

Delegating Work to AI

Authority, evidence, and oversight

Artificial intelligence can make a workflow move faster while leaving its hardest decisions unresolved. A useful delegation therefore begins with the work: what must happen, what evidence makes the result defensible, and who remains accountable when the evidence is incomplete.

My work on Sentinel examines that operating problem. The aim is to discover how work gets done, show what should change and why, and measure whether the change actually helped. This guide translates that approach into decisions a leader can make before giving an AI agent a larger role.

A practical starting point

Choose one recurring task with a clear beginning and end. Describe the human workflow first, including review, exceptions, rework, and waiting time. Then identify the limited contribution an agent could make without taking over decisions that require accountable human judgment.

The first deliverable is a delegation record. It names the permitted task, the evidence boundary, the accountable owner, the required reviewer, and the conditions that stop the work. The next deliverable is a comparison with the current process using the same acceptance standard.

Before expanding authority, ask: Can a qualified person reconstruct what happened, explain why the disposition is supported, and stop the workflow when those conditions fail?

What this guide establishes

The design and worked example are proposals for controlled evaluation. HarborBank is fictional; the case and evidence labels are synthetic. This guide reports no production deployment, customer outcome, measured savings, or validated operating threshold. An illustrative outcome is not an experimental result.

Juan A. Martinez Díaz, MBA · September 22, 2026 · Public edition 1.0. Adapted from the author's applied work during MIT Sloan Executive Education's Implementing Agentic AI program. The analysis and examples are the author's; this is not an MIT publication or endorsement.

Write the delegation before running the agent.

A prompt can describe a boundary. The surrounding system must enforce it. For the proposed Sentinel workflow, the agent prepares an evidence package within an approved case. A separate authorization layer determines which tools and actions are available.

Decision	Proposed boundary
Purpose	Organize evidence and draft gaps for a risk and control self-assessment. Success means a reviewer can reach a supported disposition.
Read access	Only the approved synthetic case, its control requirements, and explicitly permitted evidence sources.
Permitted output	A source-linked evidence index, discrepancy list, and draft review package. Drafts remain distinguishable from approved records.
Human-reserved decisions	Materiality, risk acceptance, control effectiveness, final challenge, case disposition, and production promotion.
Enforcement	A case-scoped nonhuman identity, least-privilege tools, and deterministic authorization outside the model's prompt.
Stop conditions	Missing authorization, unavailable event logging, unexpected access, conflicting policy, or unavailable review coverage.
Recovery	Hold the case, preserve the evidence and action record, revoke access if needed, and return work to the accountable human owner.

Assign people to the exceptions

The workflow owner defines the task and its acceptance conditions. A qualified reviewer challenges the evidence and owns the disposition. Engineering and security owners implement access controls, logging, and recovery. An independent assessor checks whether the evaluation supports the proposed next step. Record actual names before a trial begins.

Specify the review window and a backup reviewer. If neither can respond, the case stays on hold. A full queue or an unanswered escalation does not create permission to proceed. Narrow intake when oversight capacity cannot support the workload.

Reference basis: NIST AI Risk Management Framework 1.0, GOVERN 2.1 and 3.2 and MAP 3.5, addresses responsibilities and human oversight. The delegation record above is the author's proposed application, not a NIST certification checklist. [1]

A missing approval is a finding to resolve.

Synthetic case HB-014 concerns HarborBank’s quarterly privileged-access review. The local test requirement asks for a current population record, completed reviewer decisions, and evidence that exceptions were resolved or formally accepted by an authorized person. These are invented case requirements, not a statement of banking regulation.

Evidence received	What it supports—and what it does not
E-01 · Current quarter access roster	Identifies the submitted population. By itself, it does not establish that every access decision was reviewed.
E-02 · Reviewer decision log	Records review decisions but leaves two exceptions open. It provides no closure or authorized acceptance for those items.
E-03 · Prior quarter sign-off	Provides historical context. It does not authorize closure of the current quarter’s exceptions.

The proposed sequence

1. Intake. The system admits the three documents to HB-014, records their versions, and restricts the agent to that case. Instructions inside a document are treated as document content; they cannot change tool permissions or approval requirements.
2. Draft. The agent links the two unresolved exceptions in E-02 to the missing current-period evidence. Its draft says that the submitted package does not support closure. It requests resolution records or authorized acceptance, with citations to the relevant entries.
3. Human review. The designated reviewer checks the population, the cited entries, and the applicable requirement. In this illustrative outcome, the reviewer accepts the missing-evidence finding and records the disposition as return for evidence. The agent cannot convert that disposition into a control-effectiveness rating.
4. Hold and resume. The case remains open. When new evidence arrives, it receives a new version and another review. The old draft and the reasons for the earlier return remain visible in the record.

A defensible return for missing evidence can be a successful review outcome. An unsupported closure cannot. Neither outcome should be confused with proof that the control is effective.

This walkthrough illustrates intended behavior. It has not been presented as an executed test or a result from a real financial institution. The evidence labels identify fictional documents created for this example.

Measure the decision and the work behind it.

Faster drafting has limited value if it creates more review work or hides unresolved cases. The following measures are proposed for Sentinel evaluation. They are not validated industry standards, and their thresholds must be set for the task before testing.

Time to Defensible Decision

Measure elapsed time from evidence submission until a qualified reviewer accepts a supportable decision package. Include waiting, human review, and rework. Record active reviewer effort separately. Show the ages of open cases alongside completed-case times so unresolved work cannot disappear from the comparison.

Evidence-to-Decision Integrity Rate

Divide reviewed dispositions that pass every integrity condition by all reviewed dispositions. A pass requires evidence that is current, traceable, and complete for the stated disposition; action within approved authority; and a recorded disposition with its accountable reviewer. Report the raw numerator and denominator.

In HB-014, a correctly supported return for evidence could pass because the finding establishes why closure is not supported. Calling the control effective without the required evidence would fail. Unreviewed cases remain visible as a separate count and age profile; they are not treated as successful dispositions.

Authority breaches and blocked attempts

Count actions actually executed outside approved permissions. The proposed target is zero. Record blocked attempts separately; blocking is a control outcome, not evidence that an unauthorized action executed. If the system cannot produce the authorization record, pause and investigate rather than assuming compliance.

What the review packet must show

Always include	Why the leader needs it
Raw counts, material misses, and unresolved cases	A favorable aggregate can conceal a consequential failure or a growing queue.
Human baseline and reviewer effort	The comparison must account for the total operating burden.
Model, tools, rules, and case-set versions	The conclusion applies to a defined configuration and scope.
Disagreements and limitations	A reader must be able to distinguish a supported conclusion from an unresolved judgment.

Reference basis: the U.S. Government Accountability Office's framework connects governance, data, performance, and monitoring and supports independent assessment. These specific measures are the author's proposals. [2]

Make the test capable of changing the decision.

Start with a frozen task definition and acceptance criteria. Compare the existing human process with the proposed assisted process using equivalent cases and the same quality standard. Have qualified reviewers examine the evidence; independent challenge should be separate from front-line development.

A proposed three-week trial

Stage	Work and decision evidence
Week 1 · Define and prepare	Confirm owners, reviewer availability, data scope, permissions, stop conditions, and recovery. Build synthetic cases and record the human baseline.
Week 2 · Develop and challenge	Use 20 development cases to examine the bounded workflow, its failure paths, and the event record. Freeze the candidate configuration before final evaluation.
Week 3 · Evaluate and decide	Use 20 previously withheld cases. Preserve misses, disagreements, and unresolved work. Decide to proceed within scope, narrow, redesign, or stop.

The 40-case design is an initial proposal for learning, not a statistically representative validation. If a withheld case is used to tune the system, it becomes development evidence; replace the final evaluation set before claiming an independent check.

Include the conditions that are easy to overlook

Test stale evidence, incomplete populations, contradictory records, unsupported approvals, and instructions embedded in retrieved documents. Challenge attempts to cross case boundaries, unavailable logging, expired authorization, exhausted runtime limits, and a reviewer who does not respond.

The workbook proposes initial machine limits of 20 tool calls, two read retries, and five minutes of processing per case. Treat these as trial settings. Record why they fit the task, what happens when they are reached, and whether they need revision. A timeout ends machine execution; it does not erase the case or the human work still required.

Demonstrate stopping, access revocation, evidence preservation, and return to the human process. A recovery procedure that has never been exercised remains an assumption.

Reference basis: NIST AI Risk Management Framework 1.0, MEASURE 1.3, 2.1, and 2.5 and MANAGE 2.4, supports independent assessment, documented testing, stated limits, and intervention. The case count, schedule, and runtime settings are this author's proposed design. [1]

Expand one task at a time.

The release decision should identify the specific task, configuration, and operating conditions that the evidence supports. A successful intake test does not authorize autonomous risk acceptance, control ratings, or final case closure.

Decision	Meaning for the next operating period
Proceed within scope	Retain the tested boundary, review coverage, monitoring, and recovery path. Name the owner and the next evidence review date.
Narrow	Remove unsupported actions, case types, sources, or volume. Keep only work that the team can govern and review.
Redesign	Change the workflow or controls, then evaluate the changed configuration before expanding authority.
Stop	Suspend the affected workflow, preserve evidence, and route remaining work to the accountable human process.

Record the continuing obligations

Keep a versioned record of the model, prompts, tools, parsers, evidence collections, permission rules, and approval logic. Assign responsibility for reviewing changes. Reserve real reviewer time and a backup path; if that capacity disappears, reduce intake or narrow the task.

The leader's decision record should state the evidence considered, unresolved limits, authorized scope, accountable owners, stop conditions, and next review date. This makes delegation a continuing operating responsibility rather than a one-time approval.

Sources and provenance

[1] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0), 2023. Core: GOVERN, MAP, MEASURE, and MANAGE. [Open source](#)

[2] U.S. Government Accountability Office. Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities. GAO-21-519SP, June 30, 2021. [Open source](#)

Source material: Juan A. Martinez Diaz, Implementing Agentic AI — Sentinel Organizational Playbook, Modules 1–3, September 2026. This public adaptation removes course prompts and submission scaffolding. Its delegation design, proposed measures, and synthetic walkthrough are the author's applied work. External frameworks provide reference points; they do not establish that Sentinel satisfies a standard.

Continue reading: research, published experiments, and the M.A.R.T.I.N.E.Z. Method. [Open source](#)

Juan A. Martinez Diaz, MBA · Former Wells Fargo Vice President · Retired U.S. Army Sergeant Major · Independent work; views are my own. Contact: sgmmartinez@gmail.com