

A FIELD GUIDE TO AI PILOT TRIAGE

Kill or *Ship.*

How to decide which AI pilot survives the next quarter.

88%

of AI pilots never reach
production

42%

of companies abandoned AI in
2025

\$18B

in AI spend written off,
2023–2025

The hardest call *in AI right now.*

In 2023, 17% of companies had abandoned an AI initiative. By 2025, the number was 42%. In the same window, IDC found that 88% of AI pilots never reach production. The pilots that did ship rarely failed because of the model. They failed because nobody was paid to kill the seven that should have been killed first.

This guide is the diagnostic NS Studio runs in paid Pilot Triage engagements. The same seven signals. The same decision tree. The same kill list. Read it in 18 minutes. Apply it before next quarter's planning cycle.

If you score three or more pilots as kills, that's your next quarter's work. Most organizations cannot kill a pilot internally — the political cost of saying "this should stop" is paid by whoever says it first. An external advisor is paid to absorb that cost. That's the entire economics of the engagement.

"Pertama Partners killed seven of nine pilots for a North American utility. The surviving two delivered \$180,000 in 90 days."

What's *inside.*

- 01 **The seven kill signals** _____
The rubric. Score every active pilot in under an hour.
- 02 **The decision tree** _____
Kill, ship, escalate. One page. Use in a meeting.
- 03 **Four cases, anonymized** _____
What the rubric returned. What the org did next.
- 04 **Governance — who actually pulls the trigger** _____
The executive question every pilot needs an answer to.
- 05 **The scoring worksheet** _____
Print this. Bring it to your next QBR.
- 06 **Next step — the Pilot Triage Sprint** _____
Two weeks. Fixed price. A kill list you can defend.

The seven *kill signals.*

For each active pilot, score every signal as PRESENT or ABSENT. Three or more PRESENT means kill. The rubric is calibrated against engagements where the alternative was to keep paying a vendor through year two.

01 No defensible dollar by week six.

WHAT IT LOOKS LIKE

The pilot is six weeks in. The team can describe what it does, but not the dollar figure it moved. "Promising" appears in the latest update. The CFO has not asked yet, but will.

KILL THRESHOLD

If you cannot write the one-page memo with a specific number, kill or restart with a named target. "Promising" is a stall, not a status.

02 No nameable internal complainer.

WHAT IT LOOKS LIKE

Ask: who, by name, escalates if this project pauses for 90 days? Not a persona, not a department, not "the business." If three people on the team can't agree on the name, the project has no oxygen.

KILL THRESHOLD

Nameable complainer required within two weeks. If none surfaces, you're funding a vitamin that survives because killing it feels rude.

03 Six executives, zero owners.

WHAT IT LOOKS LIKE

The steering committee has six VPs. The pilot has none. Decisions get deferred to the next steering. The next steering defers to a working group. Six months in, no single person can be fired for this project's outcome.

KILL THRESHOLD

If the kill decision requires a meeting of more than two people, ownership is missing. Assign a single accountable executive within seven days, or kill.

04 Switching cost is unknown.

WHAT IT LOOKS LIKE

The renewal arrives. Procurement asks if it's worth it. Three teams say yes. Nobody can describe what migration would actually take — prompts, evals, integrations, internal champions. So the renewal goes through. Again.

KILL THRESHOLD

If exit takes more than 90 days or has never been measured, the vendor is locked in. Stop renewing on autopilot. Run the exit drill before the next renewal cycle.

05 The roadmap is a screensaver.

WHAT IT LOOKS LIKE

There is a 47-slide three-year AI roadmap. It has not changed in nine months despite GPT-5, Claude 4, two new vendors, and a board mandate. The deck is presented quarterly. No engineer's calendar changes the following Monday.

KILL THRESHOLD

If the roadmap survives a major model release without revision, it's theater. A real roadmap is a working document with kill criteria, not a maturity curve.

06 Data readiness is a pending ticket.

WHAT IT LOOKS LIKE

The pilot needs three data sources. Two have access. The third has been "in security review" for eleven weeks. The team has stopped asking. Engineering has built around the missing source with synthetic data that nobody trusts.

KILL THRESHOLD

Data access blocked beyond 30 days = kill or escalate. Pilots that demo on synthetic data don't ship. They graduate into production with the same gap and it bites in week three of go-live.

07 Nobody can describe success in one sentence.

WHAT IT LOOKS LIKE

Ask three people on the team what success looks like in 90 days. You get three different answers, all hedged with "and / or." The pilot is being judged on whichever answer the next slide can support.

KILL THRESHOLD

Single-sentence success criterion required, written and signed by the accountable executive, before week four. If the criterion shifts more than once after that, the pilot is being optimized for survival, not outcome.

"Three or more signals present = kill. The rubric is not a suggestion."

Kill, ship, *or escalate.*

Run every active pilot through this tree. The output is one of three verdicts: ship, escalate, or kill. Defer is not a verdict.

PILOT TRIAGE DECISION TREE

<question_1>

Has the pilot produced a defensible dollar figure in the last 60 days?

no → go to question 2 (tentative: kill)

yes → go to question 3 (tentative: ship)

</question_1>

<question_2>

Is there a single accountable executive who would be fired if this fails?

no → verdict: kill (no owner = no ship)

yes → verdict: escalate (30 days, re-score)

</question_2>

<question_3>

Is the success criterion a single sentence signed by the accountable executive?

no → verdict: escalate (define first)

yes → verdict: ship (move to production)

</question_3>

01 / Nine pilots, two survivors.

SITUATION

A North American utility ran nine concurrent AI pilots across customer service, field operations, predictive maintenance, and analytics. Twelve months in, total spend was \$4.2M. None had reached production. Pertama Partners was brought in to triage the portfolio.

RUBRIC SCORE

Seven of nine pilots scored three or more kill signals. The two survivors — a predictive-maintenance model on transformer assets and an outage-classification model in the contact center — had nameable executive owners, single-sentence success criteria, and live data access.

VERDICT

Kill 7 of 9. Ship 2.

OUTCOME

The two surviving pilots delivered \$180,000 in confirmed savings within 90 days of the triage. The seven killed pilots returned approximately \$2.1M in budget to next year's planning cycle. Public case study at [pertamapartners.com / insights / enterprise-ai-abandonment-2025](https://pertamapartners.com/insights/enterprise-ai-abandonment-2025).

02 / The retail conglomerate.

SITUATION

A Fortune 100 retailer had a generative-AI pilot for product copy. Eight months in, the model was working. The steering committee had eleven members across marketing, merchandising, legal, and three regional groups. Every shipping decision was deferred to the next monthly steering.

RUBRIC SCORE

Signals 01, 03, and 07 present. Defendable dollar figure existed (a 30% reduction in copywriting time). Owner did not. Success criterion shifted three times in eight months, each shift defended by a different VP.

VERDICT

Escalate. Single owner, 30-day deadline.

OUTCOME

Recommendation was to assign a single accountable VP and dissolve the steering. The org's actual decision: keep the steering, kill the pilot. Six months later, a competitor shipped the same capability and the retailer issued an RFP to license it.

03 / The fintech vitamin.

SITUATION

A mid-sized fintech had a "strategic" AI initiative for personalization. Eighteen months in, three engineers full-time, no shipping path. The deck described it as foundational. The deck had survived two reorgs and three CEOs.

RUBRIC SCORE

Signals 01, 02, 05, 06, and 07 all present. No defensible dollar. No nameable complainer. Roadmap unchanged through two model generations. Data access pending. Success defined differently by every team member asked.

VERDICT

Kill. Reassign engineers within two weeks.

OUTCOME

The kill conversation took 40 minutes with the new CEO. The three engineers were moved to a fraud-detection pilot that shipped in eight weeks and saved \$2.3M in year one. The personalization deck was deleted from the shared drive.

04 / The healthcare network.

SITUATION

A regional healthcare network ran a clinical-documentation copilot pilot for 14 weeks. Defendable dollar figure: 18 minutes saved per physician per shift, translating to \$4.1M annualized. Nameable complainer: the chief medical officer who had championed it.

RUBRIC SCORE

Zero kill signals present. Owner: CMO. Sentence: "Reduce after-hours documentation by 30 minutes per physician shift, year one." Vendor exit cost: 60 days, documented. Data access: live. Success: single sentence, signed.

VERDICT

Ship. Move to production within 60 days.

OUTCOME

The pilot shipped on time. The first three months in production confirmed the rubric: actual savings hit 22 minutes per shift, ahead of target. The pilot is now the network's reference deployment. Triage took four hours; spared the org the default reflex of a third pilot phase.

Who actually *pulls the trigger.*

The rubric is the easy part. The hard part is who owns the kill decision and how it gets executed without burning the engineering team that built the doomed pilot.

Three rules. They are not optional.

01 One executive, named in writing.

Every active pilot has a single accountable executive whose name is in the project charter. If two names are listed, neither is accountable. If no name is listed, the pilot is killed by default at the next quarterly review.

02 The kill memo, not the kill meeting.

Killing in a meeting invites politics. The kill is announced via a one-page memo sent to the steering committee 48 hours before the meeting. The meeting confirms execution, not the decision.

03 Engineers move within two weeks.

The team that built the killed pilot is reassigned within two weeks to a project with three or fewer kill signals. The org cannot afford to leave senior engineers in mourning. Velocity costs more than the pilot did.

The kill *list.*

"Every Pilot Triage engagement ends with a kill list. Not a deprioritize. Not a revisit-next-quarter. Kill."

Explicit kills matter because the alternative is implicit drift. A project on the deprioritized list still consumes attention; it lingers in standups, in roadmaps, in the back-of-mind of the engineer who built it. A killed project frees the calendar.

KILL LIST / COMPOSITE — Q4 SAMPLE

- × ~~Vendor X enterprise renewal — ROI undefended after 9 months~~
- × ~~Internal LLM fine tuning — duplicates a vendor offering~~
- × ~~Personalization v2 — v1 never shipped, no nameable complainer~~
- × ~~AI center of excellence weekly — zero decisions logged in Q3~~
- × ~~Generative search prototype — success sentence shifted four times~~
- × ~~Compliance copilot — data access pending eleven weeks~~
- × ~~Synthetic data demo of fraud model — production data still blocked~~

Score every *pilot.*

Print this page. One per pilot. Mark each signal PRESENT or ABSENT. Three or more present = kill. Two = escalate. One or zero = ship.

Pilot name: _____

Accountable executive: _____

#	SIGNAL	PRESENT	ABSENT
01	No defensible dollar by week six	<input type="checkbox"/>	<input type="checkbox"/>
02	No nameable internal complainer	<input type="checkbox"/>	<input type="checkbox"/>
03	Six executives, zero owners	<input type="checkbox"/>	<input type="checkbox"/>
04	Switching cost is unknown	<input type="checkbox"/>	<input type="checkbox"/>
05	Roadmap unchanged through major model release	<input type="checkbox"/>	<input type="checkbox"/>
06	Data access pending more than 30 days	<input type="checkbox"/>	<input type="checkbox"/>
07	Success criterion is not a single sentence	<input type="checkbox"/>	<input type="checkbox"/>

Total PRESENT: _____

0-1 = SHIP 2 = ESCALATE 3+ = KILL

NOTES

Apply this *next quarter.*

01 List every active pilot.

All of them. Including the ones you forgot. Including the personal-card subscriptions. If the list runs over twelve, the problem is bigger than the rubric.

02 Score in one sitting.

Block 90 minutes. Two people, ideally one who didn't sponsor any of the pilots. Score honestly. The rubric is worthless if scored generously.

03 Write the kill memo first.

Before the kill meeting. Before the steering. The memo names the pilots, the signals scored, and the reassignment plan for the engineers. Sent 48 hours ahead.

04 Reassign within two weeks.

Engineers move to projects with one or zero kill signals. The org cannot afford to leave senior people in mourning. Velocity costs more than the pilot did.

"If three or more pilots score kill, that's your next quarter's work. Stop, don't add."

When you want *us to run it for you.*

The Pilot Triage Sprint is a two-week, fixed-price engagement. You send the pilot inventory. We run the rubric, conduct executive interviews, and return a defended kill list with reassignment recommendations. Founder-led. No junior consultants on client work.

WHAT YOU GET

- A kill / ship / escalate verdict for every active pilot
- Scored rubric for each, with named evidence
- A defended kill memo — ready for your steering committee
- Engineer reassignment plan, two-week execution path
- One 60-minute executive readout (recorded)

Pilot Triage Sprint

Two weeks · fixed scope · founder-led

\$2,500

FIRST-MONTH RATE

Book the Sprint

→ [nsstudio.com / sprint](https://nsstudio.com/sprint)

Colophon.

ABOUT THIS GUIDE

This is Field Guide 02, written by Simrat Bath, founder of NS Studio LLC. Updated quarterly. Distributed free, gated only by email. The text is set in Instrument Serif and Inter; the prompt cards in JetBrains Mono. The four case sketches are anonymized composites drawn from public reporting and engagement patterns.

ABOUT NS STUDIO

NS Studio LLC is a boutique AI product advisory based in California, working with Fortune 500 product and engineering teams. Three engagements per quarter, founder-led. We tell you which AI to build, which to kill, and how to ship the rest.

SOURCES

CIO.com / IDC – 88% of AI pilots fail to reach production

<https://www.cio.com/article/3850763/88-of-ai-pilots-fail-to-reach-production-but-thats-not-all-on-it.html>

S&P Global – 42% of companies abandoned AI in 2025

https://www.linkedin.com/posts/sammy-abdullah-1292494_companies-are-pouring-billions-into-ai-activity-7361370556123766784-HgNm

Pertama Partners – Enterprise AI abandonment, 2025

<https://www.pertamapartners.com/insights/enterprise-ai-abandonment-2025>

CONTACT

hello@nsstudiollc.com

nsstudiollc.com / sprint

calendly.com / simratbath / 30min

linkedin.com / in / simratbath

<NS />

NS STUDIO LLC · CALIFORNIA · 2026