NVIDIA

Our Body of Work

NVIDIA pioneered accelerated computing to tackle challenges no one else can solve.

Our work in AI and digital twins is transforming the world's largest industries and profoundly impacting society.

Jason O'Rear

Pioneering Accelerated Computing

Accelerated computing requires full-stack optimization, from chip architecture, systems, and acceleration libraries, to refactoring the applications.

The global NVIDIA ecosystem spans 4 million developers, 40,000 companies, and over 3,000 applications.

Application Frameworks

MODULUS • MONAI • MAXINE • NEMO • AVATAR • DRIVE • ISAAC • METROPOLIS • HOLOSCAN

Platform

NVIDIA AI • NVIDIA Omniverse

Acceleration Libraries

RTX
CUDA-X
CUDA

System Software

Magnum IO • DOCA • Base Command • Forge

Hardware

RTX • DGX • HGX • EGX • OVX • AGX • MLNX

GPU • CPU • DPU • NIC • Switch • SOC

# Sparking the iPhone Moment of AI

The acceleration of deep learning ignited the big bang of AI. ChatGPT, a large language model powered by an NVIDIA DGX™ AI supercomputer, reached 100 million users in just two months. Its magical capabilities have captured the world's imagination. Generative AI is a new computing platform, like the PC, internet, and mobile-cloud. Accelerated computing and AI have fully arrived.

> What's the definition of a large language model?

> A large language model is a type of artificial intelligence system that has been trained on massive amounts of text data and can generate human-like language responses to input.
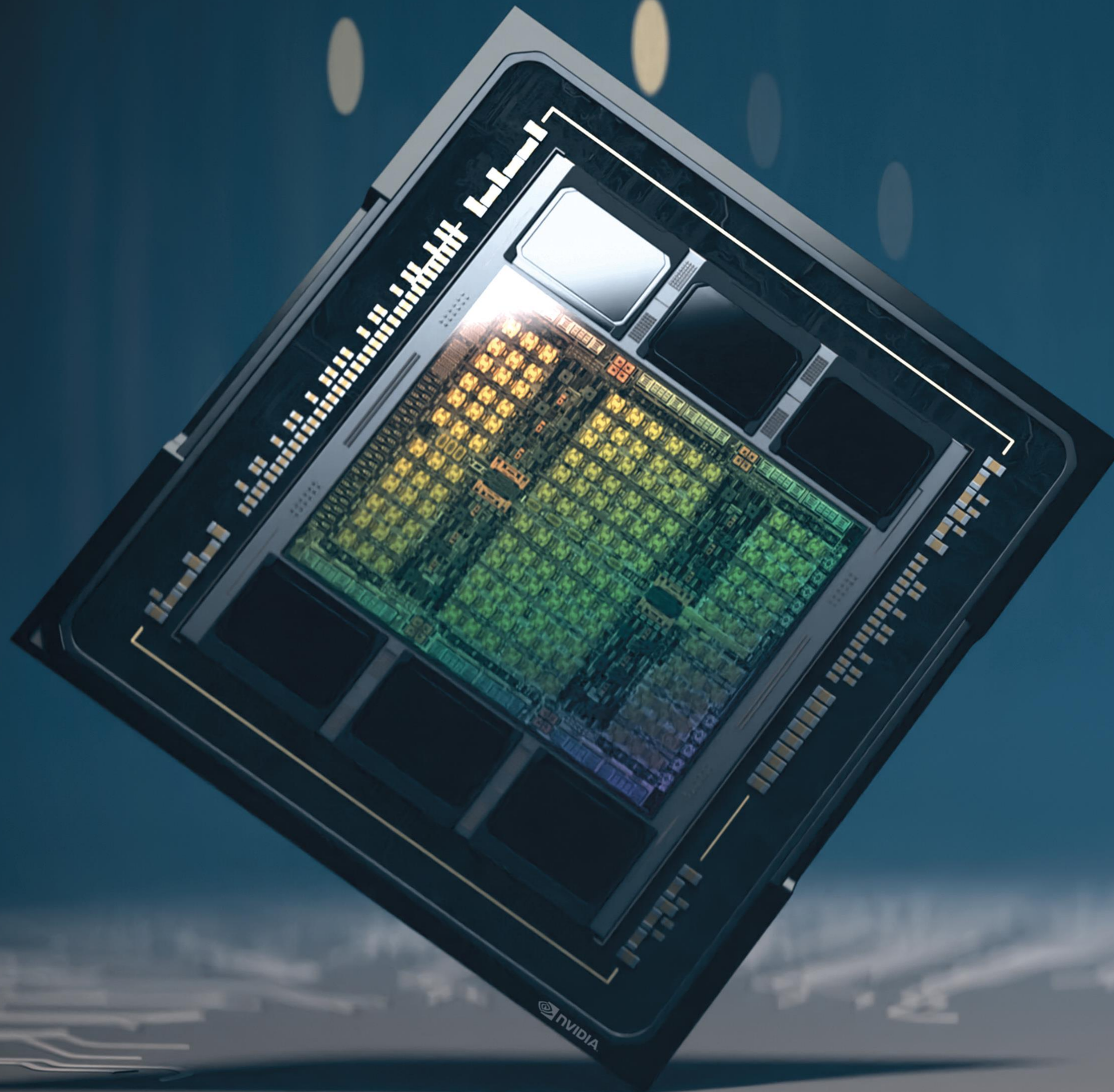
Advancing the World's Largest Industries

NVIDIA's acceleration libraries solve new challenges and open new markets. They connect to applications that connect to the world's industries, forming a network of networks.

We recently launched NVIDIA cuLitho, a new library that supercharges computational lithography, an immense computational workload in chip design and manufacturing.

NVIDIA now offers some 300 acceleration libraries.

Leaping Into the Cloud

NVIDIA accelerated computing is available in every cloud. DGX Cloud and NVIDIA AI Enterprise make our unparalleled AI capabilities accessible to everyone. NVIDIA AI Foundations is a cloud service for building custom language models and generative AI. And NVIDIA Omniverse Cloud will accelerate the digitalization of the world's largest industries.
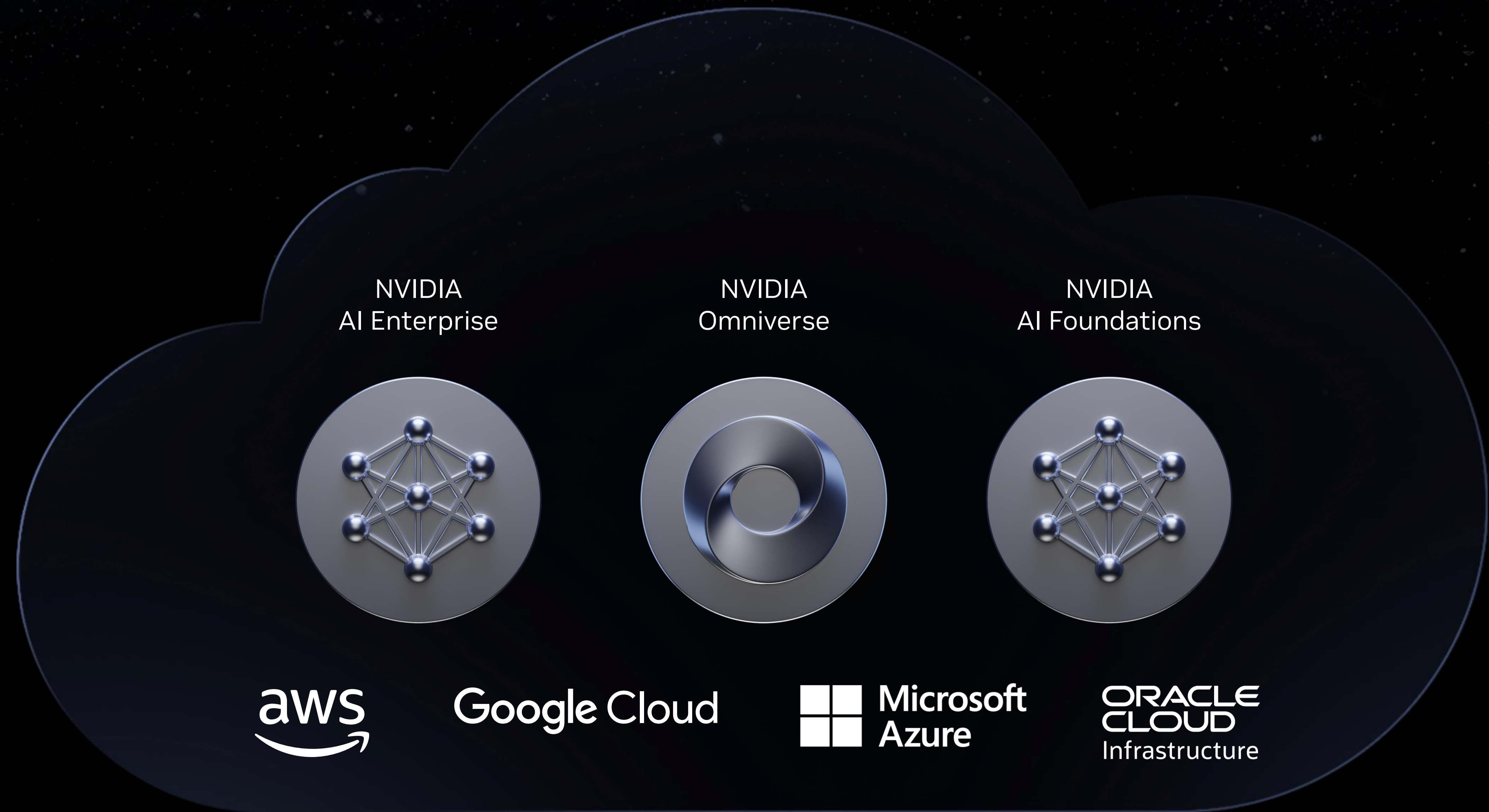
NVIDIA Accelerated Computing Is Sustainable Computing

Data centers are already about 1-2% of global electricity consumption and that consumption is expected to continue to grow. This continued growth is not sustainable.

If we switched accelerated computing workloads from CPU-only servers to GPU-accelerated systems worldwide, we estimate nearly 12 trillion watt-hours of energy savings a year, equivalent to the electricity requirements of nearly 1.7 million U.S. homes.

Acceleration is the best way to reclaim power and achieve sustainability and net zero.

FourCastNet
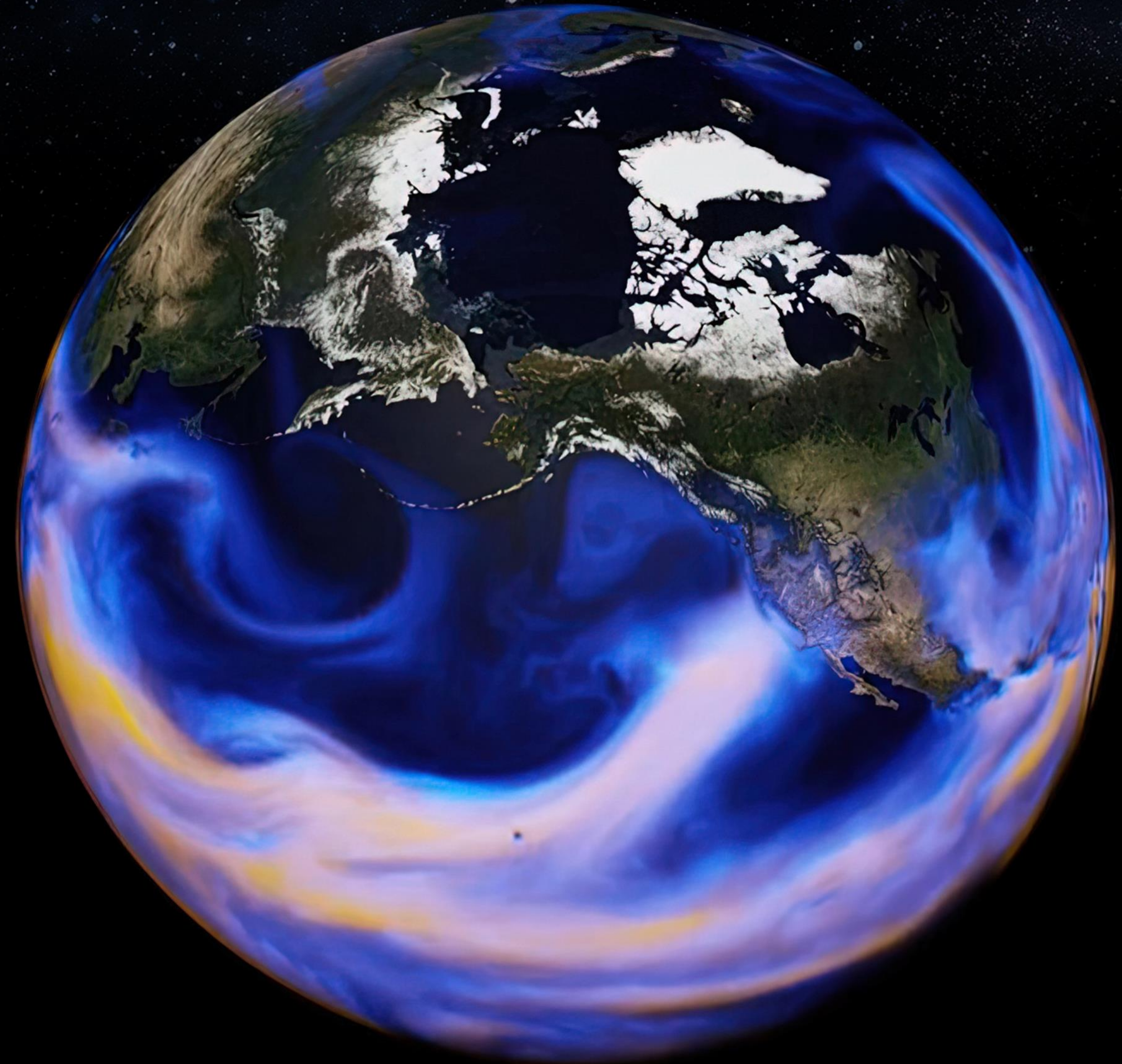
NVIDIA Reinvents Modern Graphics

We invented the programmable shading GPUs nearly a quarter century ago, defining modern real-time computer graphics.

With NVIDIA RTX™ we have reinvented computer graphics again. This new rendering approach fuses rasterization and programmable shading with ray tracing and AI to make PC games look much more beautiful and realistic—almost cinematic.

NVIDIA RTX
Resets Gaming

RTX is everywhere. More than 400 games and apps now use RTX to deliver stunning ray-traced graphics—including AAA blockbusters like *Cyberpunk 2077*, *Fortnite*, *Minecraft*, and more.
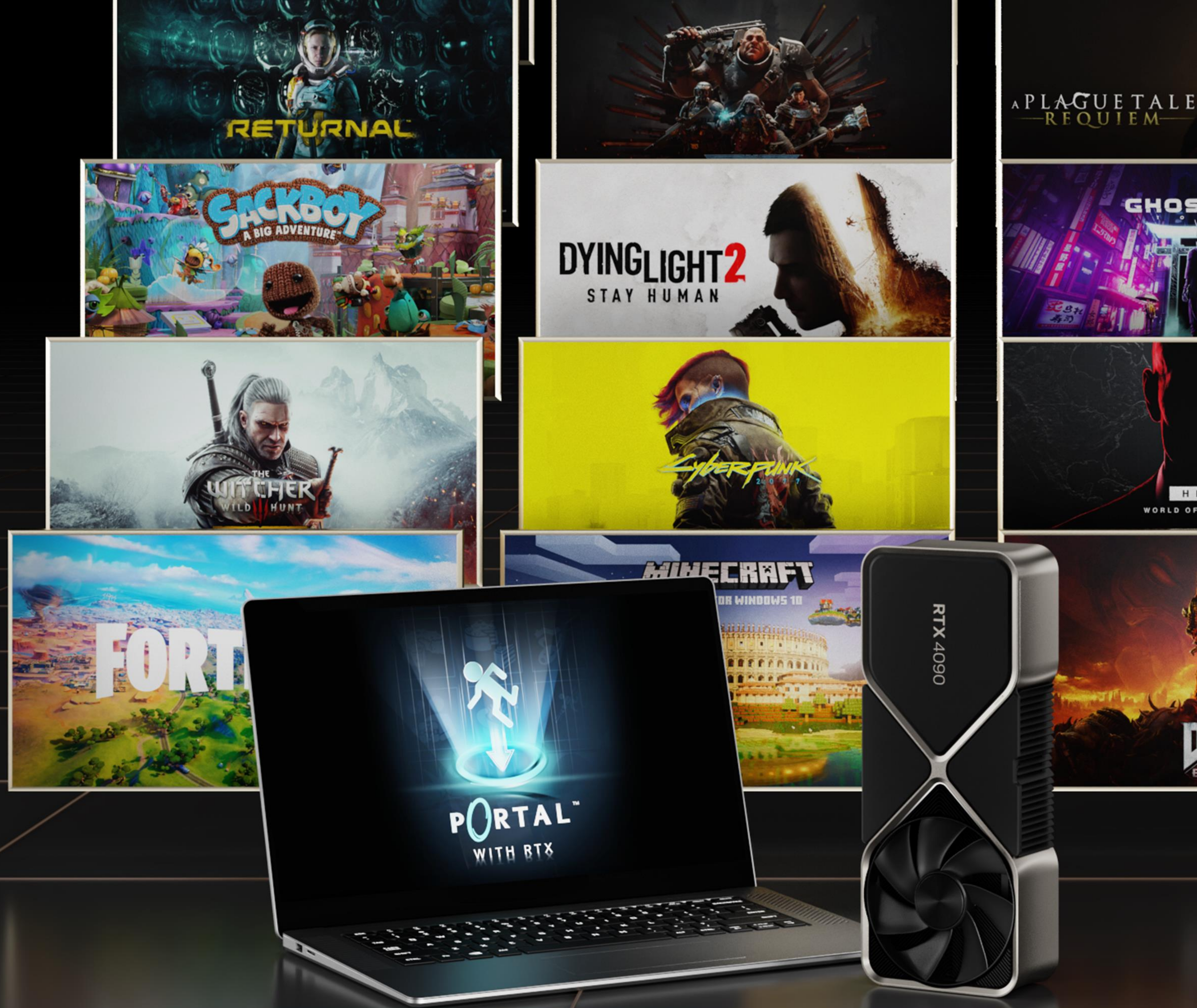
NVIDIA Cloud Gaming—Bringing RTX to Billions

With the power of NVIDIA® GeForce® GPUs in the cloud, GeForce NOW™ instantly transforms nearly any device into a powerful PC gaming machine. Any gamer can stream titles from the top digital game stores. Over 25 million members in 100+ countries now have access to more than 1,500 games.

And, recently, NVIDIA and Microsoft signed a 10-year deal to bring the Xbox PC game library to GeForce NOW.

NVIDIA Studio— Accelerated Computing Platform for Creators

Our industry-leading GPUs, paired with our exclusive driver technology and software, enhance creative apps with a level of performance and ability that is nothing short of inspiring. With NVIDIA Studio, creators are free to realize their most ambitious projects yet.

NVIDIA Powers AI Factories

Data centers process mountains of continuous data to train and refine AI software. Companies are manufacturing intelligence, and their data centers are becoming giant AI factories. NVIDIA is the engine of the world's AI infrastructure.
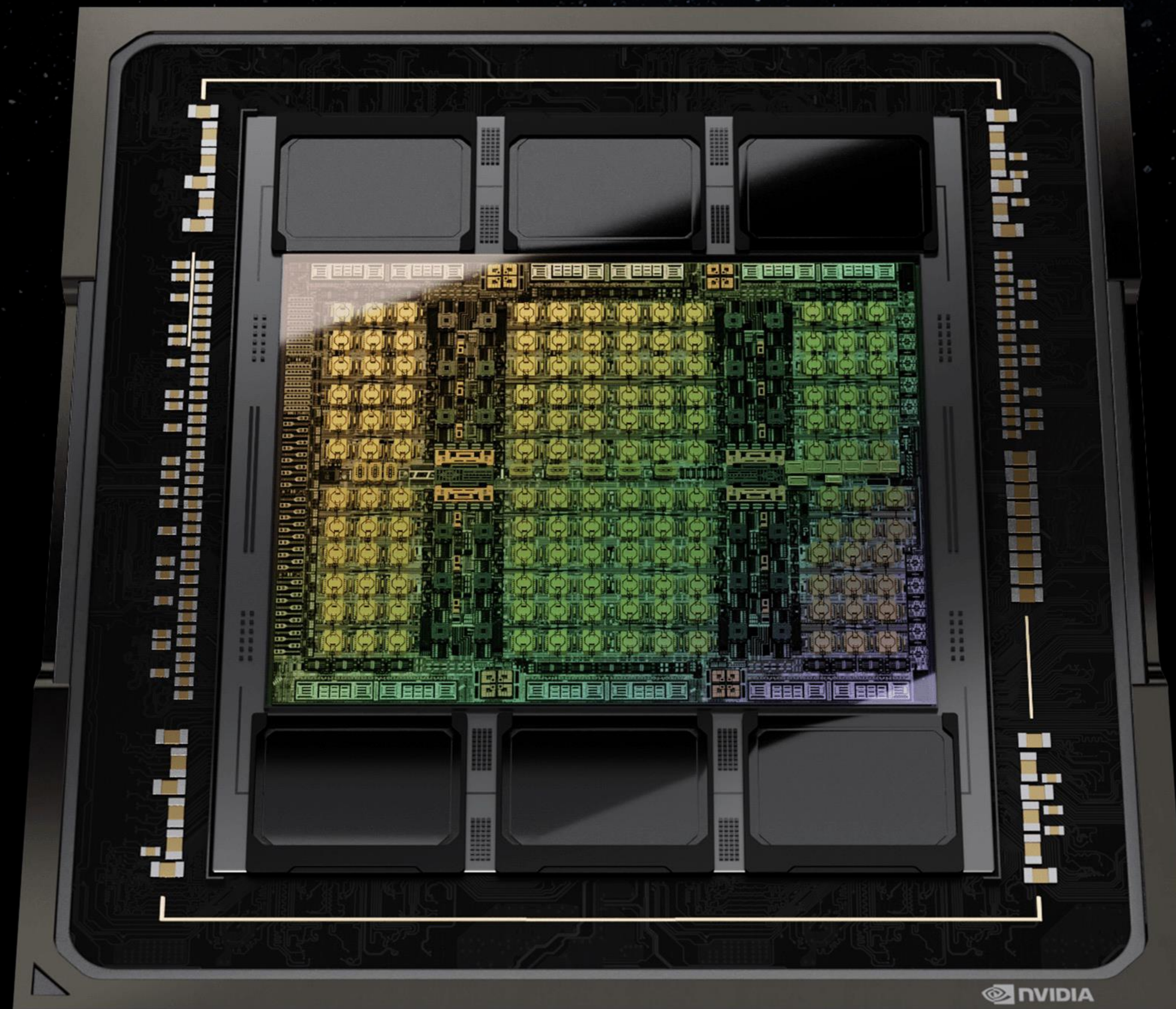
Hopper—
The Engine for the
World's AI Infrastructure

The NVIDIA Hopper™ architecture will power
the next wave of AI data centers. The first
Hopper-based GPU, the NVIDIA H100, comes
packed with 80 billion transistors and delivers
an order-of-magnitude performance leap over
its predecessor.

NVIDIA DGX—Purpose-Built for the Unique Demands of AI

Our fourth-generation NVIDIA DGX system is the world's first AI platform to be built with the new H100 GPUs. Each DGX H100 provides 32 petaflops of AI performance at FP8 precision—6X more than the prior generation. The next-generation DGX SuperPOD™ will expand the frontiers of AI with the ability to run massive workloads with trillions of parameters.

Accelerating Inference for Generative AI Models

This is NVIDIA's inference platform—one architecture for diverse AI workloads and maximum data center acceleration and elasticity: L4 GPUs for AI video; L40 GPUs for Omniverse and graphics rendering; H100 NVL for scaling out large language model inference; and Grace Hopper Superchip for recommender systems and vector databases.
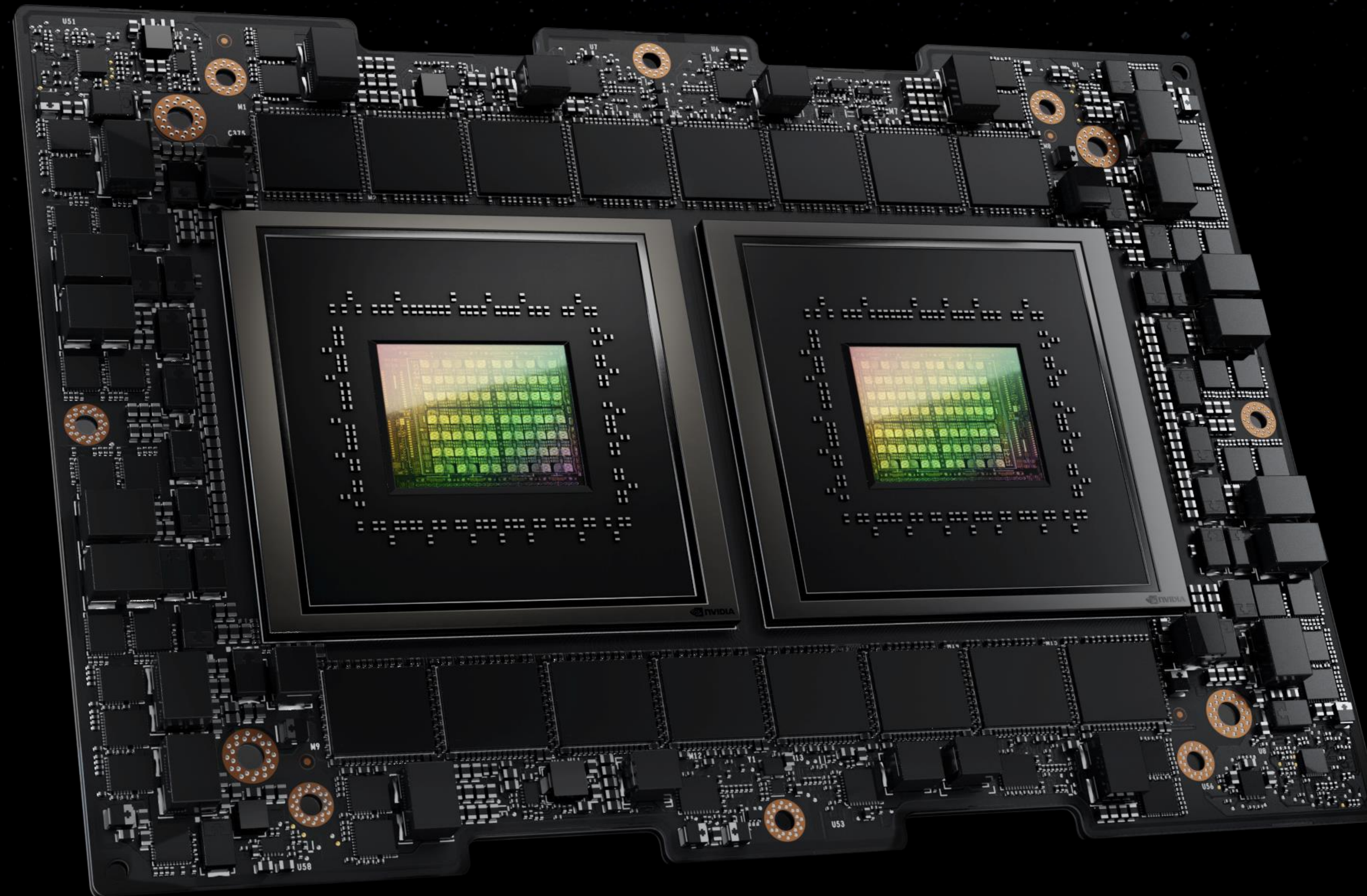

L4   L40   H100 NVL   Grace Hopper

Grace CPU Superchip—Tailor-Made for the Largest Computing Problems

Designed to process giant amounts of data, Grace will be the ideal CPU for AI factories. Grace has 144 CPU cores and one terabyte per second of memory bandwidth—over 2-3X the top Gen 5 CPUs that have yet to even ship.

Data centers must accelerate every workload. The energy saved can fuel new growth. We designed Grace for high energy-efficiency at cloud data-center scale.

Every Hyperscale Data Center Can Now Be a Generative AI Data Center

NVIDIA Spectrum-X™ is an accelerated networking platform designed to improve the performance and efficiency of Ethernet-based AI clouds. Spectrum-X is supercharged by NVIDIA acceleration software and software development kits, allowing developers to build software-defined, cloud-native AI applications.
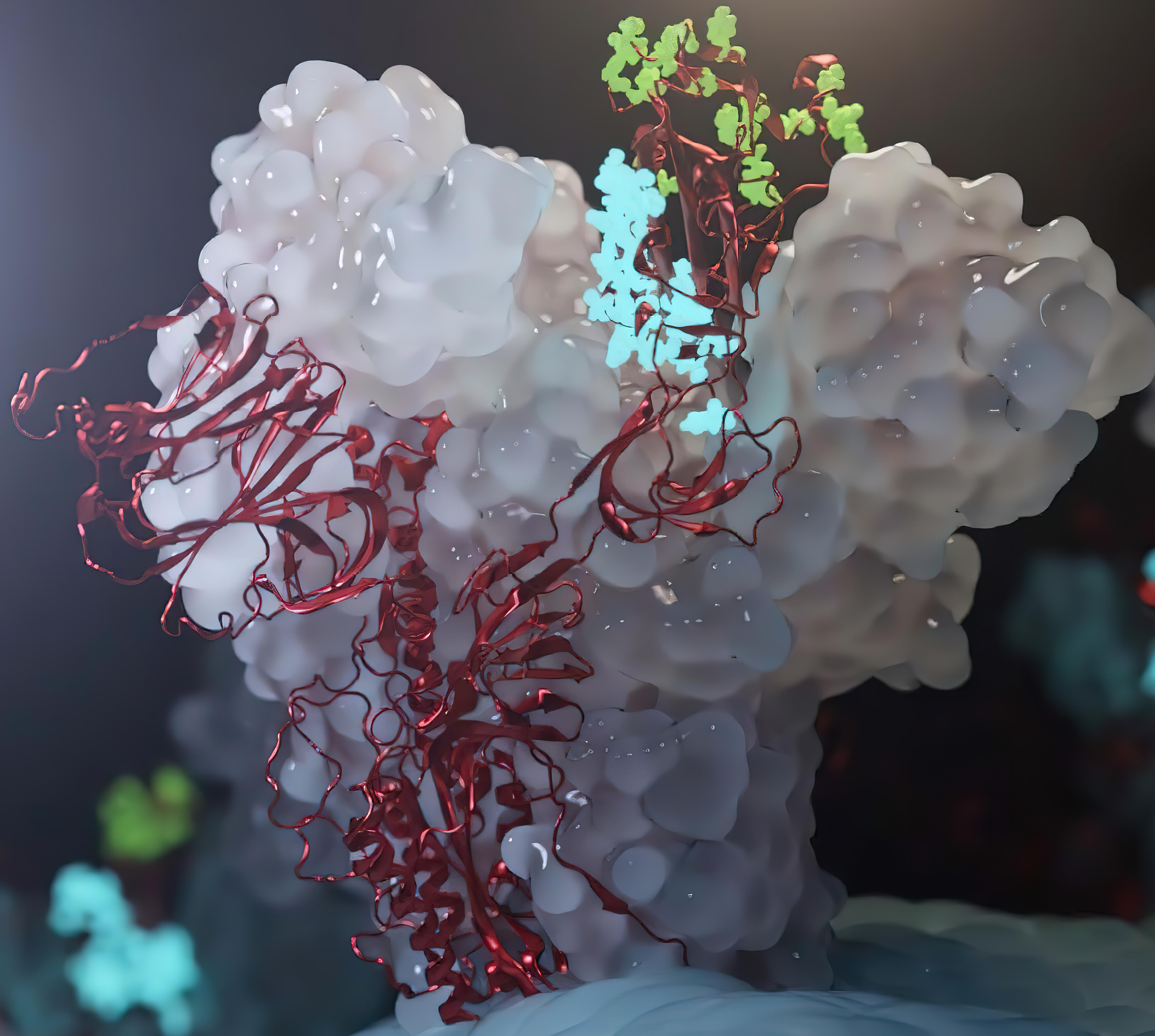

BlueField-3
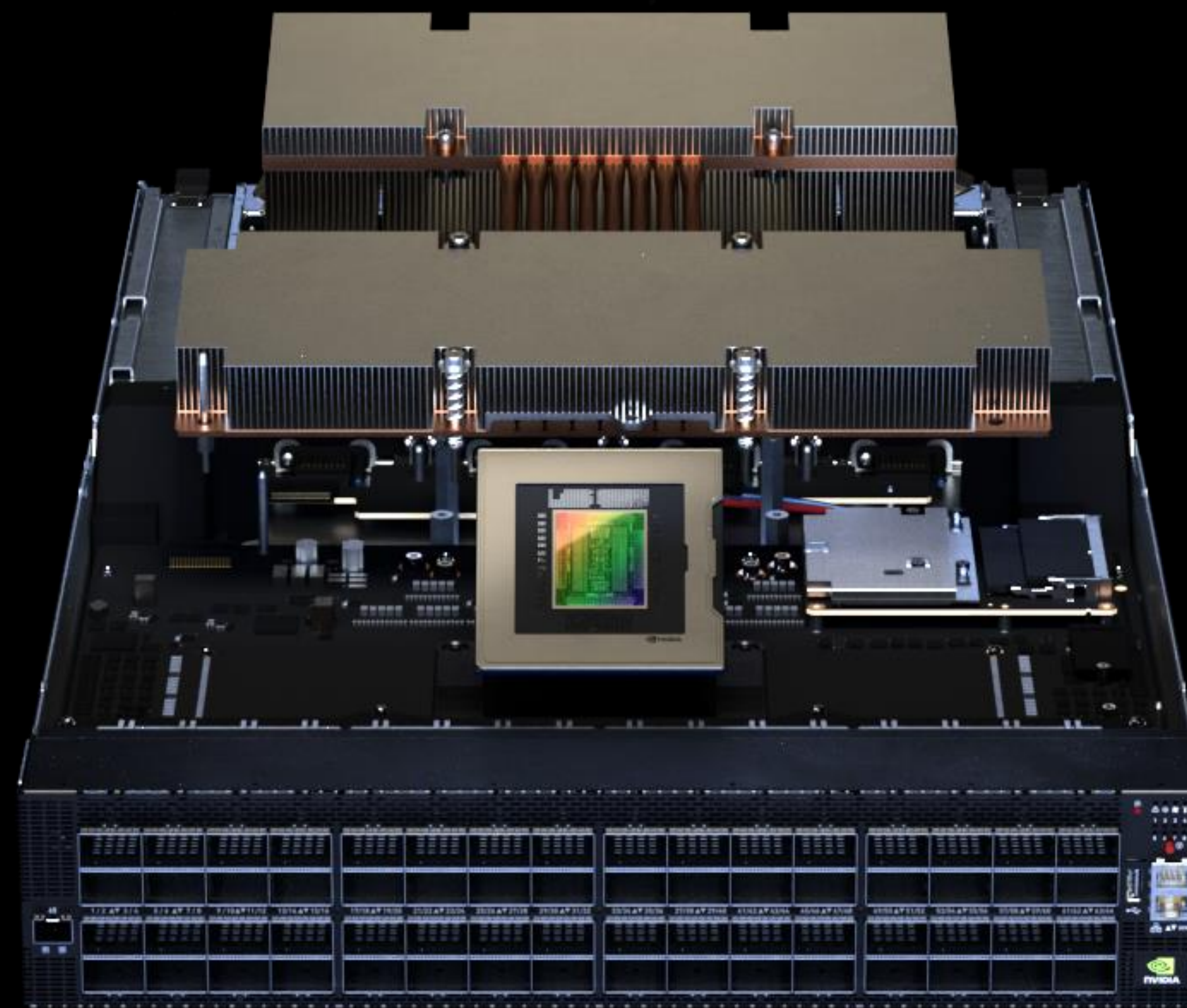
Spectrum-4

BlueField-3

NVIDIA DGX GH200—A New Class of AI Supercomputer

The most efficient large memory supercomputer, the DGX GH200 enables the development of next-generation models for generative AI language applications, recommender systems, and data analytics workloads.

NVIDIA Supercharges Healthcare

From medical imaging to drug discovery, genomics to patient monitoring, researchers across life sciences are fusing traditional simulations and AI to solve the next grand challenges.

Generative AI Will Transform the Pharmaceutical Industry

Drug discovery is a nearly $2 trillion industry with $250 billion dedicated to R&D. The industry is now jumping onto generative AI to discover disease targets, design novel molecules or protein-based drugs, and predict the behavior of the medicines in the body. NVIDIA BioNeMo provides state-of-the-art generative AI models for drug discovery, available from the cloud.
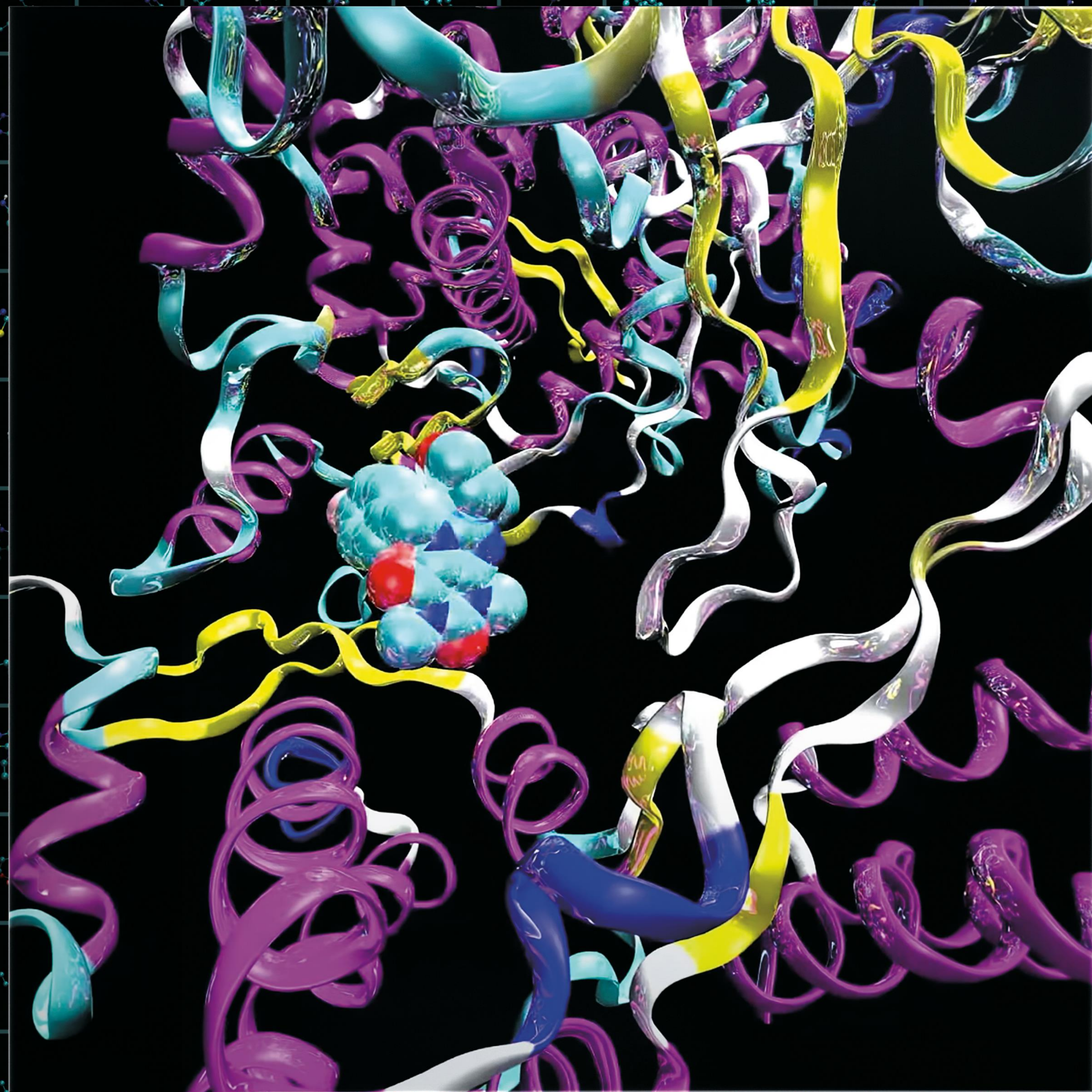
World Record-Setting DNA Sequencing Technique Helps Clinicians Rapidly Diagnose Patients

Researchers at Stanford using NVIDIA accelerated computing won the Guinness World Record for the fastest DNA sequencing technique, achieved in five hours and two minutes. The method allows clinicians to take a blood draw from a critical-care patient and reach a genetic disorder diagnosis the same day.

NVIDIA AI-Powered Medical Devices

AI-powered medical devices can help clinicians detect and measure anomalies, up-level surgical skills, enhance image quality, and optimize workflows. Here we see an augmented reality overlay of a patient's anatomy from a CT scan, rendered in real time and AI-augmented with NVIDIA Holoscan. The video feed overlay allows the surgeons to clearly view the patient's vascular and tissue structures.

NVIDIA Drives Industrial Digitalization

Omniverse, our platform for industrial digitalization, builds virtual representations of physical things and assets—creating digital twins and connecting digital and physical worlds. Omniverse enables industries grounded in physical processes to become software-defined, realize unified digitalization, and connect large, highly skilled teams.

Connecting Our Physical and Digital Worlds

As AI makes the leap to heavy industry, it needs to understand how to automate, design, navigate, and build based on the physics of our world. Digital twins via NVIDIA Omniverse enable AI to learn in a digital format. Optimizing virtually before deploying changes reduces costs and speeds deployment.
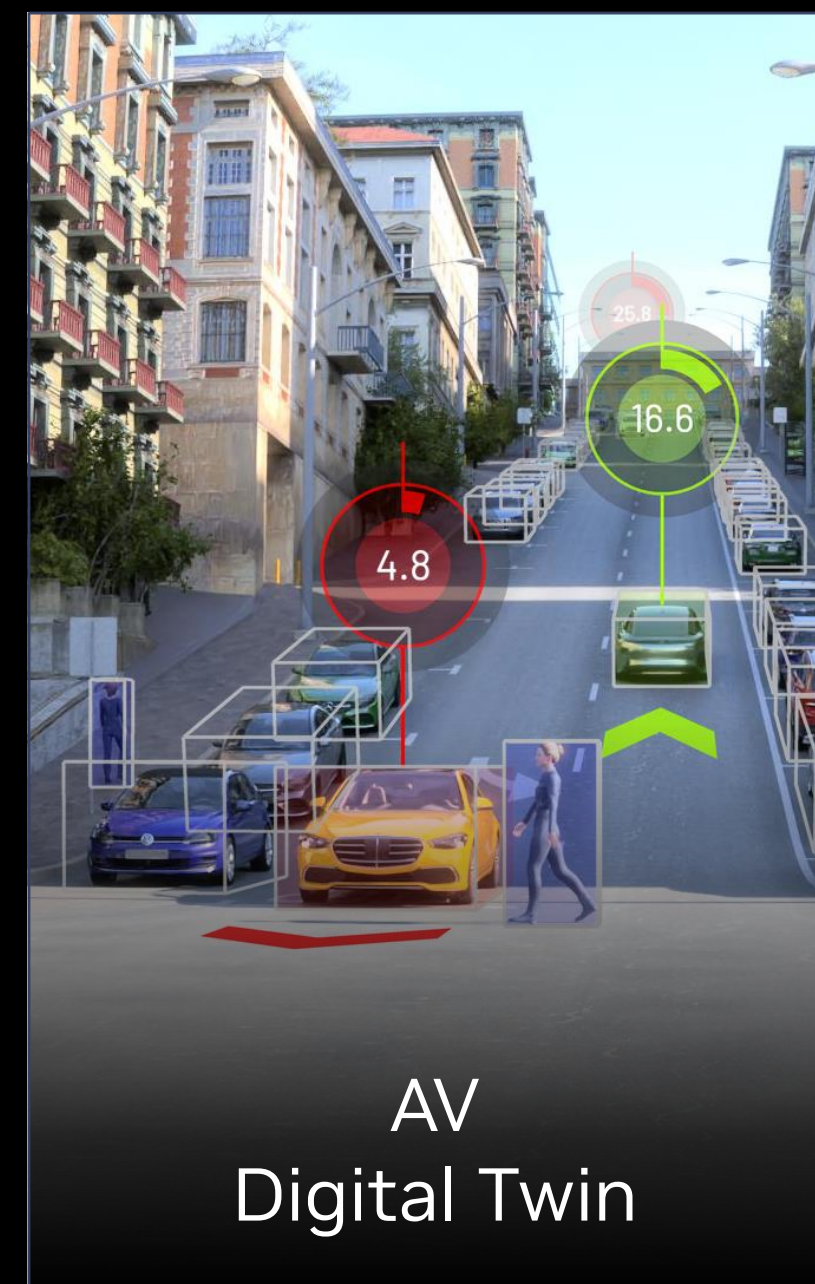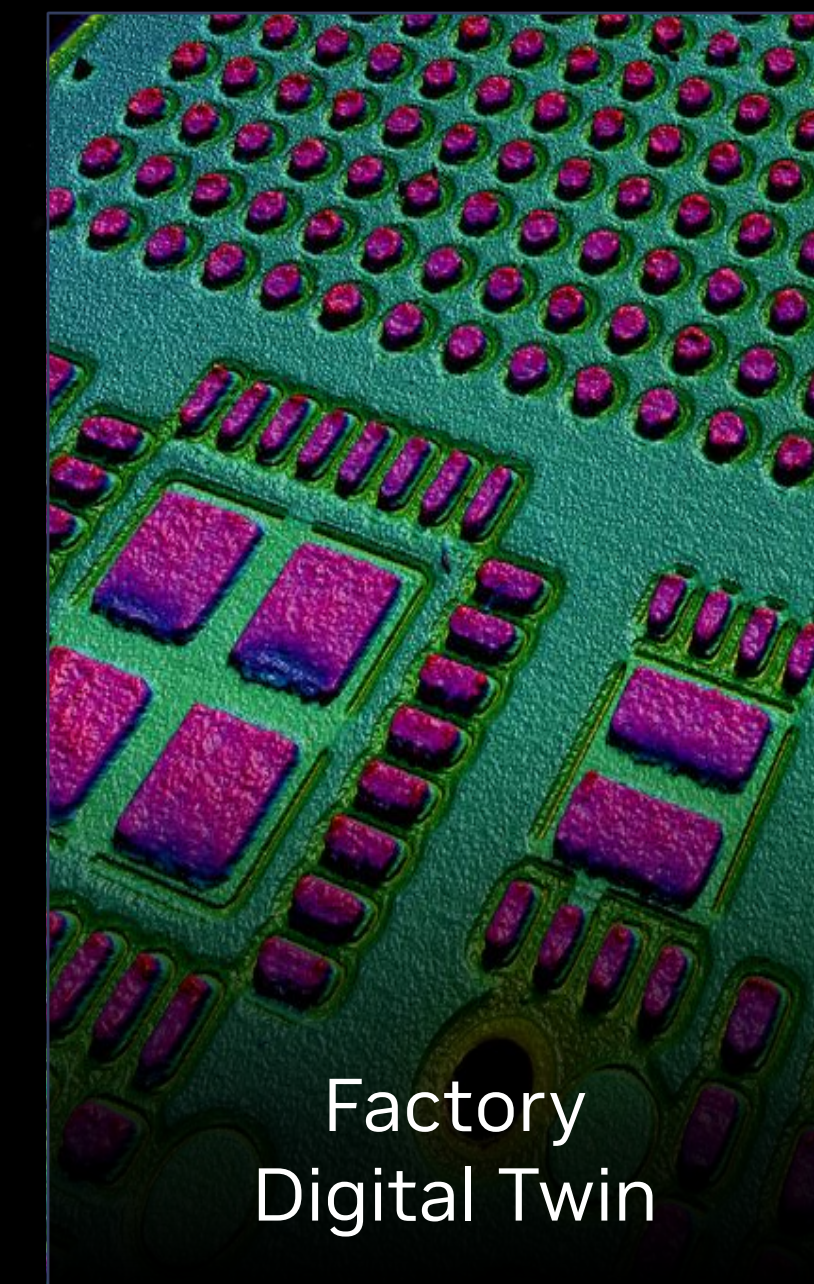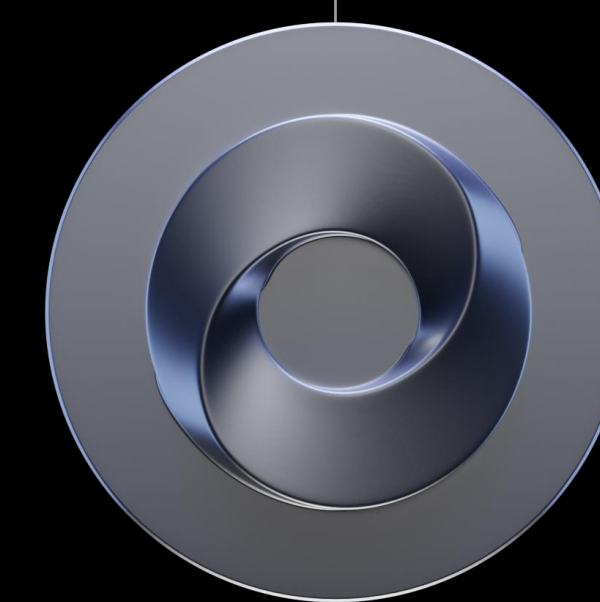
USD COMPOSER

MODULUS

DRIVE

METROPOLIS

ISAAC

METROPOLIS

Design
Digital Twin

Performance
Digital Twin

AV
Digital Twin

Factory
Digital Twin

Robotics
Digital Twin

Warehouse
Digital Twin

NVIDIA
Omniverse
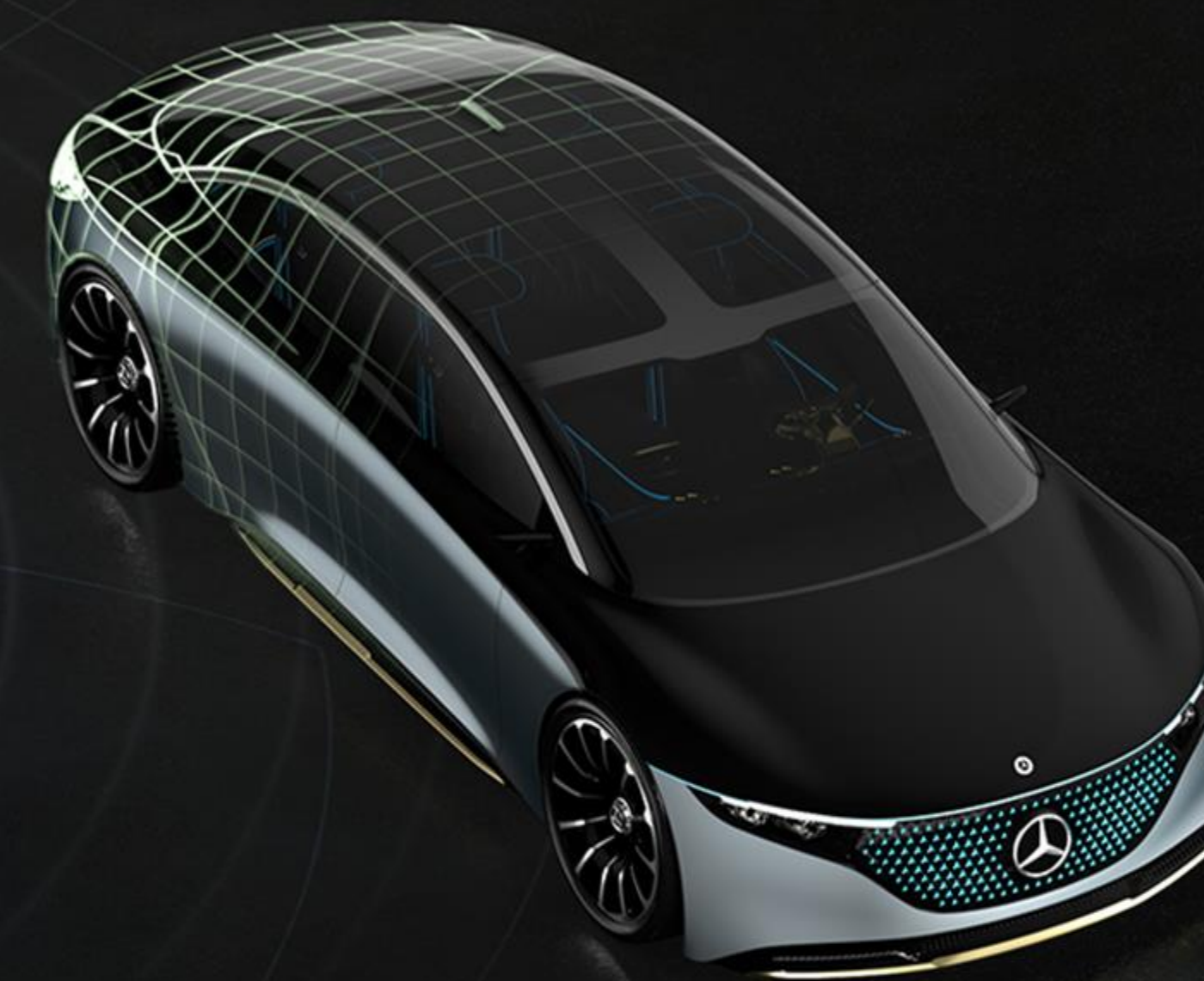
NVIDIA DRIVE—Full Stack Autonomous Driving Platform

The NVIDIA DRIVE® family of products for autonomous vehicle development covers everything from the car to the data center.

DGX Data Center

Hyperion™ 8 Development Car

Omniverse Digital Twin

NVIDIA DRIVE Sim Turbocharges Developer Productivity to Get Self-Driving Cars on the Road

With NVIDIA DRIVE™ Sim, features such as road elevation, road markings, islands, traffic signals, signs, and vertical posts are replicated at centimeter-level accuracy. Autonomous vehicles can drive millions of miles in a wide range of simulated scenarios so they hit the road running, safely.
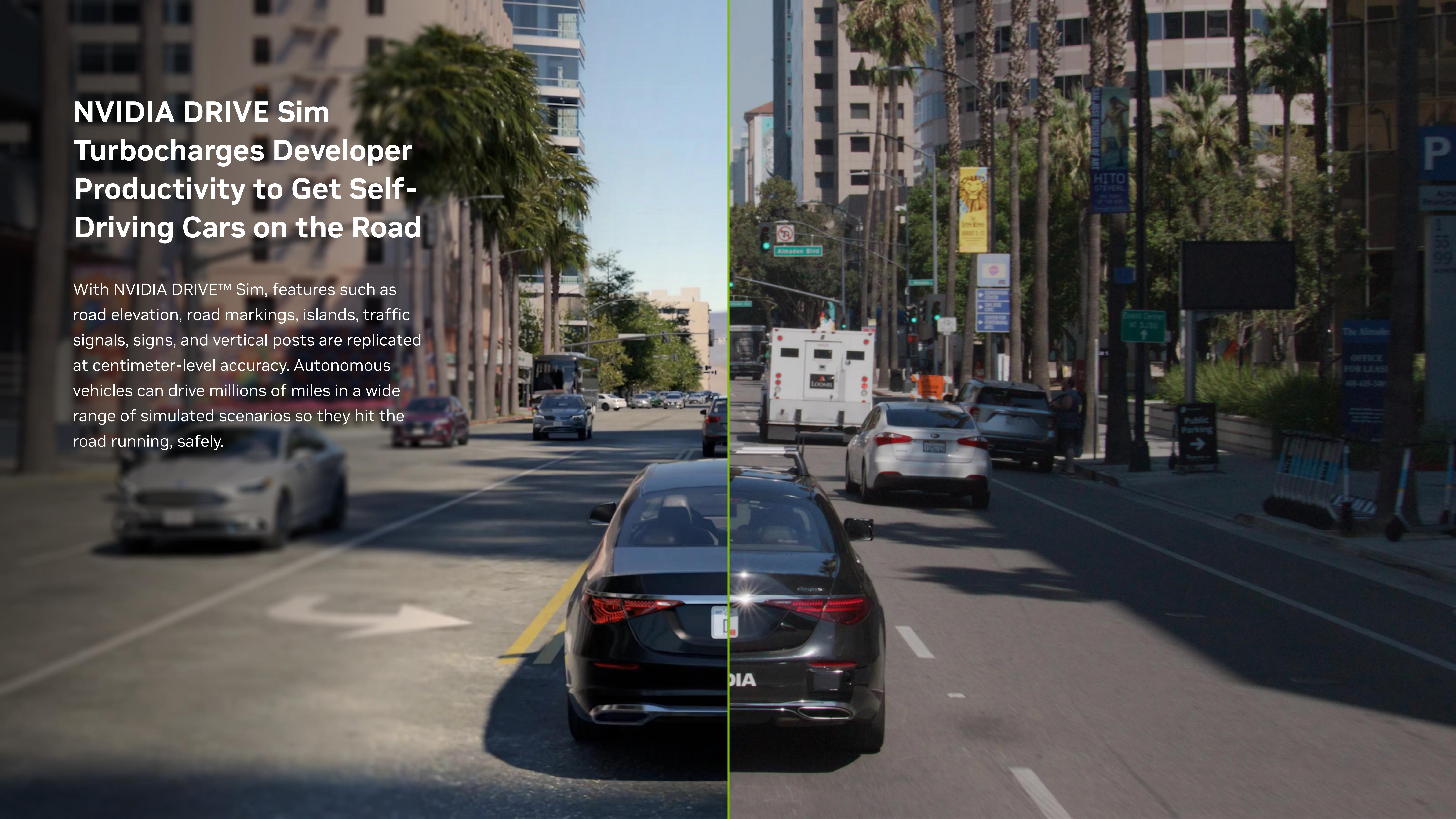
Mercedes-Benz, NVIDIA Partner to Build the World's Most Advanced, Software-Defined Vehicles

Starting in 2024, every next-generation Mercedes-Benz vehicle will include this first-of-its-kind software-defined computing architecture that includes the most powerful computer, system software, and applications for consumers. This marks the turning point of traditional vehicles becoming high-performance, updateable computing devices.

NVIDIA Powers the Software-Defined Vehicle Revolution

Automakers

Mercedes-Benz · JLR · VOLVO

HYUNDAI · KIA · GENESIS

BYD · Li Auto · XPENG

LUCID · LOTUS · Polestar

SAIC · NIO · ZEEKR

Robotaxis

cruise

DEEPROUTE.AI

DiDi

PONY.ai

ZOOX

Trucking

FAW

IVECO

kodiak

Outrider

Plus

Tier 1s

BOSCH

DESAY SV

flex

FOXCONN
HON HAI TECHNOLOGY GROUP

Quanta Computer

ZF GROUP

Omniverse Powers the Robotics Revolution

From smart automation in manufacturing to last-mile delivery, robots are becoming more ubiquitous in everyday life.

The Isaac module in Omniverse is our platform for accelerating and enhancing robotics—from development to simulation to deployment.

Robotic Systems Lab, ETH Zurich & Swiss-Mile
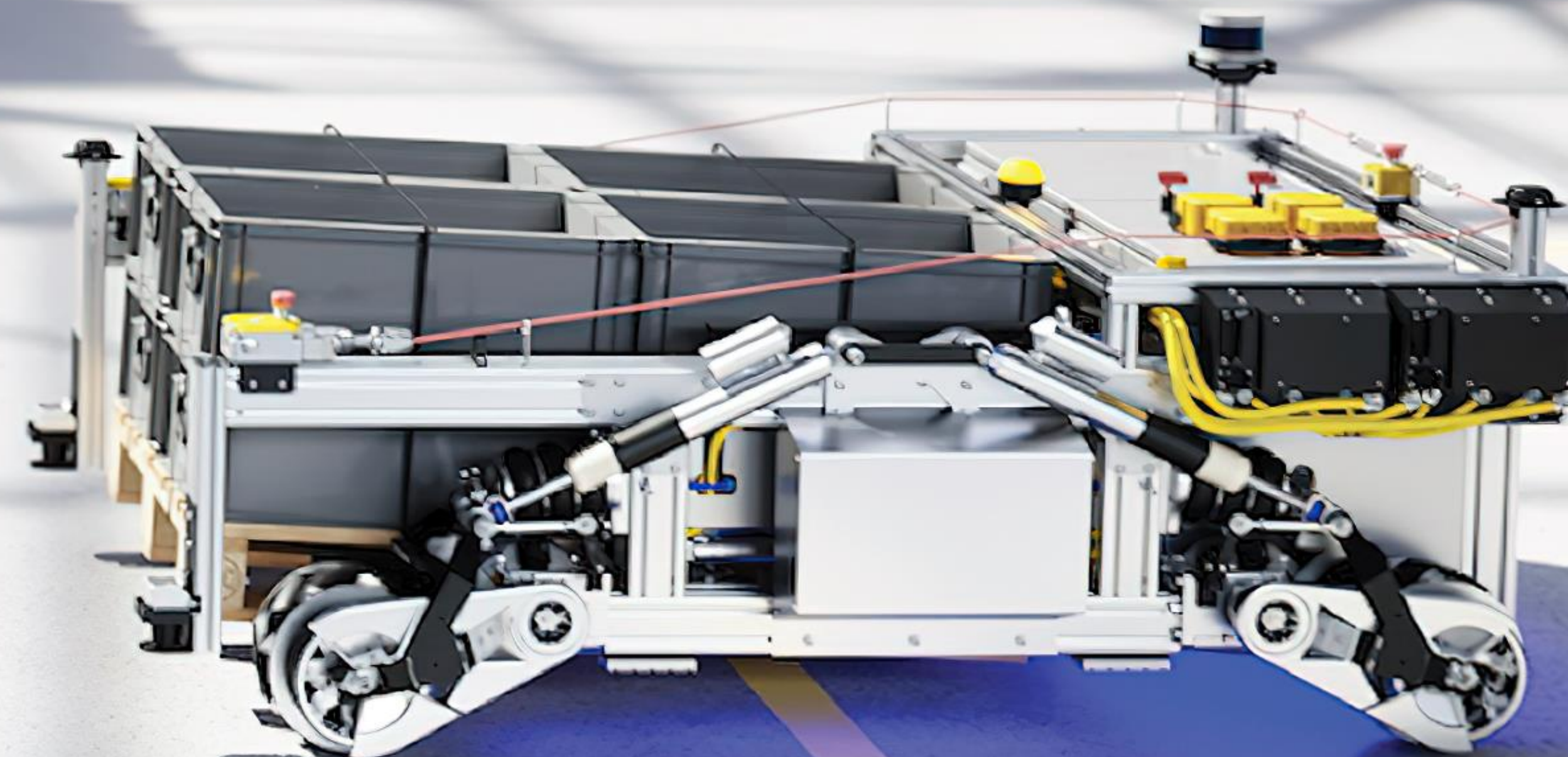
Amazon Robotics Builds Digital Twins of Warehouses in NVIDIA Omniverse

Amazon has over 200 robotics facilities that handle millions of packages each day. Using NVIDIA Omniverse™ Enterprise and Isaac Sim™, Amazon Robotics is building AI-enabled digital twins of its warehouses to better optimize warehouse design and flow, and train more intelligent robotic solutions.

Amazon Robotics
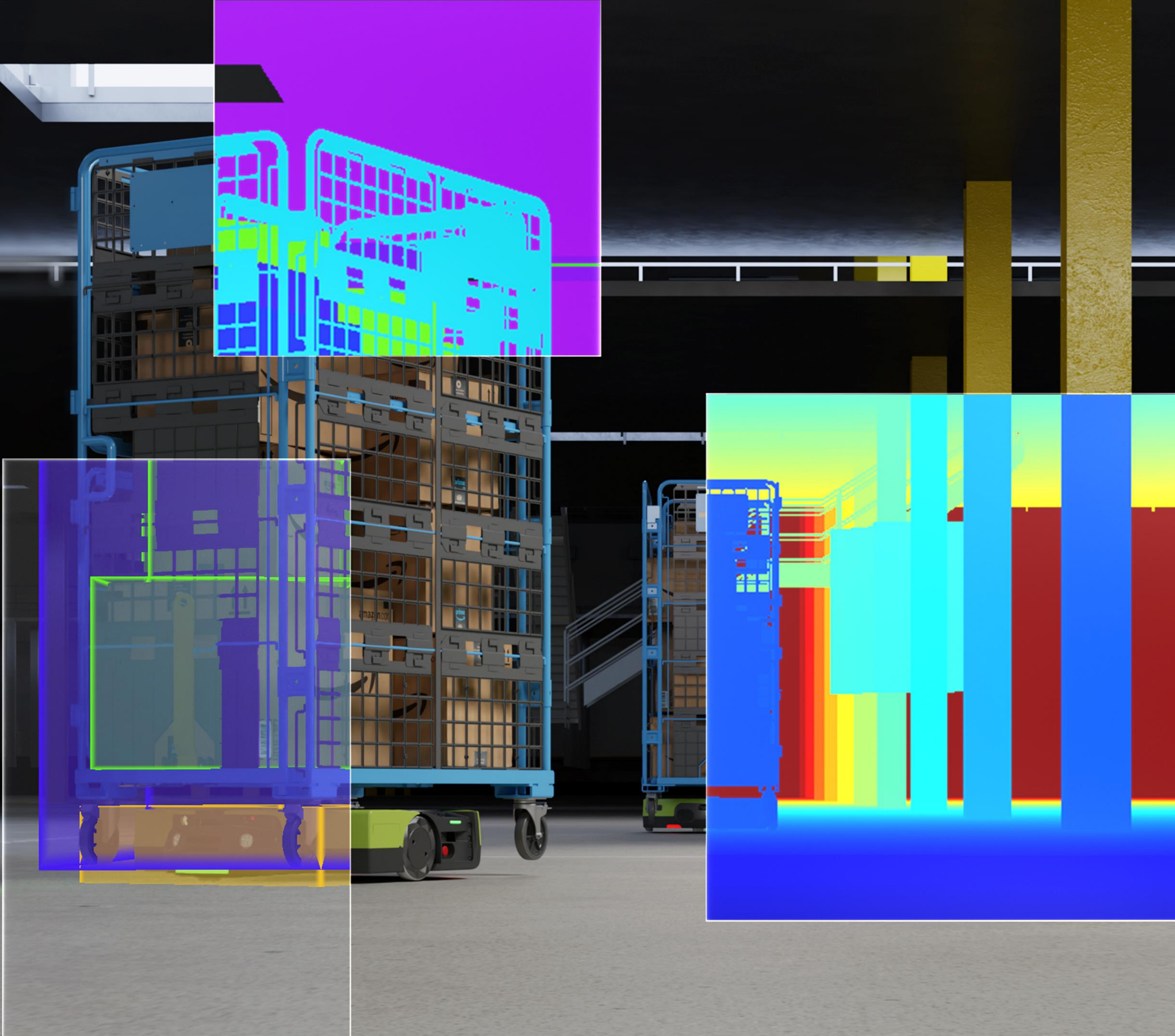
**BMW Blends
Reality and Virtual
Worlds to Build
Factory of the Future**

BMW Group is using Omniverse to build a
fully functioning factory digital twin before
building it in the real world. Using NVIDIA AI
and Omniverse has saved 20% on its factory
fleet orchestration and planning.

BMW Group

NVIDIA Robotics Across Industries

NVIDIA-powered robots are everywhere, from manufacturing and agriculture to security and home-based healthcare.
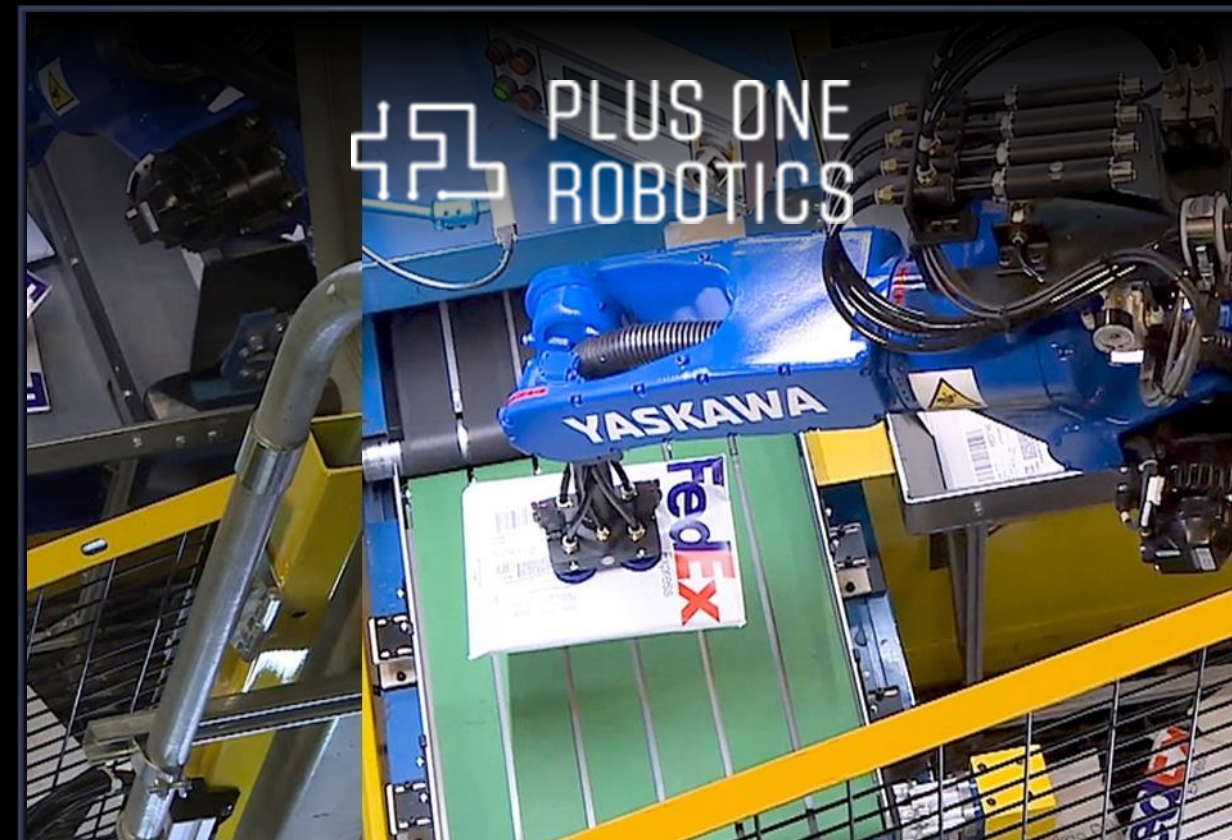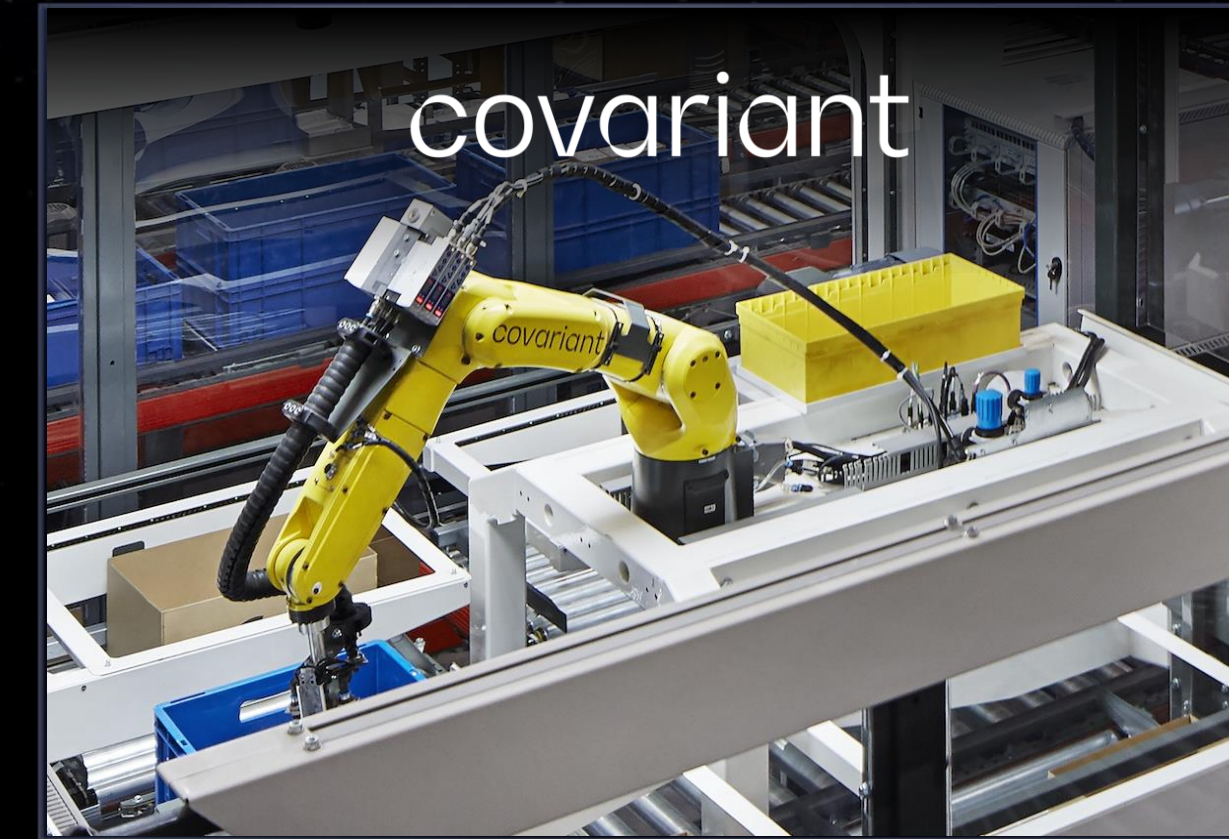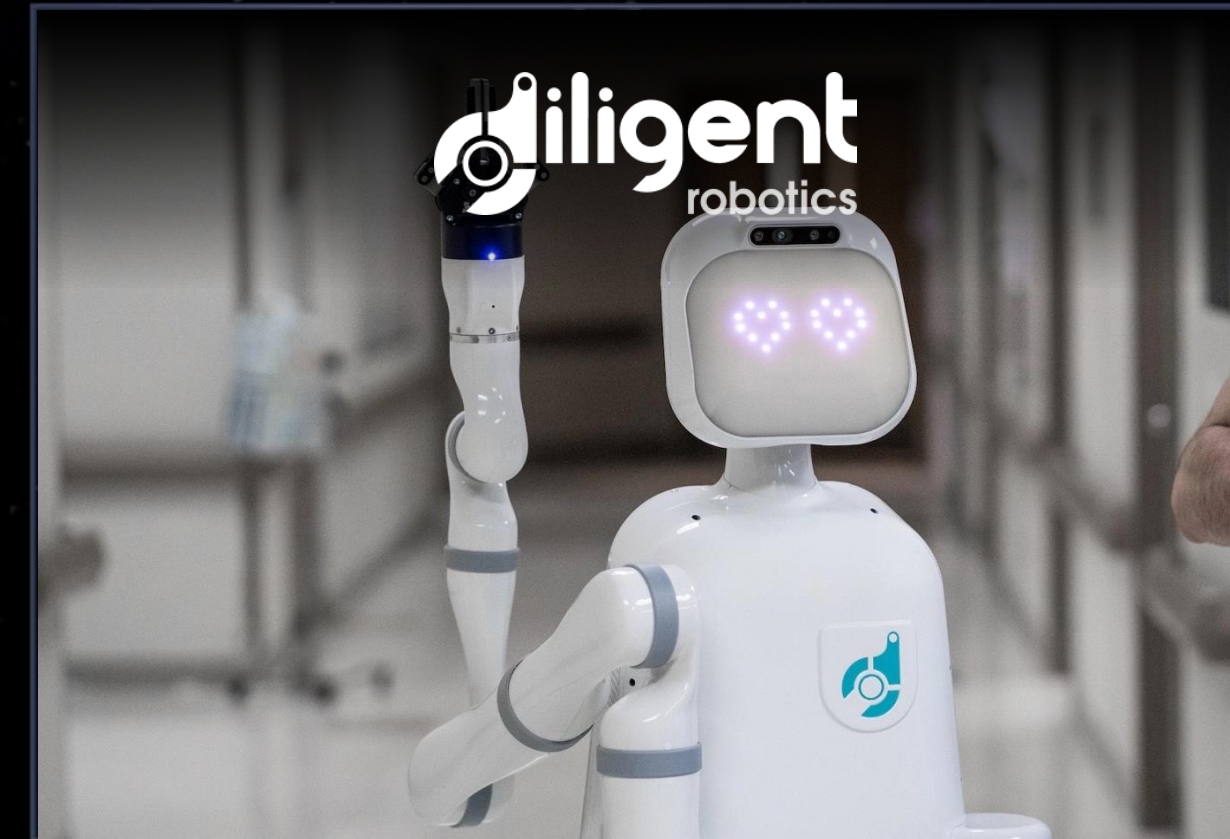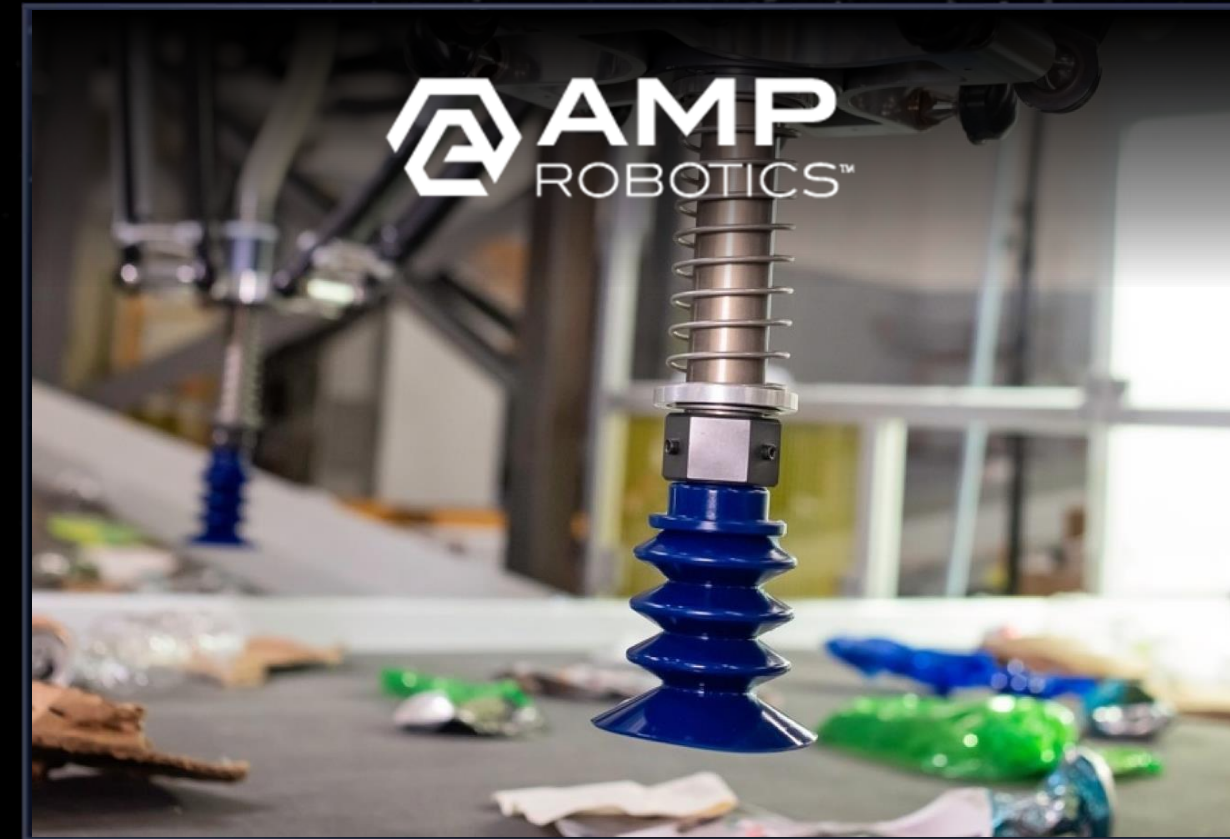
NVIDIA's Earth-2 Initiative Aims to Accelerate Climate Research

Our Earth-2 initiative will be a digital twin of Earth. This simulation of the world will help predict the complex multi-physics of Earth's atmosphere, land, sea, and ice caps at sufficiently high resolution. This will enable us to better predict the regional impacts of human actions over decades.

Earth-2

NVIDIA Is a Learning Machine

NVIDIA is united by a unique culture—the operating system of our company. We dream big, take risks, and learn from our mistakes together. Speed is the key to our success. Craftsmanship is a passion. There are no org charts—the mission is the boss.

These beliefs inform everything we do, from designing amazing products to building one of the world's great companies—a place where people can do their life's work.

We're One Team
Tackling Challenges
No One Else Can Solve

NVIDIA employees are dedicated to building technology that moves humanity forward and to supporting the communities in which they work and live.

We've been recognized as a top company in social responsibility, and our employees are passionate donors to hundreds of charities around the globe.

"Best Places to Work in 2023"

Glassdoor

"Most Innovative Companies"

Fast Company

"World's Best CEOS"

Barron's

"100 Best Companies to Work For"

Fortune

"World's Best Performing CEO"

Harvard Business Review

"50 Smartest Companies"

MIT Tech Review

Founded in 1993   Jensen Huang Founder & CEO   27,000 Employees $27B in FY23

"Nothing makes me prouder than the incredible people who have made NVIDIA the company it is today. We want our company to be where they can do their life's work.

"Together, we continue to drive advances in AI, HPC, gaming, creative design, autonomous vehicles, and robotics—some of the world's most impactful areas.

"I want to thank NVIDIA developers, partners, customers, and families for the amazing work you do. Exciting new frontiers lie ahead. Let's seek them out together."

Jensen Huang