

Simplifying GenAI development

Dell Precision AI workstations
and NVIDIA AI Workbench

[Get Started](#)



Table of contents

- Generative AI (GenAI) development challenges 3
- The solution: NVIDIA AI Workbench on Dell Precision workstations 4-7
- Success story: Invoke and Invoke Enterprise 8
- NVIDIA AI Workbench and NVIDIA AI Enterprise 9
- Next steps and resources 10

GenAI development challenges

From setting up GPUs and Jupyter notebooks to moving workloads to managing workflows, the everyday challenges of GenAI development can have a negative impact on project progress. The required time and technical expertise can seem daunting for data scientists, engineers, and developers. Fortunately, the right technology partners and tools can help address obstacles and accelerate development.

Common GenAI obstacles



Complex
GPU setup



Limited workload
portability



Time-consuming
workflow management



The solution: NVIDIA AI Workbench on Dell Precision AI-ready workstations

With NVIDIA AI Workbench on Dell Precision workstations, expert and novice developers can quickly and easily create, test, and customize GenAI projects on local workstations, then scale them up or down to any platform. Do it all on Dell Precision AI-ready workstations with the high performance required to process large data sets and complex neural networks.

NVIDIA AI Workbench

- ✓ Streamline setup
Start coding and customizing models faster with automated GPU system setup/configuration
- ✓ Run projects anywhere
Develop locally and move workloads easily to run on the best platform for cost, speed, and scale
- ✓ Accelerate AI workflows
Improve productivity with automation for container and version management

Dell Precision workstations

- ✓ Local GenAI development
Customize and develop GenAI models locally on GPU-accelerated workstations
- ✓ Scalable
Scale up from one to four NVIDIA RTX™ Ada Generation GPUs
- ✓ AI-ready
Designed and configured to meet your AI workload needs

Streamline hardware setup

Productivity suffers when developers have to deal with the complex process of setting up a GPU machine for deep learning and GenAI work. NVIDIA AI Workbench eliminates common setup issues such as OS/driver mismatches and CUDA driver and firmware incompatibilities. Developers can start coding quickly with automated software and driver installation and configuration. NVIDIA AI Workbench also helps ensure containers are configured for optimal GPU efficiency based on your workload.



Increase productivity with automated setup



Eliminate common hardware setup issues



Ensure containers are configured correctly



Run projects anywhere

NVIDIA AI Workbench gives developers the freedom to create, customize, and collaborate easily with seamless workload portability. Move NVIDIA AI Workbench projects to the best GPU-enabled system for cost, speed, and scale securely. Do it all without the time and technical skill typically required. Environment, data, model, and location changes and versioning can be addressed with ease while NVIDIA AI Workbench securely handles the authentication and secrets needed for users to access different resources.



Port workloads
easily



Move credentials and
secrets securely



Addresses changes and
versioning simply

Accelerate AI workflows



Automate project installation
and management



Track project dependencies
automatically



Simplify project management
with intuitive UI and CLI

With NVIDIA AI Workbench, novice and expert developers can quickly identify, install, and manage elements of AI and deep-learning projects. Give productivity a boost with the automated installation of multiple project resources. Easily track project dependencies for a given model and GPU so developers can focus on execution. NVIDIA AI Workbench simplifies workflow management with an intuitive user interface and powerful command-line interface.

See workflow examples on GitHub:

- ↪ [Hybrid RAG Project](#)
- ↪ [NIM Anywhere Project](#)
- ↪ [Stable Diffusion Project](#)
- ↪ [Fine-Tuning Project](#)



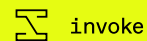
Success story: Invoke Enterprise

Invoke used an NVIDIA AI Workbench project to enable companies to train AI image models securely and collaboratively. Invoke deployed both Invoke Enterprise and NVIDIA AI Workbench on-premises, running AI compute locally on [Dell Precision 7780 mobile workstations](#) powered by NVIDIA RTX™ 5000 Ada Generation GPUs.

“Deploying generative AI models on Dell Precision AI-ready workstations, particularly the Precision 7780, has set a new standard for performance. Older hardware can struggle with the demands of AI compute, but running Invoke on the 7780 is seamless, powerful, and supports all of our most advanced AI workflows.”

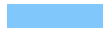
— Kent Keirse, Founder and CEO, Invoke

[Read case study](#)

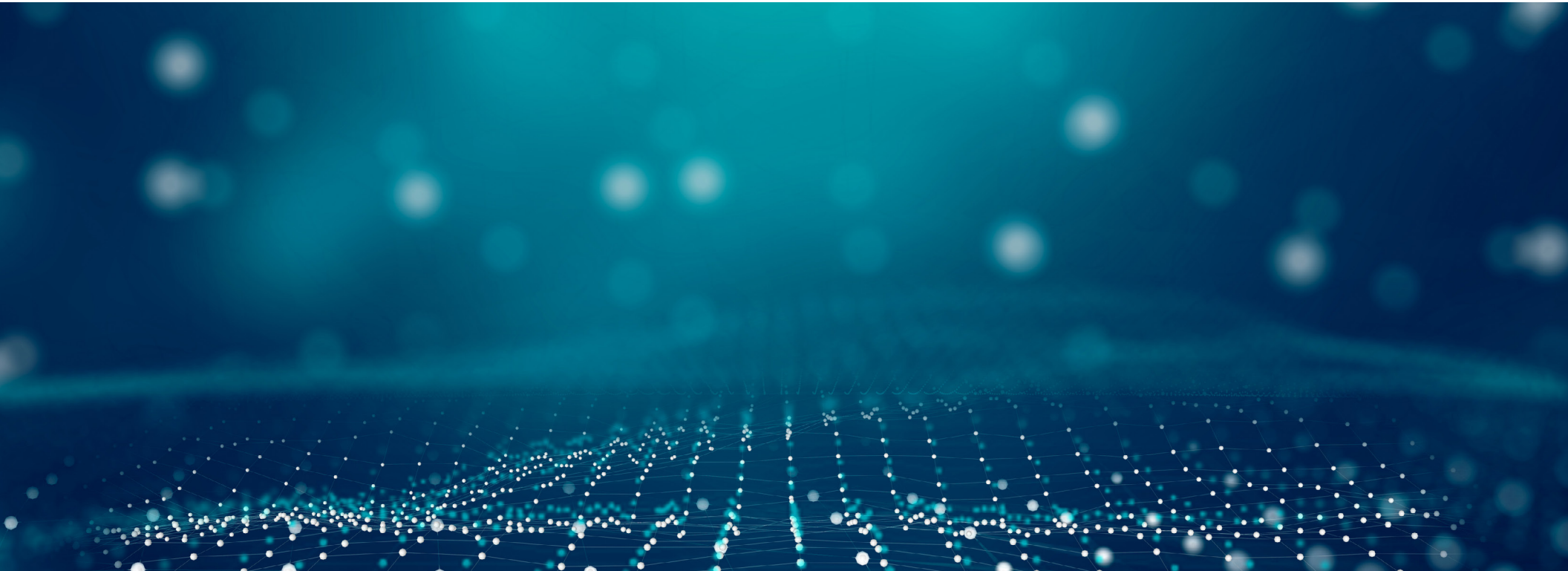




NVIDIA AI Workbench and NVIDIA AI Enterprise



Combine NVIDIA AI Workbench with NVIDIA AI Enterprise to access an end-to-end, cloud-native software platform that accelerates data science pipelines and streamlines development and deployment of production-grade AI applications. NVIDIA AI Workbench is free and comes with enterprise support for NVIDIA AI Enterprise license holders, including security features, NIM, and production branches.



Next steps and resources

Get started

- ➔ Download AI Workbench for free
- ➔ Find AI-ready workstations configurations
- ➔ Contact us to purchase NVIDIA AI Enterprise through Dell
- ➔ Try NVIDIA AI Enterprise free for 90 days

Learn more

- ➔ Harnessing NVIDIA Tools for Developers with Precision AI-Ready Workstations blog
- ➔ Decoding How NVIDIA AI Workbench Powers App Development blog
- ➔ Optimize AI Model Performance and Maintain Data Privacy with Hybrid RAG blog
- ➔ Develop and Deploy Scalable Generative AI Models Seamlessly with NVIDIA AI Workbench blog