

# VMware Private AI

## Democratizing AI for all enterprises

### At a glance

VMware Private AI is a new architectural approach for AI services that enables

- Privacy and control of corporate data
- Choice of open-source and commercial AI solutions
- Quick time-to-value
- Integrated security and management

Generative AI is no longer the next best thing—it has now achieved mainstream status and is set to transform enterprises across industries in much the same way that the PC, the internet, and the smartphone did. From creating content to generating code to automating IT operations, the applications of generative AI are virtually endless, and its potential is enormous—the impact of AI overall in the enterprise represents a **\$4.4 trillion annual economic value**.<sup>1</sup>

However, while this technology can deliver spectacular productivity gains across every aspect of an organization, several challenges must be confronted before bringing generative AI into the enterprise.

### Challenges presented by generative AI

Generative AI presents several challenges that are preventing organizations from forging forward. Let's discuss what these are.

- **Privacy** – Above all else, privacy is the core issue that enterprises need to address. Organizations want to leverage their private, proprietary data and intellectual property (IP) with large language models (LLM) to get compelling solutions, but doing this with public models and public APIs introduces the risk of IP or data leakage. Additionally, organizations need to protect and control access to the AI models themselves.
- **Choice** – Enterprises want to choose an LLM, such as Llama, Falcon, Mistral and others, that fits their needs. They also want to choose partners without getting locked into specific vendor contracts. Access to a wide variety of options is essential.
- **Cost** – Building AI systems can be expensive because of the constant changes in technology. New vendors enter the market regularly, and new software is continuously launched and deployed. Vendor lock-in may exacerbate this cost spiral.
- **Performance** – Since LLMs and AI models are trained using massive data sets, they can place significant demands on an organization's infrastructure, leading to cost and potential performance issues.
- **Compliance** – Like any enterprise solution, generative AI must meet an organization's specific compliance requirements, including access control and audit readiness.

1. McKinsey & Company. "The economic potential of generative AI: The next productivity frontier." Michael Chui, Eric Hazan, Roger Roberts, Alex Singla, Kate Smaje, Alex Sukharevsky, Lareina Yee, and Rodney Zimmel, June 2023.



According to research, **60 percent** of IT leaders believe “AI presents new risks to our data privacy that we don’t know how to address.” At the same time, **50 percent** “don’t yet plan to prevent data leakage around sensitive material and content using generative AI.”

Source: VMware FY24 Q2 Executive Pulse, N=450 Enterprise Technology Executives

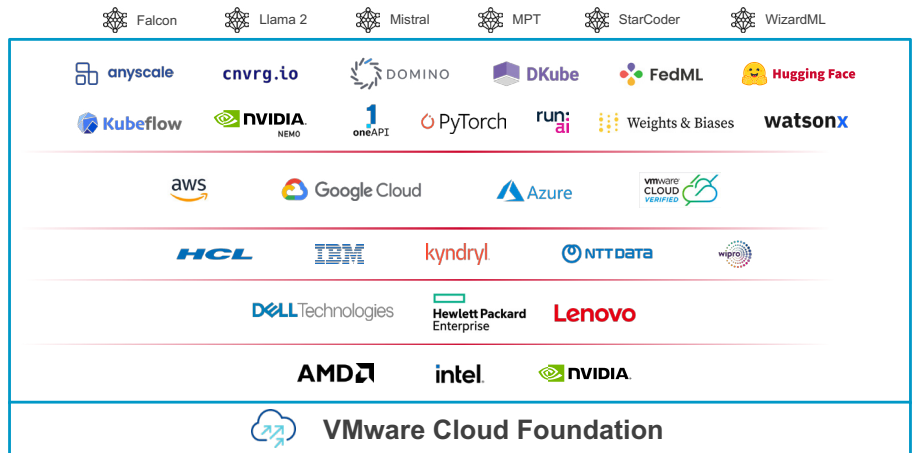
## Overcome these challenges with VMware Private AI

VMware is on a mission to democratize generative AI and ignite business innovation for all enterprises. VMware has created VMware Private AI, an architectural approach that enables you to unlock the business gains of AI while meeting your organization’s privacy and compliance requirements. Built on top of the industry leading multi-cloud platform, [VMware Cloud Foundation™](#), it is a privacy-first approach to AI. Here’s how VMware Private AI works:

1. Choose the LLM that meets your needs from a variety of openly licensed and commercially usable models.
2. Fine-tune and customize your LLM using your enterprise data.
3. Securely run your customized LLM in production and at scale across your VMware environment.

## Harness the VMware Private AI Partner Ecosystem

The VMware Private AI Partner Ecosystem is comprised of a variety of industry-leading providers to give you maximum flexibility and choice. It includes the most popular AI providers such as NVIDIA, Intel, IBM, Anyscale, Domino Data Labs and Hugging Face, enabling you to deploy with confidence. The three major server OEMs provide certified server architecture to support VMware Private AI. HCL and Wipro are GSIs (global system integrators) that help customers along their generative AI journey.



**Figure 1:** VMware partners with several industry-leading providers to create the VMware Private AI Partner Ecosystem.

## Use cases enabled by VMware Private AI



**Code generation and creation.** Accelerate developer velocity by enabling your LLM to generate code while protecting the privacy of corporate data/IP.



**Contact center resolution experience.** Improve the quality of content and feedback your contact centers provide your customers.



**IT operations automation.** Boost productivity by automating incident management, reporting, ticketing, monitoring, and more.



**Advanced information retrieval.** Leverage our advanced capabilities to enhance employee productivity by improving document search, policy, and procedure research.

## Get the most out of generative AI in the enterprise

VMware Private AI gives you the capabilities to create private, secure, customized LLMs to run inferencing and drive enhanced decision-making against your organization's private data sets. The unique architectural approach and open ecosystem of VMware Private AI combine to deliver a variety of benefits:

- **Get the flexibility of choice.** With the option of deploying a wide range of AI software tightly integrated with VMware Cloud Foundation™, it's easy to find the best fit for your application and use case.
- **Deploy with confidence.** By partnering with leading tech providers and building on the intrinsic security of VMware Cloud Foundation, VMware Private AI enables you to run your customized LLM while mitigating risk and protecting against data/IP leakage and poor model accuracy.
- **Achieve superior performance.** Get exceptional performance for your generative AI models with VMware Private AI. Our solutions support the major GPU technologies and pooling of these GPUs to extract superior performance for AI workloads.
- **Optimize costs and compliance.** VMware Private AI empowers enterprises to achieve the most cost-effective solutions for their generative AI models. Leveraging our extensive expertise and strong partnerships with industry leaders in this domain, you can rest assured that your enterprise will receive optimal value and be compliant in the strict regulatory environments in which businesses operate across the globe today.
- **Augment productivity.** VMware Private AI can help organizations with significant productivity improvements. Build private chatbots to sift through data and produce intelligent outputs. Automate repetitive manual tasks and processes. With VMware Private AI, your organization can move faster, operate smarter, and stay ahead of the competition.

## Unlock the power of AI for your organization

With VMware Private AI, you can elevate productivity across every department of your organization while maintaining the privacy and control of your corporate data and IP. To learn more about VMware Private AI, please visit [VMware.com/aiml](https://vmware.com/aiml).