

A Successful AI Pilot Is Not an Approval Case

An executive guide to turning demonstrated capability into a bounded deployment decision

Enterprise AI Adoption Gap · Research brief 06
October 4, 2026

© 2026 Eldris Inc. All rights reserved.

Executive brief

An AI pilot can produce good outputs, save time and earn enthusiastic feedback while leaving the deployment decision unresolved. The person being asked to approve it is accepting a different proposition: that a particular system, used by particular people with particular permissions, can operate within acceptable limits and deliver a worthwhile result.

Pilot success supports an approval case. It does not complete one. The missing evidence may concern ordinary working conditions, who can authorize actions, whether oversight is feasible, or what happens when the system fails. A stronger demonstration of the same capability may leave those questions untouched.

This report is for enterprise workflow owners, AI program leaders and providers preparing a production proposal. Its bounded decision is whether to authorize an AI assistant to draft routine customer-support replies for one team, with a human approving every message before it is sent. This example makes the argument concrete; different workflows need different requirements.

The central recommendation is to build an **approval evidence map** before the pilot ends. For each material claim, record what was tested, what remains unknown, who owns the operational requirement and what evidence would resolve it. Evaluate the proposed permission separately from the attractiveness of the technology.

Three principles follow:

- Define the authorized action. Approval to draft is different from approval to send, change a customer record or issue a refund.
- Test the complete operating arrangement. Human review, escalation and recovery are part of the system being approved.
- Make uncertainty actionable. A missing fact should lead to a narrower permission, a specific test or a reasoned stop decision.

The objective is a better decision, including a defensible decision to delay or decline. This report does not claim that a particular governance design increases approval rates. It provides a practical structure for discovering what the proposed deployment still needs.

The question changes at the approval boundary

A pilot may ask whether the assistant can draft a useful reply. Deployment asks whether the team should depend on that assistant under ordinary conditions. The distinction is consequential even when the model and task remain the same.

A specialist may have selected easy cases, supplied clean reference material, corrected prompts and helped reviewers recognize errors. Those are legitimate ways to explore capability. If production depends on that support, however, the support must be provided and costed. If it will disappear, the pilot must test the resulting arrangement.

There is a second boundary: permission. A draft that a trained employee checks before sending has different consequences from a message sent automatically. Changing the permission changes the approval proposition. It cannot be justified solely by a higher average score on the original drafting test.

NIST's Generative AI Profile warns that pre-deployment measurements may fail to reflect deployment contexts and that laboratory results may not transfer to real-world use. That supports examining the receiving workflow; it does not establish how often enterprise pilots fail or why particular approvers refuse them. [NIST testing guidance](#).

Decision implication: Write the production proposition before evaluating pilot success: the eligible cases, users, sources, permissions, review requirements and intended business outcome. Compare the evidence against that proposition.

Reported barriers are clues, not an approval diagnosis

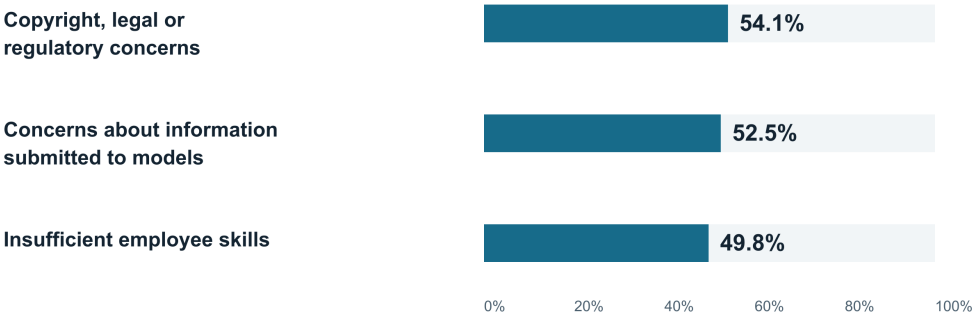
Public adoption surveys can identify concerns worth investigating. They cannot tell a program leader which concern is decisive in an organization that has already completed a successful pilot.

In OECD research, 54.1% of SMEs not using generative AI reported concerns about copyright, legal or regulatory issues; 52.5% reported concerns about information submitted to models; and 49.8% reported insufficient employee skills. These are selected, overlapping reasons for non-use. They are not percentages of successful pilots rejected for those reasons. [OECD barrier findings](#).

ENTERPRISE AI ADOPTION GAP / RESEARCH BRIEF 06

Reported concerns extend beyond output quality

Selected reasons for non-use · OECD SME survey · Late-2024 fieldwork



These are overlapping reasons for non-use, not pilot rejection rates.

Denominator: SMEs not using generative AI. Selected barriers; not exhaustive.

Source: OECD, Generative AI and the SME Workforce, 2025, chapter 4.

© Eldris Inc

Figure 1. Selected published reasons for non-use. OECD survey of 5,232 SMEs across seven countries, with main fieldwork October 14-December 6, 2024. The percentages concern non-users, not the full sample. Responses overlap; the chart is not an exhaustive ranking or a study of pilot approval. [Survey methodology](#).

For the support workflow, an approver might need evidence that reference material is current, customer information is handled appropriately or reviewers can identify misleading answers. These are different requirements. A general assurance that the AI is trustworthy does not specify which has been met.

Ask the responsible person to state the unresolved condition in operational terms: “What would need to be true for you to authorize this defined use?” Then distinguish a missing fact from a prohibited activity, a resource constraint or an unresolved allocation of responsibility.

Decision implication: Use external barriers to prepare questions. Use evidence from the actual approval process to identify the requirement that must be satisfied.

Build the approval evidence map

The map below is original Eldris guidance. It organizes a decision; it is not a validated predictor, certification scheme or additive score. A material failure in one area can outweigh strong evidence elsewhere.

ELDRIS APPROVAL EVIDENCE MAP

Five claims an approval case must examine

Original analytical guidance · Tailor the evidence to the requested permission

CLAIM	EVIDENCE TO EXAMINE	RESPONSIBILITY
Relevant capability	Performance on eligible cases; important errors and limits.	Evaluation lead + workflow owner
Bounded authority	Enforced permissions; approved data access.	System owner + authorized reviewers
Feasible oversight	Error detection, correction effort and capacity under load.	Workflow owner
Recoverable operation	Exercised stop, fallback and repair; change monitoring.	Operational owner
Worthwhile result	Accepted outcomes against baseline; total operating burden.	Business owner

Strong evidence in one area cannot cancel a material failure in another.

Role labels describe responsibilities. This is not an additive score or certification.

© Eldris Inc

Figure 2. Five claims to examine for a bounded deployment. The required evidence and accountable roles depend on the workflow. Role labels describe responsibilities, not a universal organizational structure.

Capability in the receiving workflow

Show performance on cases resembling the intended workload, including difficult and ambiguous cases. Identify excluded requests. Distinguish an answer that sounds plausible from one that correctly applies the organization's current policy.

For customer support, record incorrect policy statements, unsupported claims and responses that require material correction. Break results down where case differences could affect the decision. An average score can conceal a small category with unacceptable consequences.

The evidence should describe the evaluated system version, reference sources, reviewer qualifications and case selection. Keep uncertainty visible. A small test with no observed severe errors cannot establish that severe errors never occur.

Authority and data boundaries

State exactly what the assistant can read, produce and change. In the proposed workflow, drafting authority does not include sending messages, changing accounts or granting refunds. Test whether the system enforces these boundaries, including when a request or retrieved material asks it to cross them.

Identify the information needed for the task and the approved arrangements for accessing it. Technical demonstrations cannot settle contractual, legal or organizational permission questions. Obtain the relevant determinations from people authorized to make them.

Feasible human oversight

Specify who reviews each draft, what information they see, what they must verify and when they escalate. Then test that arrangement under realistic volume and time constraints.

A review checkbox shows that a step exists. It does not establish that the reviewer can detect important errors. Measure correction effort, missed errors and the demands placed on expertise. Include the work that the pilot team quietly absorbed.

Recovery and ongoing operation

Identify who can suspend the assistant, how work continues and who repairs affected cases. Exercise the stop and fallback procedures. Stopping future generation does not retract an already sent message or repair a changed record.

Define what changes require reevaluation: a model update, a new reference source, a new case category or broader permission. Assign operational responsibility for monitoring and responding. NIST's AI RMF addresses documented responsibilities, leadership accountability and post-deployment monitoring, including incident response, recovery and change management. It is voluntary guidance, not proof that the proposed deployment is safe. [NIST AI RMF](#).

A worthwhile operating result

Measure the outcome that justifies deployment after including review, corrections, support and operating costs. For the support team, this might be faster resolution at maintained service quality, rather than more drafts generated.

Compare with the current workflow and plausible simpler improvements. An assistant can perform well while failing to justify its total operating burden. Benefit ownership matters: identify who can change staffing, service capacity or work allocation if the result warrants it.

Decision implication: Require evidence for material claims, rather than a complete-looking document. Tailor the map to the consequence and permission being proposed.

Separate accountability from personal exposure

A named owner is useful only when the organization gives that person the authority, information, resources and escalation routes needed to act. Assigning a name to an unresolved risk does not resolve it.

In the support example, the workflow owner may control procedures and staffing; a system owner may control access and configuration; a designated leader may authorize the deployment within organizational policy. The same person can hold several roles in a small organization. A multinational may distribute them across functions and locations.

Make four responsibilities explicit: who authorizes the scope, who operates the workflow, who handles exceptions and who can stop or change the deployment. Identify how disagreements are resolved. The person asked to approve should know which commitments are funded and which remain proposals.

Avoid treating hesitation as evidence that an individual dislikes AI. Someone may accept the demonstrated capability while lacking authority to accept the residual risk. Someone else may agree in principle but lack the review capacity to operate it. These are possibilities to investigate, not explanations to presume.

Decision implication: Ask for an organizational commitment with defined decision rights. Do not ask an employee to absorb ambiguous responsibility in exchange for adopting a promising tool.

Keep permission proportional to evidence

Authorization should describe what is allowed now and what would justify a later expansion. It need not be a choice between indefinite experimentation and unrestricted deployment.

ILLUSTRATIVE CUSTOMER-SUPPORT PERMISSIONS

Each permission changes the approval question

Different propositions · No automatic progression toward greater autonomy

DRAFT FOR REVIEW Human approves before sending.	Can reviewers detect important errors at the intended volume?
SEND ROUTINE REPLIES Customer sees output before review.	Can errors be detected and addressed after a message reaches the customer?
CHANGE ACCOUNTS / REFUND Assistant changes a customer outcome.	Are action limits enforced, and can affected records or transactions be repaired?

Success at drafting does not establish readiness for sending or acting.

Some workflows should retain human approval. Every expansion needs its own decision.

© Eldris Inc

Figure 3. Illustrative permission boundaries for customer support. These are different approval propositions, not stages every organization should complete. Broader permission requires evidence about the consequences it introduces.

A limited deployment could authorize drafting for selected routine requests, require review before sending and exclude cases involving disputed charges or policy exceptions. Those exclusions would need to match the actual organization; they are examples, not universal requirements.

A proposal to send routine replies automatically introduces a new exposure: an incorrect reply can reach a customer without prior review. Evidence from reviewed drafting may inform the decision, but it does not test all consequences of removing that review. A proposal to issue refunds introduces another boundary involving changes to customer accounts and financial actions.

Broader authority may be worthwhile. It should receive its own assessment of failure consequences, controls, detection and recovery. Some workflows should retain human approval permanently.

Decision implication: Record both allowed and prohibited actions. Link any future expansion to a new decision, rather than assuming that successful use creates permission to do more.

Turn uncertainty into the next decision

Consider a constructed example, not an observed company case. A support pilot produces acceptable drafts, but the team has not tested whether ordinary reviewers detect policy errors during busy periods. The unresolved question concerns the proposed oversight arrangement.

A second showcase of high-quality drafts would add little evidence about that question. A more informative next step would test the complete review process using cases with independently established reference answers and an appropriate range of errors. Record detection, corrections, time and escalation. Protect customers through the existing approved process while the arrangement is evaluated.

Before running that test, agree what result would change the decision. Requirements should reflect the consequence of an error and the available alternatives. They should not be selected afterward to make the pilot appear successful.

Four responses are available:

- Authorize the defined scope when the material requirements are met and the responsible authority accepts the remaining uncertainty.
- Narrow the scope when exclusions or lower permissions can address a material unresolved condition.
- Run a targeted test when a missing fact can realistically change the decision.
- Decline or pause when the activity is prohibited, the operating commitment is unavailable or the value case is insufficient.

These responses are Eldris recommendations. Their effectiveness at increasing adoption or improving decisions has not been measured in this report.

What the approval proposal should contain

Use a short decision record linked to supporting evidence. Its purpose is to make the actual choice assessable, not to produce another lengthy governance document.

1. **Requested permission:** the workflow, eligible cases, users, system version, sources and allowed actions.
2. **Demonstrated capability:** the relevant tests, case coverage, performance limits and uncertainty.
3. **Operating commitments:** review capacity, access arrangements, escalation, support and accountable roles.
4. **Recovery and change:** stop authority, fallback, repair procedures and reevaluation triggers.
5. **Business result:** the baseline, intended outcome, total operating burden and owner of the benefit.
6. **Decision and conditions:** what is approved, excluded or deferred; who authorizes it; when it will be reconsidered.

For providers, this means making the deployment proposal legible to the buyer's accountable functions, not only attractive to the pilot sponsor. Product documentation and demonstrations should show how permissions, oversight and recovery work in the buyer's setting. Buyers should verify those claims.

For enterprise leaders, the useful question at the end of a pilot is: **“What deployment decision can this evidence support?”** Where the answer is unclear, identify the missing condition before funding another demonstration of the same capability.

Method and limitations

This brief selectively reviews primary sources checked on October 4, 2026. It is not an original participant study, systematic review or estimate of enterprise pilot conversion. OECD figures describe reported barriers among SME non-users in seven countries; they do not establish the causes of approval decisions in U.S. multinational enterprises.

NIST publications supply governance and evaluation guidance. They are not empirical findings about which approval designs work best. The approval evidence map, support-workflow example, permission distinctions and decision record are original Eldris analysis. The example is constructed and carries no claimed company results.

The report supports preparation and diagnosis of a bounded decision. It cannot establish legal eligibility, guarantee deployment performance or predict approval. Determining how oversight, reversibility and accountability change authorization would require research with qualified decision-makers, defined tasks and controlled comparisons, followed where feasible by evidence from real approval processes.

Sources

1. **NIST. Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1.** January 2023. GOVERN 2.1-2.3, GOVERN 3.2 and MANAGE 4.1 address responsibilities, leadership accountability, oversight and post-deployment operation. [Framework](#).
2. **NIST. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, NIST AI 600-1.** July 2024. Appendix A.1.4 discusses limitations of pre-deployment testing and mismatches between laboratory and deployment settings. [Generative AI profile](#).
3. **OECD. Generative AI and the SME Workforce: New Survey Evidence. Chapter 4, Are SMEs prepared for generative AI?** Published November 5, 2025. Selected non-use barriers: legal or regulatory and copyright concerns, 54.1%; input-information concerns, 52.5%; employee skills, 49.8%. These are overlapping responses, not pilot rejection rates. [Findings](#).
4. **OECD. Generative AI and the SME Workforce. Chapter 1, Survey methodology.** Survey of 5,232 SMEs in Austria, Canada, Germany, Ireland, Japan, Korea and the United Kingdom; firms up to 249 employees, including one-person companies. Main fieldwork October 14-December 6, 2024. Telephone interviews with stratified sampling by country, firm size and sector. [Methodology](#).