EU AI Act, Arts. 12, 14, 26	Automatic logging, effective human oversight and deployer obligations discharged through ledger records, mandate gates and approval workflows.
ISO/IEC 42001	The ledger is the management system's operational evidence base for AI operations.
Singapore IMDA Agentic AI framework (Jan 2026)	Verifiable agent identity and "who acted under whose authorization" audit trails are Layers 1, 2 and 4 of this framework.

8. Related Work and Intellectual Lineage

No serious discipline is born from nothing, and decision accounting is a synthesis with named ancestors. The academic root is the concept of decision provenance, proposed by Singh, Cobbe and Norval in 2018 and published in IEEE Access in 2019: the use of provenance methods to expose decision pipelines — the chains of inputs to, the nature of, and the flow-on effects from decisions taken within algorithmic systems — in service of oversight, audit, compliance and risk mitigation.[2] That work anticipated, at the level of research vision, what this paper renders operational: the decision record as a first-class artifact, and the joining of decisions to their downstream effects.

The internal-control profession has been converging on the same ground from its own direction. COSO's February 2026 guidance moves control expectations upstream into the AI-influenced decision layer, observing that organizations can often control final outputs while remaining unable to evidence the decision path — inputs, model output, human review, exceptions and changes — that produced them, and calling for contemporaneous evidence, traceability and accountability as elements of control design rather than documentation assembled after the fact.

Practitioner convergence through 2025–2026 is itself evidence of the gap: independent essays and early products have proposed agent decision ledgers, decision-design frameworks with explicit decision boundaries and escalation thresholds, hash-chained and cryptographically signed agent action records, and governed execution layers for agents acting on systems of record. Each addresses a fragment — a record here, a boundary there, an integrity mechanism elsewhere. The operational-risk advisory community has arrived independently at the same first principle, publishing on AI dependency under

the proposition that accountability cannot be outsourced — which is Principle I of this framework in five words, reached by a different route. We also note, for completeness, that the phrase “decision accounting” has an older and different life in the management-accounting literature, where it denotes accounting information supplied in support of management decisions; the present usage inverts that direction. It should also be distinguished from what the trade press now calls “agentic accounting” — AI agents performing accounting work. This paper concerns the inverse: accounting for what agents decide, across the enterprise, wherever delegated authority is exercised.

Against that landscape, this paper's contribution is the synthesis and its naming: a single discipline — with conservation of accountability as its first principle, materiality denominated in the enterprise's own units, mandates distinguished from technical permissions, outcomes closed against decisions, and a maturity ladder suitable for attestation — rather than a further fragment. We claim the frame, not the ingredients; the ingredients are the field's, and the citations above are offered so that adopters may trace them.

9. A Call to the Cross-Sector Community

Accounting became trustworthy when it became shared. We propose three near-term acts of convergence, each achievable within the coming year. First, a common Decision Event schema — an open, minimal record format any vendor or enterprise can emit, so that ledgers interoperate the way financial statements do. Second, a shared mandate vocabulary — a cross-sector taxonomy of agent decision rights, thresholds, and escalation classes, adoptable by coalition working groups. Third, a maturity attestation convention — a simple, auditable statement of an organization's level on the ladder, suitable for boards, insurers, counterparties, and coalition membership.

None of this requires waiting for regulation, and none of it requires any particular product. It requires what every accountability revolution has required: a small number of serious institutions deciding that the ledger is worth keeping. The enterprises that adopted double-entry bookkeeping first did not do so because it was mandated. They did so because merchants who could prove their books were trusted with more capital. The enterprises that adopt decision accounting first will be trusted with more autonomy — by their boards, their regulators, their customers, and each other. In the agentic era, that trust is the scarcest asset there is.

> A note on what this paper deliberately leaves out. The most consequential property of decision accounting emerges when it crosses firm boundaries — anonymized failure signatures contributed to and drawn from a sector pool, so that one institution's incident becomes every member's early warning. That subject deserves its own treatment, and two reviewers of earlier drafts said so. It is now a separate paper: Federated Decision Exchange.

Read it here: [Federated Decision Exchange — Sharing What Agents Get Wrong, Without Sharing What They Touched]

Notes and Sources

[1] OutSystems survey of 1,900 IT leaders, 2026: 12% of respondents reported that they can effectively govern the AI agents they deploy.

[2] J. Singh, J. Cobbe and C. Norval, “Decision Provenance: Harnessing Data Flow for Accountable Systems,” IEEE Access 7 (2019) 6562–6574. See also J. A. Kroll et al., “Accountable Algorithms,” University of Pennsylvania Law Review 165 (2017).

Further sources: interagency supervisory guidance SR 26-2 and OCC Bulletin 2026-13; Cyber Risk Institute Financial Services AI Risk Management Framework (February 2026); COSO, Achieving Effective Internal Control Over Generative AI (February 2026); NIST AI Risk Management Framework and related agent-standards work; EU Regulation 2024/1689; ISO/IEC 42001:2023; Singapore IMDA Model AI Governance Framework for Agentic AI (January 2026).

Notes on Revisions

This version incorporates review by senior practitioners in wholesale banking market infrastructure, bank data and AI leadership, and board risk governance. The principal changes since v1.0: preventive controls are given explicit precedence over recording, and mandates are positioned above rather than instead of the entitlement model; supervision is expressed as an exception-based operating model rather than universal review; the top of the maturity ladder is stated as an evidentiary rather than a reliability claim; the paper's purpose and its deliberate limits are stated at the outset; terminology has been tightened, with “framework” used throughout in place of “reference model” to avoid collision with model risk vocabulary; the Decision Event is defined once, as the record rather than the act; a non-financial worked example has been added; the federated exchange material has been removed to a separate paper; and concentration of model and runtime dependency has been added as an exposure the ledger alone does not surface. We are grateful to those reviewers, who improved the work by disagreeing with it.

Comments are welcome: info@actee.ai