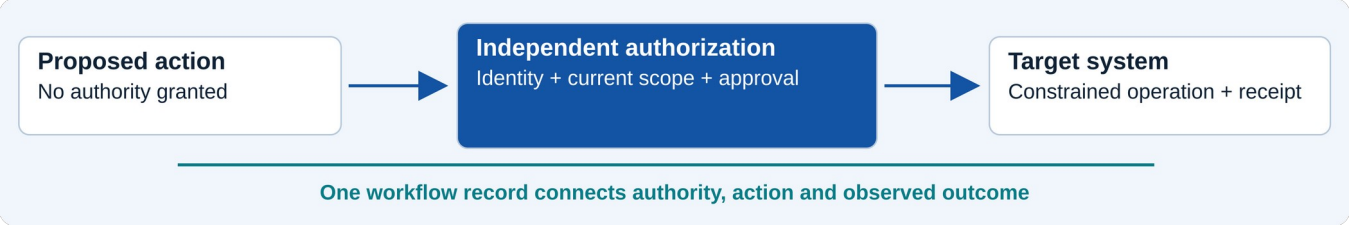
Governing enterprise agent workflows

Authorization and evidence across multi agent systems

Anwesh Rath, Founder and CTO, AI Standards Inc.



Conceptual action path. Authorization is enforced independently of model reasoning.

Executive summary

Enterprise agents need an enforceable boundary between a proposed action and its effect on the business. That boundary must survive delegation, changing permissions and uncertain outcomes. We recommend independent authorization at each tool call, narrowly scoped identities, action-specific approval where required, and records that connect the authority granted to the result observed.

For CTOs, CISOs and platform teams, the practical unit of governance is the complete transaction. This paper sets out the controls and acceptance tests needed to move from a promising demonstration to a workflow whose permissions, failures and recovery can be inspected.

The problem and why it matters now

Connecting a model to purchasing, customer records or infrastructure gives its output consequences beyond an answer. Even a read can disclose confidential information through a query or retrieved result. The system must establish whether this user and workload may perform this operation on this resource. A plausible explanation of intent cannot establish that authority.

Delegation makes the problem harder. A research agent may pass supplier details to a purchasing agent that has broader credentials. A malicious invoice can then influence a later write, even when the original request was only to compare vendors. Shared memory and queues can carry instructions beyond the task that introduced them. NIST’s Generative AI Profile identifies information security and value-chain integration risks [2]. Here, those risks become concrete paths through identities, tools and retained context.

Regulatory and standards context

NIST AI RMF 1.0 is voluntary guidance [1]. NIST SP 800-207A supports identity-based application authorization [3]; applying it to agents requires additional design choices. The EU AI Act follows a risk-based approach for specific uses and distinguishes providers from deployers [4]. Assess the intended use, actor role, jurisdiction and applicable obligations before selecting controls. A technical control framework supports that assessment but does not establish legal conformity or replace sector-specific requirements.

Design the execution boundary

Let the model propose a structured action. A separate service validates the operation, parameters, target and data classification against current policy before a narrowly privileged adapter executes it. Keep credentials outside model context. Retrieved documents and tool outputs cannot grant permissions. Apply the same boundary to sensitive reads, outbound data transfers and alternative tools that could produce the forbidden effect.

Define each transaction and its completion condition. Reading supplier records, drafting a requisition and committing it require different permissions. An agent allowed to read a document must not gain write access merely because its next step needs it. Escalate missing authority to the workflow owner. Include information sent in search queries, prompts and logs in the data-flow assessment.

Delegation and approval controls

Bind each run to its initiating user, workload identity, parent task and purpose. A child receives only the intersection of the user's rights, the parent's remaining authority, its own role and the tool's scope. Set expiry, delegation depth, fan-out and resource budgets. Check authorization when queued work executes, using current rights and task status.

Bind human approval to the action parameters, destination, approving person and expiry. Include relevant business state, such as the requisition version. A supplier substitution or amount change invalidates approval. Separate requester and approver where the consequence warrants it. Show the actual proposed effect so the reviewer can make an informed decision.

Control mapping

These are proposed acceptance requirements. Standards anchors show selected alignment rather than a complete crosswalk.

Requirement and source	Enforcement control	Evidence and acceptance test
Bound delegation [3]	Issue distinct, expiring identities and narrower child scopes.	Inspect the identity chain; reject a child requesting a parent-only operation.
Define oversight [1] GOVERN 3.2	Bind approval to action parameters and current business state.	Change the approved destination; require rejection and fresh approval.
Handle uncertainty	Use supported idempotency and reconcile target-system state.	Test a timeout after commit; confirm one transaction or an explicit reconciliation hold.
Enable disengagement [1] MANAGE 2.4	Revoke authority and stop queues and dispatch.	Revoke a parent; block queued dispatch and reconcile in-flight operations.
Evaluate and monitor [1] MEASURE 2.3 and 2.4	Test deployment-like conditions and monitor production behavior.	Inspect monitoring records; repeat representative and adversarial tests after changes.

Evidence and failure handling

Record the identity chain, policy version, model and tool versions, approved parameters, authorization result and target-system receipt under one workflow identifier. Preserve source references so investigators can distinguish retrieved material from generated content. Limit sensitive payloads and control access and retention. Tamper evidence can reveal alteration; it does not establish that an action was correct.

A timeout after dispatch may mean the target committed the action. Reuse a stable operation identifier and the target's idempotency facility where available. Query the target before retrying. If the outcome remains uncertain, hold the operation for reconciliation. When authorization is unavailable, block new consequential writes. Record completed effects, unresolved effects and any compensating action that needs separate authority.

Check permissions again before commit. Where supported, use a conditional write that requires the approved record version to match. Otherwise use a lock respected by every writer or another target-enforced safeguard. If conflicting changes cannot be prevented, retain manual control and document the remaining risk. Test revocation across queues and adapters, measure its enforcement delay, and reconcile operations already in flight.

Implementation roadmap

The timing below is illustrative; release depends on evidence. Days 1 to 15: the workflow owner and security lead inventory tools, identities, data flows and consequential writes. Draw the delegation graph, including queues. Choose a bounded pilot and define prohibited actions, fallback, stop authority and release criteria.

Days 16 to 45: the platform team implements authorization and approval binding. Test malicious document instructions, forged delegation, stale approvals, duplicate dispatch and policy-service failure. Include combinations across agents. A component can pass its own test while the complete workflow bypasses a restriction through another component.

Days 46 to 90: operations rehearses revocation and reconciliation, then reviews representative runs with the process owner. Report accepted task completions, unauthorized effects, false denials, review effort and containment time, with denominators and test conditions. Define thresholds before evaluation. Expand permissions only when remaining risks have a named owner and an explicit release decision.

Evaluate usefulness and control effectiveness separately. Fast completion does not excuse an approval bypass; blocking every request is not an operational success. Test allowed, prohibited and ambiguous requests. Retain inputs and configuration so later changes can be compared with the baseline. Investigate unexplained failures before expanding authority.

Illustrative purchasing scenario

This is an invented tabletop example, not customer evidence. A sourcing agent reads approved supplier records and drafts a requisition. A purchasing agent can submit it after a human approves the supplier, amount and destination. Neither agent can modify bank details or release payment.

First, an invoice asks the sourcing agent to replace the bank account. The expected control result is a denied unsupported operation, with the source reference retained. Second, the purchasing call times out. The operator queries the requisition system and reconciles the receipt rather than submitting again. Third, the owner revokes the parent task while a child remains queued. The child must fail authorization at dispatch. Inspect all three records together: permission enforcement, transaction integrity and revocation are separate properties.

Risks and limitations

A compromised administrator, enforcement service or adapter can defeat the boundary. Reviewers can approve harmful actions, and correlated agents can repeat the same mistake. On-premises hosting does not remove these risks. Independently test the enforcement path and repeat evaluation after material changes to models, permissions, tools or workflow purpose. Retain a serviceable manual process for cases the system cannot resolve.

Conclusion and next action

Choose one transaction for an architecture review with the workflow owner, platform team and security lead. Bring its tool inventory, approval rules, sample execution records and recovery procedure. Agree which permissions can be enabled, which tests must pass and who owns the decision to expand.



Conceptual failure handling. An uncertain result is reconciled before a retry.

AI Standards research and engineering approach

AI Standards approaches enterprise orchestration as an engineering problem of authority and evidence. Our documented work connects scoped access, governed inference and audit lineage [5]. Related research is described in filed provisional patent applications addressing orchestration, agent authorization and step-level provenance [6]. These are provisional applications, not granted patents.

The contribution is a connected view of the workflow: who authorized a step, what the agent was permitted to do, which information informed it and what happened at the target system. For an enterprise buyer, this provides a practical basis for reviewing an agent system across service boundaries and agreeing the evidence required for release.

For a proposed deployment, establish which paths enforce permissions, which actions require human approval and which events can be reconstructed. Test failures as well as successful runs. Our August 2026 architecture documentation provides a design reference; current commissioning and evaluation evidence must establish readiness for the selected workflow. Agree who controls data and administrative access, who can stop the system and who maintains its records.

References

Public sources checked 5 November 2026. Private AIS sources are identified separately.

[1] NIST AI 100-1, AI RMF 1.0, January 2023; GOVERN 3.2, MEASURE 2.3 and 2.4, MANAGE 2.4. https://doi.org/10.6028/NIST.AI.100-1

[2] NIST AI 600-1, Generative AI Profile, July 2024; sections 2 and 3. https://doi.org/10.6028/NIST.AI.600-1

[3] NIST SP 800-207A, application access control, September 2023. https://doi.org/10.6028/NIST.SP.800-207A

[4] European Commission, AI Act overview, updated 19 October 2026. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[5] AI Standards Inc., engineering architecture documentation, snapshot 25 August 2026. Private engineering source.

[6] AI Standards Inc., provisional patent application filing records. Private source; no application identifiers or implementation details reproduced.

About the author Anwesh Rath, Founder and CTO, AI Standards Inc.

Disclaimer Informational, not legal advice. Assess the intended use, jurisdiction and sector with qualified advisers. Controls and examples are recommendations unless expressly identified otherwise. Patent filings do not establish product performance, compliance or certification.