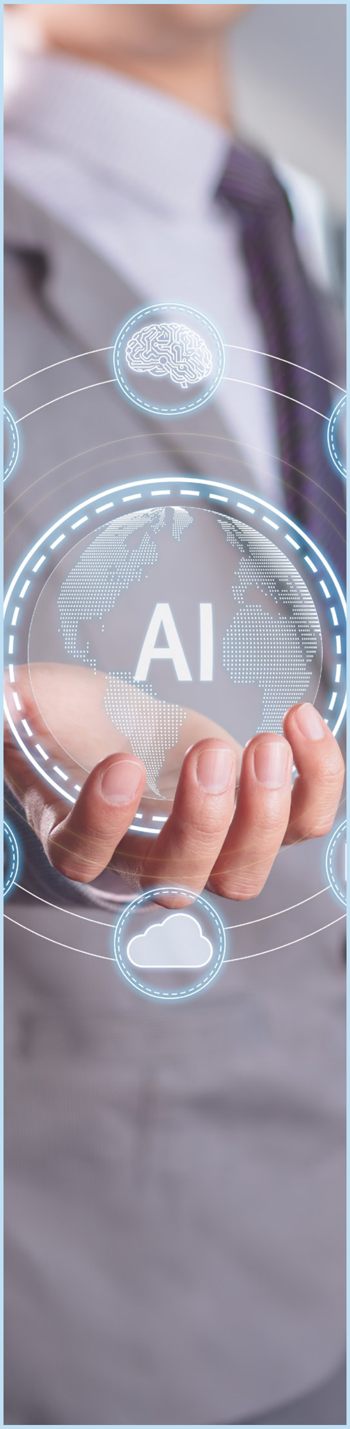
VIEW POINT



THE REAL AI CHALLENGE: BRIDGING THE GAP BETWEEN EXPERIMENTATION AND ENTERPRISE IMPACT

Abstract

Most enterprises do not fail while adopting generative AI – they fail while scaling. Pilots and proofs of concept (POCs) are launched easily, but frequently, they remain disconnected from production systems and measurable business value. The core challenge lies in the transition from experimentation to enterprise-wide deployment. Only a small proportion of organisations achieve meaningful impact at scale, with many AI initiatives stalling or failing due to poor data readiness, weak governance, and unclear ROI. Off-the-shelf models, though powerful, lack domain-specific context and require enhancement through fine-tuning, robust data pipelines, and continuous monitoring. Success depends on building an integrated AI stack, investing in high-quality proprietary data, and ensuring system observability. Equally important are organisational factors such as workflow redesign, trust-building, and disciplined prioritisation of high-impact use cases. Ultimately, scaling AI is a strategic and operational challenge, not just a technological one.

Infosys®
Navigate your next

The gap between piloting and scaling is where most enterprises are losing the artificial intelligence (AI) race. It is common knowledge that generative AI (Gen AI) can write code, draft contracts, create images, summarise earnings calls, and serve customers around the clock. The demos of new AI products are impressive, and the pilots are shipped quickly. But quite frequently, nothing much happens after that. The pilot sits in a sandbox, far from production, quietly consuming resources while executives wait for the return on investment (ROI) that never seems to arrive.

The problem in these situations is not the technology, but the transition process. And understanding the mechanics of that transition – from proof-of-concept (POC) to production-grade deployment – is the defining challenge for enterprises.

The adoption plateau is real

At first glance, the adoption of AI across enterprises appears strong. According to a McKinsey report, 88% of organisations use AI in at least one business function. A deeper look reveals a more sobering reality. Only about one-third have successfully scaled AI across the enterprise.

That gap tells the real story. True, adoption of AI is happening all around, and its usage is increasing every day, but the depth of usage and meaningful value at scale remains elusive. While many organisations can point to pilots, only a few can demonstrate measurable enterprise-wide transformation or earnings before interest and taxes (EBIT) impact.

The failure rate is not hypothetical. Initial Gartner estimates suggested that 30% of Gen AI projects would be abandoned after POC by the end of 2025 due to poor data quality, weak risk controls, escalating costs, or unclear business value. More recent analysis indicates a harsher reality. As many as 50% of projects are already failing at the POC stage, with failure rates rising even higher in organisations lacking AI-ready data. Clearly, scaling is the bottleneck, not experimentation.

Why off-the-shelf models stall in the enterprise

Foundation models from providers like OpenAI, Anthropic, and Google are powerful generalists. But their strength is also their limitation. Trained primarily on public internet data, they lack awareness of an organisation's proprietary data, domain-specific language, workflows, and compliance requirements.

This is the core problem highlighted in Scale AI's enterprise guide. For example, a wealth management AI assistant needs access to internal research reports and compliance frameworks. A healthcare summarisation tool must understand clinical protocols and institutional terminology. Generic models, however advanced, cannot deliver such specific information.

This gap often leads to hallucinations, one of the biggest risks in enterprise AI. When a model fabricates legal citations or invents medical recommendations, the consequences extend beyond technical – they become legal, financial, and reputational liabilities.

As a result, production-grade systems require far more than API access to a model. They demand fine-tuning, retrieval systems, evaluation frameworks, and continuous monitoring. These are foundational requirements, not optional enhancements.

Data security is equally important. Many off-the-shelf AI tools process data through external cloud environments, which can be unacceptable for industries handling sensitive information. In sectors like finance, healthcare, and defence, deploying AI within a controlled environment is a necessity for compliance and security reasons.

The stack that actually gets to production

Organisations that successfully move from experimentation to production think in terms of systems, not tools. They build around a layered Gen AI stack consisting of three core components: the foundation model, the data engine, and the application layer.

1. Foundation models

Contrary to popular belief, model size is not the defining factor of performance. Research shows that smaller models that are fine-tuned on domain-specific data can outperform much larger general-purpose models on specialised tasks in both accuracy and relevance. In enterprise environments – where precision, repeatability, and cost efficiency matter – task-specific models frequently deliver superior outcomes.

Industry analyses reinforce this fact – smaller, fine-tuned models are often more accurate and reliable for real-world business workflows. This is reflected in domain-specific systems such as Google's Med-PaLM 2, which demonstrates how targeted training improves performance in specialised contexts.

2. The data engine

Data offers the true competitive advantage in enterprise AI. Most organisations discover that their data is not ready for AI. Documents are unstructured, metadata is inconsistent, and labeled datasets are scarce. Building pipelines to collect, clean, annotate, and continuously refine data is labour-intensive, but essential. This is where long-term advantage is created.

While models can be accessed by anyone, high-quality, proprietary datasets cannot be easily accessed. Companies that invest early in data infrastructure are able to build capabilities that competitors cannot easily replicate.

3. Observability and monitoring

Production AI systems must be measurable and controllable. Observability includes tracking model inputs and outputs, defining performance metrics, and implementing alert systems for failures or anomalies. Without this layer, AI systems become opaque and risky. For example, a customer service bot that cannot detect hallucinations or an internal assistant that cannot flag errors could easily introduce operational and compliance risks. In a nutshell, enterprise AI is not just about generating outputs; it is about controlling and measuring them.

The organisational gap is bigger than the technical one

While technical challenges are significant, organisational barriers often prove to be equally or more difficult. Successful companies do not bolt on AI to existing workflows; they redesign processes to integrate AI effectively. This may involve restructuring how knowledge is stored, redefining approval chains, and retraining employees to work alongside AI systems. This requires coordination across functions that have historically operated in silos.

Trust is a key differentiator. According to Gartner, 57% of highly mature AI organisations report strong business trust in AI systems, compared to just 14% of low-maturity organisations. This gap is not about algorithms – it is about change management.

Another important factor is prioritisation discipline. Leading organisations recognise that not all AI initiatives deliver equal value. In practice, a small percentage of use cases, often 10–15%, generate the majority of returns. Organisations that scale successfully are those that identify these high-impact use cases early and concentrate investment accordingly.

From experimentation to production: A practical frame

The path from pilot to production is demanding but not opaque. Organisations should begin by evaluating use cases across three dimensions: business value, technical feasibility, and risk. The most promising opportunities lie where value is high and risk is manageable. Rather than spreading resources across numerous pilots, companies should focus on delivering one or two production-grade deployments that demonstrate clear ROI.

Clear requirements must first be defined:

- What defines success?
- Where does the data reside?
- What are the governance structures that are needed?
- How will performance be measured and monitored?

These are not one-time decisions. They require ongoing oversight and leadership. Gartner reports that it is for this very reason that 91% of leaders in mature AI organisations have set up dedicated AI leadership roles. Ultimately, scaling generative AI is not an IT project – it is a strategic transformation.

What is at stake and what are the opportunities?

The economic potential of gen AI is substantial. McKinsey estimates it could generate between $2.6 trillion and $4.4 trillion in annual value across industries, particularly in customer operations, marketing and sales, software engineering, and R&D. Yet capturing this value depends less on technological breakthroughs and more on execution discipline.

The organisations that succeed are not those with the most advanced models, but those that commit to the tough tasks: cleaning and structuring data, redesigning workflows, building governance frameworks, and fostering organisational trust. The gap between experimentation and production is where most companies stall. But it is also where the competitive advantage lies. The question now is whether organisations have the focus, patience, and operational discipline to move beyond the demo and pilot stage and make it real.

How can Infosys BPM help?

The Infosys Generative AI Business Operations platform comprises a suite of customised, responsible design frameworks and ready-to-use BPM-focused solutions.

AI systems must be built with robust safeguards to boost the effectiveness of generative models. We ensure a positive impact on business processes by implementing strict risk management to remove biases and hallucinations in our gen AI models. Our AI framework is responsible and has been developed after collaborating with multiple industry and regulatory bodies, as well as internal experts.

For more information, contact infosysbpm@infosys.com

Infosys®
Navigate your next