



BURAK GALİBA
Thoughts on AI Transformation

ESSAY · ENTERPRISE AI

Pilot Purgatory — Why Your AI Demos Never Ship, and How to Escape

Your AI demos impress everyone and ship nothing. Why pilots stall between demo and production — and the triage that gets one live in 90 days.

PILOTS

TRANSFORMATION

ROI



Burak Galiba · May 26, 2026 · 3 min read

<https://burakgaliba.com/blog/escaping-pilot-purgatory>

SCAN TO READ ONLINE

You know the symptoms. Twelve AI pilots, three innovation awards, a steering committee that meets monthly — and nothing in production. [Every demo impresses](#). Nothing ships.

That's pilot purgatory, and it isn't a phase you grow out of. It's a stable equilibrium. Left alone, a pilot portfolio doesn't mature into production — it just becomes [a larger pilot portfolio](#). The NANDA report makes the point at scale: [95% of enterprise GenAI pilots never produce measurable P&L impact](#).

The good news: purgatory has an exit, and it's not where most teams are looking.

How you got here (it was rational)

Nobody sets out to build a science fair. Purgatory is the sum of individually sensible choices.

Pilots got funded from innovation budgets, so they never needed a business case. They were scoped to demo well, so they skipped integration. They ran on curated data, so nobody found the real data problems. And they were "just pilots," so security, legal, and the process owner were politely not invited.

Each choice made the pilot faster to stand up. Together, they guaranteed it could never ship — because everything deferred is exactly what production requires.

The five gaps between demo and production

Triage a stalled portfolio and the same gaps appear in nearly every stuck pilot:

- **The integration gap.** The demo runs beside the workflow; production must run inside it — wired into the ERP, the claims system, the queue people actually work from.
- **The evaluation gap.** A demo needs a happy path. Production needs evals: accuracy thresholds, failure-mode testing, and a defined behavior for "the model isn't sure."
- **The accountability gap.** No named owner whose budget feels the result. Demos can be orphans; production systems can't.
- **The risk gap.** No security review, no data-protection answer, no governance sign-off. This gap is now scheduled to hurt: Gartner expects [over 40% of agentic AI projects to be canceled by end-2027](#), with weak risk controls a leading cause.
- **The adoption gap.** Nobody redesigned the job the AI touches. Software that asks people to change how they work, without anyone managing that change, gets ignored to death.

Note what's absent: model quality. In three years of this pattern, the model is almost never the blocker.

The escape route: triage, then ship one thing

You don't escape purgatory by improving all twelve pilots. You escape by shipping one — and being ruthless about which one.

1. **Triage the portfolio: ship, fix, or kill.** For each pilot, one page — owner, metric, baseline, blocker. "Ship" means a credible 90-day path to production. "Fix" means real value with one named blocker. "Kill" means everything else, and it should claim at least a third of the list.
2. **Pick one ship candidate.** Not the most impressive — the most measurable. High volume, clear success criteria, an owner who wants it. That's usually a back-office workflow, which is precisely where that report found the best returns hiding.
3. **Write the production spec the pilot skipped.** Integration points, eval thresholds, guardrails, escalation path, rollback plan, and the metric with its baseline. If this document doesn't exist, you have a demo with tenure.
4. **Ship it with a hard gate.** A go-live date, go/no-go criteria, and the owner's name on both. Ninety days is enough for a well-chosen use case, and a deadline is what separates production from purgatory.

What shipping one thing buys you

More than one workflow's ROI. The first production deployment breaks the equilibrium.

It gives finance a real number instead of projections, which changes every future funding conversation. It forces the data, security, and governance answers that the next use case inherits at a discount. And it converts AI from a topic into an operating capability.

The second deployment is measurably cheaper than the first. The tenth is routine. That compounding is the entire difference between the enterprises that escaped purgatory in 2025 and the ones still hosting demo days in 2026.

Where to start

Run the one-page triage this week — every pilot, four fields: owner, metric, baseline, blocker. The portfolio will sort itself with uncomfortable speed, and you'll know your ship candidate by Friday.

If you want an outside read before you triage, my free [AI Readiness Score](#) takes ten minutes and 20 questions to show you where your program stands — including whether your pilot portfolio looks like a pipeline or a purgatory. The score comes with your first three moves, which usually includes naming the one pilot worth shipping first.

Read the essay online at <https://burakgaliba.com/blog/escaping-pilot-purgatory>

I built a free AI Readiness Score — 20 questions across pilots, data, talent and governance, about ten minutes, no email required: <https://burakgaliba.com/readiness>

© 2026 Burak Galiba · <https://burakgaliba.com> · Views are my own.