



AI Implementation Thesis

A pilot is not a product demonstration; it is a managed learning loop around a meaningful piece of work.

WORKS WITH THE ASSISTANTS TEAMS ALREADY USE



EDITION 49 / EMBER & SKY

EXECUTIVE WORKING EDITION - 08 OCT 2026

49

AT A GLANCE

A pilot is not a product demonstration; it is a managed learning loop around a meaningful piece of work.

Many AI pilots are demonstrations in disguise: impressive, loosely scoped and impossible to judge. This thesis treats a pilot as a managed learning loop. You contain a specific workflow, design its inputs, roles and quality checks, introduce it with the people closest to the work, and learn from evidence before deciding to extend, pause or redesign. It sets out four stages (Contain, Design, Pilot, Learn), diagnostics across five dimensions, ownership, measures, risks and a first-quarter plan.

WHO IT IS FOR Operations, product and technology leaders and business owners who are turning a promising AI idea into a workflow people will actually use.

TIME TO USE 55 minutes to read; one pilot cycle to apply

What you will have at the end

- A pilot charter specific enough to test and to stop
- A workflow design with inputs, roles, quality checks and exception handling
- Clear business ownership and a review cadence
- An evidence-based decision to extend, pause or redesign

Contents

01	A pilot is a learning loop, not a demonstration	03
02	What most organizations get wrong about implementation	04
03	Four stages: Contain, Design, Pilot, Learn	05
04	Contain: choose something small enough to test	07
05	Design the workflow, not only the tool	08
06	Pilot with the people closest to the work	10
07	Learn, then decide: extend, pause or redesign	11
08	Diagnostics and ownership	13
09	Measuring progress in a pilot	14
10	Risks, trade-offs and when not to pilot	15
11	How Halden works with teams	17
12	A first quarter that produces evidence	18
13	Terms used in this publication	19

THE POINT OF VIEW

A pilot is a learning loop, not a demonstration

01



The point of a pilot is to find out whether a change works in real conditions, for real people, and what it would take to make it work well.

We hold that implementation is mostly about work, not technology. A model can produce useful output in a demonstration in minutes. Turning that into a dependable part of someone's job requires decisions about who uses it, what goes in, who checks what comes out, what happens when it is wrong, and how anyone knows it is helping. These decisions are the implementation.



A pilot is the safest place to make those decisions, provided it is designed to teach. That means a defined user group, a bounded workflow, an agreed outcome, a baseline, a review cadence and a decision at the end. Without those, a pilot becomes an open-ended trial that lingers, accumulates opinions and produces no decision.

We also believe in business ownership. Technology teams are essential, but the person accountable for the outcome should be the person accountable for the work. When the owner is a technology lead without authority over the workflow, adoption stalls even when the system functions. Put the business owner first, with technology as a partner.



- Choose a workflow, not a tool or a department
- Design the checks and exceptions before launch
- Introduce the change with the people closest to the work
- Measure against a baseline taken before the change
- End every pilot with an explicit decision: extend, pause or redesign

A USEFUL TEST

Ask what would make you stop the pilot. If nobody can answer, it is a demonstration, because a real test has outcomes that could go either way.



WORKING QUESTION

What is the workflow you would trust a pilot to teach you about, and who is accountable for it?

COMMON FAILURES

What most organizations get wrong about implementation

02

Pilots fail in predictable ways. Most of them are visible before launch.

The first failure is scope that is too wide. A pilot for a whole department, with several use cases and no boundary, cannot be evaluated. Everyone is partly involved and nobody is responsible. When it ends, it is impossible to say what worked, so the result is a vague sense that it was promising.

The second is the absence of a baseline. Without a record of how long a task took, how often it was wrong and how people rated it before the change, success is a matter of opinion. Enthusiasts claim large gains, skeptics claim none, and the organization cannot adjudicate.

The third is designing for the happy path. The tool works well on typical inputs, and the pilot is declared a success. Real work contains exceptions: missing information, unusual customers, ambiguous requests. If exceptions are not designed for, they surface as errors, workarounds and quiet abandonment.

FAILURE PATTERN	WHAT IT LOOKS LIKE	BETTER APPROACH
Scope too wide	Several use cases, many teams, one pilot	One workflow, one team, one boundary
No baseline	Success judged by impressions	Measure before the change
Happy path only	Typical cases work, exceptions break	Design review and fallback for exceptions

FAILURE PATTERN	WHAT IT LOOKS LIKE	BETTER APPROACH
Technology-owned	Owner has no authority over the work	Business owner accountable, technology partners
No end date	Pilot continues indefinitely	Fixed duration and decision meeting
Users not involved	Tool built for people, not with them	Co-design with the closest practitioners

ILLUSTRATIVE EXAMPLE

An insurer's claims team (illustrative) pilots a summarization assistant. It works for straightforward claims. Complex claims produce inconsistent summaries and handlers quietly revert to manual work. No exception route was defined, so nobody noticed the pattern for weeks.

WORKING QUESTION

Which of these six patterns is already visible in a pilot, or a proposal, you know about?

THE METHOD

Four stages: Contain, Design, Pilot, Learn

03

The stages turn a promising idea into a decision backed by evidence.

Contain chooses a user, workflow, boundary and outcome specific enough to test. Design defines inputs, roles, quality checks, exceptions, information boundaries and a measure of progress. Pilot introduces the change with the people closest to the work and creates space for direct feedback. Learn reviews evidence, improves the workflow and consciously decides whether to extend, pause or redesign.

Our free implementation assessment at halden.company/assessments scores five dimensions: tools, workflow design, integration, quality control and scaling. They are a good way to see whether you are tool-heavy and workflow-light, a very common pattern.



STAGE	CORE QUESTION	MAIN OUTPUT
1. Contain	What exactly are we testing, for whom?	A pilot charter with scope and outcome
2. Design	How will the workflow run and be checked?	Workflow design, quality checks and exception routes
3. Pilot	What happens in real use?	Observations, feedback and a measured comparison
4. Learn	What do we do next?	An evidence-based decision and an improved design



DIMENSION	QUESTION	IF WEAK
Tools	Are tools approved, understood and fit for the task?	Narrow the tool set and give guidance
Workflow design	Is the work defined, with owners and steps?	Document the workflow before piloting
Integration	How is AI connected to systems and data?	Init access and add logging
Quality control	Are review and error tracking in place?	Add checklists and sampling first
Scaling	Is there a route from pilot to rollout?	Define evidence and support needed to scale



WORKING QUESTION

Does your current pilot, or planned pilot, have a written answer to all four stage questions?



STAGE 1: CONTAIN

Contain: choose something small enough to test

04



The best pilots are narrow, real and consequential enough to matter.

Containing means drawing a boundary. Pick one workflow, one user group of perhaps five to fifteen people, and one outcome. The workflow should be frequent enough to generate evidence quickly, important enough that improvement matters, and low enough in risk that errors can be caught and corrected before harm. The charter should fit on one page.



01 Select the workflow

Choose a recurring task with a clear start, finish and owner, ideally from your strategy priorities.

02 Choose the users

Include volunteers and at least one skeptic who knows the work well. Avoid choosing only enthusiasts.

03 Set the boundary

State what is in scope, what is out, which information may be used and which decisions remain wholly human.



04 Define the outcome

Pick one or two measures the business already cares about, such as time per task, error rate or turnaround.

05 Fix the duration

Set a start date, an end date and a decision meeting, typically four to eight weeks.

The pilot charter

- Business owner and technology partner, by name
- Workflow, users and boundary in plain language
- Baseline measures and how they will be taken
- Success, change and stop conditions, written in advance



- Information rules and any approvals needed
- Start date, review dates and the decision meeting



COMMON MISTAKE

Writing a charter that cannot fail. If every outcome counts as success, the charter does not help you decide anything. Write at least one result that would lead you to stop.



WORKING QUESTION

Could you describe your pilot's boundary in three sentences, including what is explicitly out of scope?



STAGE 2: DESIGN

Design the workflow, not only the tool

05

Define who does what, what goes in, how output is checked and what happens when things go wrong.

Design is where most of the real work happens. Start by writing the workflow as it will run: the trigger, the inputs, the step where AI assists, the person who reviews, the output and its destination. Then walk through exceptions: missing inputs, unusual cases, output that fails review, a tool outage. Each needs a defined response, even if the response is simply do it manually and note it.



DESIGN ELEMENT	QUESTION	EXAMPLE ANSWER (ILLUSTRATIVE)
Inputs	What goes in, from where, and is it allowed?	Standard request form; no personal data in free text
Roles	Who prompts, reviews and approves?	Handler prompts; team lead samples reviews

DESIGN ELEMENT	QUESTION	EXAMPLE ANSWER (ILLUSTRATIVE)
Quality checks	How do we know output is acceptable?	Five-point checklist; ten percent sampled
Exceptions	What happens when it does not fit?	Route to manual process and log the reason
Information boundary	What must never be entered?	Defined categories, listed in the charter
Progress measure	How will we see improvement?	Time per task and reviewer rating

Involve the people who will use the workflow in the design. They know where it will break. A one-hour walk-through with three practitioners will often reveal an exception the project team never imagined. Record the changes you make as a result, since they show future teams what good design involves.

Match access to risk. The more an AI-assisted step is connected to real systems, the more review, logging and fallback you need. A drafting assistant that a person reads before use needs light controls. A step that writes to a customer record or sends messages needs stronger ones, and often should not be piloted until simpler steps are proven.

Design review before launch

- The workflow is written down end to end, including who does each step
- Every exception has a defined response
- There is a named reviewer and a sampling plan
- Information rules are in the charter and understood by users
- There is a manual fallback and a way to switch to it
- Users have been trained on the specific workflow, not just the tool

WORKING QUESTION

What is the most likely exception in your pilot workflow, and what does the user do when it happens?

STAGE 3: PILOT

Pilot with the people closest to the work

06

Launch deliberately, watch closely and make it easy for users to say what is not working.

Piloting is a period of intense attention. Hold a short kickoff to explain purpose, boundaries and how feedback works. Check in daily for the first week, then weekly. Sit with users while they work. The most valuable information is rarely in a survey: it is the workaround someone uses without thinking, or the moment where they hesitate.

01 Kick off

Explain the purpose, the boundary, the baseline and the review dates. Make clear that feedback, including critical feedback, is the point.

02 Observe early

In the first week, watch real use and note friction, errors and workarounds.

03 Collect feedback continuously

Provide a simple channel and a weekly ten-minute conversation.

04 Track the measures

Record time, quality ratings, exceptions and incidents in one place.

05 Adjust carefully

Make small changes between reviews and record each, so you know what caused any change in results.

SIGNAL	WHAT IT MIGHT MEAN	RESPONSE
Users revert to manual work	Design does not fit real cases	Sit with users, find the cause, redesign
Reviewers find frequent errors	Inputs, instructions or task suitability	Tighten inputs, add checks, narrow scope
Time saved but quality dips	Checking is being skipped	Reinforce review, adjust workload expectations

SIGNAL	WHAT IT MIGHT MEAN	RESPONSE
Only enthusiasts use it	Training or trust gap	Pair users, address concerns directly
Exceptions are high	Task is not as standard as assumed	Redefine the boundary

COMMON MISTAKE

Changing many things mid-pilot. If you adjust the tool, instructions, scope and users at once, the final result cannot be explained. Change one thing at a time where you can.

WORKING QUESTION

Who will sit with users in the first week, and what will they be looking for?

STAGE 4: LEARN

Learn, then decide: extend, pause or redesign

07

The end of a pilot is a decision meeting, supported by evidence and honest about what it does not show.

At the decision meeting, review the baseline, the measures, the feedback, the exceptions and the incidents. Ask what was learned about the workflow, about the tool, and about the organization. Decide one of three things. Extend if the evidence meets the conditions you wrote in advance. Redesign if the idea is promising but the design needs change. Pause stop if the benefit is not there or the risk is too high.

DECISION	WHEN IT FITS	NEXT STEP
Extend	Success conditions met, risks understood	Define what is needed to scale: training, support, integration

DECISION	WHEN IT FITS	NEXT STEP
Redesign	Promise visible, design flaws clear	Revise workflow and run a short second cycle
Pause	Conditions are not yet right, such as information or capacity	Record what must change and a date to revisit
Stop	Benefit not shown, or risk too high	Document lessons and share them

Stopping a pilot is a legitimate outcome and often a valuable one. Organizations that cannot stop anything learn to avoid honest evaluation. Celebrate a well-run pilot that ended with a clear no, because it saved the cost of a larger mistake and taught something about the work.

Scaling is a separate piece of work. A pilot with fifteen motivated users does not guarantee success with two hundred. Plan for training, support, information upkeep, monitoring, ownership and funding. Treat the first scaled group as another pilot, with its own baseline and review.

ILLUSTRATIVE EXAMPLE

A procurement team (illustrative) pilots draft supplier summaries. At review, time per task fell, but reviewers found recurring errors in contract dates. The team redesigns: dates are entered manually and checked, summaries cover the rest. A second cycle is run.

WORKING QUESTION

What evidence, agreed now, would persuade you to scale, and what would persuade you to stop?

DIAGNOSTICS AND ROLES

Diagnostics and ownership

08

K'

A self-check shows whether your pilot is ready to launch. Named roles keep it from drifting.

Use this checklist before launch and again at the midpoint. Where you answer no, fix it or consciously accept the risk and write that down. The free implementation assessment at halden.company/assessments helps you see patterns such as tools ahead of quality control, or well-defined workflows without a route to scale.

Pilot readiness check

- The workflow is specific, with a start, finish and named owner
- Users, boundary and outcome fit on one page
- A baseline has been taken
- Success, change and stop conditions were written before launch
- Quality checks and sampling are defined
- Exceptions have defined responses and a manual fallback exists
- Information rules are clear to every user
- Access to systems is the minimum needed and is logged
- A review date and a decision meeting are scheduled
- Someone has thought about what scaling would require

ROLE	RESPONSIBILITY	TYPICAL HOLDER
Business owner	Accountable for the workflow and its outcome	Manager of the affected team
Pilot lead	Runs the day-to-day cycle and the review cadence	Operations or project lead
Technology partner	Configures tools, integration and IT, data or an external partner logging	
Quality reviewer	Samples and rates output	Experienced practitioner

ROLE	RESPONSIBILITY	TYPICAL HOLDER
Users	Do the work and give direct feedback	Pilot group members
Risk contact	Advises on information and approvals	Legal, compliance or security lead

In a small organization one person may fill the pilot lead and technology partner roles. That is workable. The role that must not be merged away is the business owner, since it is accountability for the work that makes a pilot a business decision rather than a technical exercise.

WORKING QUESTION

Which role in your pilot is currently held by nobody, or by someone without time for it?

MEASURING PROGRESS

Measuring progress in a pilot

09

Measure a few things that matter, take them before and after, and be careful what you conclude.

A pilot needs measures in three groups. Outcome measures describe the business result: time, cost, turnaround or satisfaction. Quality measures describe output: errors found, reviewer ratings, rework. Adoption and experience measures describe use: how often, by whom and with what comments. Use all three, since an outcome that improves while quality or experience declines is not a success.

GROUP	EXAMPLE MEASURE	COLLECTION METHOD
Outcome	Time per task, before and after	Timing sample of ten to twenty tasks
Outcome	Turnaround time for requests	Existing workflow system

GROUP	EXAMPLE MEASURE	COLLECTION METHOD
Quality	Errors found per ten outputs	Reviewer sampling
Quality	Rework rate	Count of outputs returned for correction
Adoption	Share of eligible tasks using the workflow	Simple log or tally
Experience	User rating and comments	Weekly conversation, short form

Samples will usually be small, so avoid claiming precision you do not have. Report ranges and describe the conditions. If five users measured their own times, say so. Combine numbers with short stories from users, which explain the numbers and reveal things they cannot.

Track incidents separately and openly. An incident log with date, description, cause and action shows whether the controls are working and builds the trust needed to scale. A pilot with no logged incidents may be well controlled, or it may be badly observed. Ask which.

WORKING QUESTION

Which two measures would you want on the table at the decision meeting, and have they been baselined?

RISKS AND TRADE-OFFS

Risks, trade-offs and when not to pilot

10

A pilot has costs and can create risk. Know what you are accepting.

The key trade-off is speed against control. More design, review and logging make a pilot safer and slower. Match the effort to the risk: a drafting aid that a person reads before use needs light controls, while anything touching customers, money, health or employment decisions needs heavy ones and careful professional advice. When in doubt, narrow the pilot until the risk is manageable.

- Automation bias: people trust fluent output more than they should, so keep review real
- Information leakage: pilots often involve data that has not been classified, so settle rules first
- Shadow expansion: a successful pilot spreads informally before it is ready
- Vendor dependence: check what happens to your data and workflow if you change tools
- Legal and regulatory duties vary by sector and region, and may apply to automated or AI-assisted decisions: check current requirements and take qualified advice

WHEN NOT TO PILOT YET

Do not pilot if no one will own the result, if the workflow is undefined, if the information involved is not permitted or reliable, or if the team has no time to participate. Fix those first.

Also be honest when a simpler solution exists. Some problems are better solved by fixing a form, a template or a process than by adding AI. If a pilot succeeds mainly because it forced the team to document the workflow, record that as a finding, and consider whether the documentation alone delivers most of the value.

WORKING QUESTION

What is the worst plausible error your pilot could produce, and where would it be caught?

WORKING WITH HALDEN

How Halden works with teams

11



We help teams contain, design, run and learn from pilots, and we are candid about what the evidence shows.

Halden & Company is an AI advisory and implementation partner. For implementation, we work alongside your business owner and users to write the charter, design the workflow and quality checks, support the pilot and facilitate the decision meeting. We start from your workflow and choose or adapt tools to fit it, not the reverse.



- We can start with the free assessment at halden.company/assessments to see where tools, workflow design, integration, quality control and scaling stand
- We help with charters, workflow design, baselines and review cadence
- We support observation and feedback in the first weeks of the pilot
- We facilitate the decision meeting and record what was learned
- We will say if the evidence says stop, and we will not push you to scale

You may run much of this yourselves using this document. If you would like a partner for a first pilot, or a review of a plan before it starts, reach us at halden.company/contact. A conversation costs nothing, and a candid answer may be that your idea needs a smaller first step.



WORKING QUESTION

What would you want a partner to do in the first two weeks of your pilot that your team cannot?



30 / 60 / 90

A first quarter that produces evidence

12

A first quarter that ends with one well-run pilot and an evidence-based decision.

First 30 days: ground it

- Select one workflow and one user group and name the business owner
- Write the one-page pilot charter with success, change and stop conditions
- Take baselines for time, quality and experience
- Draft the workflow design and exception routes with practitioners

Days 31 to 60: test it

- Complete the design review checklist and fix the gaps
- Train users on the specific workflow and information rules
- Launch the pilot with a kickoff and daily check-ins in week one
- Log measures, exceptions and incidents in one place

Days 61 to 90: decide

- Hold the decision meeting using the pre-agreed conditions
- Record the decision, reasons and lessons
- Redesign, extend or stop, and write what scaling would require
- Share findings with leadership and other teams

Glossary

Terms used in this publication

13



TERM	MEANING
Pilot	A time-limited trial of a change with defined users, scope, measures and a decision at the end.
Pilot charter	A one-page document setting out owner, scope, baseline, conditions and dates.
Workflow	A repeatable sequence of steps by which a piece of work is done.
Baseline	A measurement of the current state taken before the change.
Boundary	The limits of a pilot: what is in scope, what is out and which decisions stay human.
Exception	A case that does not fit the standard workflow and needs a defined response.
Quality check	A review step or sampling routine that tests whether output is acceptable.
Fallback	A manual way to do the work if the AI-assisted step fails or is not suitable.
Business owner	The person accountable for the workflow and for the pilot's outcome.
Automation bias	The tendency to trust fluent machine output more than it deserves.
Scaling	Extending a proven workflow to more users or teams, with training, support and monitoring.
Incident log	A record of errors or near misses with cause and action taken.



A note on use

This publication is a public orientation resource, not legal, technical, financial or professional advice. Adapt it to your people, information, workflows and obligations, and check current tool terms, regulations and risks before acting.



TAKE THE NEXT STEP

Take the free Halden assessment at halden.company/assessments, or talk to us at halden.company/contact about turning this into a plan for your organization.

