

For maintenance, operations and executive leaders
in asset-intensive organisations

Your AI pilot worked.

What did it actually prove?

Before approving implementation, establish whether the pilot validated the problem, the available data, the business value and the way people will use it.

Test the assumptions.
Fund the next stage on evidence.

Working technology is only part of the investment case.

An AI assistant drafts a maintenance job plan in seconds. The demonstration uses clear records from a familiar equipment family. An experienced planner checks the output, and the supplier is on hand to resolve problems. Everyone agrees it works.

Before approving a rollout, ask what made that result possible. Would it work with ordinary records, another shift and normal support? Would the completed job improve after checking, correction and approval?

The example is illustrative, but the decision is practical: a pilot should reduce uncertainty about the investment. A demonstration of capability answers only one part of that question.

FIVE QUESTIONS BEFORE THE NEXT COMMITMENT



A pilot can succeed by telling you to stop

Finding that a problem is too small, the data is unsuitable or the workflow cannot support the solution can prevent a poor investment. The mistake is treating an untested assumption as a proven result.

This paper follows a job-plan assistant through those five questions. It combines a clearly labelled illustrative example with research and an industrial scaling account. The framework is a practical guide, not a validated scoring model.

Did you validate a problem, or build an interesting solution?

“Use AI to improve planning” is a direction, not a defined problem. Before evaluating a tool, establish where planning breaks down and who experiences the consequence.

Find the actual bottleneck

A slow job-plan process could reflect time spent drafting instructions. It could also reflect missing equipment information, unavailable parts, unclear scope or delayed approvals. Faster drafting will not necessarily resolve those constraints.

Observe representative jobs and speak with the people involved. Record how often the problem occurs, the effort it consumes and what happens downstream. Include awkward cases, not only those the team can already describe neatly.

Compare with a simpler remedy

A better template, corrected master data or a clearer approval process may address part of the problem. Compare the proposed AI approach with those alternatives. If both help, identify what each contributes rather than crediting the entire improvement to AI.

Write a testable problem statement

For example: “For this equipment family, planners spend substantial time assembling job information from several sources. We want to reduce that effort while maintaining plan quality and approval requirements.” Add a measured baseline and a success threshold agreed with the operation.

Keep the target stable during the trial. If the problem changes, revise the evaluation explicitly. A compelling demonstration should not quietly redefine what the project was meant to solve.

Evidence to take forward

A named user

Who experiences the problem?

A measured baseline

Frequency, effort and operational consequence.

An alternative

What happens if we improve the current process without AI?

A decision owner

Who agrees that solving this is worth funding?

Did it work on your data, or only on your best data?

Starting with a clean dataset can be a sensible way to test feasibility. It becomes misleading when success on that subset is presented as evidence for the entire operation.

Make preparation visible

Record what was cleaned, labelled, reconciled or excluded. If a specialist corrected equipment identifiers or selected the right drawing before each test, that effort belongs in the implementation plan and cost estimate.

Test the conditions the rollout will encounter

For a job-plan assistant, include incomplete work orders, inconsistent equipment names, different configurations and conflicting document revisions. Sample across relevant asset families, sites and shifts. Report results by group so that a good average does not hide a weak area.

Keep a separate evaluation set that was not used to refine the solution. Have qualified reviewers agree what a usable answer looks like. Test whether the assistant asks for clarification or declines to proceed when evidence is insufficient.

Check access as well as accuracy

The right information may exist but be inaccessible under normal permissions, arrive too late or require manual export. Trial the intended connection to operational systems, including update frequency and source ownership.

Distinguish “the technology can work with suitable data” from “we can supply and maintain suitable data at an acceptable cost”. The first finding can justify further work. It does not settle the second.

Keep three views

Prepared cases

What is possible under favourable conditions?

Representative cases

What happens in ordinary work?

Difficult cases

What happens when records are missing, conflicting or unfamiliar?

A faster draft is not the same as a better job.

The job-plan assistant produces a draft quickly. The planner then checks task sequencing, confirms equipment references and corrects the content before approval. The technician may still need clarification at the job.

Measure the complete task

Compare the time and quality of the existing process with the assisted process. Include preparation, generation, checking, corrections, approval and any downstream rework. A saving in one step can be offset by effort elsewhere.

Translate time into a credible benefit

Released planner hours may increase the capacity to prepare work. They do not automatically reduce payroll or downtime. Establish how that time will be used and which operational measure should change as a result.

Avoid adding overlapping benefits together. If better preparation helps reduce a delay, do not count the same avoided delay again as a separate productivity improvement without explaining the distinction.

Include the costs required to realise it

Count data preparation, integration, review, training, software usage and support. Separate one-time development from costs repeated for each site or equipment family. State the expected volume of use and test how lower adoption or higher review effort affects the case.

For an early pilot, shorter information-search time or fewer plan corrections may be credible findings. A claim about avoided failures or production gains requires evidence linking the tool to those outcomes over an appropriate period.

Use a complete comparison

Current process

Preparation + execution + review + rework.

Assisted process

Data preparation + AI use + checking + approval + rework.

Business benefit

What can the organisation do with any time or quality improvement?

A planner and a technician do not need the same change plan.

A successful trial with enthusiastic volunteers says little about wider use unless it includes the people and conditions the rollout will serve. Treat adoption as part of product design and evaluation, not a training event at the end.



THE PLANNER

Needs: editable recommendations, maintenance-system integration and clear approval steps.

Observe: correction time, duplicate entry and whether plans reach approval sooner.

Support: practise on real planning tasks, agree review rules and provide a route to report recurring errors.



THE TECHNICIAN

Needs: equipment-specific information, access where the work happens and clear limits.

Observe: retrieval and checking during the job, clarification requests and missing context.

Support: field-based practice, shift coverage and clear instructions on when to seek qualified help.

Design around the task, not just the job title

An experienced technician and a new starter may need different explanations. Connectivity, shift pattern, work location and task risk can matter as much as role. Observe actual use and include people beyond the original pilot champions.

If people avoid the tool, investigate whether it adds work, lacks useful information or creates uncertainty about accountability. Training will not repair a workflow that is poorly designed.

Illustrative role comparison. Validate these needs with the users and working conditions at your sites.

Can it keep working when the pilot team leaves?

During a pilot, a supplier may answer questions immediately, an engineer may correct records and a project lead may encourage every user. Those arrangements can conceal the effort required to keep the service useful.

Record the help behind the result

Track manual interventions, expert review time and support requests. Decide which activities can be removed, which must become routine work and who will pay for them. Test the intended support arrangements before expanding.

Manage changes after release

Equipment configurations, procedures, source systems and AI models will change. Assign owners for updates, test important changes against agreed cases and retain a way to recover or fall back if performance deteriorates.

Separate reusable work from local work

A shared evaluation method or connector may transfer to another site. Local terminology, approval steps and equipment modifications may not. Estimate that adaptation effort before presenting the pilot as a repeatable solution.

Choose a meaningful next test

Select a receiving site or equipment family that differs in a relevant way, rather than repeating the most favourable conditions. Involve the receiving team in defining success. Keep the next commitment small enough to learn from and large enough to expose the real work.

A good result may remain valuable at one site without justifying enterprise-wide rollout. Scale should follow a demonstrated need and a credible operating model.

Make ownership explicit

Maintenance

Defines acceptable output and reviews exceptions.

Operations

Owns how the process fits into daily work.

IT and engineering

Support connections, permissions and changes.

Leadership

Funds the ongoing service and reviews its value.

The evidence points beyond model performance.

RAND: solve the right problem

RAND's 2024 research interviewed 65 experienced AI practitioners. Recurring explanations for failure included misunderstood problems, inadequate data, technology-led projects and insufficient deployment infrastructure. Its scope excluded projects that simply used pretrained language models. [1]

The findings support examining purpose, data and operating conditions. They do not establish a failure rate for maintenance AI pilots or a percentage of failures caused by change management.

Deloitte: adoption and value remain implementation work

Deloitte's 2025 smart manufacturing survey covered 600 executives. It identifies change management and adoption, resource constraints, and value tracking among implementation challenges. The survey also reports training and adoption practices across manufacturers. [2]

For the pilot sponsor, the implication is to budget and test the human and organisational work alongside the software. The survey does not quantify how much any single intervention improves pilot success.

Mettler Toledo: transfer involved people as well as systems

A September 2026 article by Bosch's Rakesh Kumar Murugan, published by the World Economic Forum, describes Mettler Toledo's Changzhou site drawing on internal experts, universities, suppliers and other sites, then extending its transformation to six further locations. [3]

This broader manufacturing example illustrates organised knowledge transfer and collaboration. It is not a controlled evaluation of a maintenance AI pilot. The relevant question is what the next team receives beyond the software.

Read the evidence correctly

Interviews

Identify experienced practitioners' explanations, not population failure rates.

Survey

Shows reported priorities and challenges, not causal effect sizes.

Industrial account

Illustrates a route to replication, not a guarantee of similar results.

Fund the next stage on what the pilot established.

Before the review, agree the evidence required for the decision. Use the table below to expose gaps rather than average them into a reassuring score. A critical unresolved issue may outweigh strong results elsewhere.

QUESTION	EVIDENCE TO BRING	DECISION TO MAKE
Problem	Baseline, affected users and operational consequence.	Is this still the problem worth solving?
Data	Results on representative and difficult records; preparation effort.	Can we maintain the conditions it needs?
Value	Complete-task comparison, quality checks and full costs.	Does the benefit justify the commitment?
People	Observed use across relevant roles, shifts and experience levels.	What must change for normal adoption?
Continuity	Named owners, support, update process and adaptation costs.	Can the next team sustain the result?

Choose the next decision explicitly

Expand when the evidence supports a defined next use. **Adapt and retest** when a specific gap can be resolved. **Keep it local** when value depends on conditions unique to one site. **Stop** when the problem, feasibility or value case does not justify further investment.

For any unresolved assumption, record the owner, the next test and the decision it will inform. The purpose of the review is to decide what to fund next, not to defend the effort already spent.

Evidence behind the paper.

[1] James Ryseff, Brandon F. De Bruhl and Sydne J. Newberry / RAND

The Root Causes of Failure for Artificial Intelligence Projects and How They Can Succeed: Avoiding the Anti-Patterns of AI. 13 August 2024. [Read source](#)

Qualitative interviews with 65 practitioners. The report's cited broad failure estimates are not a measured failure rate for maintenance pilots; this paper does not use them as one.

[2] Deloitte

2025 Smart Manufacturing and Operations Survey: Navigating challenges to implementation. 1 May 2025. [Read source](#)

Survey of 600 executives, conducted August to September 2024. Covers smart manufacturing broadly, rather than only AI pilot conversion.

[3] Rakesh Kumar Murugan / World Economic Forum

How manufacturers can scale AI and digital transformation across every factory. 27 September 2026. [Read source](#)

Practitioner analysis with a Mettler Toledo replication example. Views are the author's; the account is not an independent causal evaluation.

How to use this guide

The five-question framework, illustrations, role comparisons and approval table are NextGenMaint's practical synthesis. They are not a validated diagnostic, a maturity score or a statistical model of project success.

The job-plan assistant is a recurring illustrative scenario, not a reported customer deployment. User needs and acceptance criteria should be established with the people, equipment and operating conditions involved.

The paper does not claim that most maintenance AI pilots fail or assign a universal percentage to the effect of change management. It distinguishes technical feasibility, operational adoption and demonstrated business value.

Prepared October 2026. Public sources reviewed 4 October 2026.