

Late to AI? **Five moves** decide whether you scale or stall.

Gartner puts the pilot-to-production failure rate at 70–80%. Here's the structural playbook for landing on the right side of that number.

SWIPE →

A risk-mitigated execution order, not a tool checklist.

- 01** Governance before procurement
- 02** Clean data before any model
- 03** Automate the low-risk workflow first
- 04** Buy the API, don't build the model
- 05** Monitor drift, cost, and accuracy from day one

70–80%

of AI pilots never reach production.

The two causes show up in nearly every failure: data infrastructure nobody curated, and governance nobody wrote down before the licences were signed.

Source: Gartner, 2024

700

**full-time agents' worth of workload,
handled in one quarter.**

Klarna's enterprise AI assistant took that load off its support team in its first quarter live, while cutting repeat-ticket resolution time by 25%. Tier-1 support is where the ROI shows up first.

Source: Klarna, 2024

> 92%

is the context-precision floor a governed RAG system should clear, alongside a groundedness score above 90%, before it's trusted with a customer-facing answer.

Target benchmark, automated evaluation sets



Building a proprietary foundation model burns tens of millions in compute and specialised engineering, with no guarantee of an edge over the API you could have licensed instead.

Terence Kok

Enterprise AI Strategist

Fine-tuning vs. retrieval over clean data

FINE-TUNING

- Capital-intensive, specialised engineering
- Weeks to retrain as facts change
- Opaque: hard to trace an answer to its source

RAG ON CLEANED DATA

- A fraction of the compute cost
- Updates the moment the source doc changes
- Every answer traces back to a document

Governance is the **gate**, not the paperwork after.

Before a single AI licence is issued, put ISO/IEC 42001 and the NIST AI RMF in place, under an AI Safety and Compliance Board the CRO or CISO chairs. The target: 100% of proposed use cases pass an algorithmic impact assessment before proof-of-concept, not after the pilot is already running.

Vendor terms and monitoring are the guardrails that hold once you're live.

Contract API vendors on zero-data-retention terms so your inputs never train a public model. Then watch latency, cost, and groundedness continuously; an unmonitored deployment degrades quietly until someone notices the token bill or the wrong answer first.

Full breakdown and sources:

terencekok.com/blog/late-adopters-playbook-five-imperatives-scalable-enterprise-ai

Five moves, made before budget is committed, **not after.**

The full playbook, with sources and the governance checklist, is on the blog.



Terence Kok · Enterprise AI Strategist

SAVE · COMMENT · FOLLOW