

Why 95% of AI pilots fail — and what the other 5% do differently.

A ONE-PAGE FIELD NOTE FROM THE SOULWARE TEAM. NO FLUFF.

Between 2023 and 2025, European companies poured an estimated €30–40 billion into generative AI. MIT looked at where it landed. The answer stings: 95% of initiatives produced no measurable return. Not because the models are weak — they're not. Because the pilot was set up to fail before the first prompt.

95%

of enterprise GenAI pilots return zero measurable P&L impact.

MIT NANDA, 2025

5%

of companies are extracting real value from AI. The gap is widening.

BCG, 2025

30%

of GenAI projects will be abandoned after PoC by end of 2025.

Gartner, 2024

THE FIVE WAYS PILOTS *die*

01 Built for the demo, not the desk.

The pilot runs beautifully on clean test data, then meets real inboxes, messy PDFs and legacy exports — and breaks.

02 No one owns the outcome.

"Success" was never defined in a number, so no one can say if it worked. The project drifts until the budget runs out.

03 Chosen for the vendor, not the workflow.

A tool bought top-down by IT, without the team doing the actual work. They go back to ChatGPT in a private tab — MIT calls this the "shadow AI" gap.

04 Static where the work is living.

The model doesn't learn from corrections, doesn't retain context, doesn't get smarter with use. MIT's core finding: this "learning gap" — not model quality — is what kills 95% of pilots.

05 Stuck in PoC purgatory.

Security, compliance and IT were invited to the last meeting instead of the first. The pilot works; it just never ships.

WHAT THE 5% DO DIFFERENTLY

01 They start with a time audit, not a use-case list.

Where is an hour leaking, every day, for someone specific? That's the pilot. Not "let's do AI for marketing."

02 One workflow. One team. One owner.

Not a platform. Not a transformation programme. One person whose week gets visibly better in six weeks.

03 Ship to production in weeks, not quarters.

The 5% treat the first version as a real tool the team uses Monday — not a demo for the steering committee. Iterate against real inputs from day one.

04 Measure output quality from the start.

Not just "did it run." Is it right more often than last week? Are corrections getting rarer? BCG found the winners scale fast *because* they measure operational metrics, not usage stats.

*The gap between the 5% and the rest is not talent or budget – it's **how the first six weeks were set up**. If you're planning an AI pilot this quarter, we'd rather help you scope it right than debug it later.*

TALK TO US

hello@soulware.be

Sources: MIT NANDA, *The GenAI Divide: State of AI in Business 2025* · BCG, *Closing the AI Impact Gap, 2025* · Gartner, press release, *30% of GenAI Projects Abandoned After PoC by End 2025*, July 2024.

