OCTOBER 2024

HPE Private Cloud AI Available for Virtual Assistants

Scott Sinclair, Practice Director

Overview

This month, HPE announced that its HPE Private Cloud AI is available to order. This new integrated and pre-validated solution combines hardware and software from both HPE and NVIDIA and it is designed to accelerate deployment of infrastructure for AI and shorten time to value for AI initiatives. The new solution accelerators give enterprises the ability to deploy virtual assistants with one click and operationalize them in seconds, simplifying the process from end to end.

AI Infrastructure Investment

Few workloads present such a strong opportunity to transform business operations and customer experience as artificial intelligence does, whether it is predictive or generative AI. With the rise in popularity of generative AI, business and technology leaders have become enamored with its potential. According to research from TechTarget's Enterprise Strategy Group, 54% of organizations expect to have AI in production or to have begun adoption in 2024.[1]

When organizations were asked to identify where they believe a generative AI solution could be most beneficial in the future, customer service (cited by 48% of respondents) was the most common answer.[2] With a technology as potentially transformational as generative AI, myriad use cases will likely emerge over the coming years. That being said, the creation of chatbots/virtual assistants presents a clear path to near-term return on investment (ROI).

When it comes to infrastructure, IT professionals involved with evaluating and/or managing infrastructure specifically for AI initiatives identified their top considerations (see Figure 1).[3] While high-performance compute tops the list (cited by 42%), ease of deployment (39%) and data governance (39%) are tied for close second. In other words, when it comes to generative AI, the creation of virtual assistants is a common first step toward production and a quick path to seeing a positive return. And leveraging an easy-to-deploy, pre-optimized, validated infrastructure that can keep your organization's data private and on premises should be a priority when it comes to new investment.
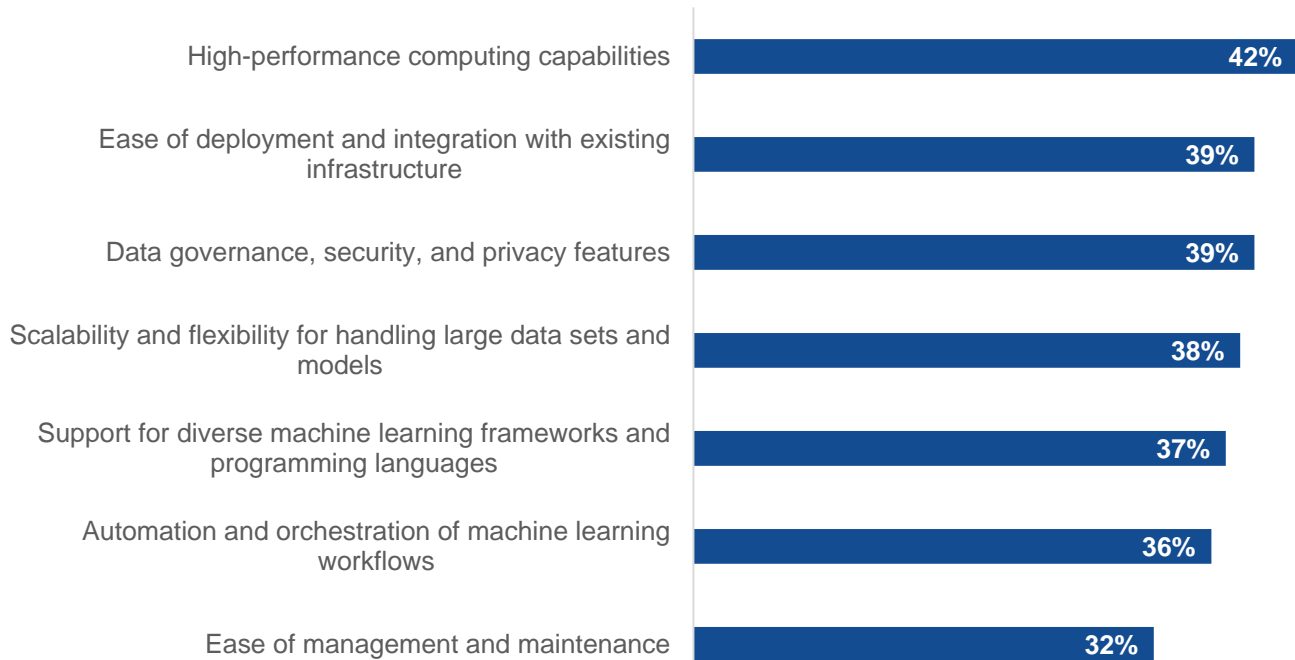
---

[1] Source: Enterprise Strategy Group Complete Survey Results, *Beyond the GenAI Hype: Real-world Investments, Use Cases, and Concerns*, August 2023.

[2] Ibid.

[3] Source: Enterprise Strategy Group Research Report, *Navigating the Evolving AI Infrastructure Landscape,* September 2023.

Figure 1. Most Important Capabilities When Selecting AI Infrastructure

When selecting AI infrastructure, which of the following capabilities are most important to your organization? (Percent of respondents, N=339, three responses accepted)

High-performance computing capabilities	42%
Ease of deployment and integration with existing infrastructure	39%
Data governance, security, and privacy features	39%
Scalability and flexibility for handling large data sets and models	38%
Support for diverse machine learning frameworks and programming languages	37%
Automation and orchestration of machine learning workflows	36%
Ease of management and maintenance	32%

Source: Enterprise Strategy Group, a division of TechTarget, Inc.

Conclusion

Enterprise Strategy Group research revealed that early success in AI is common, with 72% of respondents reporting seeing value from in AI initiatives within three months or less.[4] However, there are anecdotal examples of initial AI projects experiencing low and even negative ROI. Limitations and even failures occurring early with AI can often be the result of mistakes in infrastructure design.

HPE, in partnership with NVIDIA, is removing deployment-related guesswork. HPE claims a one-click deployment with its HPE Private Cloud AI. Of course, the number of clicks is less important than whether the deployment is simple. Simplicity is what greatly reduces the risk of deployment. To further those simplicity benefits, the AI software in HPE Private Cloud AI is also curated and integrated, including NVIDIA software and AI Essentials from HPE, reducing the complexity of managing the multiple software components necessary for AI initiatives.

In this announcement, HPE is highlighting its generative AI virtual assistant solution accelerator to help developers quickly build interactive, natural language chatbots utilizing open source models informed by an organization's private data. HPE is also expected to continue to scale its solution accelerator portfolio for HPE Private Cloud AI over time. Ultimately, organizations have a wealth of options when it comes to infrastructure for AI, both on and off premises. What I like about HPE's approach is its focus on simplicity. A best practice when starting with AI projects is to define a small, contained project through which you can get a "quick win" and provide early value to your business. HPE Private Cloud AI is designed to deliver exactly that.

---

[4] Ibid.

Enterprise Strategy Group
by TechTarget

About Enterprise Strategy Group
TechTarget's Enterprise Strategy Group provides focused and actionable market intelligence, demand-side research, analyst advisory services, GTM strategy guidance, solution validations, and custom content supporting enterprise technology buying and selling.

contact@esg-global.com

www.esg-global.com