Key Capabilities Checklist: Your Path to Operational Readiness for Enterprise AI

Achieve Operational Readiness for Enterprise AI

Introduction

AI adoption is widespread, but many deployments fall short of delivering enterprise value. SaaS AI apps provide localized productivity gains, yet they rarely transform core workflows.

Public-cloud large language model deployments can create unpredictable costs, sovereignty concerns, and management complexity. The rise of agentic AI only raises the stakes, demanding real-time orchestration and governance.

At the same time, enterprises must balance GPUs, CPUs, and accelerators across datacenter, cloud, and edge while managing power, cooling, and supply chain risks. What's needed is a platform approach that treats AI as a first-class workload— governed, resilient, and ready for scale.

Treat Enterprise AI Like Other Mission-Critical Apps

AI can't deliver enterprise value if it's limited to SaaS apps or run only in the public cloud, where costs can quickly spike and management grows complex. To achieve real ROI, enterprises need a platform approach that treats AI with the same rigor as any mission-critical application. A platform approach embeds automation, observability, and governance into operations so that models run consistently and predictably across datacenter, cloud, and edge.

Key capabilities include:

- Automated operations and lifecycle management for purpose-built AI workloads.
- Enterprise-grade observability and governance to provide trust and consistency.
- Seamless distribution of training, fine-tuning, and inference across hybrid environments.
- Policy-based placement to balance performance, cost, and sovereignty compliance.

Prepare for the Era of Agentic AI

The future of AI lies in agentic systems, where multiple agents interact with data, systems, and each other in real time. These workloads demand orchestration, modular execution, and dynamic scaling that traditional infrastructure cannot provide. Without enterprise-grade orchestration and governance, these systems overwhelm IT and amplify risk.

Key capabilities include:

- Orchestration tools to coordinate multi-agent systems in real time and seamlessly scale them across datacenter, cloud, and edge.
- Modular execution environments that support dynamic scaling.
- Secure interactions between AI agents and enterprise systems.
- Elastic infrastructure spanning datacenter, cloud, and edge to support real-time responsiveness.-

Right-Size AI Hardware for Scale

Enterprises face tough choices when planning AI infrastructure. Should inference run in the cloud for elasticity, on-prem for sovereignty, at the edge for low-latency needs, or a mix of all three?

At the same time, IT must balance GPUs, CPUs, and emerging accelerators while navigating power, cooling, and supply chain constraints. A platform approach provides the agility to right-size resources, adapt to new accelerators and extend AI reliably across environments—avoiding waste, bottlenecks, and deployment delays.

Key capabilities include:

- Unified access to GPUs, CPUs, and emerging accelerators.
- Intelligent workload placement across environments to avoid bottlenecks and waste.
- High-bandwidth infrastructure to support demanding training and inference.
- Flexibility to extend inference workloads to the edge for responsiveness and sovereignty.

Additional Guidance

For a deeper dive, read the full ebook: AI with Confidence: Why a Platform Approach is Essential to Enterprise AI.

This guide explores pressing challenges enterprises face today, ensuring AI readiness, enforcing governance and trust, and delivering predictable performance at scale. A platform-first strategy, rooted in sovereignty, security, and observability enables IT leaders to treat AI as a first-class workload— governed, resilient, and ready to deliver sustainable business value across datacenter, cloud, and edge.

NUTANIX