BGL BURAK GALİBA
Thoughts on AI Transformation

ESSAY · ENTERPRISE AI

Why 95% of AI Pilots Fail — and What the 5% Do Differently

An MIT NANDA report found 95% of enterprise GenAI pilots deliver zero P&L impact. The failure is organizational, not technical — here's what the other 5% do.

PILOTS AI STRATEGY TRANSFORMATION

Burak Galiba · May 5, 2026 · 4 min read
https://burakgaliba.com/blog/why-95-percent-of-ai-pilots-fail

SCAN TO READ ONLINE

In August 2025, a research project at MIT put a number on what most executives already suspected: **95% of enterprise GenAI pilots deliver no measurable P&L return.**

Not disappointing returns. Not slow returns. No measurable impact at all, in 19 out of 20 attempts, as Fortune reported when the study landed.

One caveat, since I lean on this number a lot: it's a project report out of MIT's NANDA initiative — roughly 300 public deployments, 52 interviews, 153 survey responses — not peer-reviewed research. Treat it as a well-built field survey, not a law of physics. I keep citing it because it matches what I've watched happen, not because the number is sacred.

Nine months later, that is still the most useful sentence in enterprise AI. Because once you accept it, the question stops being "which model should we buy?" and becomes "what exactly is the 1 in 20 doing that we're not?"

It was never a technology problem

The report is called "The GenAI Divide," and the headline stat wasn't even the important part. The cause was.

The models weren't failing. The organizations were. The researchers described a "learning gap": enterprises deploy tools that don't adapt to their workflows, hand them to teams with no mandate to work differently, and then wait for a miracle that was never on the roadmap.

Follow-up analysis in Forbes put it more bluntly: companies fail because they avoid friction. They buy AI that bolts on politely instead of AI that changes how work gets done. Polite AI produces polite results — a demo, an internal newsletter mention, and a line item that quietly vanishes next budget cycle.

Run the test on your own portfolio: how many of your current pilots required anyone to change how they work? If the answer is zero, you already know your ROI.

The budget goes where the demo shines

The report's second finding deserves far more airtime than it gets. Enterprise AI budgets skew heavily toward sales and marketing — while the strongest measured returns come from back-office automation.

The logic is embarrassingly human. Front-office AI demos beautifully in a board meeting. Back-office AI — invoice matching, claims intake, document processing, reconciliations — demos like a spreadsheet.

But the back office has the three things AI ROI actually requires: repetitive volume, clean success criteria, and a cost line someone already owns. The board demo has none of them. The 95% keep funding charisma; the 5% fund arithmetic.

What the 5% do differently

The successful minority isn't luckier, and it isn't better funded. The same behaviors show up in the research and in every stalled portfolio I've seen:

- **One P&L owner per use case.** A named executive whose budget feels the result. Committees pilot; owners ship.
- **A baseline before the build.** Cycle time, cost per transaction, error rate — captured before the pilot starts. No baseline, no ROI claim. Ever.
- **A back-office bias.** They go where the work is repetitive and measurable, not where the demo is prettiest.
- **Workflow redesign, not tool overlay.** They accept the friction everyone else avoids. The process changes, roles change, and the AI becomes load-bearing.
- **Kill criteria on day one.** Every pilot starts life knowing the conditions under which it dies. That's what keeps the portfolio honest.
- **Change management staffed as the project, not the postscript.** If failure is organizational, the winning move is organizational too.

Notice what's missing from that list: model selection, GPU procurement, prompt engineering. Nothing about the 5% requires better technology than you already have.

The 5% is a decision, not a lottery

Here's the uncomfortable implication. If failure were technical, you could wait — next year's models would rescue you. Since failure is organizational, waiting just makes you a better-funded member of the 95%.

The divide is made of decisions: who owns each use case, what gets measured, what ships first, what dies. Most enterprises haven't assigned those decisions to anyone. Which means the 95% isn't happening to them. It's being chosen, one unowned pilot at a time.

And the stakes compound. The organizations that shipped one real workflow in 2025 are shipping their third in 2026, each one cheaper than the last. The learning gap doesn't stay the same size — it widens.

Where to start

Take one page. List every AI initiative currently running, with four columns: executive owner, business metric, baseline value, target production date. If columns two and three are mostly blank, you've found the actual problem — and it isn't the technology.

For the fuller version of that exercise, my free AI Readiness Score asks 20 questions across pilots, data, talent, and governance. It takes about ten minutes, and it will tell you — with a score and a domain-by-domain breakdown — which side of the 95% you're currently on.

Read the essay online at https://burakgaliba.com/blog/why-95-percent-of-ai-pilots-fail

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