



Bring business-critical AI to life with HPE Private Cloud AI



Most organizations face the same reality: they need to move fast with AI, but they can't sacrifice security, governance, or control. Public cloud isn't always the answer, especially when proprietary data, regulatory requirements, or predictable economics are at stake. The result is that AI often gets stuck in the pilot phase, as teams try to run business-critical AI on fragile, bespoke infrastructure and integrations instead of a production-ready system.

Co-engineered with NVIDIA as part of the NVIDIA AI Computing by HPE portfolio [HPE.com/ai](https://hpe.com/ai), HPE Private Cloud AI is a turnkey AI factory that lets you move fast with AI while operating within the realities of your regulatory and security obligations. Delivered as a single SKU, it brings together everything you need to start developing and running AI use cases—typically within hours of connecting the system to your existing environment.

HPE Private Cloud AI gives AI teams self-service access to [NVIDIA® AI Enterprise](#), [NVIDIA Inference Microservices \(NIMs\)](#), curated open-source tooling, and personalized Jupyter notebooks, all from a single environment. Data scientists and engineers can spin up projects, notebooks, and models in minutes, and they can import [NVIDIA Blueprints](#) as well as custom apps or ISV solutions instead of building every project from scratch. The stack is ready, so their focus can stay on use cases, not plumbing.

Value experienced by each role

Table 1. Value of outcomes for key decision-makers for HPE Private Cloud AI

Role	Value for this role
CTO/CDO	— Provides a governed AI foundation that turns fragile pilots into a repeatable AI factory on trusted infrastructure — Enables AI to impact the same fiscal quarter instead of after prolonged experimentation
Head of AI / data science	— Ready-to-use AI factory, with NVIDIA-powered tools and curated OSS, enables teams to focus on models, RAG, and agents
Infrastructure and operations / cloud platform	— Offers pretested, lifecycle-managed AI cloud that fits existing IT processes — Reduces integration toll while keeping data on-prem and under constant security, and compliance controls

Systems sized to your AI ambitions

From early experimentation to scaled production, HPE Private Cloud AI offers two operating modes—connected and fully air gapped—and four right-sized configurations that align to your AI journey. Smaller systems give teams a low-risk way to prototype and validate use cases while larger systems deliver the performance and capacity needed to run multiple, business-critical AI services in production with consistent governance and control. HPE Private Cloud AI lets you choose the right GPU for the job, NVIDIA RTX PRO™ 6000 Blackwell Server Edition GPUs, to run a range of workloads from multimodal AI inference and physical AI to scientific computing, graphics, and video applications, and NVIDIA H200 GPUs for high performance AI inferencing at scale.

Built-in AI tools for faster outcomes

HPE AI Essentials Software is layered on top of the infrastructure to deliver self-service access to NVIDIA AI tools, curated open-source frameworks, and user-configured notebooks, allowing teams to quickly assemble and secure their development work while tapping into proprietary data. NVIDIA Blueprints, as well as custom and independent software vendor (ISV) applications, can be imported into the self-service catalog. Automated lifecycle management enforces your security and governance policies, and built-in observability makes it possible to scale AI services while retaining control. In connected mode, patching and updates for HPE AI Essentials Software are applied automatically while in air-gapped environments updates are manually downloaded, staged, and applied according to your operational policies.

Why this matters now

HPE Private Cloud AI is relevant across all industries but is especially compelling for data-sensitive and sovereignty-driven organizations. It also becomes a strong fit in sectors without strict regulations once AI moves from experiments to business-critical services, when data gravity makes it inefficient to keep shuttling data to the public cloud, or when differentiated IP and customer data become the core of your models.

HPE Private Cloud AI is designed for organizations that want to

- Accelerate revenue impact from AI by bringing high-value use cases into production up to a year earlier than DIY or public cloud-only approaches
- Stay ahead of fast-moving AI innovation with a joint, future-ready road map that reduces the need for repeated replatforming
- Increase the pace of experimentation and iteration so AI use cases can scale faster across the business
- Unlock differentiated value from proprietary data while retaining full control, governance, and ownership of models, data, and infrastructure
- Reduce risk with security and governance applied by default, rather than relying on manual hardening and one-off controls

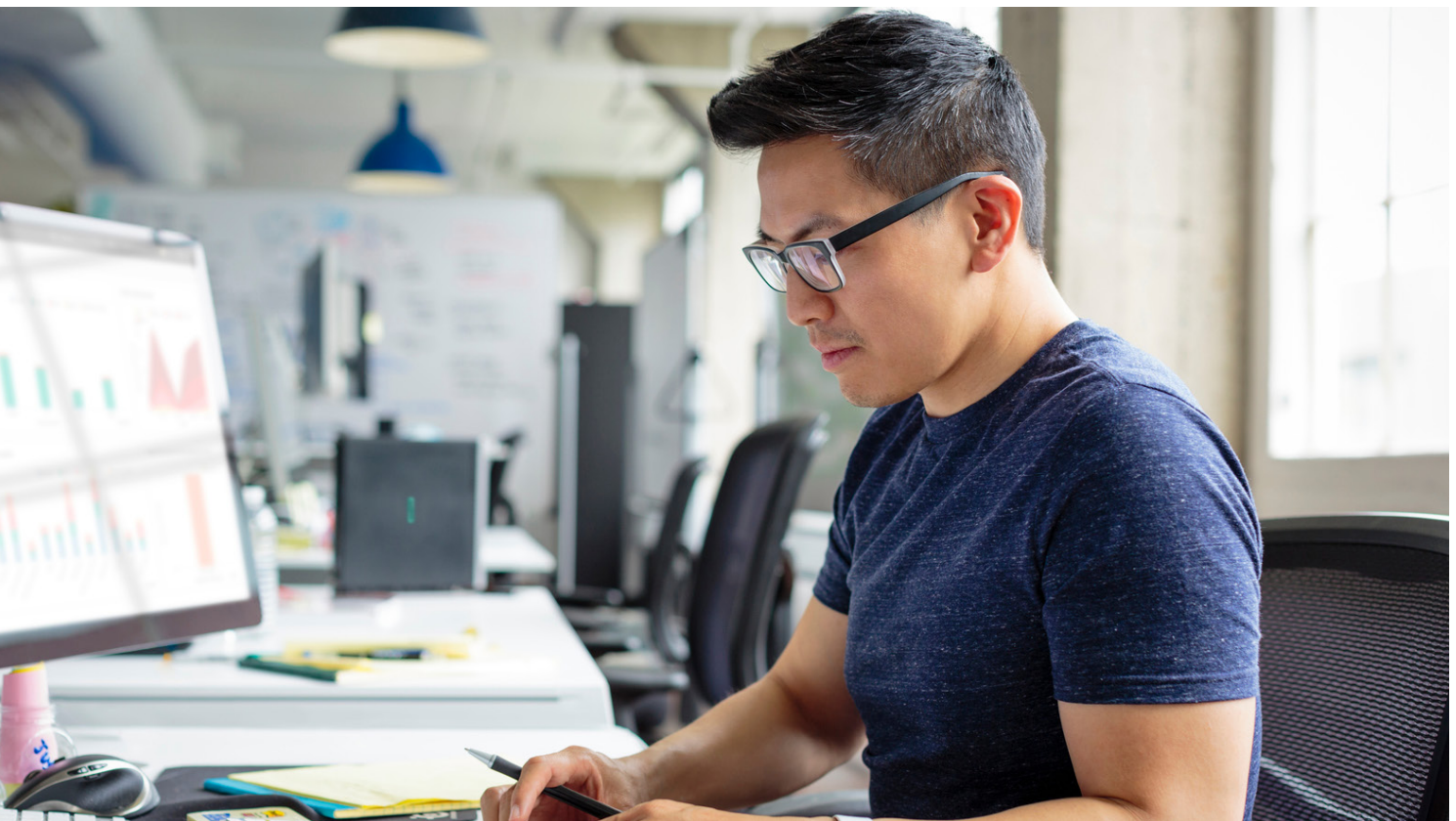
HPE Private Cloud AI gives you a production-ready AI factory that fits your security, data, and economic realities, so AI impact shows up sooner and with less risk. [Schedule time](#) with a solution expert to explore how it can support your AI strategy.

To accelerate enterprise AI, HPE and NVIDIA have joined forces to deliver AI factories integrating NVIDIA's accelerated computing with secure, scalable infrastructure from HPE. The NVIDIA AI Computing by HPE portfolio enables organizations to move beyond experiments and build robust, production-ready AI systems. Solutions include turnkey HPE Private Cloud AI for fast starts, scalable AI factories for model builders, and sovereign AI factories for specialized data protection. These offerings feature validated technology stacks, multitenant support, air-gapped and federated architectures, predictable pricing, and rapid deployment. Fully integrated infrastructure and services allow IT teams to focus on developing AI use cases.

Visit [HPE.com](https://www.hpe.com)

Learn more at

[HPE.com/Private-Cloud-AI](https://www.hpe.com/Private-Cloud-AI)



[Chat now](#)

© Copyright 2026 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

NVIDIA, NVIDIA RTX, and the NVIDIA logo are either trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. All third-party marks are property of their respective owners.

a00156964ENW

HEWLETT PACKARD ENTERPRISE

[hpe.com](https://www.hpe.com)

HPE

 **NVIDIA**