



Building AI factories — Accelerating innovation with HPE

Transforming data into AI-driven decisions

Imagine a place where innovation meets industrialization, where AI is not just a concept but a reality, where raw data is transformed into actionable intelligence at lightning speed.

That's an AI factory, an environment designed to manage the entire AI lifecycle — from data pipelines and model training to inference and real-time insights. With purpose-built infrastructure, integrated tools, scalable operations, and unparalleled AI expertise, the AI factory can revolutionize the way you harness the power of artificial intelligence.

Think of it like a traditional factory, but instead of producing physical goods, it creates value and intelligence from data. AI factories take in raw data, process it through AI models, and output actionable intelligence, predictions, or new AI solutions. The journey from data to intelligence is streamlined, efficient, and groundbreaking.

The result? Faster innovation, operational efficiency, scalability, and greater control over data and business outcomes.

Why do you need an AI factory? Because operationalizing AI can be challenging.

As organizations embrace AI's transformative potential, they face a range of complexities inherent in fully operationalizing AI.

These challenges include:

- **Complex AI workloads:** Managing diverse and resource-intensive AI workloads can overwhelm existing infrastructure, leading to inefficiencies and delays.
- **Need for multitenancy:** Efficiently managing multiple tenants and their resources is complex and resource-intensive, leading to potential conflicts and inefficiencies.
- **High costs of cloud AI:** The expenses associated with deploying AI solutions in the cloud can be prohibitive, impacting budget and ROI.

- **Efficiently scaling AI:** Scaling AI operations efficiently across different environments is complicated, causing bottlenecks and limiting growth potential.
- **Operational complexity:** Managing AI infrastructure and operations can hinder productivity and increase operational costs.
- **Security and compliance:** Data security and regulatory compliance are critical, as breaches can lead to significant financial and reputational damage.
- **Lack of AI expertise:** The need for deep technical expertise, data pipeline design, automation, and responsible AI governance complicates the deployment and management processes.
- **Data management:** Efficiently managing and accessing large volumes of data for AI applications is difficult, leading to delays and inaccuracies.

Reshape your AI landscape with HPE

HPE delivers the infrastructure, software, and services to power AI factories. Our AI offerings encompass turnkey solutions for traditional enterprises, as well as verified, composable options for model builders, service providers, and sovereigns.

HPE's full-stack portfolio provides fully integrated infrastructure, operating layer, software, control plane, networking, storage, and services to accelerate the buildout of AI factories. The HPE portfolio enables you to achieve value quickly without the hassle of navigating the stack and filling in the gaps on your own.

AI is iterative, and models can degrade over time due to data drift, changing customer behavior, and environmental shifts. To maintain relevance and performance, a high-performing AI factory infrastructure is essential for retraining models, conducting simulations, monitoring inference quality, and managing deployment pipelines for continuous improvements. The HPE AI portfolio provides precisely that.

Build your AI factory

HPE helps customers stand up AI factories that can transform your raw data into intelligence and intelligence into valuable, scalable outcomes.

— Turnkey AI factories

HPE Private Cloud AI delivers fully integrated, production-ready AI factories with trusted infrastructure and pre-integrated software. Our enterprise-ready, on-premises solution focuses on data, compliance, and performance requirements. It facilitates a straightforward, scalable pathway to production-grade generative AI and agentic AI, optimized for developer productivity and implemented as a private cloud for enterprise IT environments.

HPE Private Cloud AI delivers instant AI productivity with rapid workload deployment, secure and unified data access, an end-to-end AI software platform, and an easy to manage private cloud experience.

As part of NVIDIA AI Computing by HPE, HPE Private Cloud AI incorporates NVIDIA® AI Enterprise software and NVIDIA NIM, along with support for NVIDIA Blueprints. This turnkey solution is well-suited for enterprises implementing enterprise AI workloads at scale, such as generative AI and agentic AI applications.

— AI factories at scale

For customers who need large-scale AI factories or have complex workload needs, HPE delivers customized solutions based on proven architectures, deployed through our expert services.

Our AI factory at scale solution enables scalable AI operations with data sources across core, edge, and cloud environments. It integrates orchestration, observability, and lifecycle automation, making it ideal for model builders and service providers that deliver large-scale AI environments to multiple tenants across various locations.

HPE's customized solutions offer centralized control over decentralized AI infrastructure, optimized workload placement, GPU utilization, multitenant management, built-in automation, and observability. It is intended for organizations who are building, training, tuning, and deploying their own models. Businesses providing shared-service AI platforms, as well as sectors such as manufacturing, finance, and telecommunications will also benefit from this solution.

— Sovereign AI factories

HPE delivers AI factories designed for public sector, defense, and critical infrastructure entities within defined borders who require complete data control, sovereignty, policy compliance, and security.

With built-in multitenancy, federated access, compliance with regional regulations, and support for air-gap environments, HPE solutions accelerate national AI strategies while mitigating risks associated with dependence on foreign cloud providers. These AI factories are designed to address the interests of governments and national agencies.

Redefining the AI factory with complete solutions

HPE provides expert support and services across our entire portfolio, including comprehensive planning, design, implementation, and management of AI solutions. Each solution incorporates extensive automation and self-service capabilities, enabling you to dynamically provision and manage your environment efficiently.

AI factories are designed to manage the entire AI lifecycle, enhancing innovation, operational efficiency, scalability, and control over data. HPE delivers comprehensive AI solutions that support the ideal AI factory: robust infrastructure, integrated tools, scalable operations, and AI expertise.



© Copyright 2025 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 is a trademark and/or registered trademark of NVIDIA Corporation in the U.S. and other countries. All third-party marks are property of their respective owners.

a50013328ENW

HEWLETT PACKARD ENTERPRISE

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

Connection
we solve IT™

Contact your Connection Account Team for more information.

1.800.800.0014

www.connection.com/hpe

