Why AI pilots stall
and what to try next



Your team tried AI, so why isn’t it part of daily work? The eight common causes, a five-stage diagnostic and a worked example.

What's inside.

Your AI pilot looked promising. Use this guide to work out why it stalled, gather the evidence and agree what to try next.

01

Understand

Three tests for a pilot 03

What the research says 04

Where to put the effort 05

02

Diagnose

Eight common causes 06

Find the first domino 07

The five-stage diagnostic 08–13

03

Compare

AI+ or AI Native? 14

Six questions to ask 15

Who should be in the room 15

04

Act

Pilot post-mortem 16–17

Agree one experiment 17

About Spark Factory 18

START HERE

If your pilot has already stalled, turn to the diagnostic on page 08.

A pilot can work and still fail.

A tool can meet its technical brief and still fail to become part of daily work. A pilot needs to clear three tests before it earns a place in the business.

01

Technical success

The tool produces outputs that meet the agreed standard.

THE TEST

Is the output correct?

02

Operational success

People can use the tool reliably as part of the real workflow.

THE TEST

Does the team use it every week?

03

Commercial success

The benefit justifies the ongoing cost of running it.

THE TEST

Is the outcome worth the cost?

PASS ALL THREE

A pilot that meets all three tests has a case for surviving the next budget review.

The gap between pilot and scale.

Research points to a consistent challenge: plenty of teams try AI, but fewer can show it working at scale. These studies ask different questions across different samples.

50%+

of generative AI projects were abandoned after proof of concept by the end of 2025, Gartner reports.

Gartner, 2026

~2 in 3

of McKinsey survey respondents said their organisations had yet to begin scaling AI across the enterprise.

McKinsey, November 2025

39%

of respondents in the same survey reported any impact on company-wide earnings before interest and tax (EBIT).

McKinsey, November 2025

95%

of integrated pilots in Project NANDA's preliminary study showed no measurable profit-and-loss impact.

Project NANDA, 2025

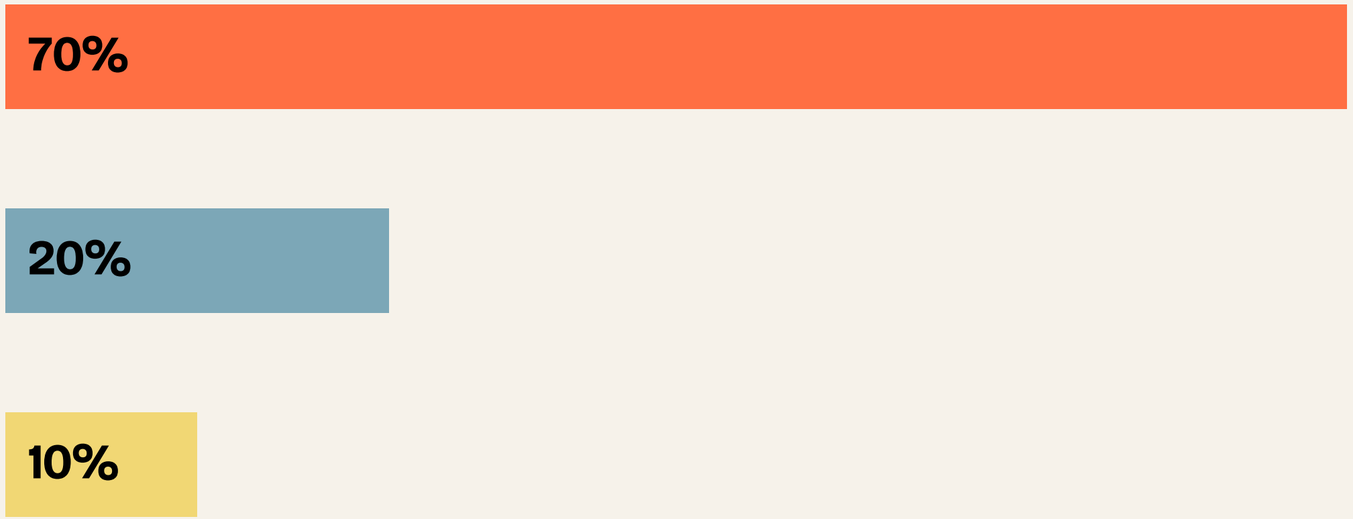
FOR YOUR PILOT

Agree what scale and value mean for your workflow before deciding whether to invest.

Where the effort should go.

BCG's 10–20–70 rule is a resource-allocation guideline. Use it to check whether the plan gives people and processes enough attention.

Source: BCG, Scaling AI Requires New Processes (2026)



People and processes

Redesigning work, training staff and making ownership clear.

Technology and data

Connecting systems, improving data quality and securing access.

Algorithms

Choosing and testing the AI model.

FOR GROWTH COMPANIES

Teams of 50–500 often fit redesign, integration and support around their day jobs.

Eight common causes.

Listen for how people describe the pilot. The symptoms below give you a starting point for investigation.

01 No business priority

SYMPTOM

“We bought Copilot before deciding which business result to improve.”

02 Workflow stayed the same

SYMPTOM

“Drafts got faster, but the same handoffs kept us behind.”

03 Data or integrations were unready

SYMPTOM

“It worked with sample data and struggled with ours.”

04 Staff lacked support

SYMPTOM

“We handed out licences. Managers had no plan to coach people.”

05 No agreed baseline

SYMPTOM

“We could count logins. Finance couldn’t measure the impact.”

06 No business owner

SYMPTOM

“IT got it live. Nobody owned the result after the demo.”

07 Reliability or governance blocked it

SYMPTOM

“Impressive, but we couldn’t trust it with customers.”

08 Too many pilots split funding

SYMPTOM

“We funded several demos without proving which one could scale.”

spark(factory)

Find the first domino.

A pilot can lose support in stages. Trace the chain back to the first failure you can verify.



01

Poor data

Records are incomplete or out of date.

02

Lower accuracy

Outputs need checking and correction.

03

Less trust

Staff stop relying on the tool.

04

Lower adoption

Usage quietly drops off.

05

Unproven return

Funding is cut.

THE LESSON

Low adoption is often the first problem you notice. The evidence can show where the trouble began.

The five-stage pilot diagnostic.

Work from the original goal through to a decision. Bring evidence from the people who do the work, the business owner and Finance.



Reconstruct

Recover the original problem, target and owner. Find out how success was meant to be measured.

PAGE 09

Observe

Watch someone do the work. Record where AI helped and where checking or rework ate the savings.

PAGE 10

Isolate

Trace the visible symptom to the underlying constraint using examples and timings.

PAGE 11

Quantify

Compare the full workflow with the baseline. Include review time and running costs.

PAGE 12

Decide

Choose whether to rescue, redesign, replace or retire the pilot. Name the next test.

PAGE 13

THE AIM

By the end, you should have a cause you can test and a decision that someone owns.

STAGE 01

Reconstruct.

Start with whoever commissioned the pilot, its business owner and the staff who used it.

Five questions to ask

01 What problem were you originally trying to solve?

02 Why use AI for this problem? What else did you consider?

03 What was the intended business outcome?

04 How were you going to measure success?

05 Who was accountable for the result?

Business case check

ROLLOUT OBJECTIVE

“We wanted everyone to have access to Copilot.”

MEASURABLE OUTCOME

“We wanted to cut the time spent preparing sales proposals by 40%.”

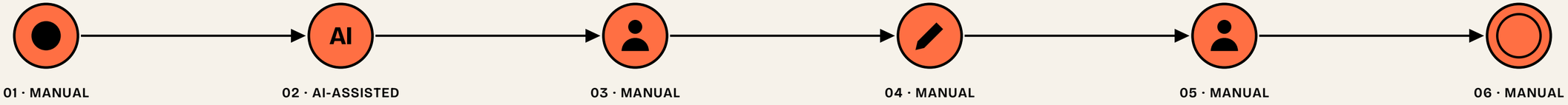
RECORD Write down the starting metric, target and owner before reviewing the pilot’s results.

STAGE 02

Observe.

Ask a staff member to show you the last real task they used the pilot for. Watch what they had to do before and after the AI step.

EXAMPLE · A SALES PROPOSAL, STEP BY STEP



AI speeds up drafting. Time steps 3–6 to see how much of that saving survives checking, correction and approval.

Trigger · AI step · Person checks · Rework · Result

COLLECT THE EVIDENCE Capture full-task timings, usage records, sample outputs, errors and staff or customer feedback.

STAGE 03

Isolate.

Use real examples to distinguish the visible symptom from the constraint underneath it.

WHAT YOU FIND	POSSIBLE CAUSE	WHAT TO TEST NEXT
Nobody uses it	Adoption or workflow fit	Watch the task and look for extra work.
Outputs are often wrong	Data, task fit or reliability	Try representative real examples.
Demo worked, production failed	Integration or controls	Test real systems and permissions.
The value is unclear	Business case or measurement	Rebuild the baseline, costs and result.
Team liked it, rollout stopped	Ownership, governance or budget	Trace who could approve the rollout.
Costs grow faster than value	Unit economics	Model realistic production volumes.

WATCH FOR

People may avoid the tool because it adds work. Watch the task before prescribing more training.

STAGE 04

Quantify.

Measure what people actually did and what the system cost to run. Use the full workflow for the before-and-after comparison.

Net time saved

Old process time − new process time

Count prompting, checking, corrections, escalations and rework.

Monthly net value

Realised benefit − total operating cost

Count saved time when it is redeployed or reduces a real cost.

WORKED EXAMPLE · MONTHLY ADMIN WORKFLOW

400 tasks per month. Each falls from 20 to 15 minutes, including checks. 400 × 5 minutes is 33.3 hours.

A$1,500	−A$1,200	A$300
Capacity value	AI running cost	Net capacity value

At A$45 per hour, assuming the team redeploys the saved time.

BE PRECISE

A$300 a month measures capacity value. Verify actual cash savings separately.

STAGE 05

Decide.

Answer the questions in order. Stop at the first “no” and identify what would have to change.

QUESTION	IF THE ANSWER IS NO	WHAT THAT MEANS
01 Is the original problem still worth solving?	Retire	Stop investing in a low-value, high-risk or uneconomic use case.
02 Is AI the right tool for it?	Replace	Try a better-fitting tool, standard software or a simpler process.
03 Does the workflow around it work?	Redesign	Change the handoffs, review steps or role of the people involved.

IF ALL THREE ARE YES — Rescue. Fix a specific issue with permissions, data quality, training or integration.

WHATEVER YOU DECIDE — Write down the next experiment, its owner, the success measure and the date you’ll review the result.

AI+ or AI Native?

AI+ adds a tool to the work people already do. An AI Native workflow redesigns how the work moves, including when a person steps in.

HOW THE WORK RUNS	AI+	AI NATIVE
What starts the work	An employee starts an existing task.	An event, such as a new request, starts the workflow.
What AI does	Staff choose when to draft, summarise or brainstorm with AI.	Connected systems carry out defined steps using approved data.
What people do	Staff check, edit and finish the work.	People handle exceptions and sign off at agreed points.
Where it fits	Varied, lower-risk tasks where people need flexibility.	Repeatable work with reliable inputs and a measurable outcome.
Sales proposal example	A rep drafts with Copilot, checks pricing and formats it.	A request pulls approved pricing into a draft. A rep reviews before sending.

WHEN TO REDESIGN Redesign the workflow when the bottleneck is the handoff, waiting time or rework around the AI task.

Six questions to ask first.

Bring the business owner, Finance, IT and the people lead. Start with the outcome the pilot was supposed to improve.

THE QUESTION	WHAT IT REVEALS
01 Which business priority was the pilot meant to support?	Whether the original goal mattered.
02 What was the baseline, and what change did Finance expect?	Whether you can measure the result.
03 What changed in the workflow, and where did the saved time go?	Whether faster tasks improved the whole job.
04 When did usage or rollout stall?	Where to start tracing the failure.
05 Would the original problem still be worth solving today?	Whether there is still a worthwhile opportunity.
06 What would you keep, and who would own it?	Which parts to rescue and who takes them forward.

NEXT STEP Use the post-mortem worksheet to turn the answers into one test with an owner and a date.

spark(factory)

AI PILOT POST-MORTEM

Diagnose the pilot.

Then decide what happens next.

HOW TO USE THIS PAGE

01 **Bring the right people.**
Include the pilot owner, Finance and the people who used it.

02 **Watch the real work.**
Write down what you saw on the job.

03 **Mark your confidence.**
Record the evidence behind each finding.

NEXT PAGE

Record your decision and one next experiment.

© 2026 Spark Factory. Free to adapt for your team.

RECONSTRUCT · OBSERVE · ISOLATE WORKED EXAMPLE

ORGANISATION	PILOT	COMPLETED BY	DATE
Wattle Quay Services (fictional)	Maintenance quote follow-up	Dana Lee, Operations	14 May 2027

Original objective 01

Name the measurable outcome, owner, baseline and target.

OUTCOME

Bring follow-up admin from 26 to 12 hours per month. Log at least 75% of eligible follow-ups in the CRM.

MEASURE	BASELINE	TARGET	OWNER
Admin time per month	26 h	12 h	Operations
Follow-ups logged	65%	75%	Sales

What actually happened 02

Observe the work directly. Record where AI entered the process.

STEP	AI / MANUAL	WHAT HAPPENED AFTERWARD
Review open quotes	Manual	Check status and assign owner
Draft reminder	AI	Sales checks dates and terms
Send reminder	Manual	Log the reply in the CRM
Escalate exception	Manual	Owner handles changes

Root cause 03

Record the evidence and mark each finding Confirmed or Suspected.

WHAT YOU FOUND	LIKELY ROOT CAUSE	EVIDENCE	CONFIDENCE
Drafts sat unused	No named owner	18 of 30 unassigned	Confirmed
Duplicate chasing	CRM status lagged	6 duplicate reminders	Confirmed
Wrong due dates	Free-text scheduling	2 dates corrected	Suspected
Replies sat unread	No shared inbox alert	3 replies found next day	Suspected

spark(factory)

AI PILOT POST-MORTEM

Diagnose the pilot.

Then decide what happens next.

HOW TO USE THIS PAGE

01 **Measure the full job.**
Include review time, corrections and operating costs.

02 **Record the decision.**
Choose one path and explain what the evidence supports.

03 **Run one experiment.**
Name an owner, a success measure and a date.

NEXT STEP

Keep the next experiment small and give someone ownership.

© 2026 Spark Factory. Free to adapt for your team.

The numbers

04

Use comparable monthly figures. Include checking and rework in the new process time.

OLD PROCESS TIME	NEW PROCESS TIME	NET TIME SAVED
26 h	22 h	4 h

REALISED BENEFIT	TOTAL RUNNING COST	MONTHLY NET VALUE
A$180	A$260	-A$80

Count saved time if it is redeployed or becomes a cost reduction.

The decision

05

Choose the option your evidence supports and explain why.

Retire · stop investing

Replace · choose a different tool

Redesign · change the workflow

Rescue · fix a specific issue

DECISION

Redesign

BECAUSE

Fix quote ownership and CRM status before adding automation. A person still approves every reminder during the trial.

Next experiment

06

Keep the test small enough to run. Agree who owns it, what success looks like and when you'll make the call.

Trial 60 open quotes in one region. Draft one reminder after three business days. Sales checks every draft. Stop on a duplicate or unauthorised send. Compare admin time and follow-ups logged with the baseline.

OWNER	SUCCESS MEASURE	GO OR NO-GO DATE
Dana Lee	≤12 h/month; ≥75% logged	14 June 2027

We build AI workflows.

We run them, too.

Spark Factory finds, builds and runs AI workflows for growth companies. We take on one workflow at a time, with a result your team can measure.

01

Find.

Find the repeatable job where AI can improve cost, speed or quality.

02

Build.

Build into the tools you use and test against real work. Fixed price, live in four weeks from sign-off.

03

Run.

Monitor, support and improve the workflow while it runs, for one monthly fee.

YOUR NEXT STEP

Bring one workflow.

Show us one job your team repeats. We'll help you choose the first workflow to trial.

Scan to book a call.

www.sparkfactory.ai

spark(factory)

© 2026 Spark Factory AI Pty Limited. Examples and figures are illustrative.