NVIDIA Non-Deal Roadshow

October 2025

Certain statements in this presentation including, but not limited to, statements as to: expectations with respect to growth, performance and benefits of NVIDIA's products, services and technologies, and related trends and drivers; expectations with respect to supply and demand for NVIDIA's products, services and technologies, and related matters; expectations with respect to NVIDIA's third party arrangements, including with its collaborators and partners such as OpenAI, Intel and CoreWeave; expectations with respect to NVIDIA's investments; expectations with respect to energy efficiency; expectations with respect to NVIDIA's strategies; expectations with respect to technology developments, and related trends and drivers; projected market growth and trends; expectations with respect to AI and related industries; and other statements that are not historical facts are forward-looking statements within the meaning of Section 27A of the Securities Act of 1933, as amended, and Section 21E of the Securities Exchange Act of 1934, as amended, which are subject to the "safe harbor" created by those sections based on management's beliefs and assumptions and on information currently available to management and are subject to risks and uncertainties that could cause results to be materially different than expectations.

Important factors that could cause actual results to differ materially include: global economic and political conditions; NVIDIA's reliance on third parties to manufacture, assemble, package and test NVIDIA's products; the impact of technological development and competition; development of new products and technologies or enhancements to NVIDIA's existing product and technologies; market acceptance of NVIDIA's products or NVIDIA's partners' products; design, manufacturing or software defects; changes in consumer preferences or demands; changes in industry standards and interfaces; unexpected loss of performance of NVIDIA's products or technologies when integrated into systems; and changes in applicable laws and regulations, as well as other factors detailed from time to time in the most recent reports NVIDIA files with the Securities and Exchange Commission, or SEC, including, but not limited to, its annual report on Form 10-K and quarterly reports on Form 10-Q. Copies of reports filed with the SEC are posted on the company's website and are available from NVIDIA without charge. These forward-looking statements are not guarantees of future performance and speak only as of the date hereof, and, except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances.

Many of the products and features described herein remain in various stages and will be offered on a when-and-if-available basis. The statements within are not intended to be, and should not be interpreted as a commitment, promise, or legal obligation, and the development, release, and timing of any features or functionalities described for our products is subject to change and remains at the sole discretion of NVIDIA. NVIDIA will have no liability for failure to deliver or delay in the delivery of any of the products, features or functions set forth herein.

NVIDIA’s Evolution From Chips to an AI Infrastructure Company

ACCELERATOR

1999
GPU

ACCELERATED COMPUTING

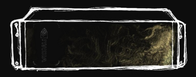
2006
CUDA

AI

2012
cuDNN

AI SUPERCOMPUTER

2016
DGX-1
NVL8



AI DATACENTER

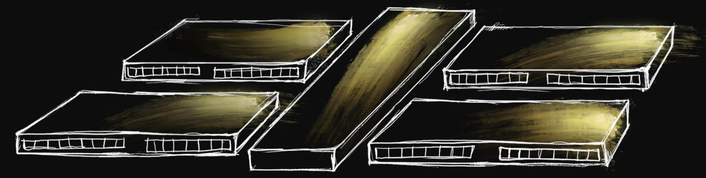
2019
MELLANOX

2020
SELENE

2024
NVL72 SCALE-UP
3-SCALING LAWS

AI INFRASTRUCTURE

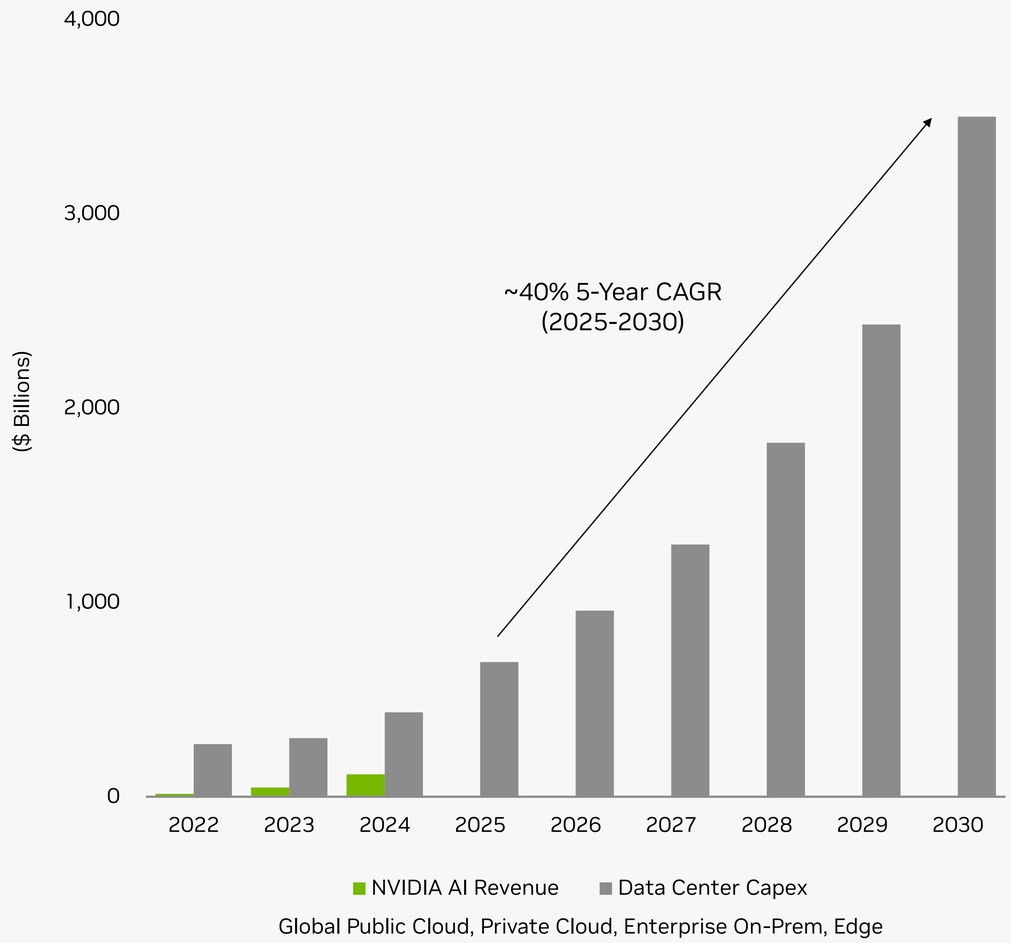
SPECTRUM-X
SCALE-OUT AND
SCALE-ACROSS



$3-4 Trillion AI Infrastructure Spend by 2030

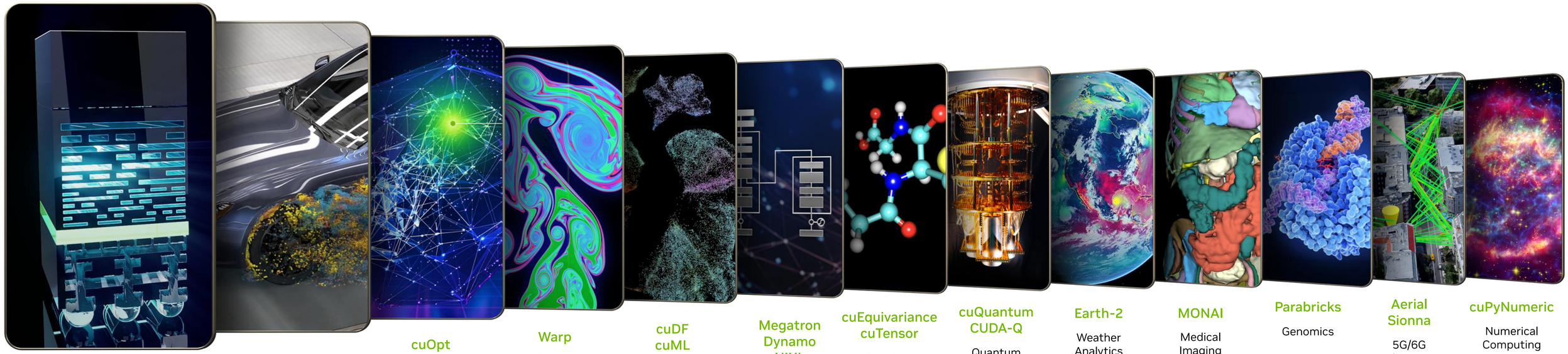
Key TAM Growth Drivers

- **End of Moore's Law drives fundamental shift from general-purpose to accelerated computing**
  - **For example, Data Processing**: NVIDIA cuDF, cuML, cuVS accelerates structured and unstructured data processing 10-100X over CPU-based methods; **Computational Lithography**: NVIDIA cuLitho accelerates computation lithography tasks, such as creating photomasks, by as much as 40-60X over CPU-based methods; **Genome Sequencing**: deciphEHR was able to achieve >5X faster alignment runtimes and >10X faster variant calling runtimes, using NVIDIA Parabricks
- **Hyperscale shift to Generative AI**
  - **For example, Ad Generation**: Google AI-powered video campaigns on YouTube deliver 17% higher ROAS (return on ad spend) than manual campaigns; **Recommender Systems**: Pinterest was able to move to 100X larger recommender models by adopting NVIDIA GPUs which increased engagement by 16%; **Search** and **User-Generated-Content** are moving to adopt LLM-powered generative AI
- **Model Makers — A new industry**
  - OpenAI, Google, Anthropic, xAI, Meta are building the foundations of AI
- **Enterprise — Agentic AI enters the labor market**
  - **Coding**: Developers using AI coding assistants could complete tasks up to 55% faster, according to an MIT study; **Vibe Coding**: Lovable opens coding to new users like product designers, creatives, marketing, IT, teachers; **Legal**: STARA, an AI designed for statutory research, conducted a task that would take two humans 8 to13.5 hours and cost ~$3,000 in 20 minutes at a cost of ~$0.86
- **Robotics, AV, Factories, and Edge powered by Physical AI**
  - Labor shortages drive everything that moves in $50T industrials to be autonomous



$3-4 Trillion AI Infrastructure Spend by 2030

NVIDIA CUDA-X Platforms Enable Shift From CPU to GPU Accelerated Computing



cuLitho
Computational Lithography

cuDSS cuSPARSE cuFFT AmgX
Computer Aided Engineering

cuOpt
Decision Optimization

Warp
Physics

cuDF cuML
Data Science and Processing

Megatron Dynamo NIXL cuDNN CUTLASS
Deep Learning

cuEquivariance cuTensor
Quantum Chemistry

cuQuantum CUDA-Q
Quantum Computing

Earth-2
Weather Analytics

MONAI
Medical Imaging

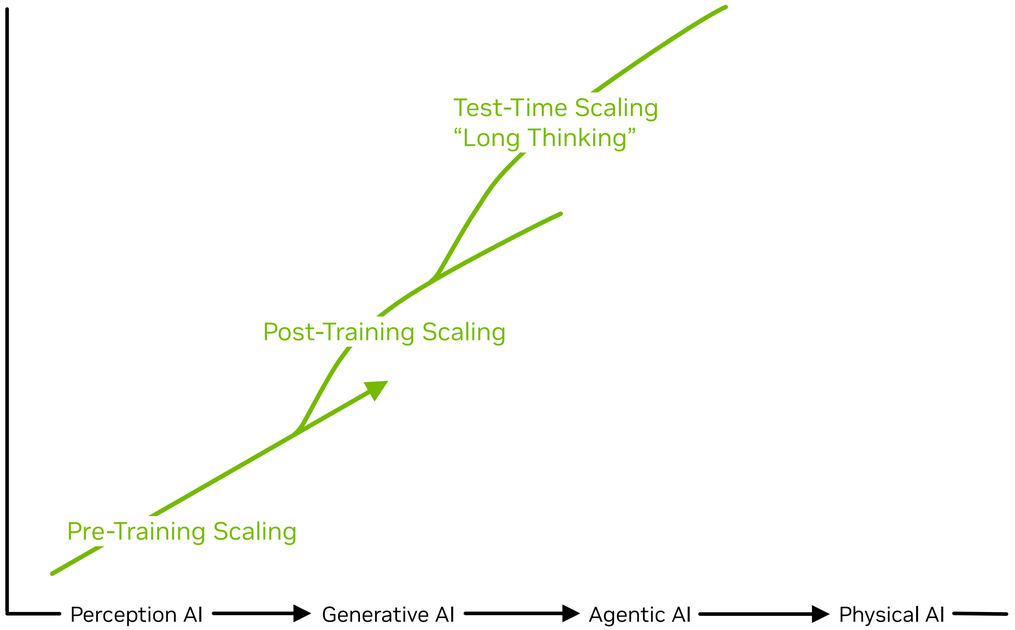
Parabricks
Genomics

Aerial Sionna
5G/6G Signal Processing

cuPyNumeric
Numerical Computing

$3-4 Trillion AI Infrastructure Spend by 2030

3 AI Scaling Laws Driving Exponential Computing Demand
NVIDIA Excels at Pre-Training, Post-Training, and Inference



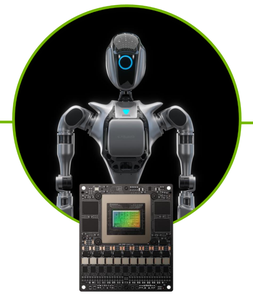
Token Generation is Doubling Every Two Months

- ChatGPT is at ~700M WAUs, with usage up ~4X y/y
- OpenAI now counts 5M paying business users, up from 3M in June
- Microsoft processed over 500 trillion tokens served by Foundry APIs in FY2025, up 7X y/y
- Alphabet processed over 980 trillion tokens in the month of June across its AI services, up from a monthly run-rate of 480 trillion tokens in May
- The Gemini app had more than 450M MAUs as of the end of July with daily requests up >50% in Q2 vs. Q1

$3-4 Trillion AI Infrastructure Spend by 2030

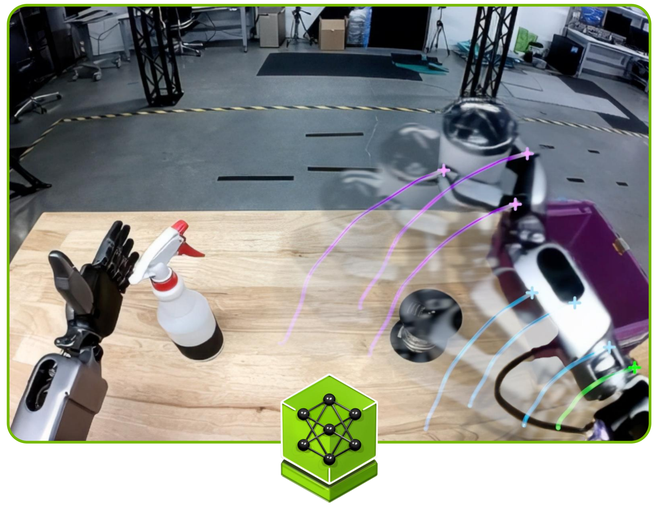
NVIDIA 3-Computers Enable Physical AI – the Next Wave of AI

Manipulating Robot



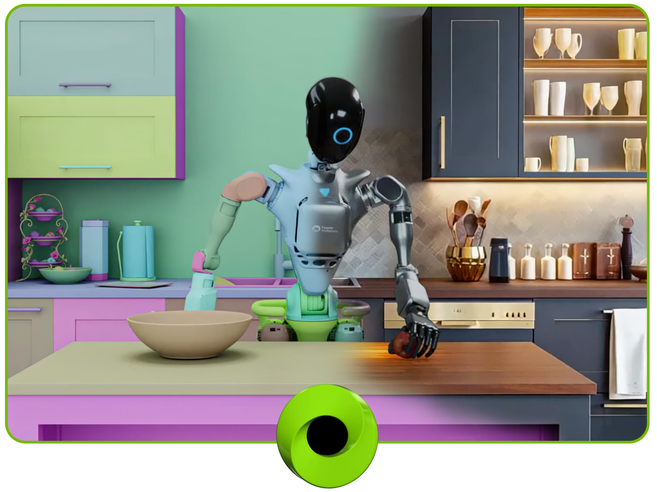
Jetson AGX Thor

Training Robot AI



GB200

Practicing in Digital Twin

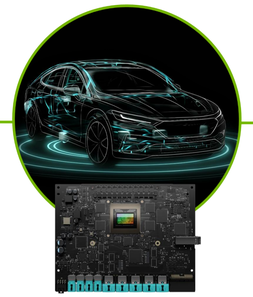


RTX PRO Server

$3-4 Trillion AI Infrastructure Spend by 2030

NVIDIA 3-Computers Enable Physical AI – the Next Wave of AI

Driving Car

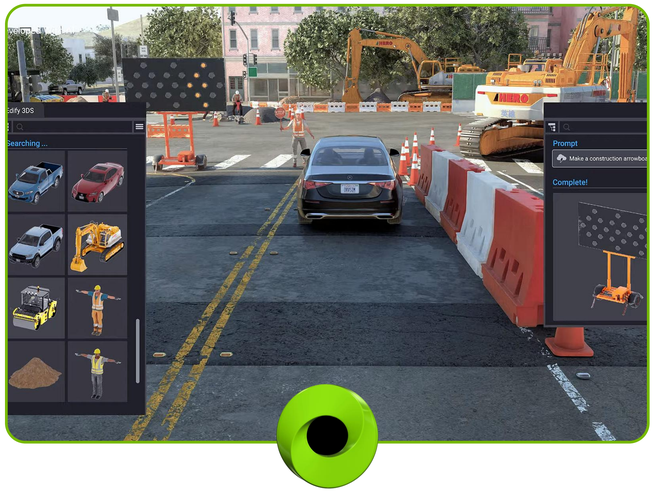


DRIVE AGX Thor

Training AV AI

GB200

Practicing in Digital Twin

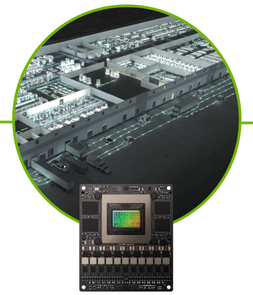


RTX PRO Server

$3-4 Trillion AI Infrastructure Spend by 2030

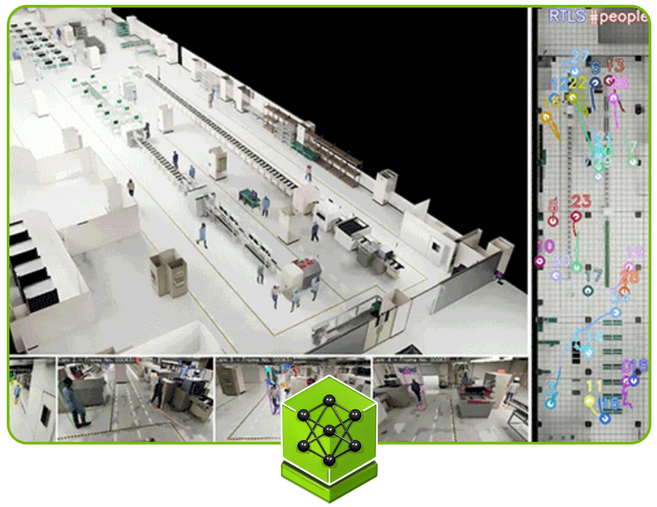
NVIDIA 3-Computers Enable Physical AI – the Next Wave of AI

Running Factories



Jetson AGX Thor

Training Factory AI



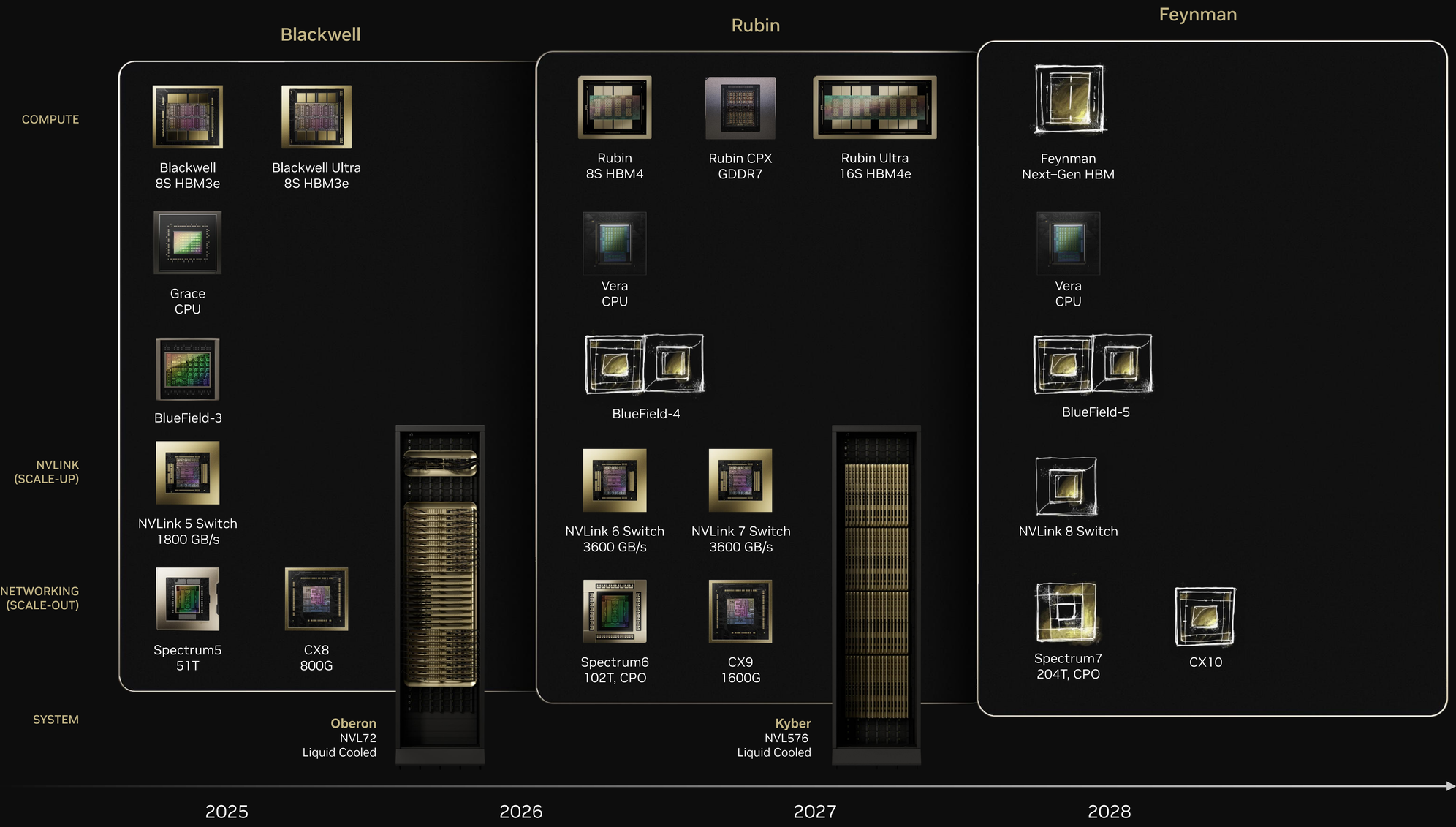
GB200

Practicing in Digital Twin

RTX PRO Server

Annual Rhythm and Extreme Co-Design for Sustained Leadership

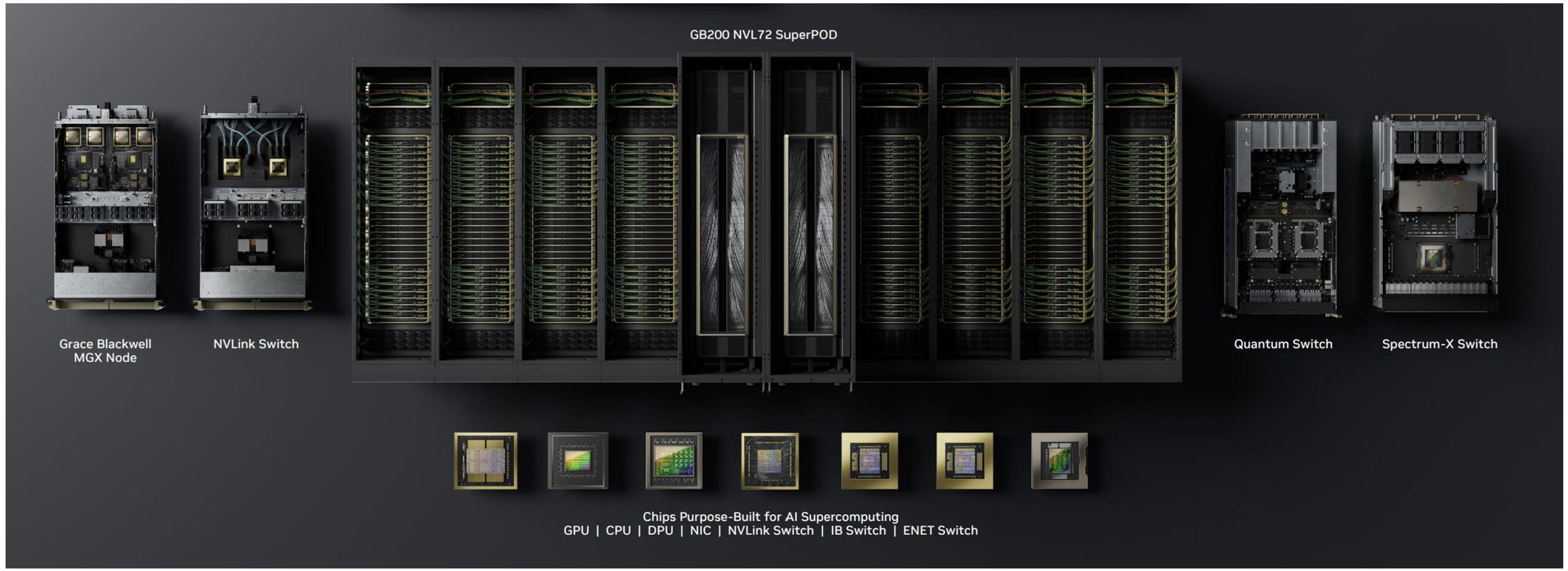
Full-Stack | One Architecture | CUDA Everywhere



Annual Rhythm and Extreme Co-Design for Sustained Leadership

Combining simultaneous breakthroughs in GPU, CPU, NIC, NVLink scale-up fabric, Spectrum-X Ethernet scale-out and scale-across network, system architecture, and a mountain of software and algorithms, we are delivering leaps in performance and cost efficiency never seen before.

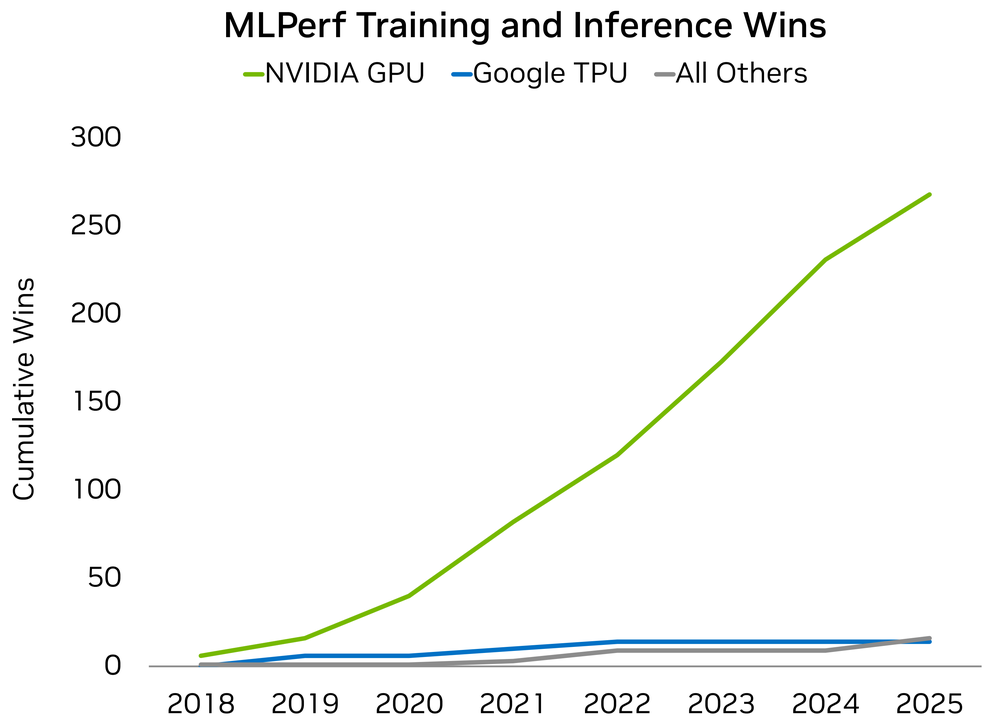
NVIDIA extreme co-design has delivered 1,000,000X performance gain in past 10 years – versus 100X of traditional Moore's law.



Chips to Systems to NVLink to networking to CUDA-X, DOCA, NCCL, TRT-LLM, NIXL, to Dynamo

Annual Rhythm and Extreme Co-Design for Sustained Leadership

NVIDIA Leads MLPerf With Hundreds of Training and Inference Wins



"NVIDIA Blackwell Ultra Sets Reasoning Records in MLPerf Debut"

– *GamesBeat*

"NVIDIA Unveils Rubin CPX Amidst Chart-Topping Blackwell Ultra MLPerf Results"

– *HotHardware*

"Blackwell GPUs Lift NVIDIA to the Top of MLPerf Training Rankings"

– *HPC Wire*

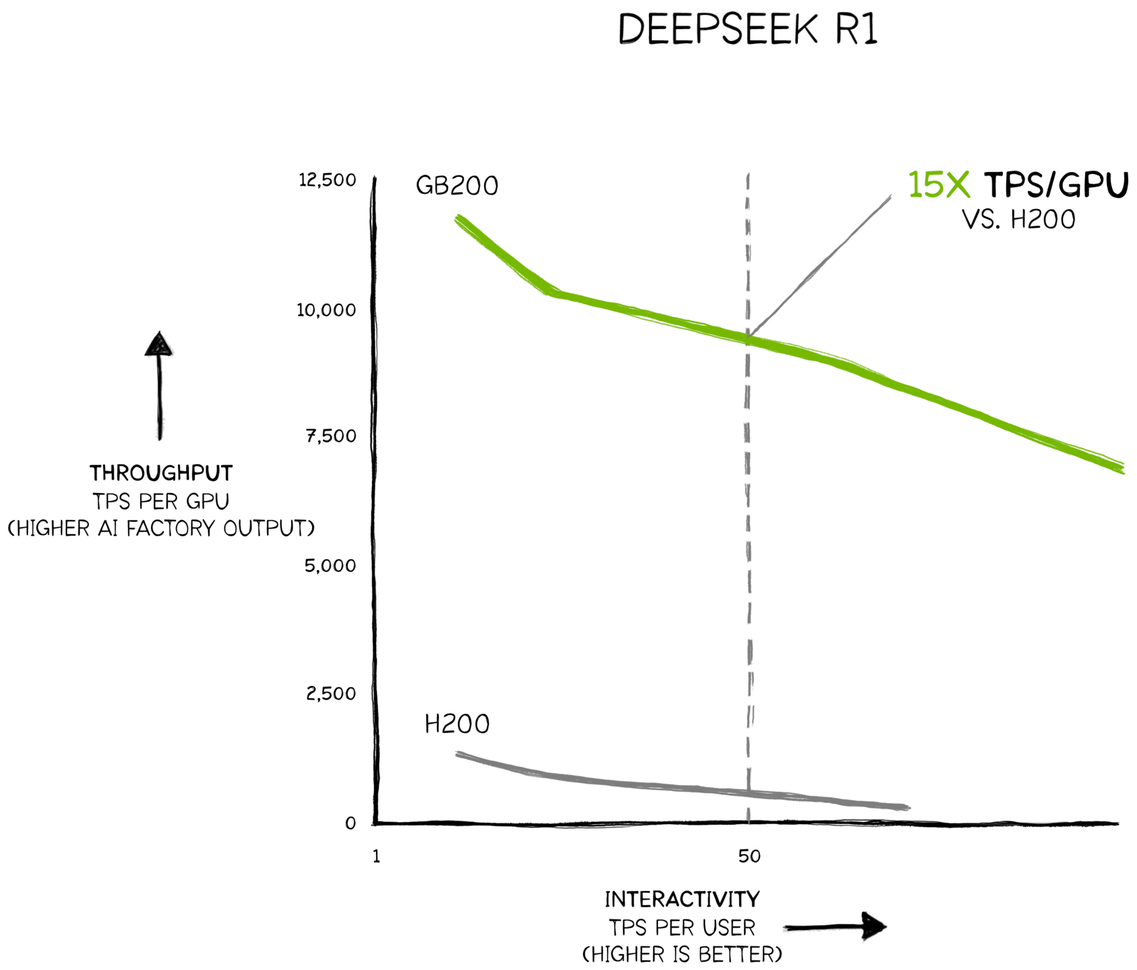
"NVIDIA's Blackwell Conquers Largest LLM Training Benchmark"

– *IEEE Spectrum*

MLPerf Training reported time-to-train and MLPerf Inference: Data Center Reported Total Score

Annual Rhythm and Extreme Co-Design for Sustained Leadership

GB NVL72 for Training to Inference
Order of Magnitude Leap in Inference Throughput and Cost Reduction



NVIDIA and OpenAI Partnership

Multi-Year, Multi-Generational Build Out of at Least 10 Gigawatts

- The OpenAI partnership is a powerful demonstration of NVIDIA's ability as an AI infrastructure partner – delivering architecture, chips, systems, networking, data centers, software, operations, and financing as one integrated solution
- The partnership extends existing collaboration to scale multi-giga-watts of infrastructure at Microsoft, OCI, and CoreWeave
- In addition to CSP capacity, OpenAI and NVIDIA will build at least 10 gigawatts – millions of GPUs – of infrastructure to be operated by OpenAI; The first gigawatt launches in 2026 on the Vera Rubin platform
- For the first time, OpenAI will buy directly from NVIDIA, secure multi-cycle supply, forging close engineering collaboration to build AI factories
- NVIDIA intends to invest in increments over time up to $100 billion in equity; Every 1GW build-out will require $50-60 billion in total spend. OpenAI will need to re-invest future revenue and/or secure other sources of financing to cover the total cost of their build outs.

OpenAI

NVIDIA