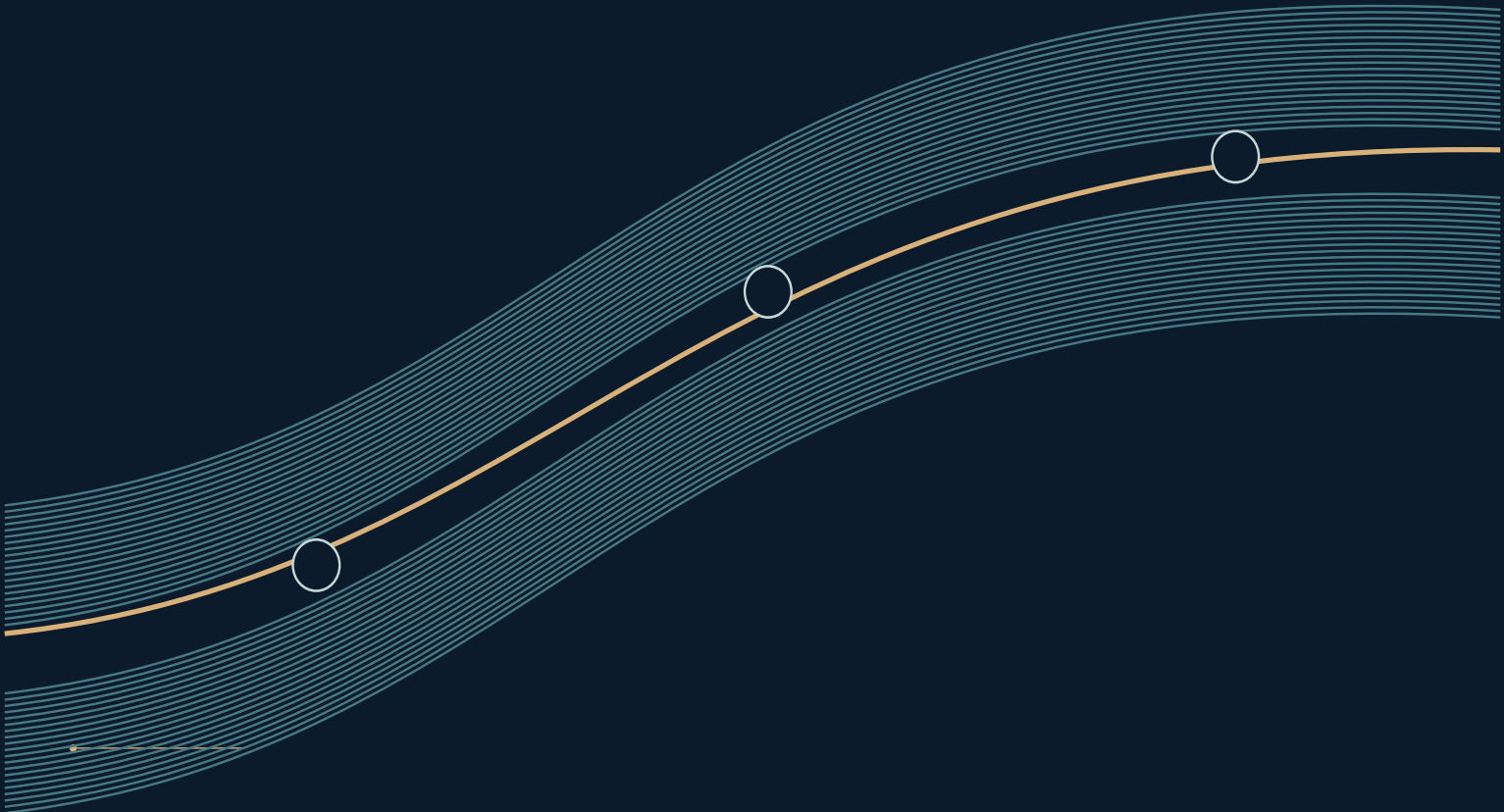


GOVERNANCE / OCTOBER 2026

Governance is not a brake

Design the conditions under which teams can move without asking again



THE DECISION IN BRIEF

Governance can support faster delivery when teams receive clear operating permissions and consequential changes trigger focused challenge.

- 01

Approve a defined operating boundary: purpose, users, evidence, permitted actions and stop conditions. Repeated work inside that boundary should not need the same decision again.
- 02

Scale controls with consequence, reach, uncertainty and reversibility. Reuse proven platform controls, then test the decisions and failure modes specific to each service.
- 03

Prove the benefit with a portfolio trial. Compare approval effort and time to useful release alongside detected failures, residual exposure and the ability to revoke permission.

THE MAYDEN VIEW

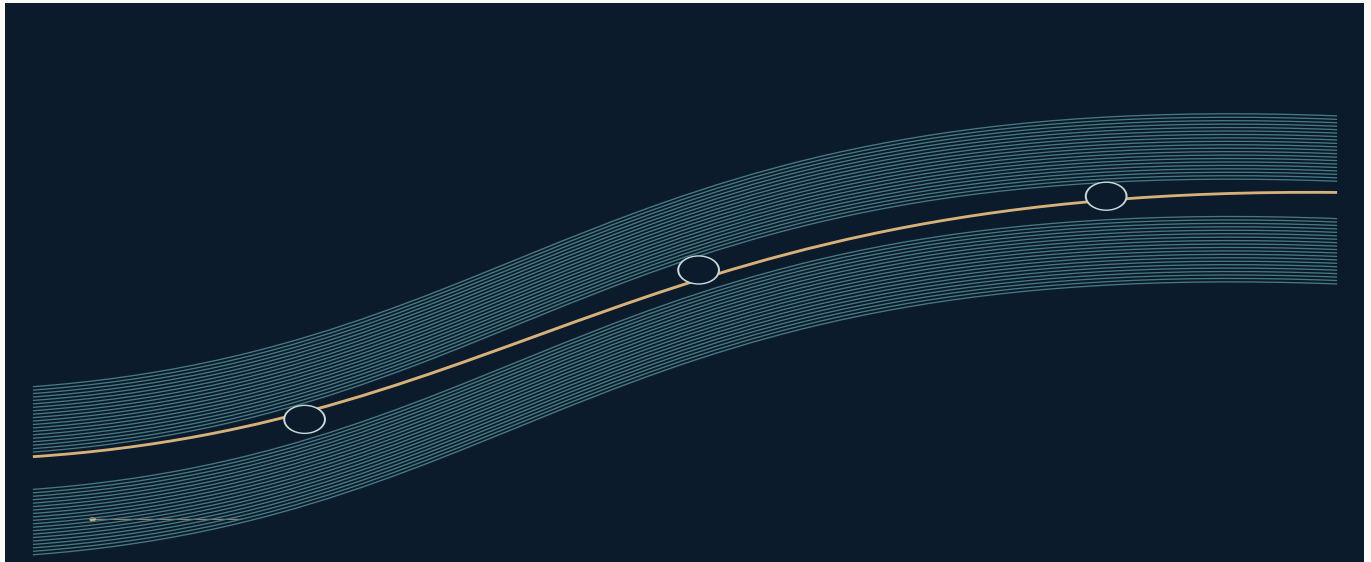
A useful approval tells a team what it can do next - and what would require a new decision.

INSIDE THIS PERSPECTIVE

The argument at a glance	3
01 / Define and calibrate reusable permission	4
Exhibit 01 / A standing permission should say when work can proceed	5
02 / Test shared controls and human oversight	6-7
Exhibit 02 / Four questions determine the intensity of control	8
03 / Make change and expiry operational	9
Exhibit 03 / Human-AI combinations did not outperform the best standalone condition on average	10
04 / Compare delivery friction with control quality	11
The leadership agenda	12
Method and references	13

ONE DECISION, FOUR CONNECTED QUESTIONS

Make responsible progress possible



GOVERNANCE RAILS / CONCEPTUAL ARTWORK

01 Define and calibrate reusable permission

Approve the boundary that lets work proceed

02 Test shared controls and human oversight

Assign owners to shared controls and their dependencies

03 Make change and expiry operational

Treat exceptions and changes as the centre of governance

04 Compare delivery friction with control quality

Make a permission expire when its assumptions expire

A useful approval tells a team what it can do next - and what would require a new decision.

CHAPTER 01

01

Define and calibrate reusable permission

Approve the boundary that lets work proceed

Consider two illustrative teams. One team waits weeks to approve another internal drafting assistant. Another adds external messaging to an approved service without a fresh decision. Both patterns reveal the same weakness: approval is attached to a project name rather than a defined operating boundary. A better design makes routine work easier to repeat and consequential changes harder to miss.

mayden proposes an operating permission stating the service's purpose, users, information, action rights, exposure limits and monitoring. Teams may improve the service inside that boundary through an agreed change process. Crossing it returns the decision to its named owner. The permission creates a practical delegation; its value must be tested in operation.

The NIST framework provides a voluntary risk-management foundation rather than a prescribed approval structure. ^[1] Our proposal is an organizational interpretation, not a claim that a framework itself produces faster delivery. It should be tested against an enterprise's existing controls and observed bottlenecks. The strategic objective is to spend scarce expert attention where it can change a consequential decision, while eliminating duplicate work that adds little assurance.

Calibrate controls to consequence, reach and reversibility

Classify the consequence before classifying the technology. A reviewed draft and an agent changing a customer's entitlement may use the same model but require different controls. Assess consequence, reach, uncertainty and reversibility together, recording the reasoning rather than compressing it into an unexplained score.

Consequence asks who or what could be harmed. Reach asks how many decisions can be affected before detection. Uncertainty asks how far the use extends beyond credible evidence. Reversibility asks whether an error can be corrected in time, including effects already experienced by customers or staff. A reversible database update may still cause an irreversible disclosure. A human approval step may be ineffective if the reviewer lacks information or is expected to process more cases than can be examined meaningfully.

mayden's proposed control matrix distinguishes bounded assistance, supervised execution and consequential autonomy. These are management categories, not legal classifications. A company can prohibit a particular autonomous action without banning all AI in that workflow. It can also approve a narrow release with transaction limits and a clear path to expanding them. The important discipline is to state what evidence would justify additional authority. Scale then becomes an earned operating decision, not the automatic next stage after a successful demonstration.

ILLUSTRATIVE DECISION TOOL

A standing permission should say when work can proceed

Fictional standing permission; no production approval

Permission field	Illustrative policy assistant	Condition that changes permission
Purpose and owner	Service owner permits internal policy drafts	New audience or consequential use requires review
Allowed action	Retrieve approved sources and draft; no external send or record change	A commit request exceeds this permission
Population and evidence	Named internal users, current corpus and tested case set	New data class or source withdrawal triggers reassessment
Validity	Recorded expiry and named revocation owner	Expiry, revocation or unresolved control failure blocks use
Monitoring and recovery	Track corrections and escalations; retain manual service	An uncontained failure restricts the affected route

THE MANAGEMENT IMPLICATION

Teams may repeat work inside the recorded boundary; changes to that boundary need a new decision.

mayden design example, not a legal risk class or validated standard. Owners set workload-specific conditions and dates.
Sources: [1], [2], [7]

CHAPTER 02

02

Test shared controls and human oversight

Assign owners to shared controls and their dependencies

Centralize controls that projects genuinely share: identity, data access, supplier checks, logging, model routing and incident handling. Make their operating evidence reusable. Business teams should inherit working services and clear obligations, rather than rebuild each requirement from a policy document.

Reuse has a boundary. An access-control mechanism can be common while the decision about who should have access remains use-specific. A central evaluation harness can be common while the cases and acceptance criteria remain local. The NCSC's lifecycle security guidance supports treating secure deployment and maintenance as part of delivery.

^[2] It does not imply that a secure platform guarantees correct business judgment. The organization needs evidence for both the shared control and its application in context.

The systems perspective matters because model quality is only one source of failure. The research paper Hidden Technical Debt in Machine Learning Systems identifies dependencies, feedback loops and changing environments as sources of long-term maintenance burden. ^[3] mayden's inference is to assign owners to those surrounding dependencies before launch. A model owner cannot be expected to resolve a changed policy document, an upstream data contract and a customer-remedy process alone. The permission boundary must name the people responsible for keeping its assumptions true.

CHAPTER 02

02

Test shared controls and human oversight

Human oversight must earn its place in the design

Putting a person in the workflow can reduce risk, but it can also create ceremonial approval. The relevant question is whether the person improves the outcome under realistic conditions. Can the reviewer see the source material, understand the limits, challenge the recommendation and stop the action? Are they evaluated only on speed? If rejecting an AI suggestion requires several extra steps while accepting it takes one click, the interface is shaping behavior regardless of the policy's language.

A study by Buçinca and colleagues found that cognitive forcing functions reduced overreliance in the studied decision task, while introducing user-experience trade-offs. ^[4] A separate research programme found that explanations could reduce overreliance under some conditions, with effects influenced by the costs and benefits of engaging with them. ^[5] These results do not establish a universal interface rule. They challenge the assumption that adding an explanation or a confirmation button is sufficient evidence of effective oversight.

A preregistered meta-analysis of 106 experiments and 370 effect sizes gives two different answers because its comparators differ. Against the better of human or AI alone, the combined condition performed worse on average: Hedges' $g = -0.23$, 95% confidence interval -0.39 to -0.07 . Against humans alone, it performed better: $g = 0.64$, interval 0.53 to 0.74 . ^[8] These standardized effects are not percentages. The studies, published January 2020–June 2023, varied widely in task and design. The findings support testing relevant alternatives; they do not show that every human-AI combination is harmful or that every reviewer improves a workflow.

Governance should therefore compare three operating designs where lawful and appropriate: a person working alone, an automated service, and their combination. Measure decision quality, review load and the consequences of error on the same cases. Human involvement may be required for reasons beyond average accuracy. Where it is, the task is to improve how the person intervenes rather than count the presence of a reviewer as evidence that the control works.

ORIGINAL MAYDEN FRAMEWORK

Four questions determine the intensity of control

mayden operating-permission framework

Consequence	Question What harm can the action cause?	Implication Focus on affected people and assets
Reach	Question How far can an error spread before detection?	Implication Limit exposure and shared dependencies
Uncertainty	Question Where does credible evidence stop?	Implication Narrow scope or acquire more evidence
Reversibility	Question Can effects be corrected in time?	Implication Retain human authority where recovery is weak
THE MANAGEMENT IMPLICATION		
Proportionality is a reasoned decision, not simply a lighter questionnaire.		

No numerical risk score is asserted. Sources: [3], [6], [7]

CHAPTER 03

Make change and expiry operational

03

Treat exceptions and changes as the centre of governance

A permission boundary is useful only if the organization can detect when it no longer holds. Changes in purpose, users, model behavior, data, tools or legal context can invalidate earlier assumptions. mayden recommends a small set of material-change triggers agreed at release. Examples include access to a new sensitive record, a new action permission, a broader customer population or an untested language. Routine cosmetic changes should not create the same burden as a new ability to transfer money.

The organization also needs a shared language for events. The OECD distinguishes actual AI harms from hazards that could produce harm. ^[6] This is useful for internal learning: a blocked attempt to exceed authority deserves investigation even when no customer was harmed. Reporting only realized losses creates a late and incomplete picture. Conversely, counting every unusual output as a severe incident can overwhelm operators and undermine trust in the process.

Design an exception route with ownership, deadlines and a decision record. An exception should say what condition is temporarily unmet, what compensating control applies and when the permission expires. It should not become an indefinite waiver hidden in email. Repeated exceptions signal a design problem: either the control is poorly calibrated or the product is not ready for the authority it seeks. Review that pattern at portfolio level, where management can address shared causes rather than asking individual teams to work around them.

Measure control quality and delivery friction together

Measure governance as an operating service. Establish a baseline for review time, missing-evidence delays, duplicate requests, post-release rework and material changes detected before deployment. These are proposed measures, not industry benchmarks. A shorter approval cycle counts as progress only while release decisions and failure detection remain credible.

A practical review can separate three delays. Some arise because a team has not supplied essential evidence. Some arise because reviewers lack capacity. Others arise because several functions ask for the same evidence in incompatible formats. The remedies differ: improve preparation, allocate expert capacity or simplify the process. Calling all delay bureaucracy hides legitimate risk work; calling all delay necessary assurance hides poor service design. The CEO should expect the governance function to make this distinction transparently.

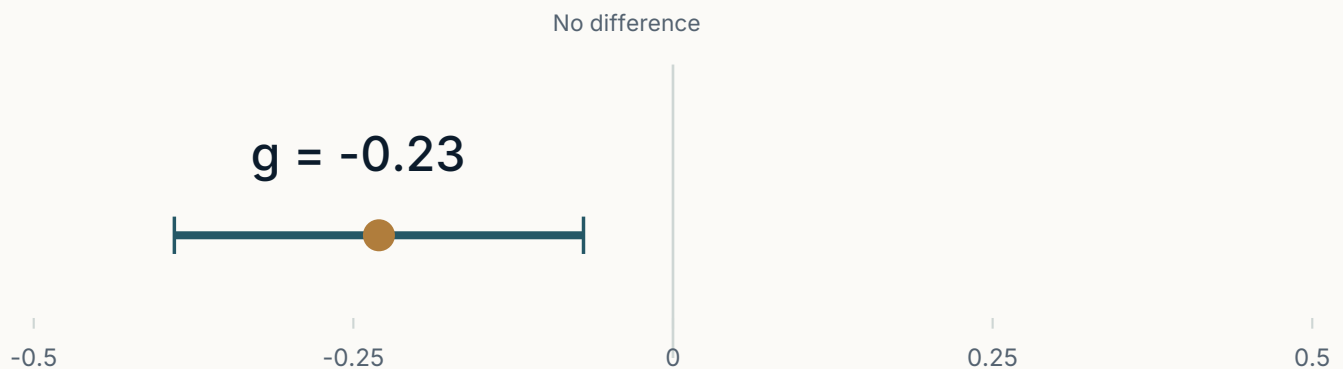
DSIT's assurance guidance describes a range of techniques for evaluating and communicating confidence. ^[7] mayden's extension is to attach each requested artefact to a decision it supports. If no reviewer can explain how an artefact changes approval, scope or monitoring, question its value. Preserve independent challenge where incentives could suppress bad news. Remove repeated requests and ambiguous requirements while preserving the expertise and authority needed to stop an unsupported release.

PUBLISHED META-ANALYSIS · STANDARDIZED EFFECT AND 95% CONFIDENCE INTERVAL

Human-AI combinations did not outperform the best standalone condition on average

Hedges' g

Human-AI combination versus the better of human or AI alone



95% confidence interval: -0.39 to -0.07

Human-AI combination performs worse

Human-AI combination performs better

THE MANAGEMENT IMPLICATION

Test the combined workflow against relevant alternatives; the presence of a reviewer does not establish a performance gain.

Vaccaro, Almaatouq and Malone (2024): preregistered meta-analysis of 106 experiments and 370 effect sizes; studies published January 2020–June 2023. Negative values favor the better standalone condition. The standardized effect is not a percentage. Heterogeneous historical studies and possible publication bias limit generalization. Sources: [8]

CHAPTER 04

04

Compare delivery friction with control quality

Make a permission expire when its assumptions expire

For a fictional purchasing assistant, one page can permit comparison drafts from approved records while excluding contract signature and bank-detail changes. Name the owner, users, sources, review requirement and reassessment triggers, with links to supporting evidence. Staff can then judge a proposed change against an actual delegation.

Suppose the team then adds a tool that emails a preferred supplier automatically. Even if the model is unchanged, the service can now make an external representation. The permission owner must decide whether that action needs review, transaction limits or a narrower message template. Reusing the earlier accuracy evaluation may be reasonable for some content, but it supplies no evidence about recipient selection or authorization. A change review should explicitly separate evidence that remains valid from evidence that the new action requires.

Expiry also needs an operational consequence. If a required supplier assessment lapses or monitoring stops working, specify whether the service continues in a narrower mode, pauses new transactions or stops entirely. Avoid a rule that operators cannot implement without an emergency software change. Rehearse the transition and assign the customer communications. Our hypothesis is that clear consequences make proportionate control more credible: management can grant useful authority because it knows how that authority contracts when the supporting conditions weaken. The organization should test this proposition through actual change and suspension exercises before expanding the model across its portfolio.

Exhibit 01 turns that principle into a standing permission for internal policy drafting. External messages and customer-record changes remain outside it. Test this route against repeated case approval using effort, time to useful release and boundary failures. Reduced administration is worthwhile only if the permitted work remains adequately controlled.

Earn the case for faster governance

Open-ended systems will not become fully predictable because a permission has been documented. Pair action constraints with monitoring and narrow the delegation where uncertainty matters. If consequences are severe and recovery is weak, assistance may remain the appropriate mode.

Test the design on a small portfolio with different consequence levels. Reuse a common platform evidence pack, write a use-specific permission for each service, and challenge the highest-consequence release independently. Rehearse a material change, a hazard report and a suspension. These exercises reveal whether the organization can apply its own distinctions.

At the agreed review date, compare delivery friction, evidence quality and detected failures with the baseline. Inspect real refusals and exceptions as well as approvals. Retain boundaries that support useful, controlled work; revise those that are vague, unnecessarily restrictive or unsupported. Expand the model only on that evidence.

FROM INSIGHT TO ACTION

Decisions to take forward

BUSINESS OWNER AND GOVERNANCE LEAD

Authorize repeat work inside the tested boundary; identify what returns for review.

Evidence to require: Permission with purpose, users, actions, expiry and revocation

INDEPENDENT EVALUATOR

Select on relevant outcomes; reviewer presence alone is not proof.

Evidence to require: Human-only, AI-only and combined designs where lawful and appropriate

PORTFOLIO OWNER

Reuse controls that hold; restrict permissions whose assumptions have expired.

Evidence to require: Approval effort, useful-release time and detected/missed boundary failures

mayden connects senior judgement, research and engineering to turn strategic choices into working capabilities.

EVIDENCE YOU CAN INSPECT

Research notes

Method and boundaries

Narrative synthesis of eight opened primary guidance and research sources, including two studies with different findings on human reliance; cutoff 10 October 2026.

Selected for proportional risk management, system dependencies, assurance, incident language and human oversight. The proposed operating model combines these strands without claiming a measured causal effect.

No organization-level trial or legal classification. Experimental human-AI findings are task-dependent and may not transfer to production. The permission model is a mayden hypothesis requiring evaluation against local controls and outcomes.

Primary sources and author research

[1] AI Risk Management Framework

NIST / 2023-01-26; accessed 2026-10-10. Overview.

[Read the source](#)

[2] Guidelines for secure AI system development

NCSC and partners / 2023-11-27; version 1.0. Lifecycle stages.

[Read the source](#)

[3] Hidden Technical Debt in Machine Learning Systems

Sculley et al., NeurIPS / 2015. Sections 2-6, dependencies and feedback loops.

[Read the source](#)

[4] To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making

Buçinca, Malaya and Gajos / 2021. Abstract and study findings.

[Read the source](#)

[5] Explanations Can Reduce Overreliance on AI Systems During Decision-Making

Vasconcelos et al. / 2022-12-13 preprint; subsequent publication. Abstract; five studies.

[Read the source](#)

[6] Defining AI incidents and related terms

OECD / 2024-05-06. Abstract; incident and hazard distinction.

[Read the source](#)

[7] Introduction to AI assurance

UK DSIT / 2024-02-12. Assurance toolkit.

[Read the source](#)

[8] When combinations of humans and AI are useful: A systematic review and meta-analysis

Vaccaro, Almaatouq and Malone; Nature Human Behaviour / 2024-10-28. Main: Overall levels of human-AI synergy; Figure 1; published PDF pages 1-3. Heterogeneous historical studies; potential publication bias; standardized effects are not percentage changes or forecasts for modern agents.

[Read the source](#)