95% of AI Pilots Fail? Here's the 90-Day Fix That Actually Works

MIT just published what your CFO suspected: 95% of AI pilots are expensive theatre.

While tech stocks tumbled and MIT revealed that 95% of GenAI pilots deliver zero P&L impact, the real story isn't about AI failing, it's about basic project discipline disappearing. The turmoil is with the hasty implementers, not the careful builders.

My view is that much of this outcome was avoidable with basic project portfolio discipline: clear business plans, cost-and-benefit tracking, and the courage to stop or pivot early. No magic, just PMO and PPM principles.

The Root Cause

Many of these pilots are really just demos. They sit on the side of a process, they do not change how work actually flows. They do not write back to the system of record. They do not remember past cases or learn from outcomes. Most importantly, they do not link to a specific GL account. If a tool does not move a GL line, in other words, if it doesn't reduce costs or increase revenue in your accounting system, it's not real value.

Rushing because of AI FOMO (fear of missing out) makes this worse. Teams pick unrealistic use cases, or too many at once. Budgets get stretched, but benefits get reported as "time saved" anecdotes, not as hard entries in the ledger.

What a Value-Centred PMO/PPM Changes

1) Name the money up-front

Every pilot must have a business case that states the GL line it intends to move and approximately by how much. Example: "Reduce contractor costs £500K per year, reduce processing errors by 40%, within 12 weeks." If the GL line is unclear, the work should not start.

2) Build a Minimum Viable Workflow, not a chatbot

Wire the tool into the process end-to-end. Read from source systems, write back decisions, and keep an audit trail. Add basic guardrails and evaluation checks so reliability, efforts, and improvements are measurable, not assumed.

3) Require state and memory

If the system never learns from prior decisions, value will not compound. Memory turns single-shot demos into repeatable improvements. A customer service bot that remembers previous interactions can reduce handle time by 60%. One that starts fresh each time? Just an expensive FAQ.

4) Limit Work in Progress (WIP)

Run three to five use cases at a time. Finish or stop before starting new ones. This forces focus and exposes weak bets early.

5) Give Finance shared ownership

Benefits must be co-signed by Finance every month. No soft proxies. Show the ledger movement: supplier invoices avoided, unit cost down, write-offs reduced, or revenue up.

6) Pre-agree kill gates

No wishful thinking. Set clear goals:

- Week 4: <30% of target = PIVOT
- Week 8: <50% of target = STOP
- Week 12: ≥70% of target = SCALE

This saves months and reputational capital.

A Simple Example: KYC Remediation (Know Your Customer)

Before: An "AI assistant" answers questions in chat. Nice demo. No write-back. No memory. Analysts still re-type data.

- Time per case: 45 minutes
- P&L movement: £0

After: The tool pulls documents, pre-populates fields, logs reasons, writes to the case system, and flags exceptions for review.

- Week 2: baseline minutes per case agreed with Finance

- Week 6: verified unit-cost reduction on live traffic (28 minutes per case)
- Week 12: decision to scale based on P&L evidence (£2.1M projected annual savings on 50,000 cases)

The 90-Day Reset Leaders Can Copy

Weeks 0–2

Pick three workflows with clean cost baselines. For each, write one sentence: "We will move GL account X from £A to £B by week 12." Set the kill thresholds and name one accountable owner.

Weeks 2–6

Build the Minimum Viable Workflow. Integrate one system of record. Add guardrails, evaluation metrics, and audit logs. Start weekly Finance sign-offs.

Weeks 6–12

Run A/B tests on real volume. Scale only pilots that hit ≥70% of target. Stop the rest. Reallocate capacity to winners. Keep WIP tight.

Why This Matters Beyond One Firm

Stakeholders are starting to ask for evidence of value, not just usage metrics (logins, queries, user counts). If enterprises slow spend until benefits are proven, markets will keep rewarding teams that can show ledger-level impact and discipline. That's healthy. It shifts the narrative from hype to economics.

Having led PMO transformations through multiple technology hype cycles, I've seen this pattern before. The winners are always those who maintain discipline when others chase trends.

How This Article Was Produced (Transparency)

I drafted this by:

1. Reviewing public reporting, including the Financial Times piece on the AI-led sell-off and the MIT analysis of GenAI pilot outcomes.
2. Using an AI assistant (ChatGPT, GPT-5 Thinking) to check phrasing, test logic, and challenge assumptions.
3. Applying practical experience from PMO and PPM leadership in financial services to outline a workable, value-centred approach.
4. Structuring the article for non-specialists, expanding acronyms, and focusing on steps any organisation can execute in 90 days.
5. Reviewed and refined with Claude (Anthropic) to ensure clarity, test logical flow, and enhance readability for a business audience.

Call to Action

Resist AI FOMO. Start fewer pilots. Tie each one to a GL line. Add memory. Limit WIP. Share ownership with Finance. Use kill gates.

Try this for one quarter and watch your success rate flip from 5% to 70%. Which of your current AI pilots would survive the GL test?

Disclaimer and Disclosure

Third-party Content and AI Assistance: This article references tools and software that are publicly available and proprietary to their respective creators. The author does not claim ownership or affiliation with these third-party products. This article was written by the author with assistance from Generative AI Language Models.

Transparency Notice: While every effort has been made to ensure accuracy, readers should verify information independently and consult official sources or documentation for the mentioned tools and software. The use of AI in the writing process is disclosed in the interest of transparency, but all opinions and analyses are the author's own unless otherwise stated.