Enterprise Strategy Group by TechTarget

Generative AI: What's the Right Infrastructure Fit?

Organizations are exploring generative AI's (GenAI's) potential to drive productivity, automation, innovation, and creativity at scale. However, enterprises face several key challenges to realizing their vision for privacy and security, cost, control, compliance, performance, and infrastructure fit. Selecting a partner that addresses these challenges is the best approach.

This Enterprise Strategy Group Infographic was commissioned by Broadcom and is distributed under license from TechTarget, Inc.

Enterprises Are Pursuing the Vast Potential of GenAI

Enterprises are investing heavily in GenAI. Research from Informa TechTarget's Enterprise Strategy Group found 30% of respondents have GenAI use cases in production, not in a proof-of-concept or experimental phase. AI is becoming a core component of enterprise workloads and is rapidly moving to mass market adoption.

30%
of respondents have GenAI use cases in production

Other indicators of how GenAI adoption is accelerating:

68%
of organizations believe GenAI budgets will increase in the near term.

12%
of organizations have GenAI budgets of $1 million or more.

Enterprise expectations:

63%
of organizations believe GenAI will deliver significant or transformative value in productivity gains.

84%
of organizations agree it is important to incorporate their own enterprise data to support GenAI initiatives.

The Problem: Security, Cost, and Scalability

Enterprises face several key challenges to realizing their vision for the application of GenAI, including privacy and security, cost, control, compliance, performance, and infrastructure fit. To maximize GenAI benefits, enterprises need to use fine-tuned LLM models. But they need to seek ways to govern and control GenAI to minimize risk, meet compliance obligations, and manage costs. Performance, ease of use, and infrastructure fit have to make sense for organizations to succeed with GenAI.

73%
of organizations named **data privacy and security** as a top GenAI risk.

65%
of organizations say changes are needed to their **supporting infrastructure** before they can proceed with GenAI.

GenAI implementation challenges:

33%
of organizations say **regulatory compliance.**

26%
of organizations say **technical complexity.**

25%
of organizations say **cost.**

The Right Approach: Flexible, Scalable, and Secure GenAI Infrastructure

GenAI infrastructure best practices criteria:

Infrastructure and data access must be tightly controlled. Proprietary data must be protected to prevent leakage outside the organizational boundary.

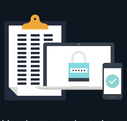
Private cloud solutions combine the scale and agility of the public cloud with the security and performance of the private cloud.

Solutions that provide a wide and evergreen range of choices for AI models, tools, and frameworks enable organizations to remain flexible and on top of rapidly evolving GenAI technology and practices.

The Solution: VMware Private AI Foundation With NVIDIA

Enterprises need AI-specific infrastructure and a flexible platform to scale GenAI successfully. Broadcom and NVIDIA have developed a GenAI platform—VMware Private AI Foundation with NVIDIA—to fill that need. VMware Private AI Foundation unlocks the power of GenAI and unleashes productivity with lower TCO for enterprises.

Key Benefits:



Enable privacy, security, and compliance

The platform leverages the built-in capabilities of VMware Cloud Foundation and NVIDIA AI Enterprise to ensure privacy, security, and compliance for AI models.

Simplify GenAI deployment and optimize costs

The specially developed advanced capabilities streamline the deployment of GenAI models and optimize costs, making it easier for enterprises to harness GenAI models.

Accelerate performance regardless of the LLM

Software and hardware capabilities enabled in the platform extract maximum performance from GenAI models. The latest Broadcom benchmark study shows that VMware and NVIDIA achieve AI workload performance on par with bare metal.

Enterprises can deploy GenAI use cases such as code generation, contact center optimization, content creation, advanced information retrieval, and agentic AI in their environment using VMware Private AI Foundation with NVIDIA.

vmware by Broadcom

LEARN MORE

Source: Enterprise Strategy Group Research Report, *The State of the Generative AI Market: Widespread Transformation Continues*, September 2024.