

Executive Summary

AI Infrastructure for Business Impact: Enabling Agentic AI with Scalable Compute



Peter Rutten
Research Vice-President, Performance Intensive Computing, Worldwide Infrastructure Research, IDC



Madhumitha Sathish
Research Manager, High Performance Computing, IDC

Overview

AI has transitioned from experimental deployments to core enterprise capabilities. In 2025, the emergence of agentic AI, systems capable of autonomous decision-making and task execution, marks a pivotal shift in enterprise strategy. These systems promise dynamic, context-aware operations across datacenter, cloud, and edge environments, enabling intelligent resource orchestration, adaptive deployments, and real-time responsiveness.

However, IDC's survey of 410 IT and AI infrastructure decision-makers reveals that over 75% of enterprises lack clarity on agentic AI use cases, underscoring a critical readiness gap. To unlock AI's full potential, organizations must align infrastructure strategy with business objectives, governance frameworks, and cost-efficiency goals.



75% of enterprises lack clarity on agentic AI use cases, underscoring a critical readiness gap.

Key Findings

Strategic alignment and use case selection

- ▶ 83% of enterprises launch fewer than 10 AI use cases simultaneously, reflecting cautious scaling.
- ▶ Only 21.7% conduct full ROI analyses; 22.2% ensure alignment with business objectives.
- ▶ Poorly defined use cases risk stalling in proof-of-concept phases, leading to wasted investment and compliance exposure.



Governance and security

- ▶ Enterprises are adopting multilayered data governance strategies, including access controls, anonymization, and life-cycle policies.
- ▶ Security testing practices include adversarial simulations, input sanitization, and prompt manipulation to safeguard model integrity.
- ▶ Organizations must embed governance across the AI life cycle to ensure trust, compliance, and operational resilience.



Comprehensive cost evaluation

- ▶ Two-thirds of GenAI initiatives begin with full cost assessments covering infrastructure, licensing, labor, and scalability.
- ▶ Top cost concerns include specialized infrastructure for training (60.7%) and inferencing (54.5%).
- ▶ Strategic planning must account for long-term feasibility, vendor ecosystems, and regulatory compliance.



Infrastructure strategy and deployment preferences

- ▶ AI training relies heavily on GPUs, high-speed interconnects, and cluster-level orchestration.
- ▶ Inferencing demands low-latency, scalable systems, often deployed across datacenters, public clouds, and edge environments.
- ▶ 77% of respondents prioritize hardware/software consistency across deployment environments to ensure performance predictability and model portability.



Strategic Imperatives



Align AI with business value: Enterprises must define use cases with measurable outcomes and conduct disciplined ROI evaluations to ensure strategic relevance and scalability.



Build governance foundations early: Businesses must integrate security and compliance, from development through to deployment. Robust governance mitigates risk and supports responsible AI innovation.



Adopt flexible infrastructure models: Organizations use a mix of in-house builds, turnkey systems, and integrator-led deployments. Unified infrastructure across datacenter, cloud, and edge environments is key to reducing latency and complexity.

Conclusion

As enterprises embrace the transformative potential of AI, success hinges not just on adoption but on strategic execution. Agentic AI introduces powerful new capabilities, but also demands clarity in use case selection, disciplined cost evaluation, and robust governance. IDC's survey reveals that while enthusiasm for AI is high, many organizations still lack the foundational infrastructure and strategic alignment to scale effectively.

To move from experimentation to enterprisewide impact, organizations must prioritize three imperatives: align AI initiatives with measurable business value, embed governance and security from the outset, and adopt flexible infrastructure models that support deployment across datacenter, cloud, and edge environments. By doing so, they position themselves to harness the full promise of agentic AI, driving innovation, operational efficiency, and sustainable competitive advantage in the years ahead.

[Read the full white paper](#)