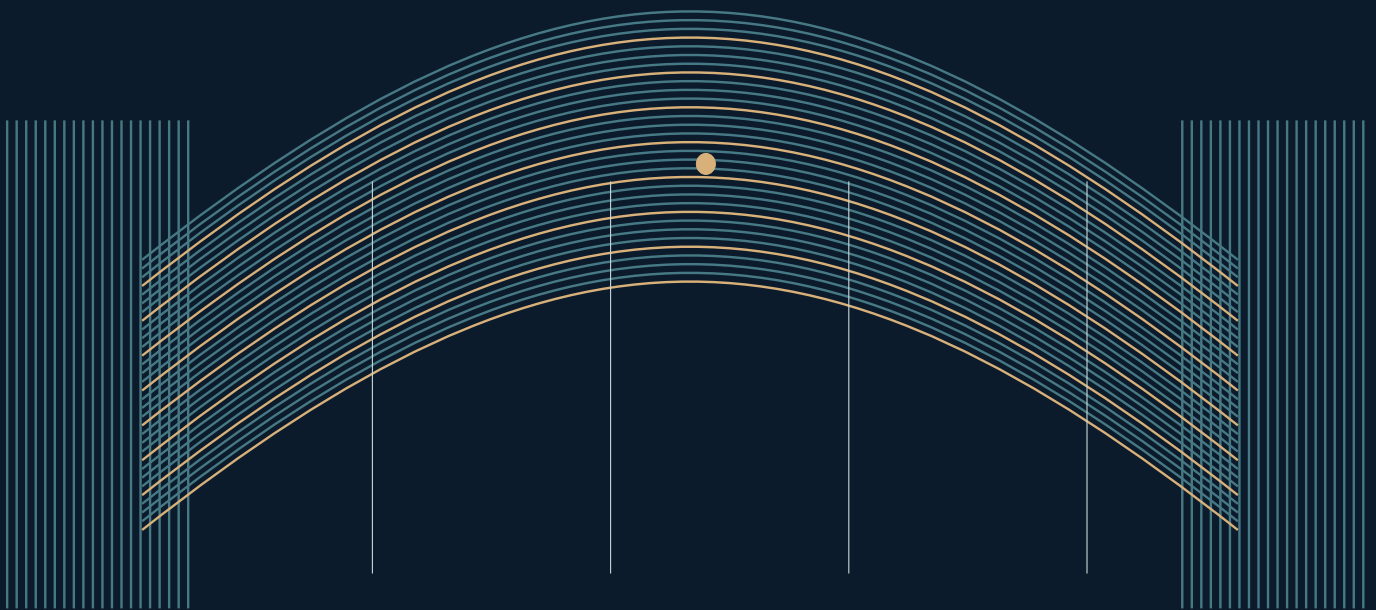




STRATEGY / OCTOBER 2026

The execution gap is a chain of unresolved constraints

How to move from promising demonstrations to repeatable business performance



THE DECISION IN BRIEF

Fund the constraint that prevents the next reliable operating outcome, then test where the bottleneck moves before expanding again.

- 01

A successful demonstration establishes feasibility under its tested conditions. Production must also withstand ordinary users, incomplete records, competing priorities and the cost of exceptions.
- 02

Locate the current limit: access, quality, ownership, integration, economics or capacity. Faster generation can worsen a reviewer queue when approval capacity remains fixed.
- 03

Use a Constraint Ledger to connect the next expansion to a named owner, a bounded intervention and evidence of removal. Choose deliberately among expansion, redesign, further learning and closure.

INSIDE THIS PERSPECTIVE

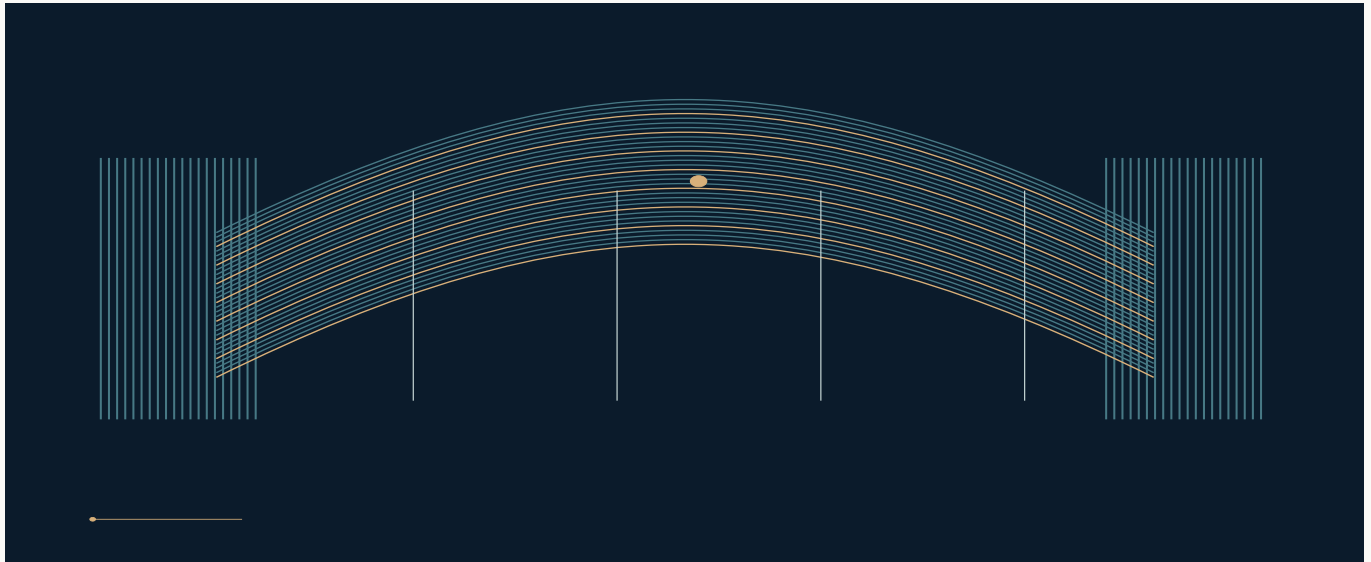
The argument at a glance	3
01 / Locate the binding constraint	4
02 / Earn operating acceptance	5
Exhibit 01 / The constraint moves after the review queue is resolved	6
03 / Test the next limitation	7
Exhibit 02 / EU enterprise adoption grew; this does not measure business impact	8
04 / Make the portfolio decision	9
The leadership agenda	10
Method and references	11

THE MAYDEN VIEW

Scale the accepted outcome, and fund the constraint that currently prevents it.

ONE DECISION, FOUR CONNECTED QUESTIONS

Connect the decision to delivery



EXECUTION BRIDGE / CONCEPTUAL ARTWORK

01 Locate the binding constraint

The next queue can absorb the first gain

02 Earn operating acceptance

Treat integration as part of the product

03 Test the next limitation

Make evaluation survive the move beyond enthusiasts

04 Make the portfolio decision

Some pilots should stop, and some should remain small

Scale the accepted outcome, and fund the constraint that currently prevents it.

CHAPTER 01

01

Locate the binding constraint

The next queue can absorb the first gain

A team prepares cases twice as fast, but the approval queue grows because the same specialists must review every case. This illustrative pattern explains why a strong pilot can fail to improve delivery. Eurostat reports AI use by 20.0% of its covered EU enterprise population in 2025, up from 13.5% in 2024. ^[1] Those figures cover enterprises with at least ten employees in specified sectors. They establish diffusion, not repeatable business impact.

The 2026 Census AI supplement distinguishes adoption from breadth. In its November 2025-January 2026 reference period, 18% of firms reported AI use in a business function; among adopters, 57% used it in three or fewer functions. ^[8] The practical implication is to ask where an initiative stops crossing operational boundaries. These are survey patterns, not evidence that broader adoption causes stronger performance or that a particular rollout has failed.

Demonstrations benefit from engaged sponsors, curated inputs and immediate specialist attention. Production adds ordinary users, staff turnover, incomplete records and competing priorities. Test those conditions before treating the demonstration as evidence of normal operating performance.

mayden defines the execution gap as the distance from demonstrated capability to a result that survives ordinary variation. Useful output must become accepted work; accepted work must become repeatable delivery; delivery must earn an economic or service benefit. Each transition needs evidence. A single pilot-success percentage hides where the chain breaks.

Fund the condition that limits the next release

Identify the condition governing the next expansion. Training cannot repair missing data access; lower model latency cannot repair a reviewer queue; a platform team cannot settle unowned business authority. Distinguish today's bottleneck from issues that matter later, then fund the action capable of changing it.

mayden's Constraint Ledger records the desired result, limiting condition, accountable owner, evidence of removal and next constraint to test. Exhibit 01 follows an illustrative approval workflow: first test routing or capacity against the reviewer queue, then test whether authorized source updates limit usable approvals. The second constraint is a hypothesis. Keep the outcome definition fixed so that narrowing the service cannot masquerade as removing the bottleneck.

The ledger should not become a list of excuses. Every constraint needs a test and a decision date proportionate to the work. If access is the issue, demonstrate a complete authorized record flow. If quality is the issue, evaluate a held-out sample with a defined acceptance rubric. If adoption is the issue, distinguish inability to use the system from rational refusal because it creates extra work. Each explanation implies a different next action.

Test interacting remedies before funding them in parallel. Better retrieval may reduce expert referrals, while sharper routing may concentrate difficult cases in the same queue. Trace a common sample through both changes. Fund the remedy to the binding constraint, then remeasure. Otherwise separate teams can claim the same released capacity or create incompatible demands on it.

CHAPTER 02

02

Earn operating acceptance

Treat integration as part of the product

Technical debt research identifies dependencies, feedback loops and maintenance burdens around machine-learning systems. ^[3] Its relevance here is structural: the model is one component inside a system that must continue operating. A proof of concept may hide its dependence on a manually prepared spreadsheet or a specialist who corrects every failed response. At scale, these hidden dependencies become visible costs and sources of inconsistency.

Integration should therefore be tested with the same seriousness as model output. Can the system identify the right record, preserve its version and return an outcome to the authoritative workflow? Can it recognize that the record has changed since the recommendation was prepared? Can a reviewer understand what happened without opening several tools? These questions determine whether the deployment removes work or creates a new reconciliation layer that staff must maintain.

A reusable platform can help, but premature generalization can also delay useful work. Build common services where several real deployments share requirements: identity, evidence retrieval, monitoring, controlled action and cost attribution. Keep domain-specific decisions close to the process owner. The platform should make complete outcomes easier to deliver, using interfaces and records that business teams can understand.

Replace launch approval with operating acceptance

A launch is an event; acceptance is a business judgement about continuing use. Before expansion, the operating owner should accept the definition of completion, the error limits, the fallback and the required support capacity. This differs from approving a presentation about model accuracy. It establishes who receives the work on an ordinary day and who bears the cost when it fails. The acceptance record should stay with the deployment as its scope changes.

GAO's accountability framework separates governance, data, performance and monitoring. ^[6] That separation is useful because a deployment can be strong in one area and weak in another. An accurate model with unclear authority is not ready for consequential action. A well-governed system with stale data is not ready either. mayden recommends a narrative readiness judgement with evidence for each material condition, rather than an average maturity score that allows strength in one area to conceal a critical gap.

The UK government's AI Playbook likewise addresses life-cycle management, human control, procurement and skills. ^[5] A private company need not copy a public-sector process, but it should recognize the same operational interfaces. Commercial terms determine whether a supplier can change a model without adequate notice. Support arrangements determine whether staff can recover from an outage. A deployment is only as repeatable as these surrounding commitments permit.

ILLUSTRATIVE DESIGN · NO MEASURED RESULTS

The constraint moves after the review queue is resolved

Illustrative decision record

Ledger field	Stage 1: review bottleneck	Stage 2: conditional next constraint
Desired operating result	Usable approvals with complete evidence	Same outcome and acceptance window
Limiting condition	Prepared cases wait for expert review	After review improves, authorised source updates may limit freshness
Accountable owner	Operating owner with expert lead	Data owner with process owner
Evidence to resolve	Accepted throughput, queue age and review effort after routing/capacity change	Dated source-to-case trace and unresolved stale evidence
Next constraint / decision	Check source freshness before expanding intake	Repair update authority or narrow scope if it is now binding

THE MANAGEMENT IMPLICATION

Resolve a tested limitation, then re-open the ledger instead of assuming the whole service can scale.

Illustrative sequence, not observed bottlenecks or a guaranteed causal chain. Stage 2 applies only if the trace confirms the next constraint. Source: mayden analysis. Proposed framework or illustrative scenario.

CHAPTER 03

03

Test the next limitation

Make evaluation survive the move beyond enthusiasts

Early adopters are useful collaborators, but they may not represent the people who will use a system next. They know the tool's limitations and often compensate for them silently. Wider rollout should deliberately include ordinary users, less familiar cases and the variation the pilot excluded. Segment the results before deciding whether the system is ready for the entire workflow. A strong mean can coexist with an unacceptable failure mode in a smaller case category.

Research on agent evaluation warns against accuracy-only assessment and weak holdout practices. ^[4] The corresponding enterprise discipline is to preserve examples the team has not optimized against and to measure the full cost of obtaining an acceptable answer. Repeated retries, human repairs and abandoned cases must remain visible. Otherwise the reported quality can improve simply because more hidden effort has been spent on each successful outcome.

HM Treasury's evaluation guidance emphasizes designing evaluation into an intervention from the start. ^[7] mayden's application is to define the decision the evidence must support before the pilot begins. An experiment designed to learn whether the system can retrieve evidence cannot simultaneously prove company-wide savings. Separate feasibility, operating and economic questions, and assign each an appropriate comparison. This prevents a technically successful pilot from being promoted into a stronger business claim than it earned.

Follow the work after the first bottleneck moves

When preparation improves, inspect the queue it feeds. More reviewer capacity, clearer evidence or authority to complete a bounded low-consequence category are different responses with different costs. Anticipate the next bottleneck, but do not assume it exists before measuring the changed workflow.

Track exception demand as an operating load. Distinguish errors that require correction, uncertain cases requiring judgement and legitimate requests outside the system's scope. They need different skills and should not share an undifferentiated queue. Record the time to detect, assign, resolve and return a case. The objective is not to drive all escalation to zero; some escalation is evidence that the system respects its boundary. The objective is reliable handling at the required volume.

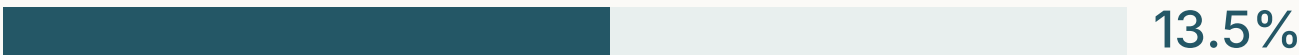
Changes to work also consume management capacity. The Census Bureau's early adoption research reports associated changes in training, workflows and cloud services. ^[2] That observation argues for budgeting implementation effort, while remaining insufficient to establish causality. Leaders should protect operational experts' time during expansion. A programme that relies on unpaid overtime from the same few specialists may appear economical until those specialists return to their normal responsibilities.

EMPIRICAL RESULT · ENTERPRISE ADOPTION

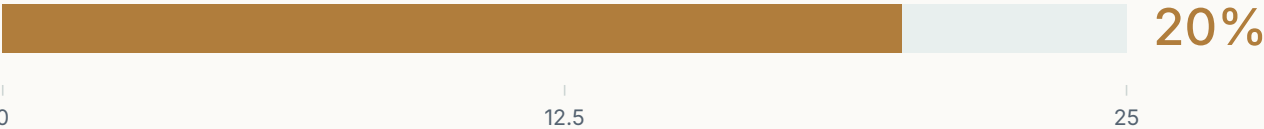
EU enterprise adoption grew; this does not measure business impact

% of enterprises using AI

2024



2025



THE MANAGEMENT IMPLICATION

Diffusion raises the urgency of operating discipline; it does not establish the return on an individual deployment.

Eurostat release 11 December 2025, covered sectors and enterprises with at least ten employees. A 6.5-percentage-point increase. This is adoption, not the proportion scaled successfully. Sources: [1]

CHAPTER 04

04

Make the portfolio decision

Some pilots should stop, and some should remain small

Scaling is not a virtue independent of value. A narrow tool may serve a small but important group exceptionally well. Forcing it into other departments can reduce quality and increase support requirements. Similarly, a pilot that establishes that a process is not economically attractive has produced useful information if it was designed to answer that question cheaply. The portfolio should distinguish successful learning from successful production, without pretending the two have identical financial value.

The strongest objection to a constraint-led approach is that it can favour incremental improvements over a bold redesign. That risk is real. A local bottleneck analysis can miss a different operating model in which the whole workflow becomes unnecessary. To address this, keep two questions in the same review: what prevents the next release under the present design, and what alternative design could remove the need for that release? Incremental evidence and strategic imagination should challenge each other.

Another boundary is weak observability. Some outcomes, such as trust or strategic judgement, are difficult to measure quickly. The answer is not to substitute a convenient usage metric and call it impact. Use multiple forms of evidence, explicit assumptions and longer observation where needed. When the consequences are material and the evidence remains weak, narrow the scope of action. The uncertainty itself belongs in the decision rather than being hidden behind a confident score.

Make the next operating decision explicit

Require each sponsor to name the current limiting condition and the smallest funded action that could resolve it. Record the owner, decision date and required evidence. Organize the portfolio around those unresolved operating decisions, so management can see why the next commitment is justified.

Choose explicitly. Expand when the accepted result repeats. Redesign when the mechanism works but the surrounding process prevents value. Fund further learning only against a specific uncertainty. Close the initiative when the next cost exceeds the expected benefit. Prior spending is not evidence for another release.

Preserve the conditions behind the result: failed cases, exception demand, integration choices and the reason for the investment decision. Reuse that knowledge without copying its assumptions. The portfolio improves when teams can identify the next constraint faster and demonstrate that removing it changes a real outcome.

FROM INSIGHT TO ACTION

Decisions to take forward

SPONSOR

Fund the smallest decision-changing action

Evidence to require: One limiting condition and a bounded test

OPERATING OWNER

Resolve the current constraint

Evidence to require: Queue, accepted throughput and failure trace

PORTFOLIO OWNER

Expand, redesign, learn or close

Evidence to require: Evidence of removal and next constraint

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 official, academic and engineering sources with an original constraint-based operating framework. No pooled effects or proprietary survey.

Sources selected to connect adoption measurement, integration, evaluation and operating accountability; accessed 10 October 2026.

Adoption statistics cannot quantify a scaling failure rate. Frameworks are normative. Illustrative operating examples are mayden analysis and require company-specific validation.

Primary sources and author research

[1] 20% of EU enterprises use AI technologies

Eurostat / 11 December 2025. Main findings and methodological notes; isoc_eb_ai. Adoption is not scale or impact; excludes most microenterprises and some sectors.

[Read the source](#)

[2] Tracking Firm Use of AI in Real Time: A Snapshot from the Business Trends and Outlook Survey

US Census Bureau; Bonney et al. / March 2024. Abstract; CES-24-16. Associations do not identify the causal effect of adoption; data September 2023–February 2024.

[Read the source](#)

[3] Hidden Technical Debt in Machine Learning Systems

Sculley et al.; NeurIPS / 2015. Abstract. Predates generative AI; used for structural insight, not current failure-rate estimates.

[Read the source](#)

[4] AI Agents That Matter

Kapoor et al.; arXiv / 1 July 2024. Abstract. Benchmark findings do not quantify enterprise return.

[Read the source](#)

[5] Artificial Intelligence Playbook for the UK Government

Government Digital Service / 10 February 2025. Principles; human oversight; life cycle and commercial guidance. UK public-sector context; not a substitute for jurisdiction-specific advice.

[Read the source](#)

[6] Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities

US Government Accountability Office / 30 June 2021. Highlights and framework principles. Normative framework, not a measured causal estimate.

[Read the source](#)

[7] Magenta Book: Central Government guidance on evaluation

HM Treasury and Evaluation Task Force / 15 May 2026. Evaluation design, impact and value-for-money sections. Adapted to company decisions here; no claim of official endorsement.

[Read the source](#)

[8] The Microstructure of AI Diffusion: Evidence from Firms, Business Functions, and Worker Tasks

Bonney et al.; US Census Bureau, CES-26-25 / April 2026; page revised 7 May 2026. Abstract; supplement reference period November 2025–January 2026. Descriptive/self-reported; correlation is not causation. New business-function definition is not directly comparable with earlier producing-goods/services question.

[Read the source](#)