



# Why IT Leaders Should Deploy Generative AI Infrastructure Now

## In this handbook:

Why IT Leaders Should  
Deploy Generative AI  
Infrastructure Now

# Why IT Leaders Should Deploy Generative AI Infrastructure Now

*SCOTT SINCLAIR, PRACTICE DIRECTOR*

In the past several months, rampant excitement about the potential benefits of generative AI technology has increased the technology's priority status across enterprise organizations worldwide.

According to a recent research report from TechTarget's Enterprise Strategy Group, "Beyond the GenAI Hype: Real-world Investments, Use Cases, and Concerns," 42% of organizations said they are in a generative AI proof of concept if they haven't already deployed it in production. Our research showed that generative AI ranks higher than cloud in overall strategic business initiatives, which highlights how critical these projects are now.

In other words, the adoption rate for generative AI projects is expected to be massive and unlike anything we've seen from enterprise technology. And as a result, there is a high likelihood that your own executive team is currently in a conflicted state: They are excited about the potential productivity benefits of generative AI, but they are concerned about the risks to data privacy.

## In this handbook:

Why IT Leaders Should  
Deploy Generative AI  
Infrastructure Now

Regardless of the pace of AI adoption within your own organization, the expected overall adoption rate means that if your organization lags in adopting and deploying generative AI products, your competition will gain an increased advantage.

## What's in the 'Box'?

Organizations need to move quickly when it comes to generative AI, but they should do so in a manner that enables them to start small, scale quickly and mitigate risk associated with data privacy, compliance and security. With that necessity in mind, Nutanix has introduced GPT-in-a-Box.

The product combines the following elements:

1. Nutanix Cloud Infrastructure hyperconverged (HCI) technology, its AHV hypervisor, Nutanix Files Storage, Nutanix Objects Storage, Kubernetes and Nvidia GPU acceleration.
2. Nutanix services to help size and deploy the infrastructure and deploy the software stack, along with learning frameworks and a curated set of large language models.

There is a lot to like in this packaging, but most important is its simplicity of design. Nutanix is known for simplicity, which is a hallmark of its HCI technology.

Overall, there is likely going to be a longer than usual "crawl" phase before you get to "run" with generative AI within your organization. But if you want to get a leg up on generative AI initiatives, don't waste time trying to deploy the perfect infrastructure for

## In this handbook:

Why IT Leaders Should  
Deploy Generative AI  
Infrastructure Now

what the ideal use will be in three to five years. In fact, few -- if any -- organizations truly have a strong grasp on what the ideal use will be.

# Best practices for generative AI use, deployment

We do have a sense of what those uses will look like in general. According to Enterprise Strategy Group's generative AI research report, the more commonly identified uses improve productivity, efficiency and the overall customer experience.

As a result, organizations should seek to speed up infrastructure deployment to enable their data science teams to get started on identifying the right data and models. As an example, the Nutanix product enables organizations to start quickly, and it gives them the flexibility to scale and adapt as needed.

The ability to deploy the product on premises is also important. While public cloud services will likely support most generative AI products, a separate Enterprise Strategy Group research study, "Multi-cloud Application Deployment and Decision Making," found that 29% of organizations identified AI/machine learning workloads as not being candidates for cloud deployment.

## In this handbook:

Why IT Leaders Should  
Deploy Generative AI  
Infrastructure Now

Some of those organizations will be launching AI initiatives that will use sensitive data or data sets with privacy concerns. Or maybe the data and compute requirements in the cloud are simply too costly for organizations just getting started. According to the multi-cloud research report, the cost of low-latency performance in the cloud is the most common reason organizations decided that an on-premises workload is not a candidate for the public cloud.

Ultimately, when it comes to generative AI, speed is of the essence. And given the increased executive-level priority on generative AI workloads, IT leaders must be proactive.

GPT-in-a-Box is simple to use and flexible, but Nutanix is not the only provider that has announced a strengthened Nvidia partnership and a product for generative AI. Always evaluate your options.

*Enterprise Strategy Group is a division of TechTarget. Its analysts have business relationships with technology vendors.*