

Choose an AI pilot worth putting to work

A business-first decision guide for B2B teams

Select a useful task, set a safe data boundary and define the tests, economics and ownership needed before an AI pilot reaches everyday work.

Version 1.0 · September 14, 2026

Original analysis + working worksheet
Not proprietary survey research

Is this task ready for a controlled pilot, and what evidence would justify continuing?

Business, operations and technology leaders evaluating a bounded internal AI use case.

Three decisions to carry forward

- Compare AI with a simpler process, search or automation fix.
- Choose a task with approved information, a named reviewer and a useful fallback.
- Measure correct completed work and review effort, not only generated output.

Evidence and limitations

Original Onyx planning guidance, not a security assessment, legal opinion, certification or assurance of model performance. No live tool, customer data or model was evaluated for this report. Risk and acceptance thresholds must be set by the accountable business and technical owners.

How to use the report

Read the analysis, use the companion worksheet and test the questions against your own operating evidence. Do not substitute an example or an outside survey for a decision by your responsible business and technical owners.

Start with the job, not the tool

Describe one recurring task in terms of its input, decision and output. Examples include finding an approved internal answer, drafting a nonbinding meeting summary or organizing material for expert review. Select a problem that matters often enough to measure and has a person responsible for the result.

Compare at least three alternatives: improve the underlying process, use a deterministic search or automation tool, or test an AI-supported approach. If the source material is contradictory, unapproved or unavailable, a new model will not resolve who owns the truth. Fix the information problem first.

Avoid a first pilot that makes unsupervised safety, employment, credit, legal or technical suitability decisions. A task can be commercially attractive and still be inappropriate for the available controls. Scope the permitted assistance narrowly and keep consequential decisions with authorized people.

Put this to work

- Write one input-to-output task statement.
- Name the non-AI alternative and why it may be sufficient.
- Define what the tool may never decide or do on its own.

Read industry evidence at the right scale

Deloitte's 2025 Smart Manufacturing Survey covers 600 executives at large companies: at least \$500 million in annual revenue and more than 1,000 employees. It discusses workforce capability, transformation and operational risk. Its findings are not a benchmark for every contractor, distributor or smaller business.

NIST's AI Risk Management Framework is voluntary guidance for managing AI risks. Its Govern, Map, Measure and Manage functions support ongoing risk work; they are not a certification checklist or a promise that completing four steps makes a system safe.

Onyx interpretation: a practical first pilot needs an accountable owner, explicit operating context, reproducible quality checks and a way to respond when the system is wrong. Use the following questions to prepare those decisions. The worksheets are an original working aid, not an official NIST assessment.

Put this to work

- Record which evidence actually applies to your task.
- Do not present survey adoption as proof of local demand or ROI.
- Have the responsible owner review the scope before selecting a tool.

Source context: [Deloitte Insights – 2025 Smart Manufacturing Survey](#); [NIST – AI Risk Management Framework](#). Full scope notes are on the final page.

Check readiness before ranking opportunities

First apply hard gates. You need permission to use the information, a defined task, a reviewer able to judge the output and a fallback when the tool fails. A missing gate is a reason to investigate or pause, even when estimated time savings are large.

After those gates, compare frequency, current handling effort, consequence of error, source stability and integration burden. Record the evidence behind each judgment. A simple ordinal ranking can help discussion, but a high score is not permission to override a failed privacy or safety gate.

Readiness gate	Evidence required	If missing
Permitted information	Owner-approved data categories and tool boundary	Use synthetic material or pause
Judgeable output	Examples of correct, incorrect and uncertain results	Define the task and reviewer first
Human ownership	Named owner with review capacity	Do not release unattended work
Recoverable failure	Manual fallback and stop condition	Keep the work outside production

Define the smallest permissible information boundary

Specify what may enter the system and what must remain outside it. For an internal knowledge aid, start with a limited set of approved, current documents. Record owner, version, effective date, access restrictions and the procedure for withdrawing superseded information.

Ask the technical owner to confirm the selected vendor's actual terms, retention, access controls and deployment settings at the time of implementation. Do not assume that every product tier or account configuration handles data the same way. This report does not approve any vendor or transfer of confidential material.

Separate retrieval from action. A tool that can suggest an internal answer does not automatically need permission to send a customer email, update a quote or change a record. Treat instructions inside retrieved documents as data to inspect, not authority to expand what the system may do.

Put this to work

- Create an explicit allowed-source list and exclusion list.
- Test a stale source, conflicting source and unauthorized request.
- Provide an answer-with-uncertainty or escalation path rather than a forced guess.

Build a test set before the demonstration

Collect representative tasks and write the expected evidence before configuring the pilot. Include routine requests, ambiguous wording, missing information, contradictory documents, restricted data and a request to exceed authority. Use synthetic or permitted redacted material.

Keep some cases separate from configuration work so the final review does not simply repeat rehearsed demonstrations. Record task, source version, expected behavior, observed output, reviewer and result. A small test set can reveal defects; it cannot establish a universal accuracy rate or guarantee future behavior.

Agree which failures are critical. An unauthorized disclosure or action should stop the affected release, even if many easy answers are correct. Measure unsupported statements, incorrect source selection, missed escalation and review effort separately. Repeat relevant tests after changes to the model, prompt, source library, access or workflow.

Put this to work

- Test normal, ambiguous, adversarial and failure cases.
- Use an independent reviewer for consequential acceptance.
- Record exact versions and retest evidence, not only a green status.

Make the economics include review and maintenance

Measure the complete job. If a draft takes seconds but checking and correcting it takes longer than the old process, the pilot has not demonstrated an efficiency gain. Include preparation, review, corrections, failures and time spent keeping sources current.

Illustrative arithmetic: 100 tasks per month at 12 minutes each is 20 baseline hours. If the pilot needs 3 minutes of preparation and 4 minutes of review per task, total effort is about 11.67 hours. The modeled capacity release is about 8.33 hours. At a hypothetical \$45 loaded hourly rate, that is \$375 of capacity value before software, setup and maintenance.

If monthly software and maintenance total an assumed \$250, the modeled residual is \$125 before implementation cost and other obligations. This is not a price recommendation, cash return, revenue forecast or observed result. A business may value better consistency or faster response even without cash savings, but should name that objective honestly.

Input	Measure it as	Do not confuse it with
Volume	Comparable tasks in a defined period	All company activity
Effort	Preparation + review + correction + failure recovery	Generation time alone
Quality	Correct, supported, authorized completed work	Fluent-looking answers
Value	Capacity or actual avoided cost, separately	Guaranteed cash or revenue

Prepare adoption, ownership and the stop decision

Train the actual users on the task boundary, how to verify an answer and when to escalate. Ask them to demonstrate the job using unfamiliar examples. Give managers a way to observe first use without encouraging people to bypass review in pursuit of adoption numbers.

Before a controlled release, agree who maintains sources, who investigates errors, who can disable the workflow and who approves a material change. Define a review date and the information needed to decide whether to continue. Software access without those responsibilities is an incomplete handoff.

Continue only when the evidence supports the agreed use case. Revise when the underlying task is useful but the controls or economics fall short. Stop when acceptable performance requires unavailable review capacity, unauthorized data or costs the business cannot justify. A decision not to automate can be a valuable outcome of the pilot.

Put this to work

- Confirm user practice and owner acceptance.
- Provide a manual fallback and an issue-reporting route.
- Decide continue, revise or stop against the original task and evidence.

Keep the evidence with the decision.

Prepared by Onyx Advisory Group. Public sources reviewed September 14, 2026. This is an original decision guide, not commissioned client work or a reproduction of the source reports.

NIST – AI Risk Management Framework

AI RMF 1.0: 2023; Generative AI Profile: July 26, 2024. Voluntary risk-management guidance, not a certification or an assurance that a particular tool is safe.

Deloitte Insights – 2025 Smart Manufacturing Survey

2025; fieldwork August–September 2024. 600 executives; represented companies have at least \$500 million annual revenue and more than 1,000 employees, with US headquarters or operations. Not representative of all small and midsize businesses.

AI pilot decision and test worksheet

The companion CSV contains 6 working rows and 10 fields. Open it in your spreadsheet application, preserve the original, and fill a separate copy with approved information.

The worksheet is an original planning aid, not a validated maturity score or a promise of commercial results.

Bring the work into a discussion

Identify the business decision, the evidence you already have and the unresolved question. Use the discussion link on the report's website page when a separately scoped engagement would help.

No third-party endorsement or affiliation is implied. The website discussion brief prepares a local file and does not yet deliver to CRM.