w

DELL Technologies | NVIDIA

Contextual Product Guide

Demystifying the technologies to help your AI journey

Setting the stage for AI success

Scaling and deploying AI across an enterprise is a disciplined exercise in timing, alignment, strategy, and execution. Success depends on deploying the right technologies at pivotal moments, ensuring each step forward is intentional and strategically sound. Without a fully baked and reasoned approach, even well-funded initiatives risk falling short of their transformative potential. Deploying AI demands clarity at every stage, from identifying opportunities to addressing operational bottlenecks and aligning technology with business outcomes.

In this guide, we'll address the pain points at each stage and demonstrate how Dell and NVIDIA have aligned to help accelerate AI success for organizations through the Dell AI Factory with NVIDIA—the industry's first and only end-to-end enterprise AI solution.[1] We'll also explore how solutions, products, and services from Dell and NVIDIA can provide the essential tools and expertise to navigate this journey confidently.

Stages

Strategy

Evaluation

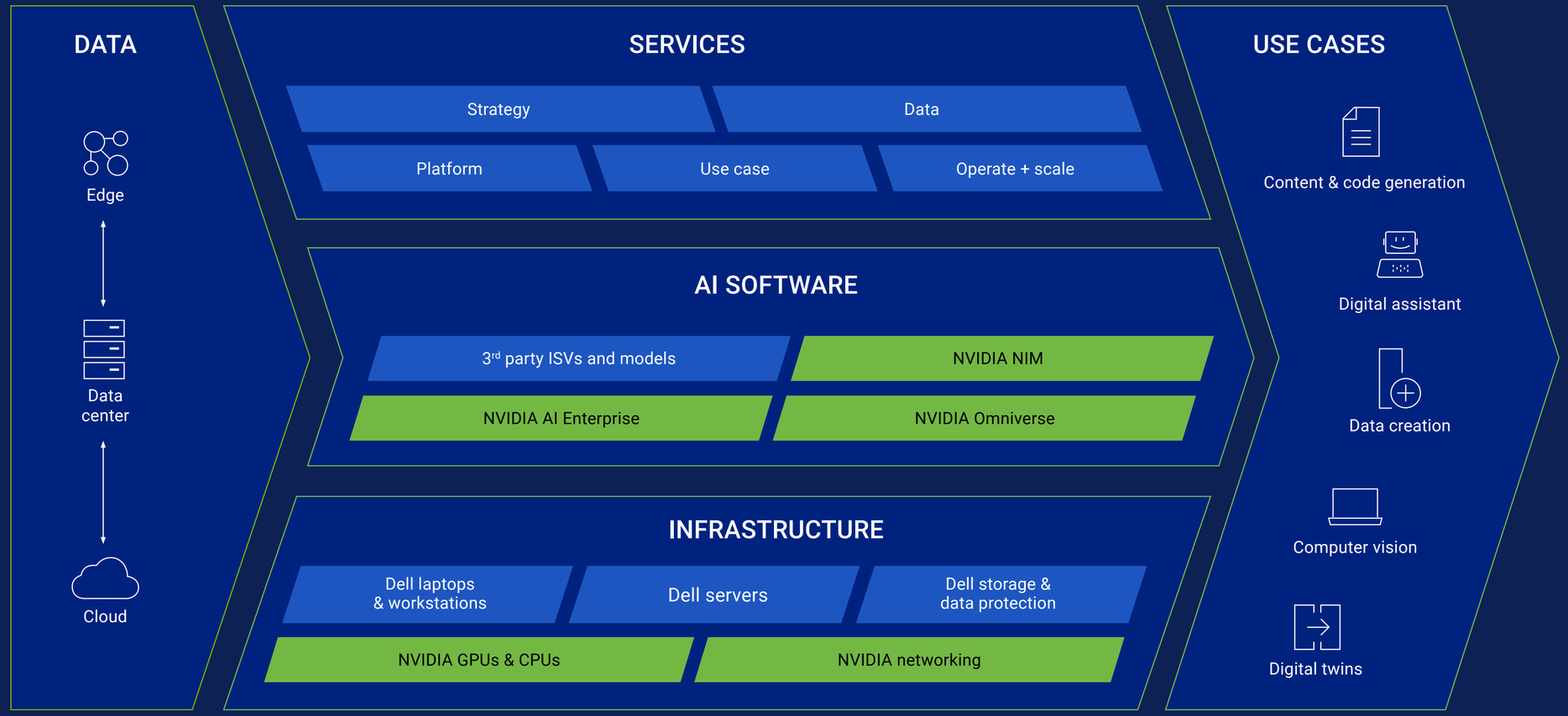
Development

Deployment

1.Based on Dell analysis, July 2024. Dell Technologies offers solutions with NVIDIA hardware and software engineered to support AI workloads from PCs with AI-powered features and workstations to servers for high-performance computing, data storage, cloud-native software-defined infrastructure, networking switches, data protection, HCI and services.

The Dell AI Factory with NVIDIA

Industry's first end-to-end enterprise AI solution

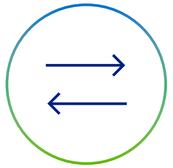


Sustainability | Security

Strategy stage: Key priorities and considerations

Focus on what matters

The **strategy** stage is where organizations establish the foundation for AI success. This involves aligning stakeholders around a shared vision, prioritizing high-value use cases, and creating a roadmap that drives measurable business outcomes. By focusing on initiatives that directly address organizational goals, such as operational efficiencies or improved customer satisfaction, enterprises set the stage for impactful AI deployment.



Break down misalignment

A common challenge during this phase is stakeholder misalignment. Different priorities across business units, IT, and leadership often lead to stalled progress and diluted focus. Without transparent collaboration, organizations risk losing momentum and pursuing initiatives that fail to deliver expected outcomes. Structured engagement, such as workshops or advisory sessions, fosters consensus and shared ownership of the AI strategy.

From vision to execution

Once stakeholders are aligned, the focus shifts to identifying and evaluating use cases. Prioritizing initiatives based on feasibility, complexity, and ROI helps organizations concentrate on what will drive the most value. Data-driven frameworks and realistic roadmaps reduce uncertainty, while clear governance structures facilitate sustainable progress as the strategy evolves.

Strategy stage: Establish a strategic AI vision

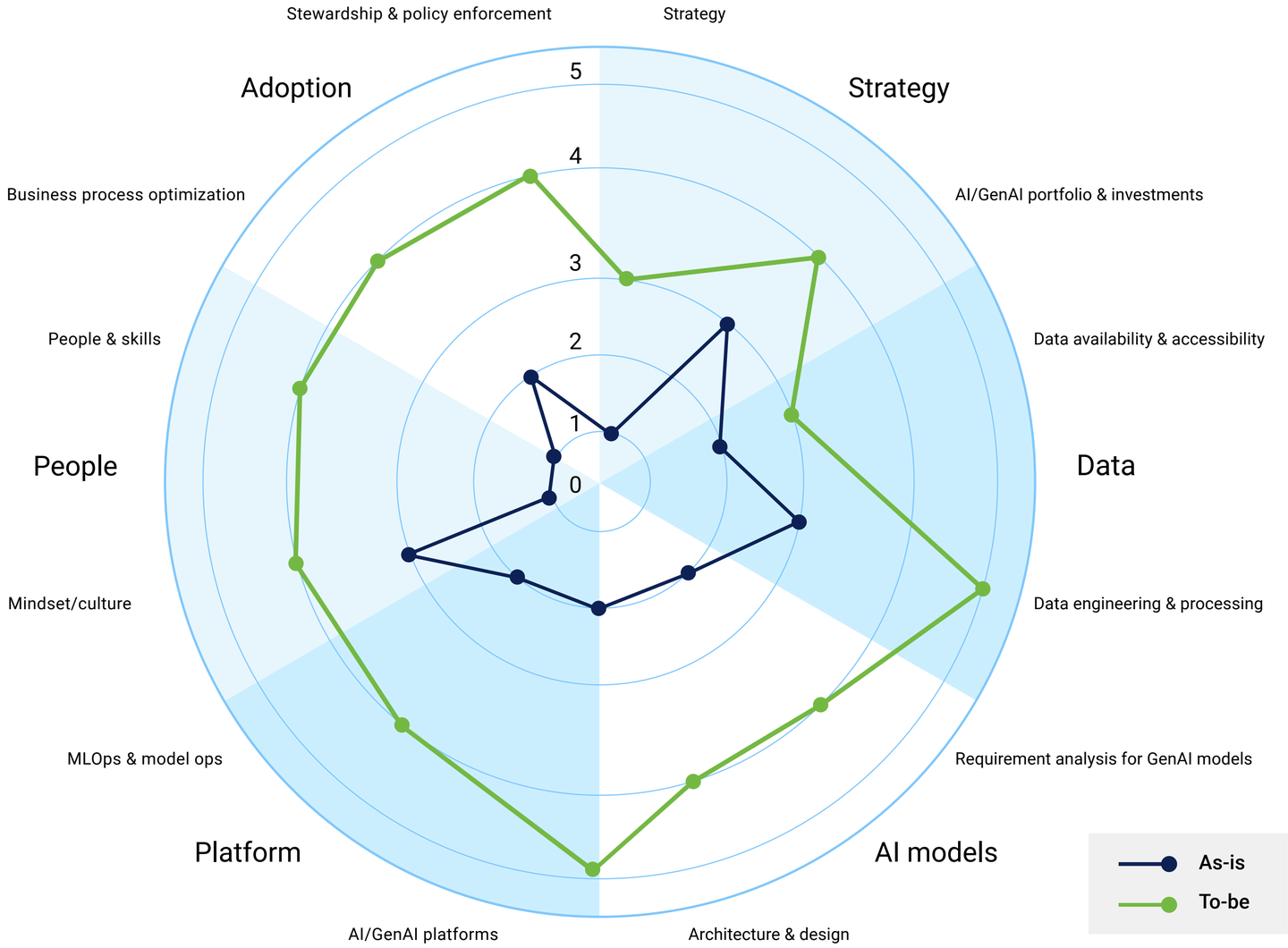
Build your strategy with Dell and NVIDIA

During the strategy phase, your organization can take advantage of the following Dell Professional Services, which are part of the Dell AI Factory with NVIDIA, to unlock tailored guidance and success, utilizing an “as-is” and “to-be” methodology.

Accelerator Workshop with Dell Services
Facilitates structured, collaborative sessions to align stakeholders, prioritize use cases, and define actionable steps, resulting in a cohesive strategy.

ProConsult with Dell Advisory Services
Offers expert guidance on scalability, readiness, and infrastructure, helping organizations assess feasibility, address gaps, and build a strong foundation for AI initiatives.

Contact your Dell representative to learn more about the services outlined above.



Evaluation stage: Key priorities and considerations

Prioritize the right fit

The **evaluation** stage is critical to identifying the most suitable AI model for a given use case. Choosing the right model can drive efficiency, accuracy, and business impact, while a poorly selected one risks inefficiencies, wasted resources, or suboptimal outcomes. The key is to align technical performance with organizational objectives, balancing complexity with real-world applicability.

Navigate complexity and challenges

Evaluating AI models is challenging due to the vast and ever-growing array of options, from pre-trained solutions to custom architectures. Decision-makers must determine whether a model aligns with their goals, considering factors like accuracy, scalability, and interpretability.

Drive success through benchmarking and collaboration

Evaluation begins with clear success criteria and robust benchmarking. Testing models against representative datasets allows organizations to assess performance metrics such as precision, recall, and latency. Collaboration between technical teams and business leaders ensures the model meets business goals while maintaining technical excellence.

Manage risks during experimentation

Organizations often struggle to balance innovation and caution during testing, rushing to implementation without adequate security, validation, and risk analysis, leading to inefficiencies and reputational damage. A significant challenge is establishing sandbox environments that replicate real-world conditions while isolating risks and enabling iterative refinement without impacting production systems.

Evaluation stage: Simple development launchpad

Accelerate and simplify evaluation with the Dell AI Factory with NVIDIA

Partnering with Dell and NVIDIA enables customers to access cutting-edge infrastructure, expert guidance, and scalable solutions to evaluate and develop AI projects with confidence. These resources simplify AI project evaluation by offering secure environments for testing, preconfigured hardware, and tailored strategies for rapid prototyping and deployment.

Dell and NVIDIA provide a comprehensive ecosystem of tools and resources to streamline the evaluation process.

Contact your Dell or NVIDIA representative to learn more about these tools and resources.

NVIDIA API Catalog
A library of pre-trained models and APIs that accelerates testing and integration for various use cases, saving time and resources.

NVIDIA AI Workbench
A versatile platform for testing and fine-tuning models, enabling rapid iterations and data-driven decision-making.

NVIDIA AI Enterprise
A robust software suite offering tools and frameworks optimized for AI deployment in hybrid and multi-cloud environments.

Dell AI Workstations powered by NVIDIA
High-performance Dell Pro Max Series workstations deliver the computational power needed for rigorous model evaluation and refinement.

NVIDIA GenAI GitHub and Launchpad AI Workbench
These platforms offer pre-built models, scalable resources, and pre-configured tools for testing prototypes in controlled, secure environments.

Dell PowerEdge Servers with NVIDIA GPUs
Scalable servers simulate real-world conditions, providing robust, isolated environments for enterprise-scale testing.

The Dell Enterprise Hub on Hugging Face
Offers simplified on-premises deployment of the most popular large language models (LLMs) on Dell's infrastructure.

The Dell Customer Solution Centers
The Dell Technologies Global Customer Solution Centers provide trusted environments where world-class IT experts assist customers in designing, validating, and building innovative solutions through hands-on engagements and proofs-of-concepts (POCs), remotely or in person.

Dell Professional Services for GenAI
A great first step on any AI journey, enabling customers to leverage Dell's expertise to create a GenAI strategy and roadmap, assess current capabilities, and set up a low-risk sandbox for rapid prototyping and experimentation. Tailored guidance from Dell experts in areas like effective testing strategies and data preparation also help to simplify this stage of AI adoption.

Development stage: Key priorities and considerations

Scale and integrate

The **development** stage focuses on transforming tested prototypes into enterprise-ready solutions. At this point, organizations must address scalability, interoperability, and operational readiness to prepare for AI applications to handle real-world demands, integrate seamlessly into existing systems, and deliver consistent value.

Bridge complexity and readiness

Challenges in this phase often stem from scaling AI beyond controlled environments. Prototypes must handle increased data volumes, diverse workloads, and unpredictable operational conditions. Additionally, integrating AI into complex enterprise ecosystems, often comprising legacy systems and multicloud platforms, poses significant interoperability hurdles.

Reliability and compliance

In the development stage, AI projects transition into fully functional applications capable of delivering meaningful, real-world outcomes according to specific organizational or vertical needs. The focus is on creating systems that are robust, reliable, and compliant with regulatory standards while meeting organizational objectives.

Privacy and reliability concerns

During development, challenges often arise when assessing if AI applications can handle diverse real-world scenarios without compromising accuracy or performance. Developers must mitigate potential failures through comprehensive testing and optimization. Additionally, data privacy concerns require robust measures to comply with laws like GDPR, HIPAA, and CCPA.

Development stage: Prepare to scale AI

Accelerate development with the Dell AI Factory with NVIDIA

Dell brings its expertise in reliable, high-performance computing infrastructure, while NVIDIA provides accelerated computing platforms and AI frameworks that are critical for tackling complex machine learning and deep learning tasks. This partnership ensures access to industry-leading expertise, tailored solutions, and seamless integration of hardware and software that accelerate innovation, enhance productivity, and drive impactful business results.

Organizations can leverage a variety of Dell and NVIDIA tools, frameworks, and services to streamline AI development.

Contact your Dell or NVIDIA representative to learn more about these tools, frameworks, and services.

NVIDIA NIM Microservices
Simplifies AI infrastructure management, reducing deployment times from weeks to minutes and ensuring resources are orchestrated efficiently.

NVIDIA NeMo™
Offers pretrained models and scalable tools for developing advanced AI applications, including conversational AI and natural language processing. By simplifying fine-tuning, NeMo accelerates application development while maintaining high precision.

NVIDIA® CUDA®
Provides a parallel computing platform to accelerate training and inference processes, allowing organizations to build high-performance applications faster.

NVIDIA DOCA™
A comprehensive framework for secure, high-performance AI workloads, DOCA facilitates privacy-preserving workflows and optimized data center operations.

Dell PowerEdge Servers with NVIDIA GPUs
Deliver computational power and security to handle large-scale AI workloads while supporting robust data protection.

NVIDIA NGC™ Catalog
A library of pre-validated, GPU-optimized models and containers simplifies the transition to production by streamlining deployment processes and integrating seamlessly into enterprise ecosystems.

NVIDIA Blueprints
These reference workflows provide pretrained AI models and microservices, equipping developers with customizable tools to build scalable applications tailored to real-world use cases, including Blueprints for customer service, drug discovery, and computer-aided engineering.

Dell Data Lakehouse
Powered by Starburst, this modern data architecture unifies hybrid and multi-cloud data sources, simplifying data access and enabling the scaling of AI models with clean, optimized datasets.

The Dell AI Data Platform
A platform that unifies Dell industry-leading file and object storage with PowerEdge Servers and the Dell Data Lakehouse to drive efficiency in AI model training and fine-tuning, via a single point of data access and high-performance across environments.

Dell AI Workstations powered by NVIDIA
High-performance Dell Pro Max Series workstations deliver the computational power needed to refine and optimize AI models for production workloads, ensuring scalability and efficiency.

Development stage: Prepare to scale AI (cont.)

Dell GitHub Repositories
Dell and NVIDIA GitHub repositories offer open-source tools, scripts, and reference architectures for automating infrastructure deployment, optimizing workloads, and integrating AI and machine learning workflows.

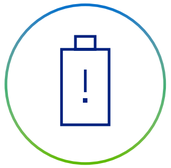
Dell Infrastructure as Code offerings
Help to relieve the management burden on DevOps teams by automating the provisioning of infrastructure.

Dell Professional Services for GenAI
Offers support throughout this stage, including preparing and securing enterprise data, implementing Dell Data Lakehouse for analytics, and building an AI platform optimized for seamless model integration and training with validated GenAI designs. Dell's expert teams guide organizations through the complexities of building AI solutions, making sure applications are reliable, scalable, and compliant.

Deployment stage: Deploy in the enterprise

Scale operational excellence

The **deploy** stage takes AI solutions from isolated use cases to fully integrated enterprise applications, focusing on achieving scale, training teams, and operationalizing AI for long-term success wherever the AI solution needs to be processed and analyzed.



Limited capacity and expertise

Scaling AI presents challenges in infrastructure, team readiness, and operational continuity. Many organizations lack the capacity or specialized expertise to deploy AI at scale, making it challenging to manage increased data volumes, build seamless integration, and maintain high performance across distributed environments.

Iteration and agility

AI solutions should be continuously refined to meet changing business demands, improve performance, and improve results based on feedback. The focus is on maintaining agility-fine-tuning models, adapting deployment patterns, and making certain the system remains aligned with organizational priorities and emerging challenges.

Manage drift and evolving needs

AI models are not static; they may encounter scenarios that differ from their training data over time, leading to drift and reduced accuracy. Additionally, as business priorities shift, models must be updated to address new objectives and incorporate features to remain relevant.

Deployment stage: Deploy in the enterprise (cont.)

Address challenges with Dell and NVIDIA

Dell offers robust, scalable hardware designed to handle the demands of AI applications, while NVIDIA delivers industry-leading GPUs and optimized software frameworks for machine learning and deep learning. Together, they ensure seamless integration, high performance, and faster deployment, helping customers quickly deploy and customize their AI initiatives.

Dell and NVIDIA provide advanced infrastructure, tools, and services that address these challenges and support enterprise-wide AI deployment.

Extend your enterprise with AI and GenAI at scale, powered by the broad Dell portfolio of AI infrastructure, tools, and professional services with NVIDIA industry-leading accelerated computing. The full stack includes: GPUs, networking, software, microservices, models, and agent blueprints.

Contact your Dell or NVIDIA representative to learn more about these tools, frameworks, and services.

NVIDIA NIM Microservices
Simplifies AI infrastructure management, reducing deployment times from weeks to minutes and ensuring resources are orchestrated efficiently.

NVIDIA NeMo
Integrated with NVIDIA GPUs and cloud-native services, NeMo optimizes AI performance, supporting scalable, cutting-edge applications.

NVIDIA Spectrum-X™ and InfiniBand Networking
Designed for AI workloads, these networking solutions deliver high performance, reduced power consumption, and seamless data transfer across scalable environments.

Dell Reference Designs
Provides engineering-validated, pre-sized solutions for common AI use cases, shortening deployment times and reducing the time needed to validate and test the integration of components.

Dell PowerScale
Offers a secure and scalable storage solution with high-speed network access, enabling organizations to manage massive AI data workloads efficiently.

Dell NativeEdge
Automates NVIDIA AI Enterprise delivery at the edge, allowing AI deployment where data is created and processed.

Dell PowerProtect
Provides cyber-resilient data protection, enabling recoverability and compliance while safeguarding AI workloads from potential threats.

Dell PowerEdge Servers powered by NVIDIA
These high-performance servers are designed to support frequent model retraining and scalability, facilitating responsive and effective systems.

iDRAC
Integrated Dell Remote Access Controller is embedded onto Dell servers, providing a centralized interface to deploy, update, and manage servers.

Dell AIOps
Centralizes management of all facets of an infrastructure stack with a cloud-based solution, powered by AI, to consolidate management of servers, storage and networking.

Dell Professional Services
Helps organizations deploy tailored GenAI solutions aligned with specific business use cases, simplify ongoing operations, enhance workforce skills via training, and scale GenAI adoption with proactive support to drive ROI. Organizations can continue to rely on Dell experts to empower your workforce with training and certification programs, managed services and other types of personalized support.

Scaling AI for sustainable growth

Deploying AI at an enterprise scale is not just a technological endeavor but a strategic journey that requires thoughtful planning, collaboration, and execution. By embracing a phased approach, organizations can confidently navigate the complexities of AI adoption, addressing challenges at every stage while staying focused on business objectives.

Each phase of the AI lifecycle brings unique opportunities and considerations, but can be overcome with a strategic approach and alignment between the lines of business and IT leaders. By following this structured journey, and working with Dell and NVIDIA experts, organizations can accelerate their AI journey and help to overcome challenges at all stages of adoption.

Dell Technologies and NVIDIA have a long-standing partnership with more than 25 years of joint innovation, focused on accelerating innovation and delivering cutting-edge platforms, solutions, and software that enable transformative results for our joint customers.

Getting started

The Dell AI Factory with NVIDIA brings together two trusted technology leaders to deliver a comprehensive and secure AI solution customizable for any business. With a portfolio of products, solutions, and services tailored for AI workloads—from desktop to data center to cloud—it paves the way for AI to work seamlessly for you.

Take the first step in your AI journey with Dell's fee-waived Accelerator Workshop and let our experts help you align priorities and achieve clarity for your AI vision.

Get started with AI

Explore the Accelerator Workshop for GenAI