



NVIDIA



# Our Body of Work

NVIDIA pioneered accelerated computing to tackle challenges no one else can solve. Our work in AI and the metaverse is profoundly impacting society and transforming the world's largest industries—from gaming to robotics, self-driving cars to life-saving healthcare, climate change to virtual worlds where we can all connect and create.



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.



# Sparkling the iPhone Moment of AI

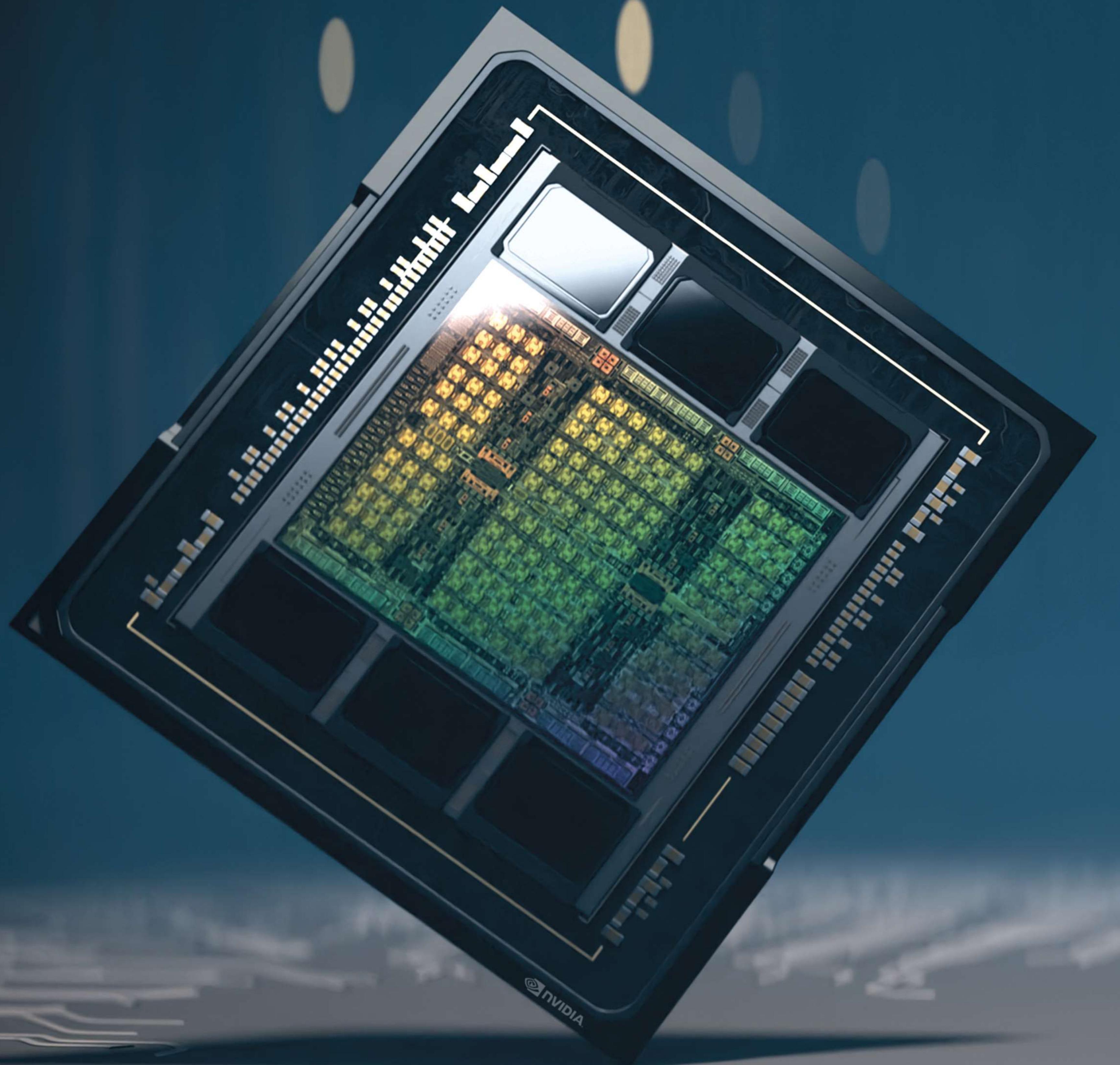
The acceleration of deep learning ignited the big bang of AI. ChatGPT, a large language model powered by an NVIDIA DGX™ AI supercomputer, reached 100 million users in just two months. Its magical capabilities have captured the world's imagination. Generative AI is a new computing platform, like the PC,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

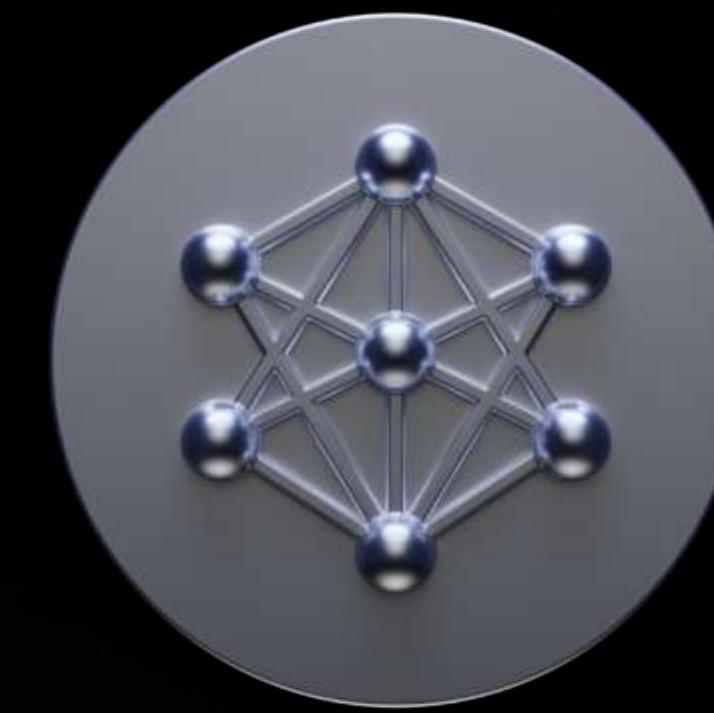
NVIDIA now offers some 300 acceleration libraries.



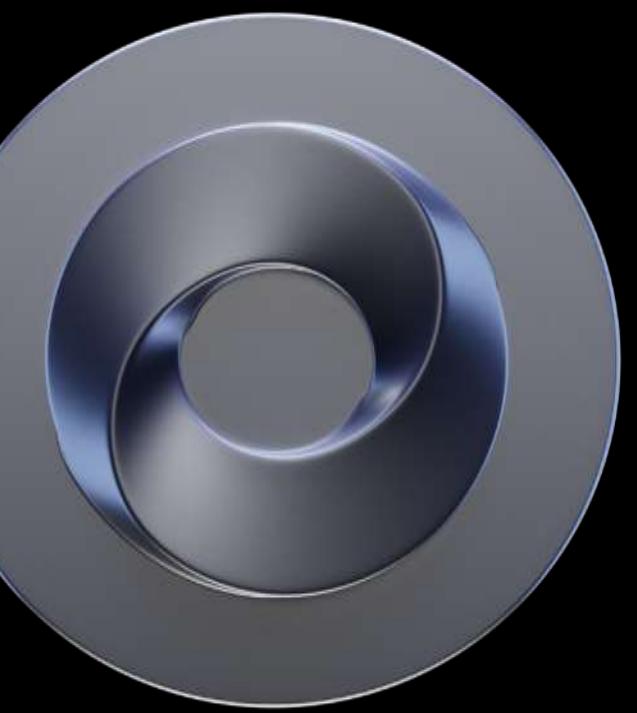
# Leaping Into the Cloud

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.

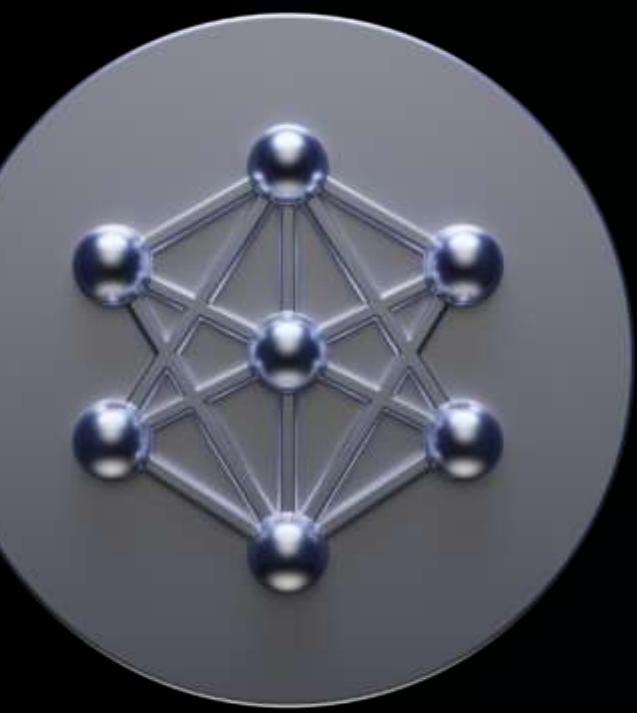
NVIDIA  
AI Enterprise



NVIDIA  
Omniverse



NVIDIA  
AI Foundations



Google Cloud

Microsoft  
Azure

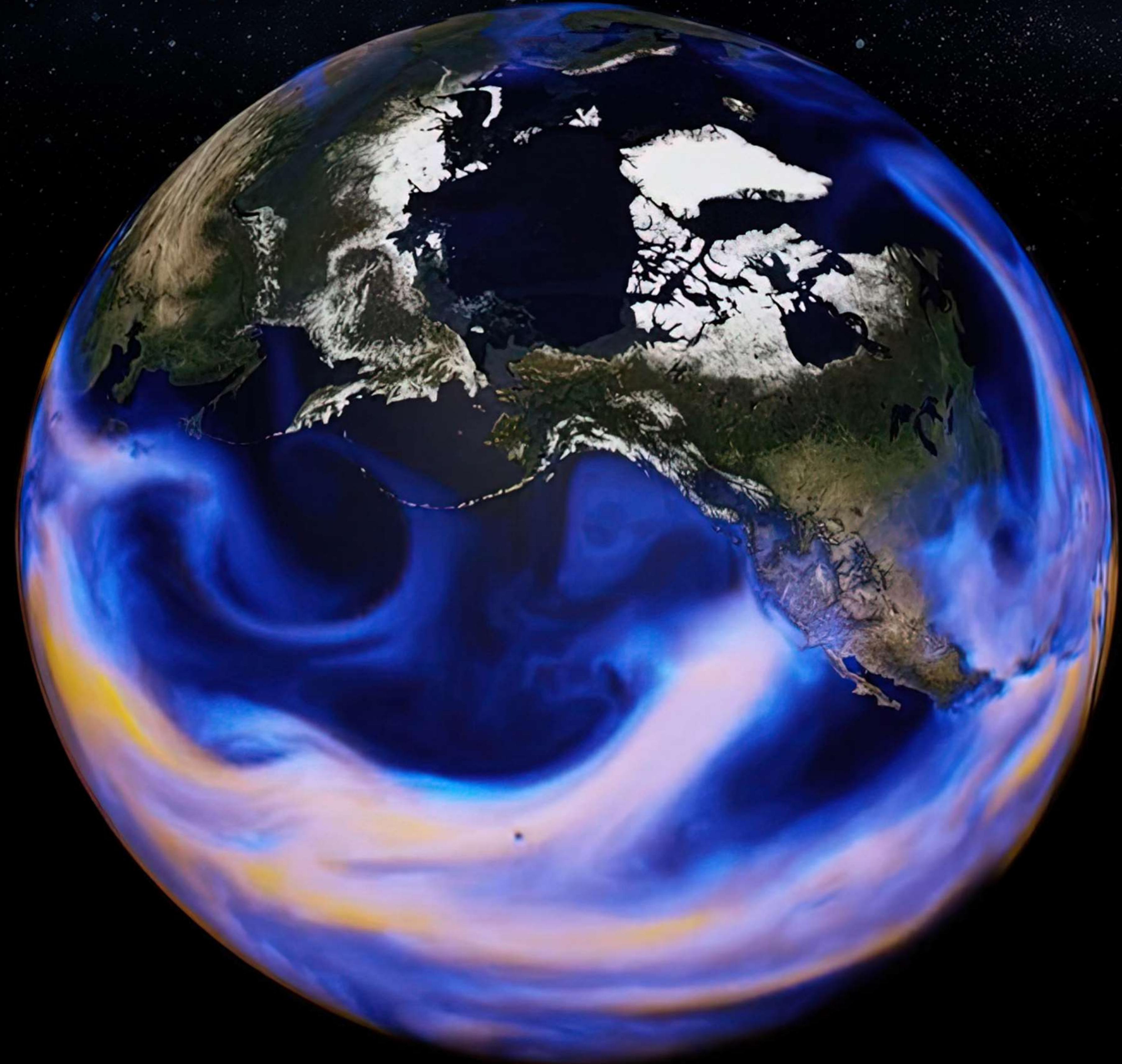
ORACLE  
CLOUD  
Infrastructure

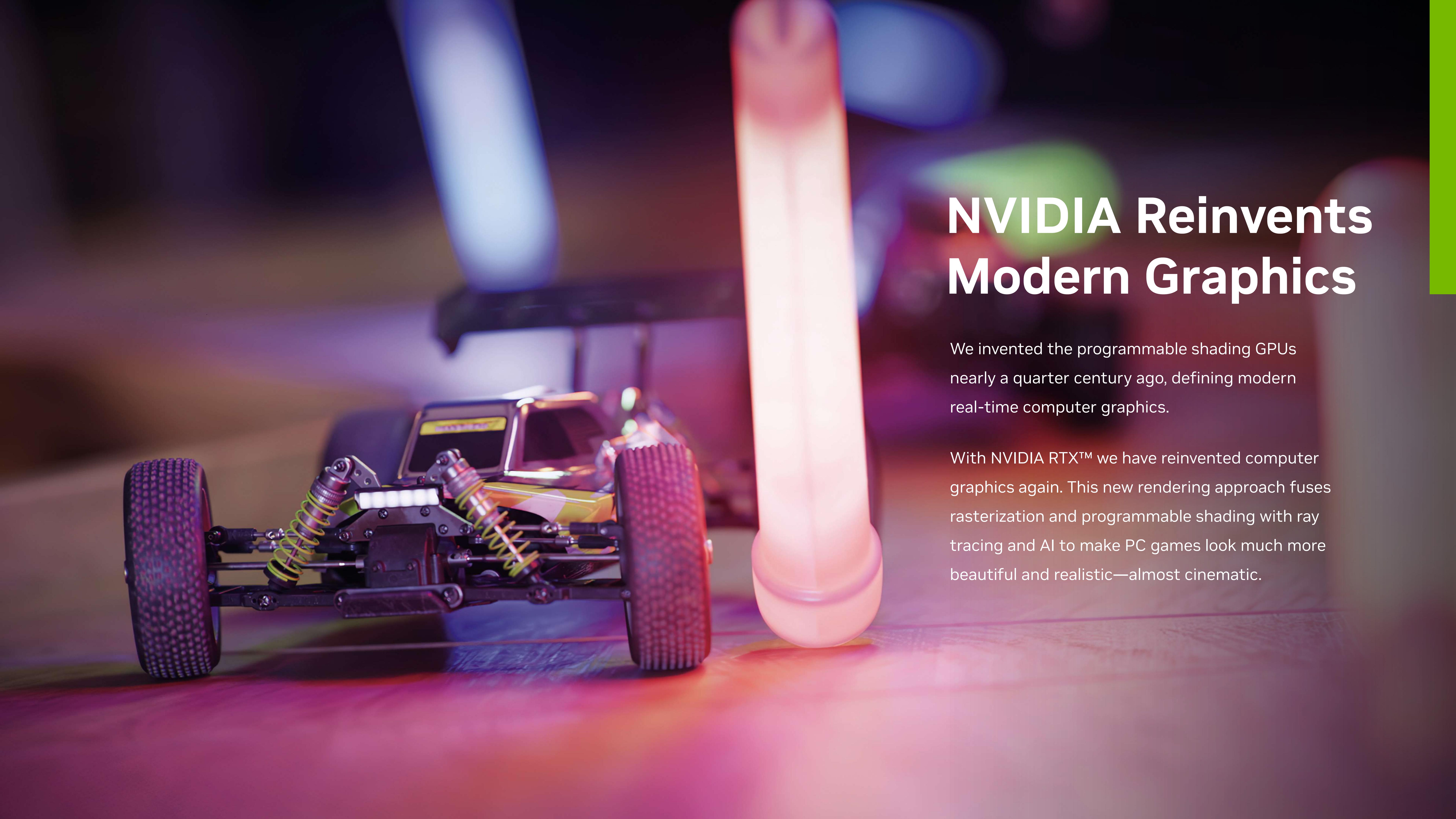
# NVIDIA Accelerated Computing Is Sustainable Computing

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.





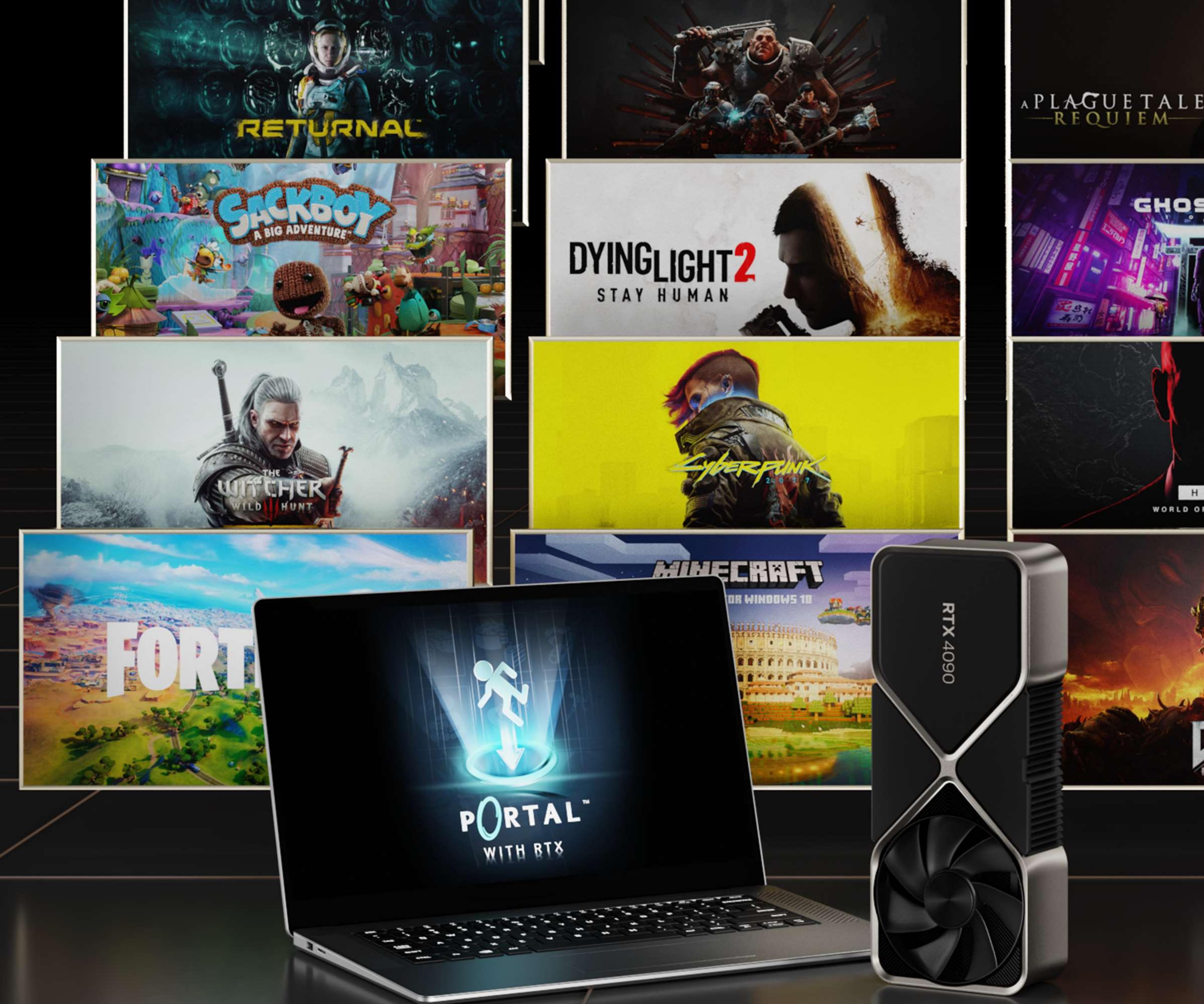
# NVIDIA Reinvents Modern Graphics

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.

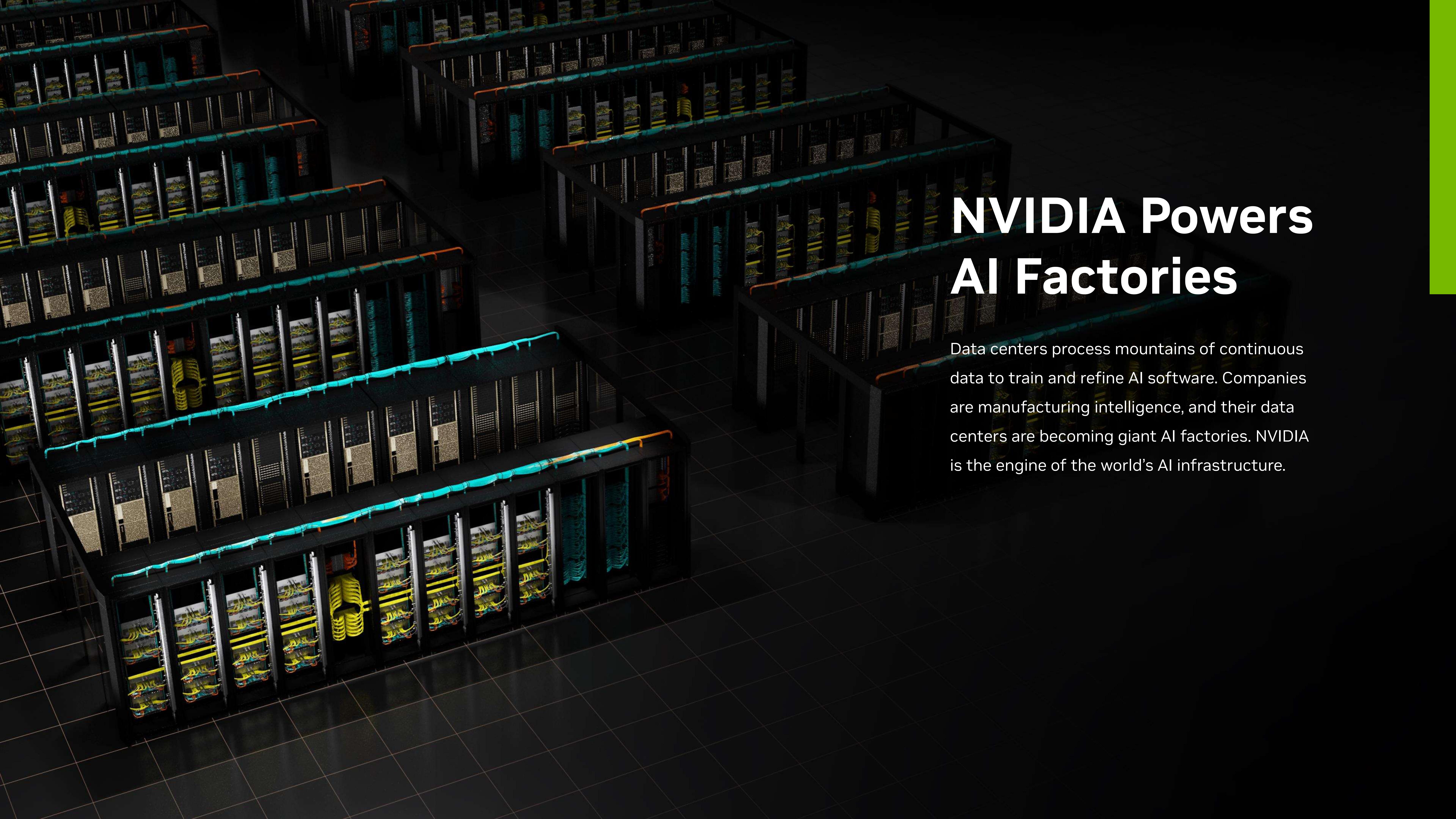
And, recently, NVIDIA and Microsoft signed a 10-year deal to bring the Xbox PC game library to GeForce NOW.



# 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.

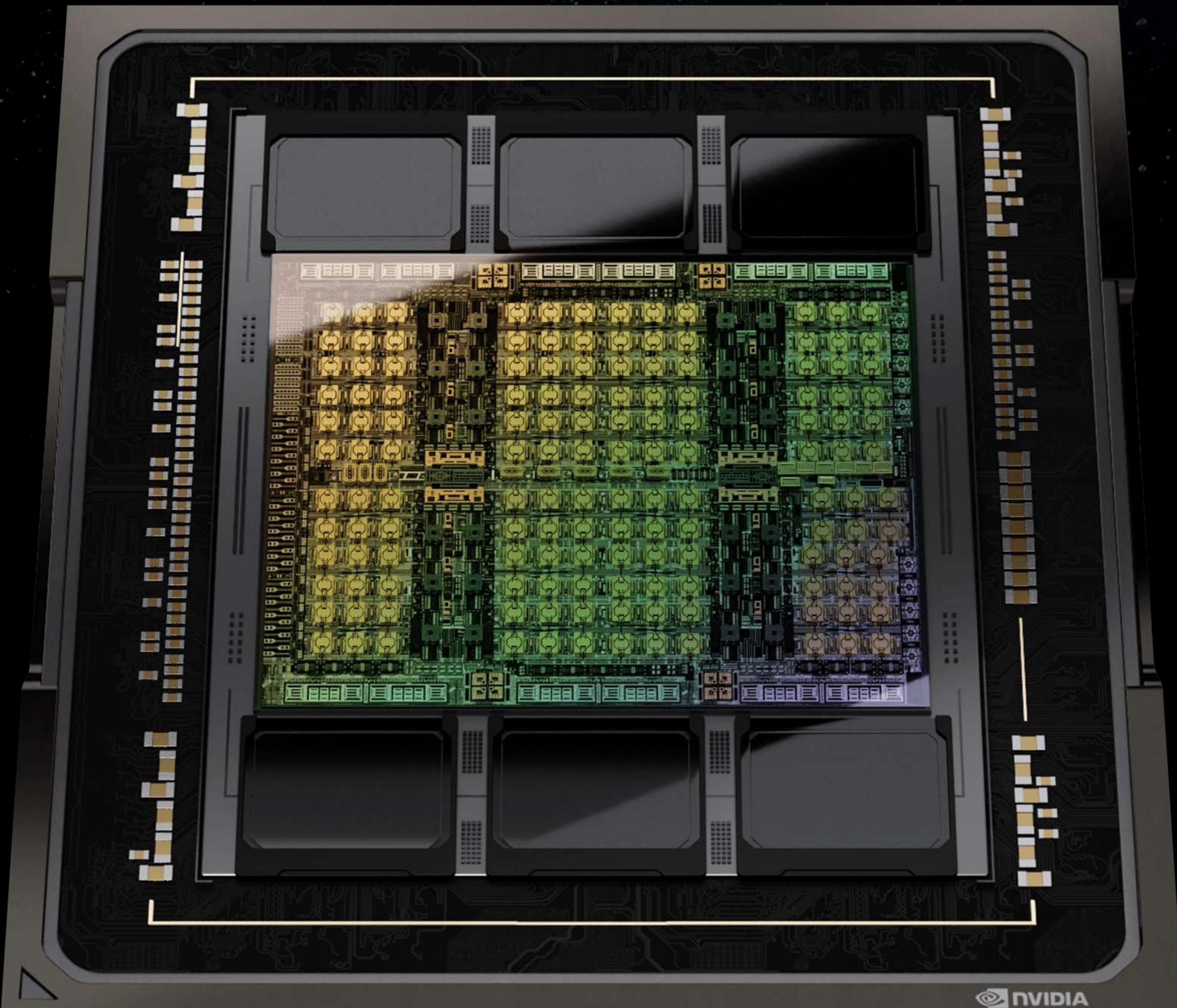




# NVIDIA Powers AI Factories

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.

