NVIDIA Non-Deal Roadshow

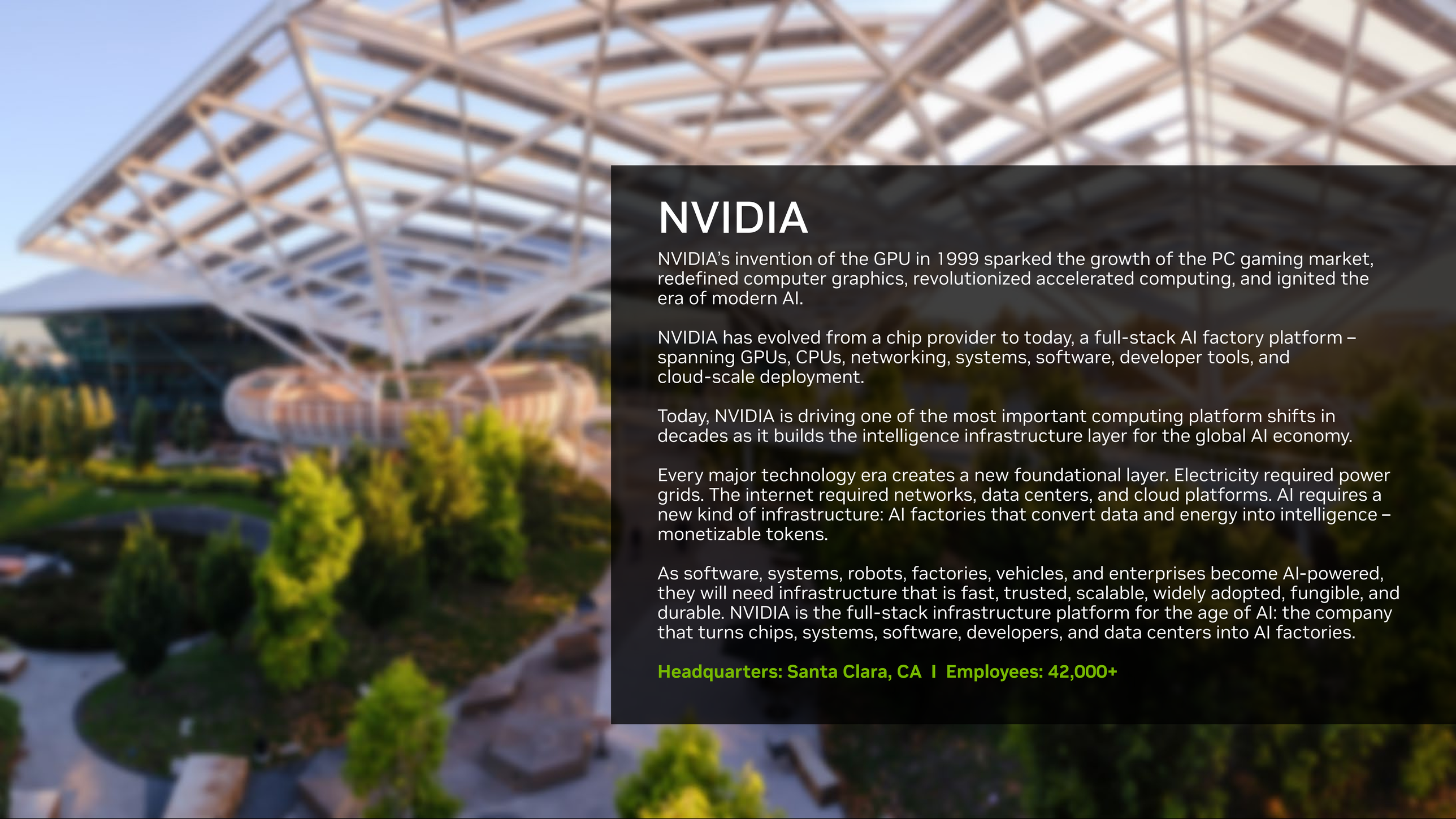
September 2026

Certain matters in this presentation including, but not limited to, statements as to: expectations with respect to growth, demand, performance, benefits, value and availability of and opportunities for NVIDIA's products, services, and technologies, including Vera Rubin, Vera and Groq3 LPX; NVIDIA's financial position and financial and business outlook; NVIDIA's future cash dividends and other returns to shareholders, including share repurchases; expectations with respect to NVIDIA's third party arrangements, including with its collaborators and partners and its proposed Hugging Face acquisition; expectations with respect to NVIDIA's strategies, plans and investments; third parties adopting our products and technologies; expectations with respect to technology developments, and related trends and drivers; projected market growth, opportunities and trends; expectations with respect to NVIDIA's share in the market; expectations with respect to AI and related industries and demand for AI; 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 products 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; NVIDIA's ability to realize the potential benefits of business investments or acquisitions; 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 uses certain non-GAAP measures in this presentation including free cash flow. NVIDIA believes the presentation of its non-GAAP financial measures enhances investors' overall understanding of the company's historical financial performance. The presentation of the company's non-GAAP financial measures is not meant to be considered in isolation or as a substitute for the company's financial results prepared in accordance with GAAP, and the company's non-GAAP measures may be different from non-GAAP measures used by other companies. Further information relevant to the interpretation of non-GAAP financial measures, and reconciliations of these non-GAAP financial measures to the most comparable GAAP measures, may be found in the slide titled "Reconciliation of Non-GAAP to GAAP Financial Measures."



NVIDIA

NVIDIA's invention of the GPU in 1999 sparked the growth of the PC gaming market, redefined computer graphics, revolutionized accelerated computing, and ignited the era of modern AI.

NVIDIA has evolved from a chip provider to today, a full-stack AI factory platform – spanning GPUs, CPUs, networking, systems, software, developer tools, and cloud-scale deployment.

Today, NVIDIA is driving one of the most important computing platform shifts in decades as it builds the intelligence infrastructure layer for the global AI economy.

Every major technology era creates a new foundational layer. Electricity required power grids. The internet required networks, data centers, and cloud platforms. AI requires a new kind of infrastructure: AI factories that convert data and energy into intelligence – monetizable tokens.

As software, systems, robots, factories, vehicles, and enterprises become AI-powered, they will need infrastructure that is fast, trusted, scalable, widely adopted, fungible, and durable. NVIDIA is the full-stack infrastructure platform for the age of AI: the company that turns chips, systems, software, developers, and data centers into AI factories.

Headquarters: Santa Clara, CA I Employees: 42,000+

NVIDIA’s Unique Position in the AI Economy

NVIDIA combines CUDA programmable compute, full-stack AI factory co-design,
a developer ecosystem and global deployment channels in one AI infrastructure platform.

AI Workload Breadth

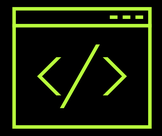
One programmable platform spans training, inference and agentic AI across open and closed models, capturing more of AI’s expanding workloads.

Existing customer demand growing >100% year over year, with an expanding customer base.

Extreme Co-Design

Compute, networking and software co-designed at factory scale improve tokens per watt and tokens per dollar.

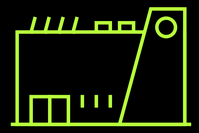
Vera Rubin delivers up to 45X lower cost per token.



CUDA Ecosystem

CUDA compounds developer innovation and installed-base value, keeping NVIDIA compute productive, fungible and durable.

~10 years A100 productive life. Excellent cloud rental economics and strong residual value.



Full-Stack AI Factory

A complete and optimized platform enables more customers to deploy AI factories and expands NVIDIA content per factory.

NVIDIA content growing: Hopper $18B/GW to Rubin $40B/GW.

Global Market Reach

Open models, enterprise agents, sovereign AI and physical AI unlock new builders, industries, and markets.

ACI&E approx. half of the business and growing ~100% year over year.

NVIDIA

Since Our Q2 Earnings Announcement…

1. New models/agents on NVIDIA infrastructure – Astra (OpenAI), Muse (Meta), Grok Bot (SpaceXAI)
2. NVIDIA announced Open Agent Safety Platform with OpenShell and Sentry with broad industry support
3. NVIDIA to acquire Hugging Face for $12.9 billion
4. Anthropic contracts 2.6 GW to date in NVIDIA AI infrastructure to be shipped through 2028
5. Anthropic's total reported contracted value across multiple CSPs/neoclouds now exceeds $180B
6. NVIDIA's neocloud partners in Australia deliver up to 2GW by 2027
7. Nscale and Firmus file for IPO
8. Palantir to deliver Ontology for Cybersecurity via Cisco's Security AI factory with NVIDIA as a preferred full-stack foundation
9. d-Matrix to connect its next-gen XPUs to NVIDIA AI infrastructure via NVLink Fusion
10. Cisco adds rack-scale AI computing solutions to its portfolio

NVIDIA Announces a $150B Share Repurchase Authorization Increase

NVIDIA makes strategic investments across the ecosystem to develop and accelerate growth of the market and CUDA ecosystem. Today, NVIDIA's portfolio consists of 13 public companies and 229 private companies. Realized ROI on exits-to-date exceeds 3X. We are committed to sharing the company's success with shareholders and will return excess free cash flow net of strategic uses in the form of share repurchases and a gradually growing dividend. On September 28, we announced a $150 billion increase to our share repurchase authorization, bringing our total authorization to $235 billion, which we expect to utilize through FY28.

Strategic Investments Across All Layers of the AI Cake

Applications

ABRIDGE · AGILITY ROBOTICS · NOKIA · Orbital Industries · Revolut · SYNOPSYS · WAYVE

Models

ANTHROPIC · Mistral AI · OpenAI · SPACEX · ssi · THINKING MACHINES

Infrastructure

CoreWeave · Crusoe · firmus · Lambda · NEBIUS · NSCALE · VOLTA

Chips

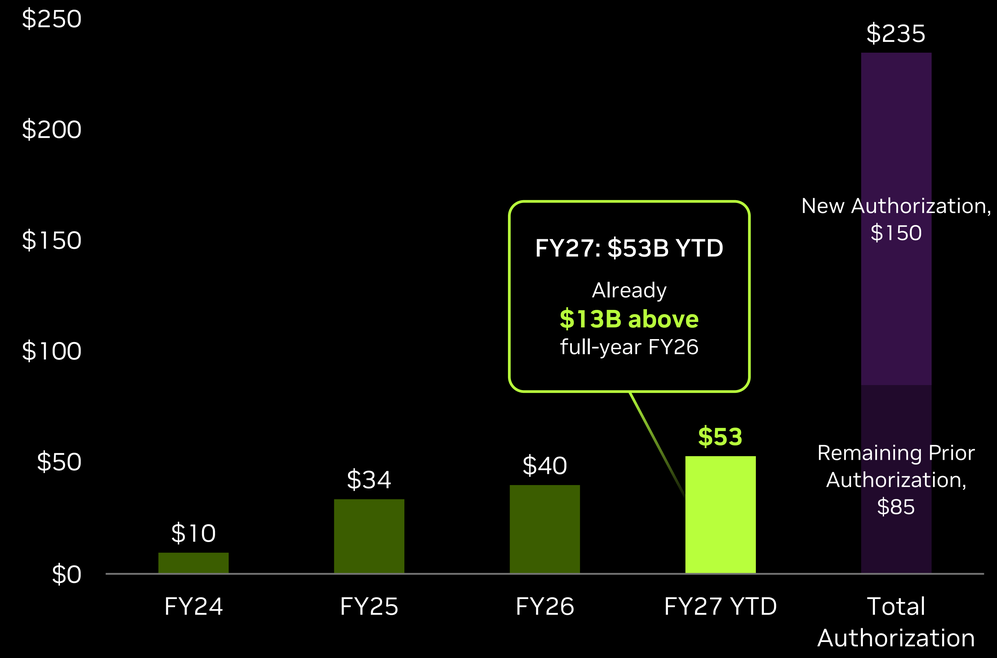
COHERENT · CORNING · intel. · LUMENTUM · MARVELL · MEDIATEK · SiFive

Energy / LPS

ATOMIC CANYON · LANCIUM · REDWOOD MATERIALS · SB Energy · TerraPower

Note: Strategic investments are not exclusive to those included in the table.

Increasing Stock Repurchases ($ Billions)



NVIDIA

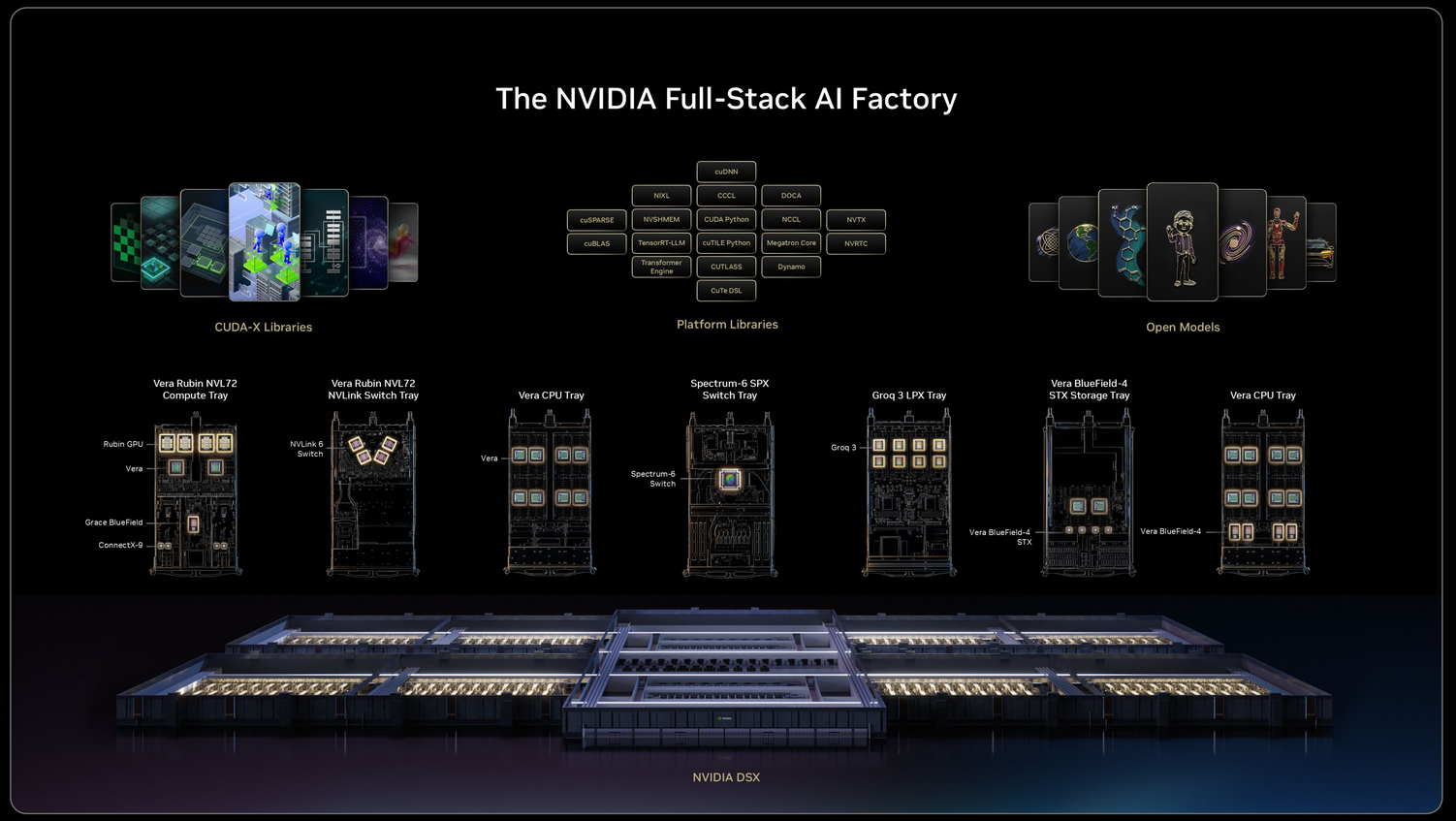
NVIDIA Full-Stack AI Factory Platform, Not the Chip, Determines Economic Output

What NVIDIA builds is fundamentally different from what other chip companies build.

The Vera Rubin platform integrates seven purpose-built chips across five rack-scale systems. And, through the DSX AI factory platform, NVIDIA helps design, build and operate efficient, resilient and grid-ready AI factories.

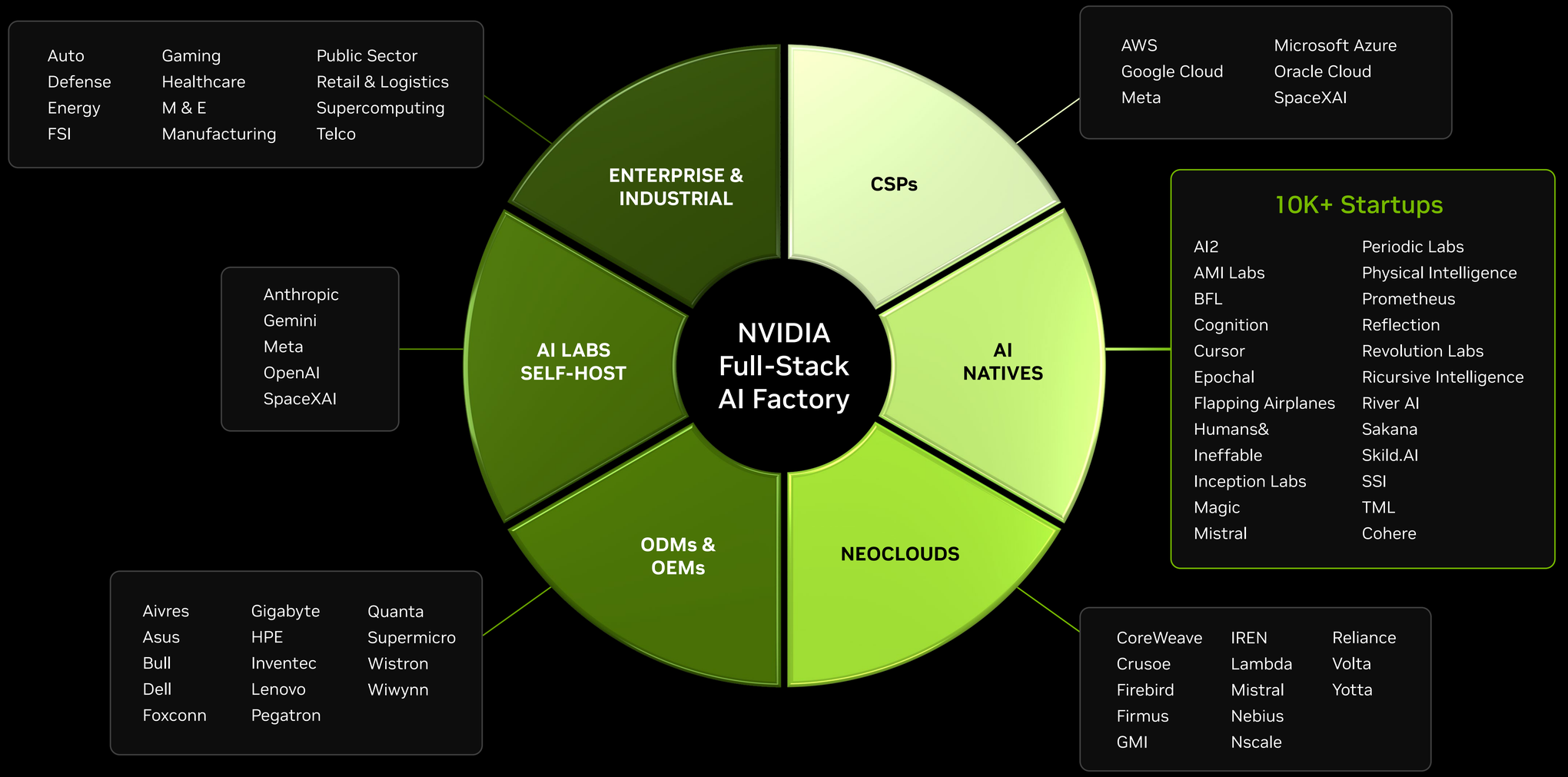
The ability to extreme co-design at AI factory scale is unique to NVIDIA, and the result is integrated yet customizable infrastructure that can be deployed across every major cloud, inside enterprise data centers, at the edge, and into robotics and physical AI systems.

NVIDIA delivers x-factor performance gains every year and continuously improves the performance of its installed base through software enhancements.



CUDA Programmability and Full-Stack AI Factory Expand the Global Diverse AI Market

Programmability open new growth | CUDA-X libraries connect ecosystems | Full-stack AI factory open go-to-market options



NVIDIA Will Leverage Its Financial Capacity to Support the Ecosystem

Future commitments by fiscal year as of July 26, 2026, were as follows:

	Remainder of 2027	2028	2029	2030	2031	2032 and thereafter	Total
				(In billions)			
Supply and capacity	$ 92	$ 87	$ 88	$ 6	$ 5	$ 1	$ 279
Cloud service agreements	3	8	7	6	4	1	29
Data center leases not commenced	—	1	1	2	1	20	25
Equity investments	18	3	2	2	—	—	25
Capital expenditures	7	1	—	—	—	—	8
Total	$ 120	$ 100	$ 98	$ 16	$ 10	$ 22	$ 366

We've partnered with our extensive network of suppliers to secure the critical components needed to meet demand for the next several years. Our commitments increased from $119 billion last quarter to $279 billion, primarily related to the procurement of memory.

Future commitments by fiscal year as of July 26, 2026, were as follows:

	Remainder of 2027	2028	2029	2030	2031	2032 and thereafter	Total
				(In billions)			
AI cloud agreements	$ —	$ 6	$ 8	$ 7	$ 6	$ 9	$ 36
Data center leases not commenced for third party	—	—	1	1	1	17	20
Total	$ —	$ 6	$ 9	$ 8	$ 7	$ 26	$ 56

We've partnered with leading AI clouds to enable broader access to our AI infrastructure to serve AI startups, model builders, enterprises, research organizations and sovereign customers. Under these agreements, we will earn revenue on the upfront sale of our infrastructure and, if certain criteria are met, we will participate in revenue share generated by the AI clouds from their third-party customers.

We signed data center lease agreements with terms of ~15 years that are expected to commence between FY2028 and FY2029. We expect to reassign these data center leases to third parties.

The following table summarizes the maximum gross exposure related to our guarantees, including the SB Energy Corp. guarantees signed in August 2026 (in billions):

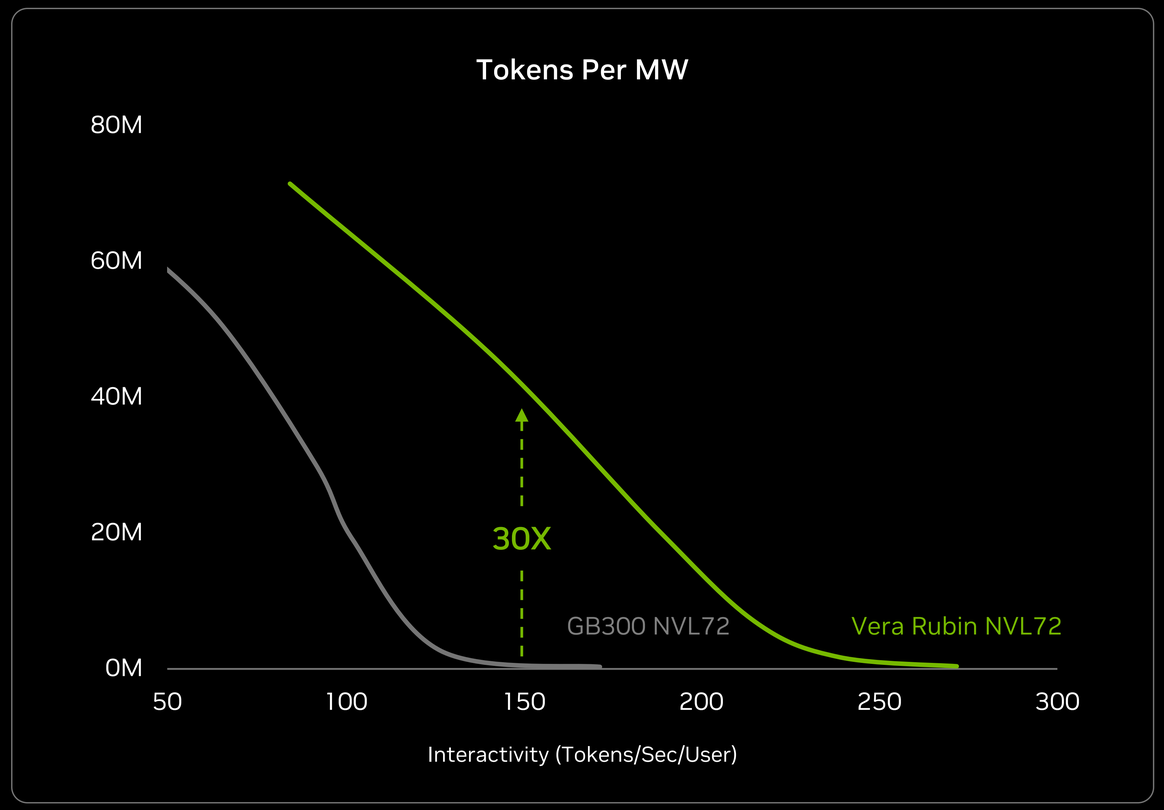
Land, power, and shell guarantees for AI clouds	$ 3.5
SB Energy Corp. guarantees	105.0
Total	$ 108.5

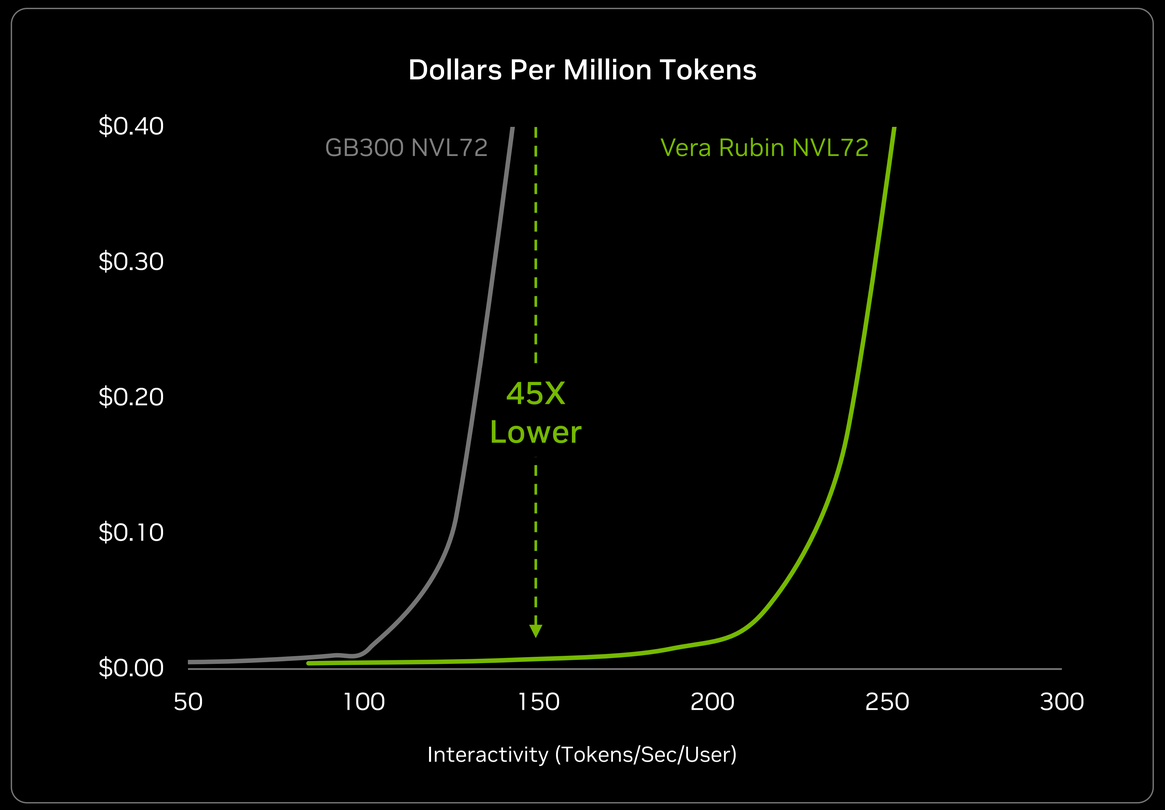
We entered into guarantees to provide credit support on the land, power, and shell buildout to secure approximately 4.25 GW at SB Energy's PORTS-Pike Technology Campus in Ohio, which will exclusively host several generations of NVIDIA infrastructure under 20-year leases to OpenAI, subject to limited exceptions.

AI Factories Drive Token Economics — Compute is Revenue

Vera Rubin NVL72 delivers up to 30X better AI factory token throughput and 45X lower cost per million tokens.

DeepSeek-V4-Pro | SemiAnalysis AgentX

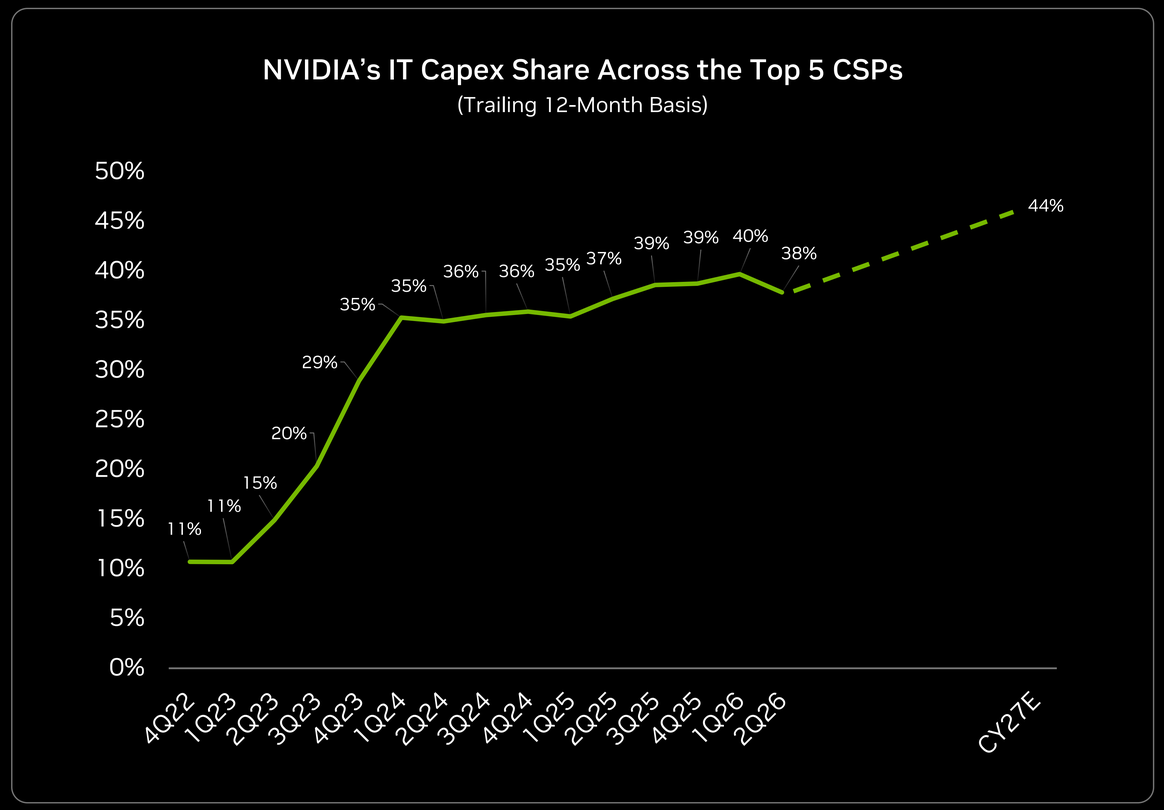


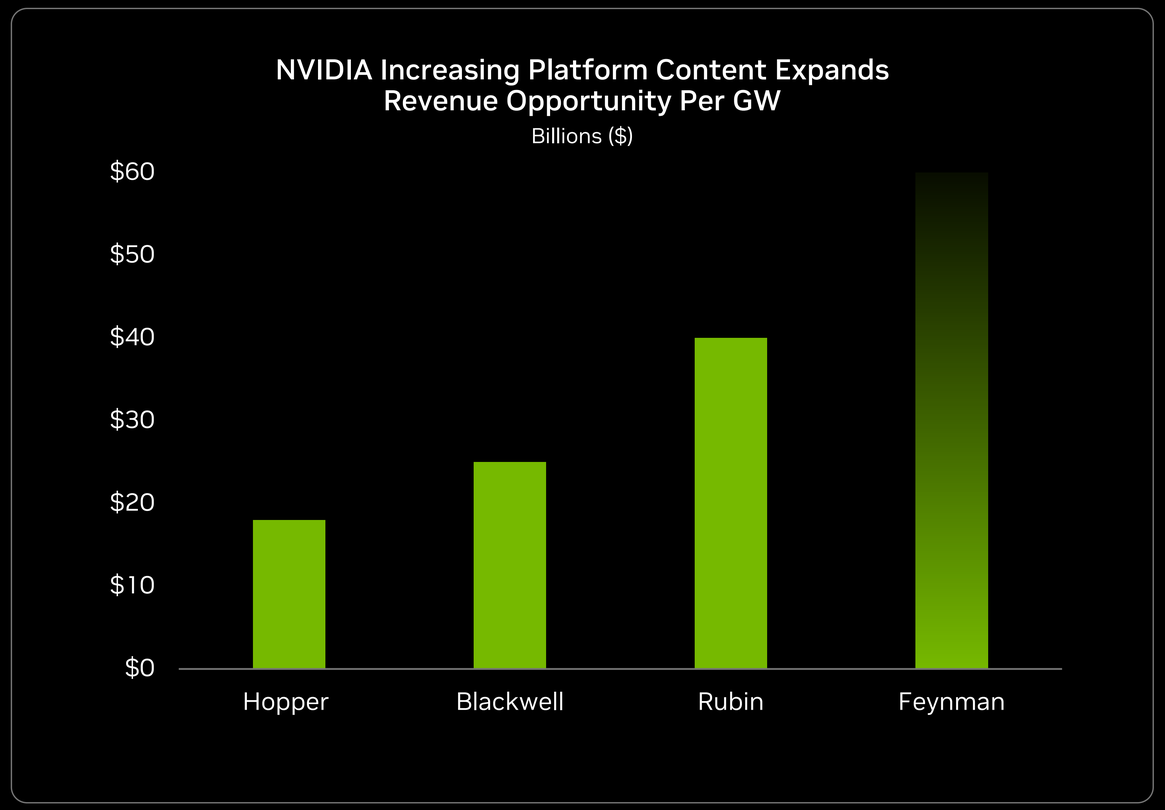


NVIDIA

NVIDIA Expanding AI Market While Expanding Revenue Opportunity Per GW

NVIDIA continues to increase its share within the overall AI infrastructure market. NVIDIA's per gigawatt revenue opportunity – based on its reference design – has expanded from $18 billion during the Hopper generation to $25 billion with Blackwell and now $40 billion with Rubin. The growth in NVIDIA's SAM (served available market) has been driven by advancements in the GPU as well as a broadening in the company's portfolio – which now includes stand-alone CPU racks for agentic AI, LPU racks for low-latency, large-context inference, and scale-in networking for secure, performant AI factories. This is all a reflection of NVIDIA's strategy in becoming a full-stack AI factory platform.

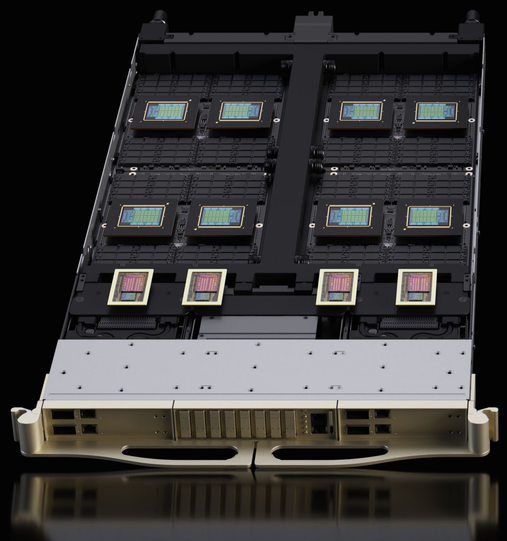




Increasing Share of Data Center CapEx

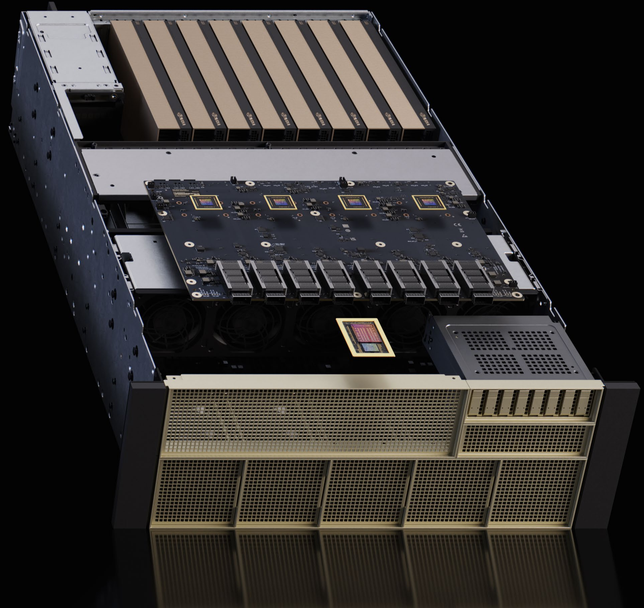
Vera Rubin NVL72
Compute Node
Model Processing

Vera MGX
Compute Node
Secure Agent Processing



Vera CMX
Storage Node
Context KV$ Offload

Vera DPX
Storage Node
Data Processing



Announcing NVIDIA Open Agent Security Platform

An Open Reference Design with NVIDIA OpenShell and NVIDIA Sentry

OpenShell is a policy-controlled runtime between an agent and the enterprise systems it can affect – data, files, credentials, tools, APIs, models, and networks.

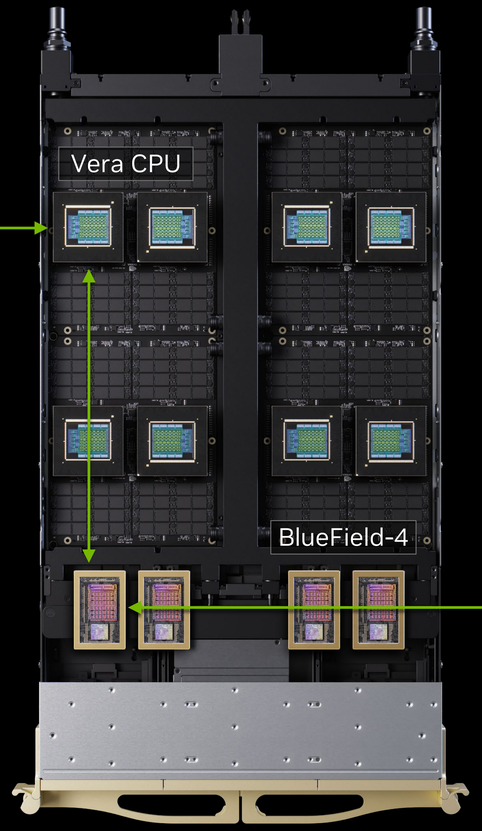
OpenShell addresses the agent security problem: 1) Prevents uncontrolled access to enterprise data, 2) Limits credential exposure and misuse, 3) Controls agent tool use and code execution, 4) Applies default-deny network egress where appropriate, 5) Helps contain prompt injection, malicious retrieved content, and harmful tool calls, and 6) Creates an auditable runtime boundary for long-running autonomous workflows.

Secure Runtime Boundary

Governs and Mediates All Actions, Enforces Policy at Runtime

NVIDIA OpenShell

- Gateway
- Supervisor
- Policy Prover
- Agent Harness
  - Agent Sandbox



Vera MGX Compute Node

Secure Agent Processing

Prompt
Tools
Data
Model
Policies

In-Silicon Out-of-Band Monitor

Agent Reasoning Inspection and Tamper-Proof Telemetry

NVIDIA Sentry

- Hardware-Based Enforcement
- Continuous Agent Monitoring and Threat Detection
- Trusted Telemetry and Detection Pipeline
- Millisecond-Scale Containment and Quarantine

Prompt | Tools | Data | Model

NVIDIA

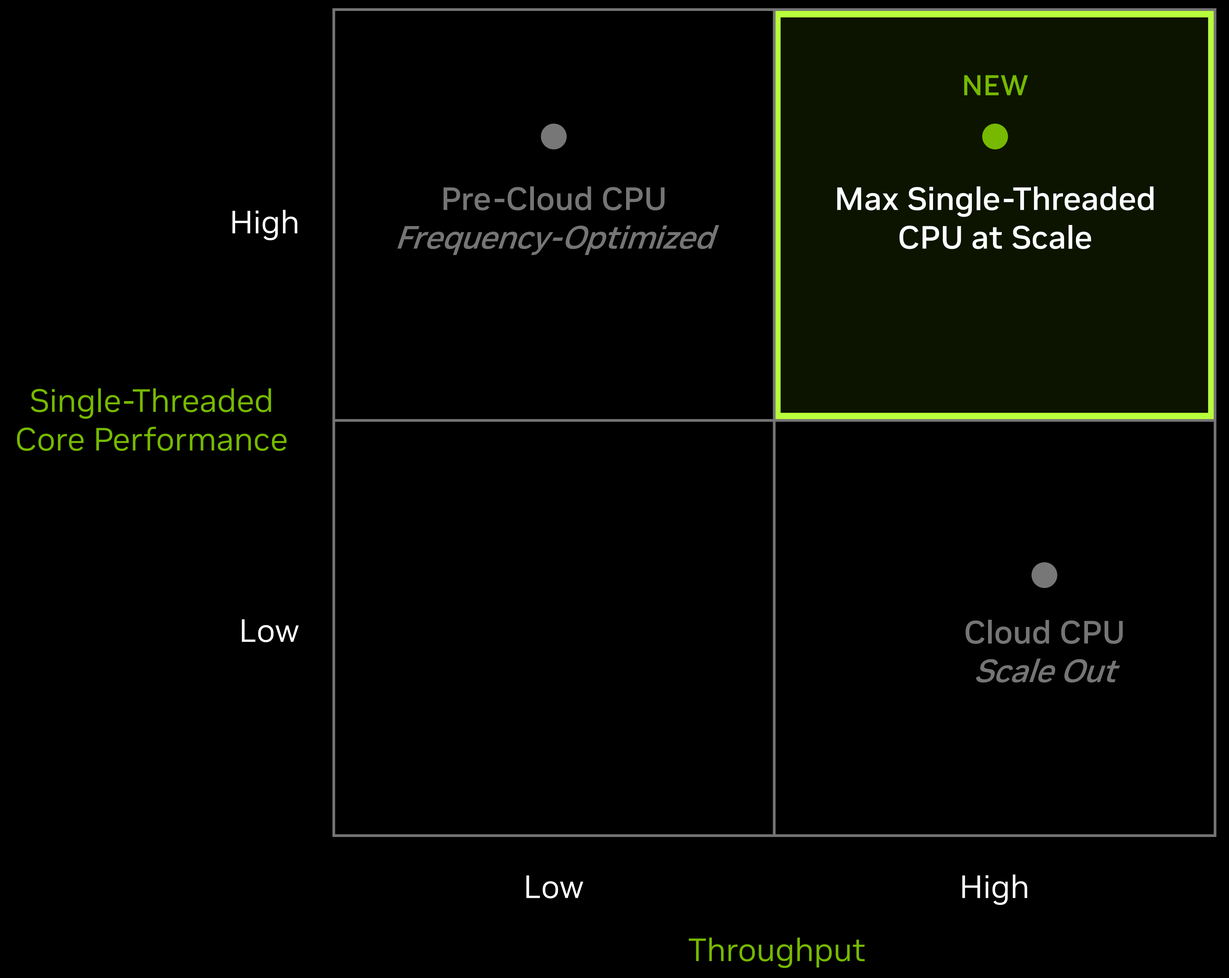
NVIDIA Vera Delivers Maximum Single-Threaded Performance at Scale

NVIDIA's data center CPU journey began in 2021 with the introduction of the Grace CPU.

Today, we are in full production of our next-generation Vera CPU. As a stand-alone product, Vera expands NVIDIA's SAM even further.

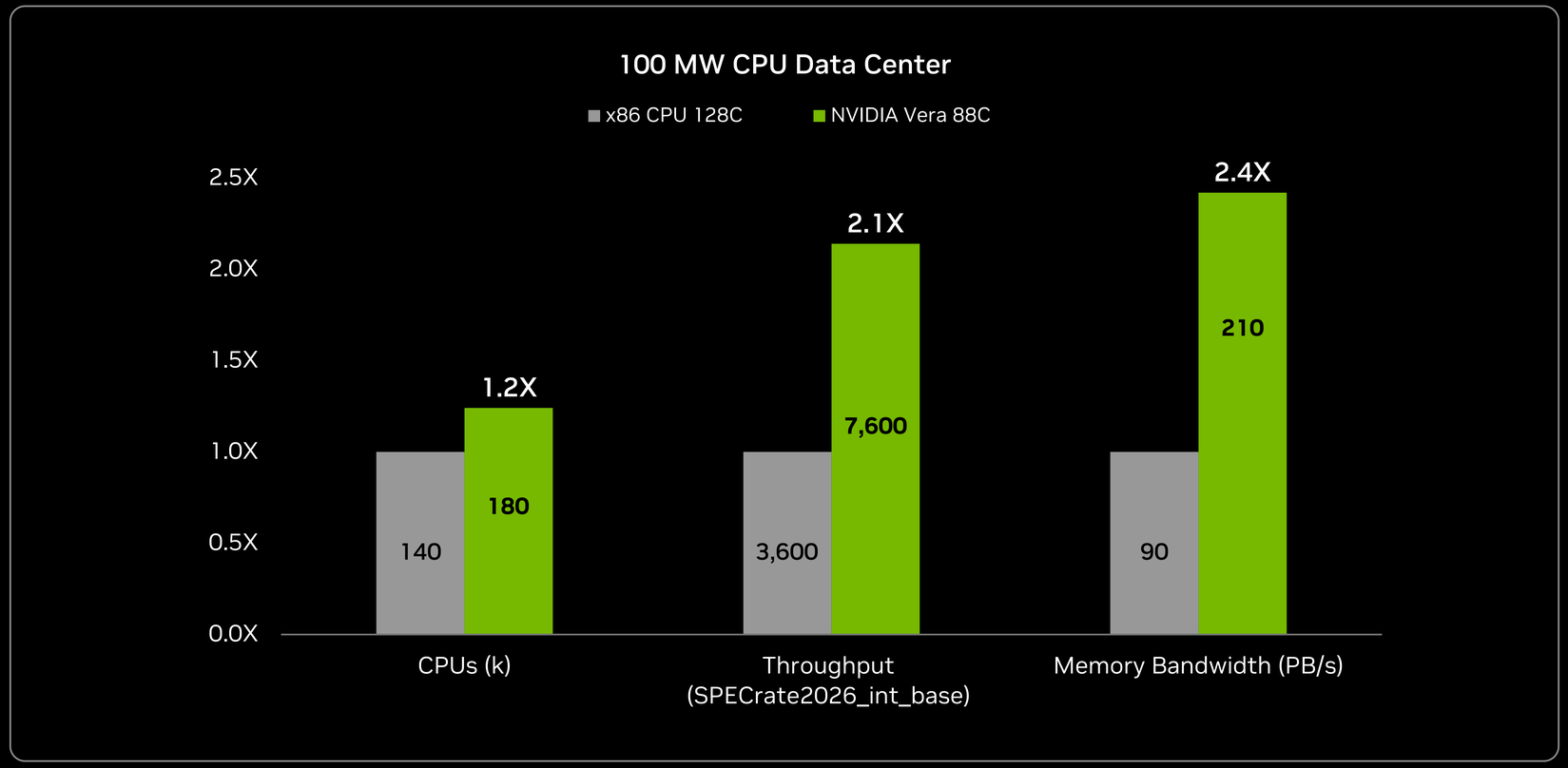
CPUs have historically traded speed against volume: how fast one task finishes versus how many run at once. Pre-cloud CPUs chose speed, while cloud CPUs chose volume (more but slower cores).

Agentic AI needs both, and Vera is the first data center CPU purpose-built for these workloads.



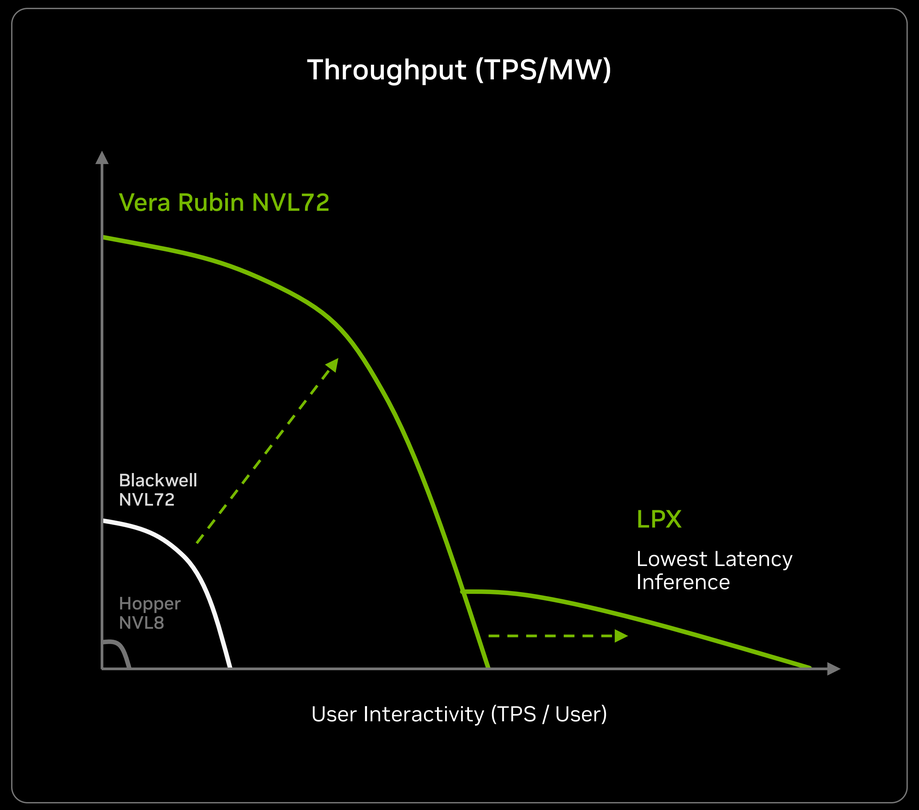
NVIDIA Vera Delivers Over 2X More Performance and Memory Bandwidth for a 100 MW CPU Data Center

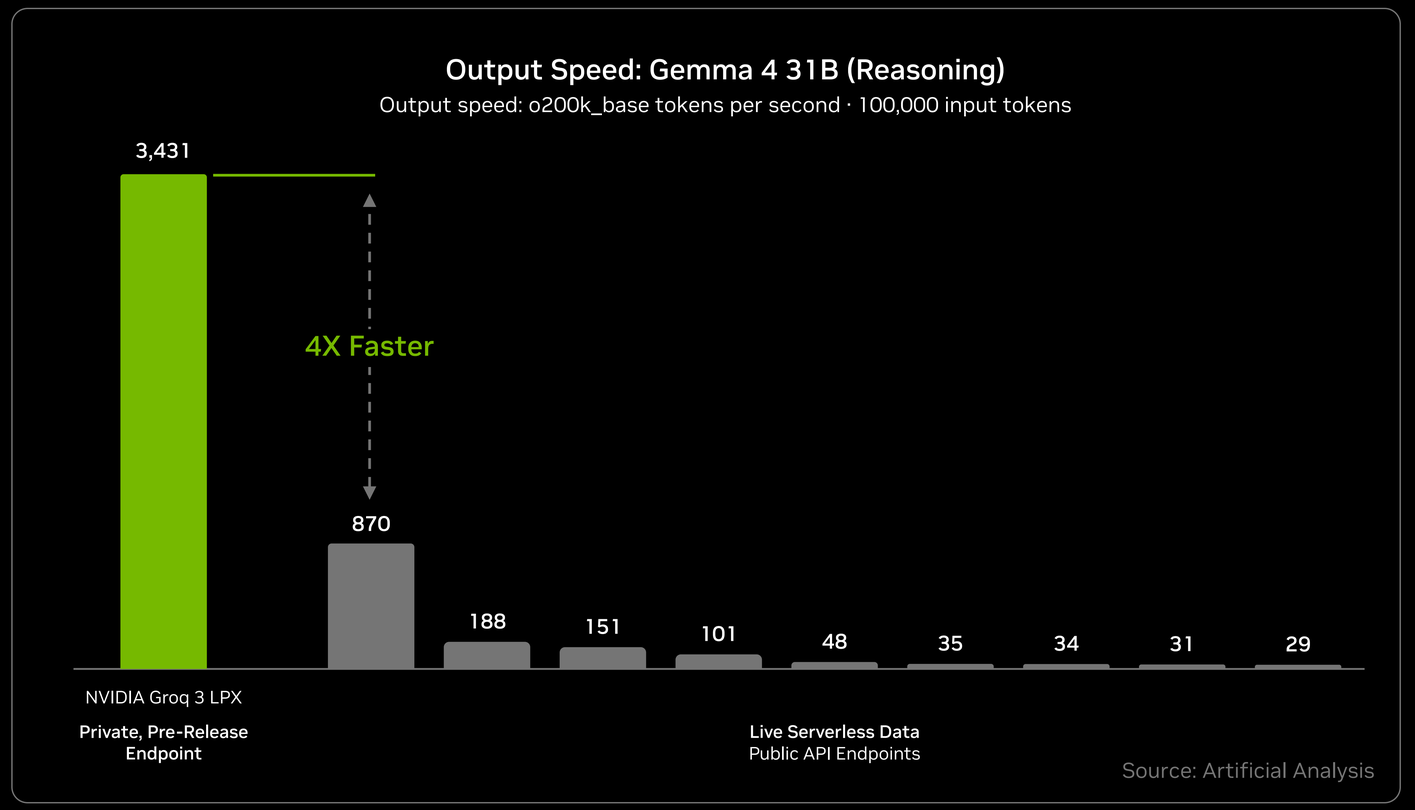
Vera completes agentic tasks 2.2X faster on the SPEC benchmark and provides 2.5X more bandwidth than leading x86 data center CPUs. With shipments already underway to lead partners including OCI, SpaceXAI and AWS, Vera is expected to be deployed by every major hyperscaler, neocloud, AI lab, and system OEM. With CPU revenue expected to more than double in FY2028, NVIDIA is set to become one of the world's leading CPU suppliers.



Groq 3 LPX Co-Designed with Vera Rubin Delivers Leadership Low-Latency Inference Performance

NVIDIA is in full production of Groq 3 LPX, the company's first rack-scale LPU system. Together with the Vera Rubin NVL72 system, LPX expands the frontier addressable by NVIDIA's full-stack computing platform – specifically, the low-latency inference segment of the market. Groq 3 LPX is already setting records, demonstrating nearly 4x the number of tokens per second against the next best alternative on Artificial Analysis' Gemma 4 31B benchmark.





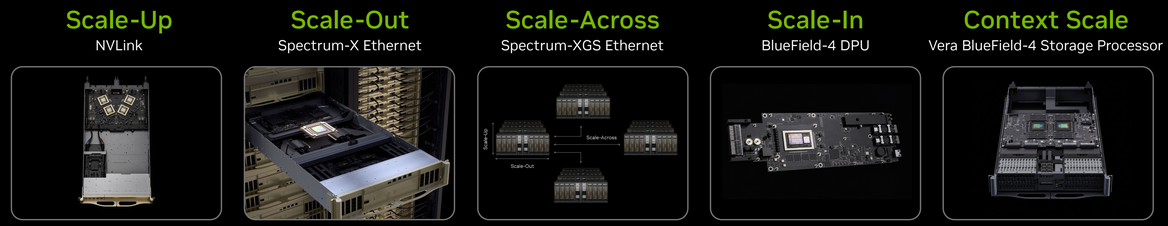
NVIDIA is the Largest and Fastest Growing AI Networking Company in the World

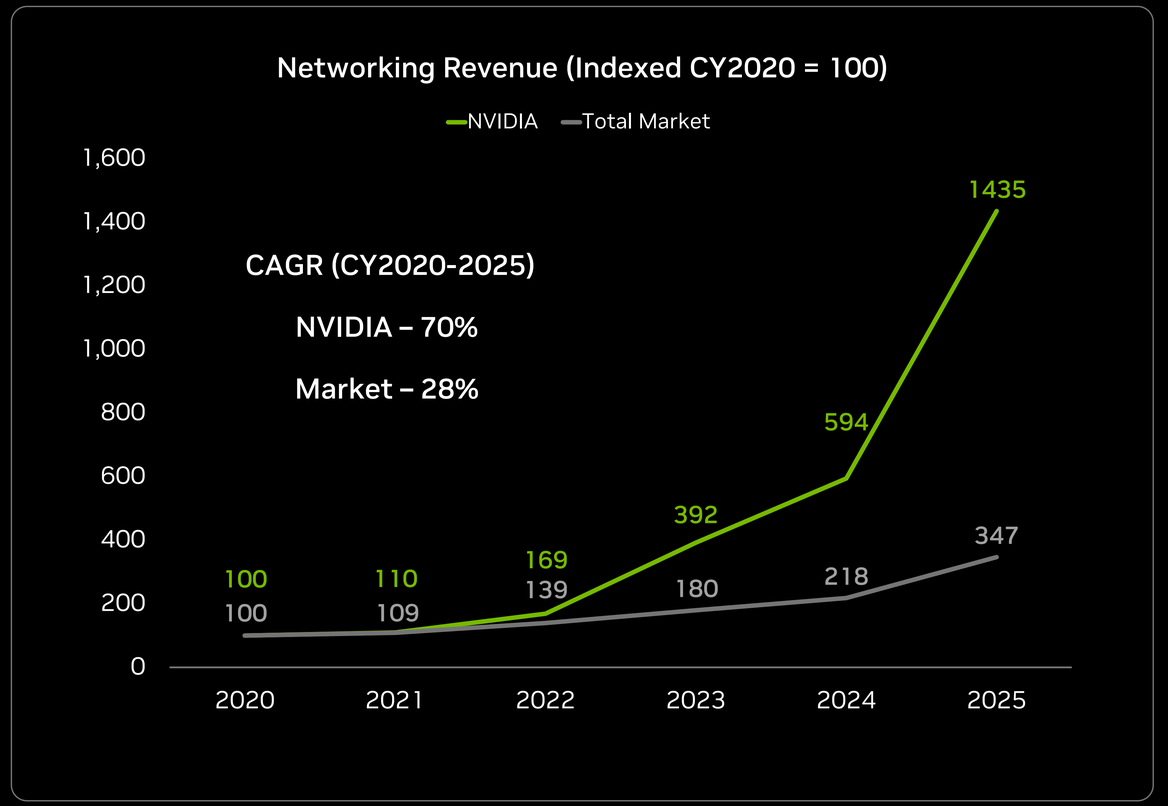
NVIDIA is the only company that offers five networking technologies:

1. Scale-Up (NVLink)
2. Scale-Out (InfiniBand, Spectrum-X Ethernet)
3. Scale-Across (Spectrum-XGS Ethernet)
4. Scale-In (BlueField-4 DPU and DOCA over Spectrum-X Ethernet)
5. Context Scale (Vera BlueField-4 Storage Processor)

Since the acquisition of Mellanox in 2020, NVIDIA's Networking revenue has grown at a CAGR of 70%, materially outperforming industry peers.

Today, NVIDIA is the largest and fastest growing AI Networking company in the world.



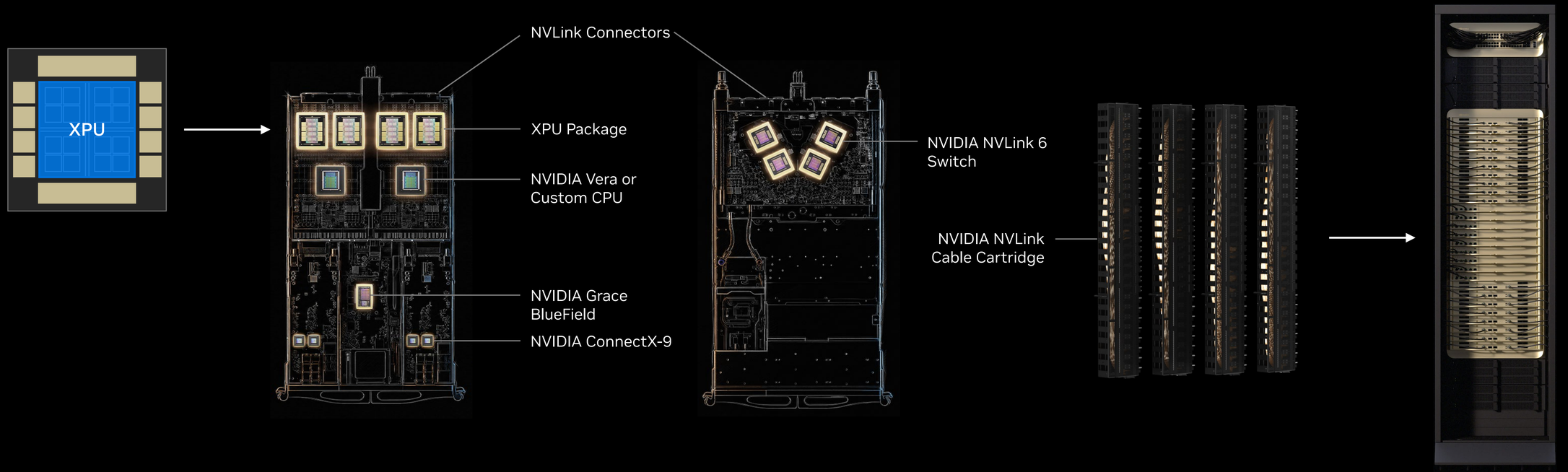


NVLink Fusion Expands NVIDIA Opportunities

NVIDIA NVLink Fusion offers customers the option to build semi-custom AI factories by integrating their custom XPUs or CPUs with NVIDIA's proven rack architecture. And to accelerate adoption, NVIDIA is partnering with various CPU, custom silicon, EDA and optical interconnect suppliers.

Customer reception has been strong with AWS, Intel, RIKEN, and others already committed to adopting NVLink scale-up technology.

NVLink Fusion opportunity to approach $250 billion by the end of the decade.



Compute Tray x18
Cableless, Hoseless, Modular Design
45C Inlet Temp

NVIDIA NVLink Switch Tray x9
3.6 TB/s per XPU All-to-All Bandwidth
3X Lower Latency
Built-In Factory Resiliency

NVIDIA NVLink Spine x1
Connects 72 XPUs in a Single Domain

NVIDIA MGX Rack
Third-Generation Design
100% Liquid Cooled

NVIDIA

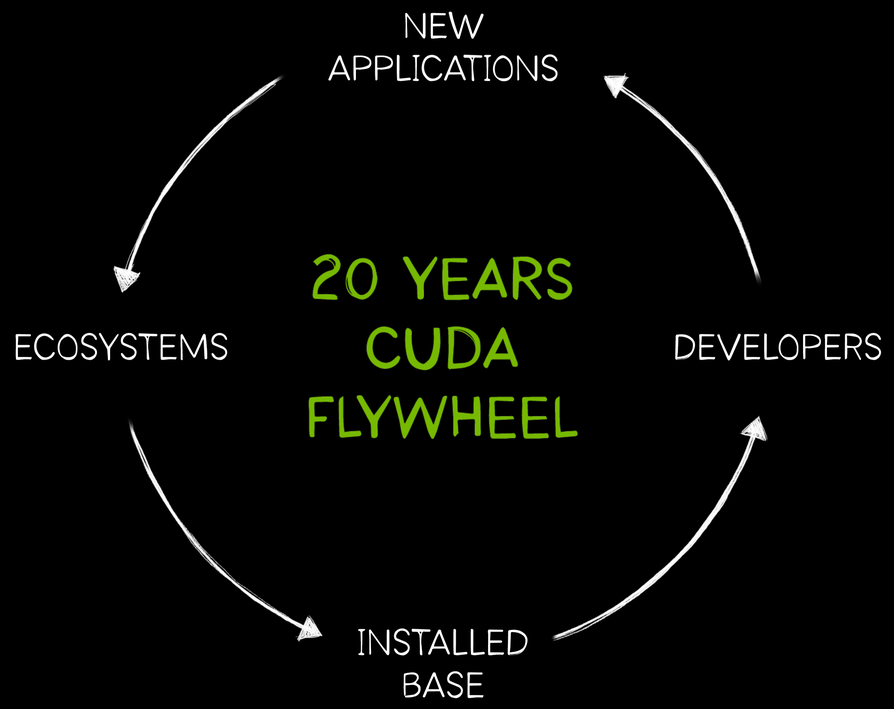
NVIDIA CUDA Turns Programmability into Ecosystem Scale

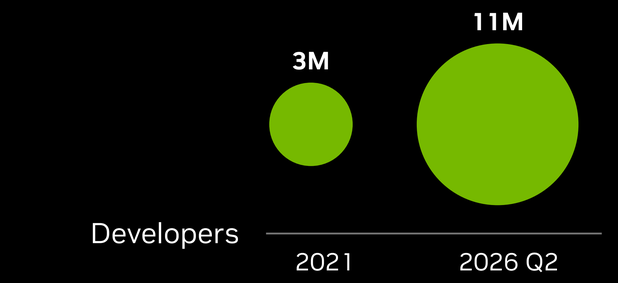
NVIDIA CUDA defines NVIDIA's strategy and powers the flywheel. Developers drive new applications. New applications create new ecosystems. New ecosystems grow the installed base. And a larger installed base invites more developers.

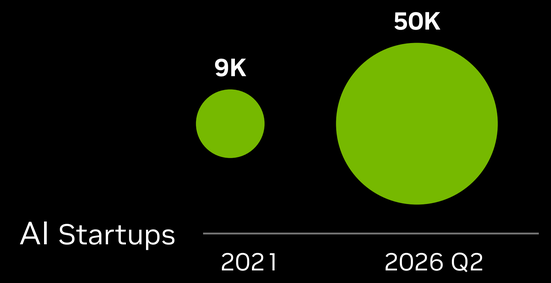
NVIDIA compute accelerates every workload. AI and non-AI. All fields of science.

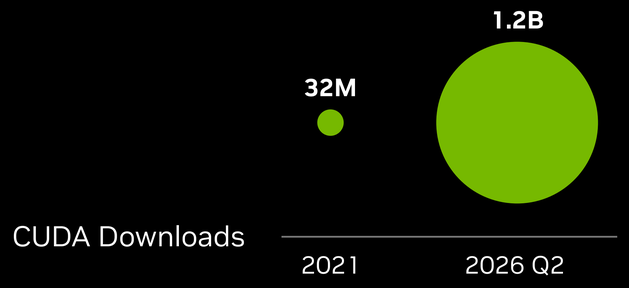
NVIDIA continuously updates its software, driving down the cost of computing, and improving the installed base's productivity.

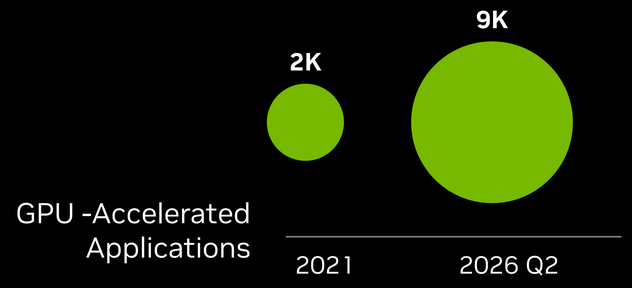
Built on CUDA, NVIDIA compute is versatile, fungible, durable, and therefore investable.







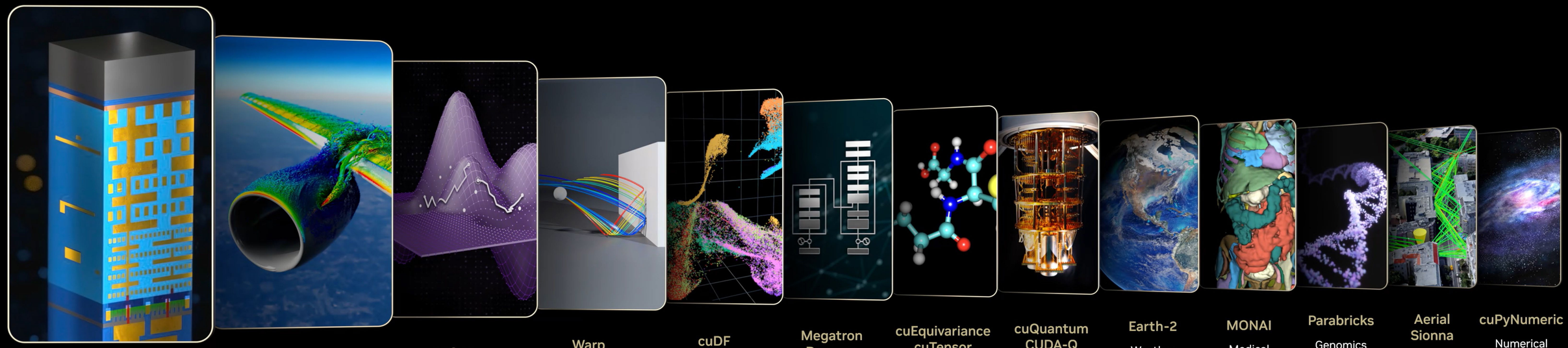




NVIDIA CUDA-X Accelerates Apps and Reduces the Cost of Computing

CUDA-X libraries accelerate apps, enabling them to run orders of magnitude faster, more efficiently, and at lower cost than on CPUs. Over time the introduction of new libraries expands our platform's addressable set of applications and use cases, unlocking new markets and increasing our long-term growth opportunity.

With ~9,000 apps, AI models, and libraries running on CUDA, NVIDIA addresses the largest market opportunity of any computing platform.



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

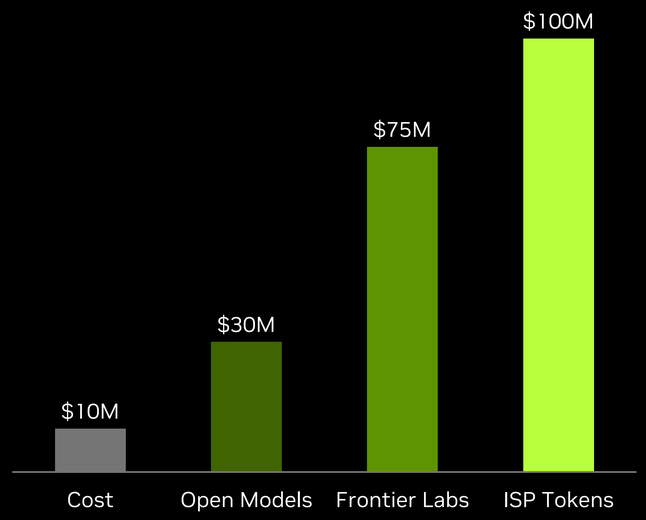
Aerial
Sionna
5G/6G Signal Processing

cuPyNumeric
Numerical Computing

NVIDIA AI Factories Are Productive, Durable, and Fungible

Productive
Profitable Tokens

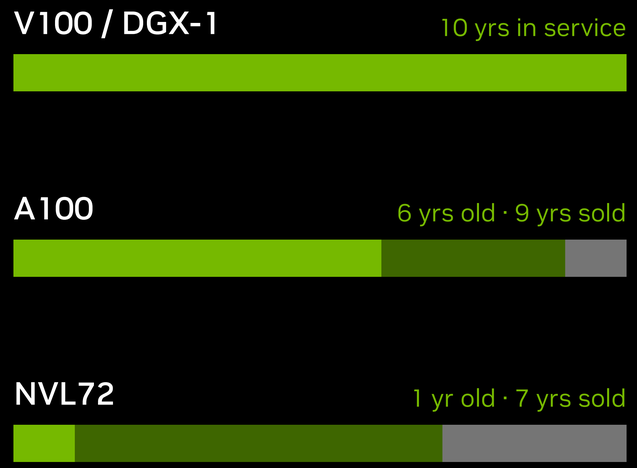
Payback well inside depreciation window



Revenue $M/MW/year
Cost: $60M/MW build · 6-yr depreciation

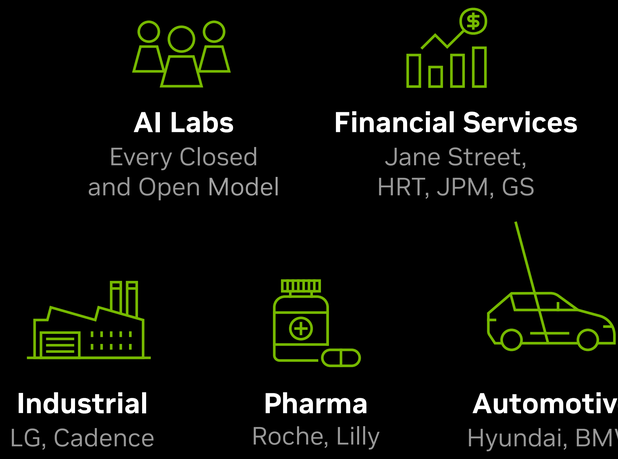
Durable
CUDA-Driven Installed-Base Enhancement

Earns after full depreciation and improves over time through software



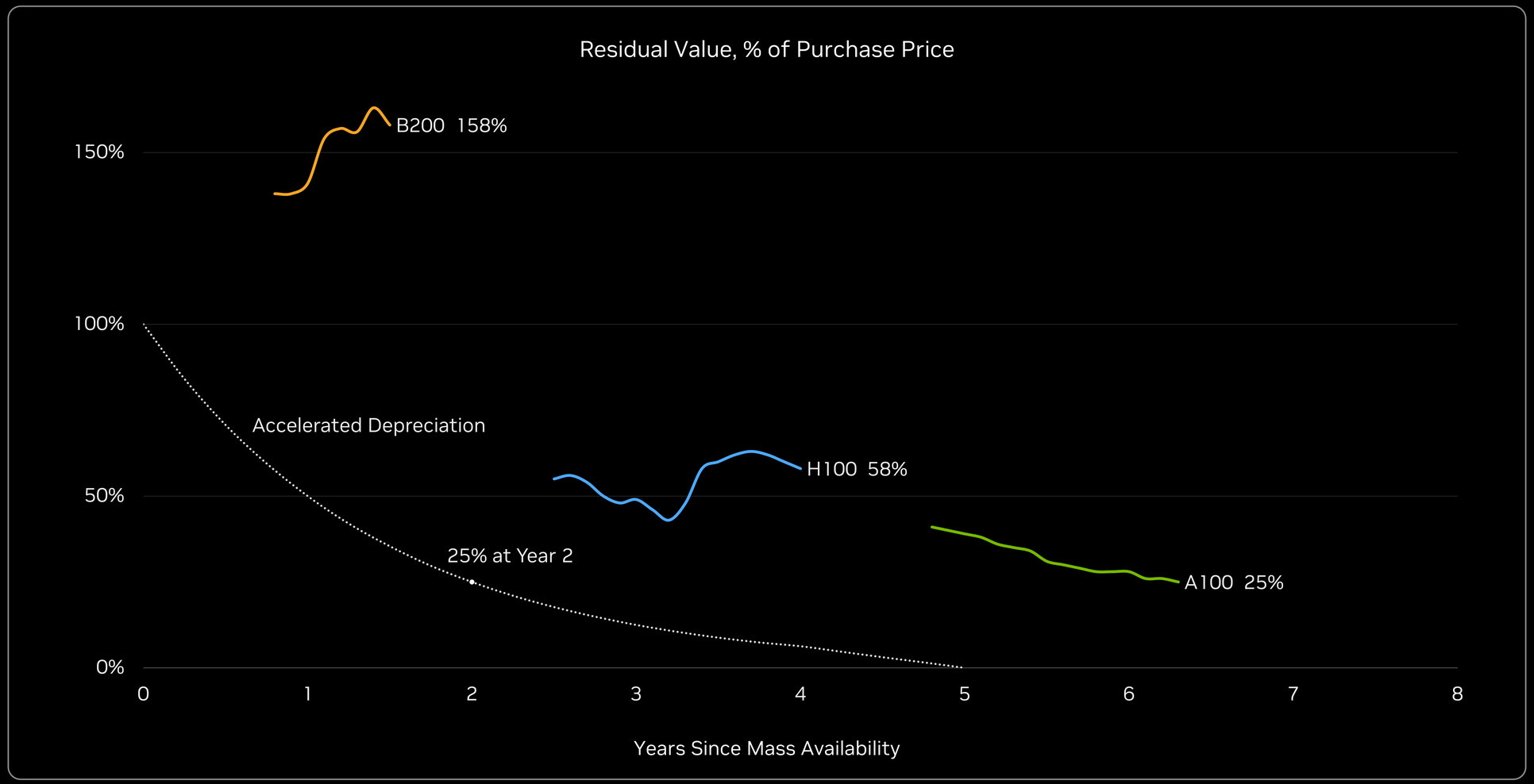
Fungible
One CUDA Architecture – Global Ecosystem

One platform, many offtakers, any workload



NVIDIA AI Infrastructure Retains Value Beyond Accelerated Depreciation Schedules

8-GPU HGX System Cost Per GPU · Dotted Curve = Accelerated Schedule, 25% Residual at Year 2, Zero by Year 5



Residual value: Silicon Data, monthly averages. Purchase price basis: A100 $20,000 · H100 $35,000 · B200 $45,000.
Mass-availability dates assumed; purchase prices are third-party estimates, not Silicon Data data.

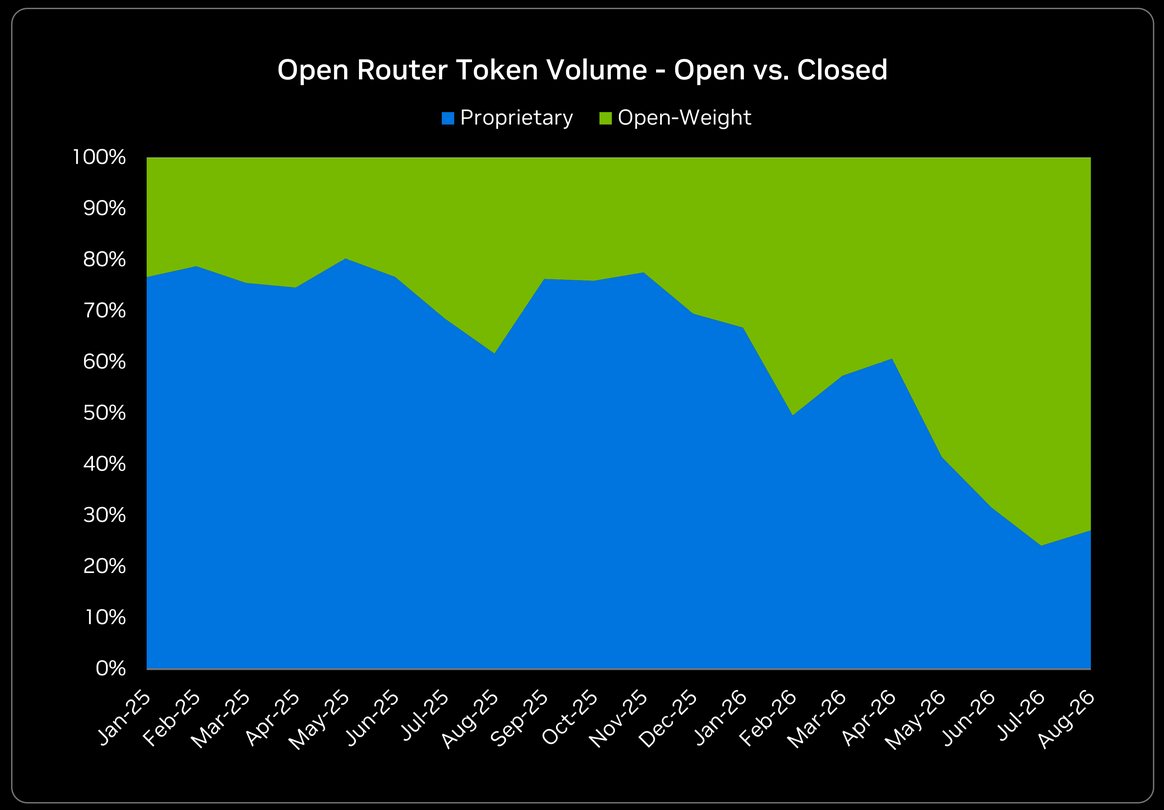
NVIDIA

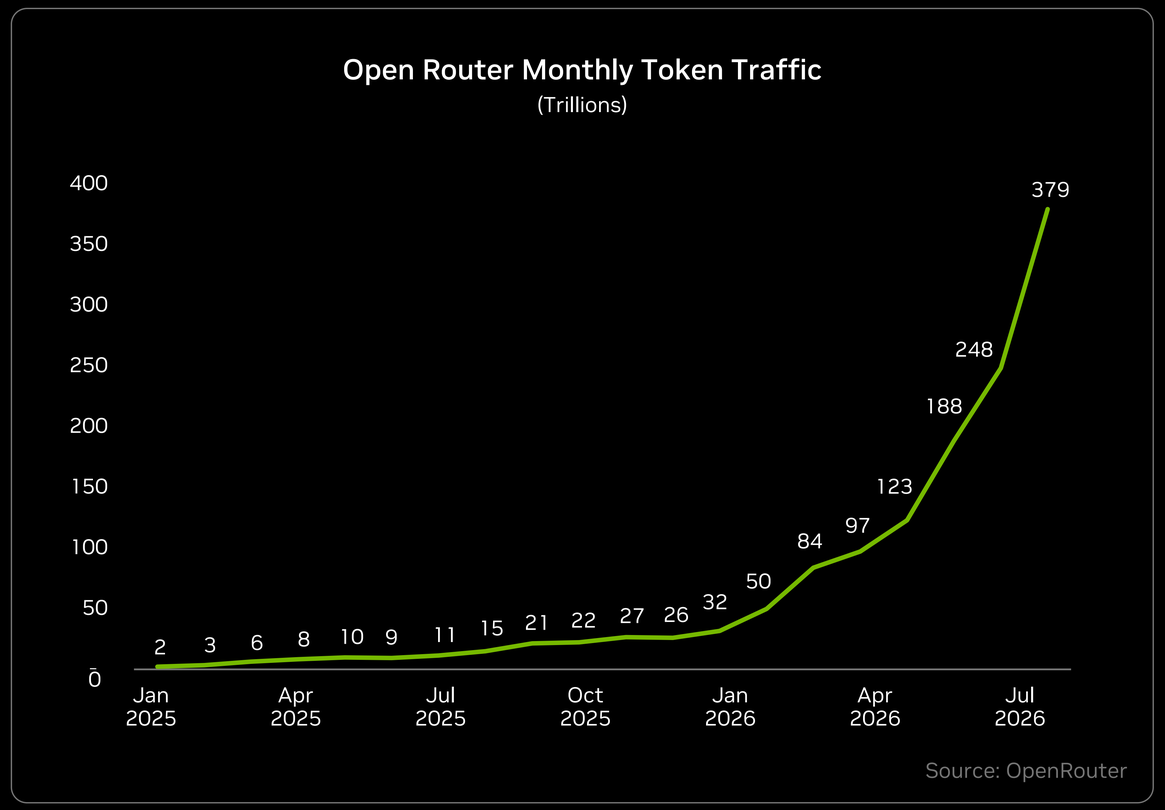
Open Models Enable New Builders and Create New Markets

Advancements in open models are contributing to the usefulness and proliferation of AI.

Open-weight models run primarily on NVIDIA and account for ~75% of token generation today, up from ~40% a year ago.

Token consumption is growing exponentially – up 25X over the past year.





NVIDIA is the Largest Contributor to Open Models

NVIDIA builds frontier open models and contributes data, software, and standards across the AI ecosystem. We are committed to investing in and growing the open source community to enable the ecosystem and drive accelerated growth.

NVIDIA's investments in open models span Nemotron (Language and Agentic AI), Cosmos (Physical AI), Isaac GR00T (Robotics), Alpamayo (Autonomous Vehicles), and more.

On September 3rd, NVIDIA announced its intent to acquire Hugging Face. Together, we will scale Hugging Face's platform, strengthen its infrastructure and expand access to AI for developers and institutions worldwide.

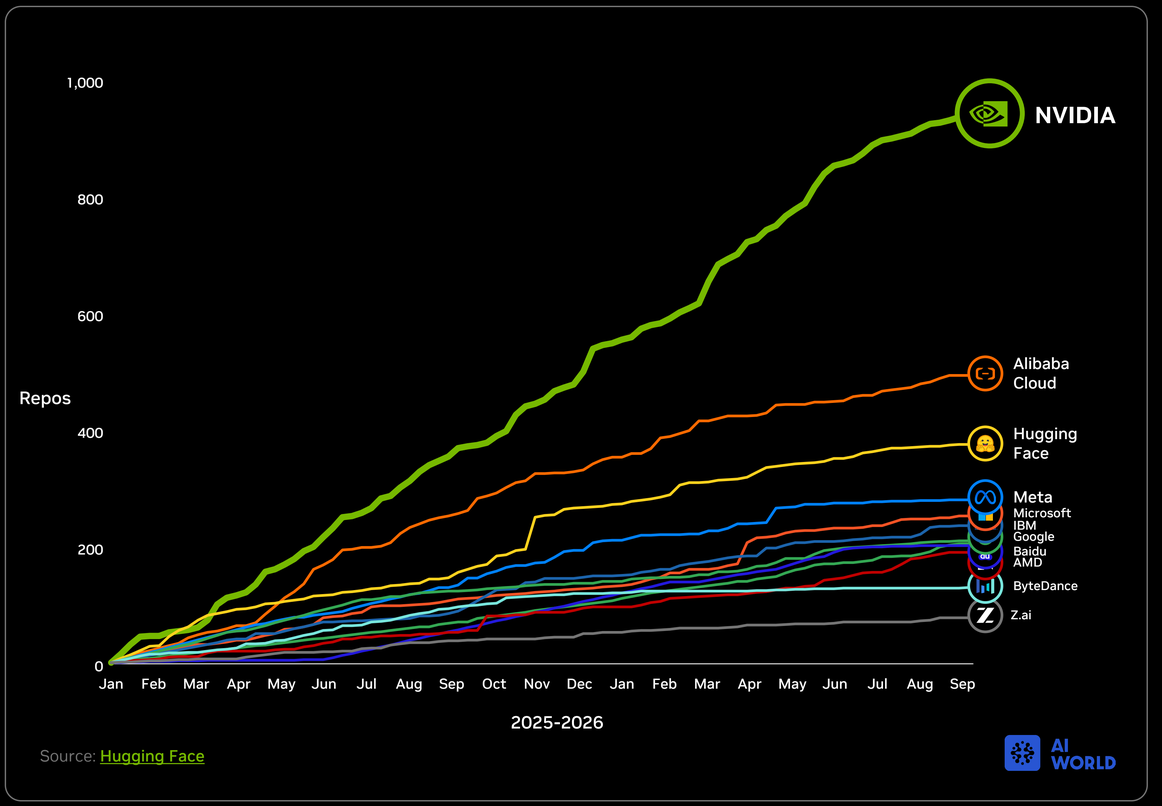
Omniverse
Libraries & microservices for developing Physical AI applications

Cosmos 3
Open world foundation model

Isaac GR00T
Open humanoid-robot reference platform

BioNeMo
Proteina-Complexa frontier generative protein model

Alpamayo
The world's first reasoning autonomous vehicle model



NVIDIA Nemotron at the Frontier

Nemotron models are the most open, post-trainable, and efficient U.S. open-weight models on the Artificial Analysis Intelligence Index.

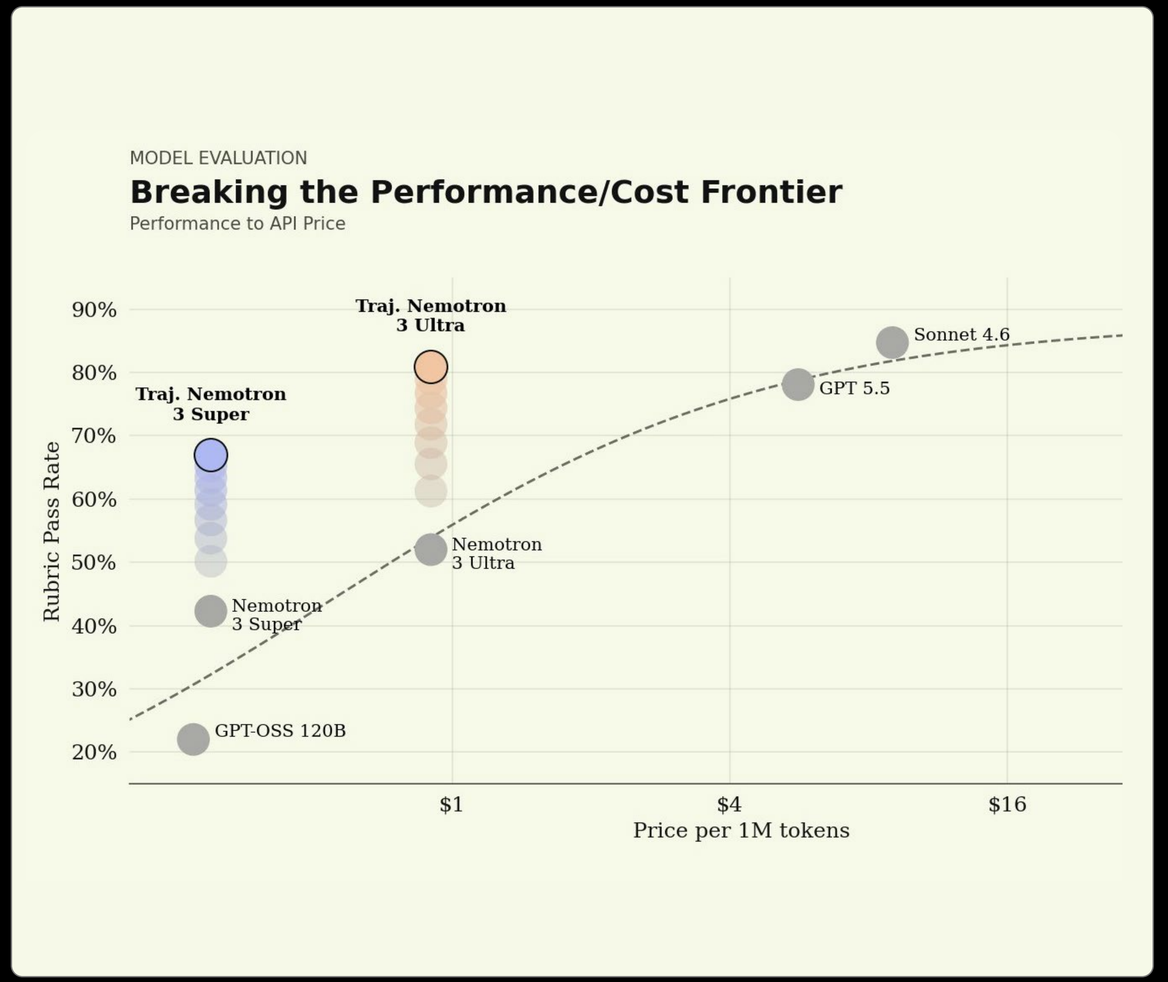
Delivers intelligence at ~5X faster output speed and ~30% lower cost to task completion than the best alternative open model.

NVIDIA releases weights, training data, and recipes, so enterprises, governments, and startups can specialize models on their own data and own the result.

Nemotron adoption is broad and accelerating across Enterprise SaaS, AI Natives/ Startups, and Sovereign.

4 out of 5 top Enterprise SaaS Platforms have deployed Nemotron models.

AI Natives and post-training partners are fine-tuning Nemotron across industries and domains.



Future Arriving

Space

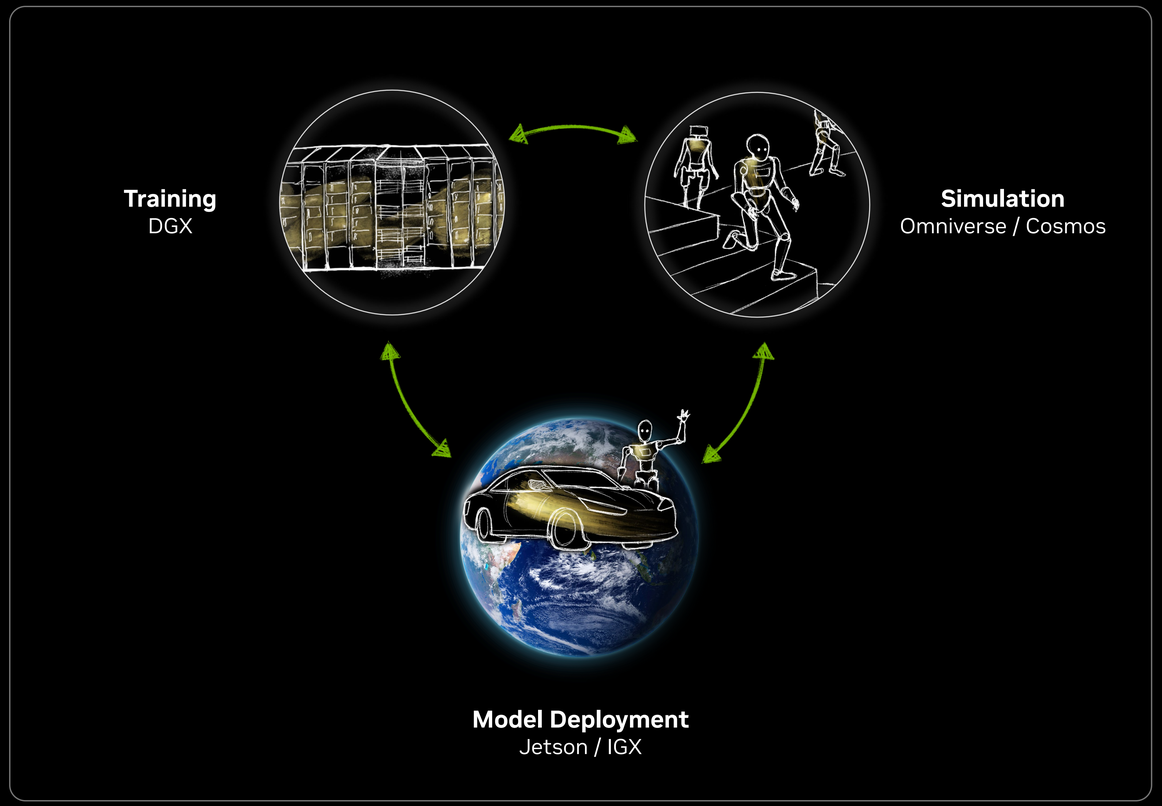
Robotaxis

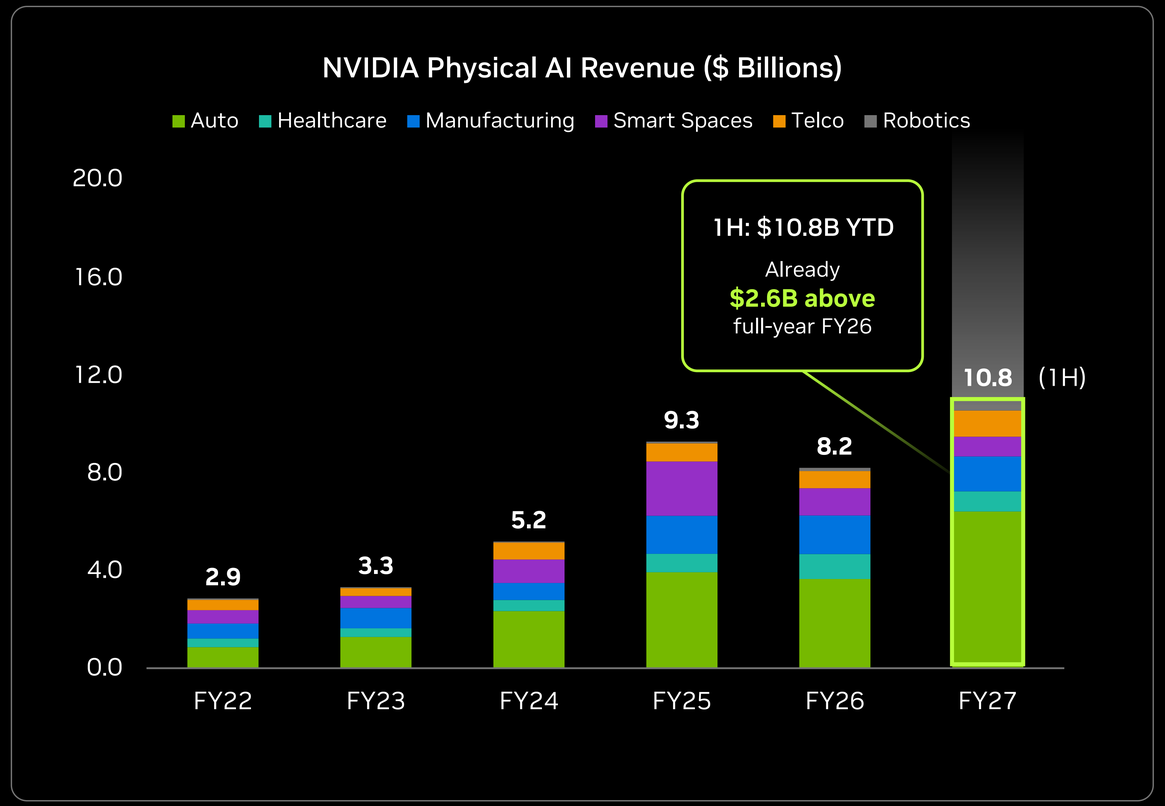
Robotics

Telco

Physical AI Adds a New Workload on AI Factories

Physical AI extends AI into the physical economy – manufacturing, robotics, autonomous vehicles and heavy industry.
It addresses a $50 trillion market and represents a multi-trillion dollar long-term opportunity for NVIDIA.
Only NVIDIA delivers all three computers necessary to build physical AI – Train (DGX), Simulate (Omniverse + Cosmos), and Run (Jetson Thor).





NVIDIA: A Platform for Compounding Growth and Value

NVIDIA is a platform that delivers compounding growth and value to investors.

CUDA programmability underpins the productivity, fungibility, and durability of NVIDIA compute. Our platform addresses every model, workload, and deployment. As a result, **NVIDIA is increasing share within an expanding AI market.**

NVIDIA is also capturing more value per GW driven by advancements in the GPU and a broadening product portfolio – a reflection of NVIDIA's strategy in becoming a full-stack AI factory platform.

And as the platform broadens from chip to a full-stack AI factory, **NVIDIA captures a growing % of customers' IT CapEx.**

Together, these three growth pillars will drive an increase in NVIDIA's cash generation and capital return to shareholders.


NVIDIA