



# The Case for AI On Prem: Control, Compliance, and Confidence

Featuring Teradata AI Factory and NVIDIA

teradata.

 NVIDIA



# Enterprise AI is at a crossroads

The promise of artificial intelligence—scalability, speed, and transformative insights—is undeniable, but so are the challenges: data security, cost volatility, and regulatory compliance. For highly regulated industries like healthcare and financial services, these concerns are magnified, driving a resurgence in on-premises AI solutions. Organizations are seeking platforms that offer control over data residency, predictable costs, and confidence in compliance. **Teradata AI Factory**, in collaboration with NVIDIA, delivers a secure, scalable, and compliant on-prem AI platform designed to meet these demands.





# Why AI on prem?

---

The rise of data sovereignty and expanding regulatory pressure has made “private AI” a strategic imperative. While cloud-first AI has dominated recent years, many enterprises—especially those with legacy on-prem infrastructure—find themselves without viable options for secure, compliant AI adoption. Regulations like GDPR in Europe and HIPAA in the US continue to expand, requiring organizations to prove compliance through audits and maintain clear data lineage throughout the AI lifecycle.

Private AI keeps data local, reducing compliance risk and simplifying audits. With data in one spot, organizations can easily document policies and workflows, making compliance less resource intensive. Cloud-only AI introduces challenges: unpredictable costs, latency issues, and governance gaps. Mission-critical use cases demand low-latency inference and reliability that cloud networks cannot always deliver. Proprietary networks in industries like telecom, defense, and energy often cannot use public or cloud networks, further underscoring the need for on-prem AI.



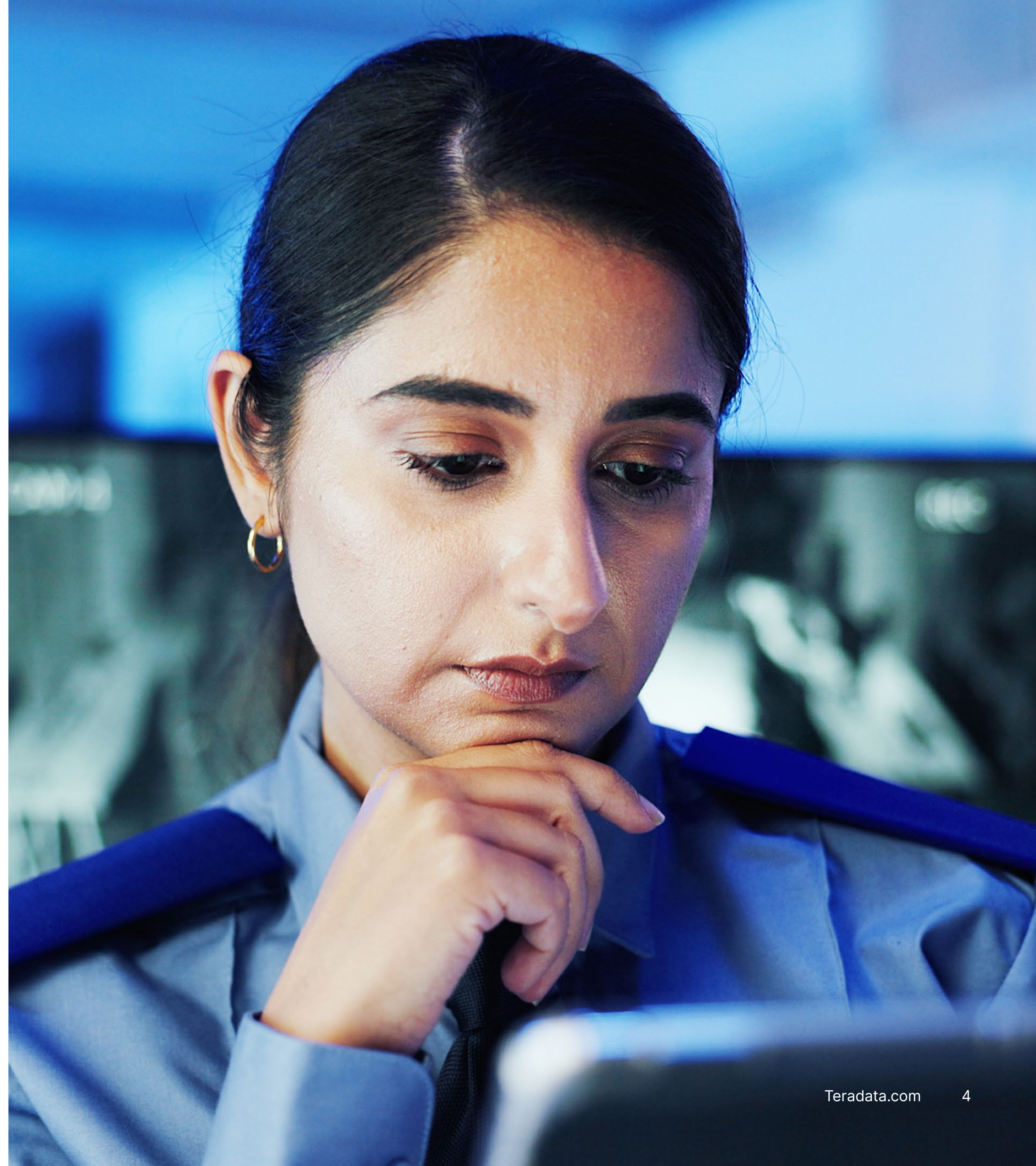


# Introducing the AI Factory: Secure, on-prem AI built for enterprises

---

Teradata AI Factory is a turnkey platform for enterprises seeking AI innovation with control and without compromise. Designed for speed, security, and scale, it enables organizations to launch AI initiatives quickly. For existing Teradata customers, familiar tools like ClearScape Analytics® are integrated, allowing teams to build, train, and deploy models seamlessly.

The AI Factory's integrated stack includes hardware, software, and microservices packaged in a uniform solution. The AI Workbench empowers teams to focus on outcomes, not infrastructure. NVIDIA's accelerates computing and AI software stack is built in for demanding AI workloads, from training to inference, ensuring high performance and scalability. NVIDIA's accelerated computing and AI software stack is built in for demanding AI workloads, from training to inference, ensuring high performance and scalability.





# AI Factories with Teradata in collaboration with NVIDIA

---

The collaboration between Teradata and NVIDIA brings together the best enterprise analytics and advanced GPU-accelerated AI infrastructure. By integrating NVIDIA's accelerated hardware and software into the Teradata AI Factory, organizations gain the ability to accelerate model training, deliver high-performance inference, and unlock sophisticated generative AI capabilities such as retrieval-augmented generation (RAG).

This synergy means enterprises can efficiently process vast volumes of both structured and unstructured data, supporting mission-critical use cases like fraud detection, claims processing, and pharmaceutical research and development.

**The Enterprise Vector Store**, a key component, enables secure semantic search across diverse data types, laying the foundation for robust generative AI workflows. This will enable companies to provide context to their AI workflows.

Looking ahead, Teradata is working with NVIDIA for continuous innovation. Future releases will deepen integration with NVIDIA's advanced AI software stack, empowering organizations to rapidly adopt new features while maintaining strict control, compliance, and predictable costs.

---

By leveraging NVIDIA's AI infrastructure within the Teradata AI Factory, enterprises not only achieve speed and scalability, but also benefit from enhanced security and governance—positioning themselves to confidently meet emerging regulatory and operational demands.



# Control, compliance, and confidence

With private AI, organizations know exactly where their data is, how it's used, and can easily audit workflows. This transparency reduces the burden on compliance and security teams.



## AI Factory is built to ensure:



### Data sovereignty

Sensitive data remains in place, never leaving the organization's infrastructure.



### Regulatory compliance


Built-in governance, auditability, and lineage features simplify compliance.



### Predictable cost and performance

Consistent SLAs and cost control, avoiding the volatility of cloud expenses.

# Real-world use cases



Teradata is already driving innovation across multiple industries which can be seamlessly orchestrated within the Teradata AI Factory.

## **Fraud detection**

Real-time anomaly detection with AI Factory enables rapid response to suspicious activity. Learn more about [financial services use cases](#).

## **Claims processing**

Automating document classification and routing streamlines operations. Learn more about [financial services use cases](#).

## **Pharma R&D**

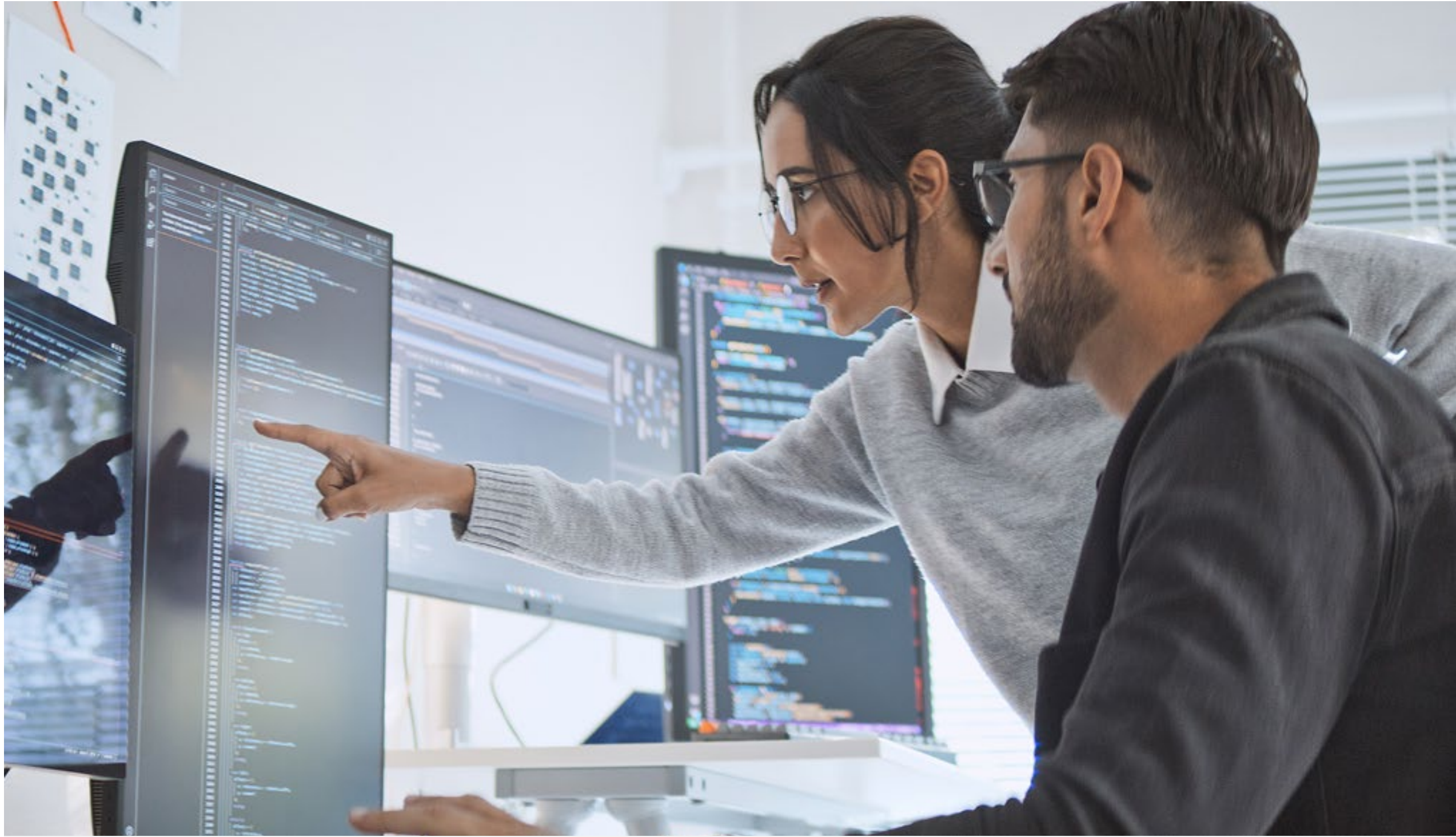
Accelerating drug discovery while maintaining strict data privacy. Learn more about [healthcare use cases](#).



# Building your AI Factory

**Best practice**

Involve compliance officers early to build governance into the AI lifecycle from day one. Visuals showing the AI Factory workbench and deployment steps can enhance understanding.



**Deploying AI Factory follows a clear, enterprise-ready path:**

- 1

**Plan/  
Assessment**  
Evaluate data, infrastructure, and compliance needs.
- 2

**Alignment**  
Coordinate IT, data science, and compliance teams.
- 3

**Deployment**  
Install the AI Factory platform with integrated Teradata + NVIDIA stack.
- 4

**Pilot**  
Launch targeted AI use case to demonstrate ROI.
- 5

**Scale**  
Expand production workloads, integrating into enterprise operations.



# What's next: Evolving with confidence

---

**The AI journey doesn't stop at deployment.  
AI Factory's roadmap includes:**

- **Agentic AI:** Moving from predictive models to autonomous, decision-making agents.
- **Expanded NVIDIA GPU support:** Harnessing the latest GPU architectures for future workloads.
- **Hybrid deployment:** Seamless integration with cloud when and where it makes sense.

Teradata will continue to innovate with NVIDIA to ensure enterprises remain at the forefront of AI.





Teradata is the AI platform built for the autonomous era. Driving business value through innovation has always been our core focus. With our AI + Knowledge Platform, supported with multifaceted AI Services, enterprises can deploy solutions that combine deep domain expertise with full enterprise context. No matter where data resides—across cloud, on-premises, or hybrid environments—Teradata uniquely connects and scales, delivering the performance AI needs.

Learn more at [Teradata.com](https://www.teradata.com).

17095 Via Del Campo, San Diego, CA 92127    [Teradata.com](https://www.teradata.com)

The Teradata logo is a trademark, and Teradata is a registered trademark of Teradata Corporation and/or its affiliates in the U.S. and worldwide. Teradata continually improves products as new technologies and components become available. Teradata, therefore, reserves the right to change specifications without prior notice. All features, functions and operations described herein may not be marketed in all parts of the world. All other trademarks are the property of their respective owners. Consult your Teradata representative or [Teradata.com](https://www.teradata.com) for more information.

© 2025 Teradata Corporation    All Rights Reserved.    Produced in U.S.A.    12.25



**teradata.**