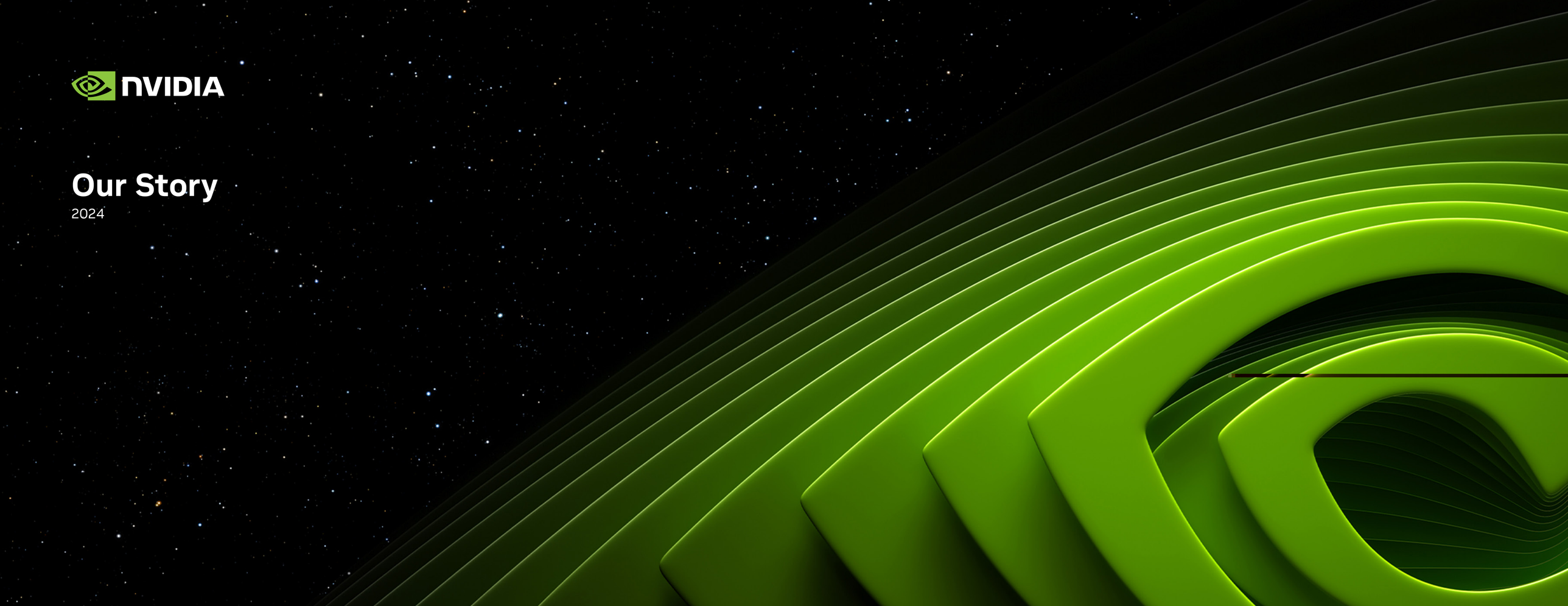
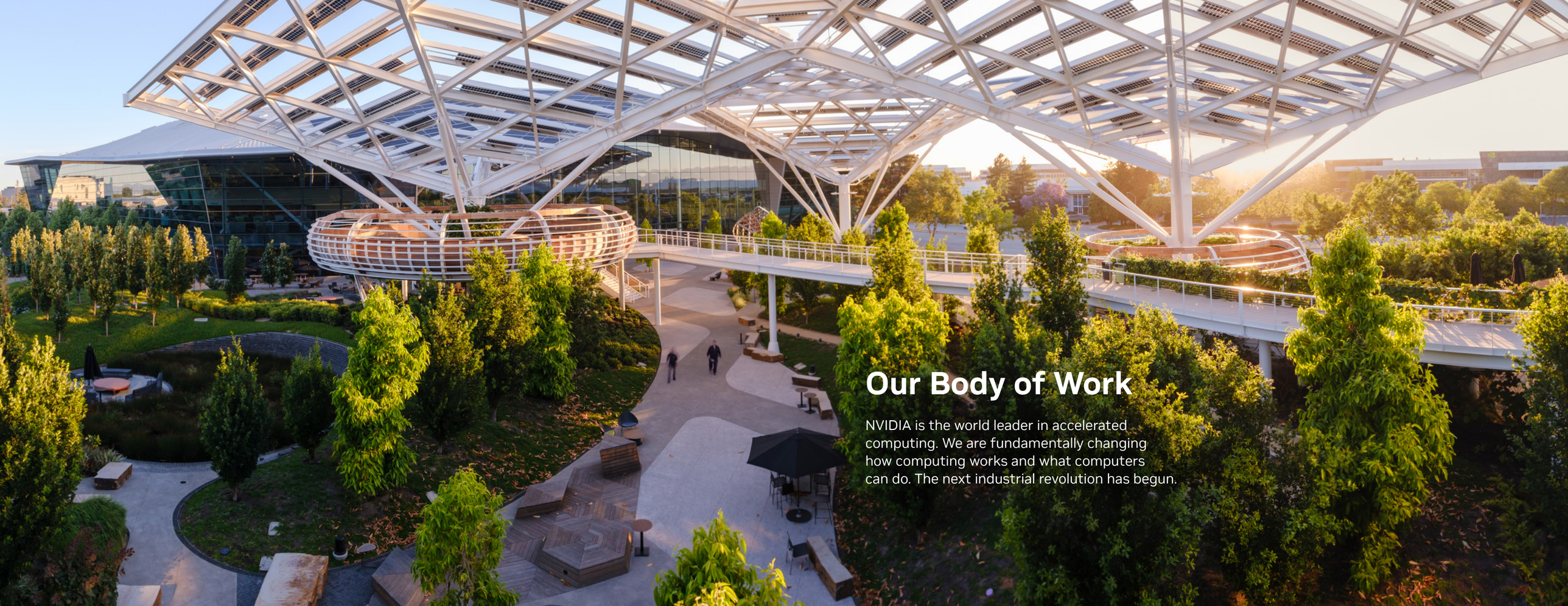

NVIDIA
Our Story
2024

NVIDIA is the result of the decades-long pursuit of a vision and the life's work of our employees. The company is well known for the caliber of our people, who have worked together for a long time, and our determined and resilient culture that prioritizes our people.

When we began this endeavor, no one knew how far deep learning would take us or how far it would scale. We are now beginning to realize the immense, transformative potential before us. NVIDIA is at the center of the simultaneous computing and industrial revolution.

I want to thank the many researchers and scientists, developers, ecosystem partners, customers, and employees who dedicated themselves to exploring what was a zero-billion-dollar market with us. We've reinvented NVIDIA, the computer industry, and very likely, the world.

Jensen Huang



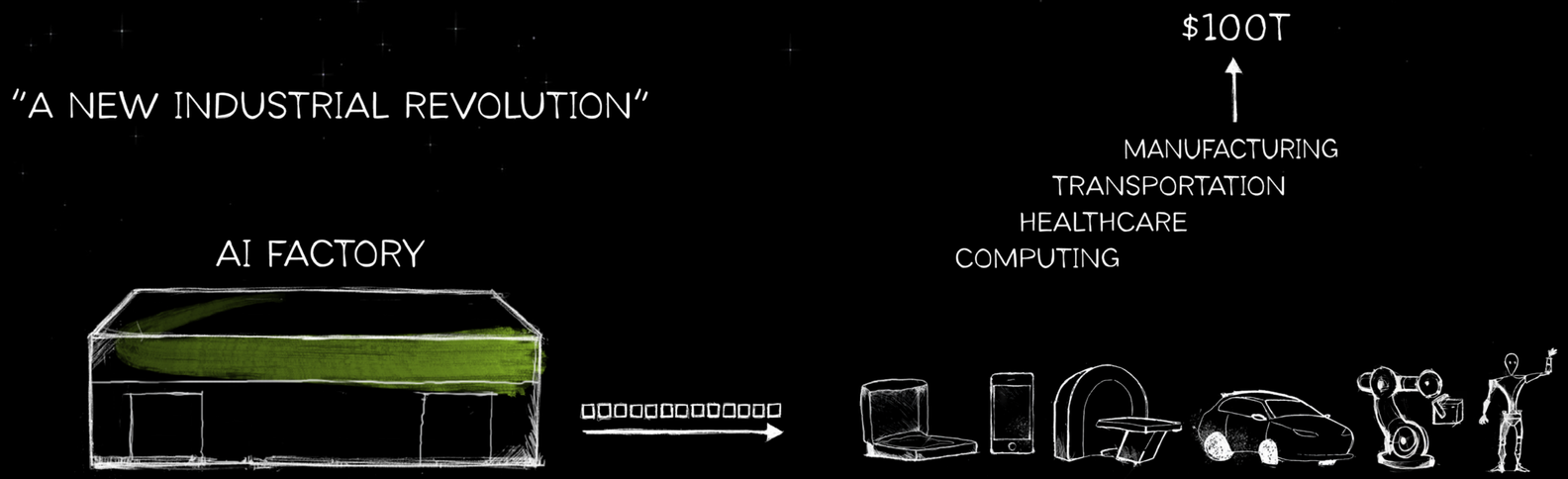
Our Body of Work

NVIDIA is the world leader in accelerated computing. We are fundamentally changing how computing works and what computers can do. The next industrial revolution has begun.

Pioneering Accelerated Computing

Modern computing started sixty years ago with the IBM System/360. For the last decade—as the performance scaling of general-purpose processing has slowed while computing demand has continued—an exponentially growing performance-to-demand deficit has built up. NVIDIA accelerated computing has created a path forward at just the right time.

Accelerated computing starts with the most advanced processors and ends with AI factories. From chip architecture to advanced networking to acceleration libraries, NVIDIA builds the entire computing system at data center scale to produce intelligence at scale.



A New Computing Platform Sparks a New Industrial Revolution

Accelerated computing paved the way for generative AI. For the very first time, we can manufacture our most valuable resource—intelligence.

To seize this generational opportunity, companies and countries around the world are building NVIDIA-powered AI factories to process, refine, and manufacture intelligence from data. Intelligence, in the form of digital tokens, is a new kind of commodity, a new kind of product, and it will create new revenue opportunities for the world's $100 trillion of industries.

NVIDIA CUDA's Virtuous Cycle Propels AI Innovation

NVIDIA accelerated computing has reached a tipping point and achieved a virtuous cycle. The significant CUDA® installed base attracts developers and applications, which attracts resellers reaching customers, which expands the installed base to attract more developers.

CUDA, our parallel computing model launched in 2006, offers developers an unparalleled toolkit with over 300 libraries, 600 AI models, numerous SDKs, and support for 3,700 GPU-accelerated applications. It has more than 53 million downloads.

The success of the CUDA model has led to the creation of a thriving ecosystem that now includes over 5 million developers, 40,000 companies, and thousands of generative AI companies—all building on the NVIDIA platform.

Accelerated Computing Is Sustainable Computing

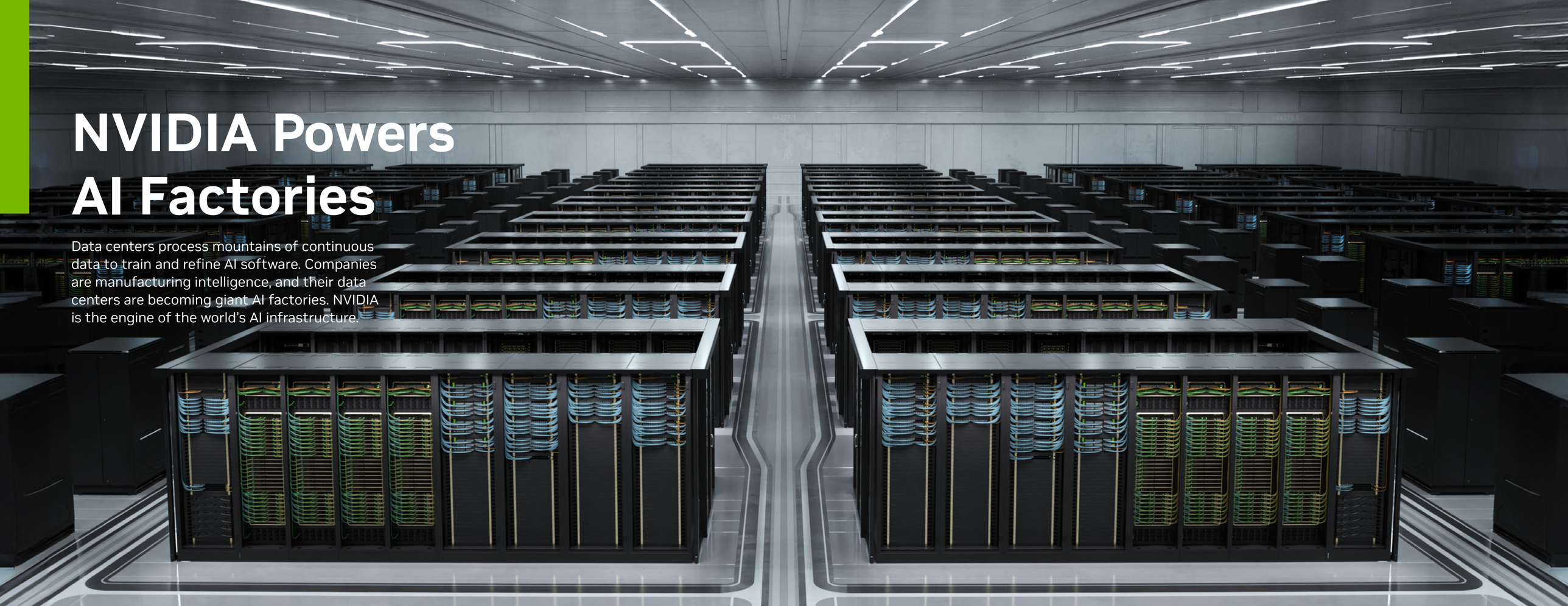
NVIDIA is modernizing the world's trillion-dollar data center industry. By accelerating workloads with NVIDIA GPUs operating in parallel, we exponentially increase throughput while driving down the total energy used to complete a task and, ultimately, the total cost of ownership. The energy savings are incredible.

That is precisely why, as CPU scaling has slowed, we must transition to accelerated computing. We cannot continue scaling out the traditional way. Accelerated computing is essential, and every application that can be accelerated should be.

Earth-2

NVIDIA Powers AI Factories

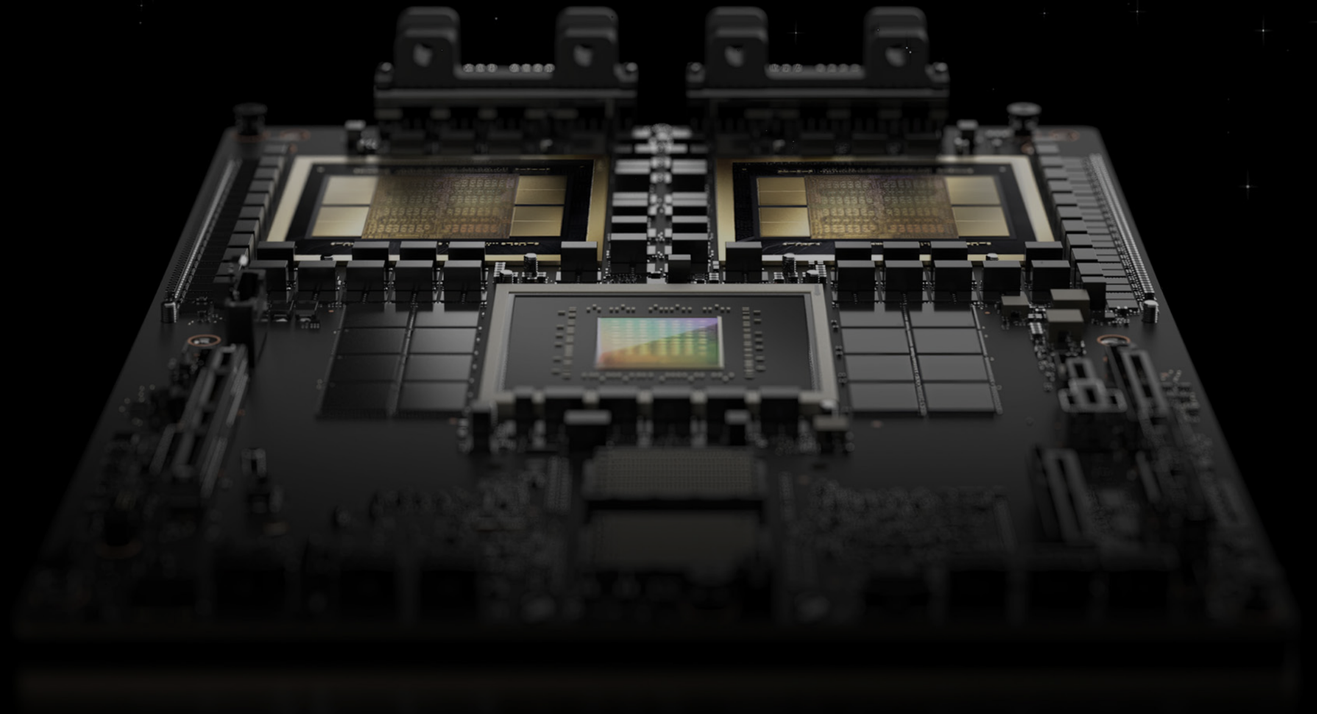
Data centers process mountains of continuous data to train and refine AI software. Companies are manufacturing intelligence, and their data centers are becoming giant AI factories. NVIDIA is the engine of the world's AI infrastructure.



Tailor-Made for the Generative AI Revolution

Blackwell is one of the most important products in our history, boasting technologies that power AI training and real-time LLM inference for models scaling up to 10 trillion parameters.

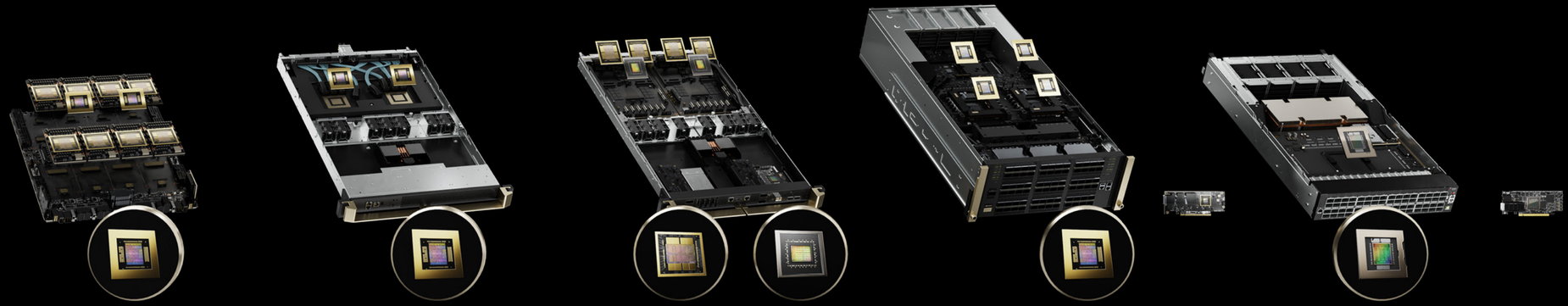
The second-generation transformer engine enables Blackwell to support double the compute and model sizes. Fifth-generation NVLink™ delivers groundbreaking 1.8TB/s bidirectional throughput per GPU for seamless high-speed communication among up to 576 GPUs for handling trillion-parameter LLMs.



The Platform for a New Era of Computing

In the future, almost all of our experiences will be generative. Blackwell enables organizations everywhere to build and run real-time generative AI on trillion-parameter large language models at up to 25X less cost and energy consumption than its predecessor.

AWS, Google Cloud, Microsoft Azure, and Oracle Cloud Infrastructure will be among the first cloud service providers to offer Blackwell-powered instances, as will NVIDIA Cloud Partner program companies Applied Digital, CoreWeave, Crusoe, IBM Cloud, and Lambda.



HGX™ B100

NVLink Switch

GB200 Superchip
Compute Node

Quantum X800 Switch
ConnectX-8 SuperNIC

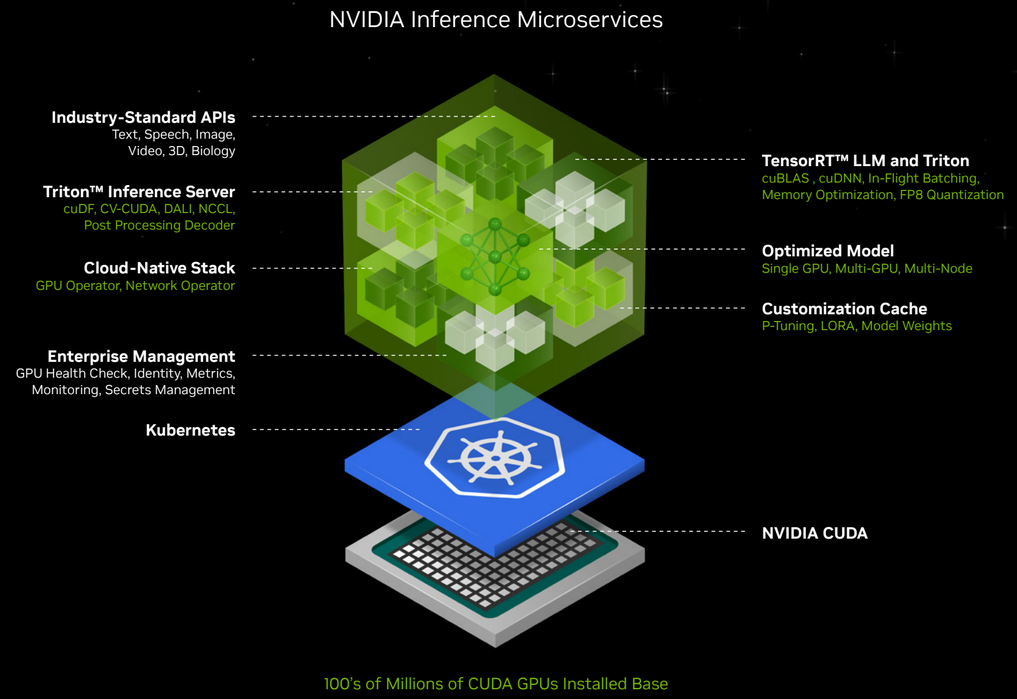
Spectrum X800 Switch
BlueField®-3 SuperNIC

NIM Brings AI to Every Enterprise

NVIDIA NIM™ inference microservices are a new way to distribute AI models.

NIM delivers enterprise-grade optimized generative AI that runs on CUDA everywhere, from the cloud to on-prem data centers to NVIDIA RTX™ AI PCs, through our expansive network of ecosystem partners.

These containerized, pretrained AI-in-a-box microservices integrate CUDA libraries, optimized inference engines, industry-standard APIs, and AI model support into containers for straightforward deployment.



NVIDIA Reinvents Modern Graphics

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

With NVIDIA RTX, we reinvented computer graphics again. We developed a pioneering, real-time rendering system that accurately simulates physical properties. This allows us to create virtual environments for training and testing AI perception models, leading the way for the AI-powered assistants of the future.

NVIDIA RTX Resets Gaming

RTX is everywhere. More than 500 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 30 million members in more than 110 countries now have access to more than 1,900 games.

NVIDIA Brings Generative AI to Millions With Tools for RTX PCs and Workstations

Running generative AI locally on a PC is critical for privacy, latency, and cost-sensitive applications. With over 100 million RTX AI PCs and workstations, NVIDIA has a massive installed base, and our developer tools provide the horsepower to tune and optimize AI models for the PC platform. NVIDIA's full-stack RTX AI innovations accelerate over 500 PC applications and games and 200 laptop designs from manufacturers.

ASUS TUF A14 / A16

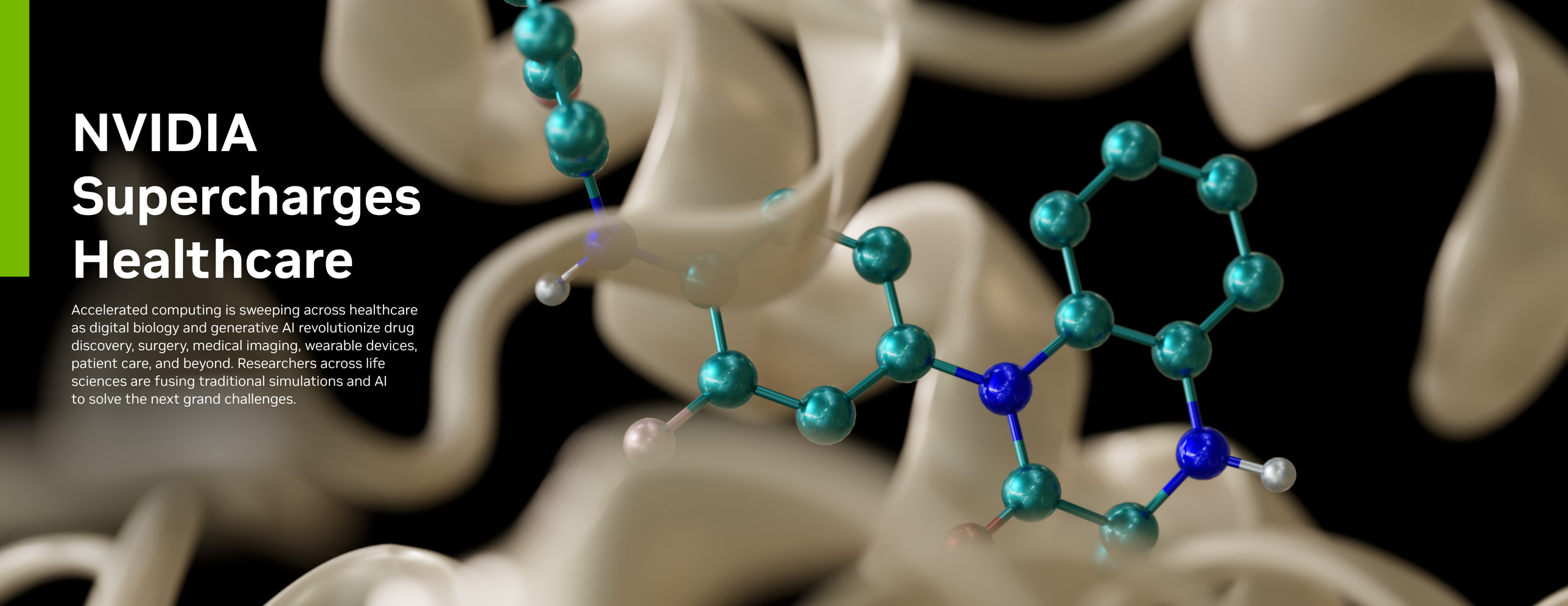
ASUS Zephyrus G16

ASUS ProArt PX13 / P16

MSI Stealth A16 AI+

NVIDIA Supercharges Healthcare

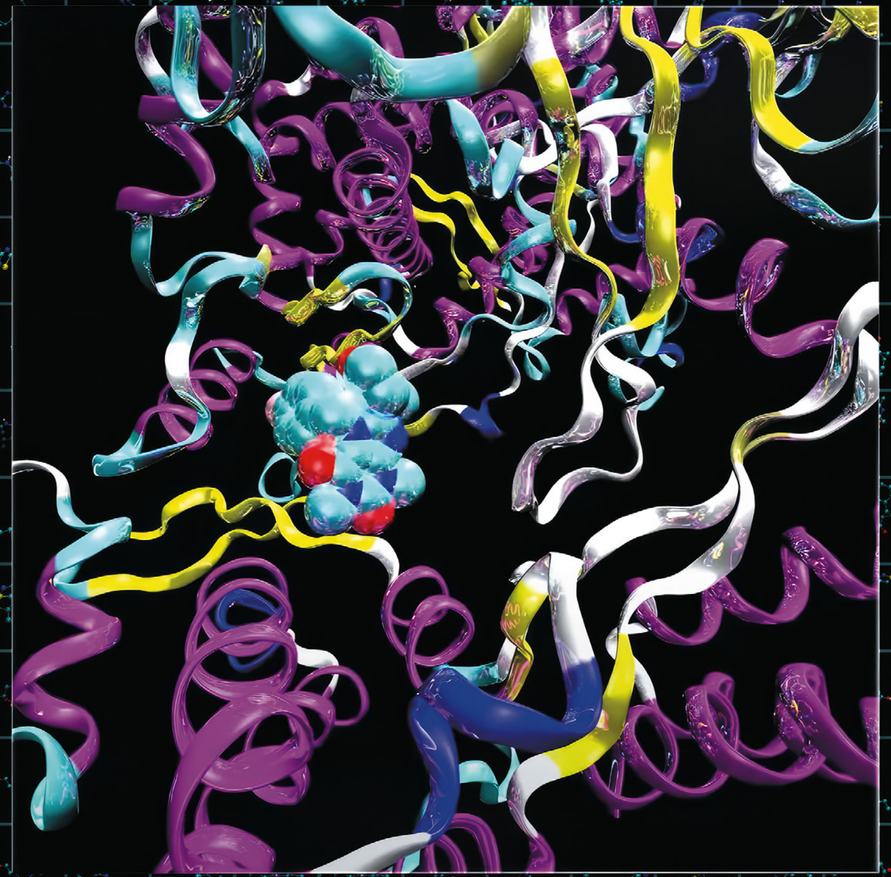
Accelerated computing is sweeping across healthcare as digital biology and generative AI revolutionize drug discovery, surgery, medical imaging, wearable devices, patient care, and beyond. Researchers across life sciences are fusing traditional simulations and AI to solve the next grand challenges.



Generative AI Is Transforming Life Sciences and the Pharmaceutical Industry

Researchers use generative AI to compose billions of chemical compounds and virtually test them as drugs against a disease target. NVIDIA AI is powering the next era of drug discovery and advances in life sciences.

NVIDIA Clara™, our suite of computing platforms, software, and services for healthcare and life sciences, and NVIDIA BioNeMo™, our platform for state-of-the-art generative AI models for drug discovery, are driving remarkable breakthroughs.



NVIDIA Leads Industrial Digitalization

Heavy industries are racing to become software-defined. NVIDIA Omniverse™ is the fundamental operating system for building digital twins crucial to unlocking new potential in heavy industries worldwide. With Omniverse Cloud APIs, developers can simplify and speed up the development of digital twins for almost any industrial application, seamlessly integrating Omniverse into their existing apps.

Siemens is integrating NVIDIA AI and Omniverse technologies into TeamCenter X, their leading project lifecycle product software, to develop and deliver products at scale. Omniverse APIs enable data interoperability and space rendering for industrial scale design and manufacturing projects.

Siemens Teamcenter X with NVIDIA Omniverse Cloud APIs

Foxconn’s Robotic Factory Ecosystem Runs on NVIDIA

Foxconn, one of the world’s largest makers of electronics, uses Omniverse to build their robotic factories. This lets them orchestrate robots running on NVIDIA Isaac™ to build NVIDIA AI supercomputers, which in turn train Foxconn’s robots.

Omniverse Digital Twin

Real Factory

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 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

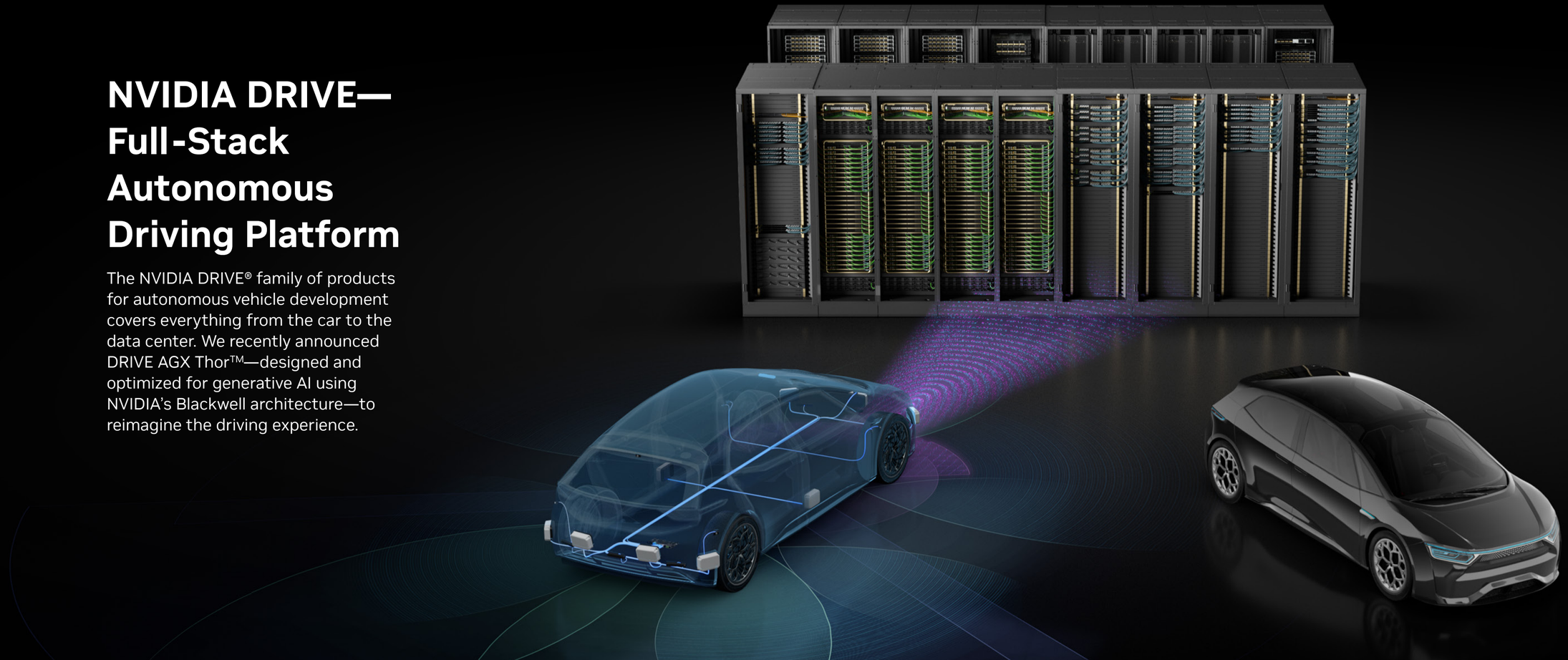
BMW Blends Real and Virtual Worlds to Build AI-Enabled Factories

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

BMW Group

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. We recently announced DRIVE AGX Thor™—designed and optimized for generative AI using NVIDIA's Blackwell architecture—to reimagine the driving experience.

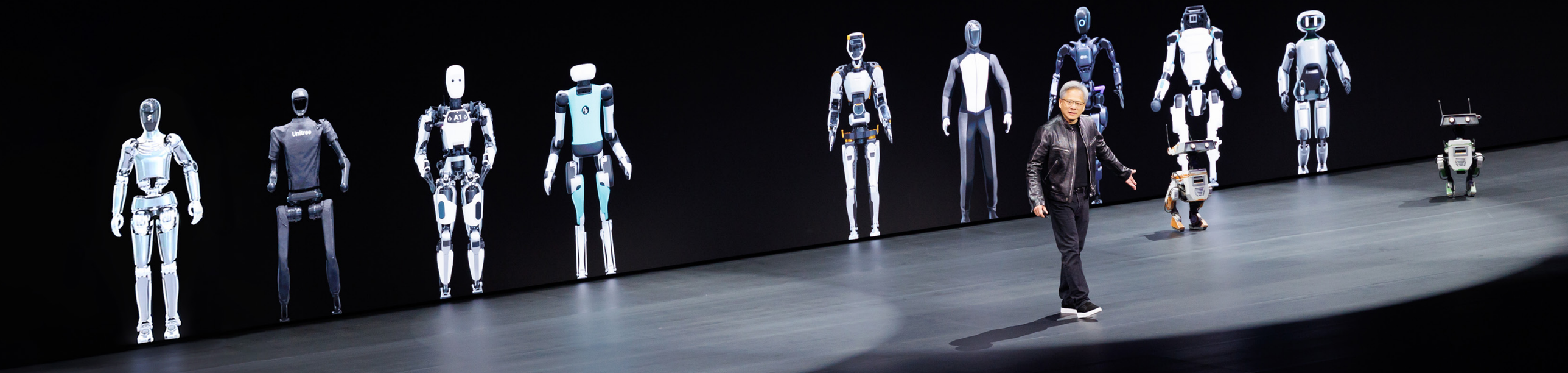


NVIDIA Omniverse Turbocharges Self-Driving Car Development

NVIDIA Omniverse Cloud APIs deliver large-scale, high-fidelity sensor simulation, paving the path to autonomous driving. By bringing together a rich ecosystem of simulation tools, applications, and sensors, these APIs let developers safely explore the wide variety of real-world scenarios autonomous systems will encounter. This enables vehicles to drive millions of miles in a wide range of simulated scenarios, so they hit the road running safely.

The ChatGPT Moment for Robotics Is Coming

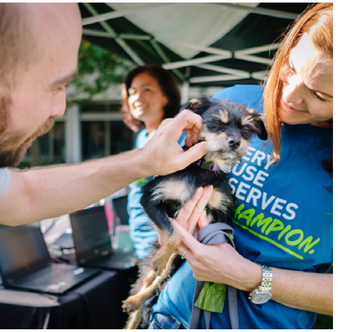
NVIDIA is fueling the next wave of AI—robotics and industrial digitalization. That new generation of robots that will learn in NVIDIA Omniverse. Over 1.2 million developers and 10,000 customers and partners are tapping into the NVIDIA Isaac and Jetson™ platforms to develop and deploy AI-driven robots. And Project GR00T, a general-purpose foundation model for humanoid robots, will help them understand natural language and emulate movements by observing human actions.



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 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 2024”
Glassdoor
“100 Best Companies to Work For”
Fortune
“Most Innovative Companies”
Fast Company
“Best CEO of 2023”
Economist
“Best-Led Companies”
Fortune
“Most Influential Company”
TIME

Photo by Jason O'Rear