

Except for the historical information contained herein, certain matters in this presentation including, but not limited to, statements as to: our financial position; our markets, market opportunity, demand and growth drivers; the benefits, impact, performance, features and availability of our products and technologies; the benefits, impact, features and timing of our collaborations or partnerships; NVIDIA accelerated computing being broadly recognized as the way to advance computing as Moore's law ends; data centers making a platform shift from general purpose to accelerated computing; trillion dollars of installed global data center infrastructure transitioning to accelerated computing; Al driving a platform shift in computing new, never-before-possible applications; broader enterprises driving the next wave of computing, followed by autonomous machines and industrial digitalization; accelerated computing being needed to tackle the most impactful opportunities of our time; NVIDIA's value to every stakeholder in the ecosystem; the ROI of high compute performance; enterprise as the next big generative Al opportunity; NVIDIA's expanding accelerated computing ecosystem; Al as the greatest technology force of our time; data centers becoming Al factories; generative Al unlocking new opportunities; the next wave of Al being robotics and industrial digitalization; NVIDIA's acceleration stacks and ecosystems helping to bring Al to the world's largest industries; NVIDIA's Al expertise and scale helping to revolutionize businesses; generative Al being the most important computing platform of our generation; full-stack and data center scale acceleration driving significant cost savings and workload scaling; our dividend program plan; and our Automotive design win pipeline and ramp expectations are forward-looking statements.

These forward-looking statements and any other forward-looking statements that go beyond historical facts that are made in this presentation are subject to risks and uncertainties that may cause actual results to differ materially. Important factors that could cause actual results to differ materially include: global economic conditions; our reliance on third parties to manufacture, assemble, package and test our products; the impact of technological development and competition; development of new products and technologies or enhancements to our existing product and technologies; market acceptance of our products or our partners' products; design, manufacturing or software defects; changes in consumer preferences and demands; changes in industry standards and interfaces; unexpected loss of performance of our products or technologies when integrated into systems and other factors.

NVIDIA has based these forward-looking statements largely on its current expectations and projections about future events and trends that it believes may affect its financial condition, results of operations, business strategy, short-term and long-term business operations and objectives, and financial needs. These forward-looking statements are subject to a number of risks and uncertainties, and you should not rely upon the forward-looking statements as predictions of future events. The future events and trends discussed in this presentation may not occur and actual results could differ materially and adversely from those anticipated or implied in the forward-looking statements. Although NVIDIA believes that the expectations reflected in the forward-looking statements are reasonable, the company cannot guarantee that future results, levels of activity, performance, achievements or events and circumstances reflected in the forward-looking statements will occur. Except as required by law, NVIDIA disclaims any obligation to update these forward-looking statements to reflect future events or circumstances. For a complete discussion of factors that could materially affect our financial results and operations, please refer to the reports we file from time to time with the SEC, including our most recent Annual Report on Form 10-K, Quarterly Reports on Form 10-Q, and Current Reports on Form 8-K. Copies of reports we file with the SEC are posted on our website and are available from NVIDIA without charge.

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 non-GAAP gross profit, non-GAAP gross margin, non-GAAP operating income, non-GAAP operating margin, and 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 pioneered accelerated computing to help solve impactful challenges classical computers cannot. A quarter of a century in the making, NVIDIA accelerated computing is broadly recognized as the way to advance computing as Moore's law ends and AI lifts off.

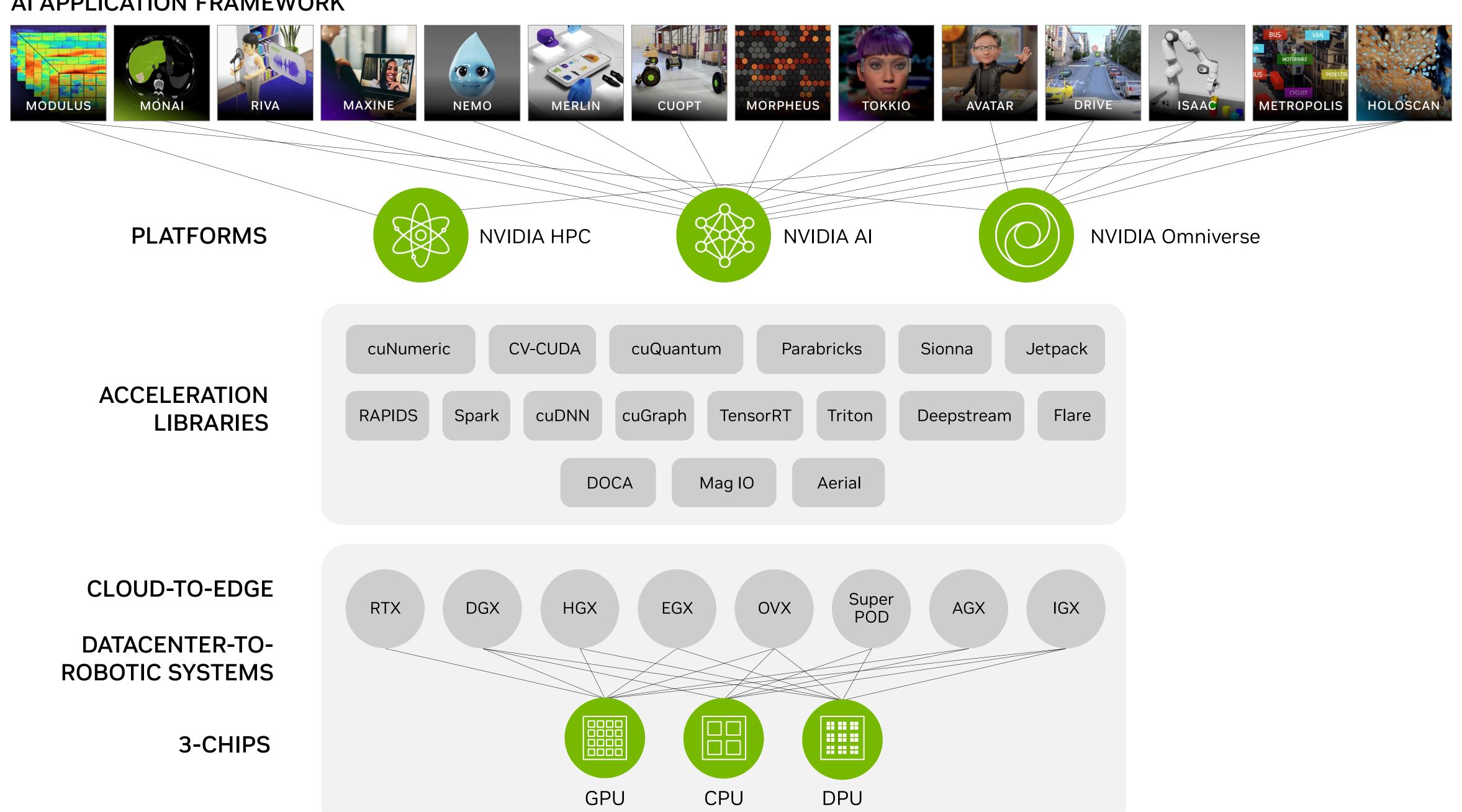
NVIDIA's platform is installed in several hundred million computers, is available in every cloud and from every server maker, powers 74% of the TOP500 supercomputers, and boasts over 4 million developers.



# NVIDIA's Accelerated Computing Platform

Full-stack innovation across silicon, systems and software

### AI APPLICATION FRAMEWORK



With nearly three decades of a singular focus, NVIDIA is expert at accelerating software and scaling compute by a Million-X, going well beyond Moore's law.

Accelerated computing requires full-stack innovation – optimizing across every layer of computing - from silicon and systems to software and algorithms, demanding deep understanding of the problem domain.

Our full-stack platforms - NVIDIA HPC, NVIDIA AI, and NVIDIA Omniverse accelerate high performance computing, Al, and industrial digitalization workloads.

We accelerate workloads at data center scale, across thousands of compute nodes, treating the network and storage as part of the computing fabric.

Our platform extends from the cloud and enterprise data centers to supercomputing centers, edge computing and PCs.

# Why Accelerated Computing?

Advancing computing in the post-Moore's Law era

Accelerated computing is needed to tackle the most impactful opportunities of our time—like AI, climate simulation, drug discovery, ray tracing, and robotics.

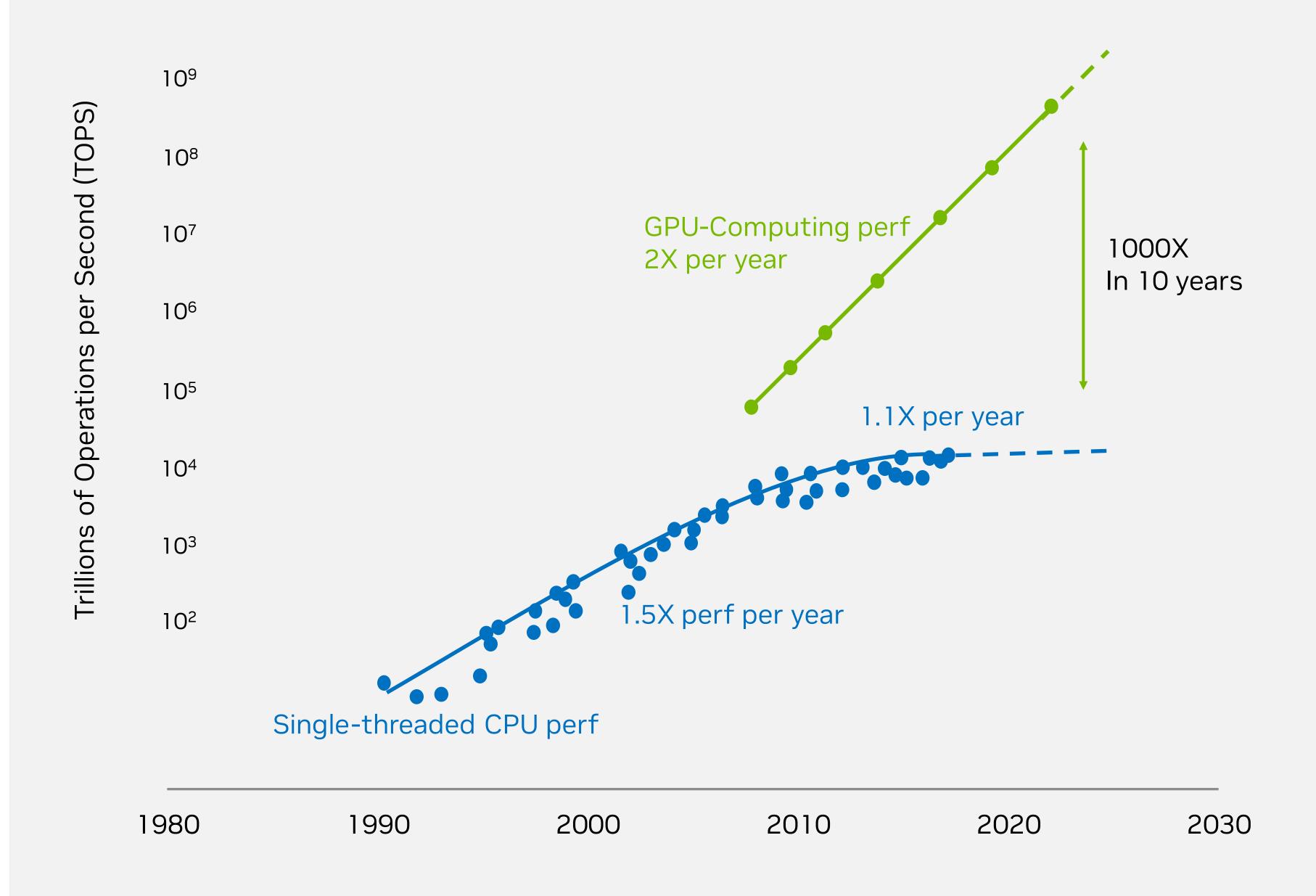
NVIDIA is uniquely dedicated to accelerated computing —working top-to-bottom—refactoring applications and creating new algorithms, and bottom-to-top—inventing new specialized processors, like RT Core and Tensor Core.

"It's the end of Moore's Law as we know it."

- John Hennessy, Oct 2018

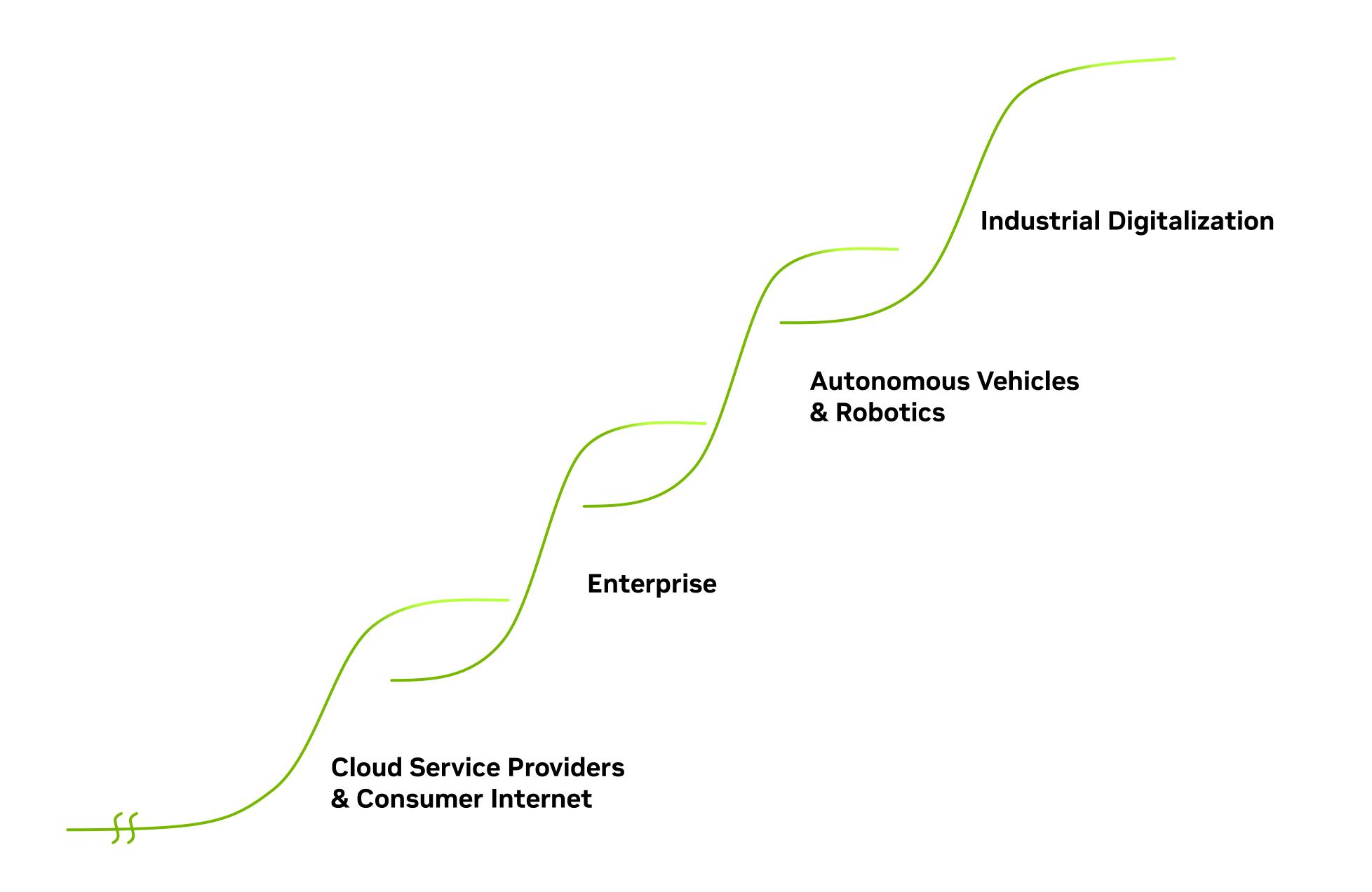
"Moore's Law is dead."

- Jensen Huang, GTC 2013



# Waves of Adoption of Accelerated Computing

A generational computing platform shift



A new computing era has begun.

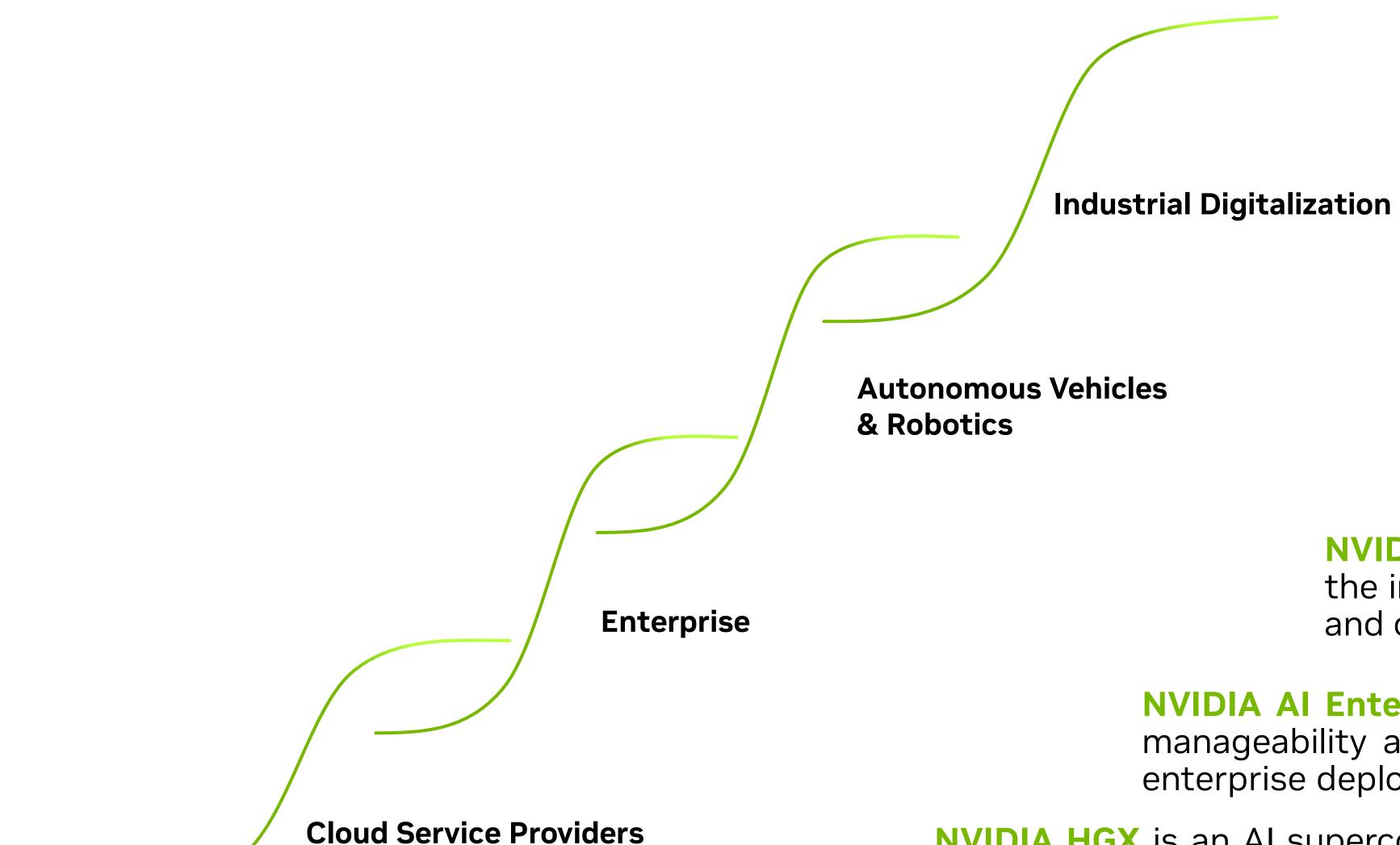
Accelerated computing enabled the rise of AI, which is driving a platform shift from general purpose to accelerated computing, and enabling new, never-before-possible applications.

The trillion dollars of installed global data center infrastructure will transition to accelerated computing to achieve an order of magnitude better performance, energy-efficiency and cost.

Hyperscale cloud service providers and consumer internet companies have been the early adopters of Al and accelerated computing, with broader enterprise adoption now under way.

Al and accelerated computing will also make possible the next big waves – autonomous machines and industrial digitalization.

# **NVIDIA Accelerated Computing for Every Wave**



& Consumer Internet

**NVIDIA Omniverse** is a software platform for designing, building, and operating 3D and virtual world simulations. It harnesses the power of NVIDIA graphics and AI technologies and runs on NVIDIA-powered data centers and workstations.

**NVIDIA DRIVE** is a full-stack platform for autonomous vehicles (AV) that includes hardware for in-car compute, such as the Orin system-on-chip, and the full AV and AI cockpit software stack.

**NVIDIA DGX Cloud** is a cloud service that allows enterprises immediate access to the infrastructure and software needed to train advanced models for generative Al and other groundbreaking applications.

**NVIDIA AI Enterprise** is the operating system of AI, with enterprise-grade security, stability, manageability and support. It is available on all major CSPs and server OEMs and supports enterprise deployment of AI in production.

**NVIDIA HGX** is an AI supercomputing platform purpose-built for AI. It includes 8 NVIDIA GPUs, as well as interconnect and networking technologies, delivering order-of-magnitude performance speed-ups for AI over CPU servers. It is broadly available from all major server OEMs/ODMs. **NVIDIA DGX**, an AI server based on the same architecture, along with NVIDIA AI software and support, is also available.

# NVIDIA's Accelerated Computing Ecosystem

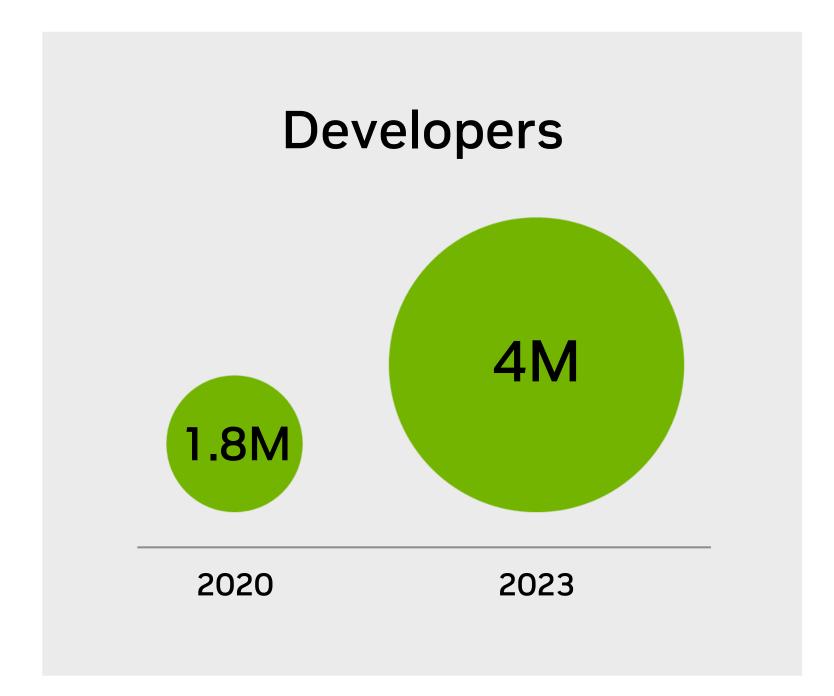
The NVIDIA accelerated computing platform has attracted the largest ecosystem of developers, supporting a rapidly growing universe of applications and industry innovation.

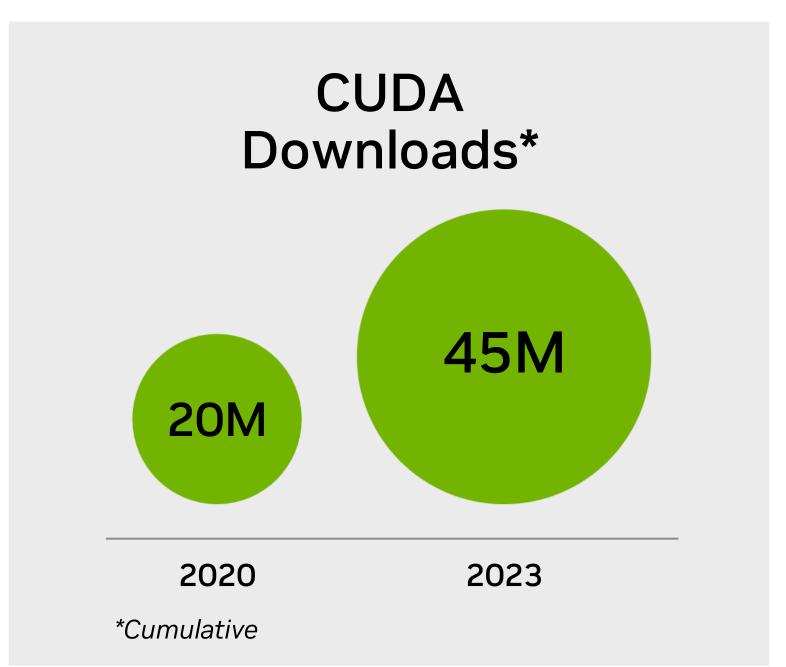
**CUDA** – our parallel computing programming model introduced in 2006 – or at higher layers of the stack, including libraries, pre-trained Al models, SDKs and other development tools.

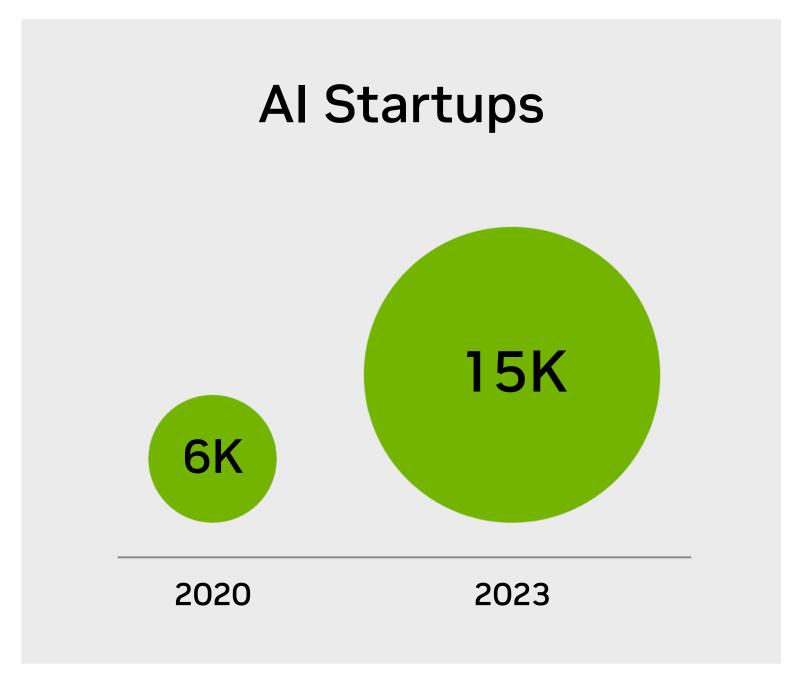
300 Libraries

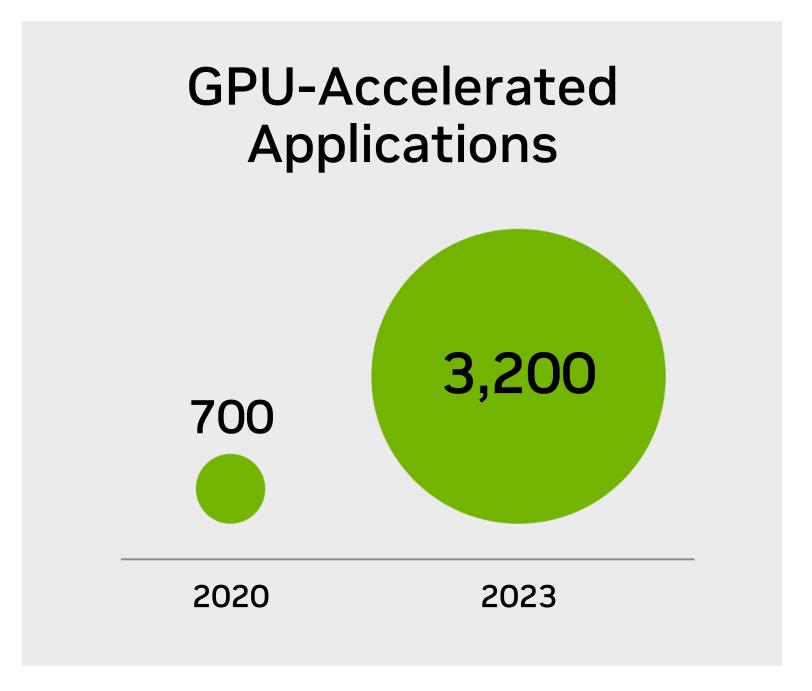
400 Al Models

100 Updated in the Last Year

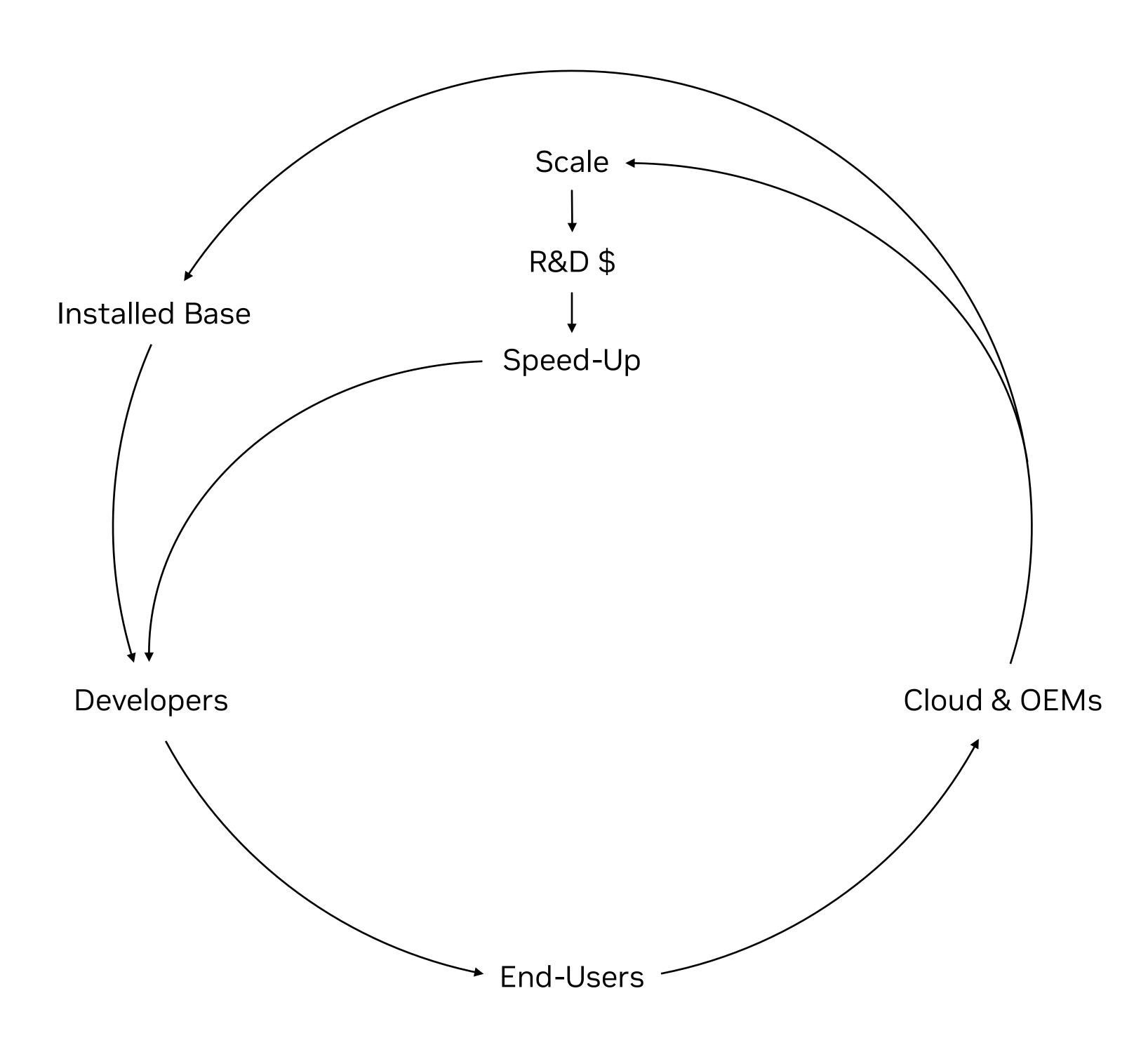








# NVIDIA's Multi-Sided Platform and Flywheel



NVIDIA Accelerated Computing Virtuous Cycle

The virtuous cycle of NVIDIA's accelerated computing starts with an installed base of several hundred million GPUs, all compatible with the CUDA programming model.

- For developers NVIDIA's one architecture and large installed base give developer's software the best performance and greatest reach
- For end users NVIDIA is offered by virtually every computing provider and accelerates the most impactful applications from cloud to edge
- For cloud providers and OEMs NVIDIA's rich suite of Acceleration Platforms lets partners build one offering to address large markets including media & entertainment, healthcare, transportation, energy, financial services, manufacturing, retail, and more
- For NVIDIA Deep engagement with developers, computing providers, and customers in diverse industries enables unmatched expertise, scale, and speed of innovation across the entire accelerated computing stack propelling the flywheel

# Huge ROI from AI Driving a Powerful New Investment Cycle

Al can augment creativity and productivity by orders of magnitude across industries

Knowledge workers will use copilots based on large language models to generate documents, answer questions, or summarize missed meetings, emails and chats – adding hours of productivity per week.

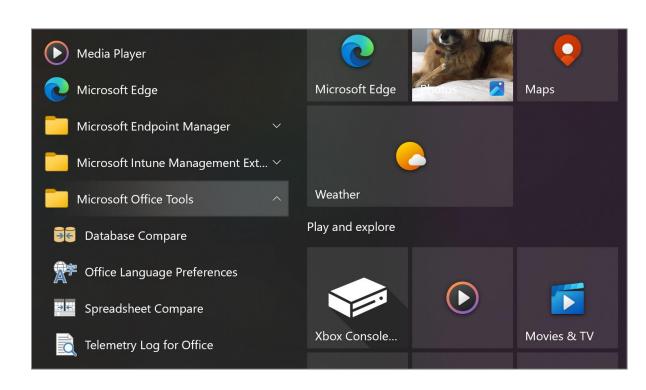
Copilots specialized for fields such as software development, legal services or education can boost productivity by as much as 50%.

Social media, search and e-commerce apps are using deep recommenders to offer more relevant content and ads to their customers, increasing engagement and monetization.

Creators can generate stunning, photorealistic images with a single text prompt – compressing workflows that take days or weeks into minutes in industries from advertising to game development.

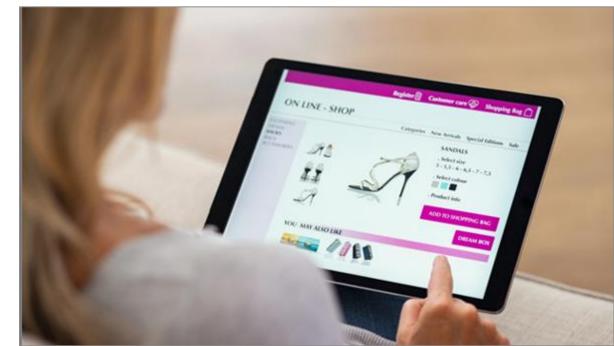
Call center agents augmented with AI chatbots can dramatically increase productivity and customer satisfaction.

Drug discovery, financial services, agriculture and food services and climate forecasting are seeing order-of-magnitude workflow acceleration from Al.

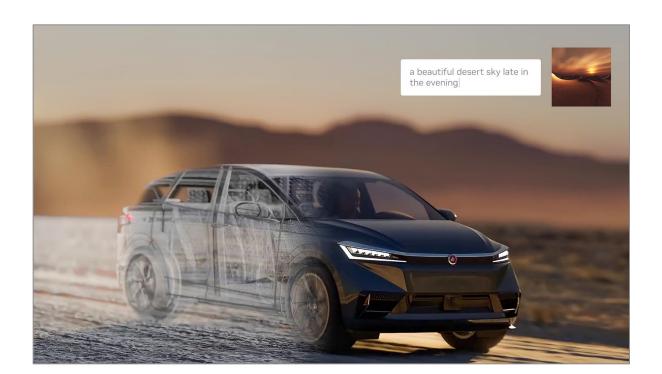


Office Al Copilots

Over 1B knowledge workers



Search & Social Media \$700B in digital advertising annually



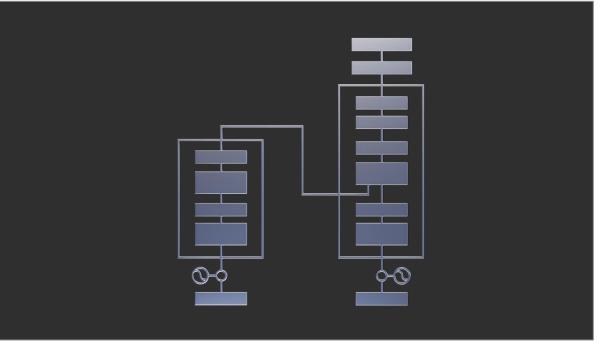
Al Content Creation
50M creators globally



Legal Services, Education ...

1M legal professionals in the US

9M educators in the US



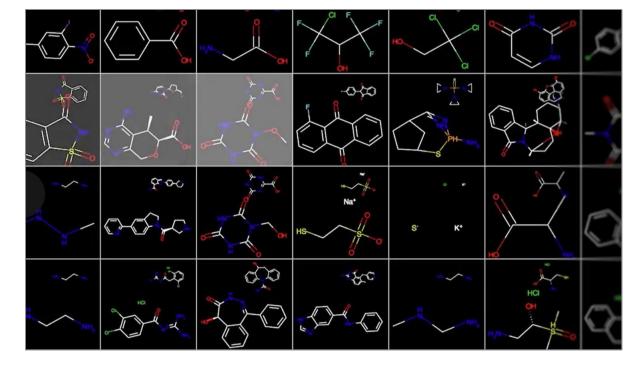
Al Software Development
30M software developers globally



Financial Services
678B annual credit card transactions



Customer Service with Al
15M call center agents globally



Drug Discovery

10<sup>18</sup> molecules in chemical space
40 exabytes of genome data



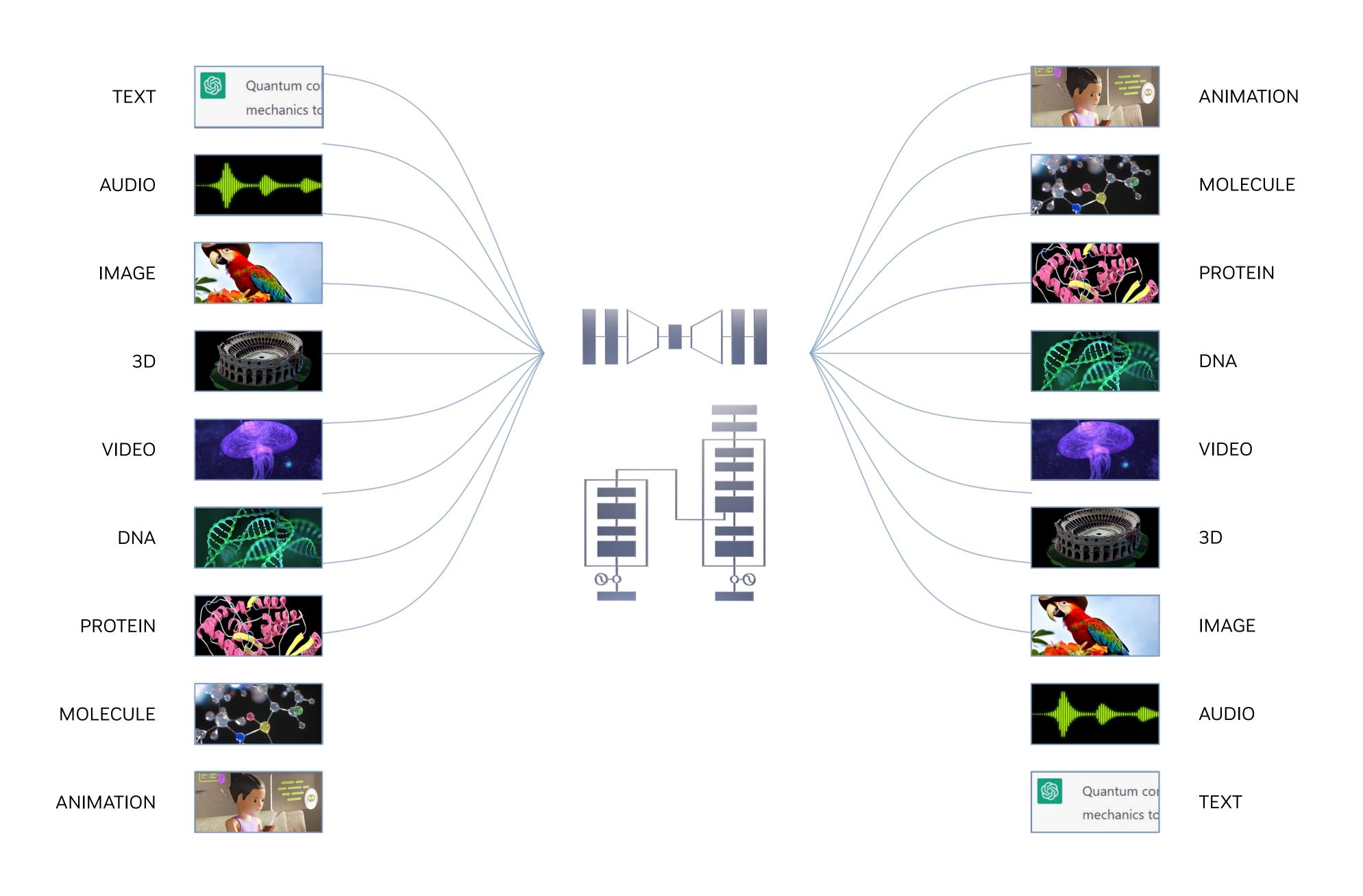
Agri-Food | Climate

1B people in agri-food worldwide

Earth-2 for km-scale simulation

# Generative Al

### The most important computing platform of our generation



The era of generative AI has arrived, unlocking new opportunities for AI across many different applications.

Generative AI is trained on large amounts of data to find patterns and relationships, learning the representation of almost anything with structure.

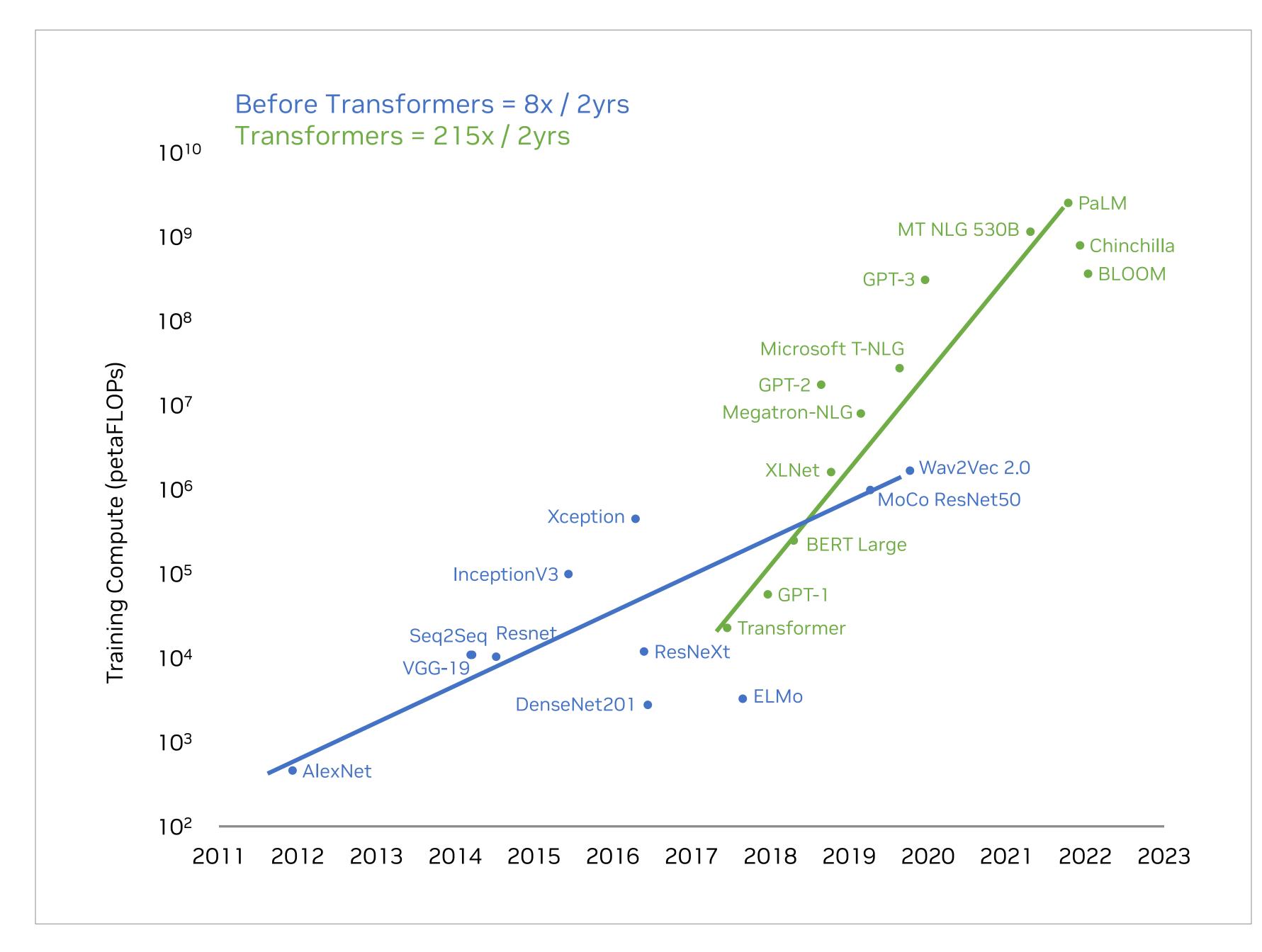
It can then be prompted to generate text, images, video, code, or even proteins.

For the very first time, computers can augment the human ability to generate information and create.

1,600+ Generative AI companies are building on NVIDIA.

# Modern Al is a Data Center Scale Computing Workload

Data centers are becoming AI factories: Data as input, intelligence as output



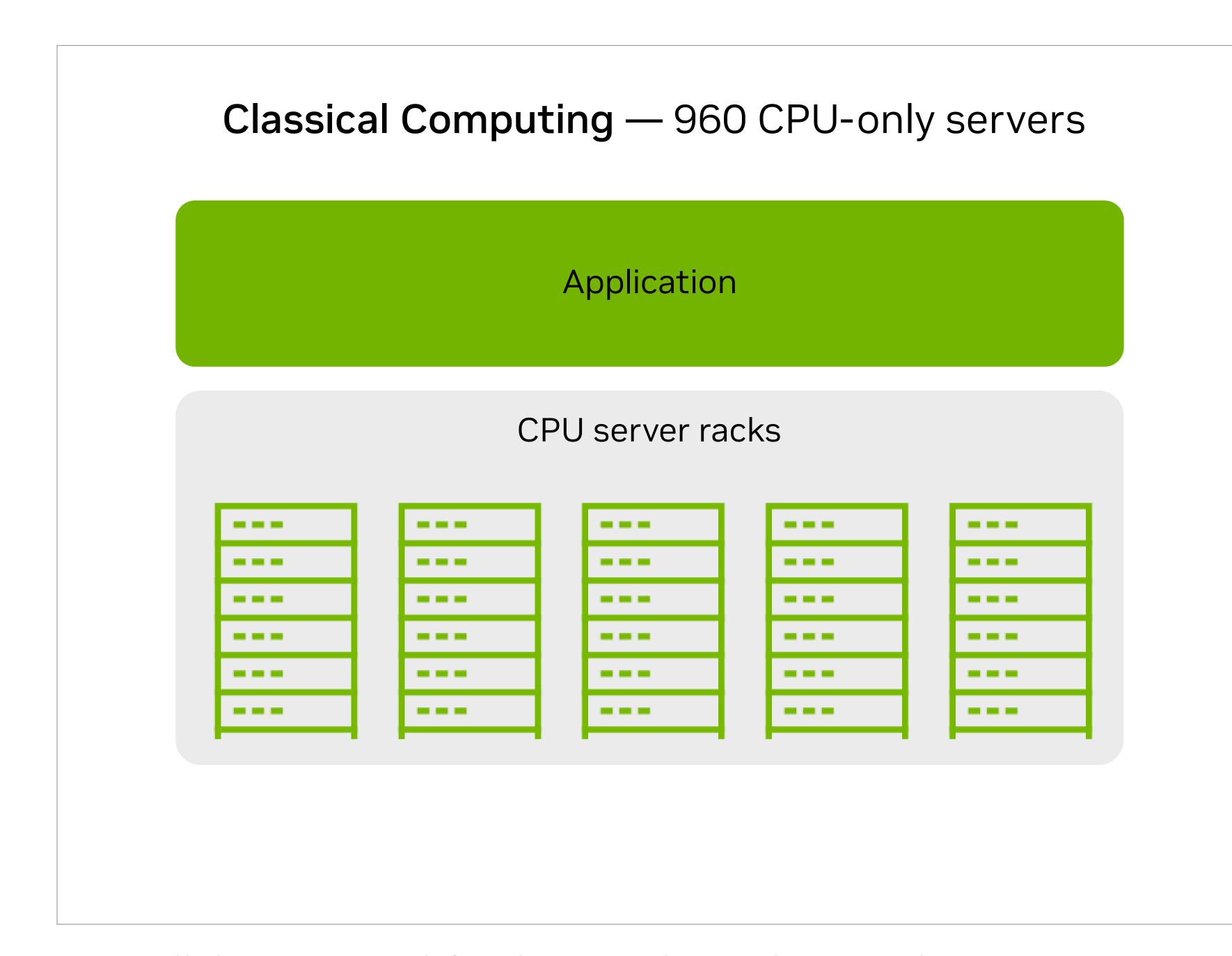
Large Language Models, based on the Transformer architecture, are one of today's most important advanced AI technologies, involving up to trillions of parameters that learn from text. Developing them is an expensive, time-consuming process that demands deep technical expertise, distributed data center-scale infrastructure, and a full-stack accelerated computing approach.

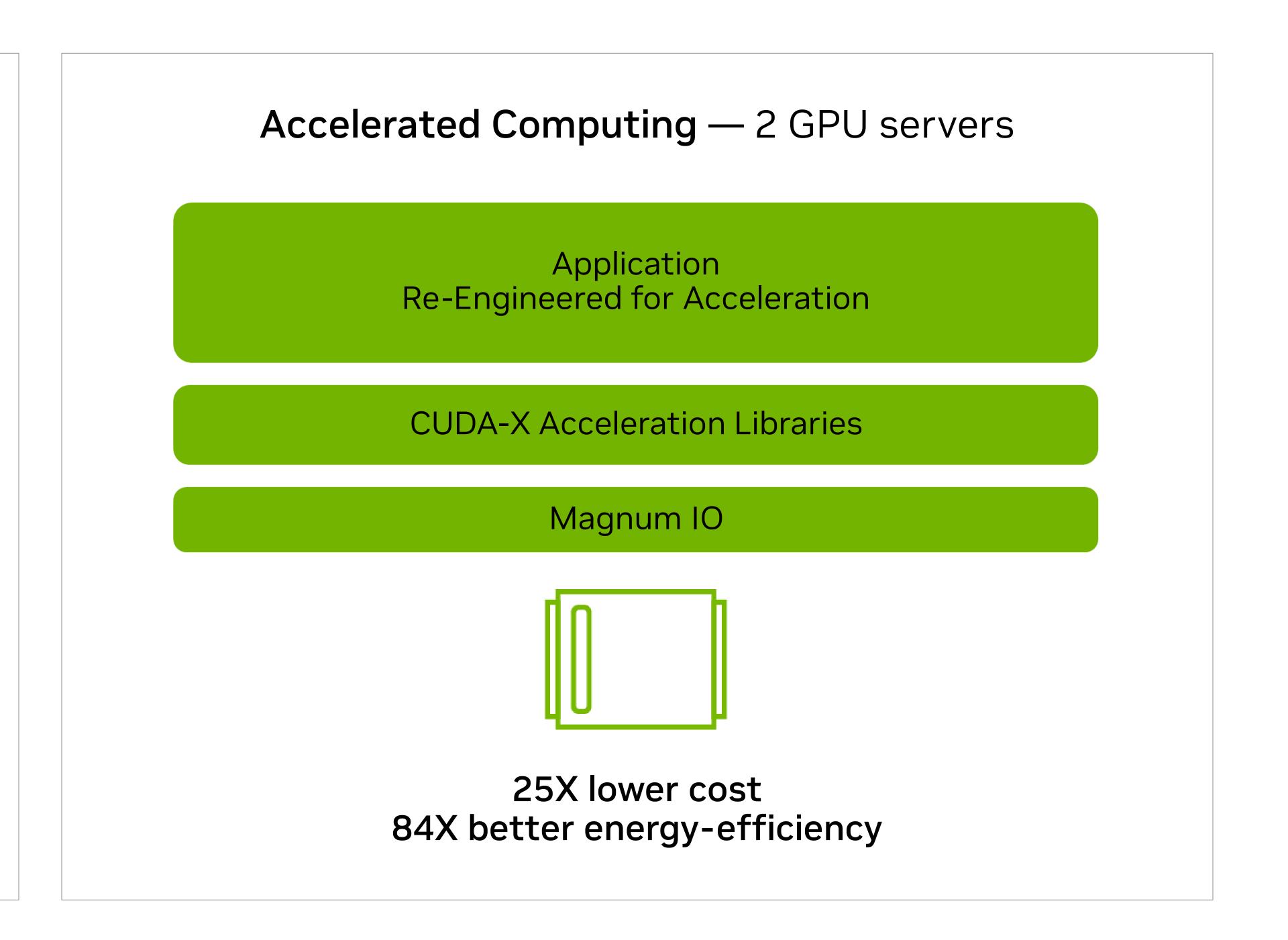
Al Training Computational Requirements

Fueling Giant-Scale Al Infrastructure NVIDIA compute & networking GPU | DPU | CPU

# Full-Stack & Data Center Scale Acceleration

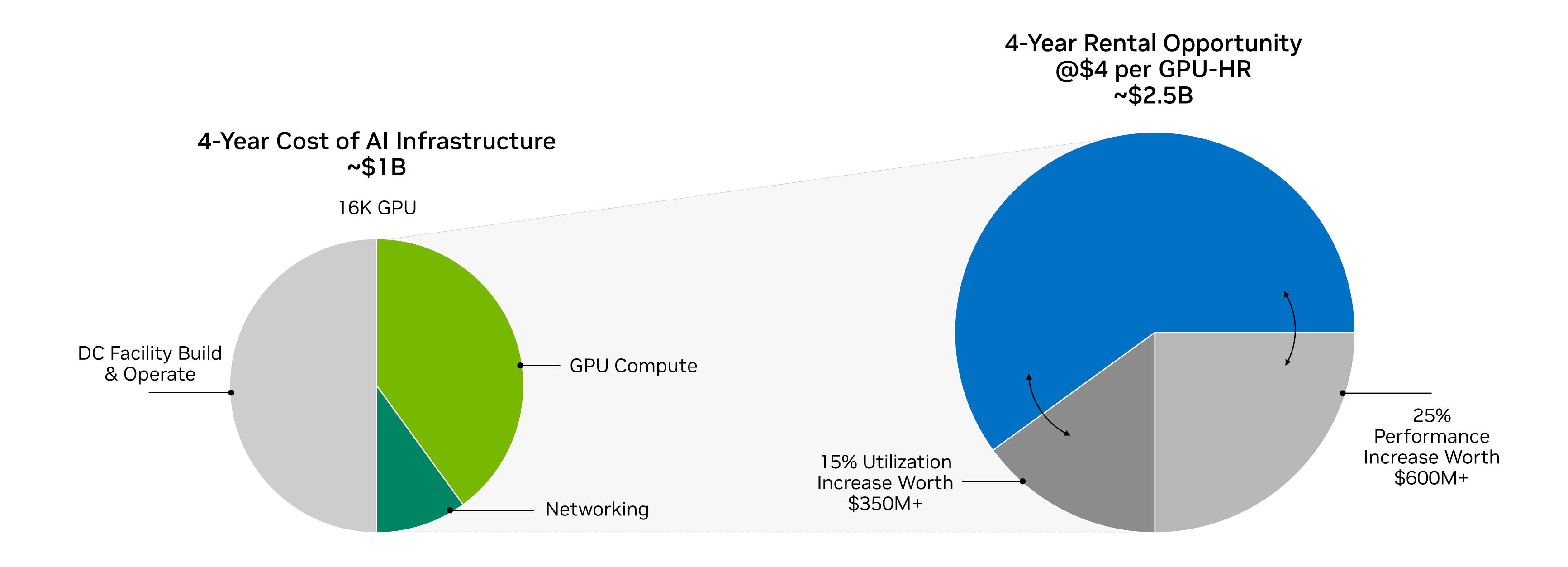
Drive significant cost savings and workload scaling





LLM Workload: Bert-Large Training and Inference | CPU Server: Dual-EYPC 7763 | GPU Server: Dual-EPYC 7763 + 8X H100 PCIe GPUs

# The High ROI of High Compute Performance

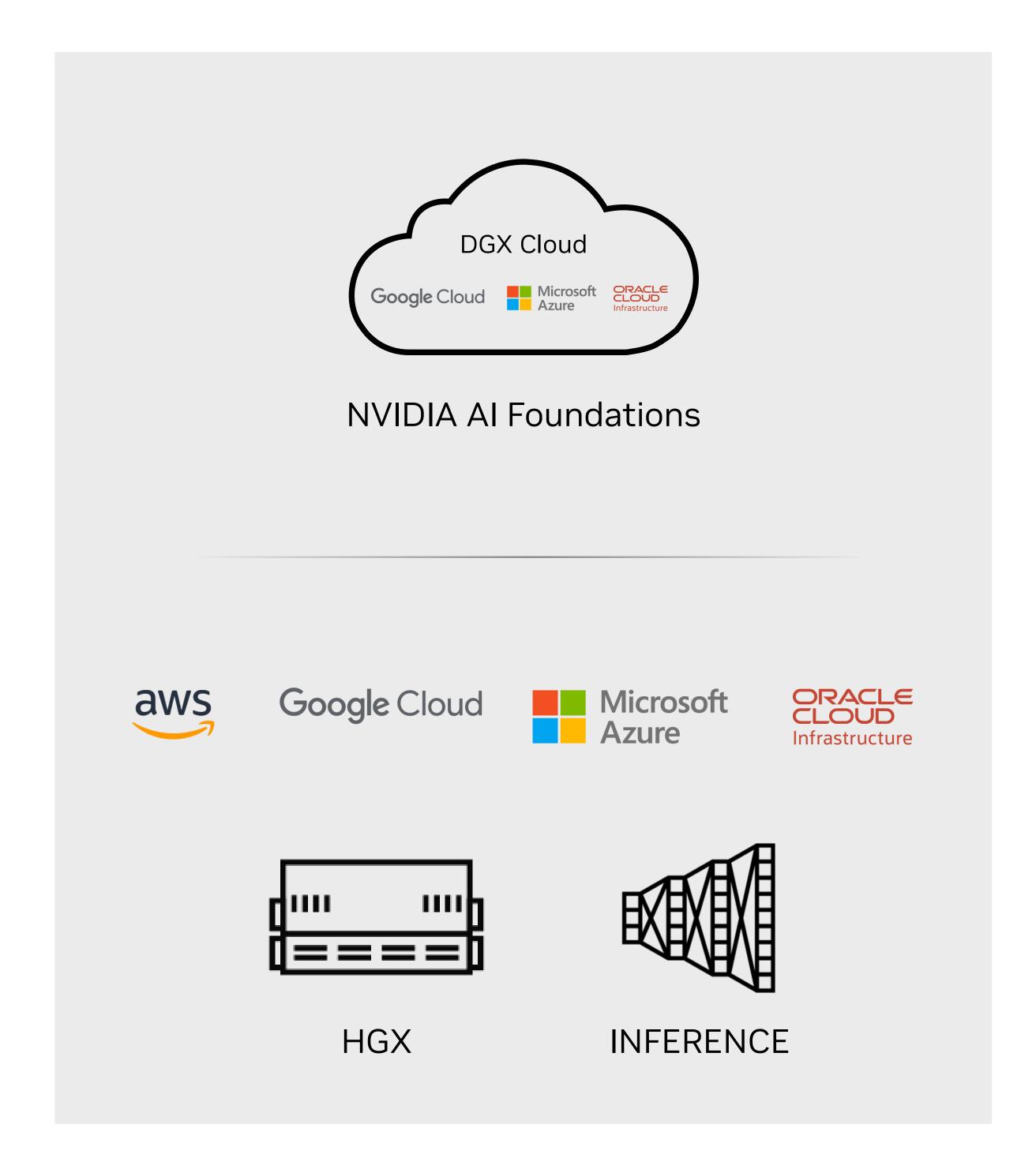


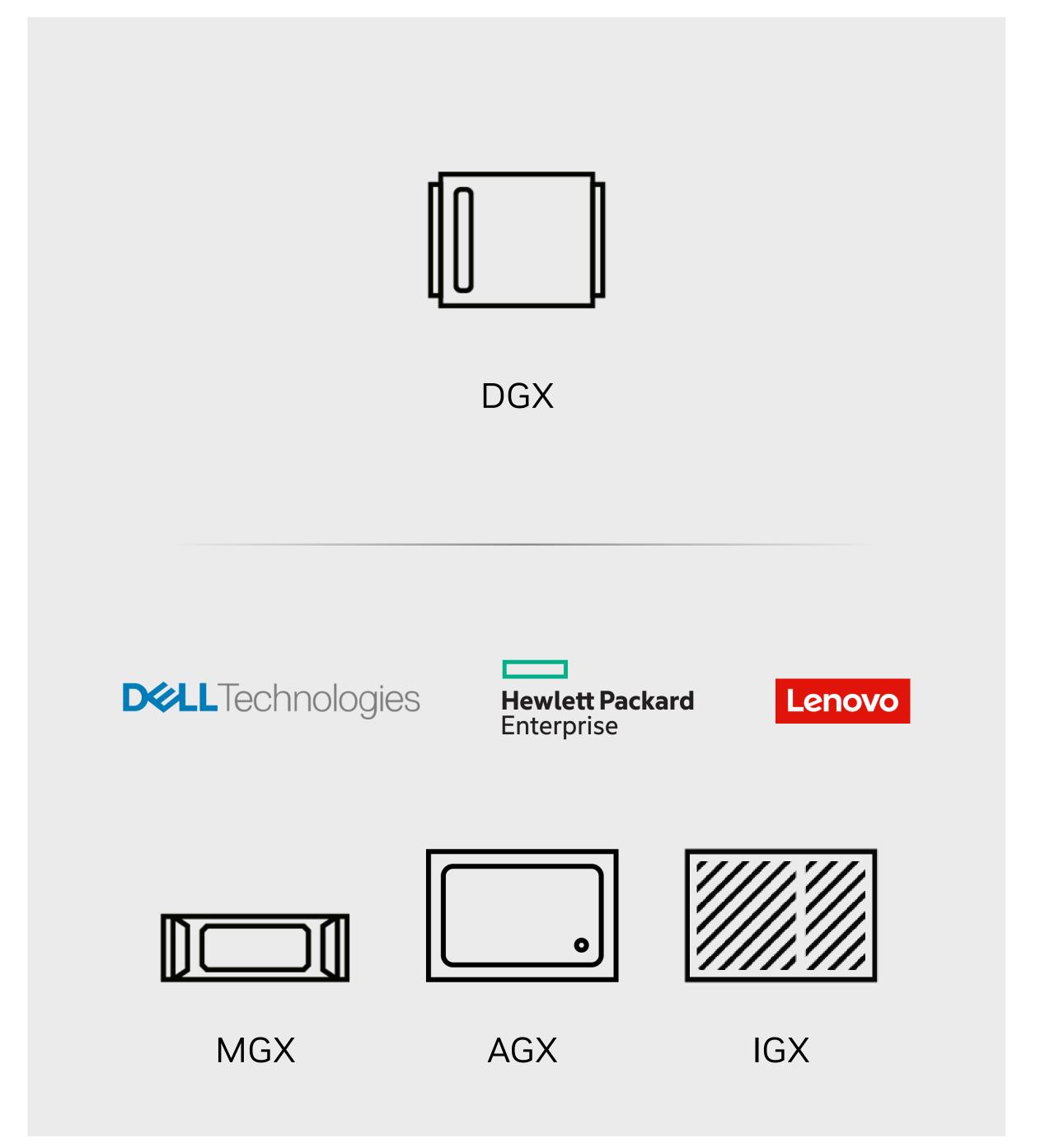
# NVIDIA Go-to-Market Across Cloud and On-Premises

Reaching customers everywhere

CLOUD ON-PREM







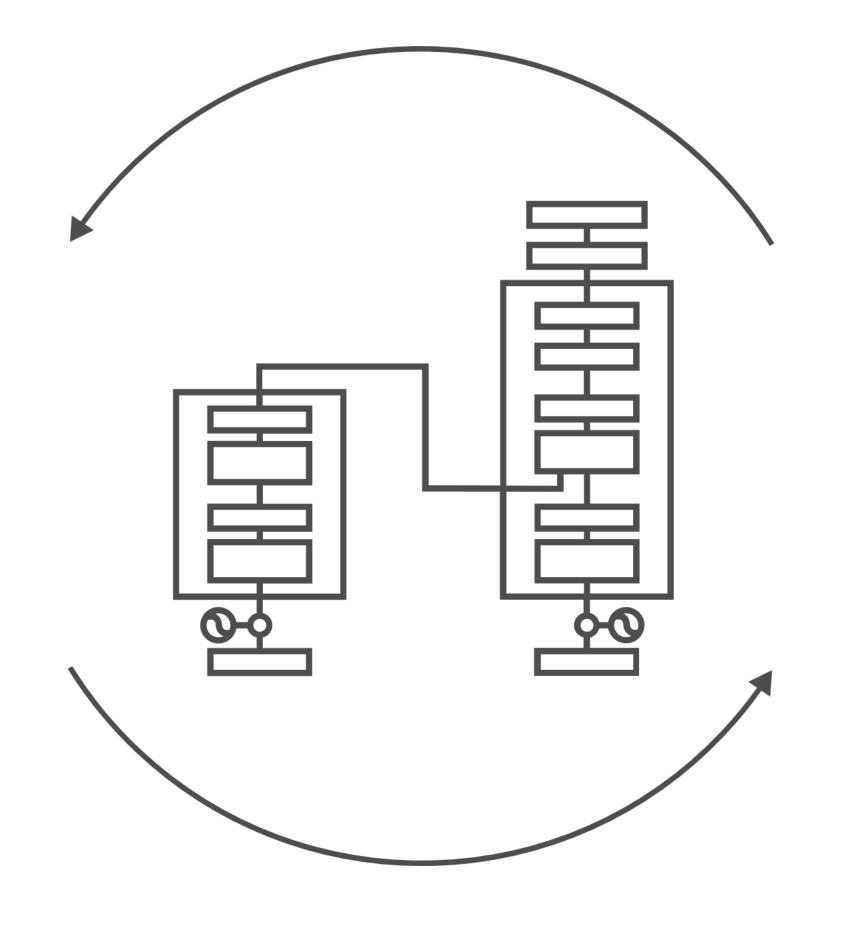
**PARTNERS** 

# Training & Inference — One Architecture

Cloud | On-Prem | Edge

### **TRAINING**





### **INFERENCE**

IN THE DATA CENTER

NVIDIA L40
Image Generation

NVIDIA L4
Al Video

NVIDIA H100 | L40S
Universal GPUs

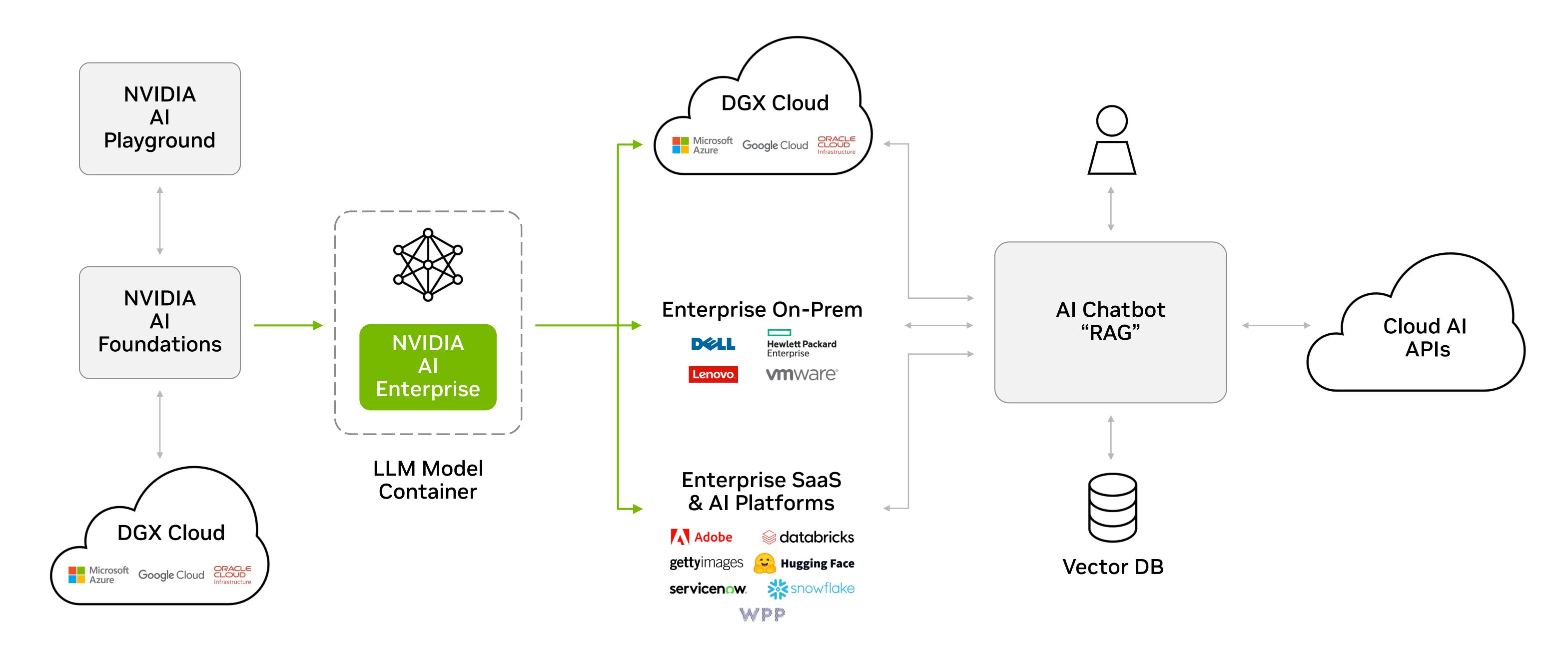
NVIDIA Grace Hopper
RecSys, Gen Al

# IGX Industrial-Grade System for Healthcare, Logistics, Manufacturing AGX Functionally-Safe System for Autonomous Vehicles

AT THE EDGE

# **NVIDIA AI Enterprise**

### Enterprise is the Next Big Generative Al Opportunity



Enterprise AI Chatbots are built as Retrieval Augmented Generation (RAG) workflows, which augment the knowledge in the LLM with vectorized Enterprise data. These Chatbots serve as apprentices, improving the productivity of every employee in every Enterprise company.

NVIDIA delivers this capability to Enterprises by packaging LLMs with NVIDIA AI Enterprise, the runtime for hosting the LLMs, into containers that can be deployed anywhere – on any cloud, on premises, or within Enterprise SaaS applications.

# **NVIDIA DGX Cloud**

Al-training-as-a-service platform for the era of generative Al

# **NVIDIA AI FOUNDATIONS** NeMo | Picasso DGX CLOUD ORACLE CLOUD Microsoft Google Cloud Azure Infrastructure

NVIDIA DGX Cloud is a cloud service that allows enterprises immediate access to the infrastructure and software needed to train advanced models for generative AI and other groundbreaking applications.

DGX Cloud provides dedicated clusters of NVIDIA DGX AI supercomputing, paired with NVIDIA AI software.

Enterprise customers can also use the NVIDIA AI Foundations model making service, which includes NVIDIA NeMo for training custom LLMs and NVIDIA Picasso for custom generative AI models for visual design.

The service is equipped with models, tools, and accelerated computing for training, customizing, optimizing, and deploying Al.

# **NVIDIA AI Enterprise**

## The operating system for enterprise Al

### **NVIDIA AI Enterprise**

NVIDIA AI Enterprise is software for deploying and running AI with enterprise-grade security, API stability, manageability and support.

Cloud-native and available in every major cloud marketplace — AWS, Microsoft Azure, Google Cloud Platform and Oracle Cloud.

Certified to run on servers and workstations from all major OEMs.

### Al Use Cases and Workflows



LLM



Medical Imaging



Speech Al



Video Analytics



Recommenders



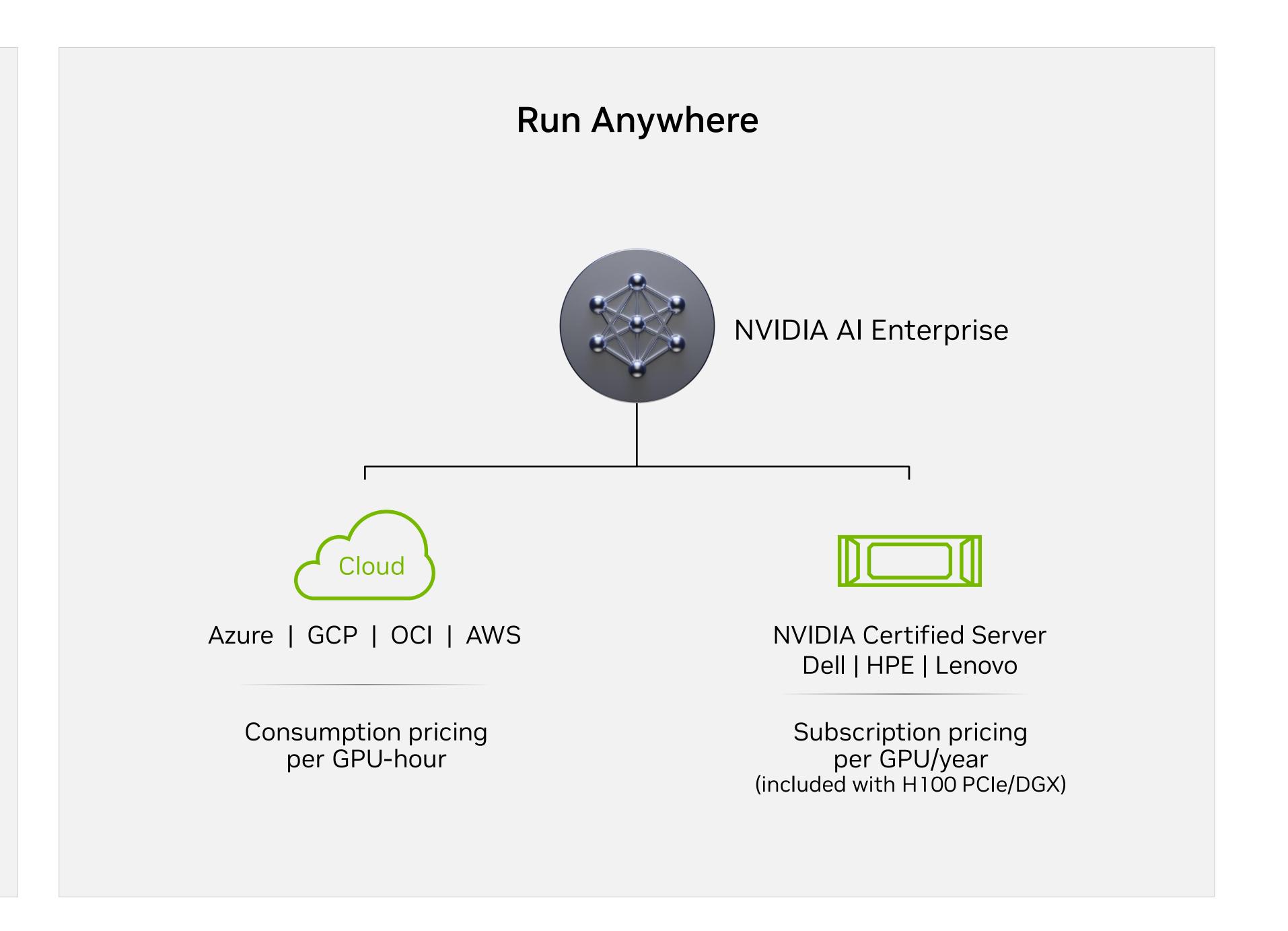
Route Optimization



Cybersecurity

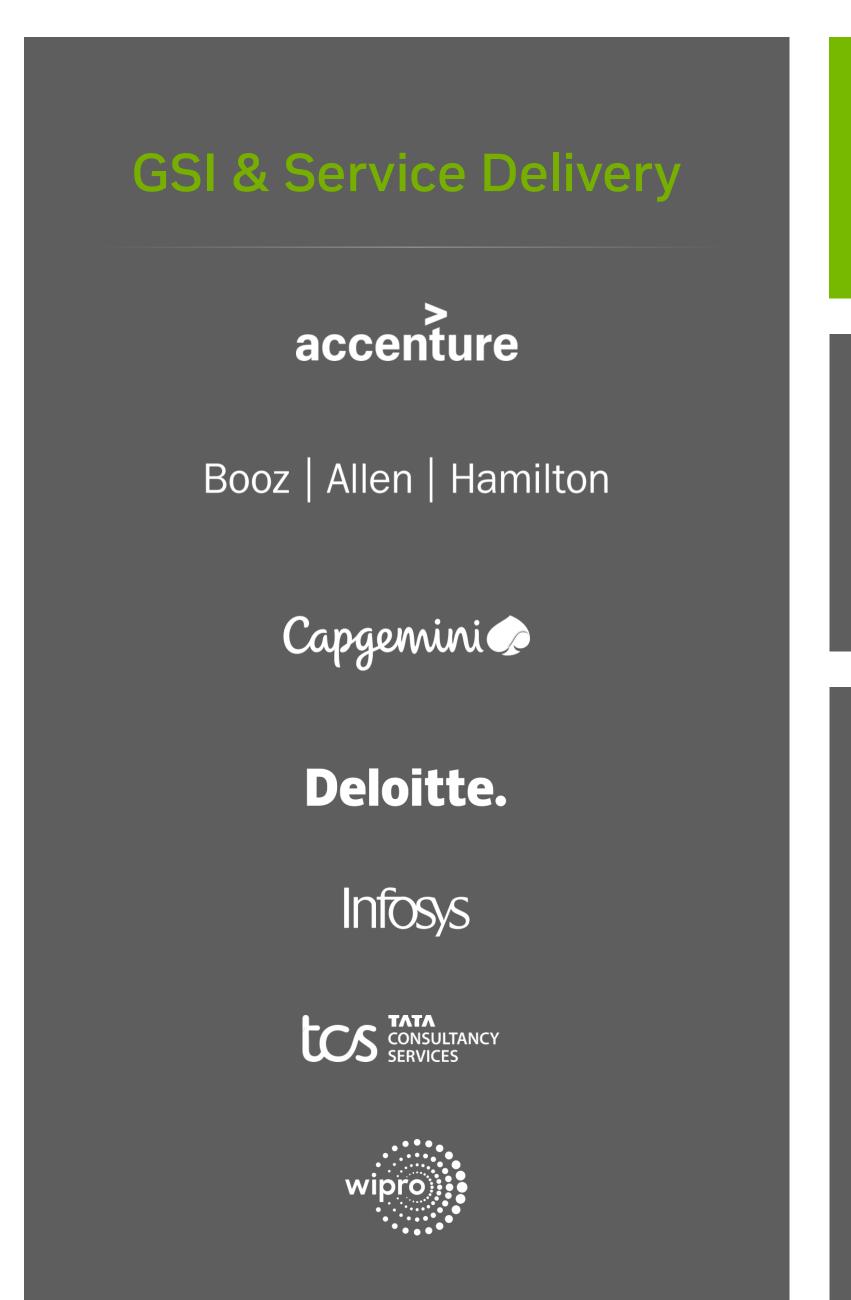
• • •

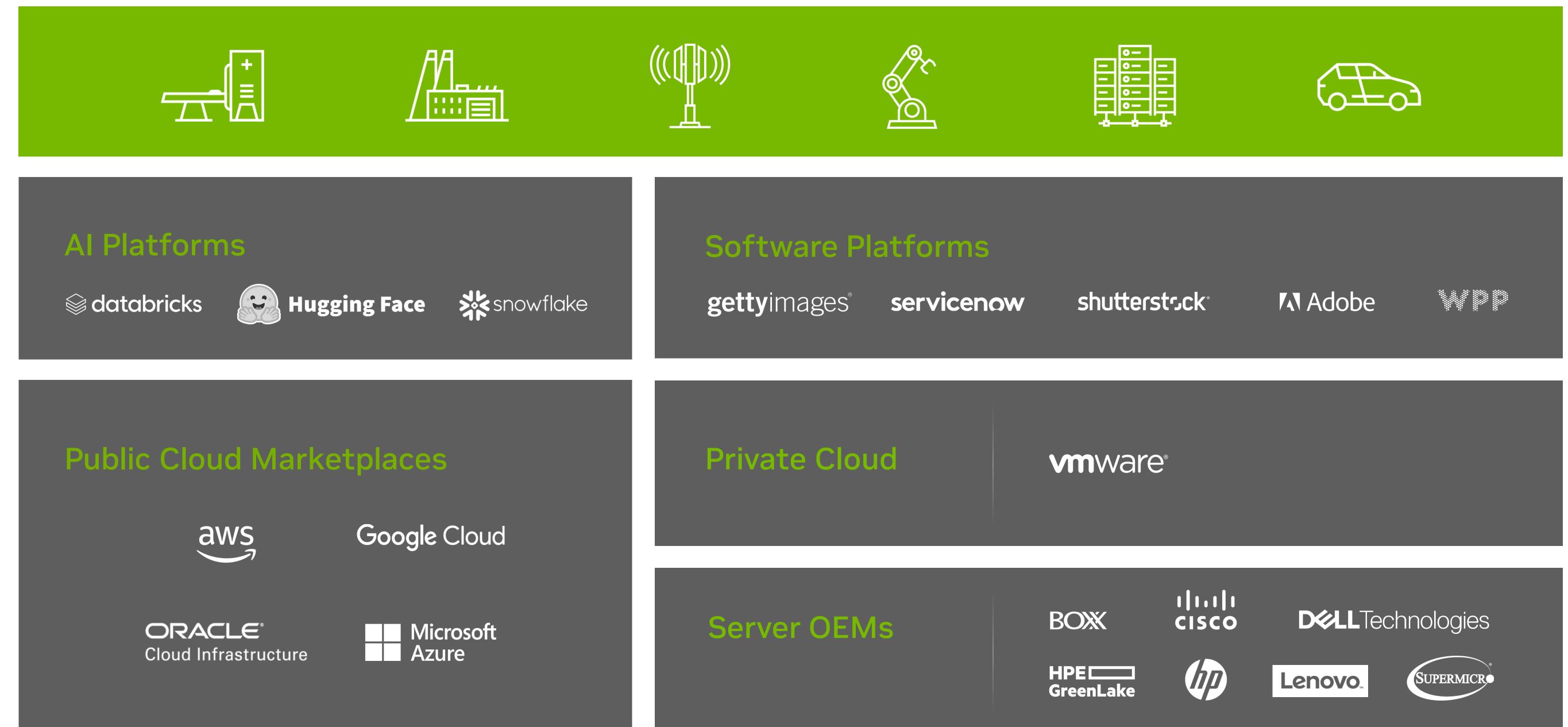
More



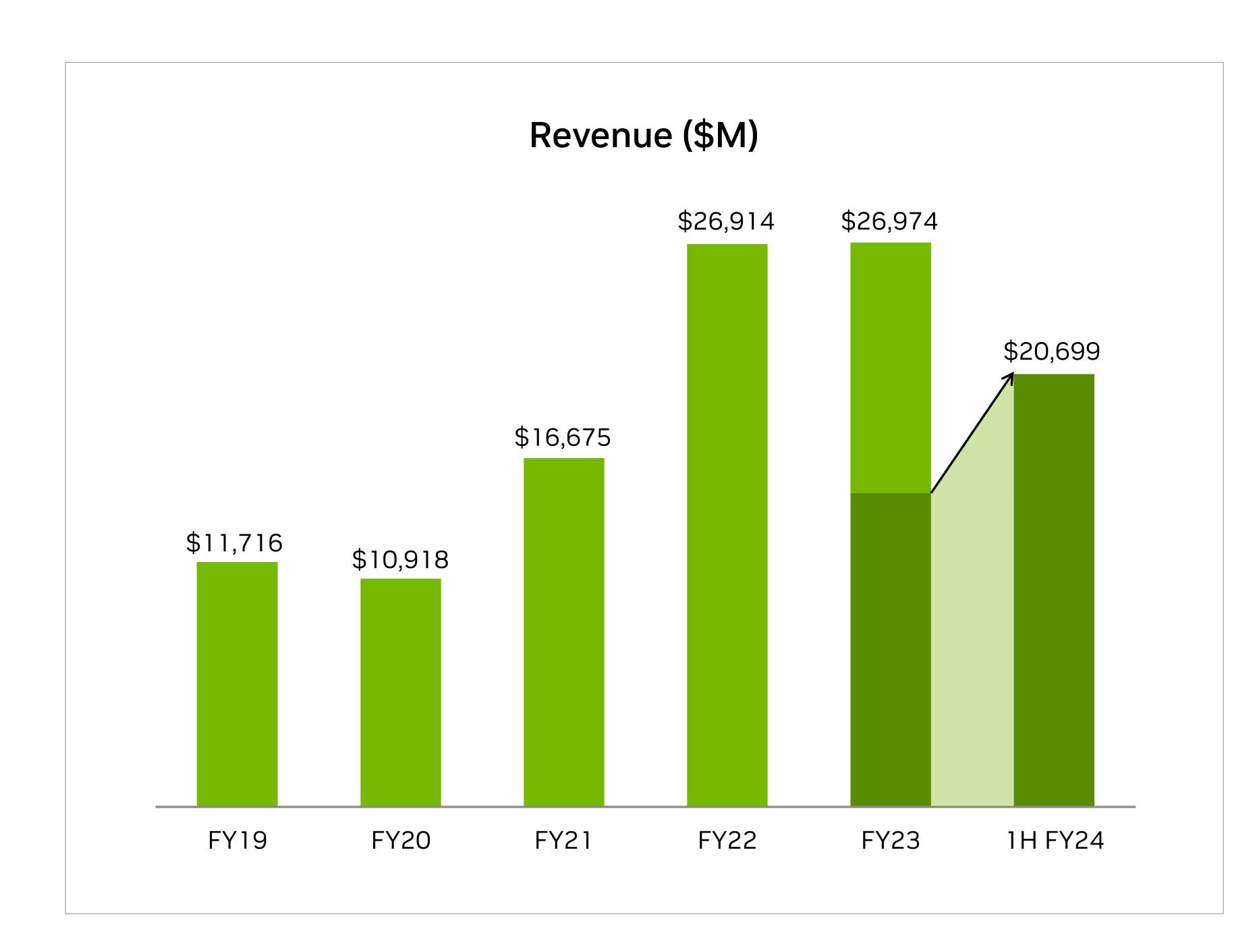
# **NVIDIA AI Enterprise**

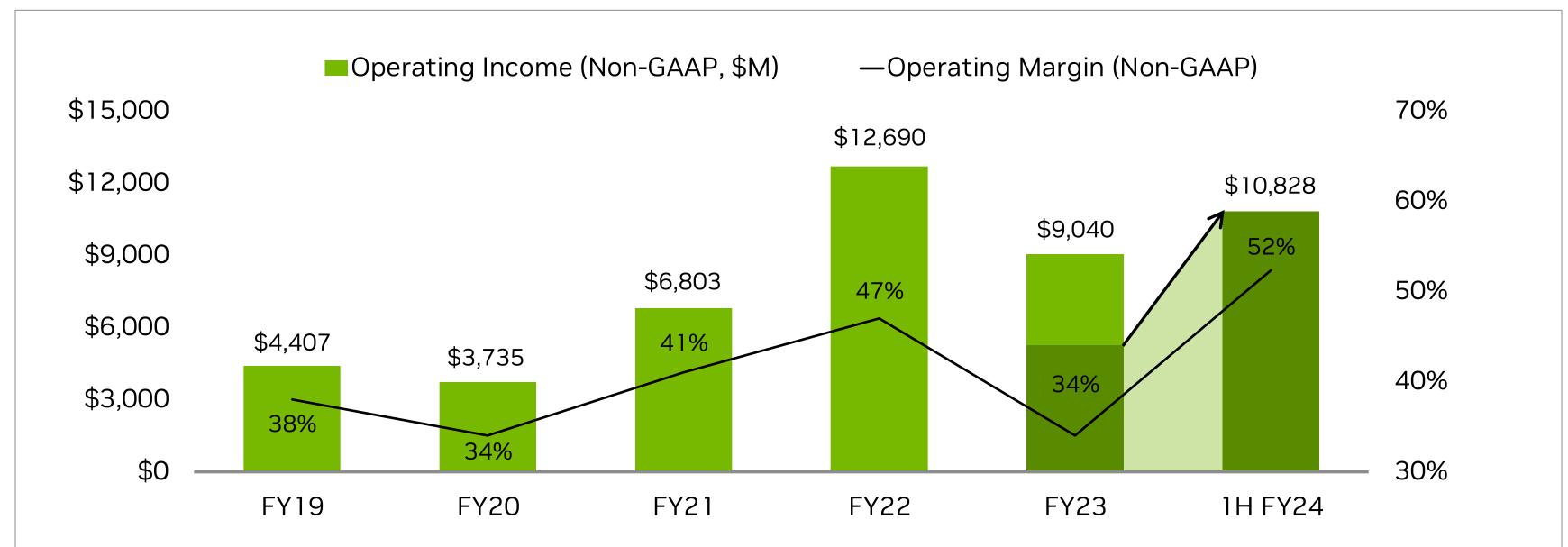
Broad and deep ecosystem and distribution to reach every enterprise



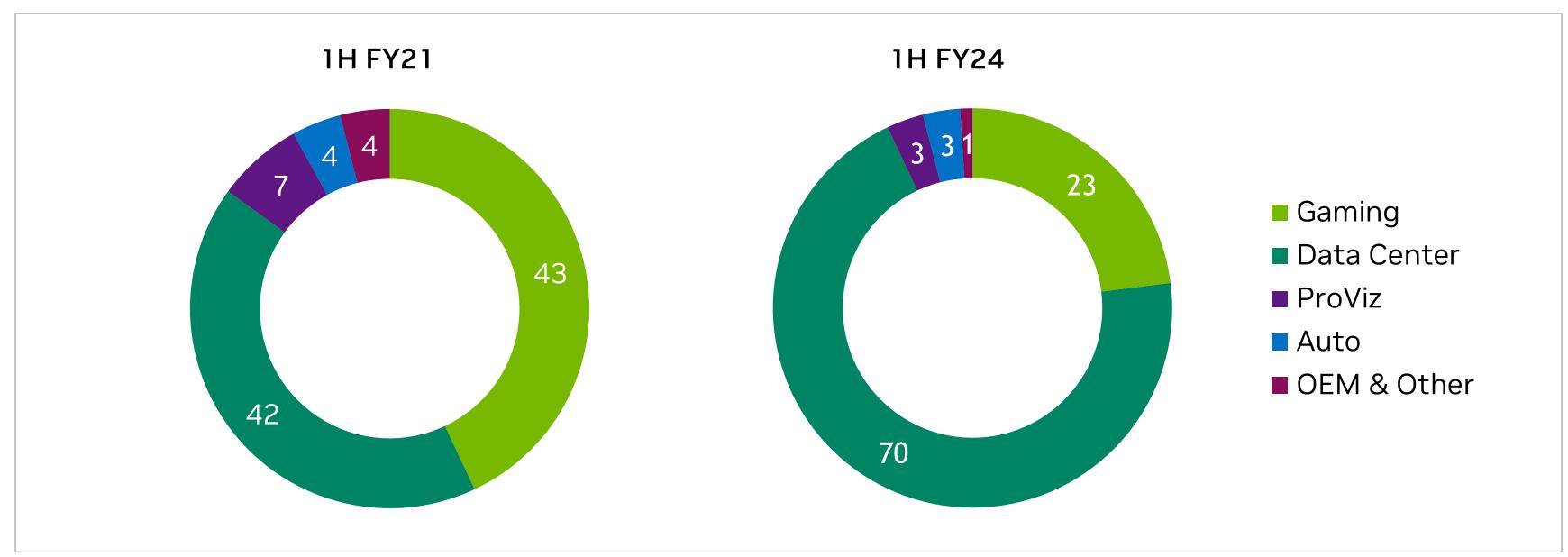


# Driving Strong & Profitable Growth





Fiscal year ends in January. Refer to Appendix for reconciliation of Non-GAAP measures. Operating margins rounded to the nearest percent.



FY23 financial metrics reflect a \$2.2B charge for inventory and related reserves primarily related to Data Center and Gaming.

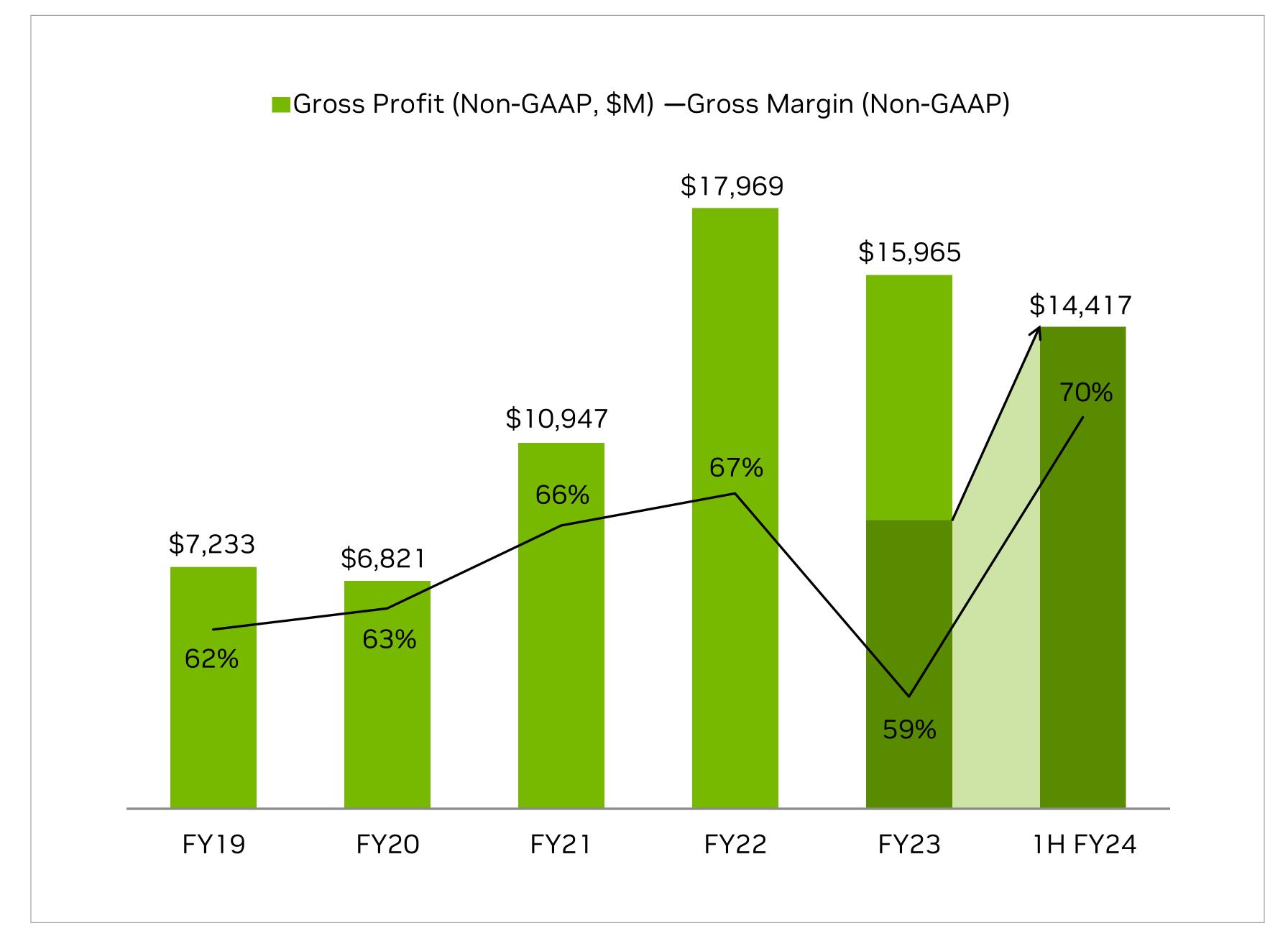
# **NVIDIA Gross Margins Reflect Value of Acceleration**

Accelerated computing requires full-stack and data center-scale innovation across silicon, systems, algorithms and applications.

Significant expertise and effort are required, but application speed-ups can be incredible, resulting in dramatic cost and time-to-solution savings.

For example, 2 NVIDIA HGX nodes with 16 NVIDIA H100 GPUs that cost \$400K can replace 960 nodes of CPU servers that cost \$10M for the same LLM workload.

NVIDIA chips carry the value of the full-stack, not just the chip.

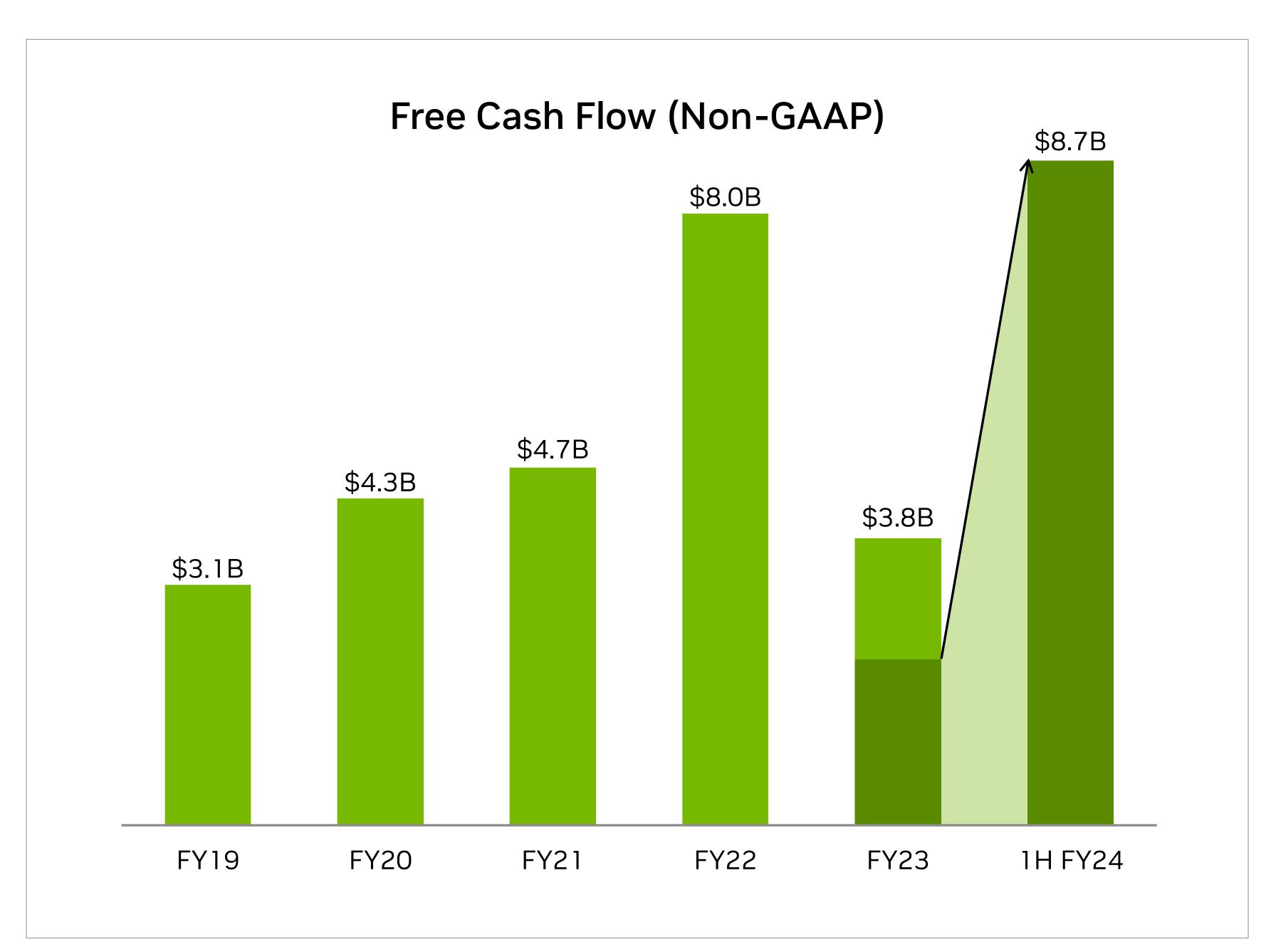


FY23 financial metrics reflect a \$2.2B charge for inventory and related reserves primarily related to Data Center and Gaming.

Fiscal year ends in January. Refer to Appendix for reconciliation of Non-GAAP measures. Gross margins are rounded to the nearest percent.

Cost comparison example based on latest available NVIDIA A 100 GPU and Intel CPU inference results in the commercially available category of the MLPerf industry benchmark; includes related infrastructure costs such as networking.

# Strong Cash Flow Generation



### Fiscal year ends in January. Refer to Appendix for reconciliation of Non-GAAP measures. <sup>1</sup> Subject to continuing determination by our Board of Directors.

### Capital Allocation

### **Share Repurchase**

\$10B repurchased in FY23
Additional \$25B in stock repurchases authorized,
adding to \$4B which remained as of end of Q2

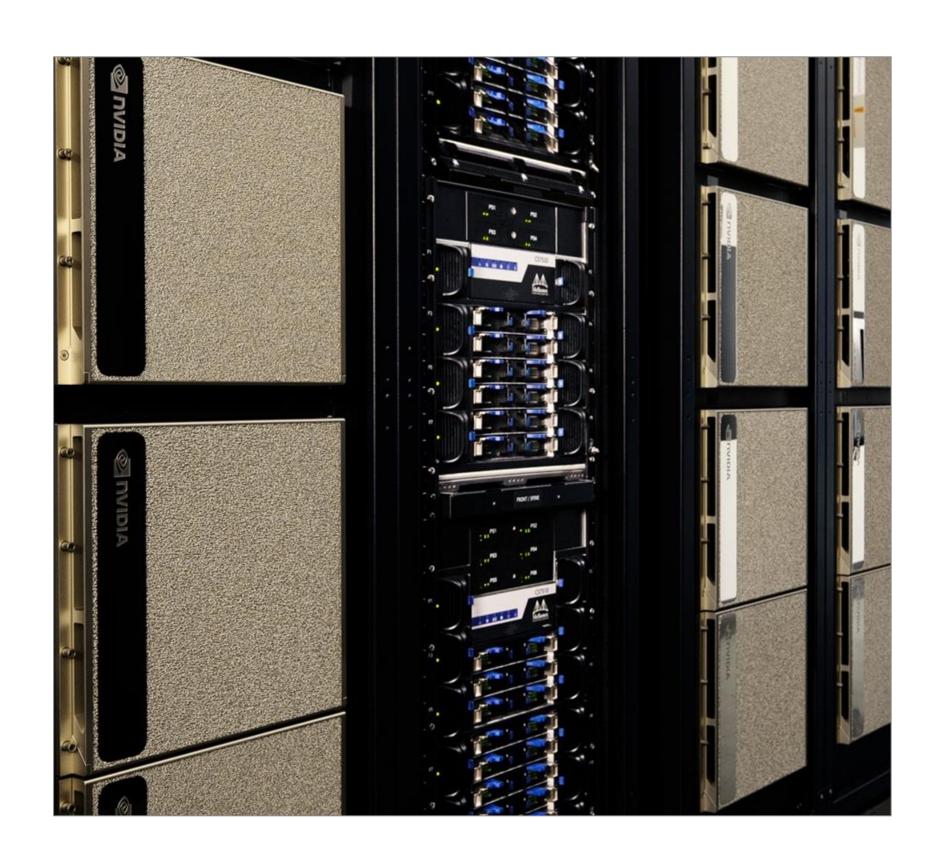
### Dividend

\$398M in FY 2023 Plan to Maintain<sup>1</sup>

### Strategic Investments

Growing Our Talent
Platform Reach & Ecosystem

# Our Market Platforms at a Glance









### **Data Center**

56% of FY23 revenue

FY23 Revenue \$15.0B

5-yr CAGR 51%

DGX/HGX/MGX/IGX systems
GPU | CPU | DPU | Networking
NVIDIA AI software

### Gaming

33% of FY23 revenue

**FY23 Revenue \$9.1B** 5-yr CAGR 10%

GeForce GPUs for PC gaming
GeForce NOW cloud gaming

### **Professional Visualization**

6% of FY23 revenue

FY23 Revenue \$1.5B

5-yr CAGR 11%

NVIDIA RTX GPUs for workstations

Omniverse software

### **Automotive**

3% of FY23 revenue

FY23 Revenue \$0.9B

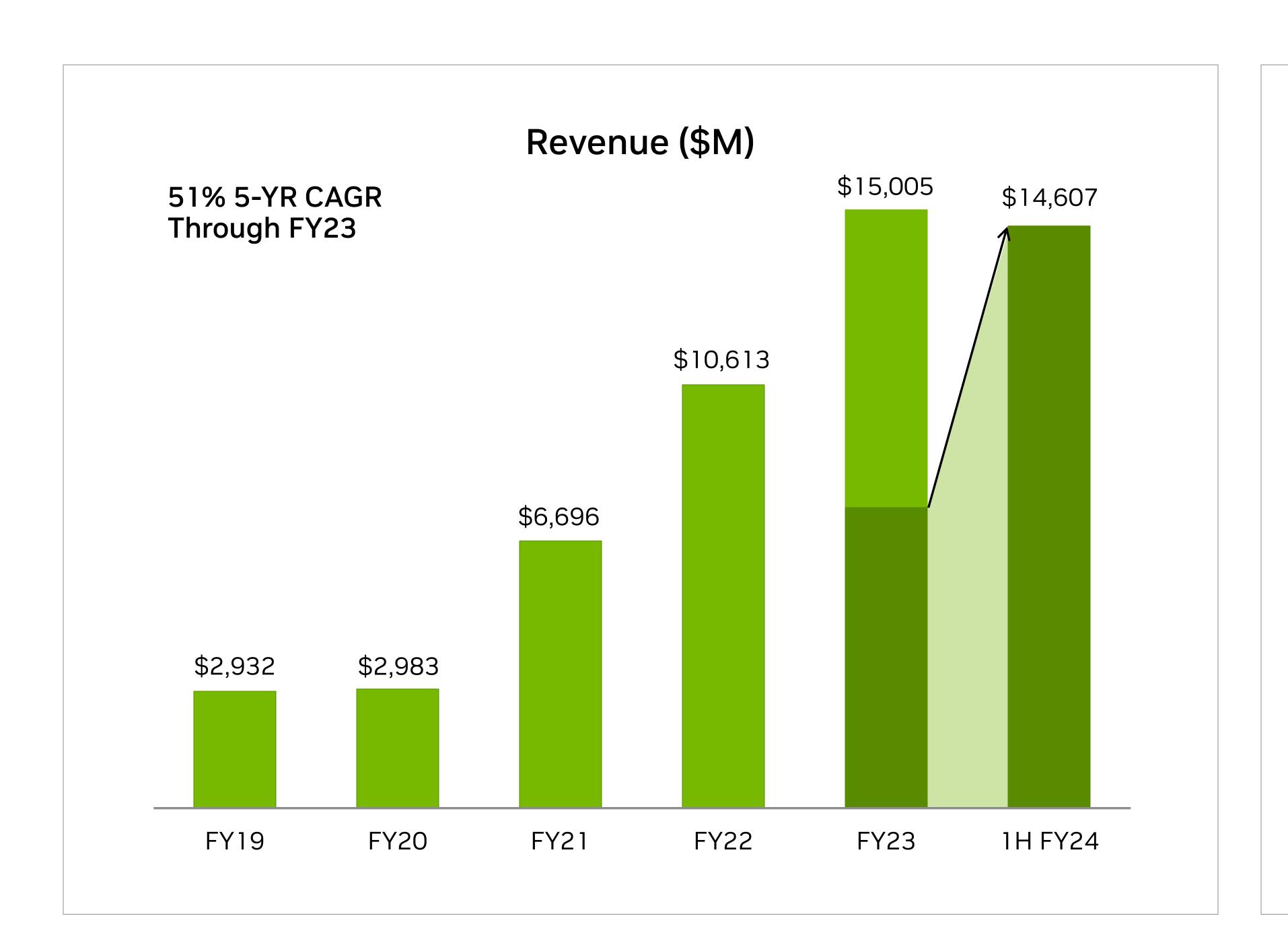
5-yr CAGR 10%

DRIVE Hyperion sensor architecture with AGX compute

DRIVE AV & IX full stack software for ADAS, AV & AI cockpit

# Data Center

### The leading computing platform for AI, HPC & graphics



### Leader in AI & HPC

#1 in Al training and inference

Used by all hyperscale & major cloud computing providers and 40,000 enterprises

Powers 74% of the TOP500 supercomputers

### **Growth Drivers**

Rapid AI adoption across industries Full-stack AI | Software

Three chip strategy — GPU | CPU | DPU

Rising computation requirements for modern Al

Data-center scale innovation

Omniverse

# NVIDIA AI - One Architecture | Train and Deploy Everywhere

From Two-Year Rhythm to

One-Year Rhythm | Training & Inference | x86 & Arm | Hyperscale & Enterprise



# **NVIDIA Grace Hopper Superchip**

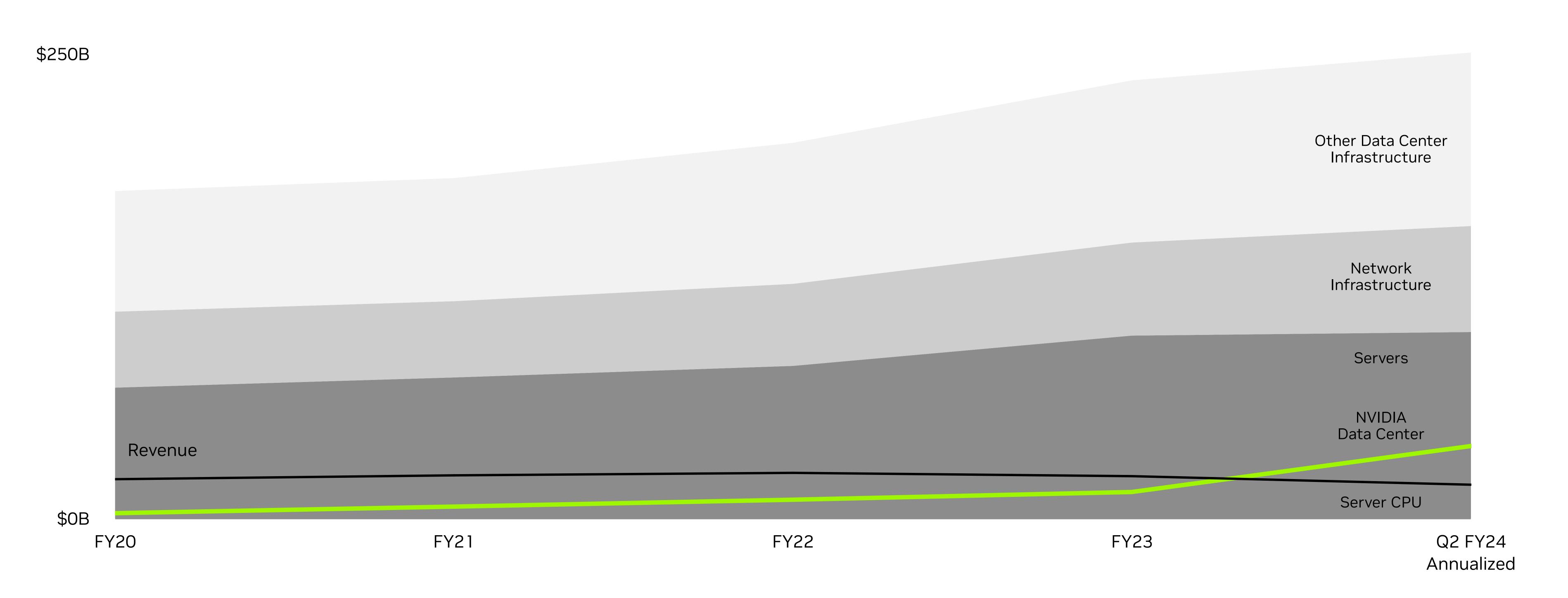
**NVIDIA BlueField-3** 

### **NVIDIA GH200**

72-Core Grace CPU
500 GB LPDDR5X
4 PFLOPS Hopper GPU
141 GB/5 Tbps HBM3e

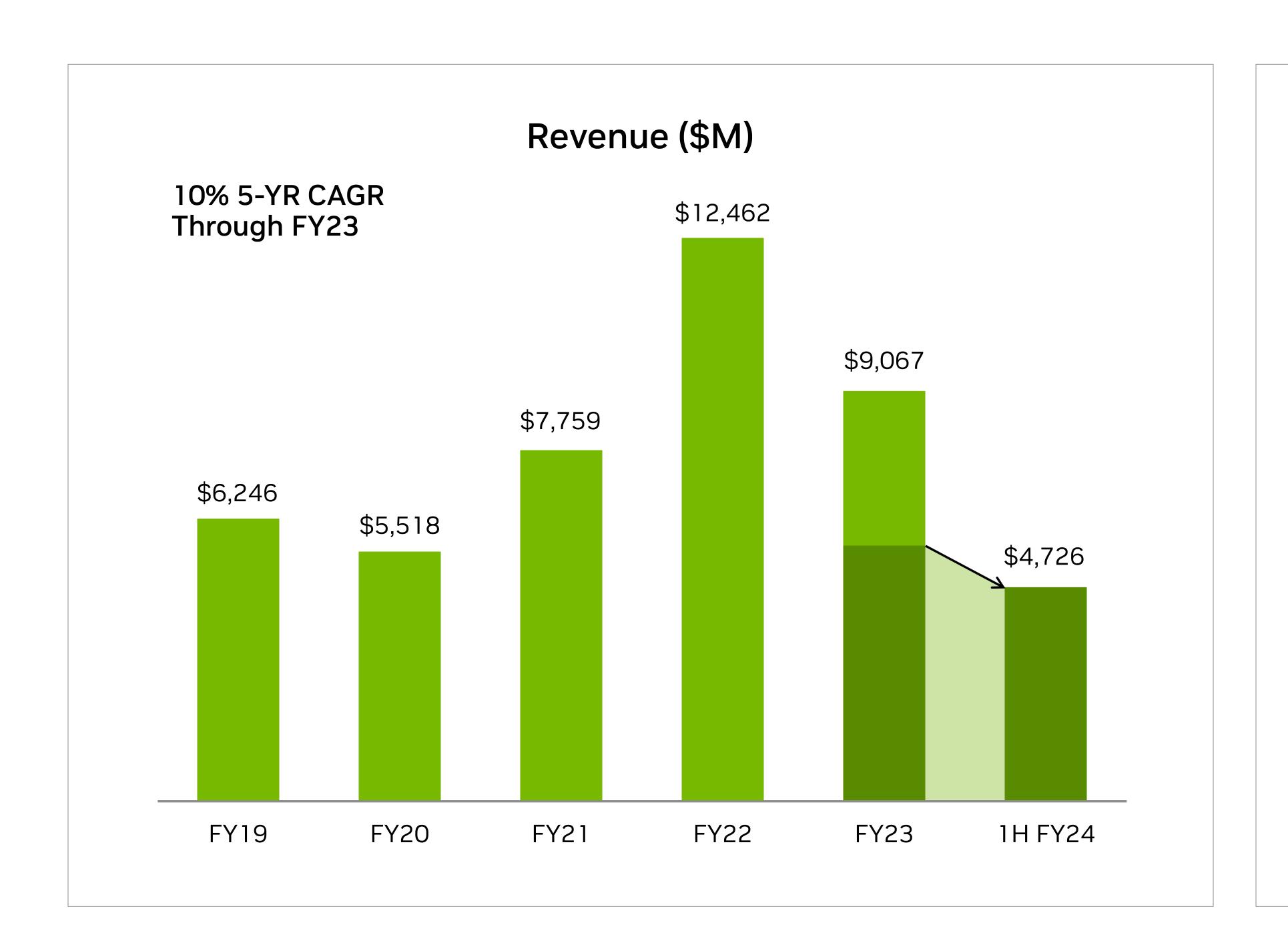
# Addressing the Entire Data Center

\$1T+ data center infrastructure installed base



# Gaming

### GeForce — the world's largest gaming platform



### Leader in PC Gaming

Strong #1 market position

15 of the top 15 most popular GPUs on Steam

Leading performance & innovation

200M+ gamers on GeForce

### **Growth Drivers**

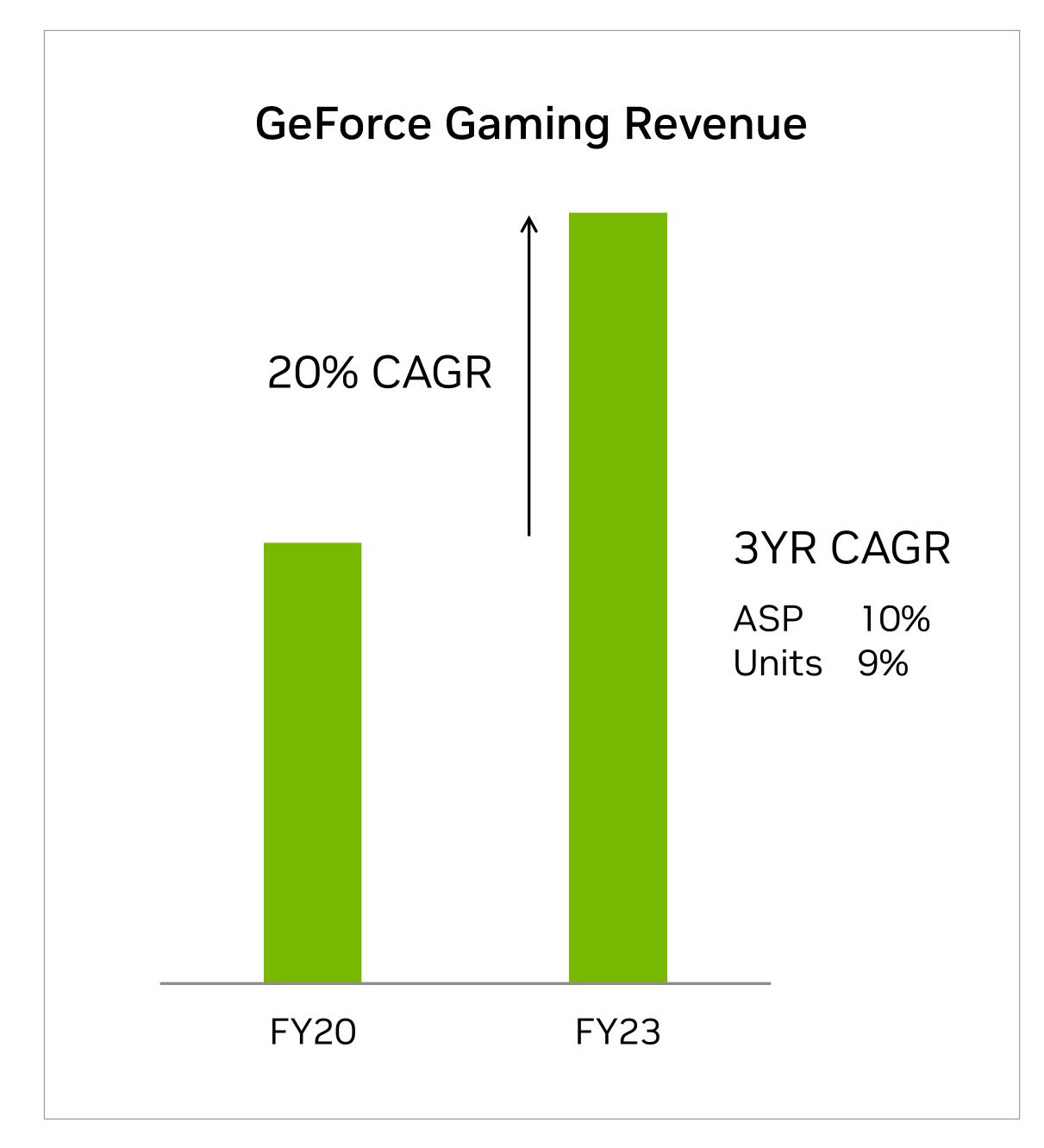
Rising adoption of NVIDIA RTX in games

Expanding universe of gamers & creators

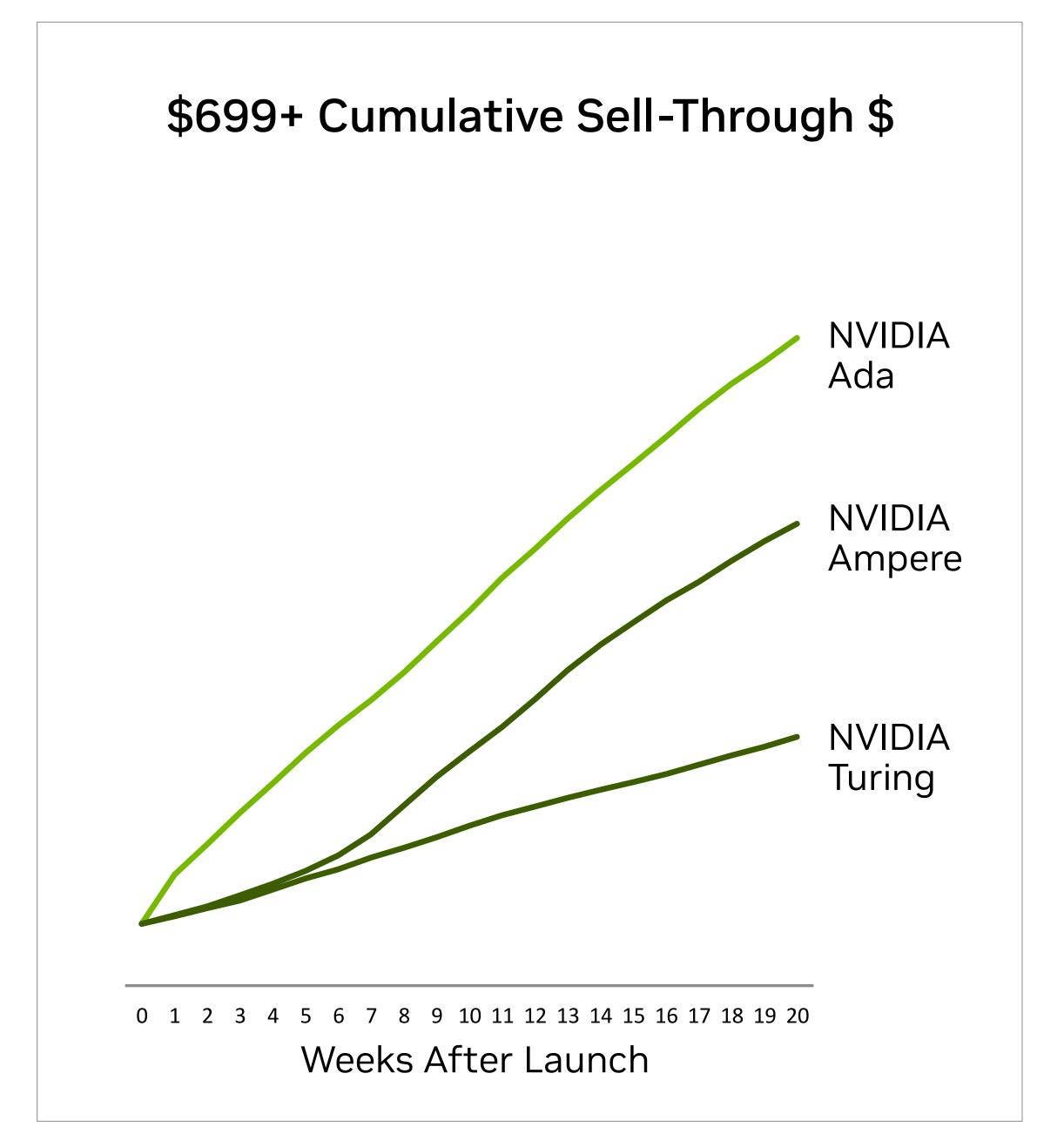
Gaming laptops & Gen AI on PCs

GeForce NOW Cloud gaming

# GeForce Extends Growth, Large Upgrade Opportunity



**Installed Base** 47% RTX RTX 3060+ 20% RTX3060+ Performance



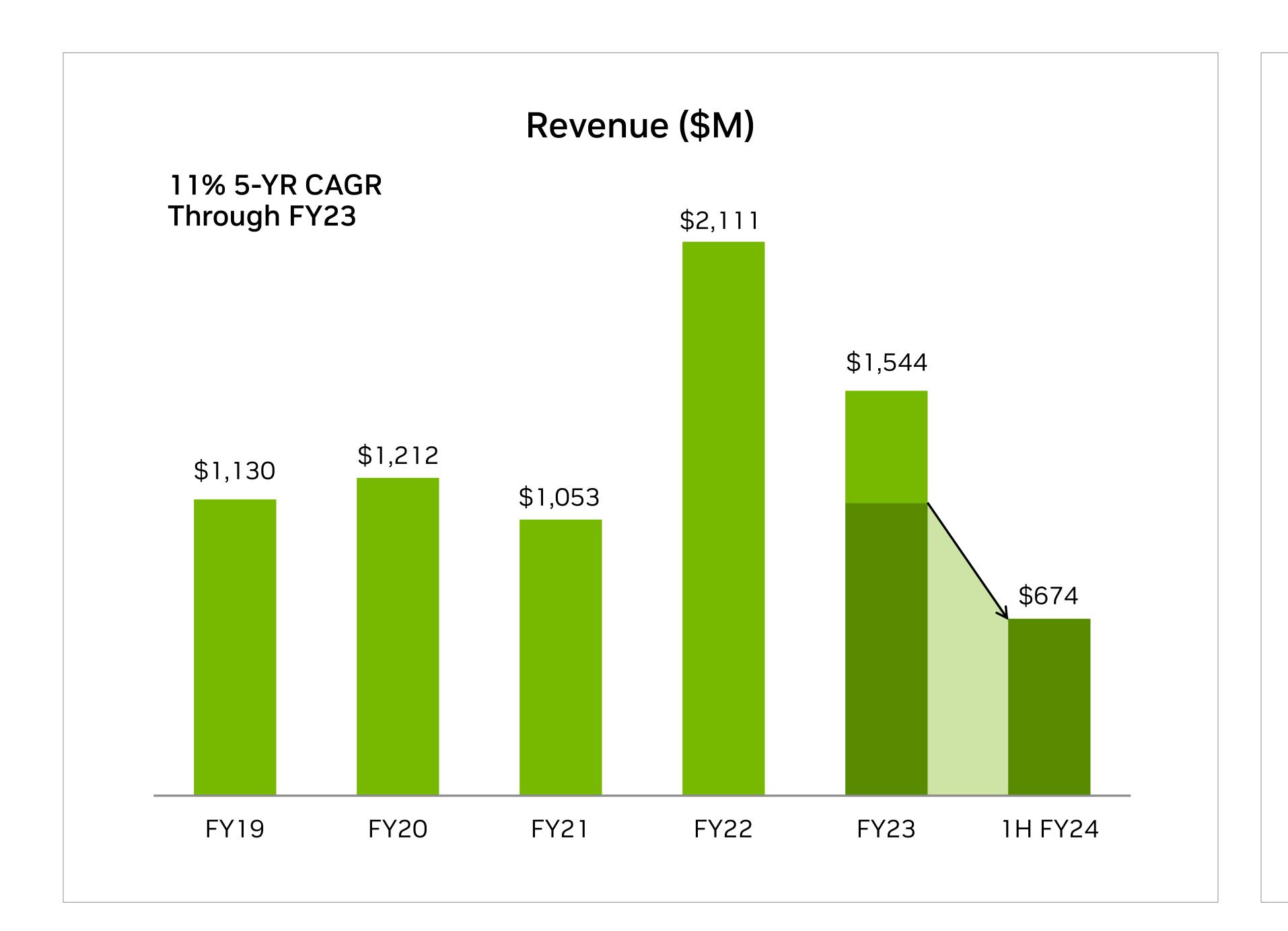
More Gamers, Richer Mix

Installed Base Needs Upgrade

Ada: 3X Turing Ramp at \$699+

# Professional Visualization

Workstation graphics



### Leader in Workstation Graphics

95%+ market share in graphics for workstations

45M Designers and Creators

Strong software ecosystem with over 100 RTX accelerated and supported applications

### **Growth Drivers**

Ray Tracing and generative AI revolutionizing design and content creation

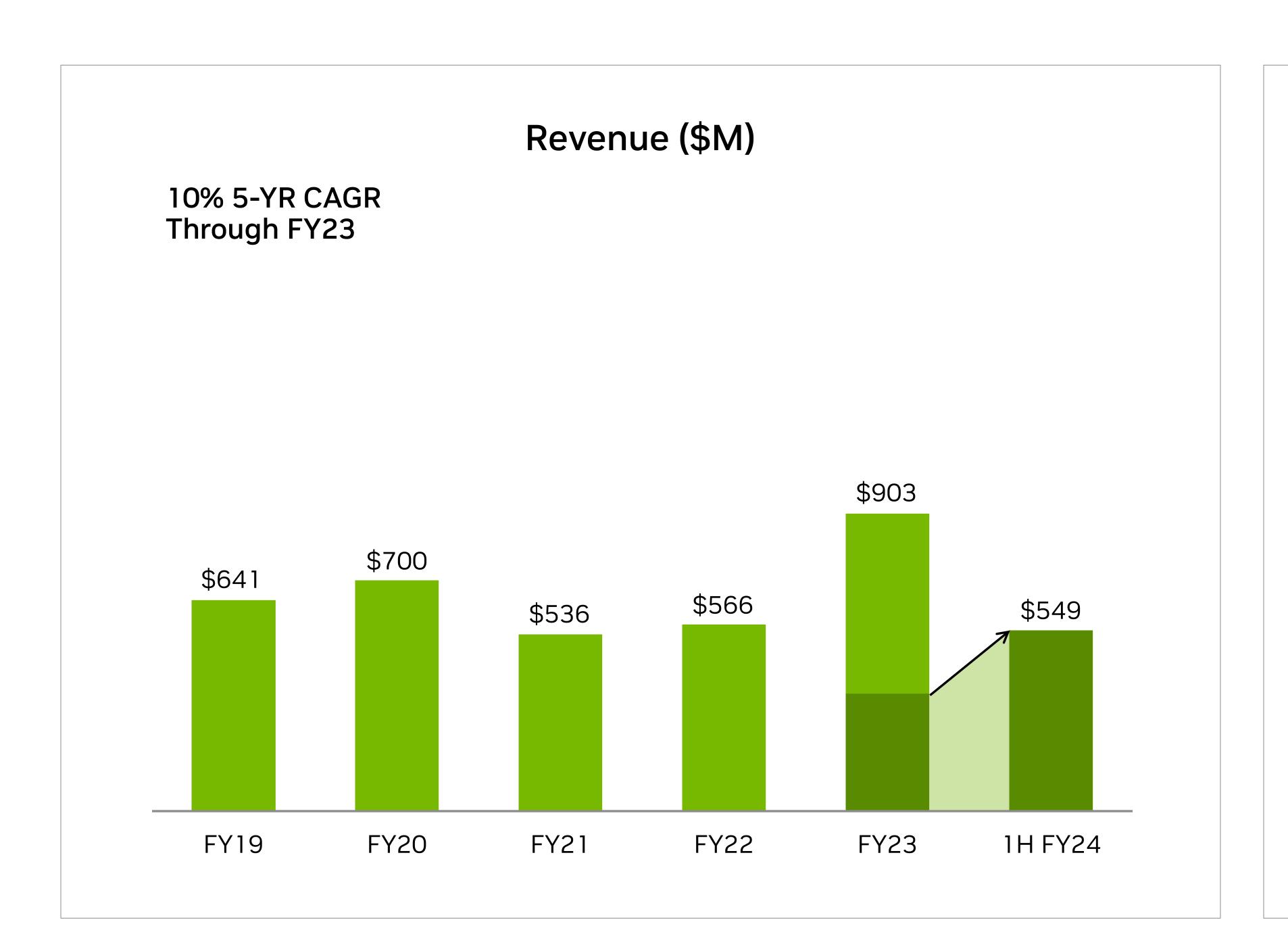
Expanding universe of designers and creators

Collaborative 3D design / Omniverse

Hybrid work environments

# Automotive

### Autonomous Vehicles (AV) & Al Cockpit



### Leader in Autonomous Driving

Revenue growth primarily fueled by NVIDIA DRIVE, our AV and AI cockpit platform with full software stack

Inflection in FY23 driven by AV as DRIVE Orin SoC began to ramp Next-generation DRIVE Thor SoC to ramp in FY26

### **Growth Drivers**

Adoption of centralized car computing and software-defined vehicle architectures

AV software and services:

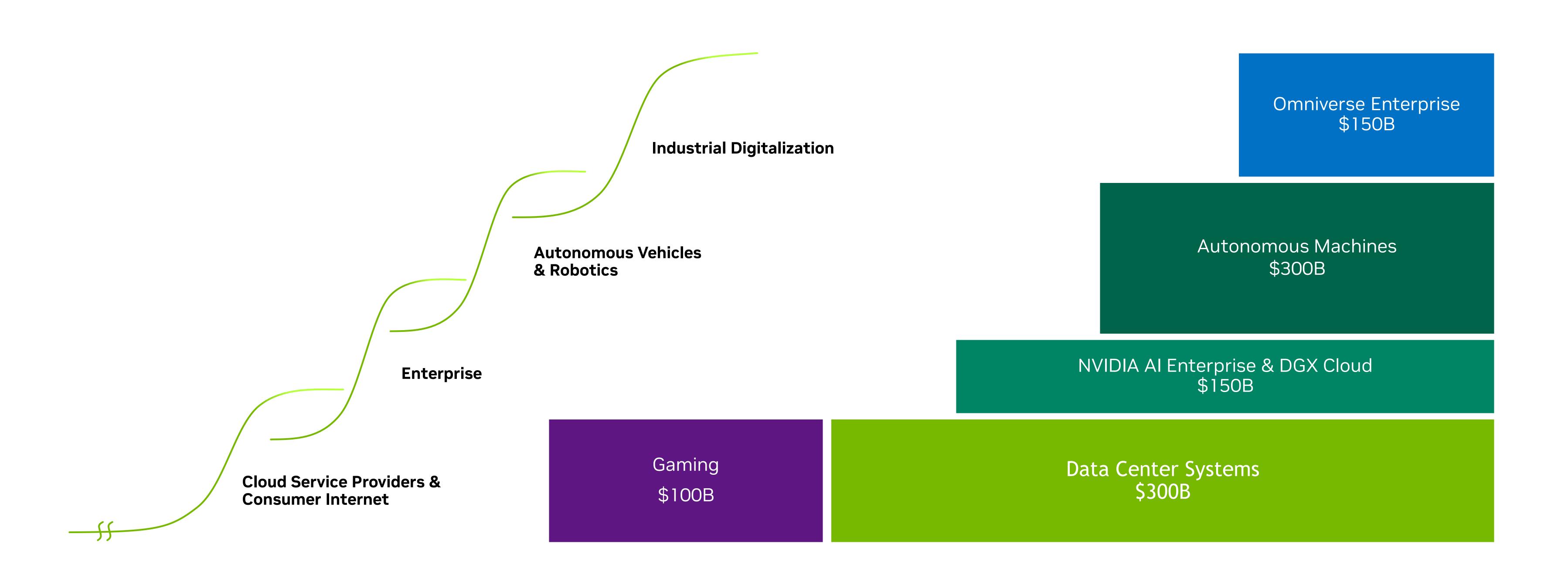
Mercedes Benz

Jaguar Land Rover

\$14B Design Win Pipeline Through FY29



# \$1 Trillion Long-Term Annual Market Opportunity



# Summary

Gen Al is the tipping point for the new computing era

Al is the new software and Accelerated Computing the new hardware

Huge ROI from Gen AI – from new revenue or dramatically lower costs - is driving a powerful new investment cycle

NVIDIA's accelerated computing platform delivers unmatched performance and TCO savings

Strong revenue, operating profit, and cash flow growth

\$1T market opportunity

# Reconciliation of Non-GAAP to GAAP Financial Measures

# Reconciliation of Non-GAAP to GAAP Financial Measures

Gross Margin  (\$ in Millions & Margin Percentage)	Non-GAAP	Acquisition-Related and Other Costs (A)	Stock-Based Compensation (B)	IP-Related Costs	GAAP
EV 2010	\$7,233		(27)	(35)	\$7,171
FY 2019	61.7%		(0.2)	(0.3)	61.2%
EV 2020	\$6,821		(39)	(14)	\$6,768
FY 2020	62.5%		(0.4)	(O.1)	62.0%
FV 2021	\$10,947	(425)	(88)	(38)	\$10,396
FY 2021	65.6%	(2.6)	(0.5)	(0.2)	62.3%
FV 2022	\$17,969	(344)	(141)	(9)	\$17,475
FY 2022	66.8%	(1.4)	(0.5)		64.9%
EV 2022	\$15,965	(455)	(138)	(16)	\$15,356
FY 2023	59.2%	(1.7)	(0.5)	(O.1)	56.9%
111 EV 2022	\$8,636	(214)	(76)		\$8,346
1H FY 2023	57.6%	(1.4)	(0.5)		55.7%
111 57 202 4	\$14,417	(239)	(58)	(10)	\$14,110
1H FY 2024	69.7%	(1.2)	(0.3)		68.2%

A. Consists of amortization of intangible assets and inventory step-up

B. Stock-based compensation charge was allocated to cost of goods sold

# Reconciliation of Non-GAAP to GAAP Financial Measures (contd.)

Operating Income and Margin  (\$ in Millions & Margin Percentage)	Non-GAAP	Acquisition Termination Cost	Acquisition-Related and Other Costs (A)	Stock-Based Compensation (B)	IP-Related Costs	Other (C)	GAAP
FY 2019	\$4,407		(2)	(557)	(35)	(9)	\$3,804
	37.6%			(4.7)	(0.3)	(O.1)	32.5%
EV 2020	\$3,735		(31)	(844)	(14)		\$2,846
FY 2020	34.2%		(0.3)	(7.7)	(O.1)		26.1%
FY 2021	\$6,803		(836)	(1,397)	(38)		\$4,532
	40.8%		(5.0)	(8.4)	(0.2)		27.2%
EV 2022	\$12,690		(636)	(2,004)	(9)		\$10,041
FY 2022	47.2%		(2.5)	(7.4)			37.3%
FY 2023	\$9,040	(1,353)	(674)	(2,710)	(16)	(63)	\$4,224
	33.5%	(5.0)	(2.5)	(10.0)	(O.1)	(0.2)	15.7%
111 51/ 2022	\$5,280	(1,353)	(324)	(1,227)		(9)	\$2,367
1H FY 2023	35.2%	(9.0)	(2.2)	(8.2)			15.8%
1H FY 2024	\$10,828		(311)	(1,576)	(10)	10	\$8,941
	52.3%		(1.5)	(7.6)		<u>——</u>	43.2%



A. Consists of amortization of acquisition-related intangible assets, inventory step-up, transaction costs, compensation charges, and other costs

B. Stock-based compensation charge was allocated to cost of goods sold, research and development expense, and sales, general and administrative expense

C. Comprises of legal settlement costs, contributions, restructuring costs and assets held for sale related adjustments

# Reconciliation of Non-GAAP to GAAP Financial Measures

(\$ in Millions)	Free Cash Flow	Purchases Related to Property and Equipment and Intangible Assets	Principal Payments on Property and Equipment and Intangible Assets	Net Cash Provided by Operating Activities
FY 2019	\$3,143	600		\$3,743
FY 2020	\$4,272	489		\$4,761
FY 2021	\$4,677	1,128	17	\$5,822
FY 2022	\$8,049	976	83	\$9,108
FY 2023	\$3,750	1,833	58	\$5,641
1H FY 2023	\$2,171	794	36	\$3,001
1H FY 2024	\$8,691	537	31	\$9,259

