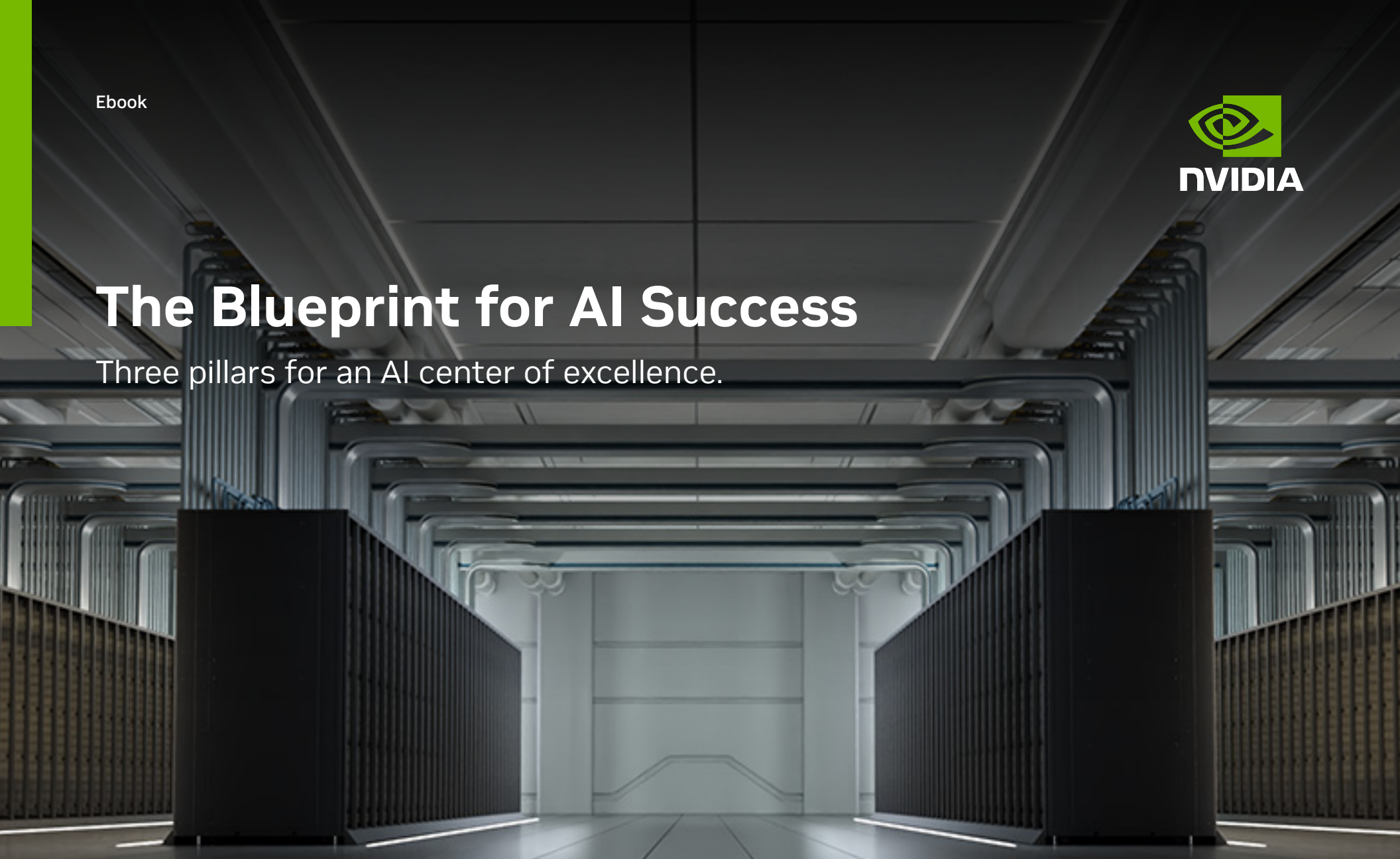
Ebook



The Blueprint for AI Success

Three pillars for an AI center of excellence.

The AI Center of Excellence: A Strategic Imperative

In a landscape where AI is a core driver of competitive advantage, many enterprises struggle with fragmented, ad-hoc AI initiatives that fail to deliver measurable returns. The solution is an AI center of excellence (CoE), a strategic framework that unifies talent, processes, and infrastructure under a single, cohesive vision. This is a critical need, as a 2025 S&P Global Market Intelligence report reveals a dramatic surge in project failures. The average organization scrapped 46% of their AI proofs-of-concept before they reached production, and the companies that abandoned most of their AI initiatives jumped from 17% to 42% in a single year.[1] The most commonly identified challenges involve data privacy, security risks, skills shortages, and cost.

To meet these challenges, building an AI CoE requires focusing on three core pillars:

- **People:** Unifying a multidisciplinary team to centralize expertise and foster collaboration.
- **Process:** Establishing repeatable workflows and governance to ensure scalable, predictable outcomes.
- **Infrastructure:** Architecting a high-performance, on-premises stack—such as the NVIDIA DGX SuperPOD™, powered by GPU-driven NVIDIA DGX™ systems and Intel® Xeon® CPUs—to serve as the core of an AI factory that turns raw data into real-time insights at scale.

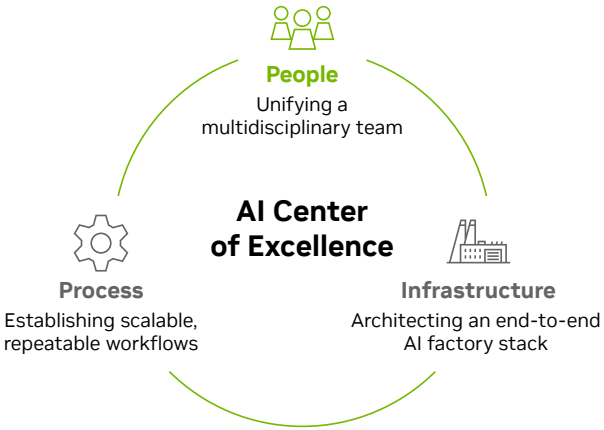
By centralizing AI efforts around these three pillars, an AI CoE helps organizations transform business process, leading to greater operational efficiency, faster time-to-market for AI solutions, and a significant increase in business value and ROI.

1. S&P Global. Generative AI Experiences Rapid Adoption, but With Mixed Outcomes—Highlights From VotE: AI & Machine Learning. May 2025.

The First Pillar: People

Unify expertise for strategic impact.

People form the core of an effective AI CoE. Centralizing AI talent provides a pool of specialized and cross-functional expertise—from data scientists, engineers, and domain experts to business leaders—enabling organizations to tackle complex challenges collaboratively. This model ensures consistent access to knowledge, whether from internal teams or trusted partners like NVIDIA AI experts, who offer deep expertise spanning system design, model optimization, and practical deployment. A multidisciplinary AI team breaks down silos and fosters cross-functional initiatives, producing solutions that are more innovative, robust, and applicable across business domains. By cultivating a unified, agile team, a CoE ensures that AI initiatives aren't only technically sound but also strategically aligned with key business objectives.



SONY

Enabling research across Sony Group companies.

As one of the global leaders in entertainment and electronics, Sony has leading R&D teams distributed worldwide. With developers using individual PCs and GPUs for distributed deep learning, Sony wanted to ensure their developers could focus on groundbreaking AI innovations without being hindered by cloud computing costs or hardware limitations.

To achieve this, Sony developed GAIA, their AI Center of Research., It provides centralized AI computing resources across all Sony Group business segments—from entertainment and electronics to financial services— leveraging DGX SuperPOD as a key component.

The results have been transformative. One team accelerated their deep-learning model training for music creation by 100x, reducing a process that once took several years to just twelve days. DGX SuperPOD has enabled higher performance on larger datasets, facilitated concurrent training runs, and supported continuous hyperparameter tuning.

Sony leveraged NVIDIA's expertise and resources—including DGXperts to optimize multi-GPU node training and NVIDIA NGC™ for containerized AI applications. This streamlined approach has enabled GAIA to maintain a consistently high utilization rate of over 95%, supporting diverse workloads such as music creation, speech synthesis, and human face, hand, and pose recognition.

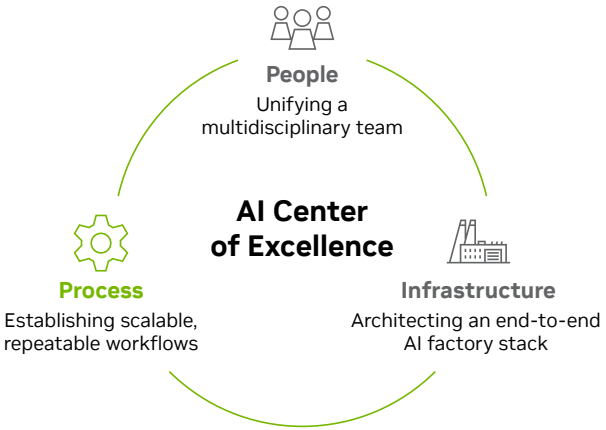
> By centralizing their AI infrastructure, Sony has shifted their developers' focus to innovation, fueling significant advances in AI integration across their various business domains.

Learn More >

The Second Pillar: Process

Establish an engine for repeatable success.

Standardizing processes is essential for repeatable, reliable AI outcomes. An AI CoE creates frameworks, tools, and documented workflows that can be replicated across departments and business lines, eliminating the need to “reinvent the wheel” and common errors. With centralized governance, best practices, and shared learnings, organizations minimize redundant work and can scale AI initiatives more efficiently. This repeatability ensures lessons from one team or use case benefit the entire enterprise. Standardized processes accelerate ROI, reduce time to market for AI-powered solutions, and support compliance, auditability, and continuous improvement at enterprise scale.

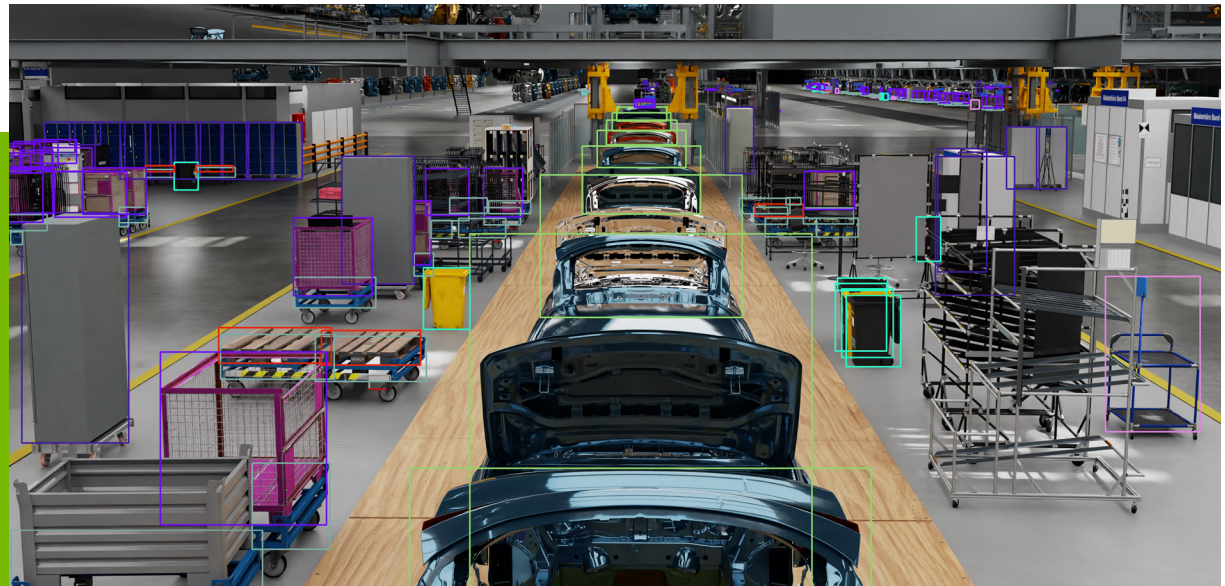


BMW GROUP

Boosting data science productivity by 8x.

The BMW Group, a global leader in premium and luxury automotive manufacturing, has a century-long history of innovation. The company uses AI in their manufacturing processes to optimize production efficiency, enhance quality control, and streamline supply chain management. To deploy AI at scale, BMW needed a centralized platform with powerful compute capabilities for faster model training and development.

Leveraging their AI factory powered by the NVIDIA DGX platform, BMW established a comprehensive deep learning operations pipeline. This infrastructure supports various industrial AI applications, allowing numerous developers to run training jobs seven days a week and manage large datasets exceeding 500,000 images. The computing power of the DGX systems meant larger, more complex models could be trained, resulting in an impressive 8x boost in data scientists' productivity.



Embracing the concept of Industry 5.0, BMW is using generative AI to improve quality control, simulate production scenarios, and enhance supply chain management. To overcome the challenges of limited data availability and quality, the company is using NVIDIA Omniverse™ to create virtual factories and simulate complex scenes and DGX systems to create synthetic data based on those simulations. A significant outcome of this initiative was the development of the SORDI.ai dataset, the largest photorealistic, open-source dataset for industrial AI.

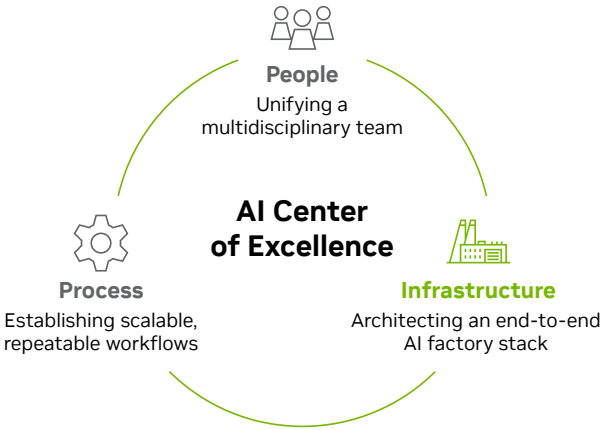
> BMW's AI factory has drastically reduced AI model development time, empowering employees to efficiently create and deploy AI solutions across various downstream tasks.

Learn More >

The Third Pillar: Infrastructure

Build a foundation for high-performance AI.

World-class infrastructure is the backbone of a modern AI CoE. On-premises systems—built with optimized components for training, retraining, inference, and deployment—ensure high performance, security, and cost predictability for enterprise AI. Solutions like DGX SuperPOD, powered by GPU-driven DGX systems and Intel Xeon CPUs, offer multi-node compute, leading-edge GPUs, low-latency networking, high-speed storage, and robust orchestration. These components form an AI factory capable of powering deep learning and generative AI at scale. While cloud resources can offer flexibility for temporary, seasonal training surges, core, steady-state AI operations are most efficiently and predictably managed on a centralized, purpose-built, on-premises system



MITRE

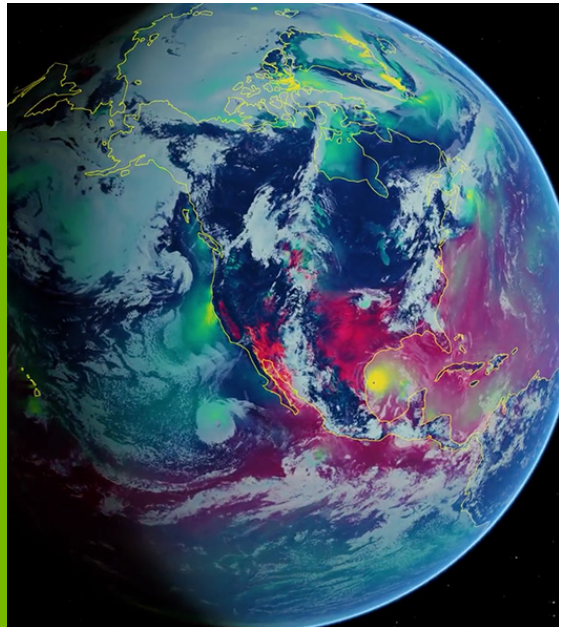
Powering AI Innovation in the public sector

MITRE's Federal AI Sandbox serves as a cornerstone of public-sector AI excellence, transforming how federal agencies innovate with AI. Powered by DGX SuperPOD, the Federal AI Sandbox delivers a 300x increase in computing performance over MITRE's previous systems, giving thousands of researchers a unified, high-performance environment for rapid model development, training, and deployment. This advanced infrastructure seamlessly integrates three critical computing components: edge AI systems for real-time data collection and processing, large-scale AI model training on DGX SuperPOD, and Omniverse technologies for robust simulation and modeling. The resulting "virtuous loop" accelerates breakthroughs in domains from ocean to meteorological science.

MITRE's AI infrastructure advancements are transforming federal missions. With it, agencies and their teams are developing AI-powered weather maps over water with one-kilometer precision, optimizing shipping routes and securing critical supply chains against disruptions, pioneering a cybersecurity foundational model that enhances threat detection and response across 190 nations, and streamlining benefits administration by automating complex workflows and improving service delivery across departments.

> By securing control over models and data and enabling rapid AI development and deployment, MITRE's Federal AI Sandbox is fast-tracking innovation in national security, weather forecasting, and public service—powering solutions with real impact for the public good.

Learn More >



An AI CoE is not just a strategic initiative—it's an essential foundation for sustained competitive advantage in the era of AI.

As part of their AI CoE initiative, organizations can invest in a centralized, full-stack solution such as NVIDIA DGX SuperPOD. This approach can unlock productivity for developers and IT operations, spur optimal business-aligned outcomes from AI initiatives, and increase ROI. By strategically combining the three pillars of people, process, and infrastructure into a unified strategy, a business can transition from fragmented AI projects to a cohesive, impactful, and sustainable AI-driven future.

Ready to Get Started?

Build your AI CoE on NVIDIA DGX SuperPOD, powered by NVIDIA GPUs and Intel Xeon CPUs: nvidia.com/dgx-superpod

© 2025 NVIDIA Corporation and affiliates. All rights reserved. NVIDIA, the NVIDIA logo, DGX, DGX SuperPOD, NGC, and Omniverse are trademarks and/or registered trademarks of NVIDIA Corporation and affiliates in the U.S. and other countries. Other company and product names may be trademarks of the respective owners with which they are associated. 4262500. DEC25

NVIDIA