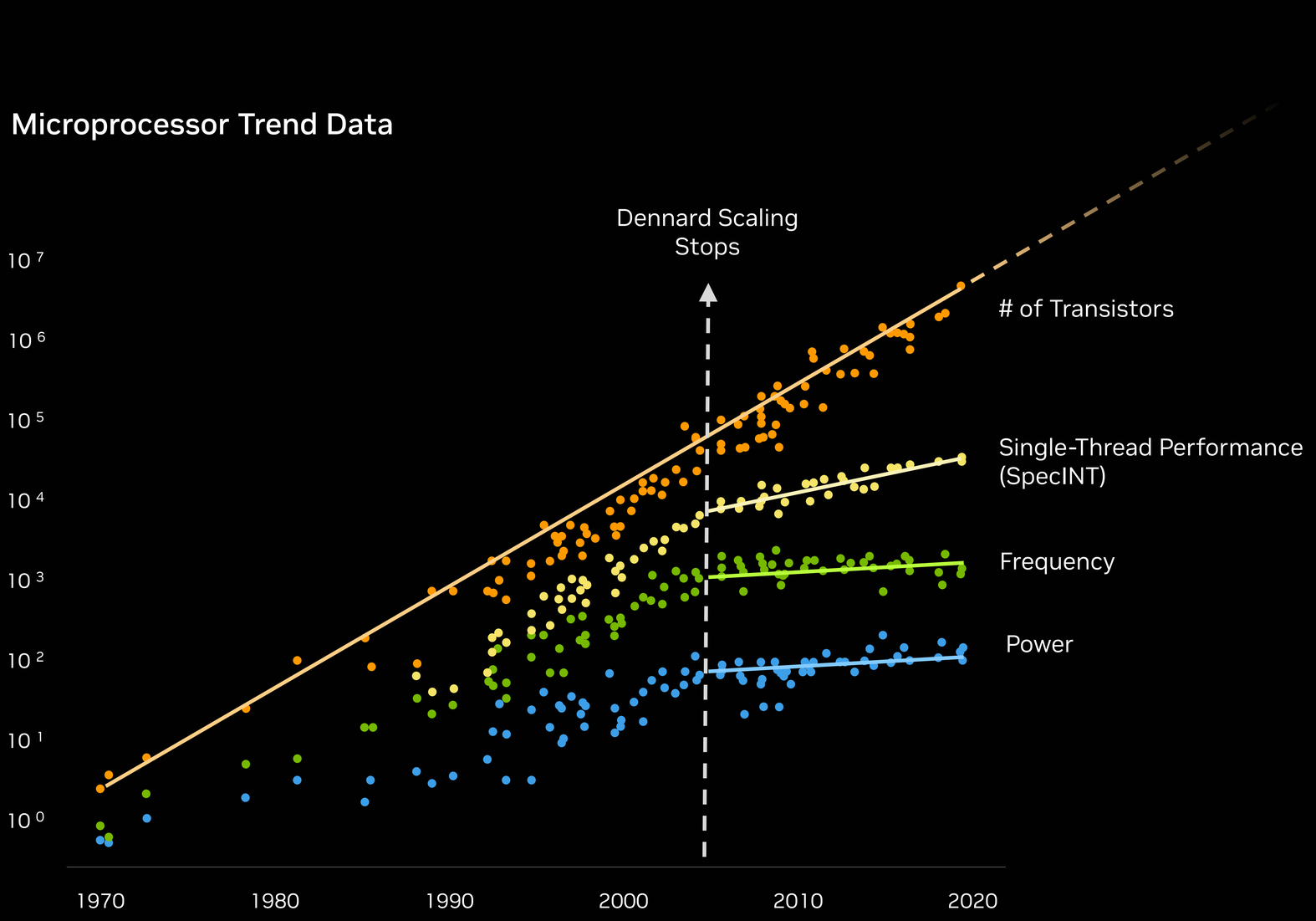
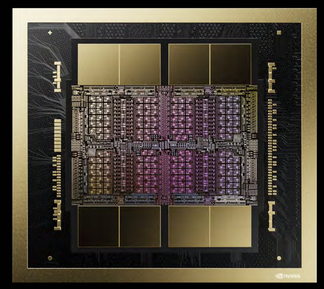

NVIDIA


GTC25

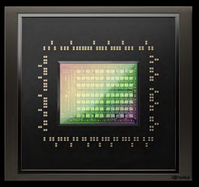
CPU General Purpose to GPU Accelerated Computing Platform Shift





GPU

+



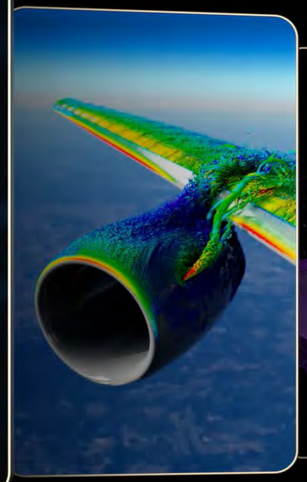
CPU

Accelerated Computing Begins with CUDA-X Libraries



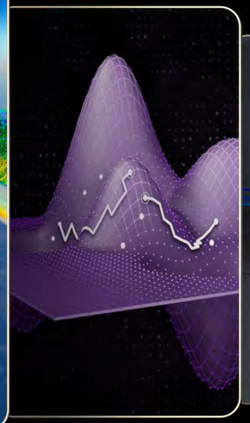
cuLitho

Computational Lithography



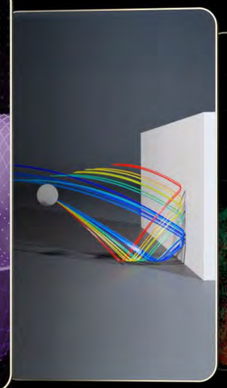
cuDSS
cuSPARSE
cuFFT
AmgX

Computer Aided Engineering



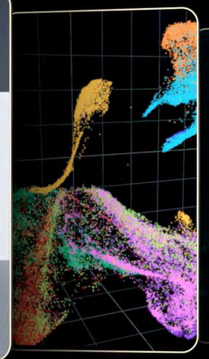
cuOpt

Decision Optimization



Warp

Physics



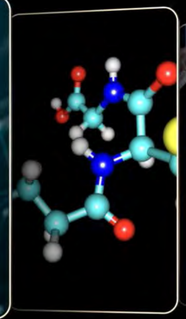
cuDF
cuML

Data Science and Processing



Megatron
Dynamo
NIXL
cuDNN
CUTLASS

Deep Learning



cuEquivariance
cuTensor

Quantum Chemistry



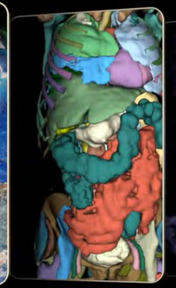
cuQuantum
CUDA-Q

Quantum Computing



Earth-2

Weather Analytics



MONAI

Medical Imaging



Parabricks

Genomics



Aerial
Sionna

5G/6G Signal Processing



cuPyNumeric

Numerical Computing



AI 6G Quantum Models Enterprise Robotics Factories



Announcing
NVIDIA and Nokia Pioneer the AI Era of Telecommunications

NOKIA | NVIDIA

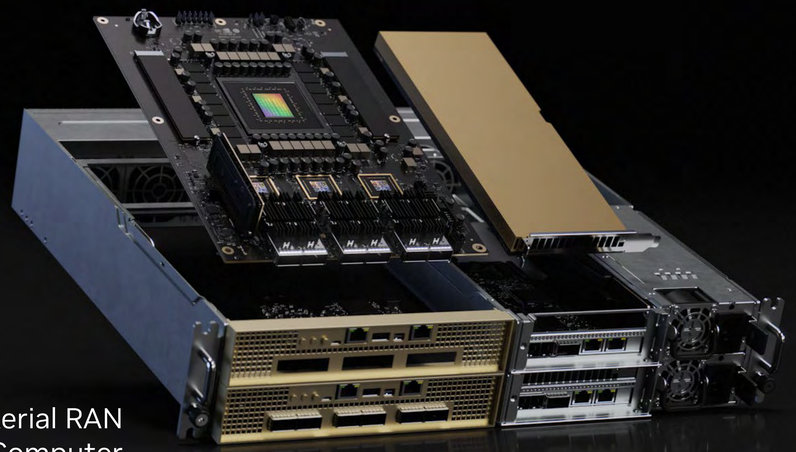


Announcing
Nokia to Build AI-Native 6G on New NVIDIA ARC
“Aerial RAN Computer”

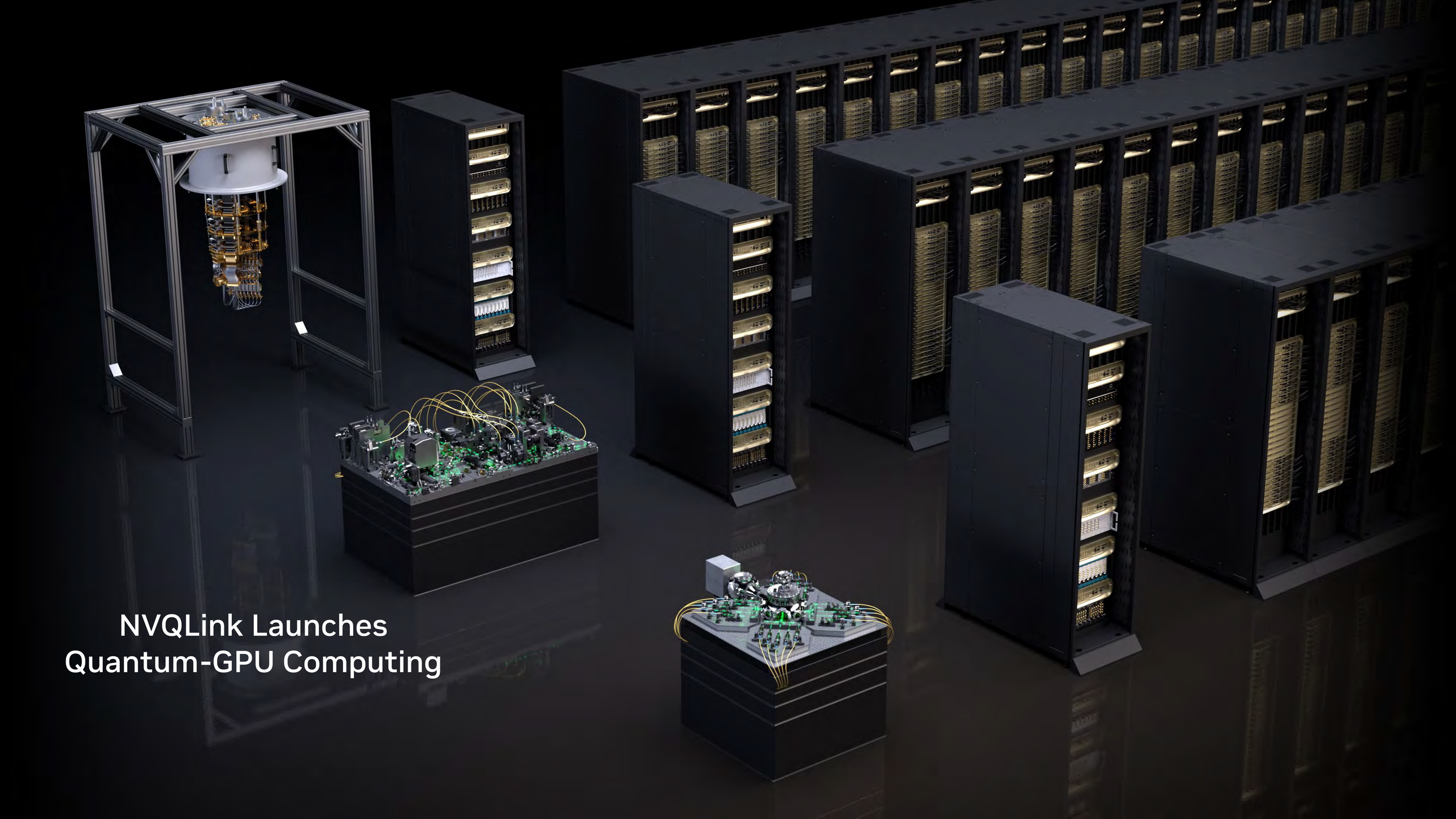
NOKIA | NVIDIA

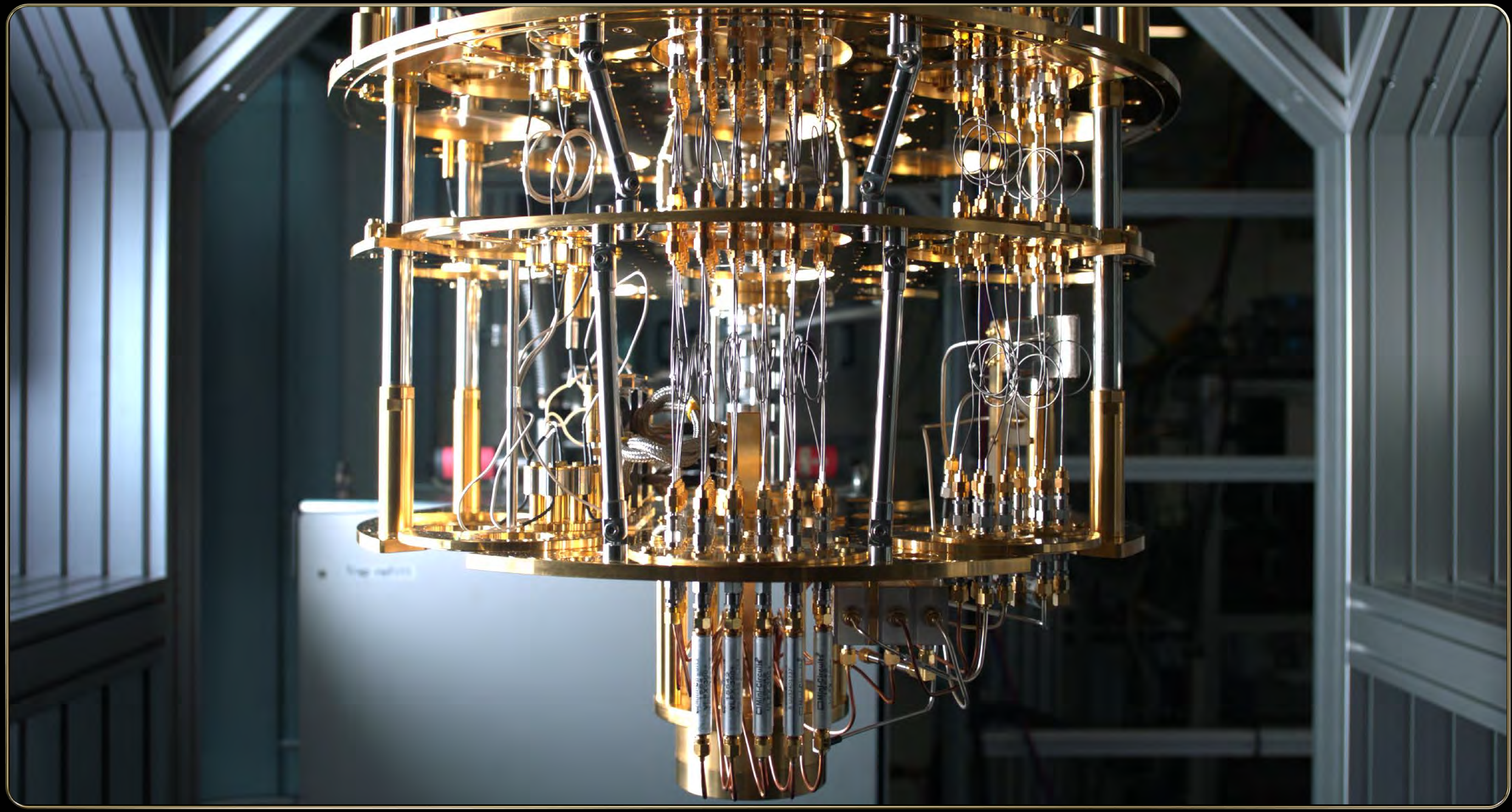


Nokia MIMO Radio



Aerial RAN
Computer


NVQLink Launches
Quantum-GPU Computing





NVQLink

QUANTINUUM

QUANTUM MACHINES

BERKELEY LAB

Zurich Instruments

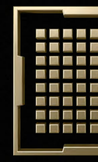
QBLOX

KEYSIGHT

Announcing NVIDIA NVQLink — Quantum-GPU Interconnect



NVQLink



NVLink



QPU GPU CPU

CUDA-Q | cuQuantum | CUDA

ALICE & BOB · Anyon Technologies · atom computing · diraq · Infleqtion · IONQ · IQM · OQC · ORCA COMPUTING · Pasqal · qci Quantum Circuits, Inc · QUANDELA · QUANTUM MOTION · IQuEra · rigetti · seeqc · Silicon Quantum Computing

BERKELEY LAB · Brookhaven National Laboratory · Fermilab · LINCOLN LABORATORY MASSACHUSETTS INSTITUTE OF TECHNOLOGY · Los Alamos NATIONAL LABORATORY · OAK RIDGE National Laboratory · Pacific Northwest NATIONAL LABORATORY · Sandia National Laboratories

Oak Ridge National Laboratory and the oak leaf symbol are registered trademarks of the U.S. Department of Energy. Use of this mark does not constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof or its contractors or subcontractors.



Announcing Department of Energy Partners with NVIDIA to Build 7 New AI Supercomputers

Argonne National Laboratory

Solstice – 100K Blackwell GPUs

Equinox

Tara

Minerva

Janus

Los Alamos National Laboratory

Powered by Vera Rubin

Mission

Vision

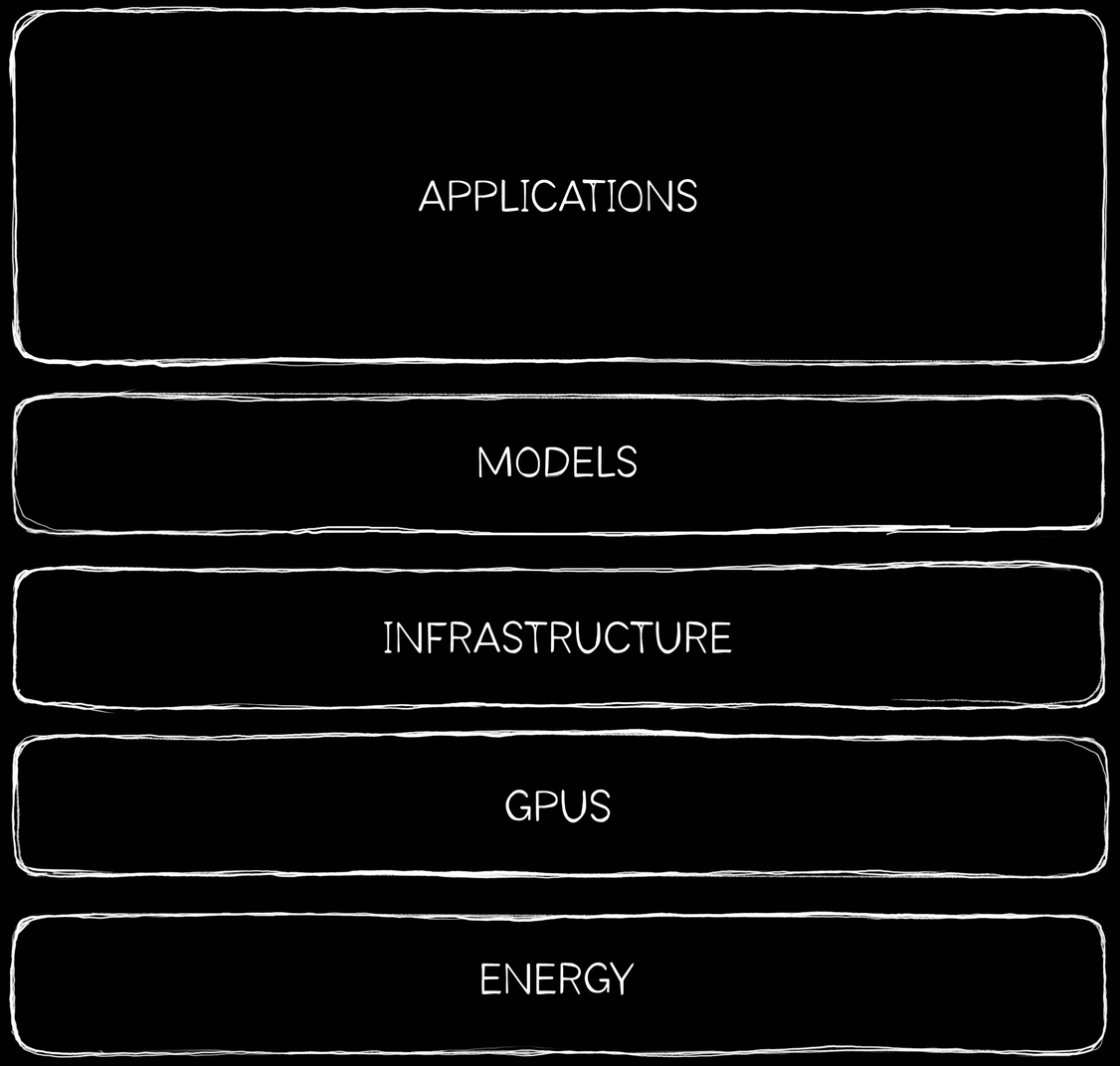
NVIDIA

Argonne NATIONAL LABORATORY

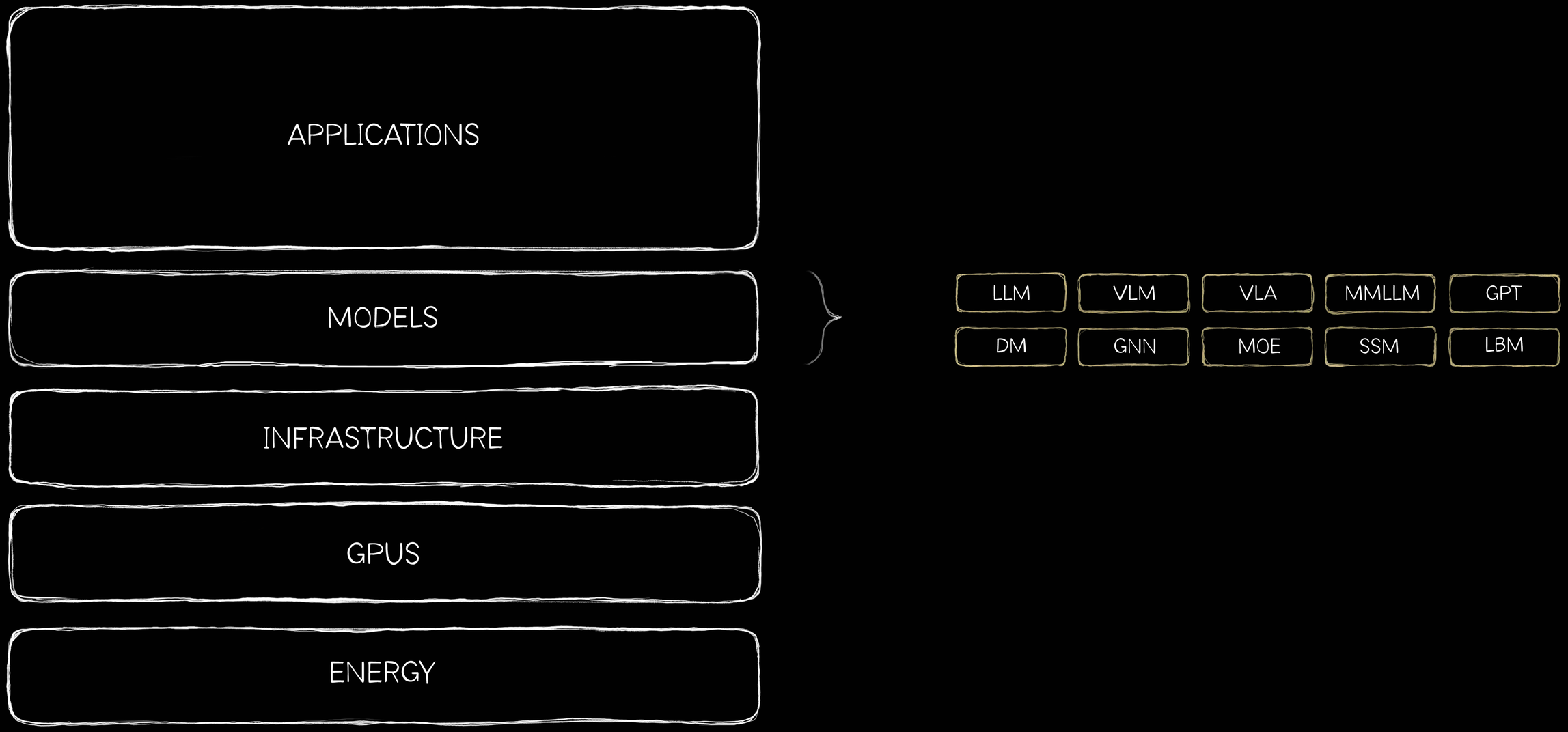
Los Alamos NATIONAL LABORATORY

U.S. DEPARTMENT of ENERGY

AI - THE NEW INDUSTRIAL REVOLUTION



AI - THE NEW INDUSTRIAL REVOLUTION



AI - THE NEW INDUSTRIAL REVOLUTION

APPLICATIONS

MODELS

INFRASTRUCTURE

GPUS

ENERGY



CHATBOTS



DIGITAL BIOLOGY



ROBOTAXI



ENTERPRISE AI AGENTS



SCIENCE



ROBOTICS



MANUFACTURING



AI CODER

LLM VLM VLA MMLLM GPT

DM GNN MOE SSM LBM

AI - THE NEW INDUSTRIAL REVOLUTION

APPLICATIONS

MODELS

INFRASTRUCTURE

GPUS

ENERGY

CHATBOTS

DIGITAL BIOLOGY

ROBOTAXI

ENTERPRISE AI AGENTS

SCIENCE

ROBOTICS

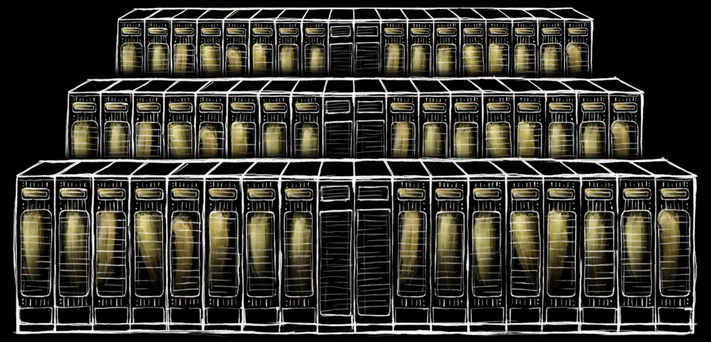
MANUFACTURING

AI CODER

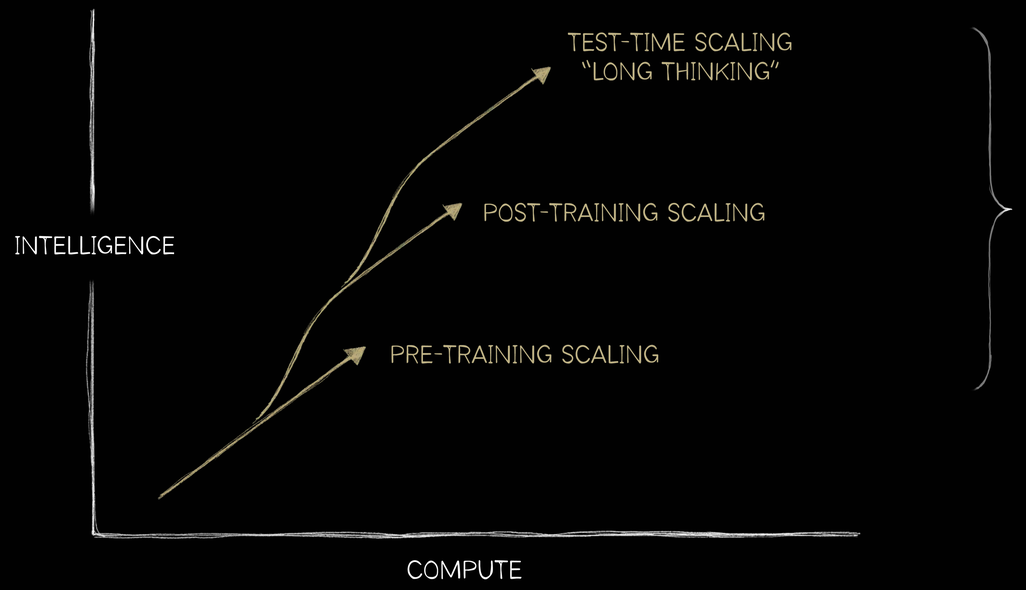
LLM | VLM | VLA | MMLLM | GPT

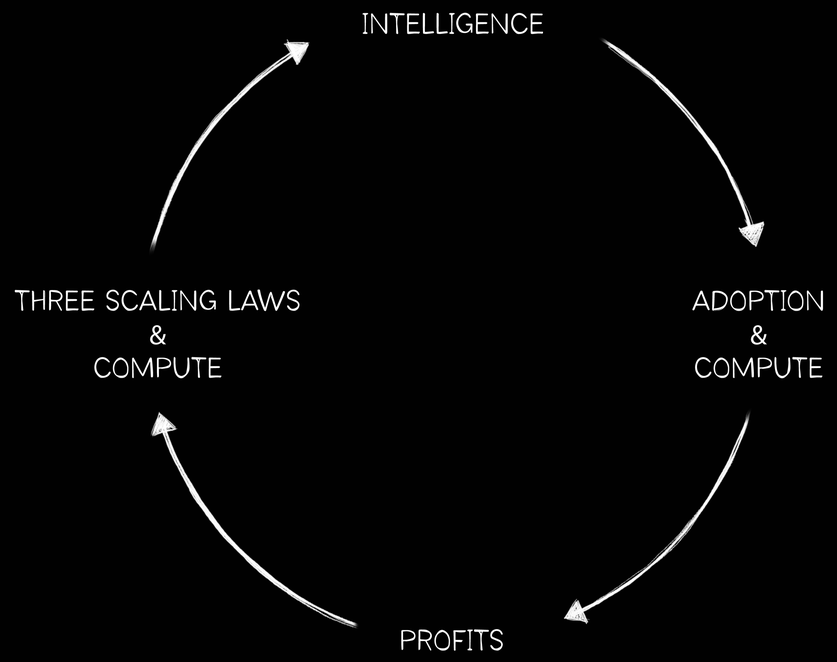
DM | GNN | MOE | SSM | LBM

AI FACTORIES

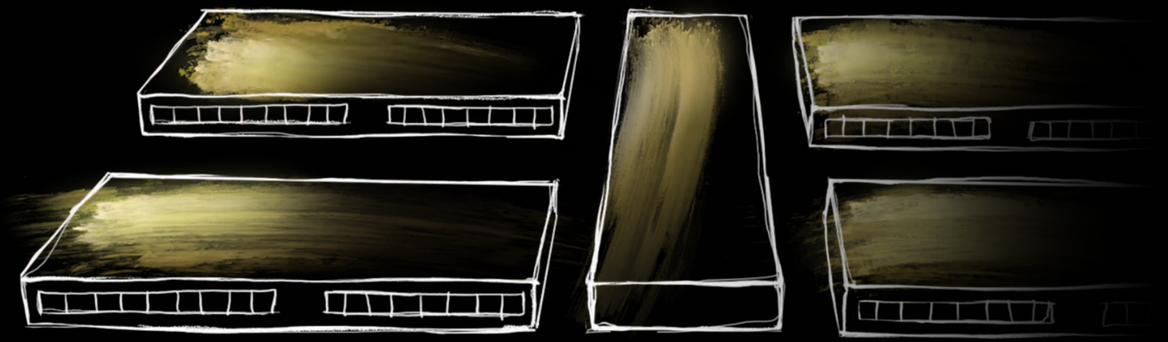


THREE SCALING LAWS
EXPONENTIAL GROWTH IN COMPUTE, ADOPTION AND USAGE



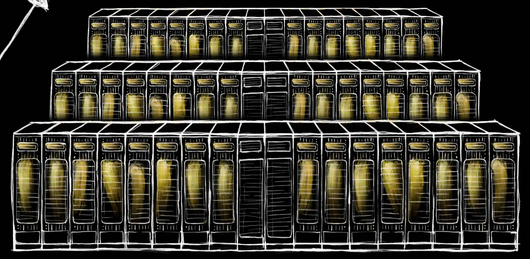


EXTREME CO-DESIGN — BEYOND MOORE'S LAW



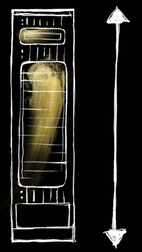
SCALE-ACROSS

SCALE-OUT



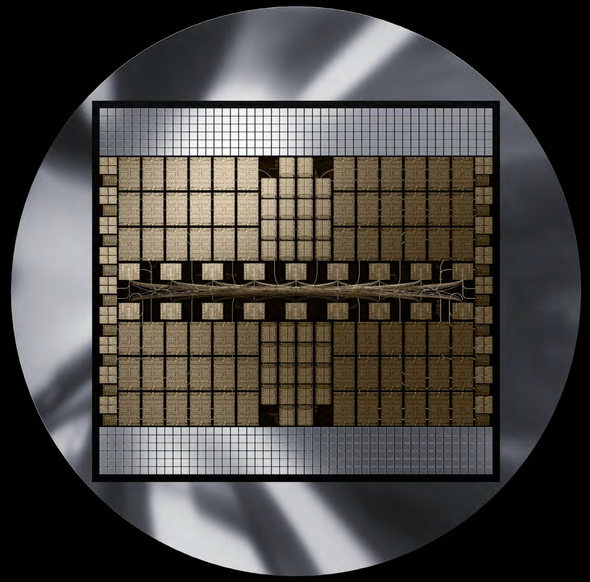
CO-DESIGN

- AI
- MODELS
- SYS SW
- SYS
- CHIPS

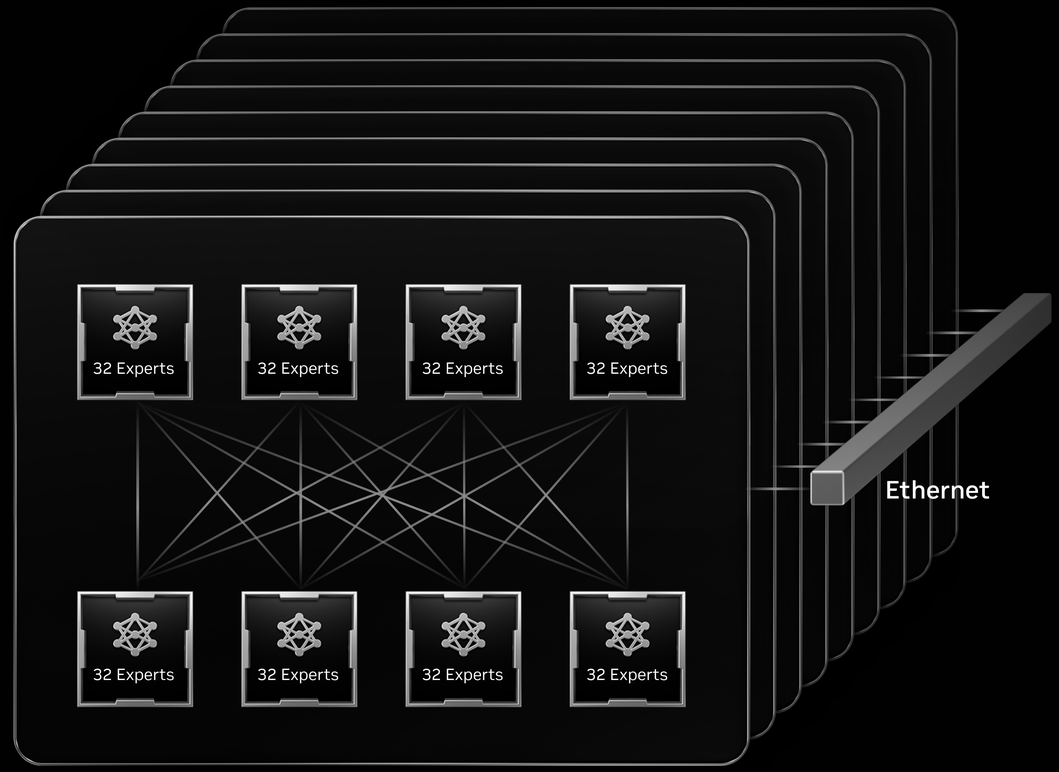


SCALE-UP

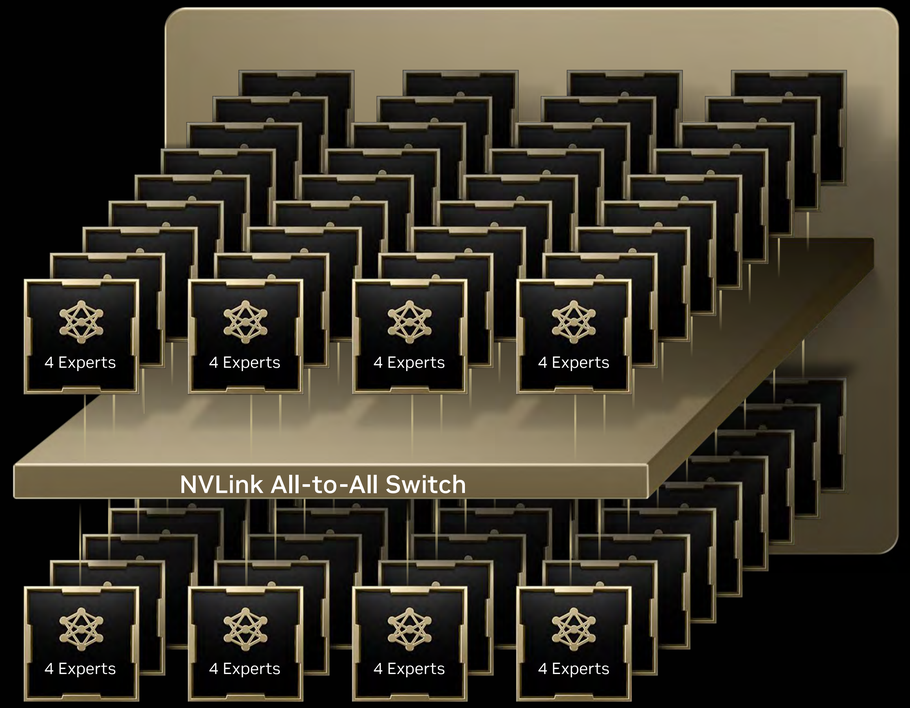
Grace Blackwell NVL72
“A Thinking Machine”



Extreme Co-Designed Blackwell NVL72
Game-Changer for Mixture-of-Experts Models



Point-to-Point
8 GPUs x 9 Servers



All-to-All NVLink Switch
72 GPUs x 1 Rack

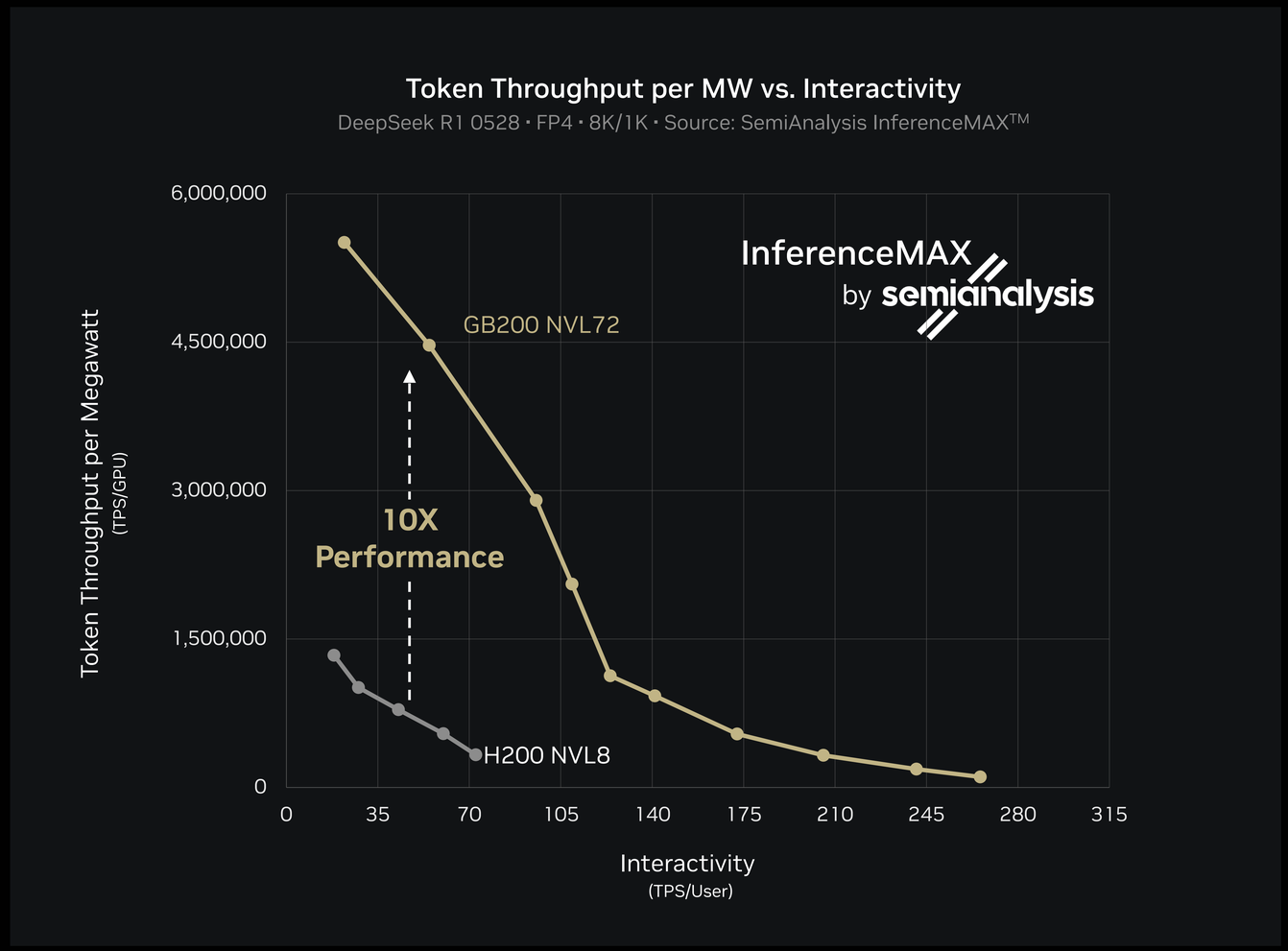
Extreme Co-Designed Blackwell NVL72
Game-Changer for Mixture-of-Experts Models

10X
Perf-per-Dollar

10X
Perf-per-Watt

10X
Throughput

10X
Revenues



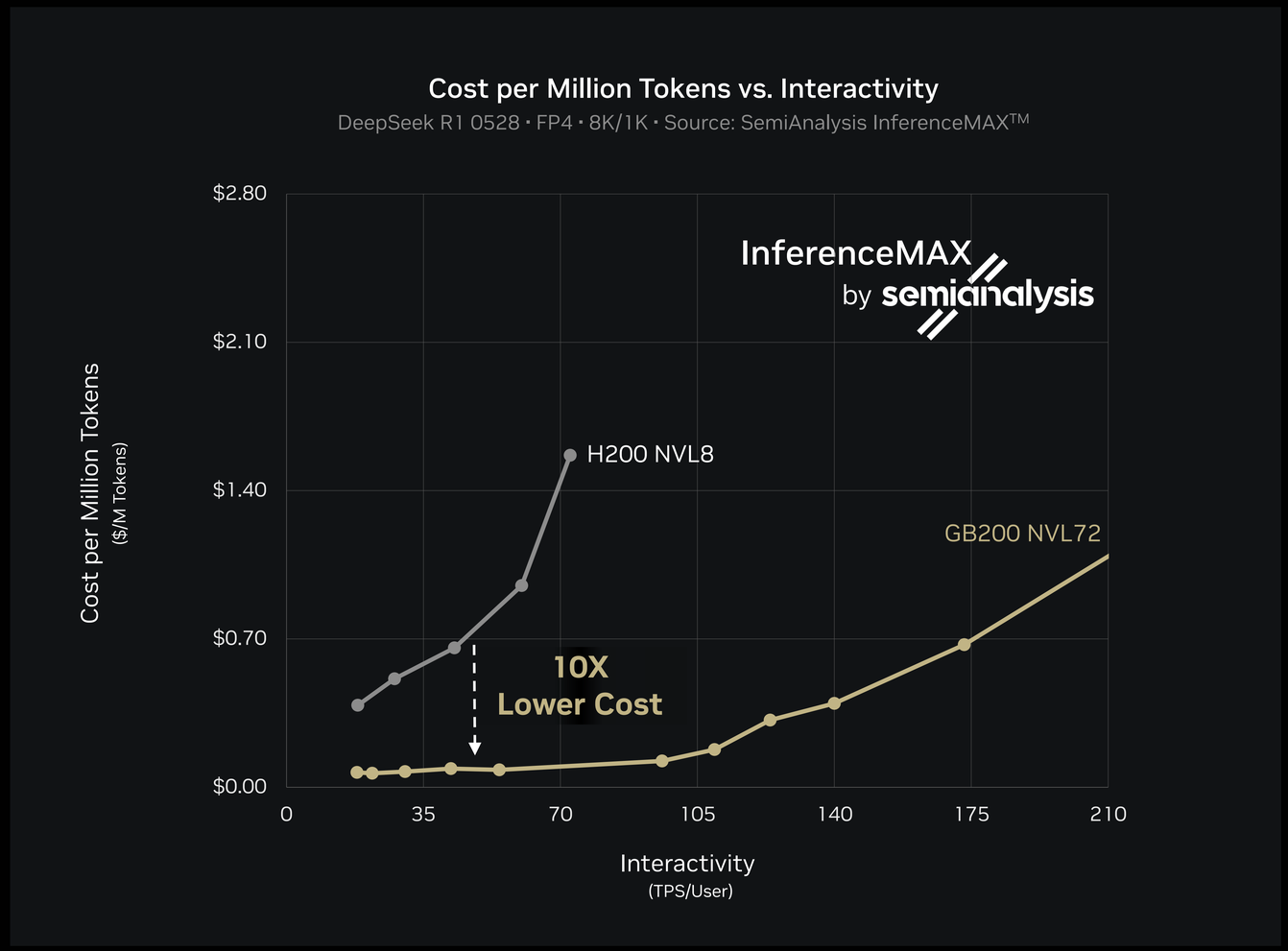
Extreme Co-Designed Blackwell NVL72
Game-Changer for Mixture-of-Experts Models

10X
Perf-per-Dollar

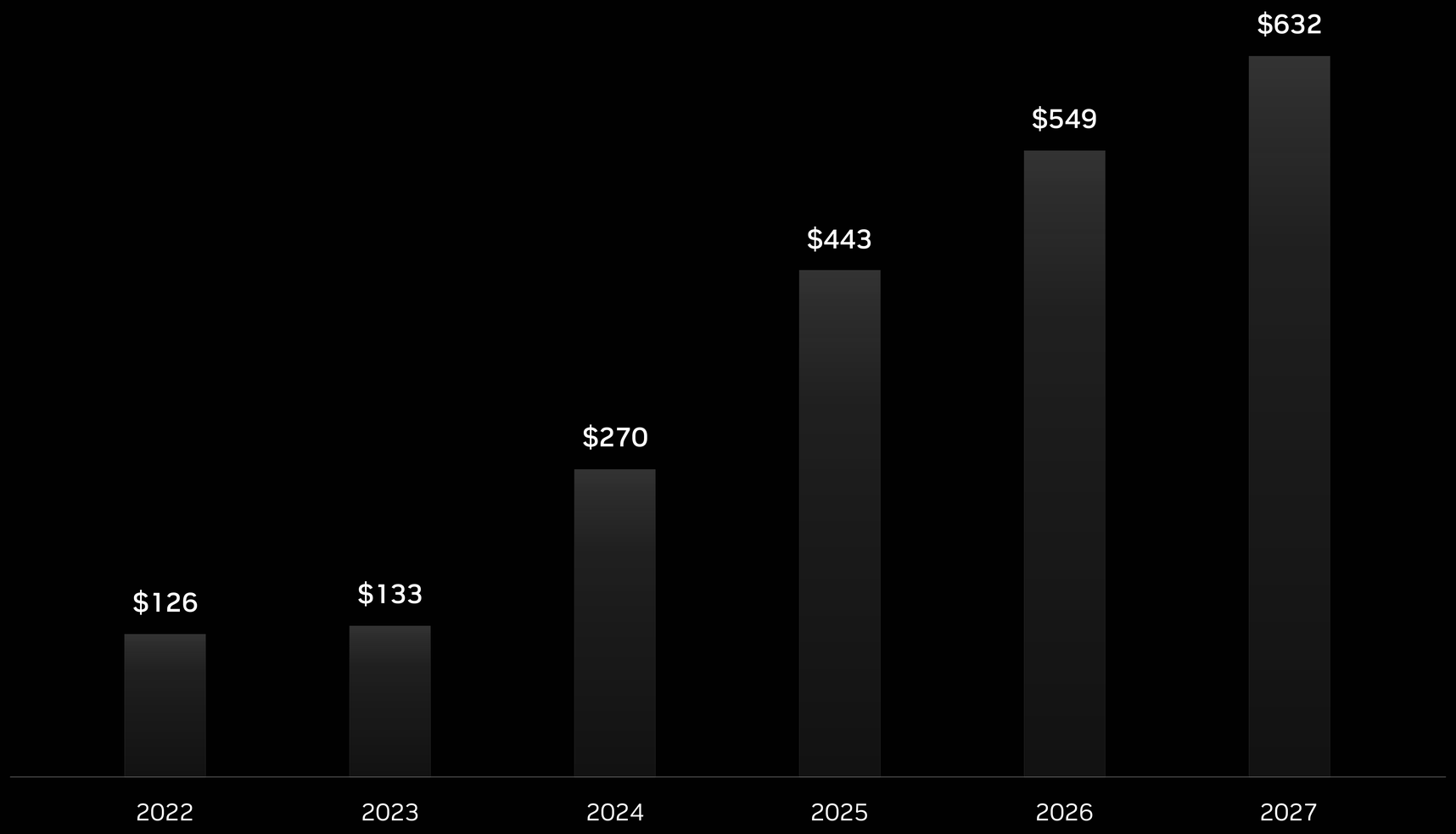
10X
Perf-per-Watt

10X
Throughput

10X
Revenues

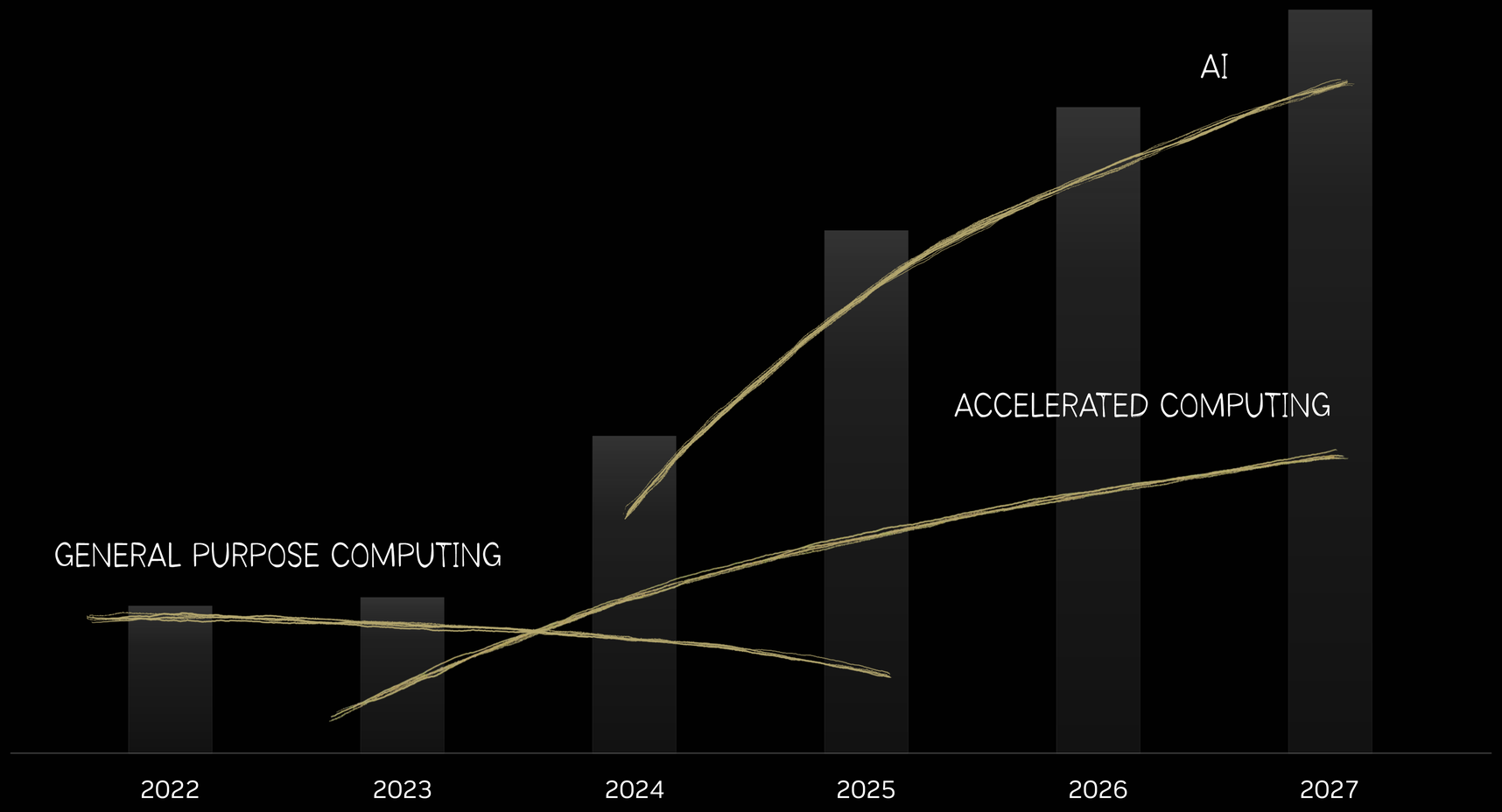


CSP CAPEX SPEND



Source: Morgan Stanley Research
Amazon, CoreWeave, Google, Meta, Microsoft, Oracle

CSP CAPEX SPEND

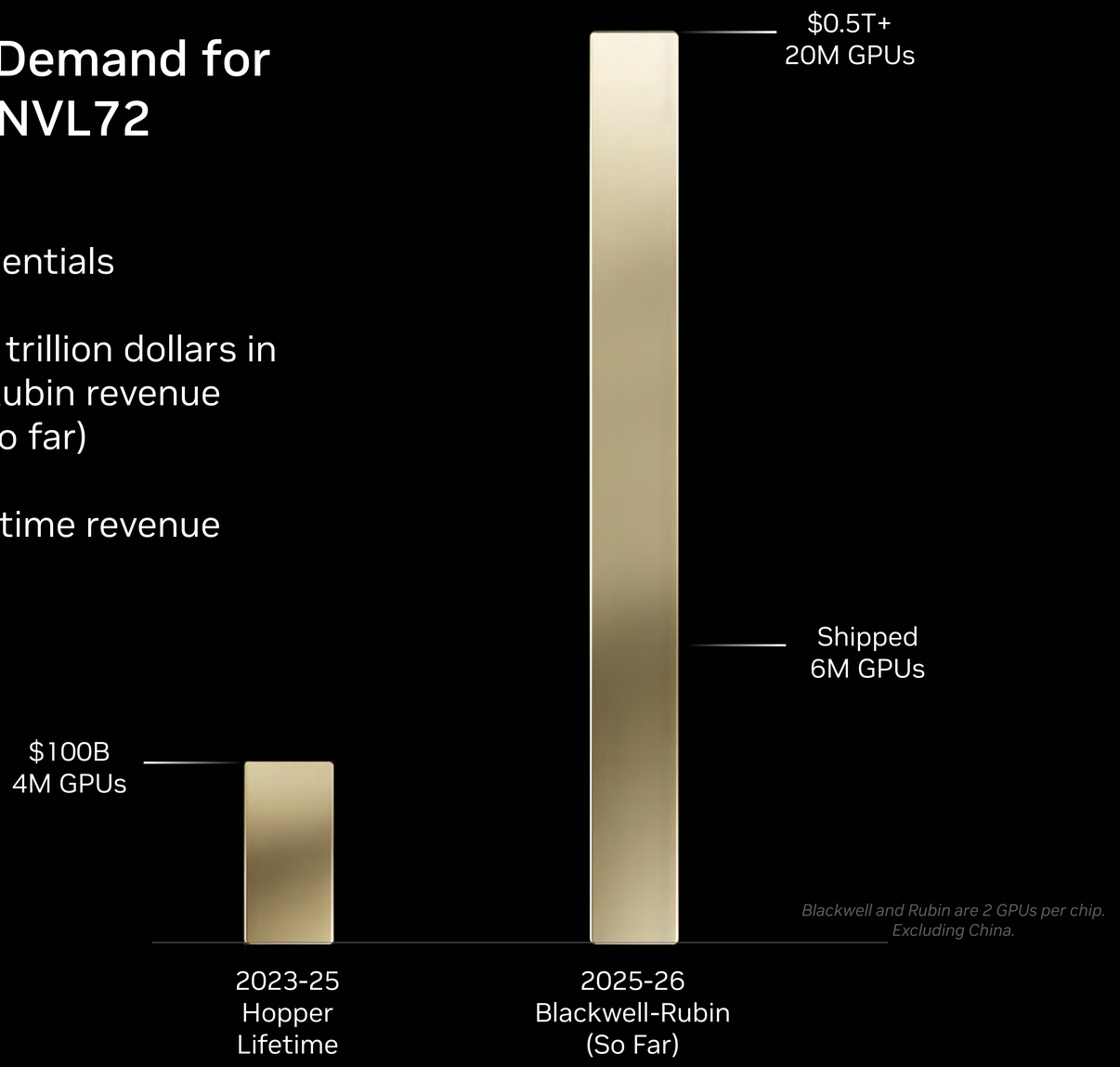


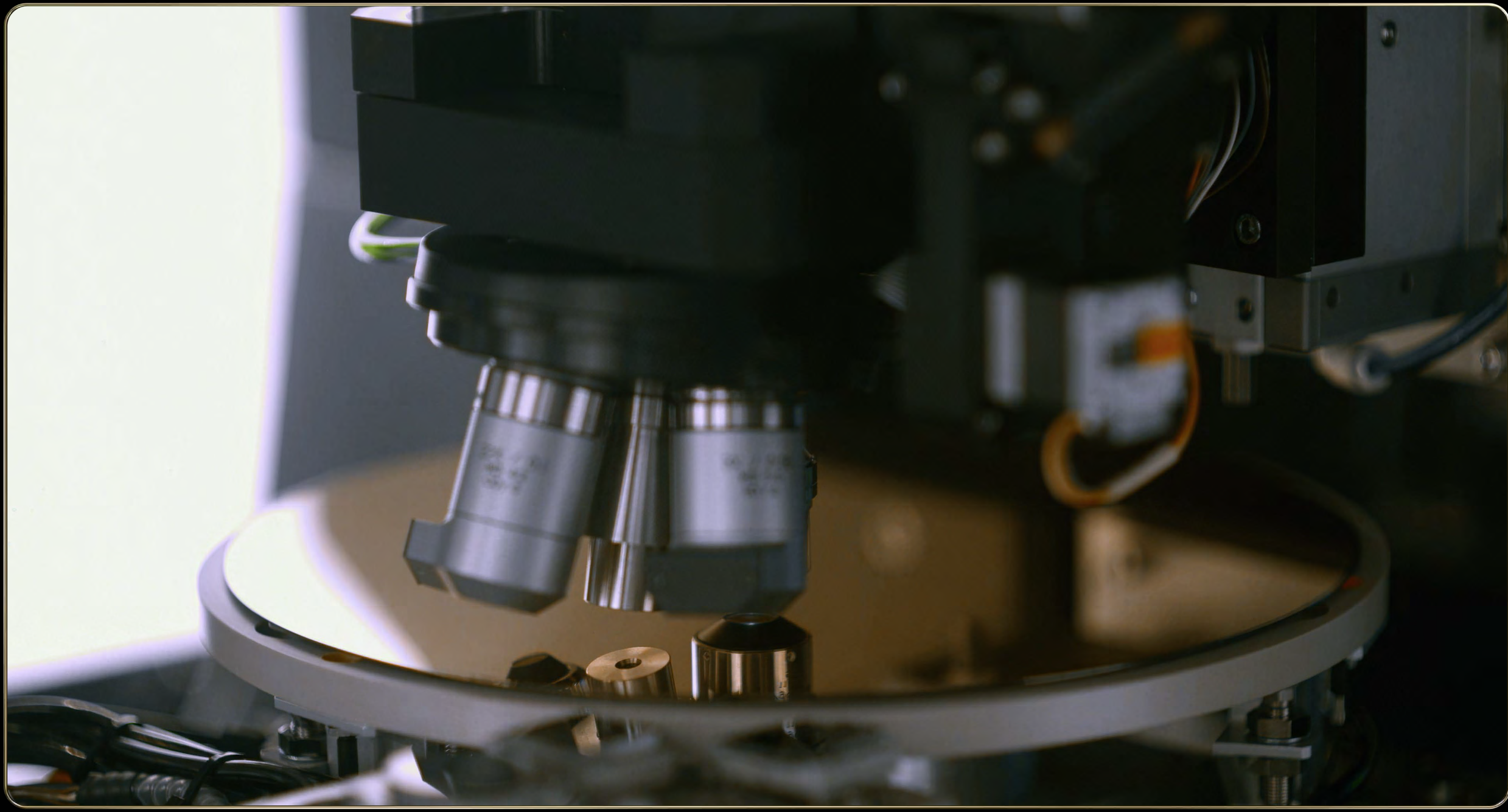
Exceptionally Strong Demand for Grace Blackwell NVL72

Driven by Two Exponentials

Visibility into more than half a trillion dollars in cumulative Blackwell and Rubin revenue through CY2026 (so far)

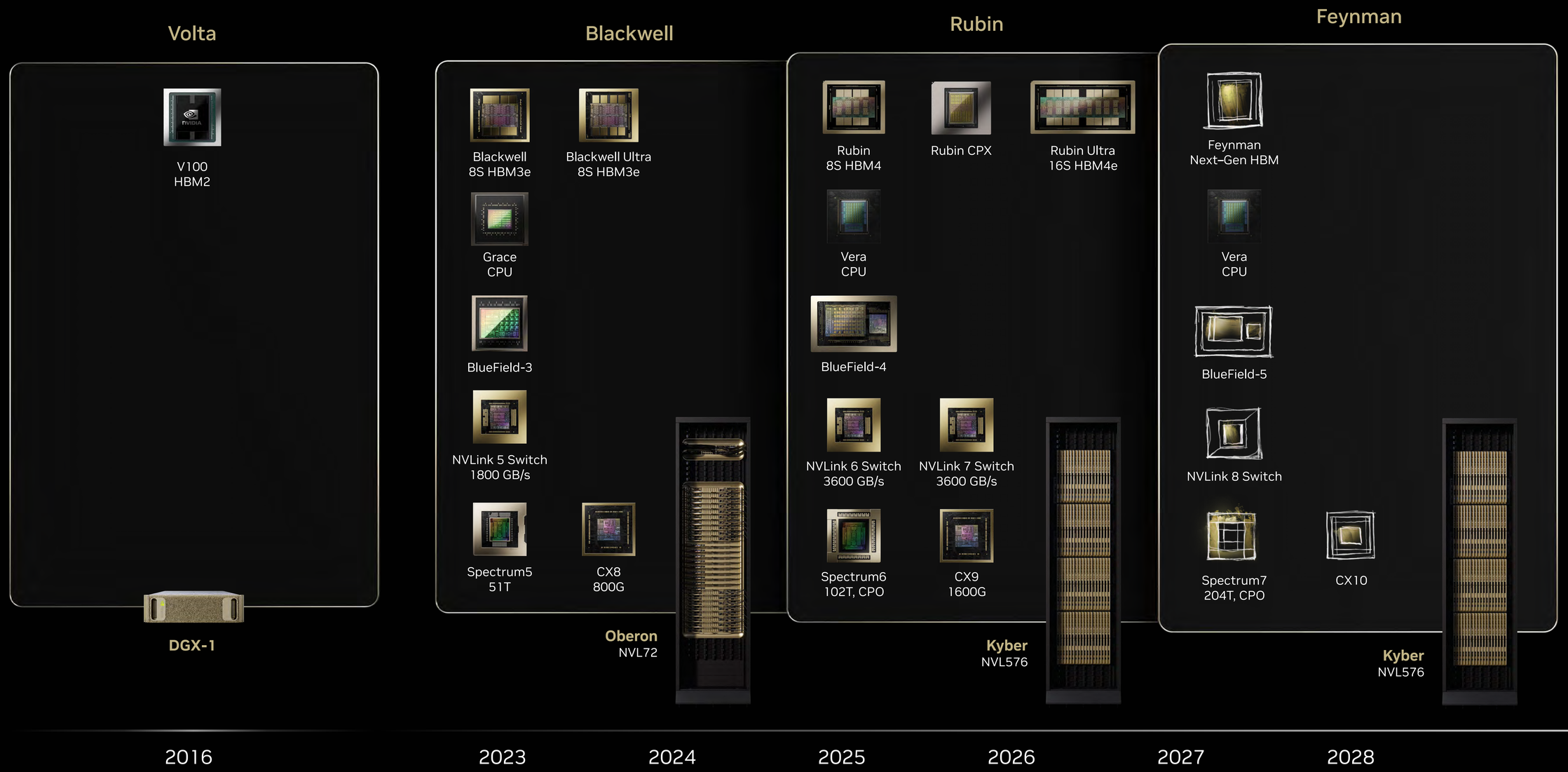
More than 5X Hopper's lifetime revenue

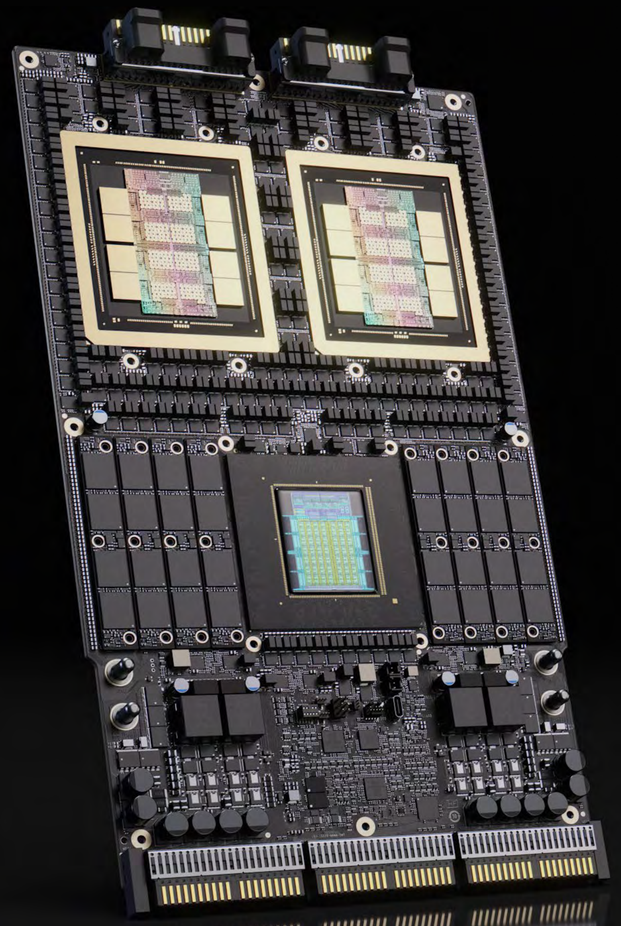




NVIDIA Extreme Co-Design Delivers X-Factors on One-Year Rhythm

Full-Stack | One Architecture | CUDA Everywhere





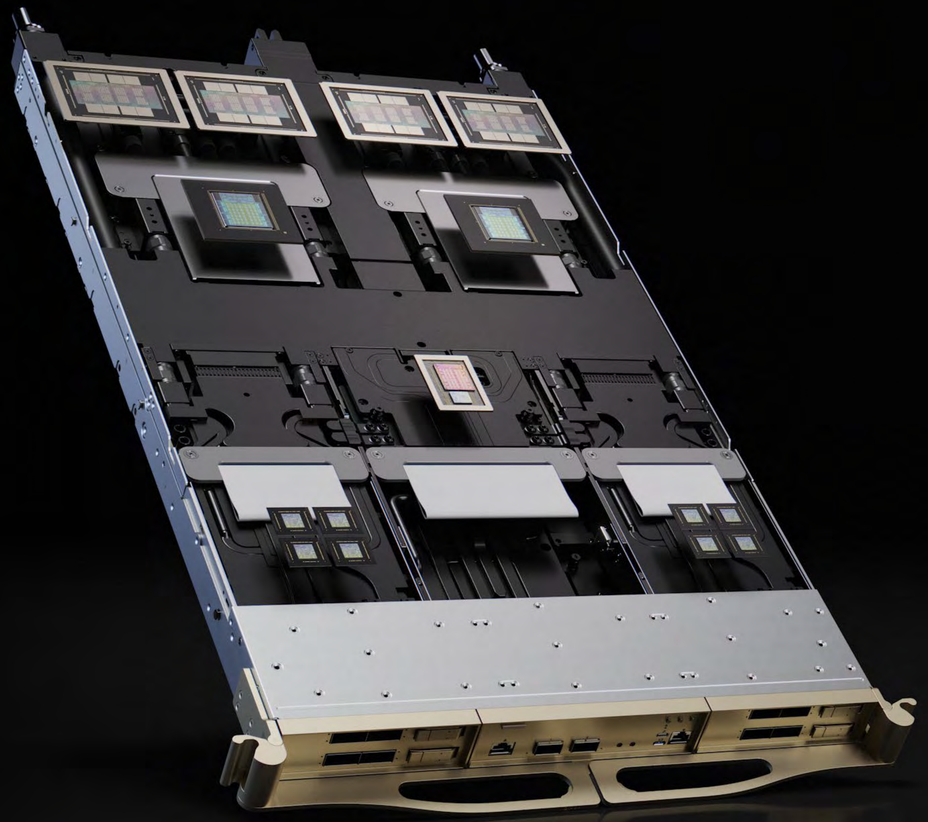
Vera Rubin Superchip

Processor for Gigascale AI Factories

100 PF AI with 88 Custom Arm Cores

2 TB Fast Memory

6 Trillion Transistors



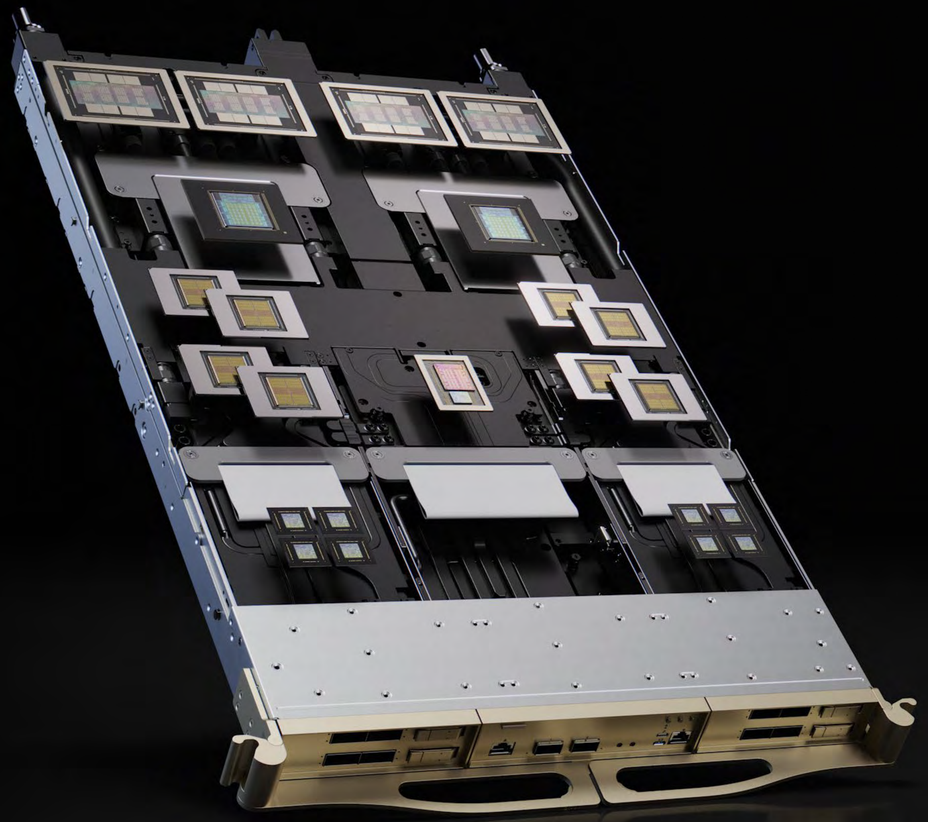
Vera Rubin Compute Tray

200 PF NVL72 Building Block

14.4 TB/s NVLink | 800 GB/s ConnectX-9

100% Liquid-Cooled

12 Trillion Transistors



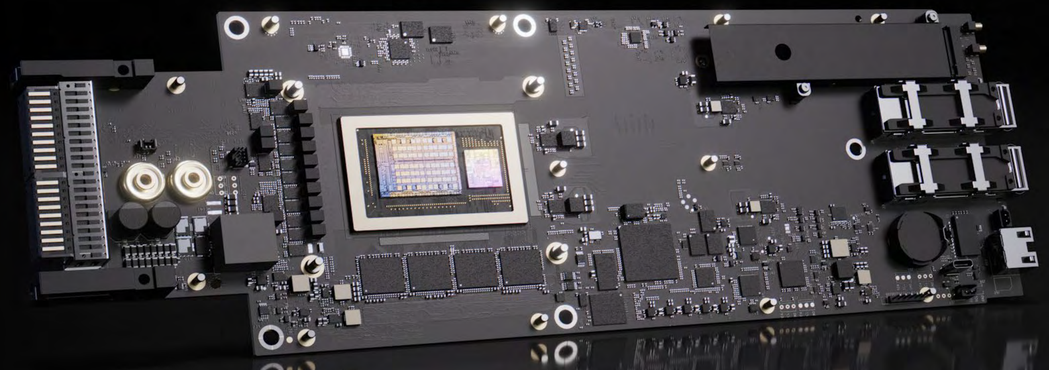
Vera Rubin CPX Compute Tray

1M+ Token Context Accelerator

440 PF Inference Performance

8 Additional Rubin CPX GPUs

Converged and CPX-Only Configurations



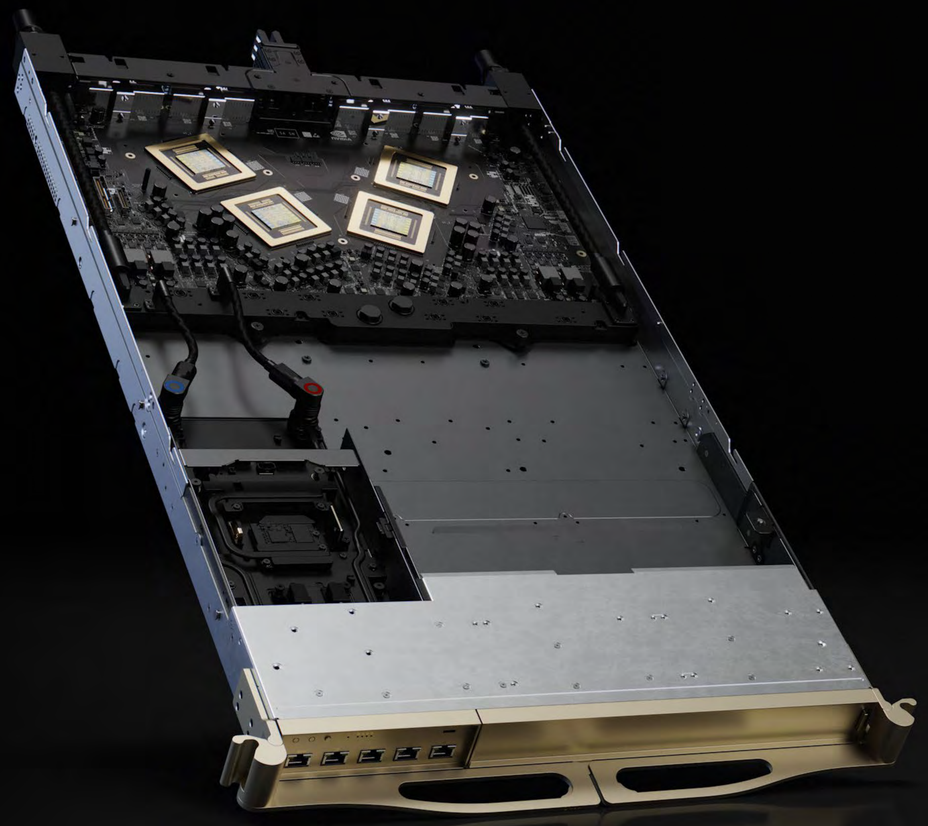
Announcing NVIDIA BlueField-4

800G SmartNIC for AI Factories

64 Core Grace CPU with ConnectX-9

AI Data Storage Acceleration

126 Billion Transistors



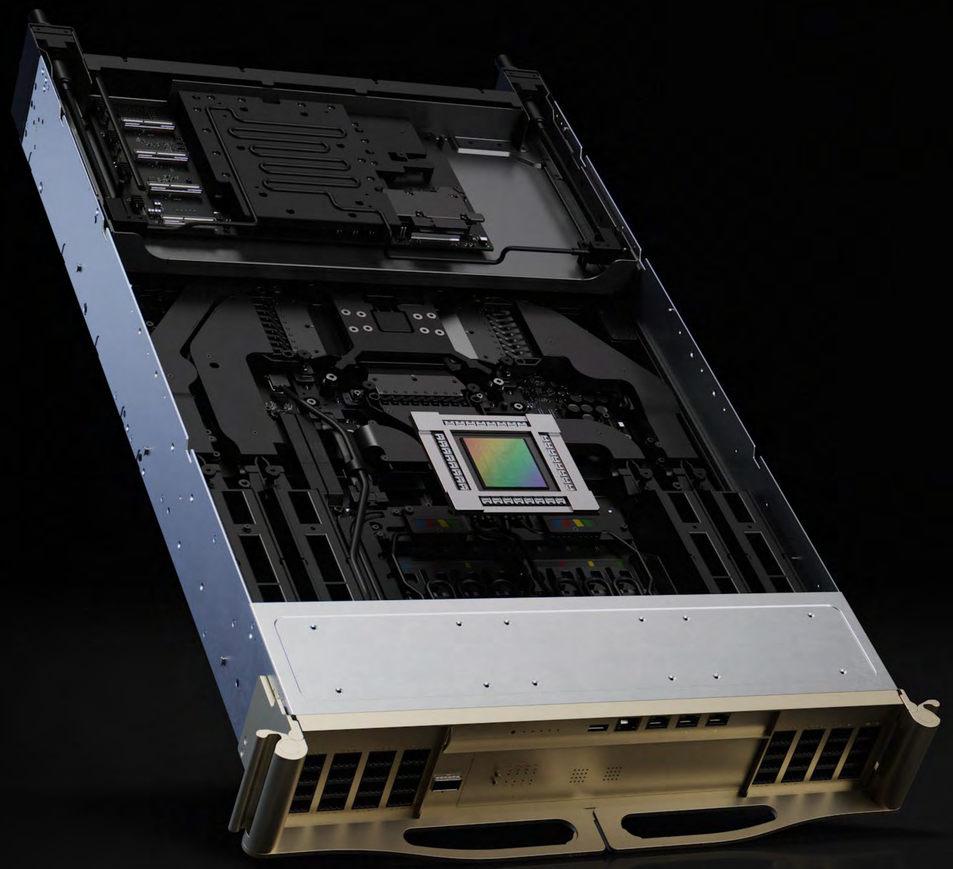
NVLink Switch Tray

Scale-Up Fabric with 3.6 TB/s Per-GPU All-to-All BW

NVLink6 400G Custom SerDes

28.8 TB/s Total Bandwidth

432 Billion Transistors



Spectrum-X Ethernet

102.4 TB/s Scale-Out Switch Infrastructure

Integrated 200G Silicon Photonics

95% Effective Bandwidth at Scale

352 Billion Transistors


NVIDIA

RTX PRO Server



Announcing NVIDIA Omniverse DSX Blueprint for Gigascale AI Factories

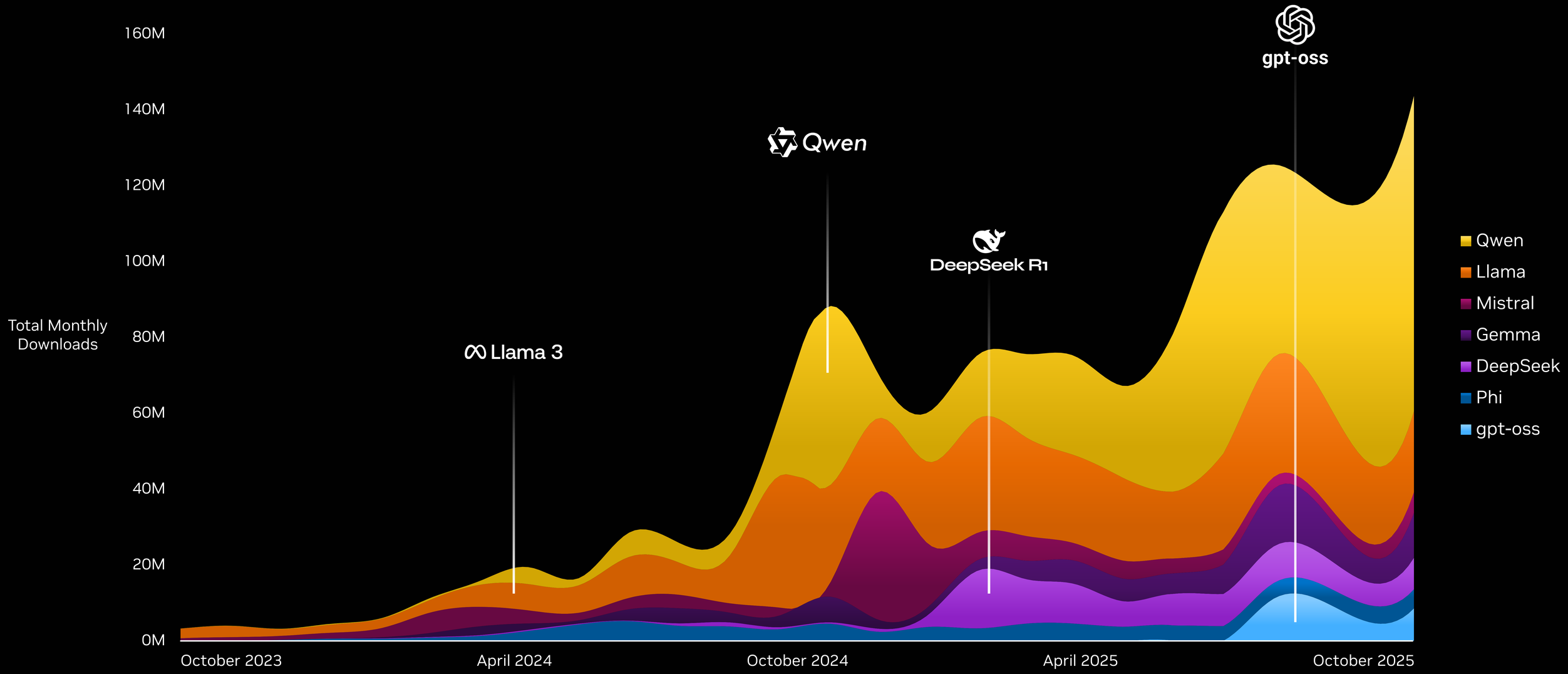


Omniverse | DSX | CUDA-X

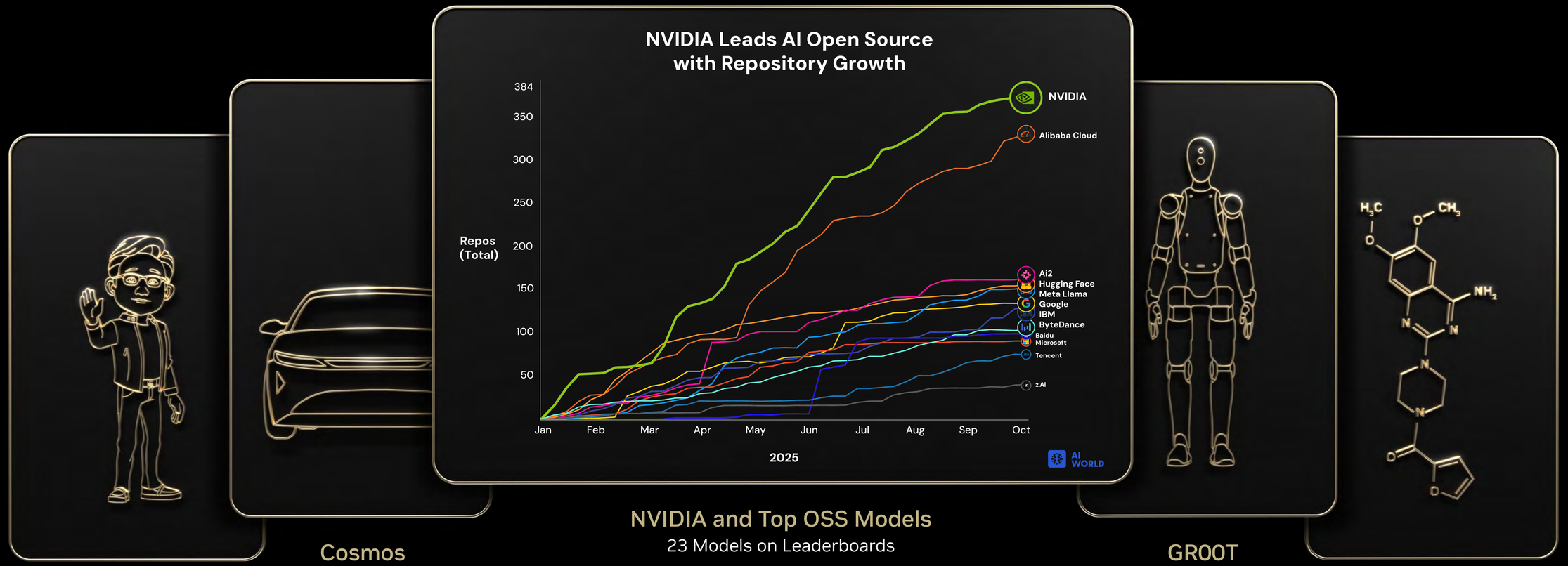
BECHTEL · cadence · DIGITAL REALTY · EATON · emeraldai · etap · GE VERNOVA · HITACHI

Jacobs · phaidra · ptc · SIEMENS · Schneider Electric · switch · TRANE TECHNOLOGIES · VERTIV

Developer Adoption of Open Models



NVIDIA Open Models, Data, Libraries Top Leaderboards



NVIDIA and Top OSS Models

23 Models on Leaderboards

Nemotron
Agentic AI

Cosmos
Physical AI

GR00T
Robotics

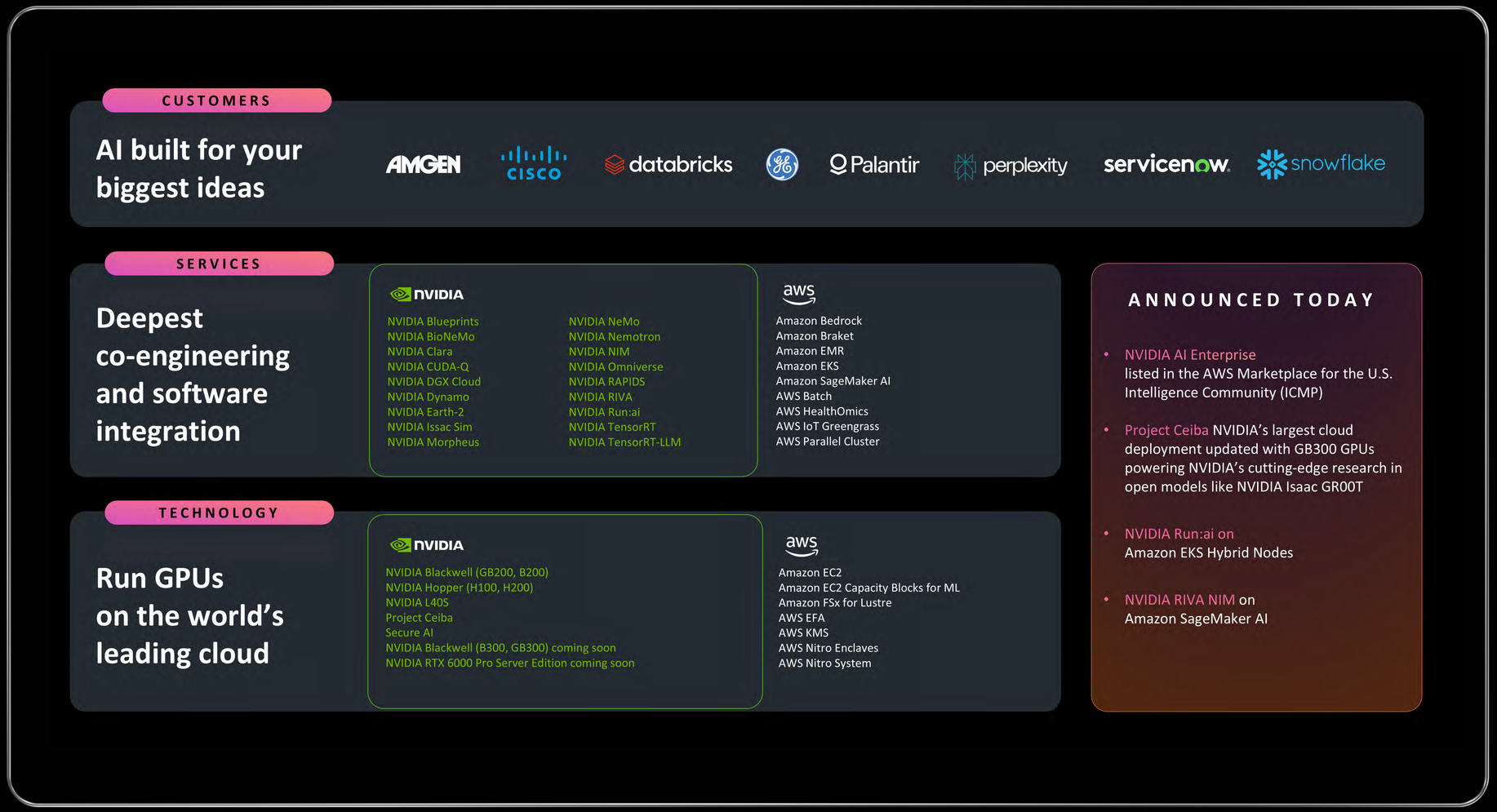
Clara
Biomedical AI

#1	#1	#1	#1	#1	#1	#1	#1
LiveCodeBench	AIME 2025	ViDoRe, MTEB, MMTEB	LLM Safety Leaderboard	OpenASR	PAIBench	MVPBench, IntPHys, CasualVQA	VLM3D

AI Startups Build on NVIDIA

ABRIDGE
Black Forest Labs
CodeRabbit
Cognition
CURSOR
Eleven Labs
exa
FieldAI
FIGURE
Fireworks AI
Generalist
Harvey
Isomorphic Labs
L I L A
Neura AI
NSCALE
nuro
OpenEvidence
perplexity
Physical Intelligence
poolside
Reflection
runway
SKILD AI
synthesia
THINKING MACHINES
together.ai
waabi
WAYVE
World Labs




CUSTOMERS
AI built for your biggest ideas
AMGEN
cisco
databricks
Palantir
perplexity
servicenow
snowflake
SERVICES
Deepest co-engineering and software integration
NVIDIA
NVIDIA Blueprints
NVIDIA BioNeMo
NVIDIA Clara
NVIDIA CUDA-Q
NVIDIA DGX Cloud
NVIDIA Dynamo
NVIDIA Earth-2
NVIDIA Issac Sim
NVIDIA Morpheus
NVIDIA NeMo
NVIDIA Nemotron
NVIDIA NIM
NVIDIA Omniverse
NVIDIA RAPIDS
NVIDIA RIVA
NVIDIA Run:ai
NVIDIA TensorRT
NVIDIA TensorRT-LLM
aws
Amazon Bedrock
Amazon Braket
Amazon EMR
Amazon EKS
Amazon SageMaker AI
AWS Batch
AWS HealthOmics
AWS IoT Greengrass
AWS Parallel Cluster
TECHNOLOGY
Run GPUs on the world's leading cloud
NVIDIA
NVIDIA Blackwell (GB200, B200)
NVIDIA Hopper (H100, H200)
NVIDIA L40S
Project Ceiba
Secure AI
NVIDIA Blackwell (B300, GB300) coming soon
NVIDIA RTX 6000 Pro Server Edition coming soon
aws
Amazon EC2
Amazon EC2 Capacity Blocks for ML
Amazon FSx for Lustre
AWS EFA
AWS KMS
AWS Nitro Enclaves
AWS Nitro System
ANNOUNCED TODAY
NVIDIA AI Enterprise listed in the AWS Marketplace for the U.S. Intelligence Community (ICMP)
Project Ceiba NVIDIA's largest cloud deployment updated with GB300 GPUs powering NVIDIA's cutting-edge research in open models like NVIDIA Isaac GR00T
NVIDIA Run:ai on Amazon EKS Hybrid Nodes
NVIDIA RIVA NIM on Amazon SageMaker AI

Google Cloud | NVIDIA

Deep engineering collaboration powers every level of the AI stack

AI-Optimized hardware

Cutting-edge performance, planet-scale deployment, and robust security.



G4 VMs
Powered by NVIDIA RTX PRO 6000



A4X Max VMs
Powered by NVIDIA GB300 NVL72

Leading software, open frameworks

Democratizing how AI gets built with developer choice, openness and stack control.

New

Vertex AI Training with NeMo Framework

GKE Inference Gateway and NeMo Guardrails

Google Cloud Recipes for NVIDIA Dynamo

AI anywhere

Sovereign cloud, on-premises, air-gapped environments, and the edge.

New

Gemini on Google Distributed Cloud with NVIDIA Blackwell

Industry breakthroughs

AI innovations for areas such as drug discovery, physics, robotics, and energy grid.

New

NVIDIA Omniverse on Google Cloud with G4 VMs

Microsoft Azure | NVIDIA

Ecosystem Partnership on Azure

amdocs · Epic · esri · servicenow

Accelerated Azure Services

Microsoft Azure · Microsoft Bing · Azure AI Foundry · Copilot

NVIDIA libraries integration throughout Azure

- NVIDIA Dynamo Inference Server
- NVIDIA TensorRT-LLM
- NVIDIA AI Enterprise
- NVIDIA NIM
- NVIDIA cuVS
- NVIDIA Omniverse on DGX Cloud



NVIDIA Blackwell Ultra
NVIDIA Blackwell



NVIDIA
RTX PRO



NVIDIA
Hopper



NVIDIA Networking
Quantum InfiniBand

New at GTC:

NVIDIA Nemotron and Cosmos Models available in Azure AI Foundry

Run.ai available on Azure Marketplace

NVIDIA Dynamo available on Azure Kubernetes Service

NVIDIA RTX PRO 6000 Blackwell Server Edition on Azure Local

ORACLE | NVIDIA

New Announcements:

Zettascale Supercluster

- 800,000 NVIDIA GPUs
- 16 ZettaFLOPS of AI

Oracle Database 26ai

- Using NVIDIA NeMo Retriever and NVIDIA NIM to build RAG pipelines

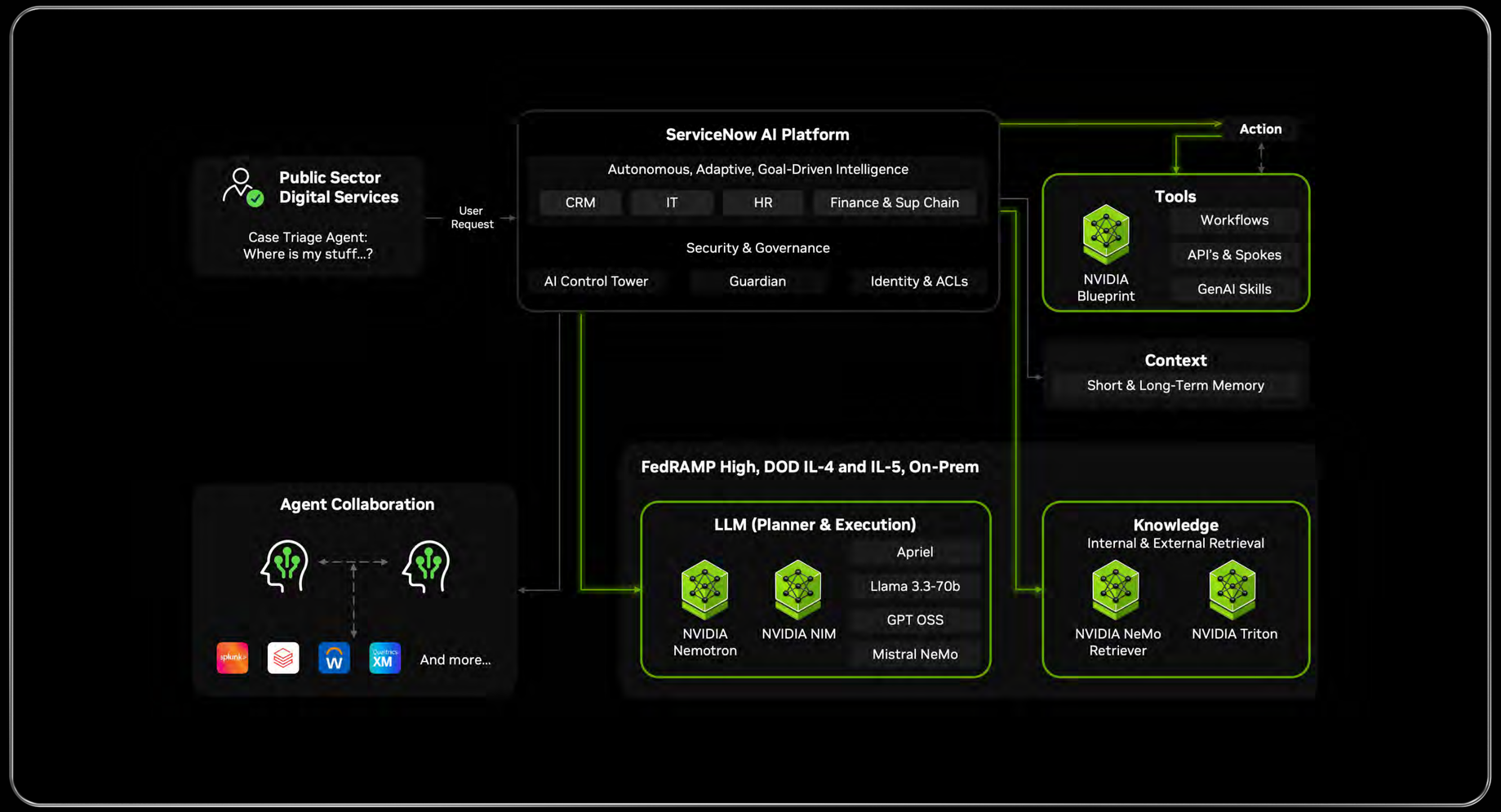
Oracle AI Data Platform

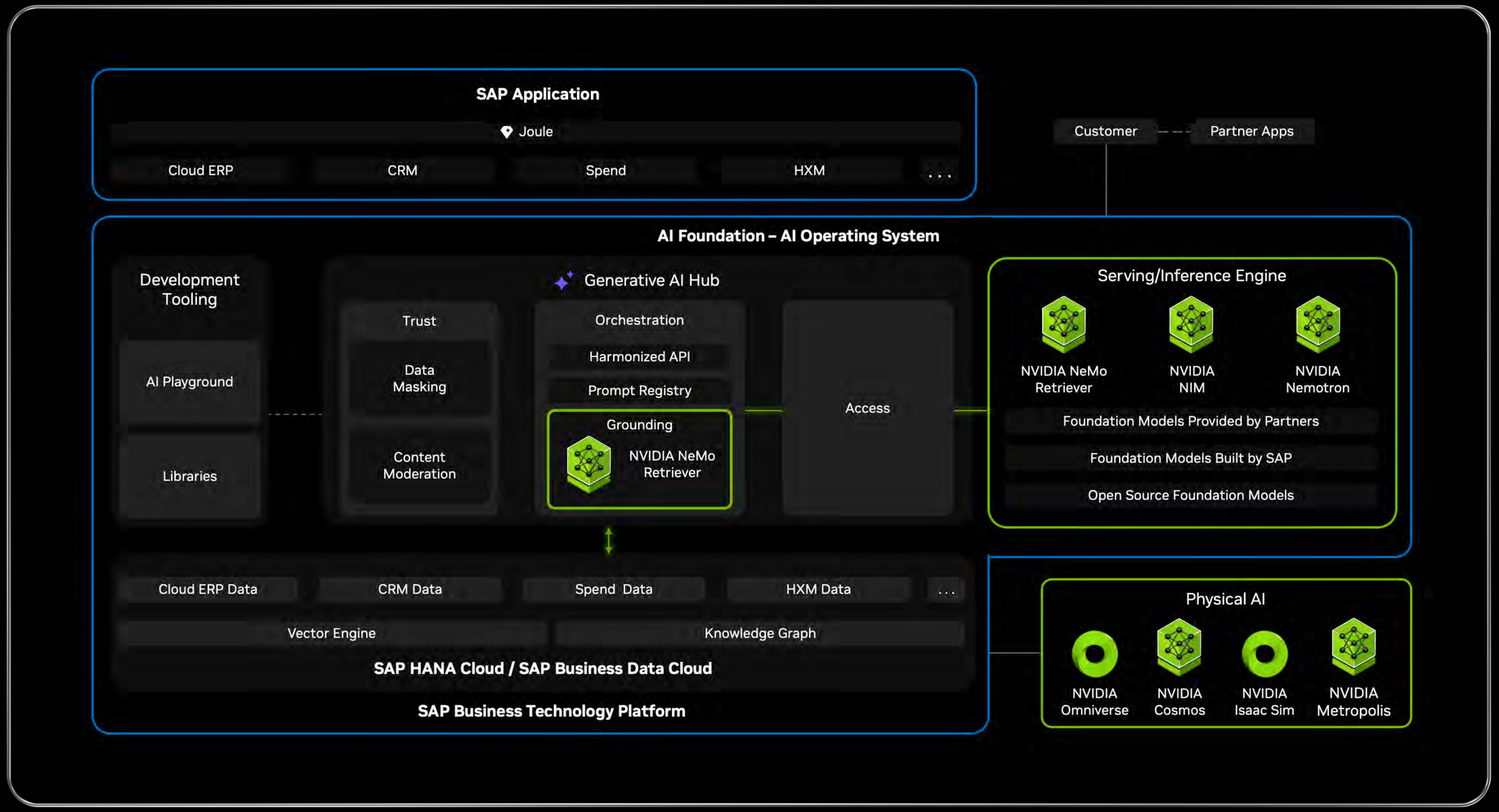
- NVIDIA RAPIDS for Apache Spark
- Faster analytics, ETL, and machine learning

ALTAIR | BEAMR | cohere | NRI | Stanford University | Uber | zoom

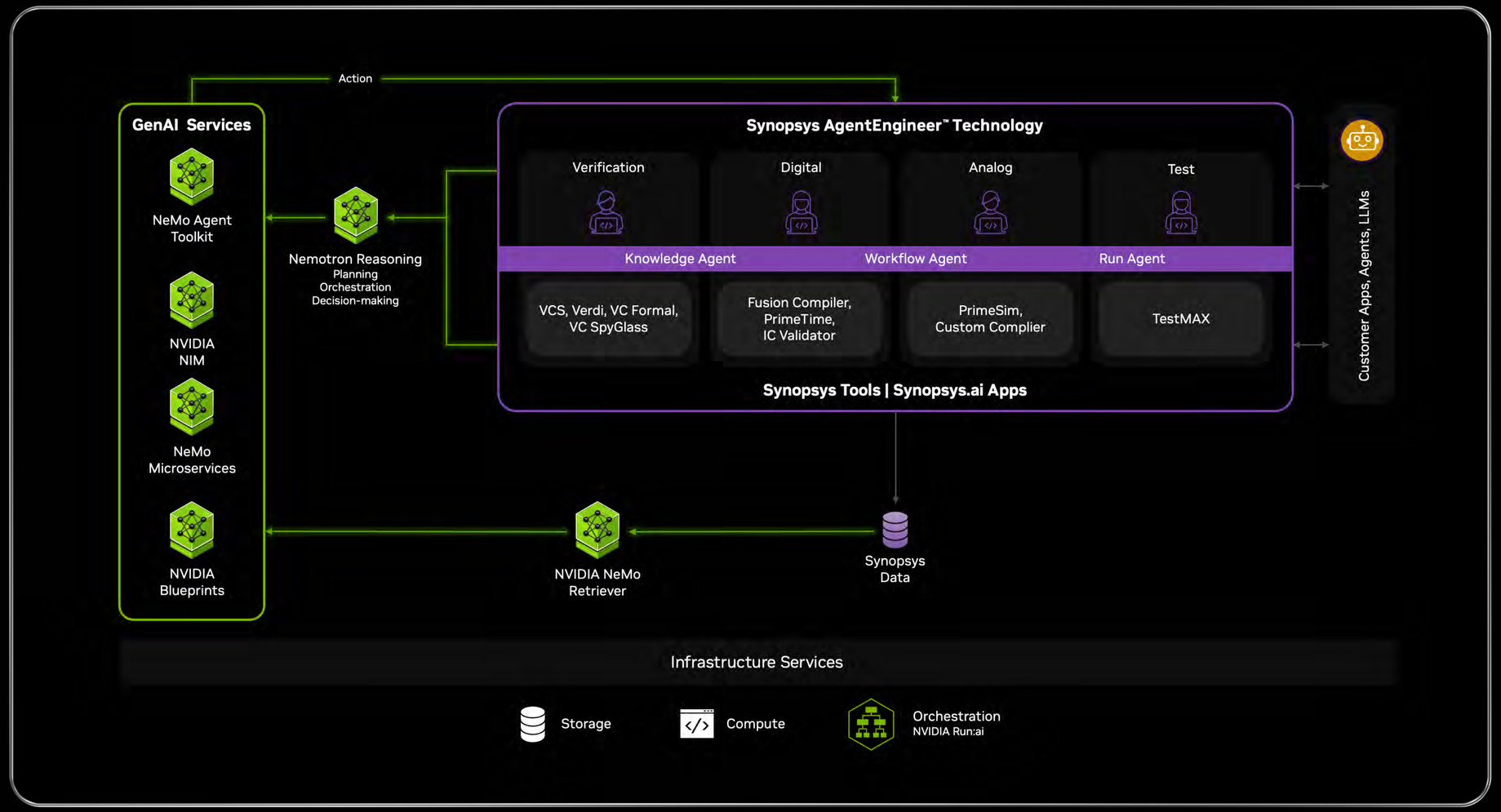
Oracle Cloud Infrastructure	Oracle AI Database, Autonomous, Vector	OCI Live Labs; LaunchPad	OCI Data Science	OCI Superclusters	OCI OKE
NVIDIA AI Enterprise	NVIDIA NIM, NeMo Retriever	NVIDIA DGX Cloud	NVIDIA TensorRT, TensorRT-LLM	NVIDIA Blueprints	NVIDIA NeMo	NVIDIA Dynamo Inference Server
 NVIDIA Blackwell	 NVIDIA Hopper	 NVIDIA Grace Hopper	 NVIDIA Networking Spectrum-X, InfiniBand, CX7, BF3	Sovereign AI DRCC, Gov. Cloud, EU Cloud	

servicenow | NVIDIA

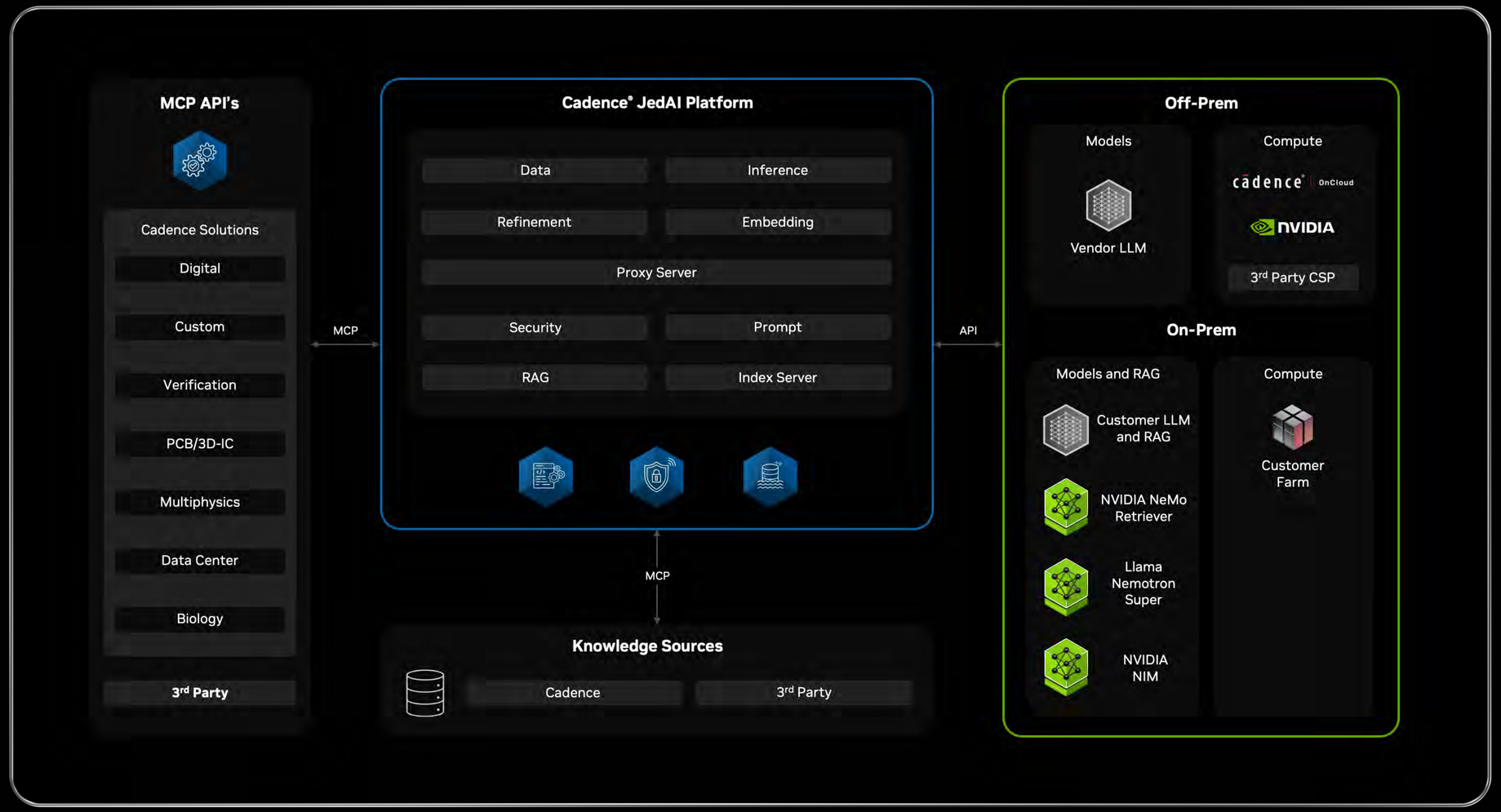



SAP Application
Joule
Cloud ERP
CRM
Spend
HXM
...
Customer
Partner Apps
AI Foundation – AI Operating System
Development Tooling
AI Playground
Libraries
Generative AI Hub
Trust
Data Masking
Content Moderation
Orchestration
Harmonized API
Prompt Registry
Grounding
NVIDIA NeMo Retriever
Access
Serving/Inference Engine
NVIDIA NeMo Retriever
NVIDIA NIM
NVIDIA Nemotron
Foundation Models Provided by Partners
Foundation Models Built by SAP
Open Source Foundation Models
Cloud ERP Data
CRM Data
Spend Data
HXM Data
...
Vector Engine
Knowledge Graph
SAP HANA Cloud / SAP Business Data Cloud
SAP Business Technology Platform
Physical AI
NVIDIA Omniverse
NVIDIA Cosmos
NVIDIA Isaac Sim
NVIDIA Metropolis

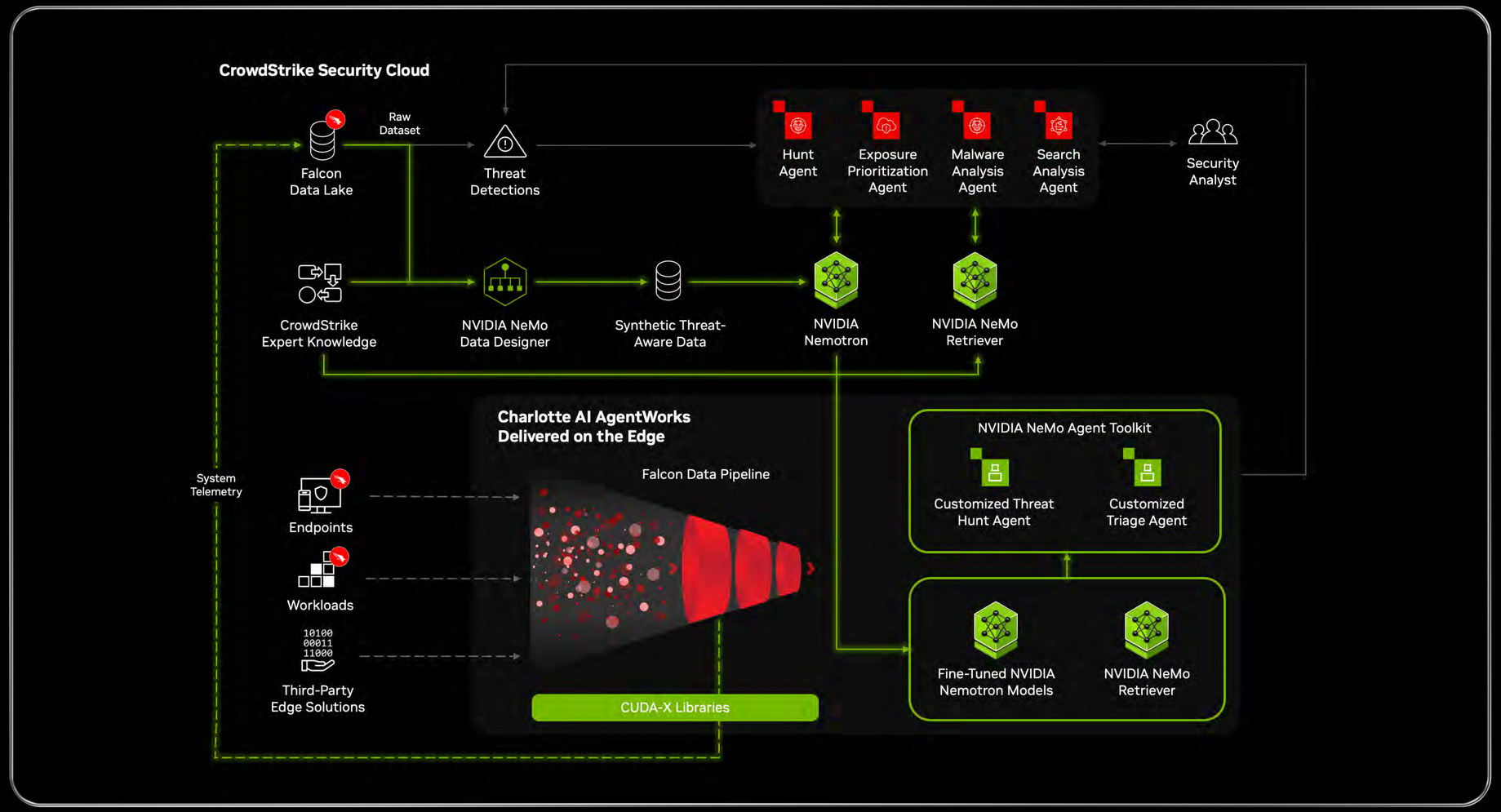
SYNOPSYS® | NVIDIA



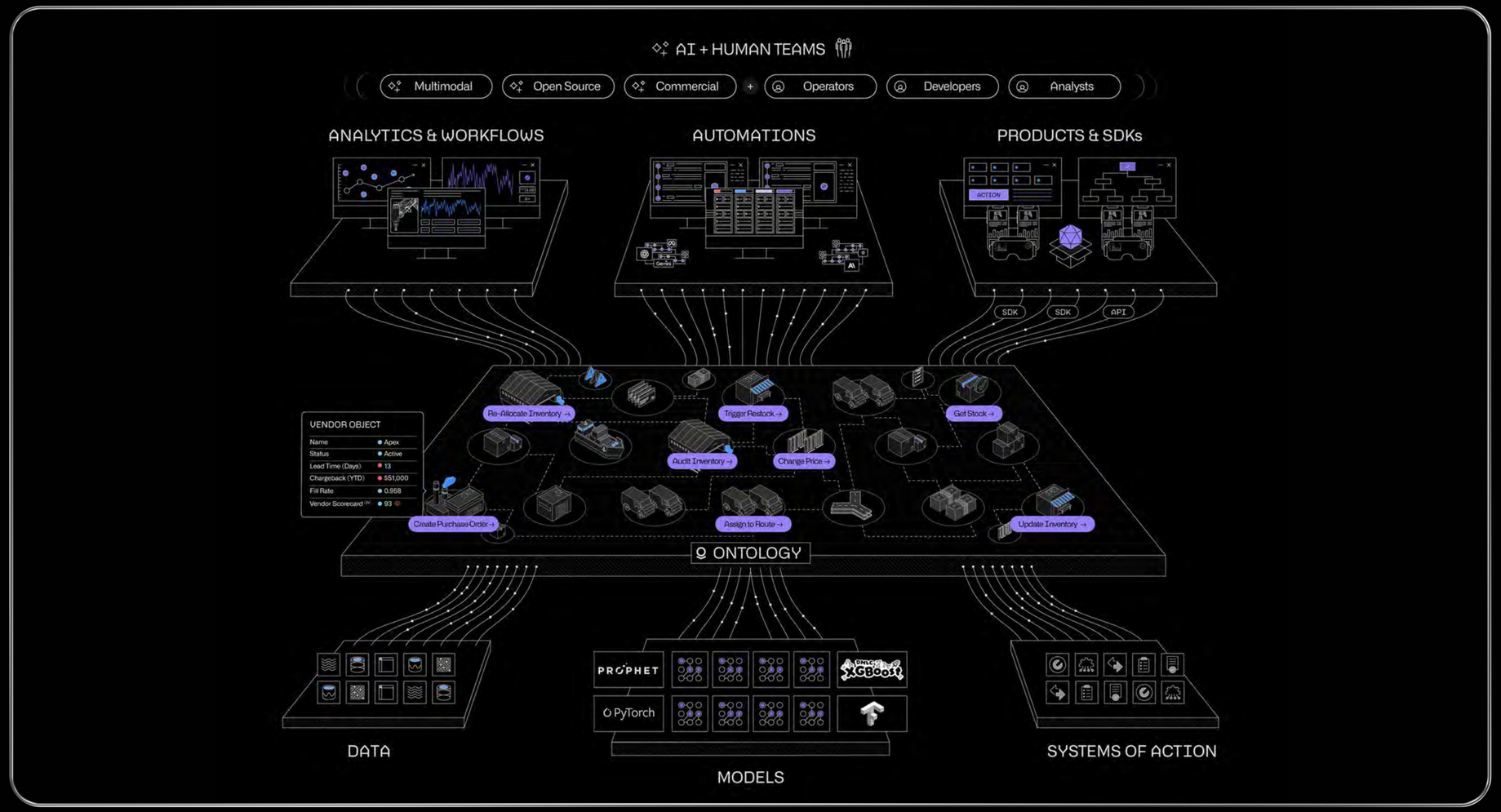
cadence® | NVIDIA



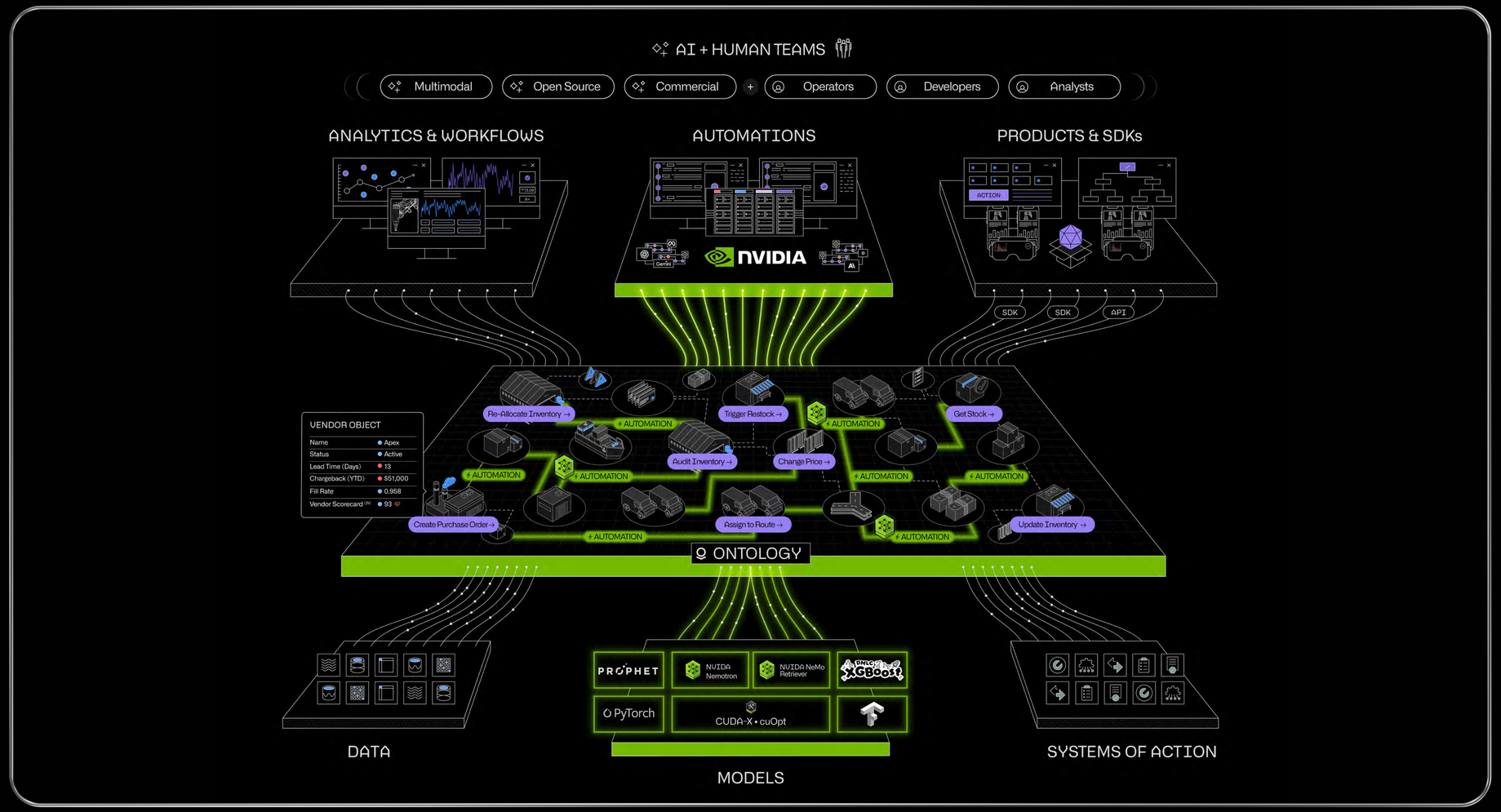
CROWDSTRIKE | NVIDIA



Palantir | NVIDIA



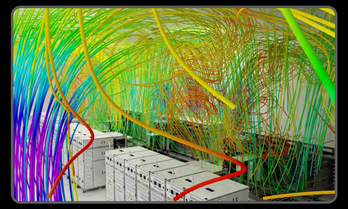
Palantir | NVIDIA



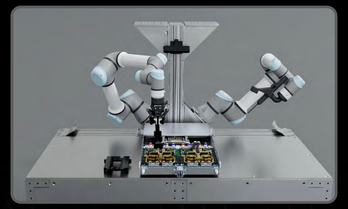
NVIDIA and Global Leaders Build Digital Twin Platform for US Reindustrialization



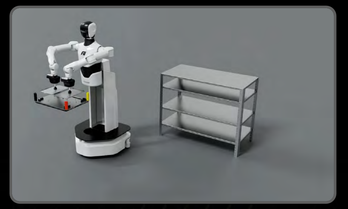
SIEMENS



cadence



SKILD AI



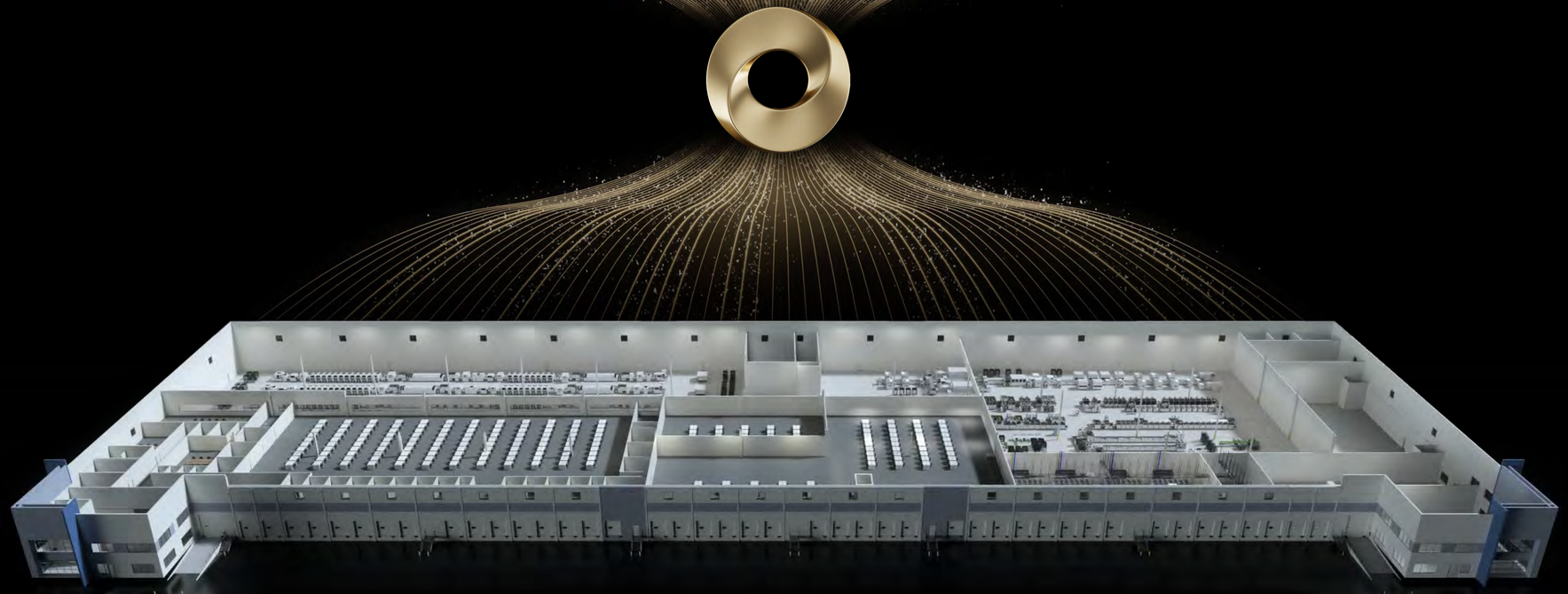
Foxconn Industrial Internet



FANUC



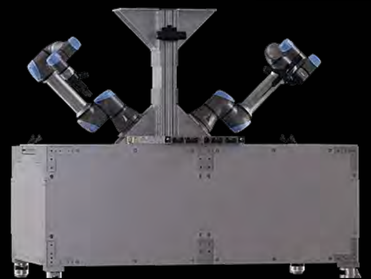
Foxconn Industrial Internet



NVIDIA and Global Leaders Build Digital Twin Platform for US Reindustrialization



SKILD AI



Foxconn Industrial Internet



FANUC



Foxconn Industrial Internet




Front Camera
Rear Camera
Left Camera
Right Camera
Front Camera
Rear Camera
Left Camera
Right Camera
C03
C04
C02

RTX PRO Server



Announcing Foxconn Builds New Blackwell Factory in Texas



Omniverse | Isaac Sim | Cosmos | Metropolis

FOXCONN® HON HAI TECHNOLOGY GROUP | NVIDIA

cadence FANUC Foxconn Industrial Internet SIEMENS SKILD AI

RTX PRO Server

Caterpillar Builds Factory Digital Twin on NVIDIA



Omniverse | Mega

CATERPILLAR | NVIDIA

GB200
RTX PRO Server

Jetson Thor

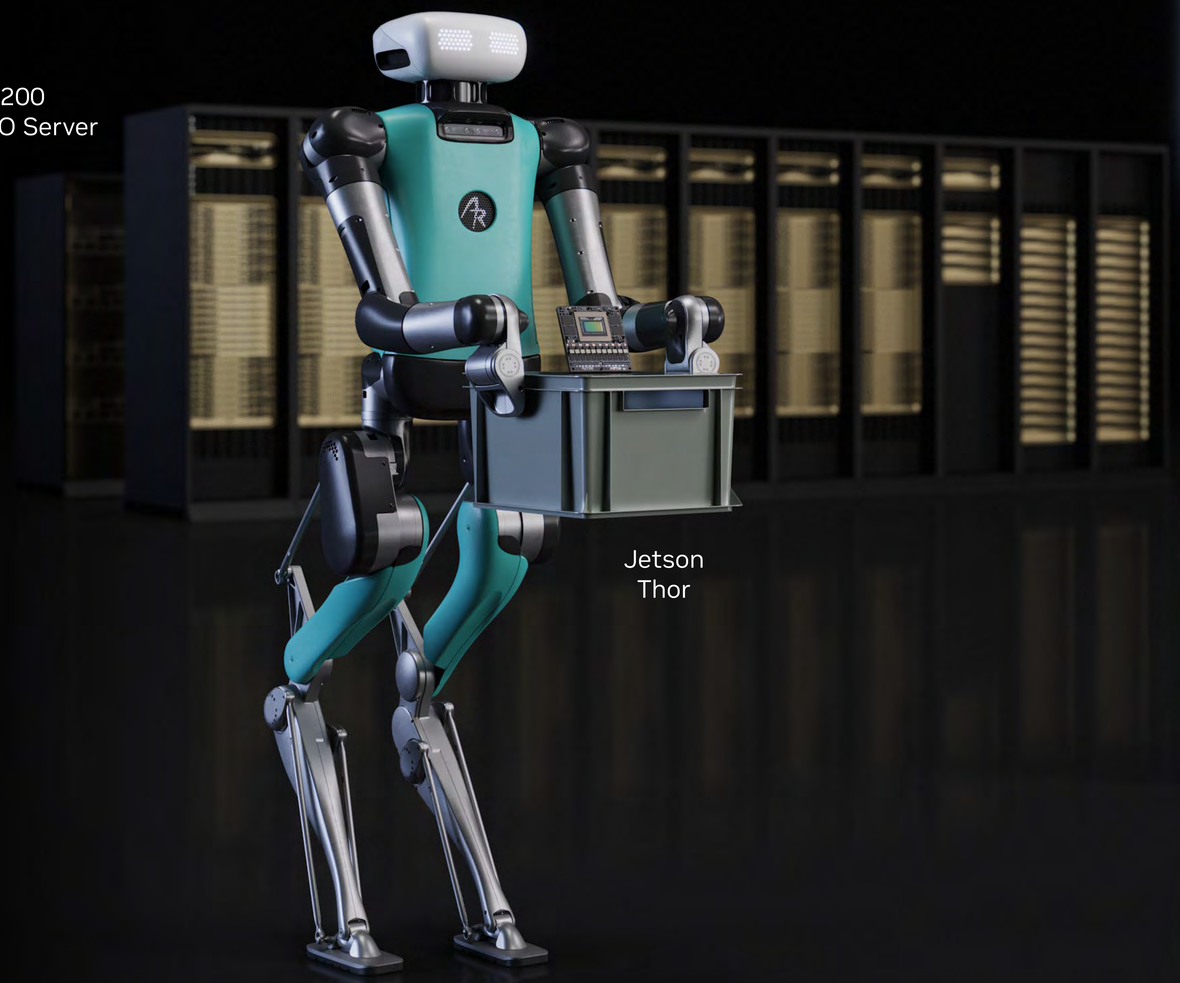
Figure Builds Humanoid Robots on NVIDIA



Omniverse | Isaac Lab | Cosmos

FIGURE | NVIDIA

GB200
RTX PRO Server



Jetson
Thor

Agility Robotics Builds Humanoid Robots on NVIDIA



Omniverse | Isaac Sim | Isaac Lab

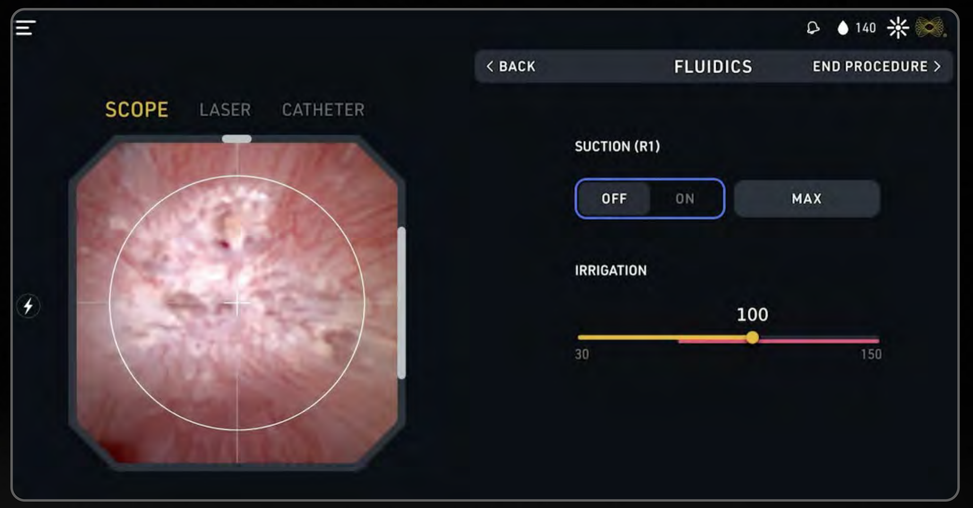
AGILITY ROBOTICS | NVIDIA

Johnson & Johnson Builds Medical Robots on NVIDIA

RTX PRO Server

IGX Thor

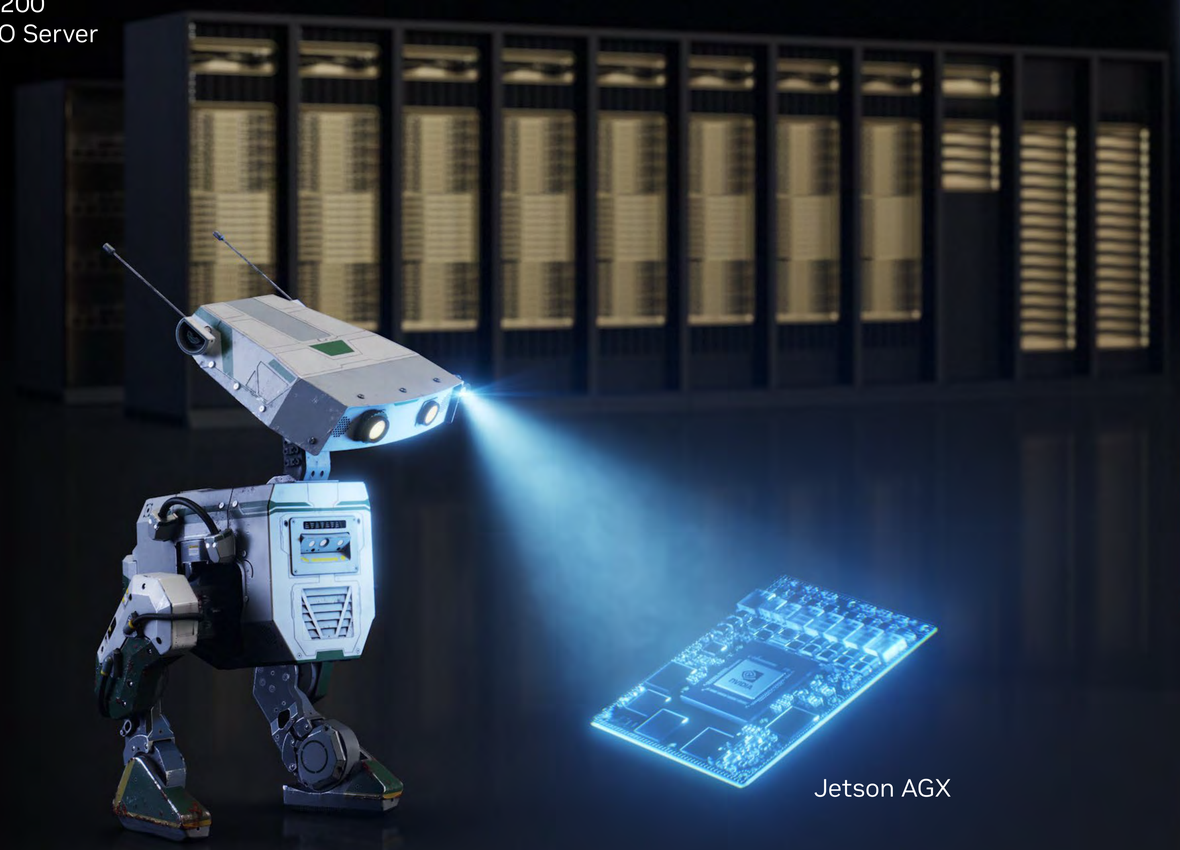




Omniverse | Isaac for Healthcare | Cosmos

Johnson&Johnson | NVIDIA

GB200
RTX PRO Server



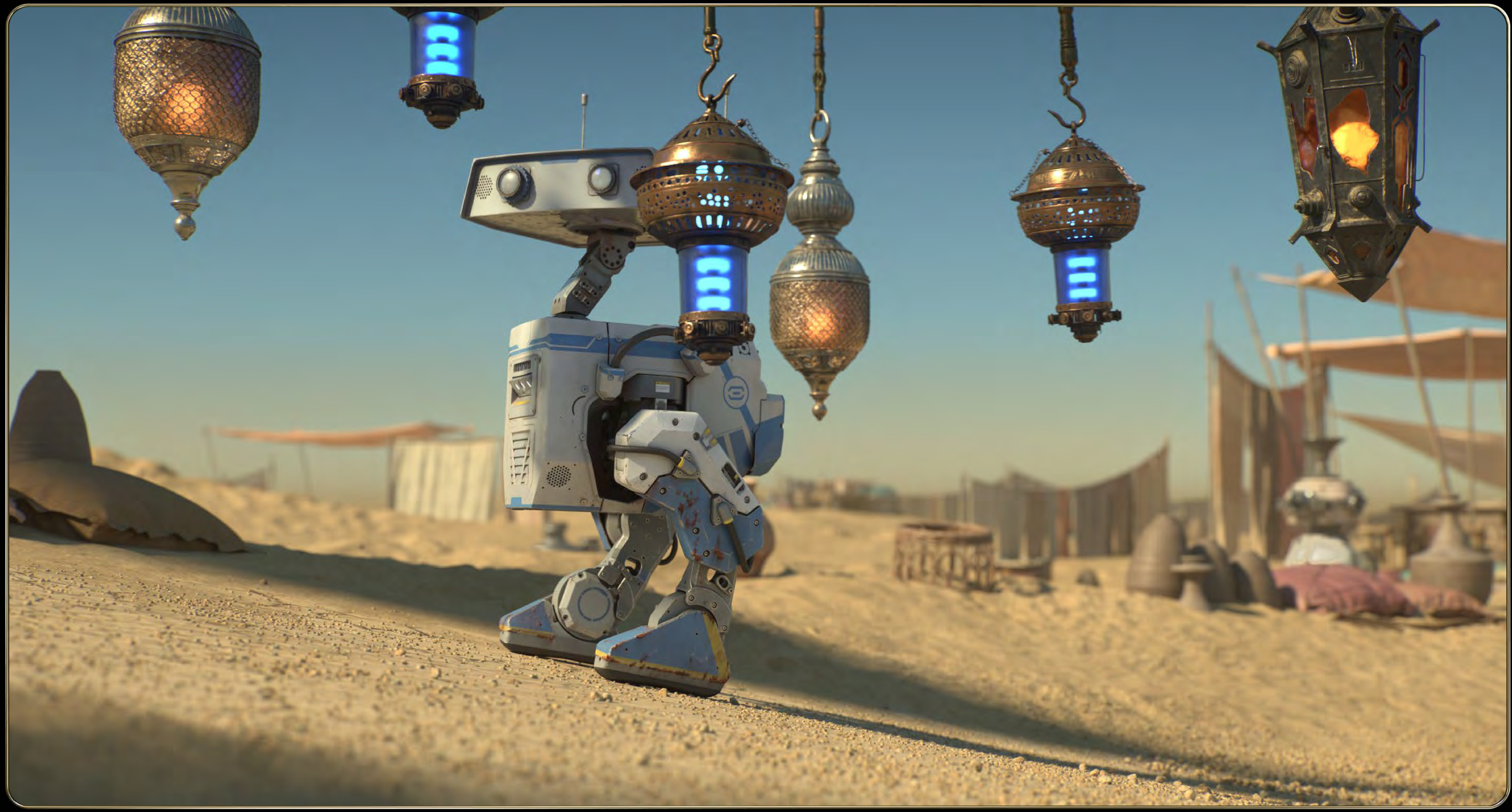
Jetson AGX

NVIDIA and Disney Research Advance Robot Learning



Omniverse | Isaac Sim | Newton

Disney Research | NVIDIA



1X
NEO



Boston Dynamics
E-Atlas



Agility Robotics
Digit



Disney Research
BDX Droid

NVIDIA
Jensen

Figure
Figure 03



Mentee Robotics
MenteeBot



NEURA Robotics
4NE-1



Apptronik
Apollo



Announcing Global Automakers Adopt NVIDIA DRIVE Hyperion

LUCID

Mercedes-Benz

STELLANTIS



DRIVE AGX ECU

momenta

nuro

pony.ai

waabi

WAYVE

WeRide


23
MPH
30
SPEED
LIMIT
D

Announcing Global Robotaxi Ecosystem Built on NVIDIA DRIVE Hyperion

Uber | NVIDIA



AI
Blackwell NVL72

Quantum
NVQLink

U.S. DEPARTMENT of ENERGY

Open Models
Nemotron, Cosmos, GR00T, Clara

Hugging Face

Made in America
Omniverse

FOXCONN HON HAI TECHNOLOGY GROUP

Jacobs

tsmc

Enterprise
CUDA-X, Nemotron

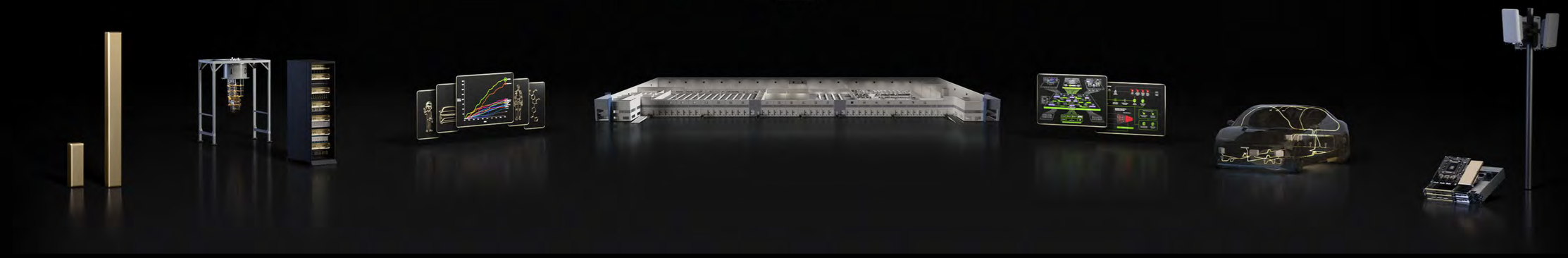
Palantir

Robotics
DRIVE Hyperion

Uber

6G
NVIDIA ARC

NOKIA




NVIDIA