

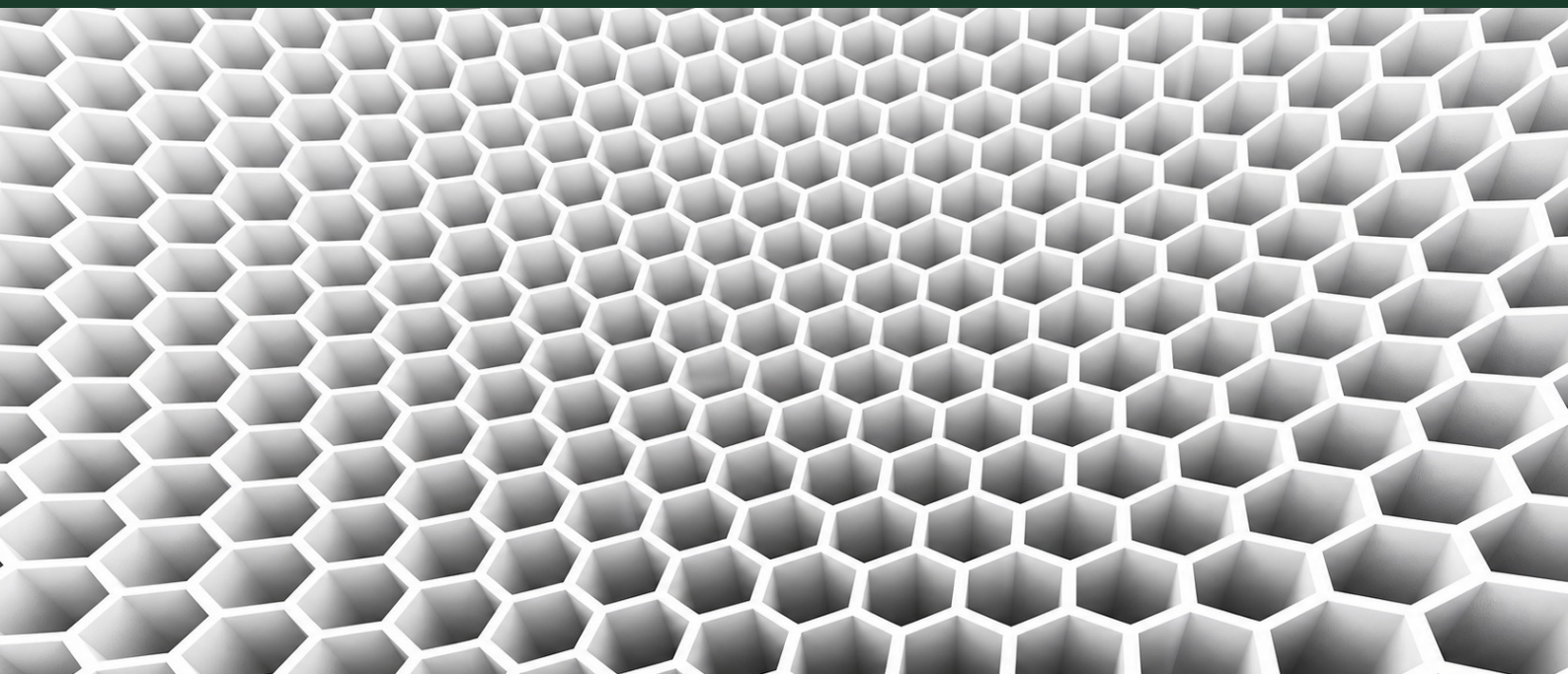
BEYOND THE DEMO

The AI POC Graveyard — Why Proofs of Concept Never Reach Production

A Practitioner's Perspective on the Industrialisation Gap in Enterprise AI Programmes

Giovanni Leonardi

January 2023



The AI POC Graveyard — Why Proofs of Concept Never Reach Production

Programme · No. 215

Originally written January 2023 | Re-edited and re-rendered for the WhereBeyond Edition, 2025

“The proof of concept succeeded because it was designed to succeed; the production environment failed because nobody designed for it at all.”

1. The Pattern Nobody Admits To

There is a graveyard in every large organisation, and it is growing. It is not populated by failed projects in the conventional sense — projects that ran over budget, missed their deadlines, or delivered the wrong thing. The occupants of this particular graveyard all succeeded. They demonstrated that AI could summarise documents more quickly than analysts. They proved that machine learning could predict equipment failure with impressive accuracy. They showed that natural language processing could extract entities from contracts in a fraction of the time a paralegal would need.

And then they stopped. The proof of concept was declared a success. The demonstration was given to the steering committee. The slide deck was filed. And the capability never reached a single operational user.

In my experience, this is not the exception. It is the norm. Across financial services, healthcare, energy, and the public sector, the pattern is remarkably consistent: organisations are running dozens, sometimes hundreds, of AI proofs of concept, and converting almost none of them into production capability. The success rate from POC to production — when organisations are honest enough to measure it — rarely exceeds ten per cent.

The question is not whether this pattern exists. It is why it persists, and what it tells us about how AI programmes are structured.

2. The Proof of Concept Was Never Designed to Scale

The most common explanation for the POC-to-production gap is technical: the proof of concept was built on clean data, in a sandboxed environment, by a small team with direct access to domain experts. Production requires messy data, enterprise integration, security compliance, monitoring, retraining pipelines, and operational support. The technical gap is real.

But it is not the real problem.

The real problem is that the proof of concept was never designed with production in mind. It was designed to answer one question: can this technology do this thing? That is a useful question, but it is the wrong stopping point. The questions that matter for production — can we do this at scale, with real data, within our existing architecture, under our governance constraints, with our current skills, and at a cost that the business case supports — are systematically deferred.

This deferral is not accidental. It is structural. POCs are typically funded from innovation budgets, run by data science teams, and governed (loosely) by innovation boards. Production deployment requires capital investment, operational ownership, IT architecture decisions, and business process redesign. These are different conversations, with different stakeholders, different timelines, and different risk appetites. The handoff between them is where AI initiatives go to die.

3. Three Structural Failures

The POC graveyard is sustained by three structural failures in how AI programmes are designed and governed.

3.1 The Ownership Vacuum

The first is an ownership vacuum. A proof of concept is typically owned by a data science or innovation team. Production capability must be owned by a business function, supported by IT operations. The transition requires a willing recipient — a business owner who accepts accountability for an AI capability they did not commission, built on technology they may not understand, with operational requirements they have not planned for.

The pattern I have observed is that this recipient rarely exists when the POC is initiated. Nobody has asked the business function whether they want this capability in production. Nobody has assessed whether they have the budget, the skills, or the appetite to operate it. The POC team assumes that a successful demonstration will create demand. It almost never does — because demand requires not just proof of possibility, but proof of operational viability, and that is precisely what the POC did not test.

3.2 The Architecture Gap

The second structural failure is an architecture gap. Enterprise AI in production requires infrastructure that most organisations have not built: model serving environments, data pipelines that can deliver production-quality data at the cadence the model requires, monitoring systems that detect drift and degradation, retraining workflows, versioning, and rollback capabilities.

These are not exotic requirements. They are the basic infrastructure of operational machine learning. But they are expensive, they take time to build, and they require skills that most IT functions do not yet have. A POC sidesteps all of this by running in a notebook, on a curated dataset, with manual intervention whenever something breaks.

The gap between a working notebook and a production-grade deployment is not a small step. It is a different engineering discipline entirely, and most AI programmes do not acknowledge it until they try to cross it.

3.3 The Business Case That Never Materialises

The third failure is economic. A proof of concept demonstrates technical feasibility. It does not demonstrate economic viability. The cost of building and running an AI capability in production — including the data infrastructure, the operational support, the ongoing retraining, the governance overhead, and the change management required to embed it in business processes — is routinely underestimated by a factor of five to ten.

When the full production cost is honestly assessed, many POCs that looked promising become marginal or unviable. The document summarisation tool that saved an analyst two hours a day costs more to run in production than the analyst's time. The predictive maintenance model that caught failures with ninety per cent accuracy in the lab drops to sixty per cent on real-world data and requires a data engineering team to keep it fed.

This is not a failure of AI. It is a failure of programme discipline — the failure to build a credible, production-costed business case before the POC begins.

4. What Good Looks Like

The organisations that are beginning to close the POC-to-production gap share a common characteristic: they have stopped treating proofs of concept as the starting point of AI adoption and started treating them as one phase in a longer, more disciplined programme.

A proof of concept should not be a standalone exercise in possibility. It should be one gate in a defined path that runs from hypothesis through technical validation, operational design, business case confirmation, and production deployment — with explicit decision points and kill criteria at every stage.

This means several things in practice:

- Production ownership is identified before the POC begins. There is a named business owner who has agreed, in principle, to accept the capability if the technical and economic case holds. This does not guarantee adoption, but it eliminates the ownership vacuum.
- The POC tests production viability, not just technical feasibility. The dataset is representative. The environment approximates production constraints. The governance requirements are assessed, not deferred.
- The business case is built on production costs, not POC costs. The data infrastructure, operational support, retraining, monitoring, and change management costs are estimated honestly before the POC is funded, and the business case must survive those numbers.

- There is a defined kill gate. If the POC reveals that production deployment is not viable — because the data is not ready, the architecture gap is too large, the business case does not hold, or the operational owner is not willing — the initiative is stopped cleanly, and the learning is captured.

5. The Deeper Problem

The AI POC graveyard is, at its root, a symptom of a deeper problem: the tendency to treat AI as a technology initiative rather than a business capability investment. Technology initiatives can succeed in the lab. Business capability investments must succeed in the operation.

Until AI programmes are structured, governed, and funded as capability investments — with the same rigour around operational readiness, change management, and business case discipline that any major capability investment requires — the graveyard will keep growing.

“The proof of concept succeeded because it was designed to succeed; the production environment failed because nobody designed for it at all.”
