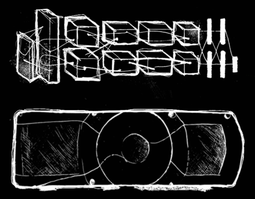

NVIDIA


accenture
Adobe
airbnb
Alibaba Cloud
amazon
AMGEN
Ansys
aramco
arm
AT&T
Baidu 百度
BARCLAYS
BlackRock
BMW GROUP
BNP PARIBAS
Bristol Myers Squibb
cadence
CapitalOne
cerence ai
cisco
Coca-Cola
COHESITY
CoreWeave
CORTEVA agriscience
databricks
DataRobot
ddn
DELL
Deloitte.
DELTA
Disney Research
EA
ebay
EPIC GAMES
EQUINIX
ERICSSON
EY
Flipkart
Ford
FOXCONN
FUJITSU
GE Healthcare
Google
GSK
hp
Hewlett Packard Enterprise
HEXAGON
HITACHI
HRT
Hugging Face
IBM
GTC25
Infosys
JD.COM
JLR
Kroger
Lambda
LangChain
Lenovo
LinkedIn
LOCKHEED MARTIN
LOWE'S
mastercard
MARS
MAYO CLINIC
Meta
Micron
Microsoft
MISTRAL AI_
Nasdaq
NAVER Cloud
NetApp
NOVARTIS
NTT Group
NYSE
OpenAI
ORACLE
PEGATRON
PEPSICO
PETROBRAS
Pfizer
PG&E
PINGAN
QCT
Red Hat
rescale
RIVIAN
ROBLOX
Roche
Rockwell Automation
Royal Bank of Canada
SAAB
salesforce
SAP
SAMSUNG ELECTRONICS
Schneider Electric
servicenow
Shell
shopify
SIEMENS
SK hynix
slb
SoftBank
S&P Global
telenor
Tencent 腾讯
The Weather Company
T-Mobile
together.ai
tsmc
Uber
UBS
Unilever
usbank
verizon
VISA
vmware
VOLVO
Waabi
Walmart
WAYMO
WAYVE
Weights & Biases
wipro
xiaomi
Yum!
ZOOX


GEFORCE RTX 5090

2012 ALEXNET

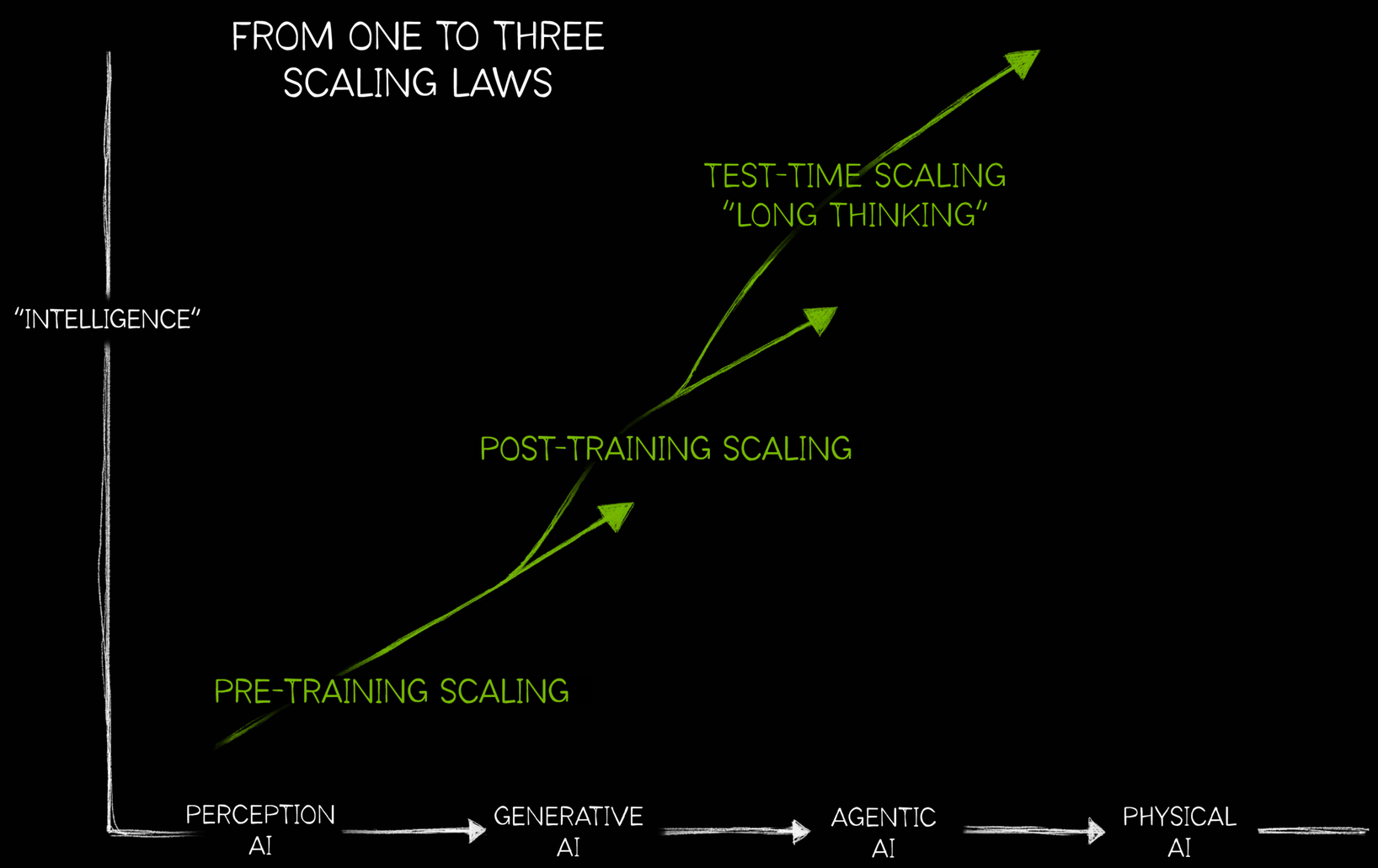


PERCEPTION AI
SPEECH RECOGNITION
DEEP RECSYS
MEDICAL IMAGING

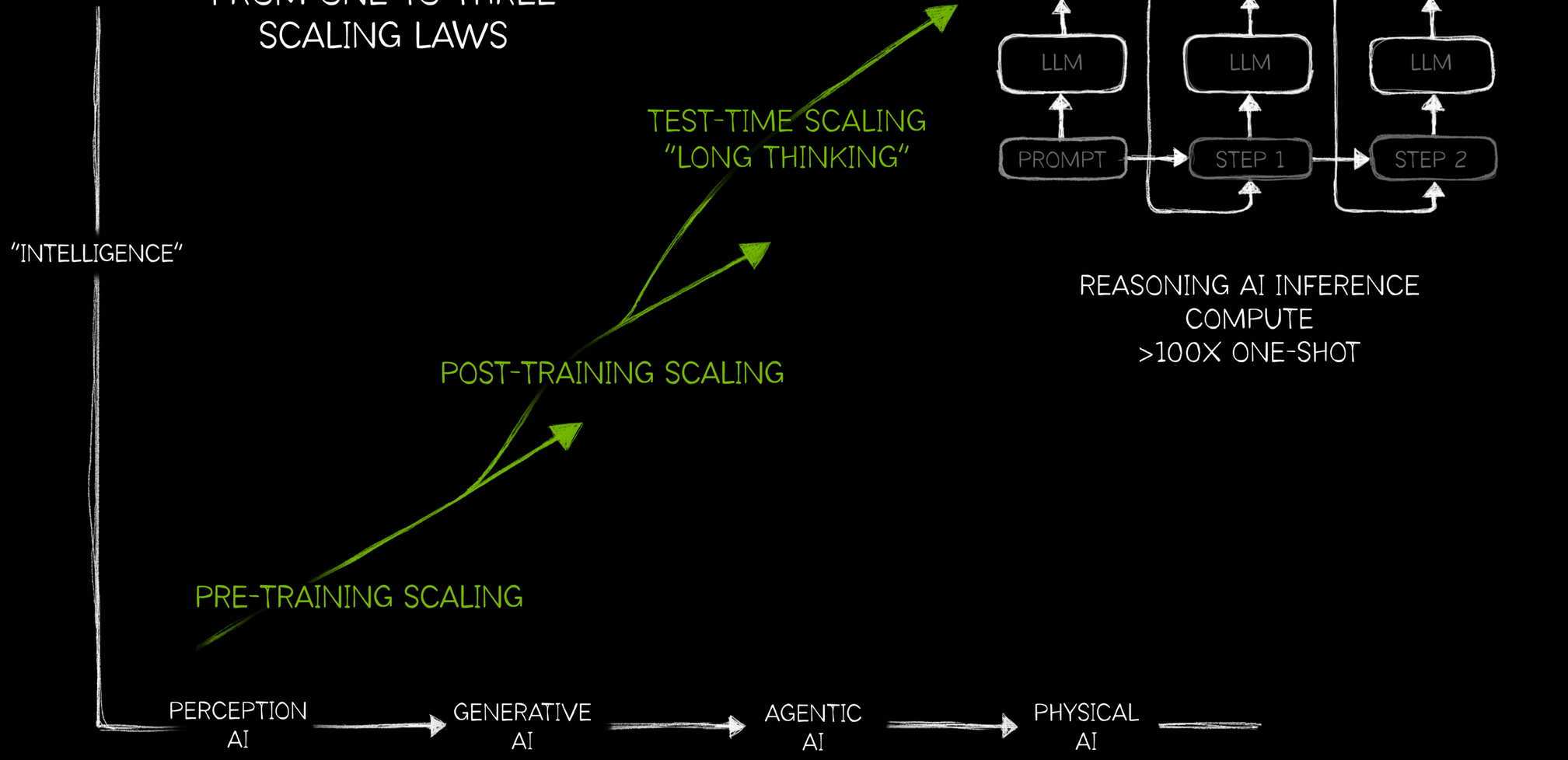
GENERATIVE AI
DIGITAL MARKETING
CONTENT CREATION

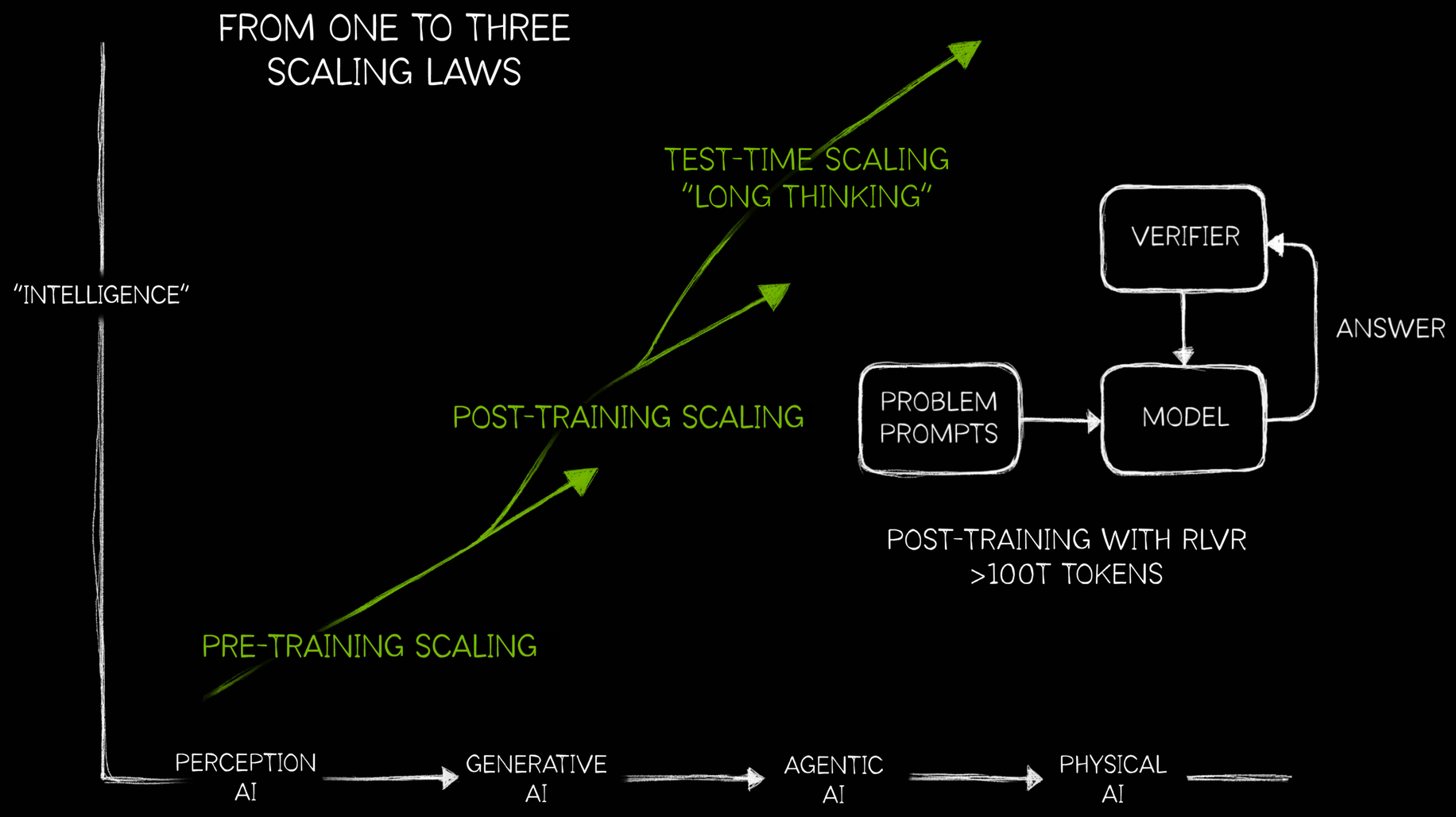
AGENTIC AI
CODING ASSISTANT
CUSTOMER SERVICE
PATIENT CARE

PHYSICAL AI
AUTONOMOUS VEHICLES
GENERAL ROBOTICS

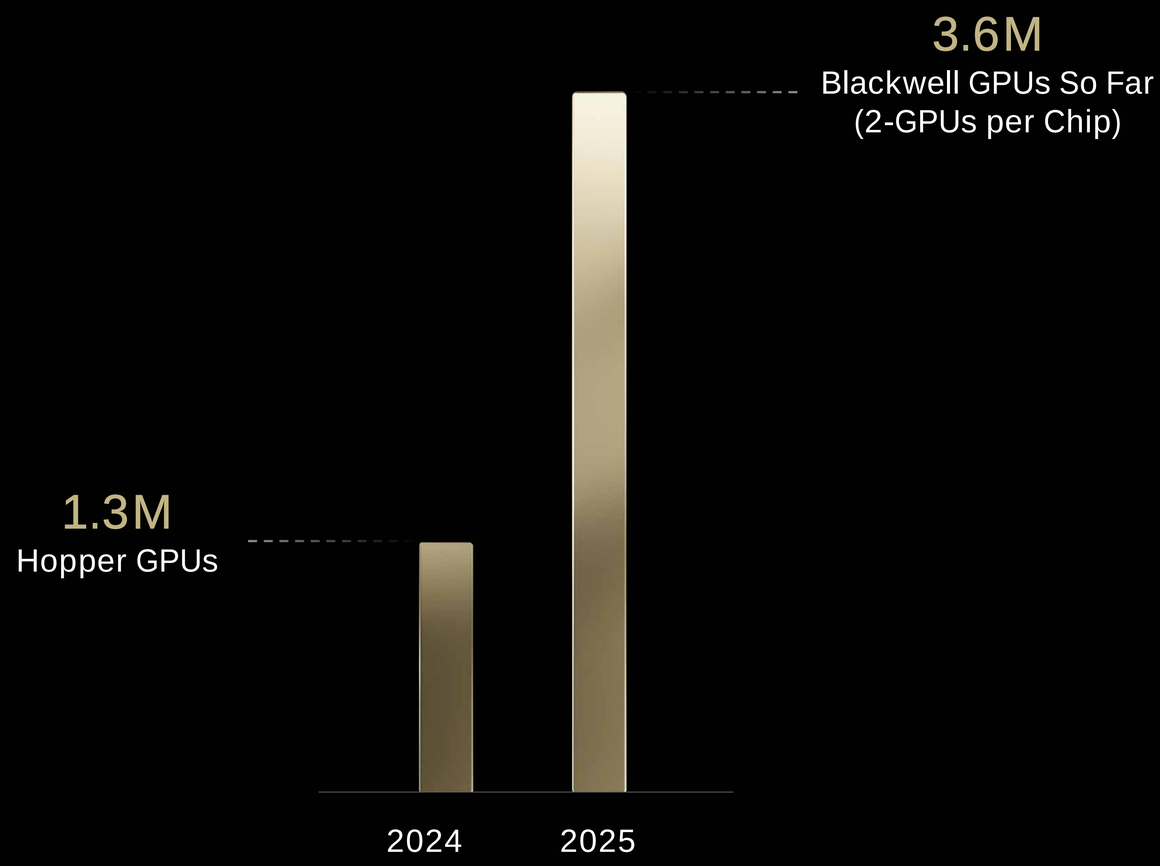

FROM ONE TO THREE
SCALING LAWS
"INTELLIGENCE"
TEST-TIME SCALING
"LONG THINKING"
POST-TRAINING SCALING
PRE-TRAINING SCALING
PERCEPTION AI
GENERATIVE AI
AGENTIC AI
PHYSICAL AI

FROM ONE TO THREE SCALING LAWS

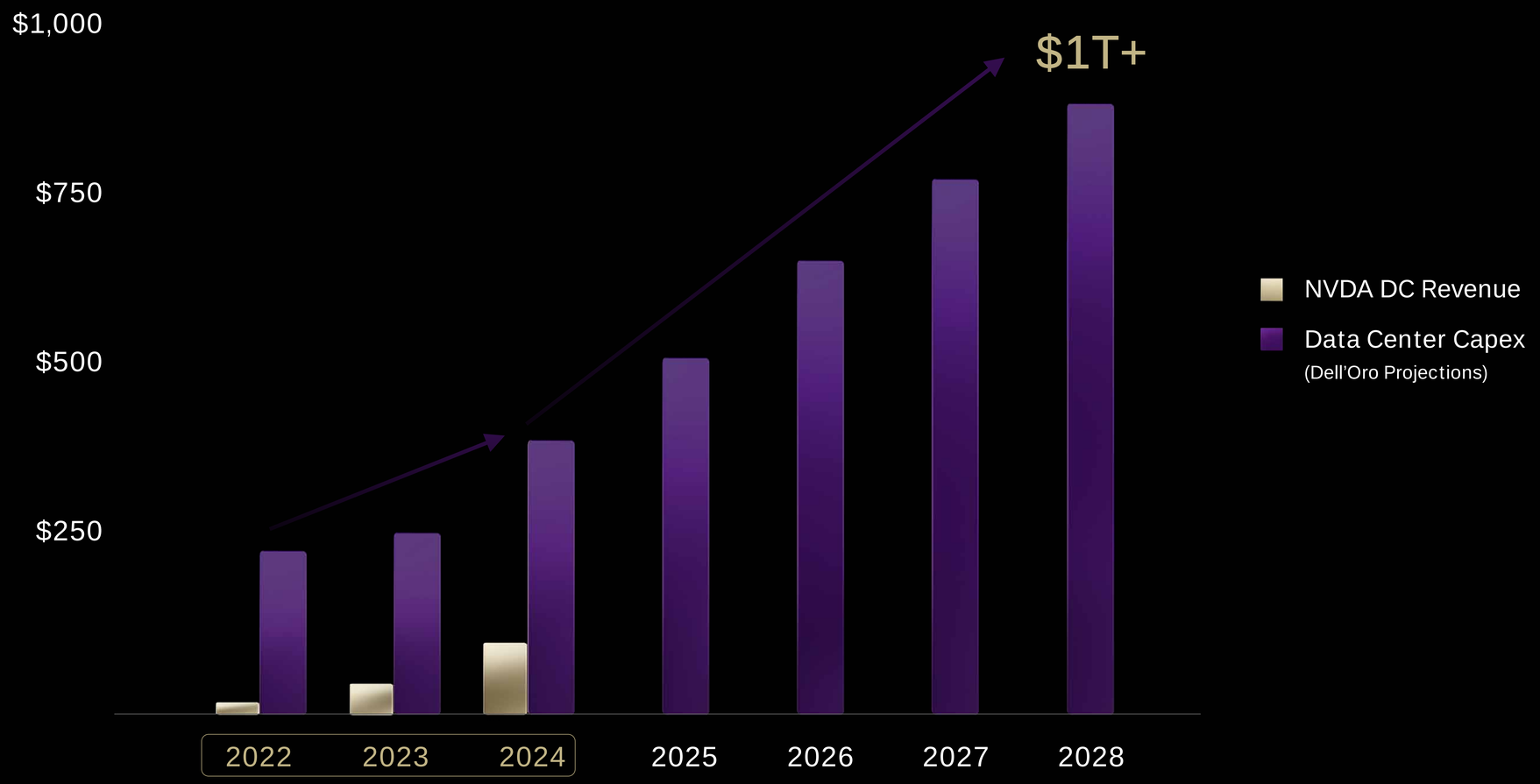



FROM ONE TO THREE
SCALING LAWS
"INTELLIGENCE"
TEST-TIME SCALING
"LONG THINKING"
POST-TRAINING SCALING
PRE-TRAINING SCALING
VERIFIER
ANSWER
PROBLEM
PROMPTS
MODEL
POST-TRAINING WITH RLVR
>100T TOKENS
PERCEPTION
AI
GENERATIVE
AI
AGENTIC
AI
PHYSICAL
AI

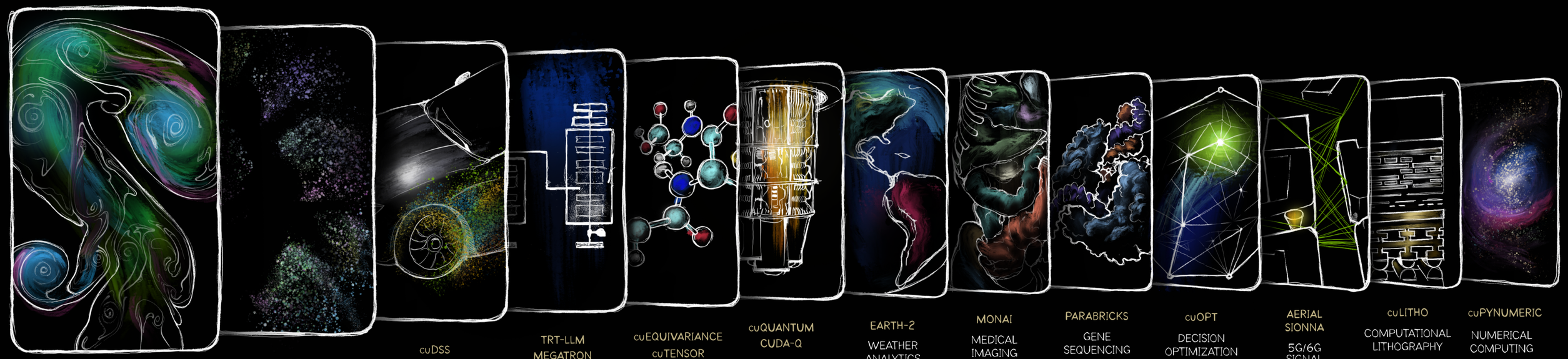
Top 4 US CSPs


3.6M
Blackwell GPUs So Far
(2-GPUs per Chip)
1.3M
Hopper GPUs
2024
2025

Computing At Inflection Point



CUDA-X FOR EVERY INDUSTRY



WARP

PHYSICS

cuDF
cuML

DATA SCIENCE
AND PROCESSING

cuDSS
cuSPARSE
cuFFT
AMGX

COMPUTER AIDED
ENGINEERING

TRT-LLM
MEGATRON
NCCL
cuDNN
CUTLASS
cuBLAS

DEEP
LEARNING

cuEQUIVARIANCE
cuTENSOR

QUANTUM
CHEMISTRY

cuQUANTUM
CUDA-Q

QUANTUM
COMPUTING

EARTH-2

WEATHER
ANALYTICS

MONAI

MEDICAL
IMAGING

PARABRICKS

GENE
SEQUENCING

cuOPT

DECISION
OPTIMIZATION

AERIAL
SIONNA

5G/6G
SIGNAL
PROCESSING

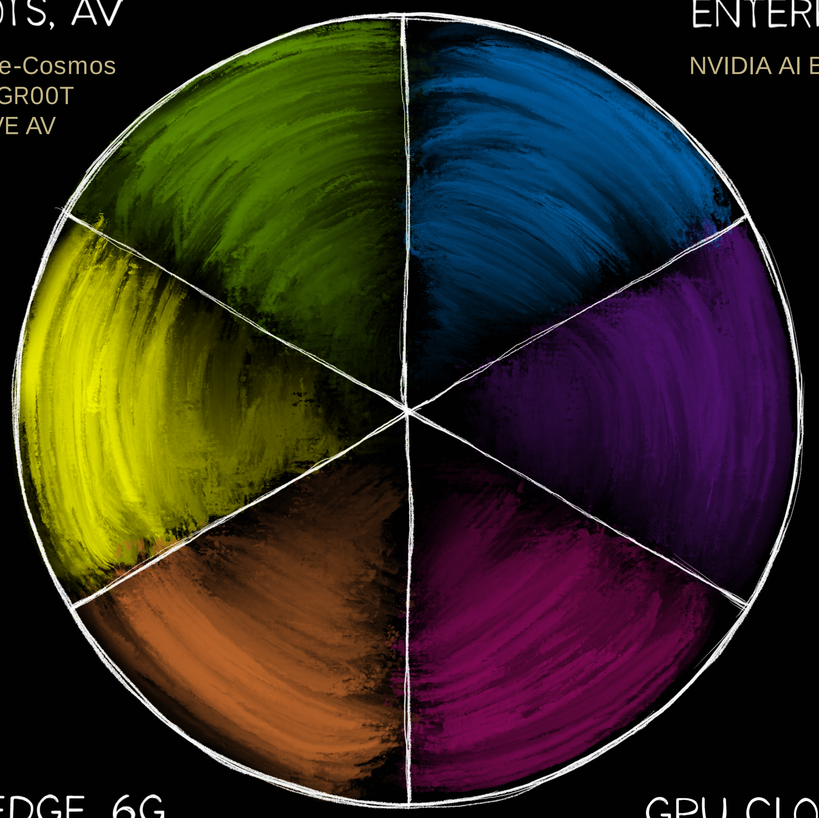
cuLITHO

COMPUTATIONAL
LITHOGRAPHY

cuPYNUMERIC

NUMERICAL
COMPUTING

AI FOR EVERY INDUSTRY



ROBOTS, AV

Omniverse-Cosmos
Isaac GR00T
DRIVE AV

ENTERPRISE IT

NVIDIA AI Enterprise

ROBOTIC MANUFACTURING

Omniverse-Cosmos

CLOUD SERVICE PROVIDERS

Leading Technology

Full-Stack AI

CUDA-X Diverse Applications

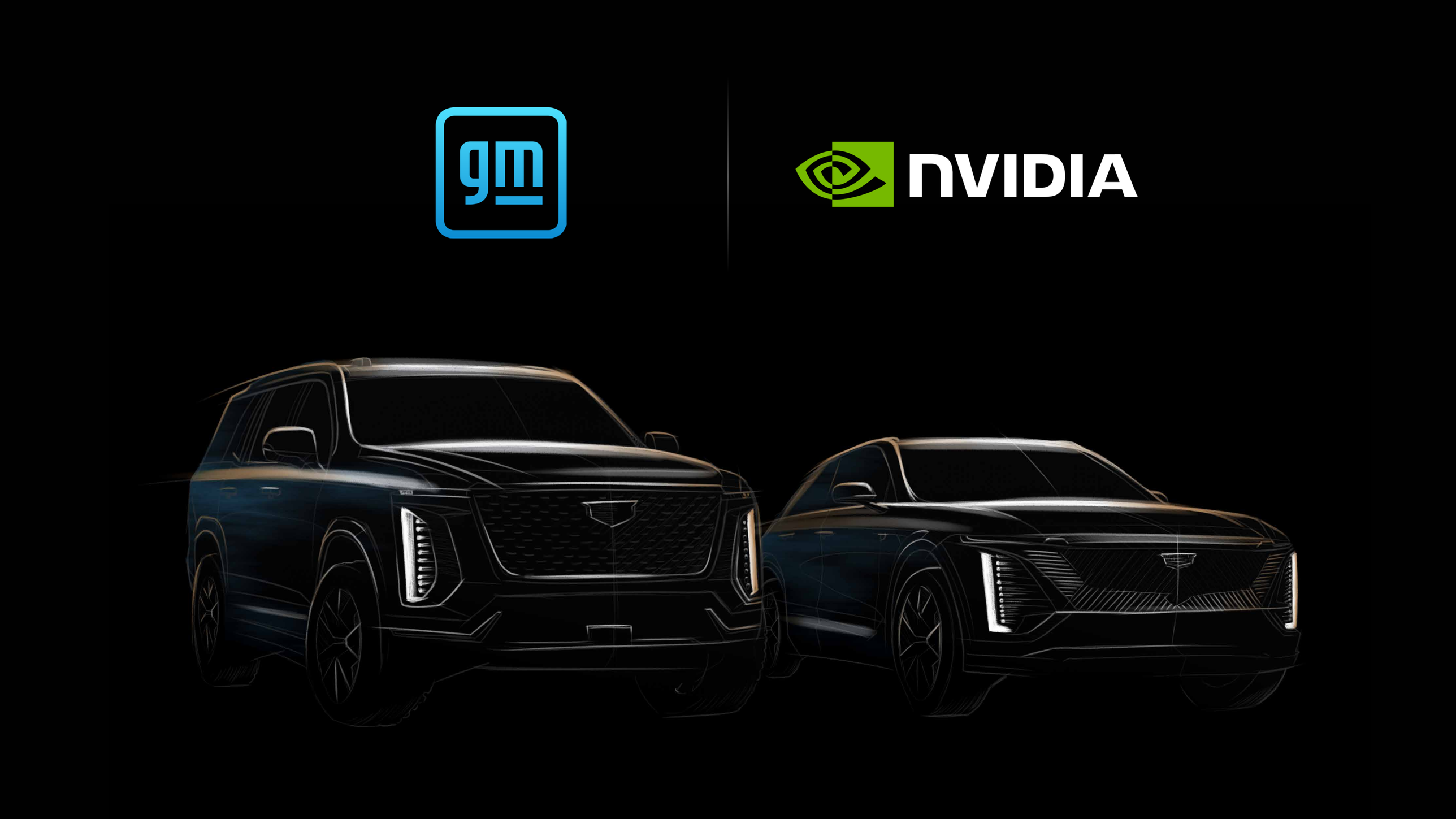
Rich Developer Ecosystem

EDGE, 6G

Aerial-Sionna
6G AI-RAN

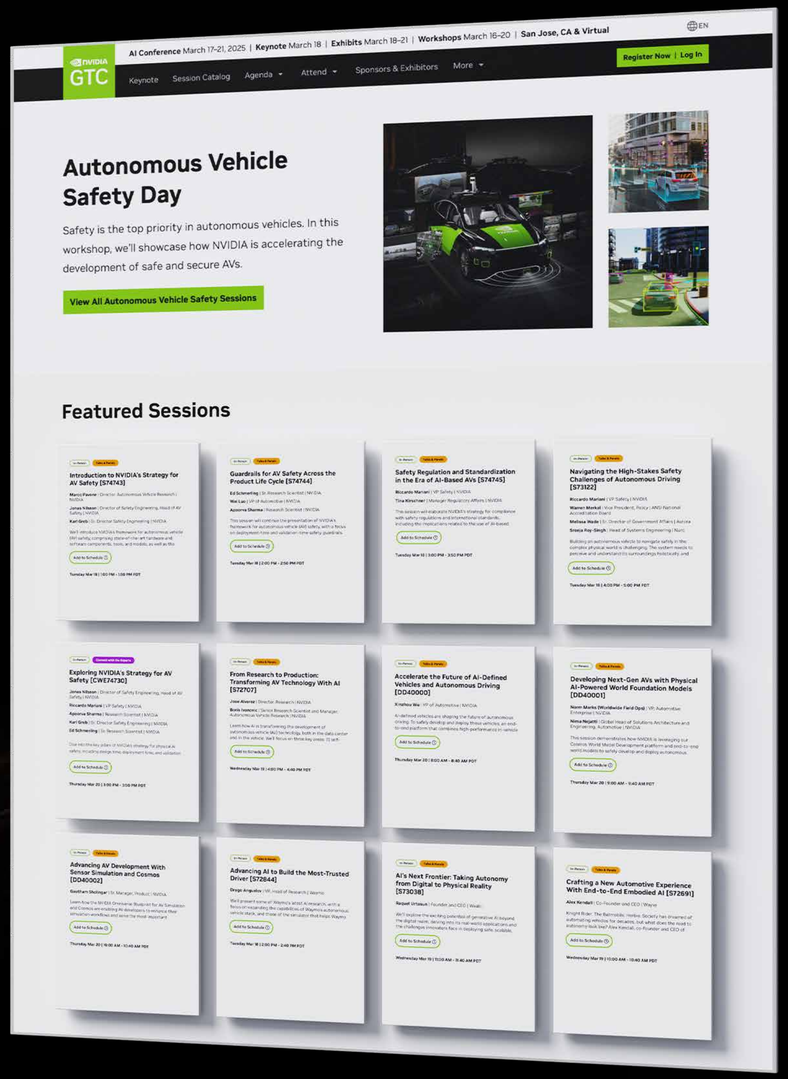
GPU CLOUDS

AI Factory
Blueprint and Full Stack

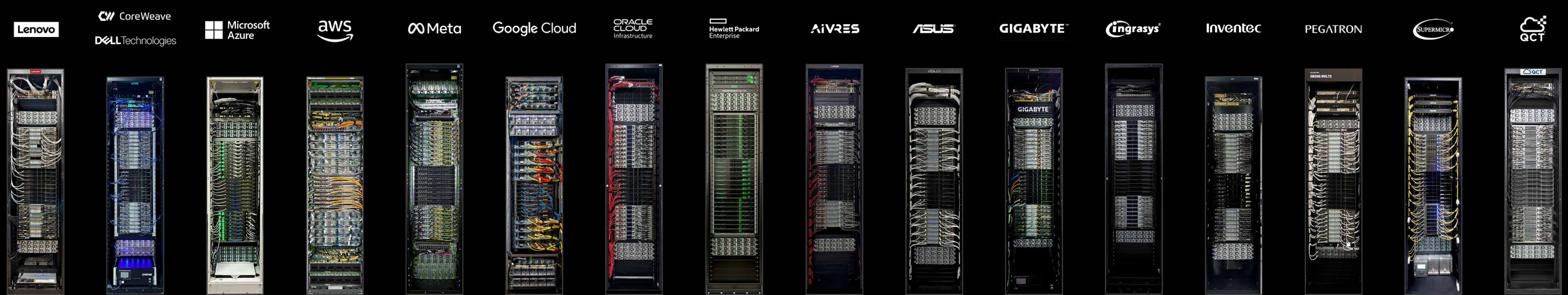

gm
NVIDIA

Announcing NVIDIA Halos

Chip-to-Deployment AV Safety System



Grace Blackwell in Full Production

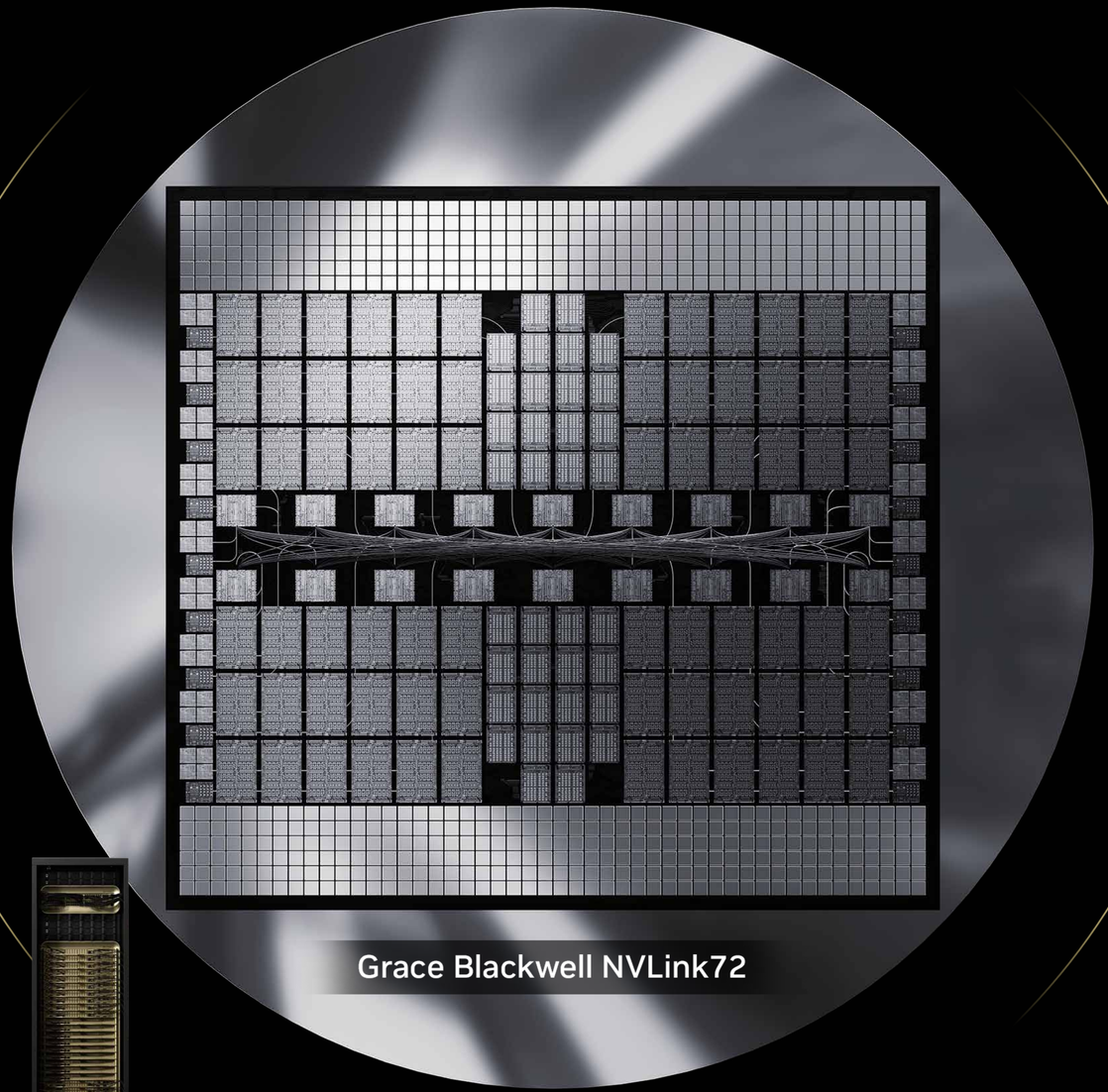


NVIDIA Blackwell System

72 Blackwell GPUs
1.1 EF FP4

576 Memory Chips
20 TB
576 TB/s

18 NVLink Switches
130 TB/s All-to-All



Grace Blackwell NVLink72

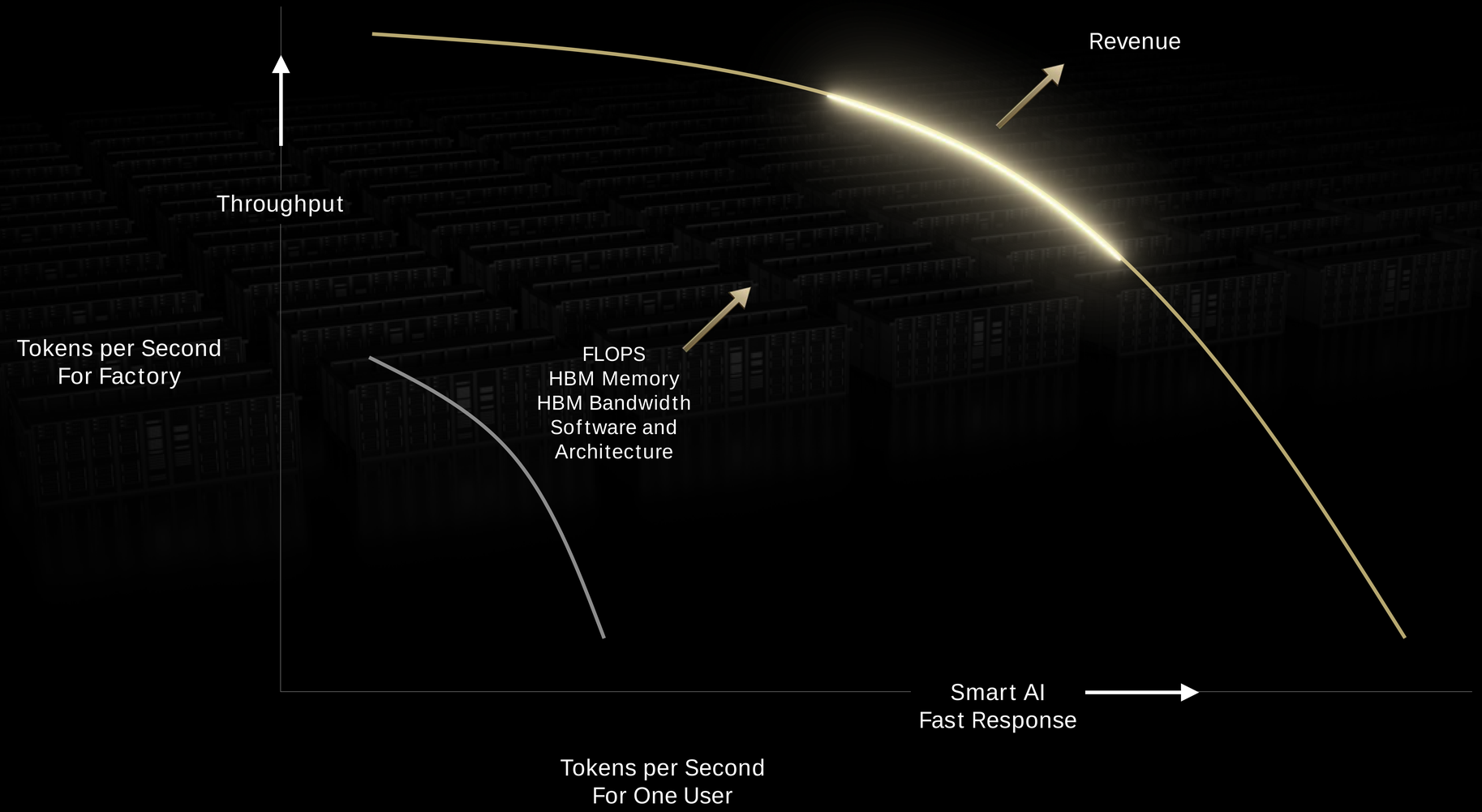
130 Trillion Transistors

2,592 Grace CPU Cores

72 ConnectX-8 NICs

18 BlueField DPUs

Inference At-Scale is Extreme Computing



I need to seat 7 people around a table at my wedding reception, but my parents and in-laws should not sit next to each other. Also, my wife insists we look better in pictures when she's on my left, but I need to sit next to my best man. How do I seat us on a round table? But then, what happens if we invited our pastor to sit with us?

Send

Traditional LLM Model

Tokens: **439**

Reasoning Model

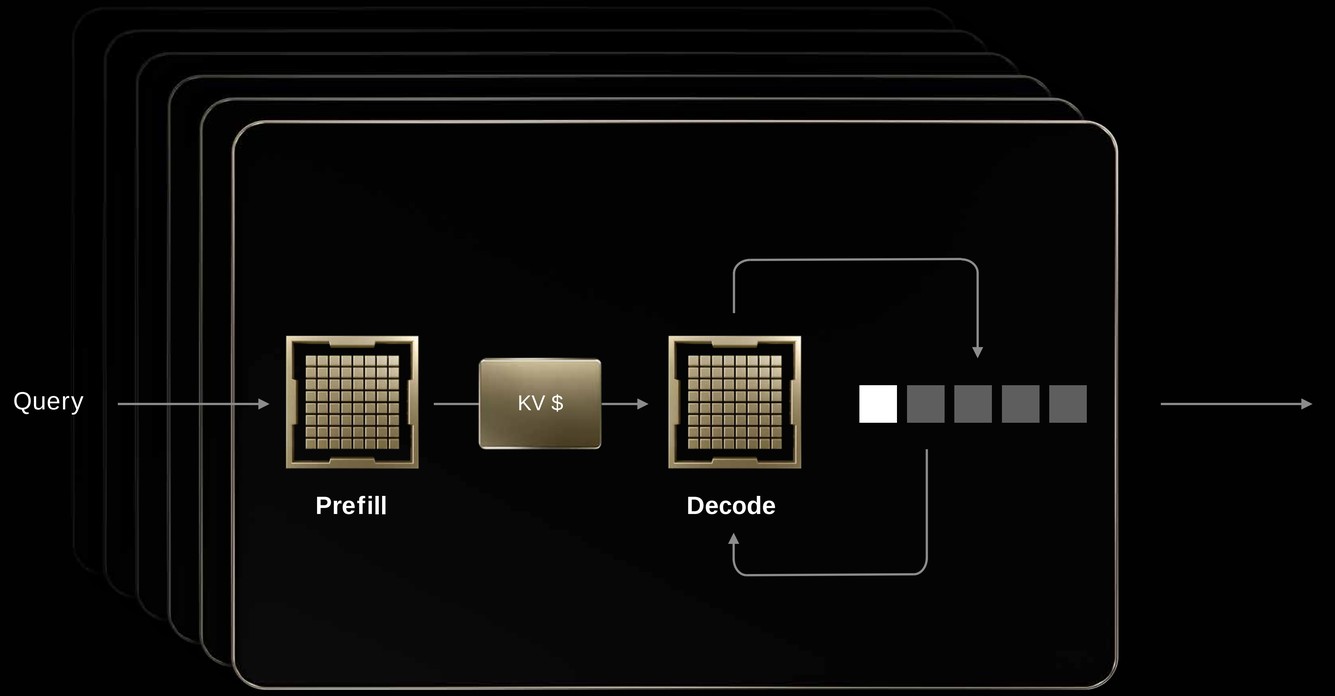
Tokens: **8,559**

Reasoning On

20x More Tokens
150x More Compute

Announcing NVIDIA Dynamo

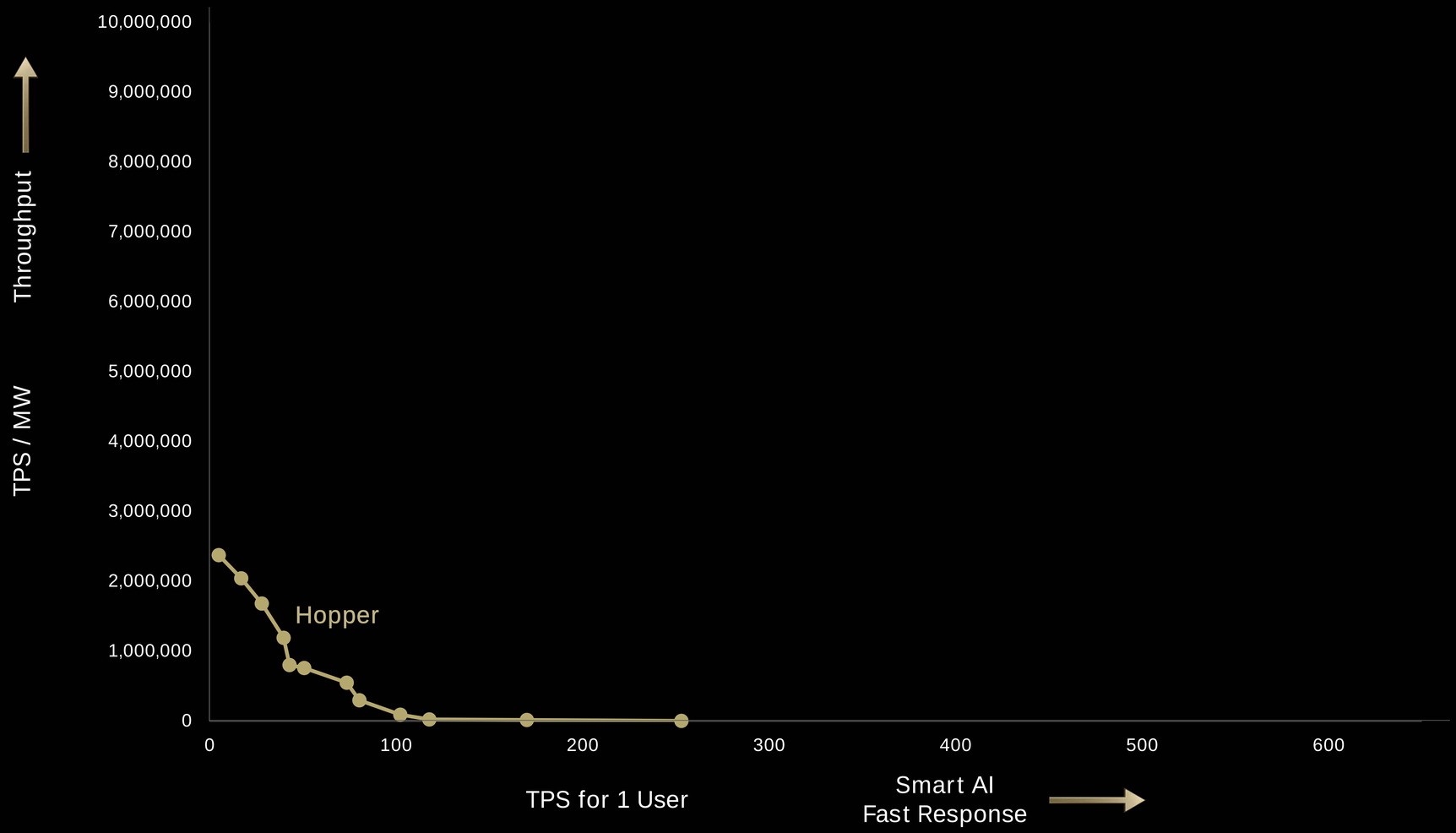
Distributed Inference Serving Library



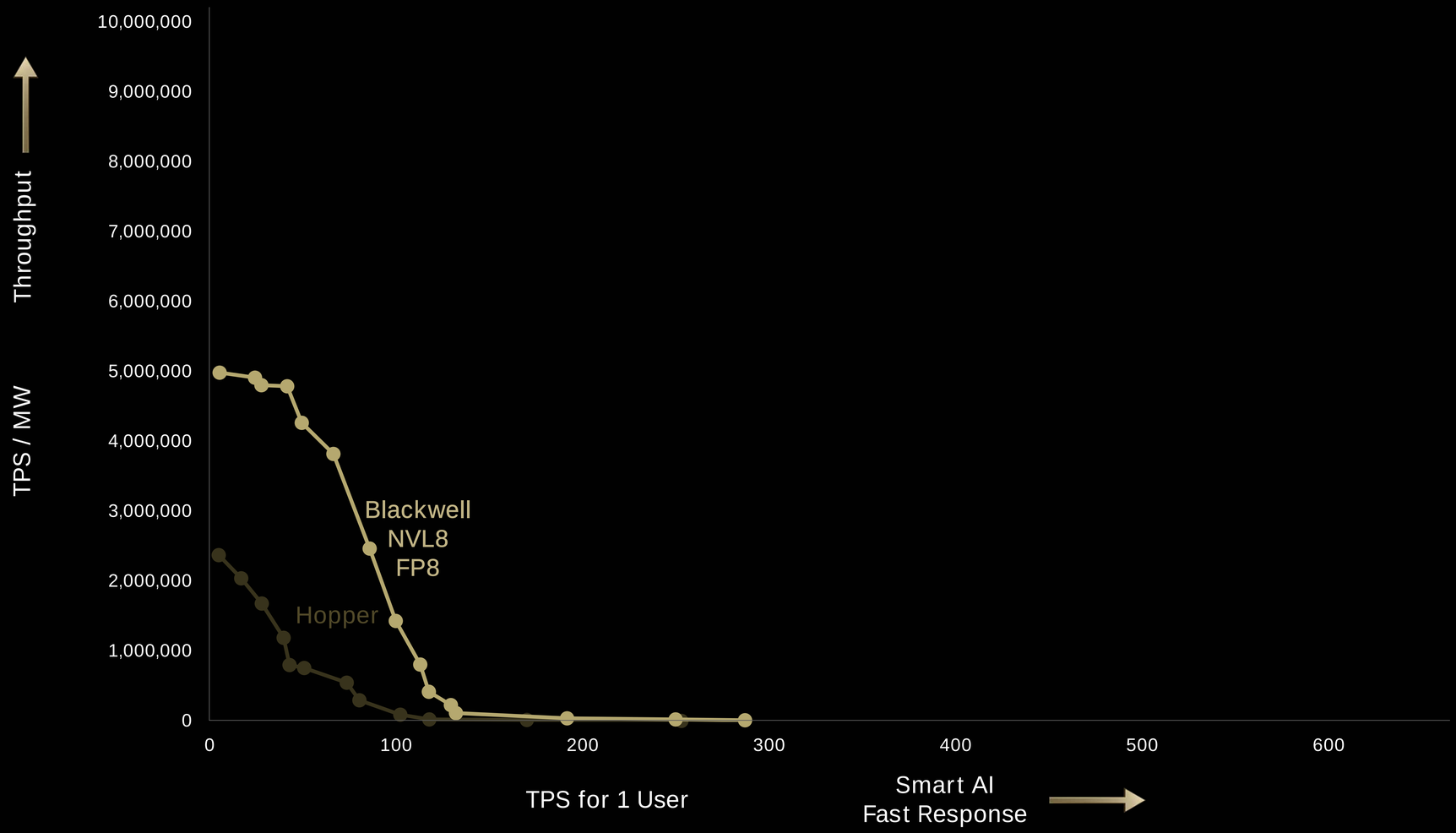
Disaggregated Inference | GPU Resource Allocation | KV Cache Routing | Communication Library (NIXL)

cohere · Fireworks AI · Meta · Microsoft Azure · perplexity · PyTorch · SGL · together.ai · VAST · vLLM

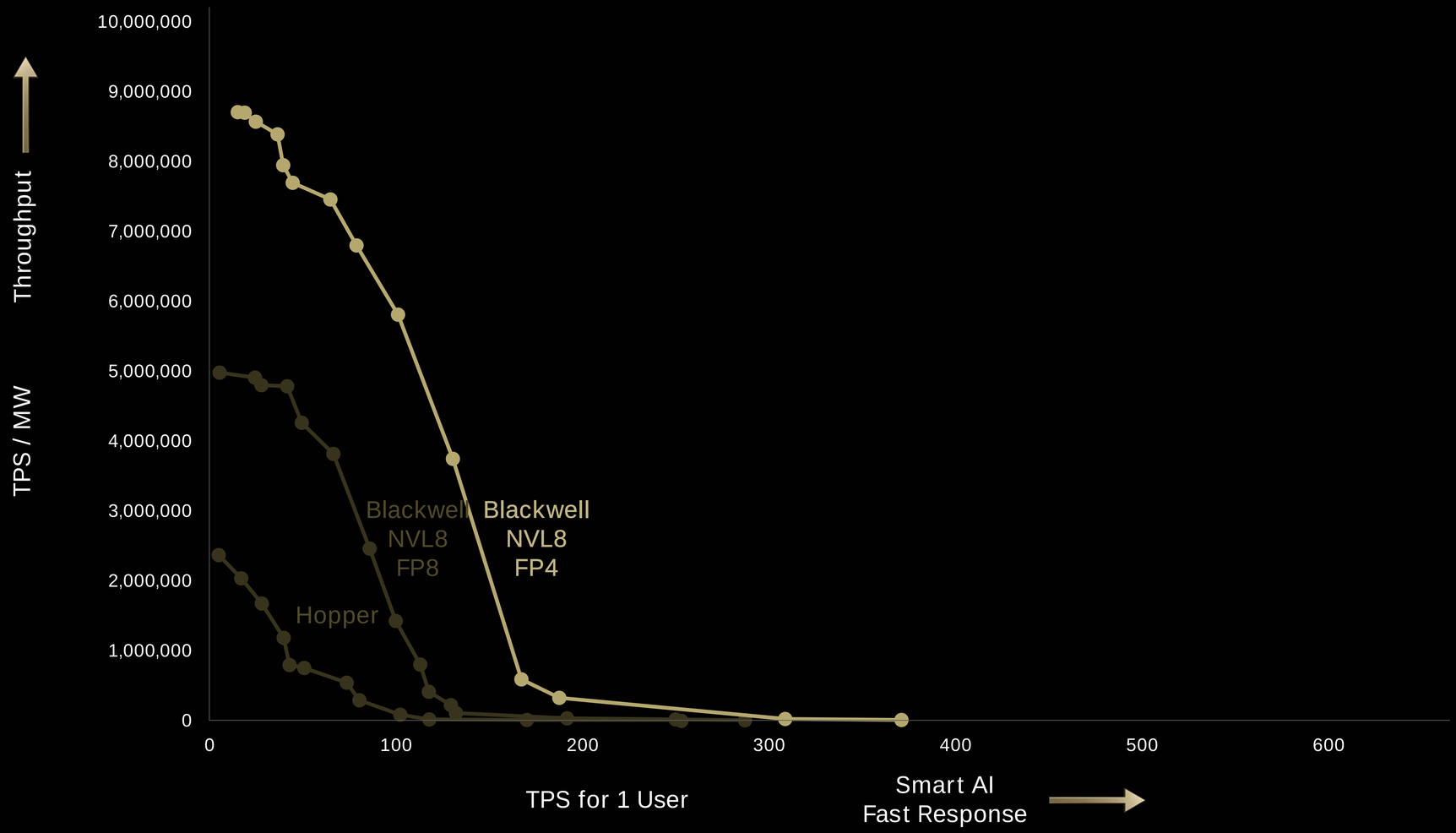
Blackwell Giant Leap in Inference Performance



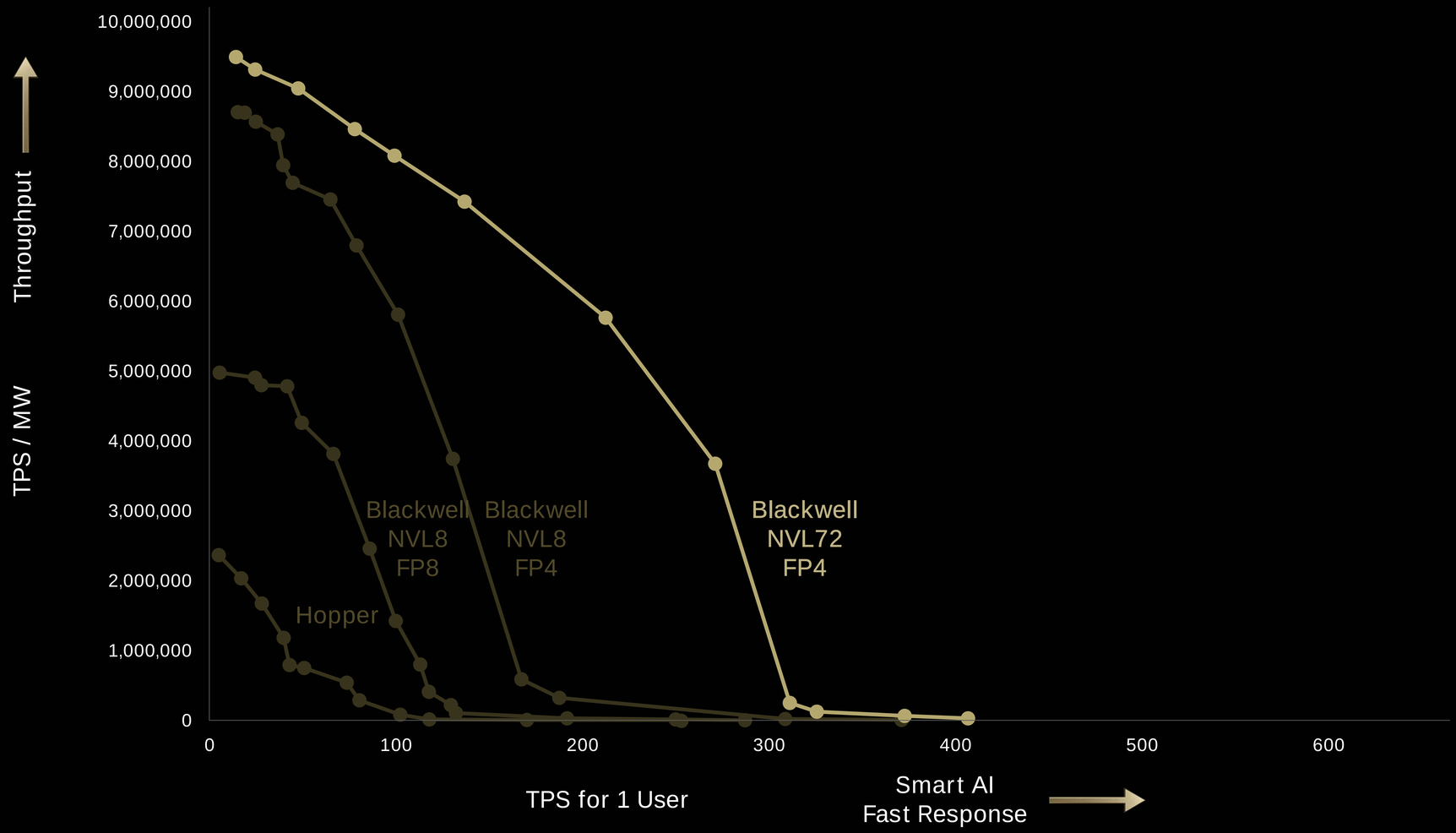
Blackwell Giant Leap in Inference Performance



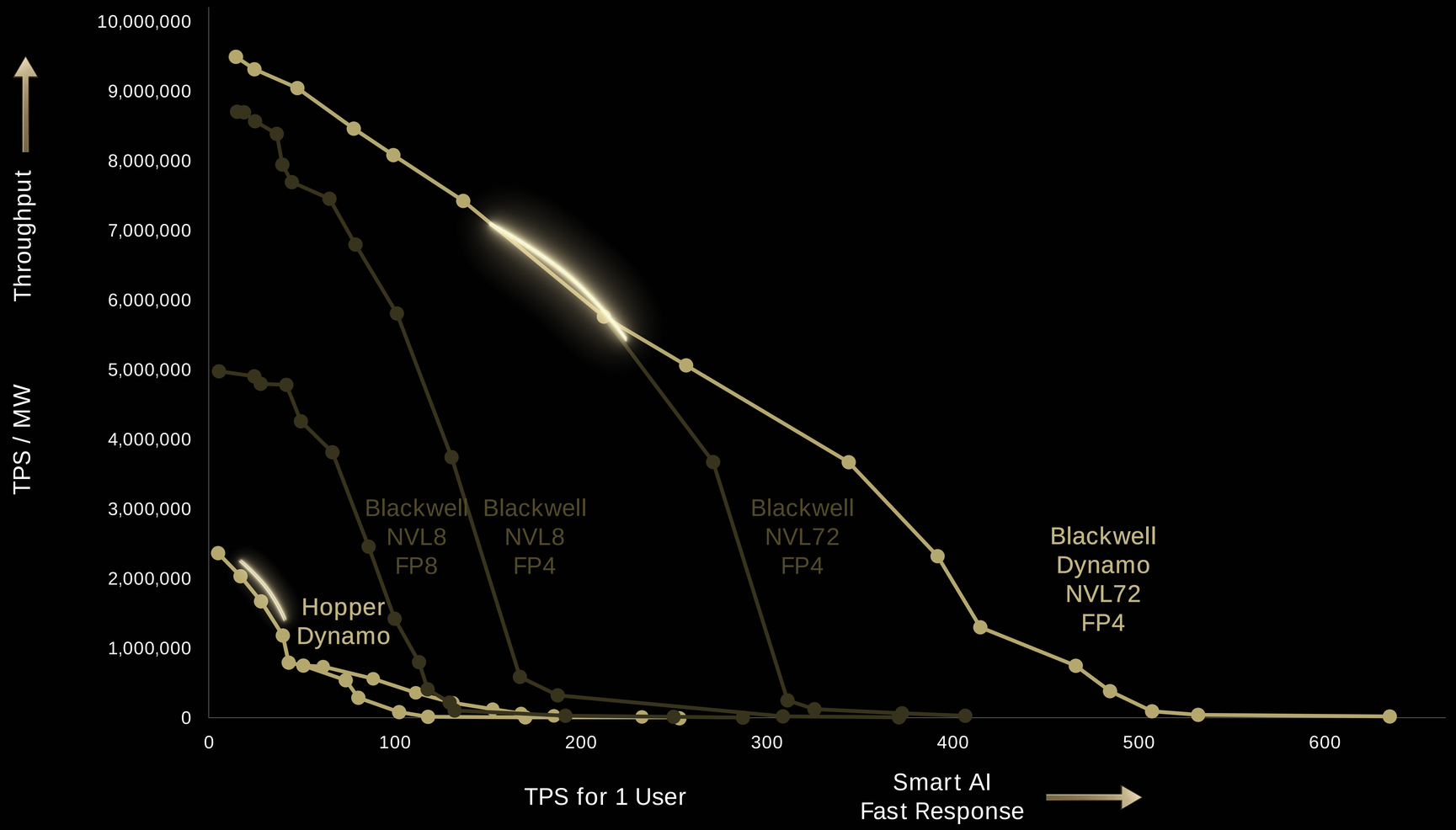
Blackwell Giant Leap in Inference Performance



Blackwell Giant Leap in Inference Performance



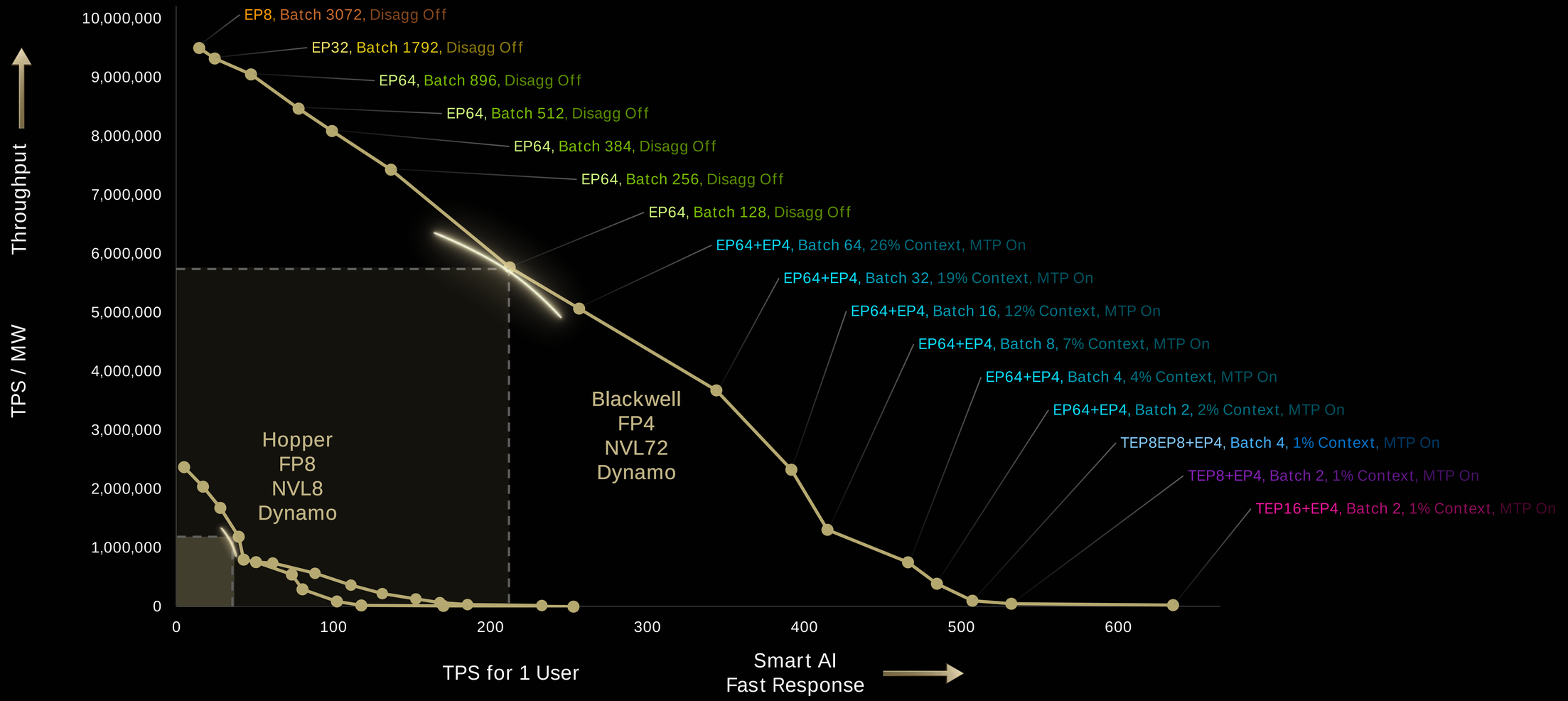
Blackwell Giant Leap in Inference Performance



Blackwell 25X Hopper

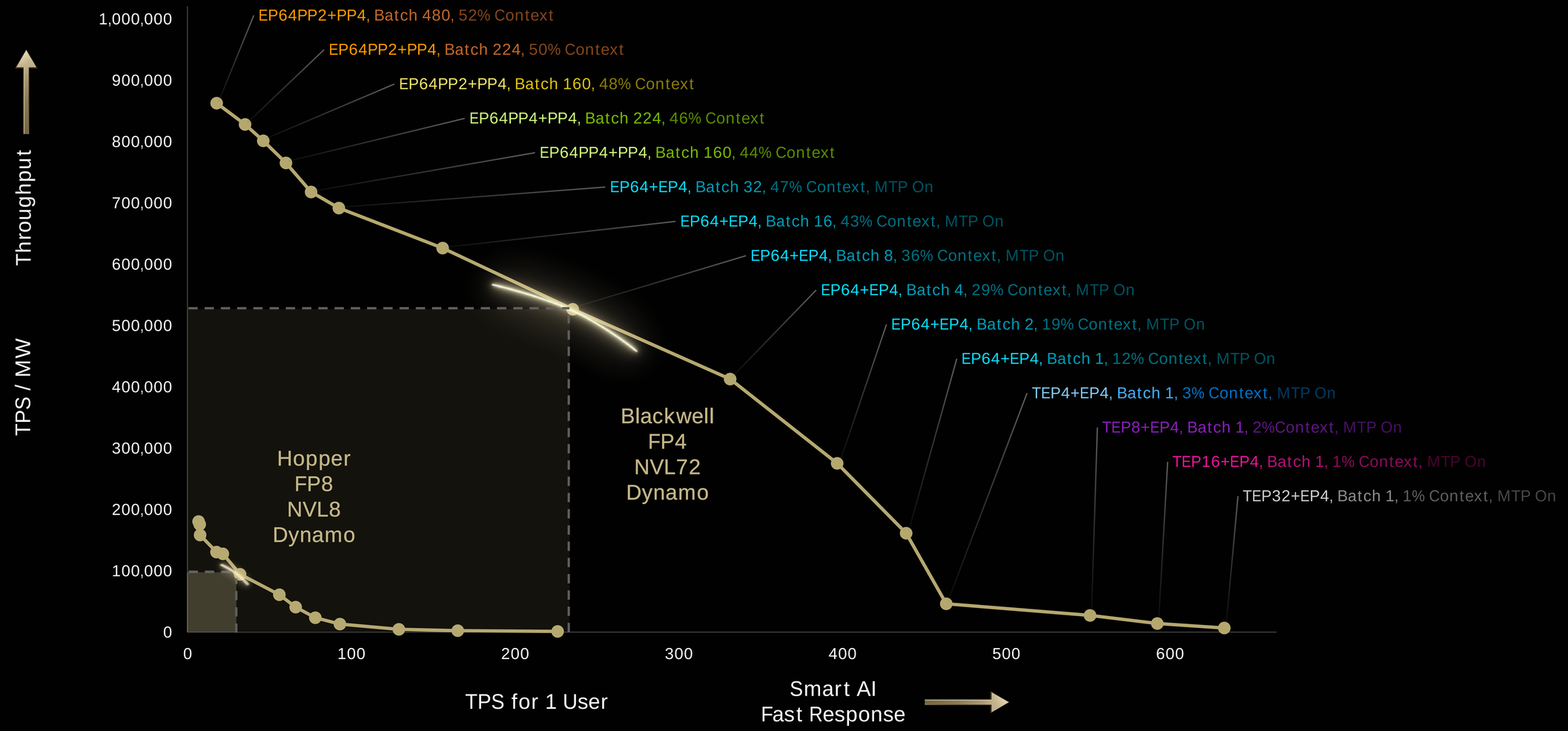
FP4, NVL72, Dynamo, and TRT-LLM Continuous Optimization

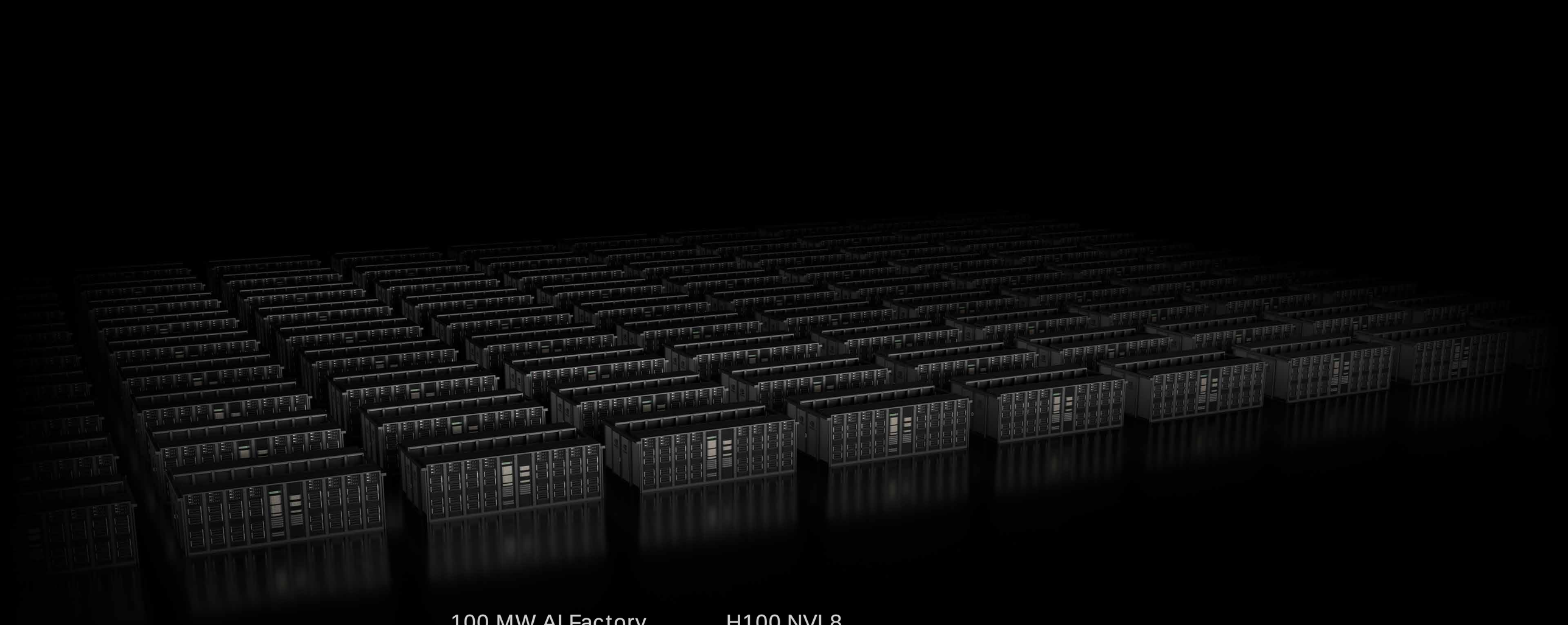
1K ISL / 2K OSL



Blackwell 40X Hopper

FP4, NVL72, Dynamo, and TRT-LLM Continuous Optimization
32K ISL / 8K OSL



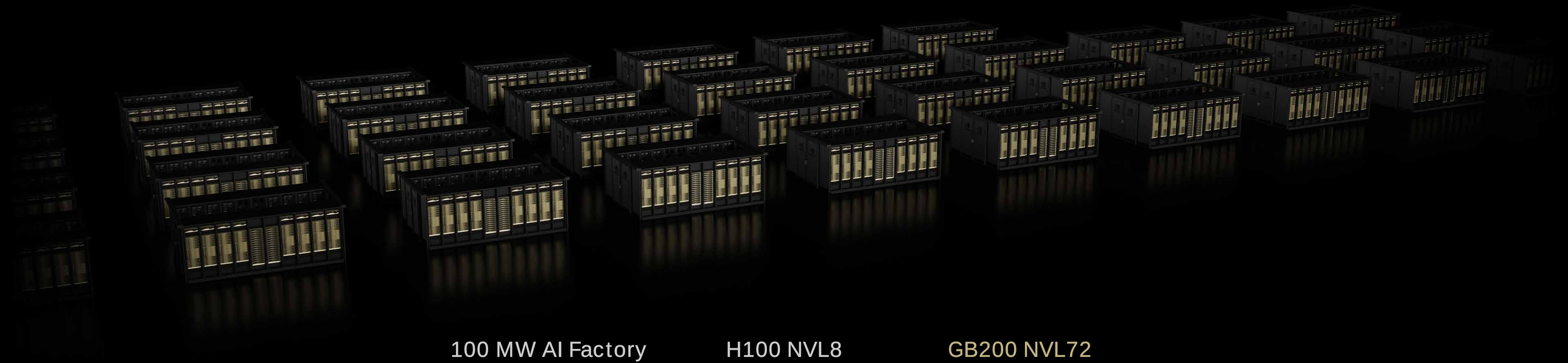


100 MW AI Factory	H100 NVL8
GPU Dies	45K
Racks	1,400
Data Center Productivity	300M

DeepSeek R1 Context and Generation, , ISL=32K, OSL=8K @ Pareto optimal TPS/User

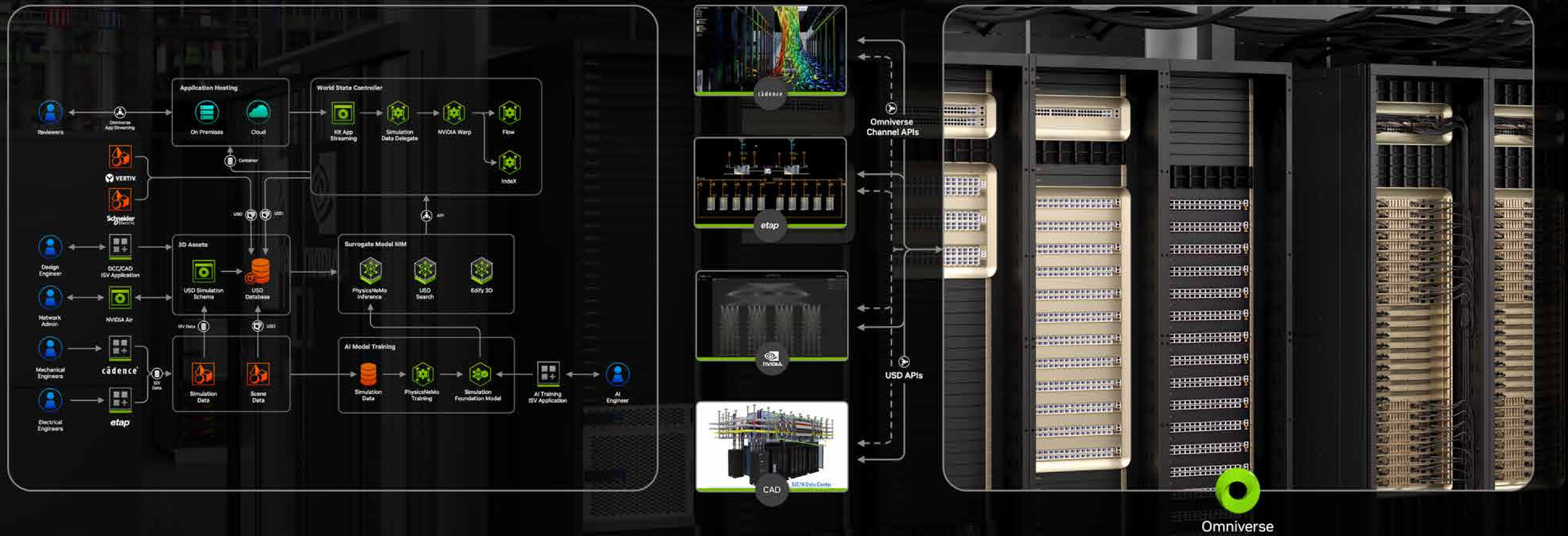
Blackwell 40X Hopper Inference Performance

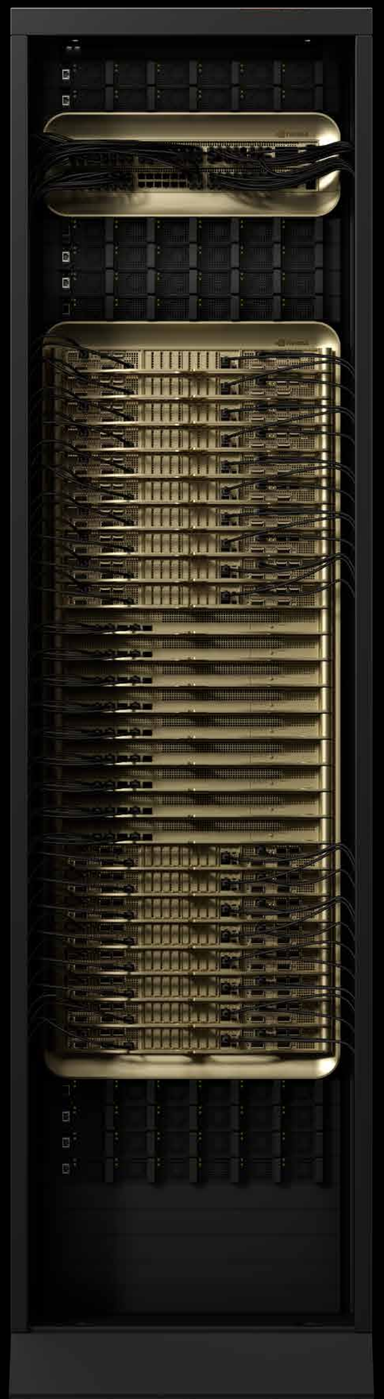
NVLink Flops Per Watt ~ AI Factory Output



100 MW AI Factory	H100 NVL8	GB200 NVL72	
GPU Dies	45K	**85K**	
Racks	1,400	**600**	
Data Center Productivity	300M	**12,000M**	40X More Token Revenue

DeepSeek R1 Context and Generation, , ISL=32K, OSL=8K @ Pareto optimal TPS/User


Application Hosting
World State Controller
Reviewers
Omniverse App Streaming
On Premises
Cloud
Container
Kit App Streaming
Simulation Data Delegate
NVIDIA Warp
Flow
IndeX
VERTIV
Schneider Electric
USD
API
Design Engineer
DCC/CAD ISV Application
3D Assets
USD Simulation Schema
USD Database
Surrogate Model NIM
PhysicsNeMo Inference
USD Search
Edify 3D
Network Admin
NVIDIA Air
ISV Data
Mechanical Engineers
cadence
Simulation Data
Scene Data
AI Model Training
Simulation Data
PhysicsNeMo Training
Simulation Foundation Model
AI Training ISV Application
AI Engineer
Electrical Engineers
etap
cadence
etap
NVIDIA
CAD
Omniverse Channel APIs
USD APIs
Omniverse



Oberon Rack
Liquid Cooled

Blackwell Ultra NVL72

Second Half 2025

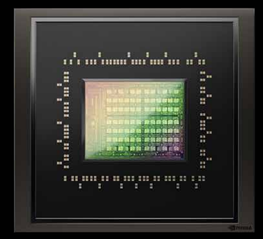
1.1 EF Dense FP4 Inference
0.36 EF FP8 Training
1.5X GB200 NVL72

New Attention Instructions
2X

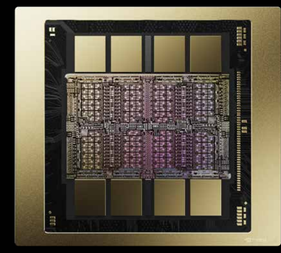
20 TB HBM | 40 TB Fast Memory
1.5X

14.4 TB/s CX8
2X

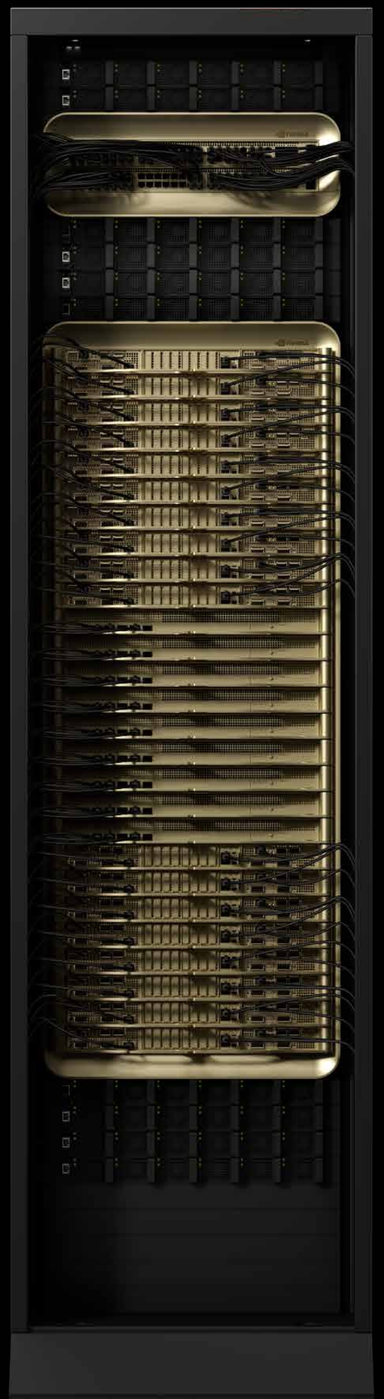
Grace



Blackwell Ultra



2 Reticle-Sized GPUs
15PF Dense FP4 | 288GB HBM3e



Oberon Rack
Liquid Cooled

Vera Rubin NVL144

Second Half 2026

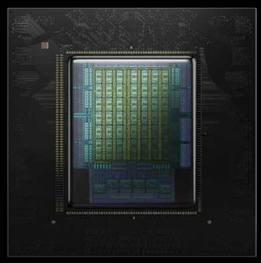
3.6 EF FP4 Inference
1.2 EF FP8 Training
3.3X GB300 NVL72

13 TB/s HBM4
75 TB Fast Memory
1.6X

260 TB/s NVLink6
2X

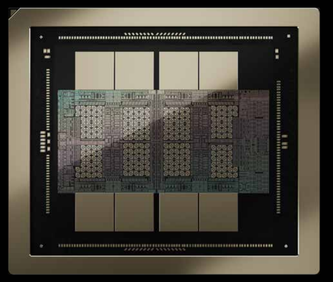
28.8 TB/s CX9
2X

Vera



88 Custom Arm Cores
176 Threads
1.8 TB/s NVLink-C2C

Rubin



2 Reticle-Sized GPUs
50PF FP4 | 288GB HBM4

Kyber Rack
Liquid Cooled

Rubin Ultra NVL576

Second Half 2027

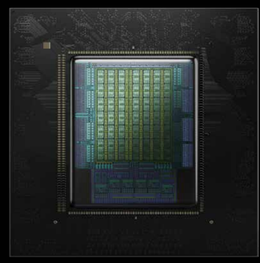
15 EF FP4 Inference
5 EF FP8 Training
14X GB300 NVL72

4.6 PB/s HBM4e
365 TB Fast Memory
8X

1.5 PBs NVLink7
12X

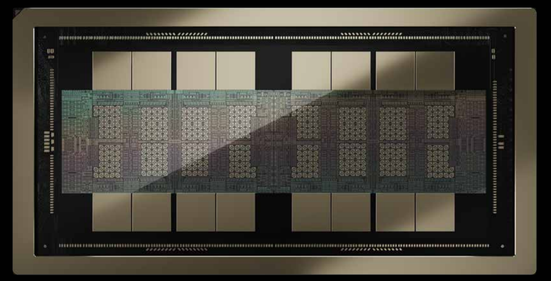
115.2 TB/s CX9
8X

Vera



88 Custom Arm Cores
176 Threads
1.8 TB/s NVLink-C2C

Rubin Ultra



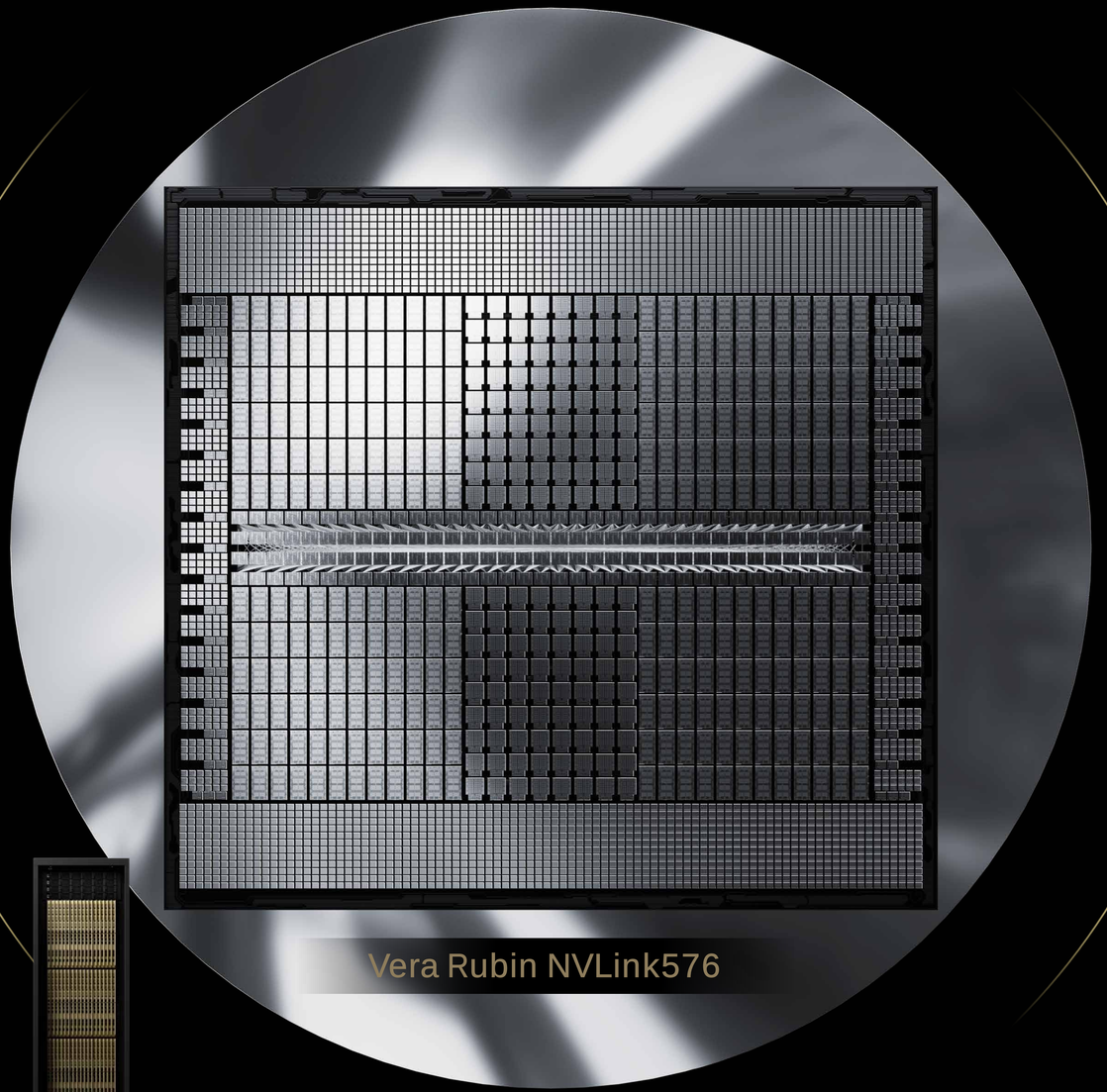
4 Reticle-Sized GPUs
100PF FP4 | 1TB HBM4e

NVIDIA Rubin System

576 Rubin GPUs
15 EF FP4

2,304 Memory Chips
150 TB
4600 PB/s

144 NVLink Switches
1500 PB/s



Vera Rubin NVLink576

1,300 Trillion Transistors

12,672 Vera CPU Cores

576 ConnectX-9 NICs

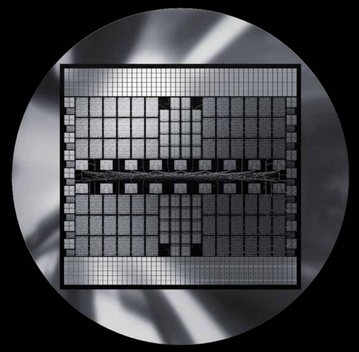
72 BlueField DPUs

NVIDIA Blackwell System

72 Blackwell GPUs
1.1 EF FP4

576 Memory Chips
20 TB
576 TB/s

18 NVLink Switches
130 TB/s



130 Trillion Transistors

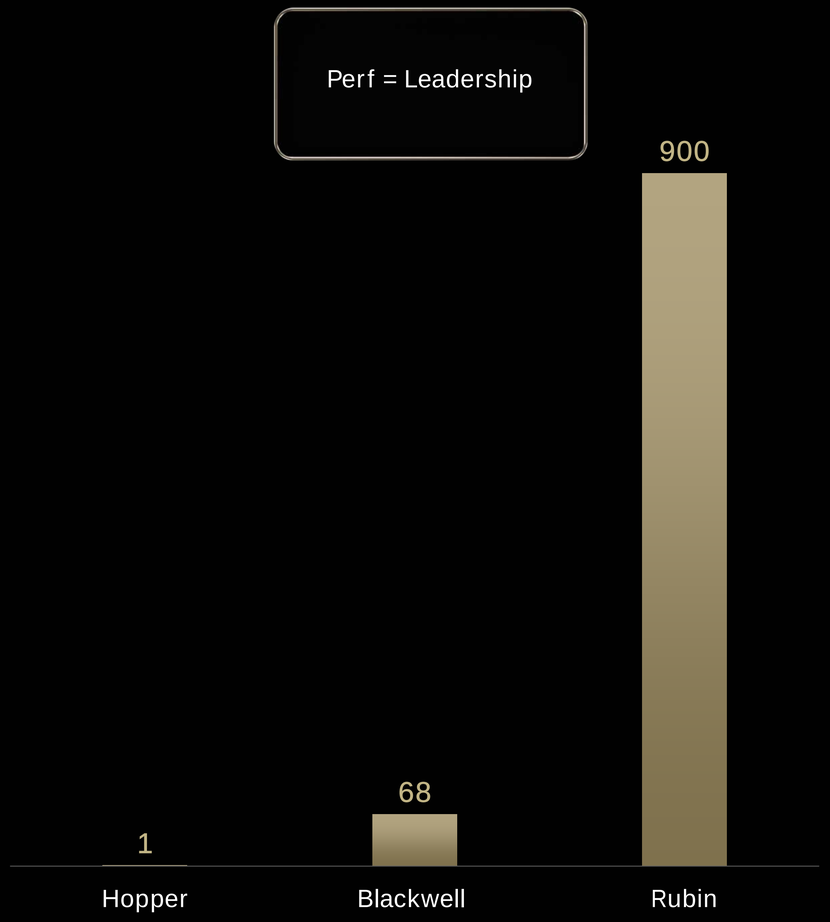
2,592 Grace CPU Cores

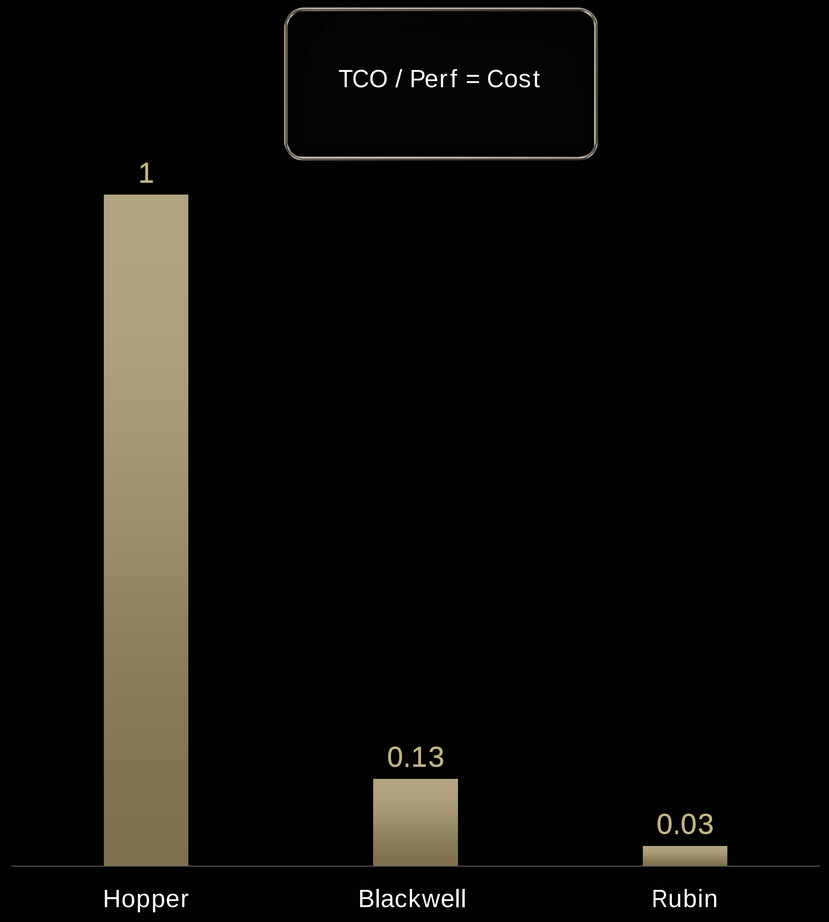
72 ConnectX-8 NICs

18 BlueField DPUs

Grace Blackwell NVLink72

Fast Roadmap Drives AI Factory Economics

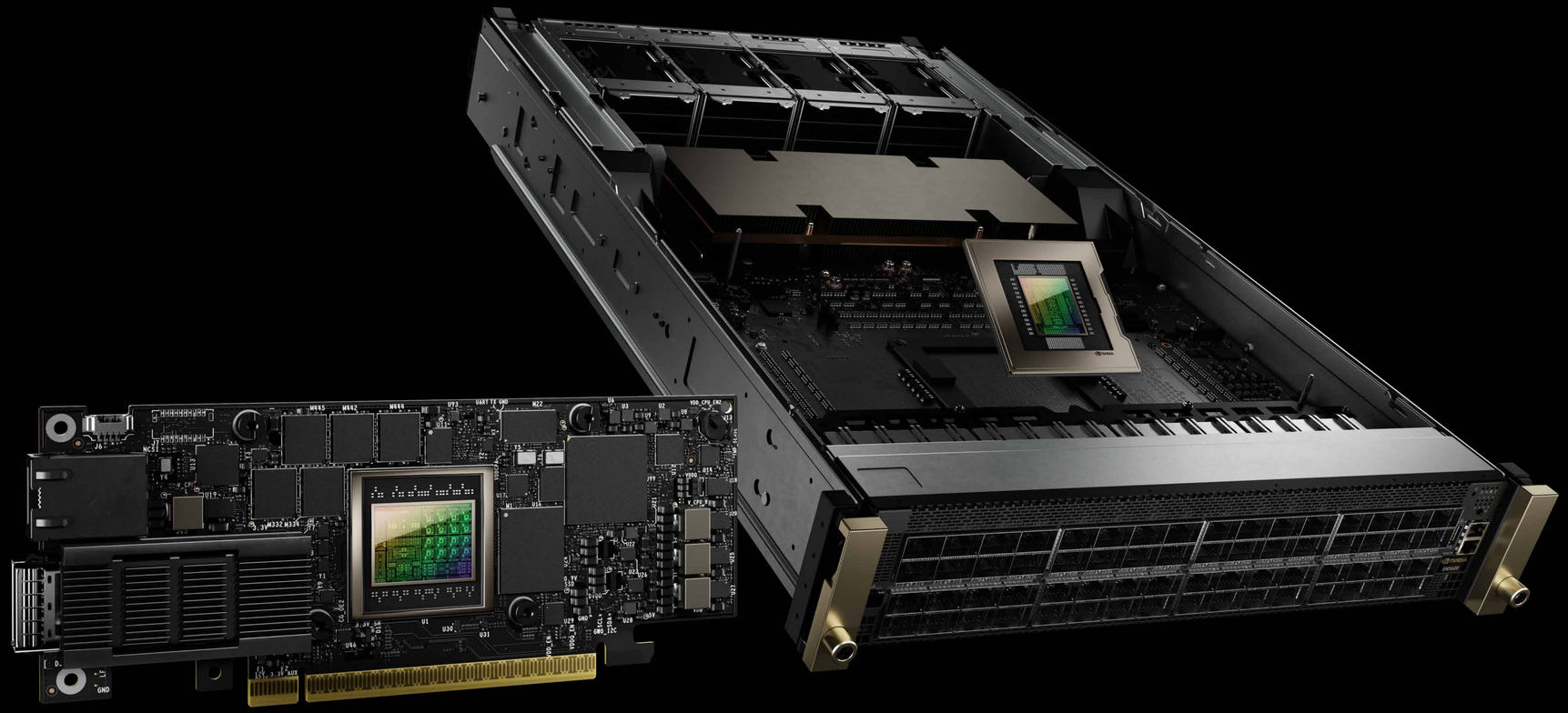




*Without sparsity

Spectrum-X “Supercharged” Ethernet

Powering the World’s Most Advanced AI Clouds and Factories



Spectrum-X SuperNIC

Spectrum-X Ethernet Switch

cisco

CoreWeave

DELL Technologies

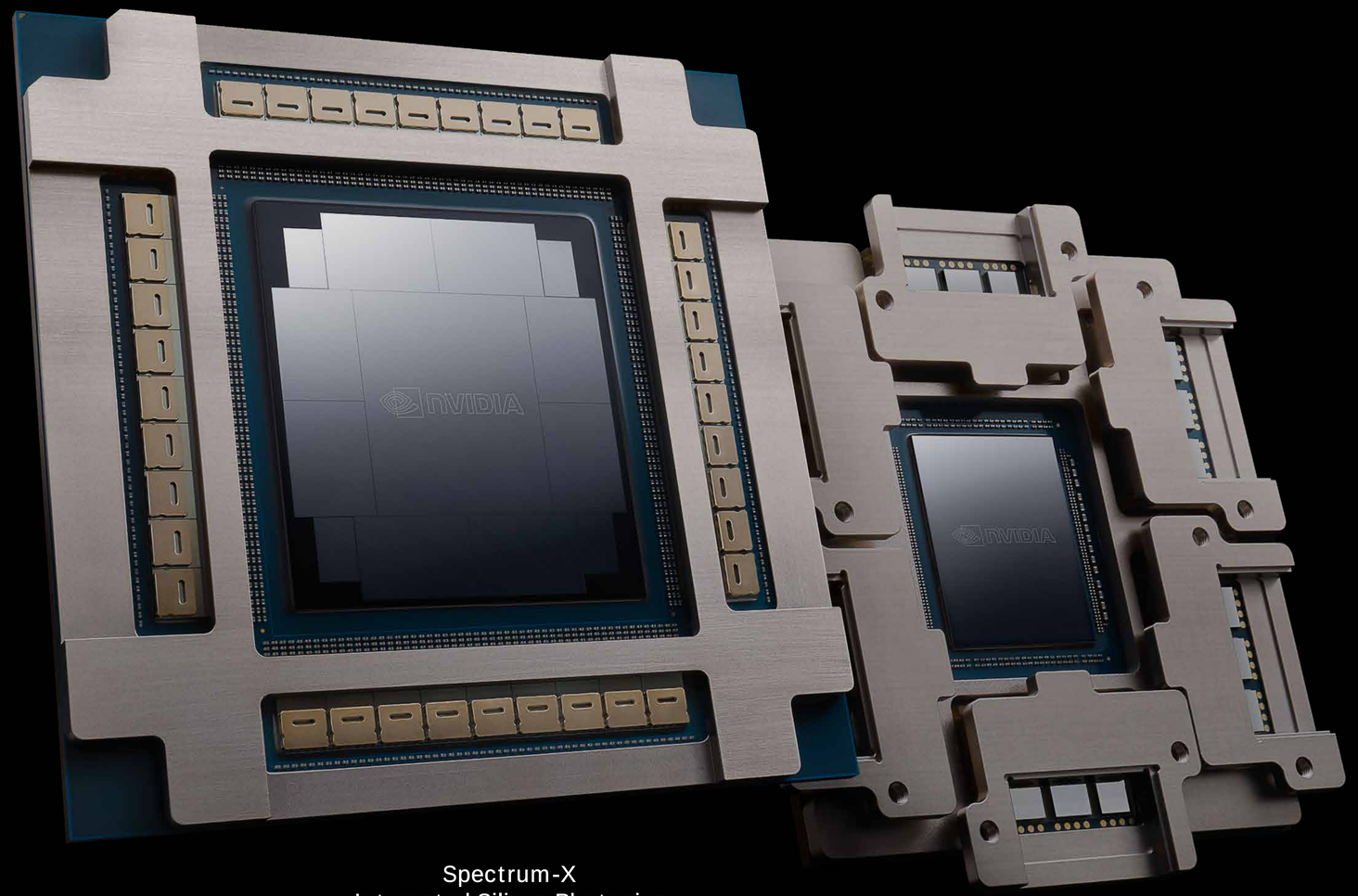
Google Cloud

Meta

Microsoft Azure

ORACLE CLOUD Infrastructure

xAI



Spectrum-X
Integrated Silicon Photonics
2nd Half 2026

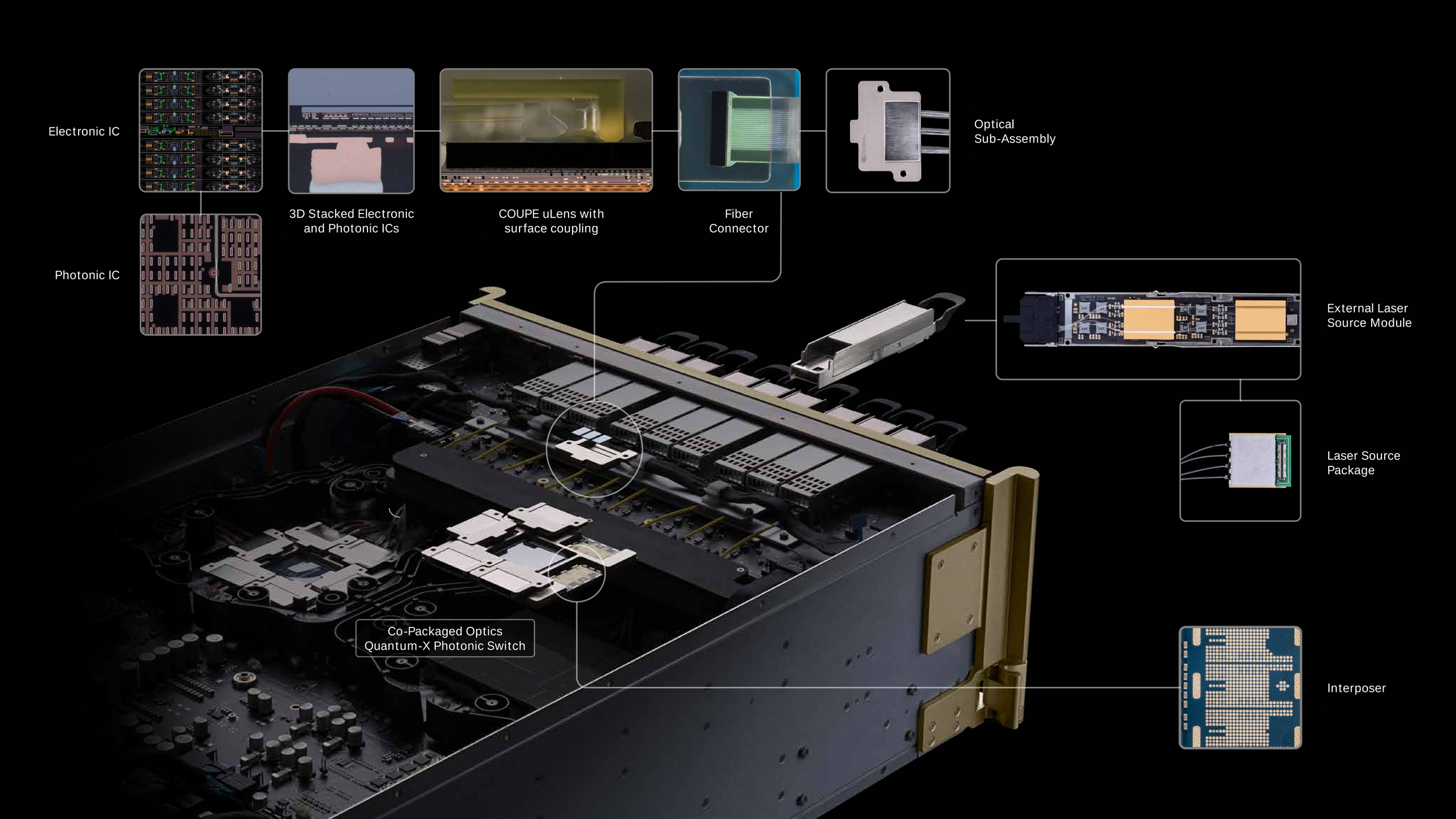
Quantum-X
Integrated Silicon Photonics
2nd Half 2025

NVIDIA Photonics

CPO Co-Invention With Ecosystem Partners

- 1st 1.6T Silicon Photonics CPO Chip - New Micro Ring Modulators (MRM)
- 1st 3D-Stacked Silicon Photonics Engine with TSMC Process
- High-Power, High-Efficiency Lasers
- Detachable Fiber Connectors
- 100’s of Patents, Licensed to Partners

波若威科技 BROWAVE CORPORATION · COHERENT · CORNING · fabrinet · FOXCONN HON HAI TECHNOLOGY GROUP · LUMENTUM · SENKO Advanced Components · SPIL · SUMITOMO ELECTRIC · TFC · tsmc


Electronic IC
Photonic IC
3D Stacked Electronic and Photonic ICs
COUPE uLens with surface coupling
Fiber Connector
Optical Sub-Assembly
External Laser Source Module
Laser Source Package
Co-Packaged Optics Quantum-X Photonic Switch
Interposer

Announcing NVIDIA Photonics Switch Systems

World’s Most Advanced Co-Packaged Optical Switches

Spectrum-X Photonics

2nd Half 2026

512 Ports of 800G | 2K x 200G

128 Ports of 800G | 512 x 200G

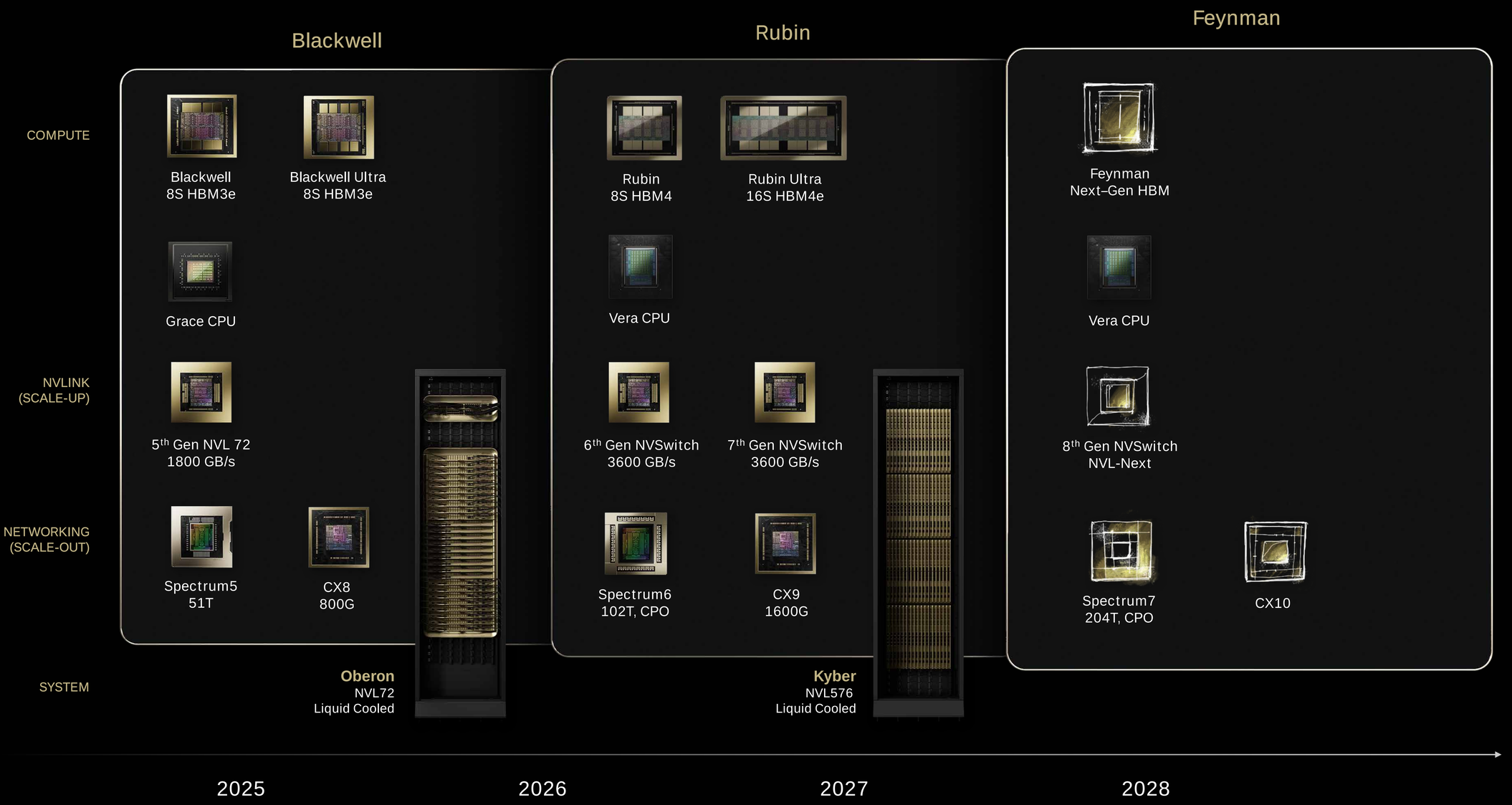
Quantum-X Photonics

2nd Half 2025

144 Ports of 800G | 576 x 200G

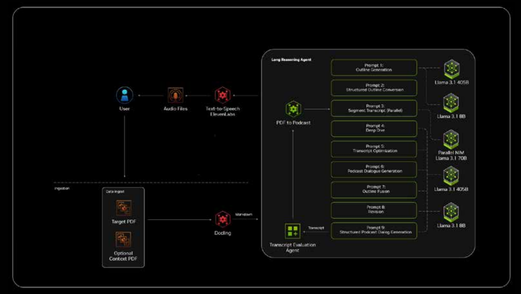
NVIDIA Paves Road to Gigawatt AI Factories

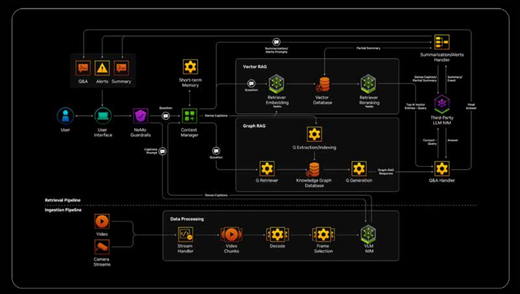
One-Year Rhythm | Full-Stack | One Architecture | CUDA Everywhere

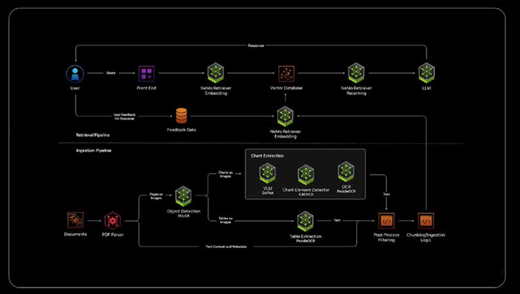


Reinventing $500B Enterprise IT For the Age of AI

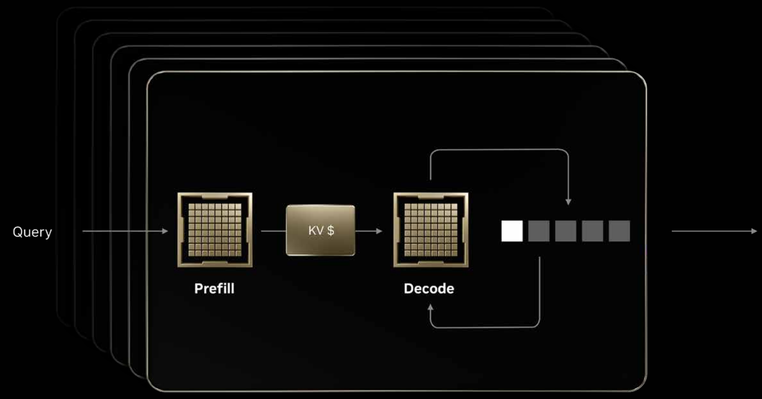
AI



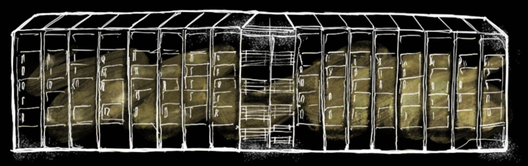




OS

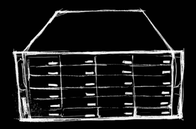


INFRASTRUCTURE



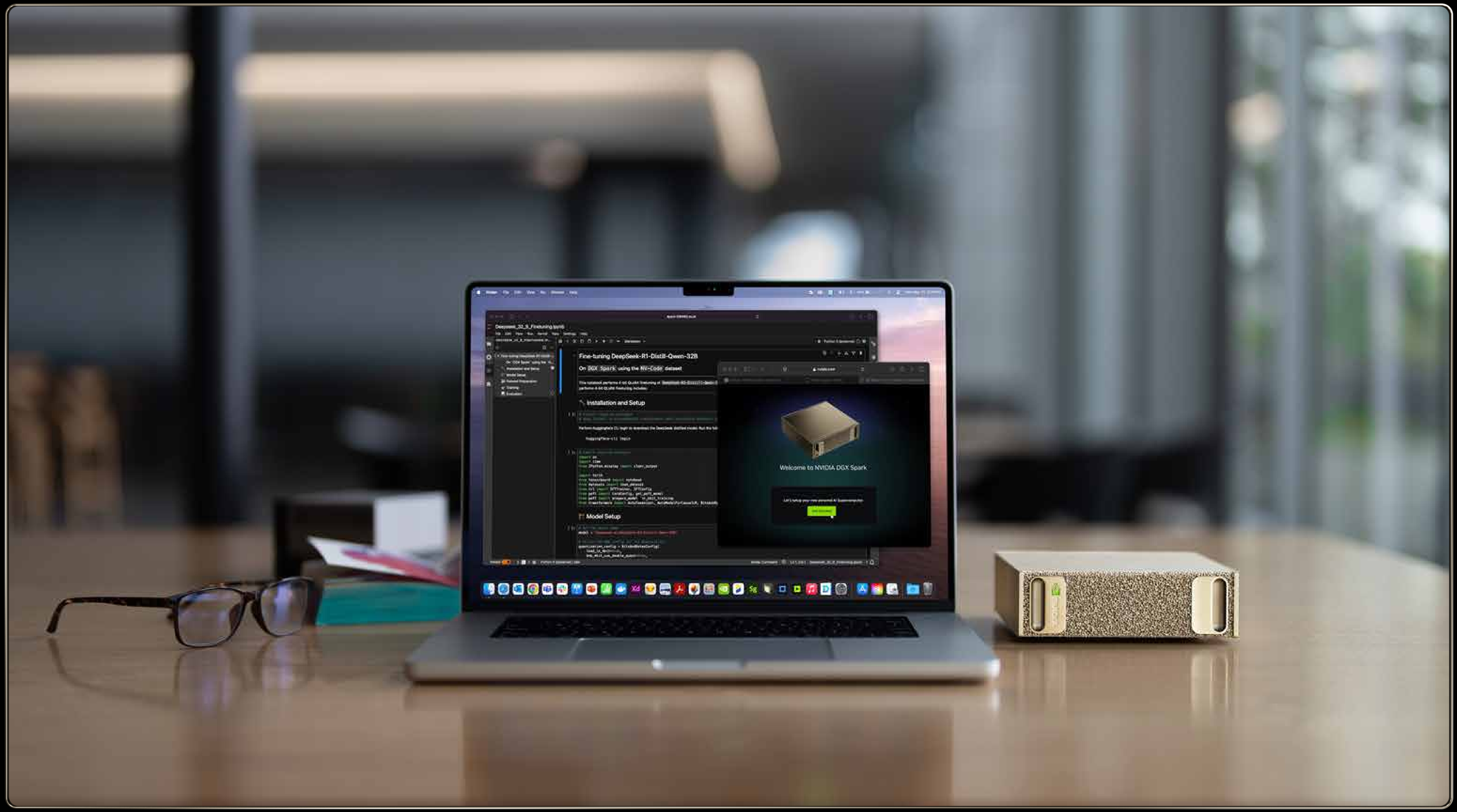
Compute

Networking



Storage


NVIDIA


Fine-tuning DeepSeek-R1-Distill-Qwen-32B
On DGX Spark using the NVIDIA-Code dataset
Installation and Setup
Model Setup
Welcome to NVIDIA DGX Spark

DGX Station

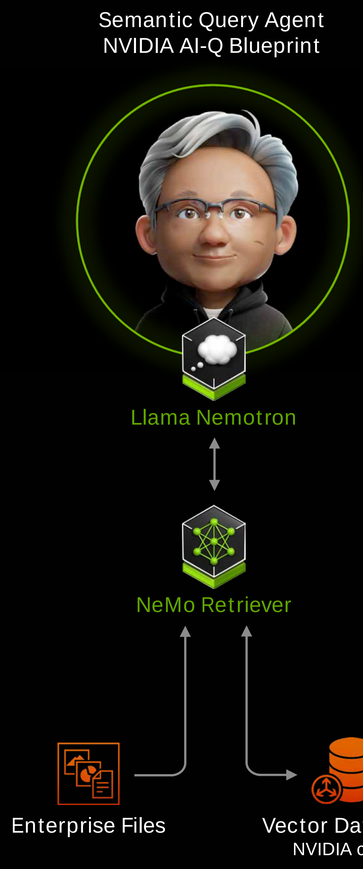
The Ultimate Workstation for AI and Data Science

- GB300 Superchip
- 784 GB Unified System Memory
- 20,000 AI TFLOPS
- ConnectX-8 SuperNIC

ASUS BOXX DELL Technologies hp Lambda Supermicro

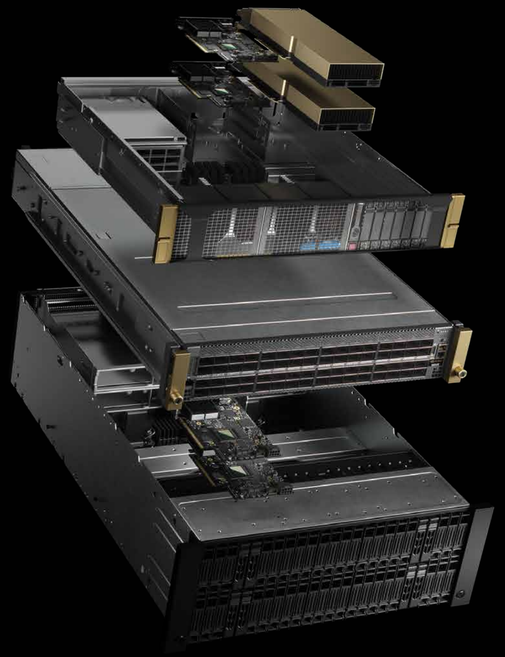
NVIDIA AI Infrastructure for Enterprise Computing

Announcing Storage Leaders Build AI Data Platforms for Enterprise AI



ddn · DELL Technologies · Hewlett Packard Enterprise · Hitachi Vantara · IBM

NetApp · NUTANIX · PURESTORAGE · VAST · WEKA



Compute Node

NVIDIA Blackwell

NVIDIA BlueField

NVIDIA Spectrum-X

Storage Node

NVIDIA BlueField / ConnectX

DELL Technologies + NVIDIA

INDUSTRY'S BROADEST END-TO-END NVIDIA AI ENTERPRISE INFRASTRUCTURE


AI PCs


Spark


Station


Compute


Networking


Storage


Blackwell Compute Racks

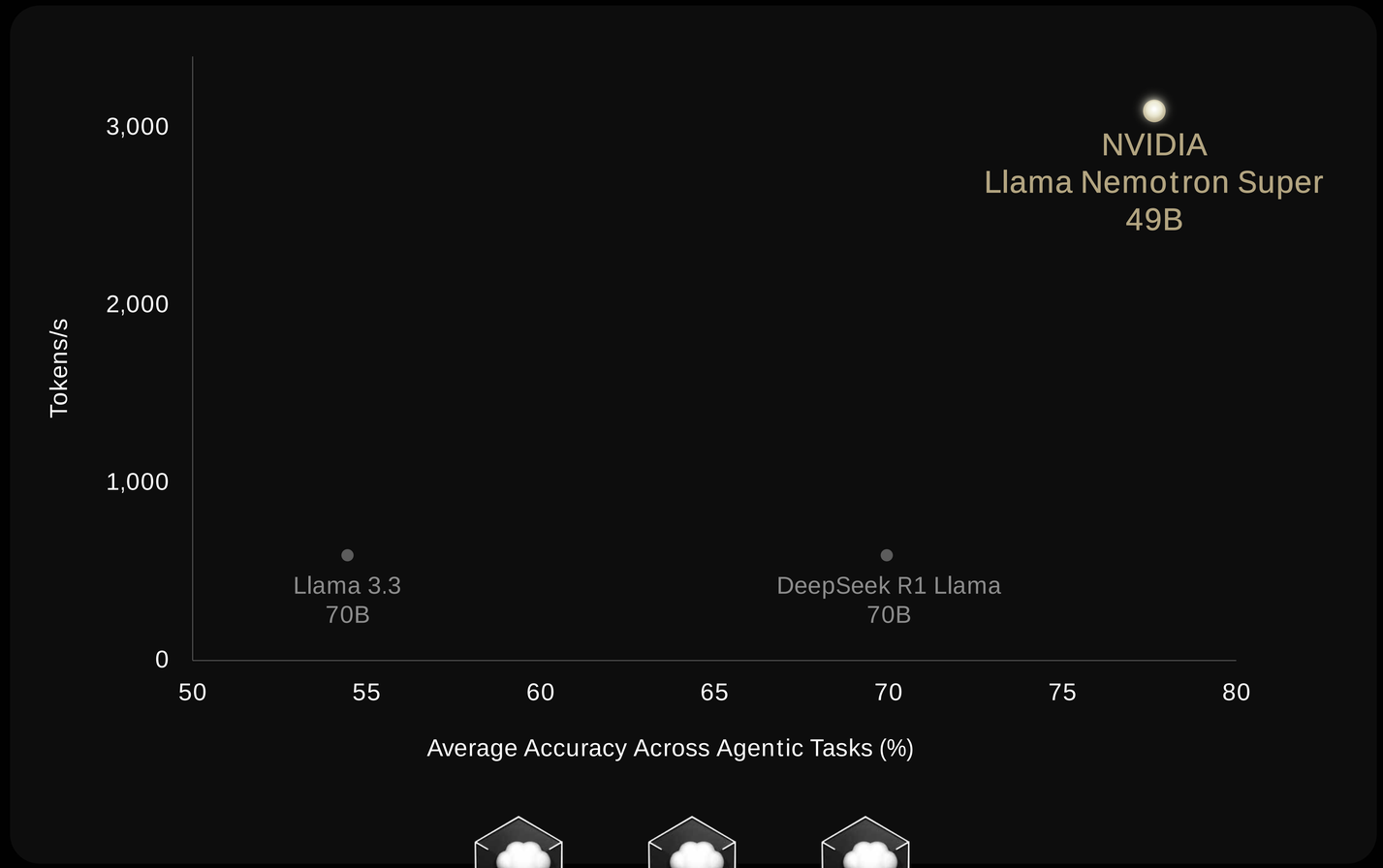
2,000+ Customers

100+ New releases

49+ Exaflops

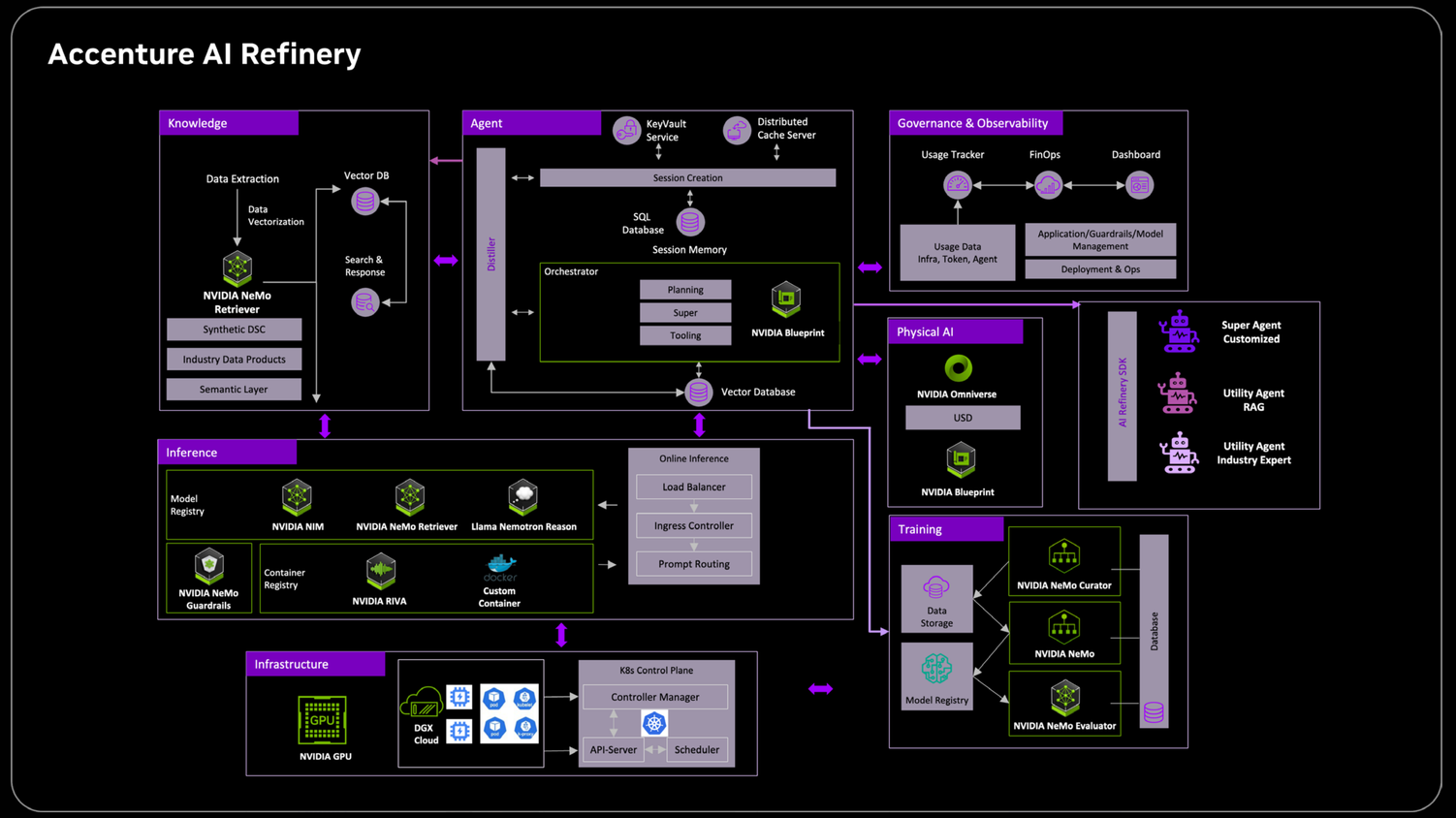
Announcing NVIDIA Llama Nemotron Reasoning

Distilled, Quantized, Aligned, and Optimized by NVIDIA

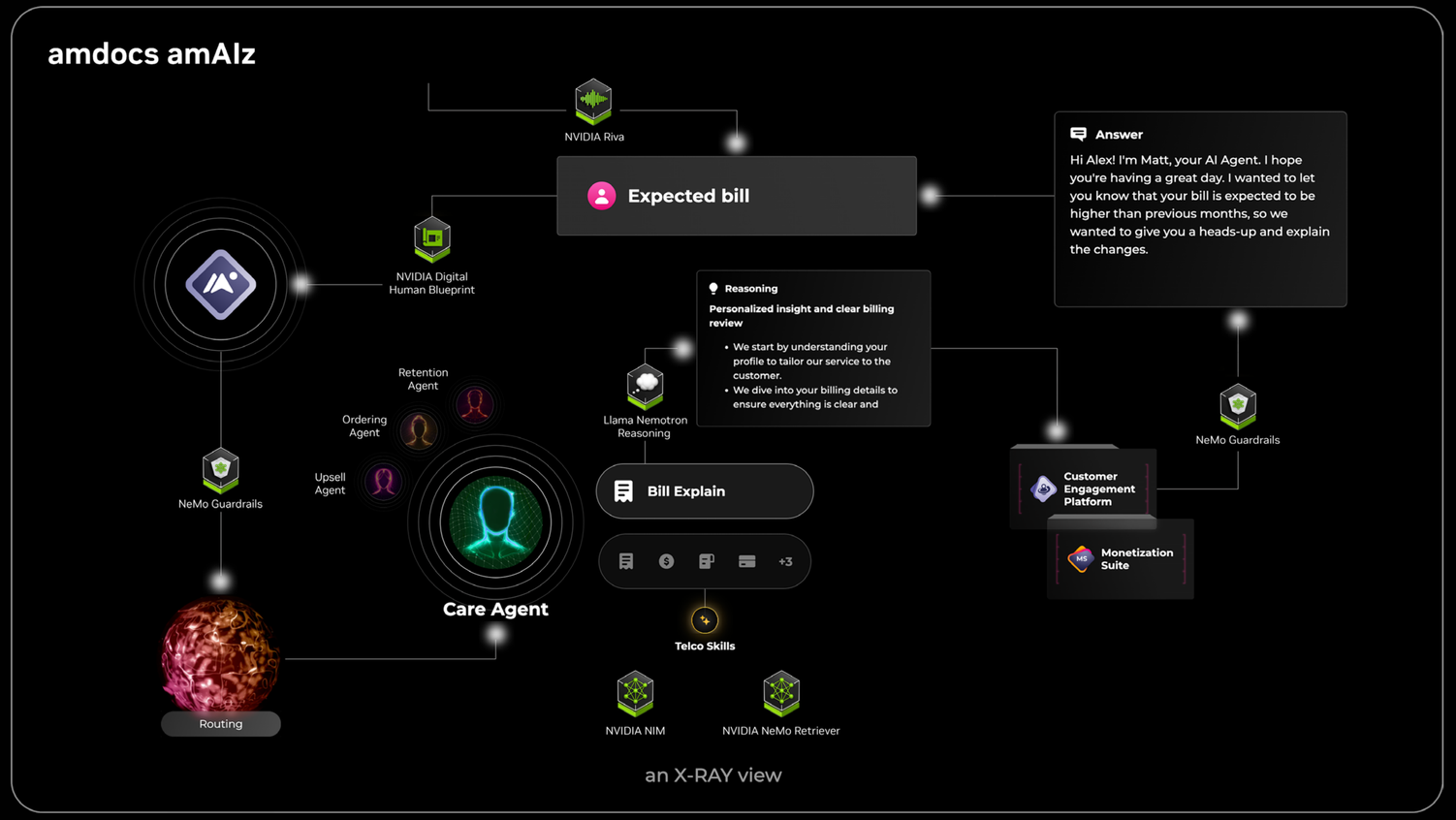


Nano Super Ultra

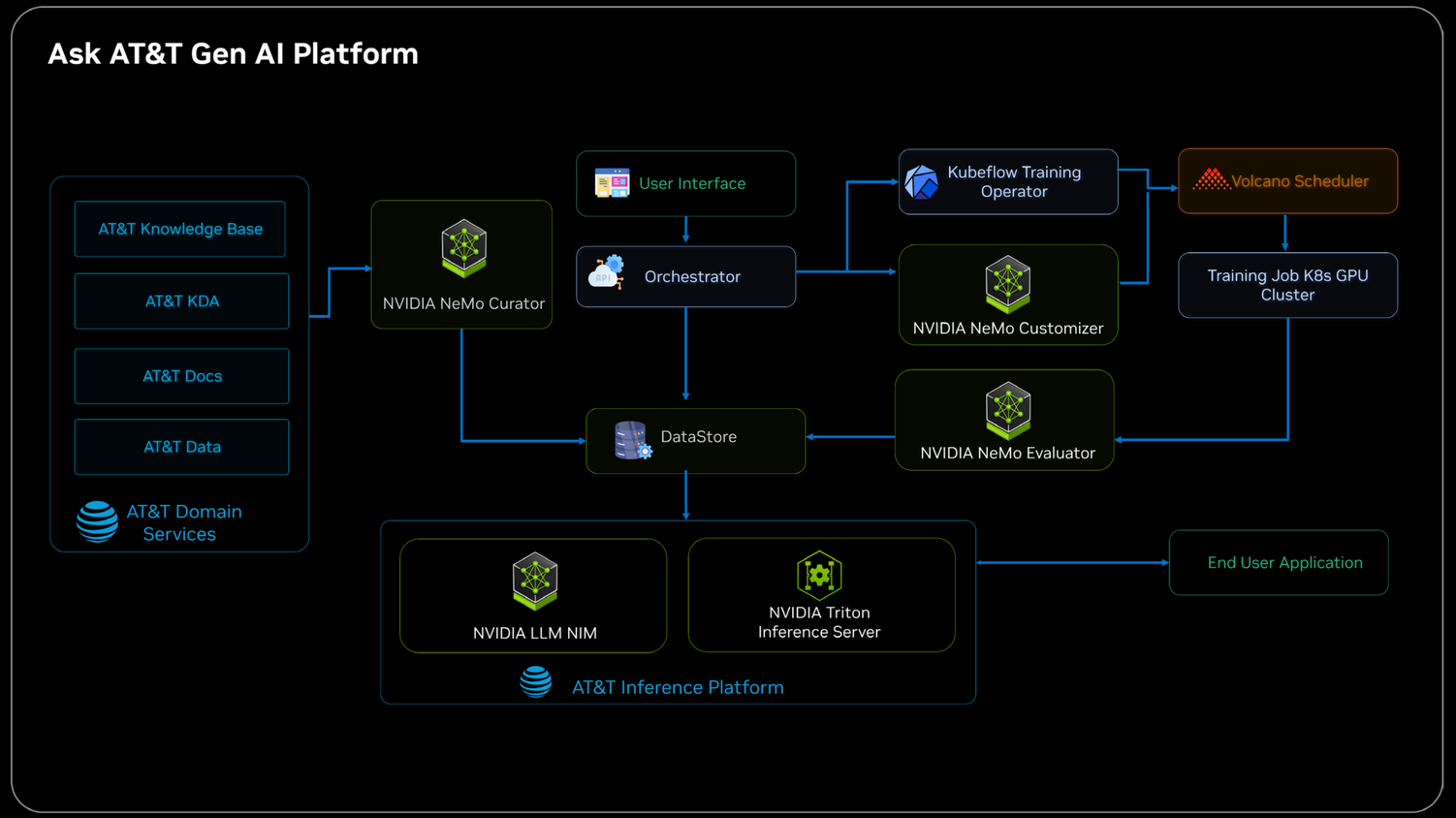
accenture


Accenture AI Refinery
Knowledge
Data Extraction
Data Vectorization
Vector DB
Search & Response
NVIDIA NeMo Retriever
Synthetic DSC
Industry Data Products
Semantic Layer
Agent
KeyVault Service
Distributed Cache Server
Distiller
Session Creation
SQL Database
Session Memory
Orchestrator
Planning
Super
Tooling
NVIDIA Blueprint
Vector Database
Governance & Observability
Usage Tracker
FinOps
Dashboard
Usage Data Infra, Token, Agent
Application/Guardrails/Model Management
Deployment & Ops
Physical AI
NVIDIA Omniverse
USD
NVIDIA Blueprint
AI Refinery SDK
Super Agent Customized
Utility Agent RAG
Utility Agent Industry Expert
Inference
Model Registry
NVIDIA NIM
NVIDIA NeMo Retriever
Llama Nemotron Reason
Online Inference
Load Balancer
Ingress Controller
Prompt Routing
NVIDIA NeMo Guardrails
Container Registry
NVIDIA RIVA
Custom Container
Training
Data Storage
Model Registry
NVIDIA NeMo Curator
NVIDIA NeMo
NVIDIA NeMo Evaluator
Database
Infrastructure
GPU
NVIDIA GPU
DGX Cloud
K8s Control Plane
Controller Manager
API-Server
Scheduler

amdocs

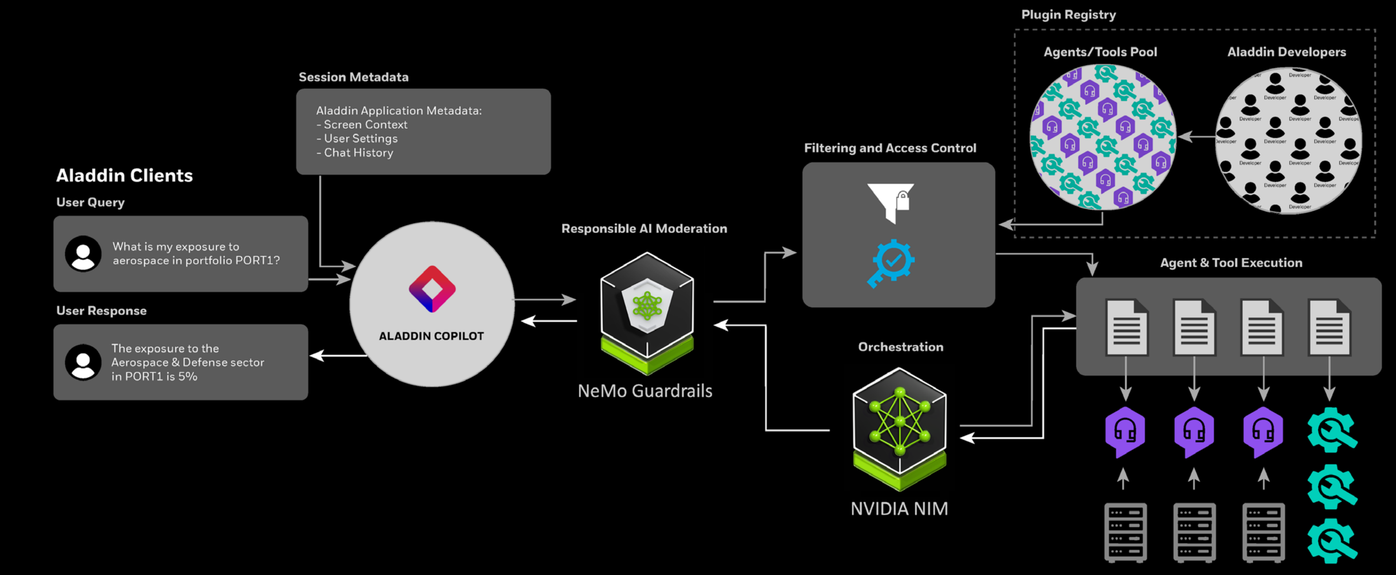

amdocs amAIz
NVIDIA Riva
Expected bill
Answer
Hi Alex! I'm Matt, your AI Agent. I hope you're having a great day. I wanted to let you know that your bill is expected to be higher than previous months, so we wanted to give you a heads-up and explain the changes.
NVIDIA Digital Human Blueprint
Reasoning
Personalized insight and clear billing review
• We start by understanding your profile to tailor our service to the customer.
• We dive into your billing details to ensure everything is clear and
Retention Agent
Ordering Agent
Upsell Agent
Llama Nemotron Reasoning
NeMo Guardrails
NeMo Guardrails
Bill Explain
Customer Engagement Platform
Monetization Suite
+3
Care Agent
Telco Skills
Routing
NVIDIA NIM
NVIDIA NeMo Retriever
an X-RAY view

AT&T

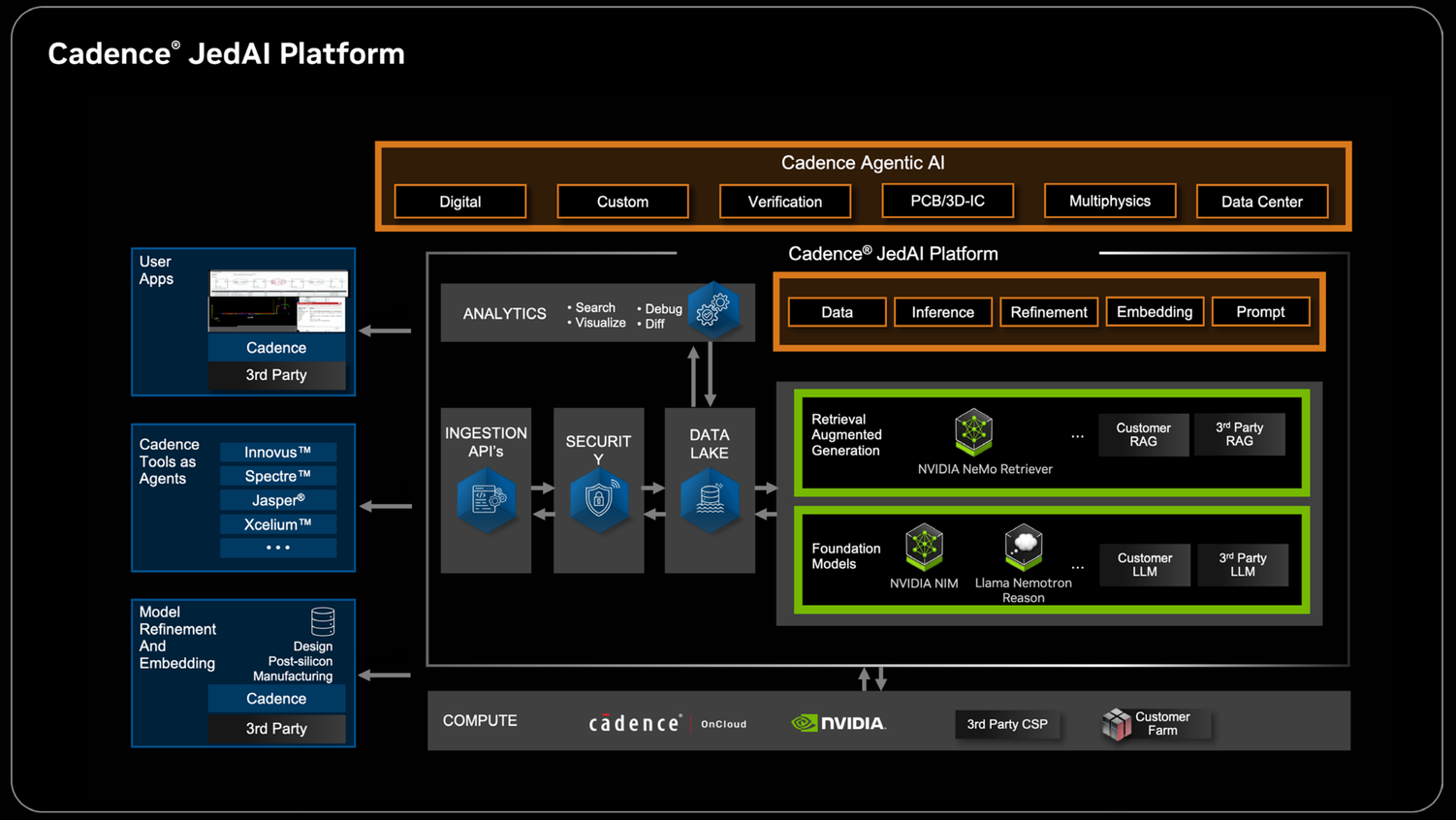

Ask AT&T Gen AI Platform
AT&T Knowledge Base
AT&T KDA
AT&T Docs
AT&T Data
AT&T Domain Services
NVIDIA NeMo Curator
User Interface
Orchestrator
Kubeflow Training Operator
Volcano Scheduler
NVIDIA NeMo Customizer
Training Job K8s GPU Cluster
DataStore
NVIDIA NeMo Evaluator
NVIDIA LLM NIM
NVIDIA Triton Inference Server
AT&T Inference Platform
End User Application

BlackRock®

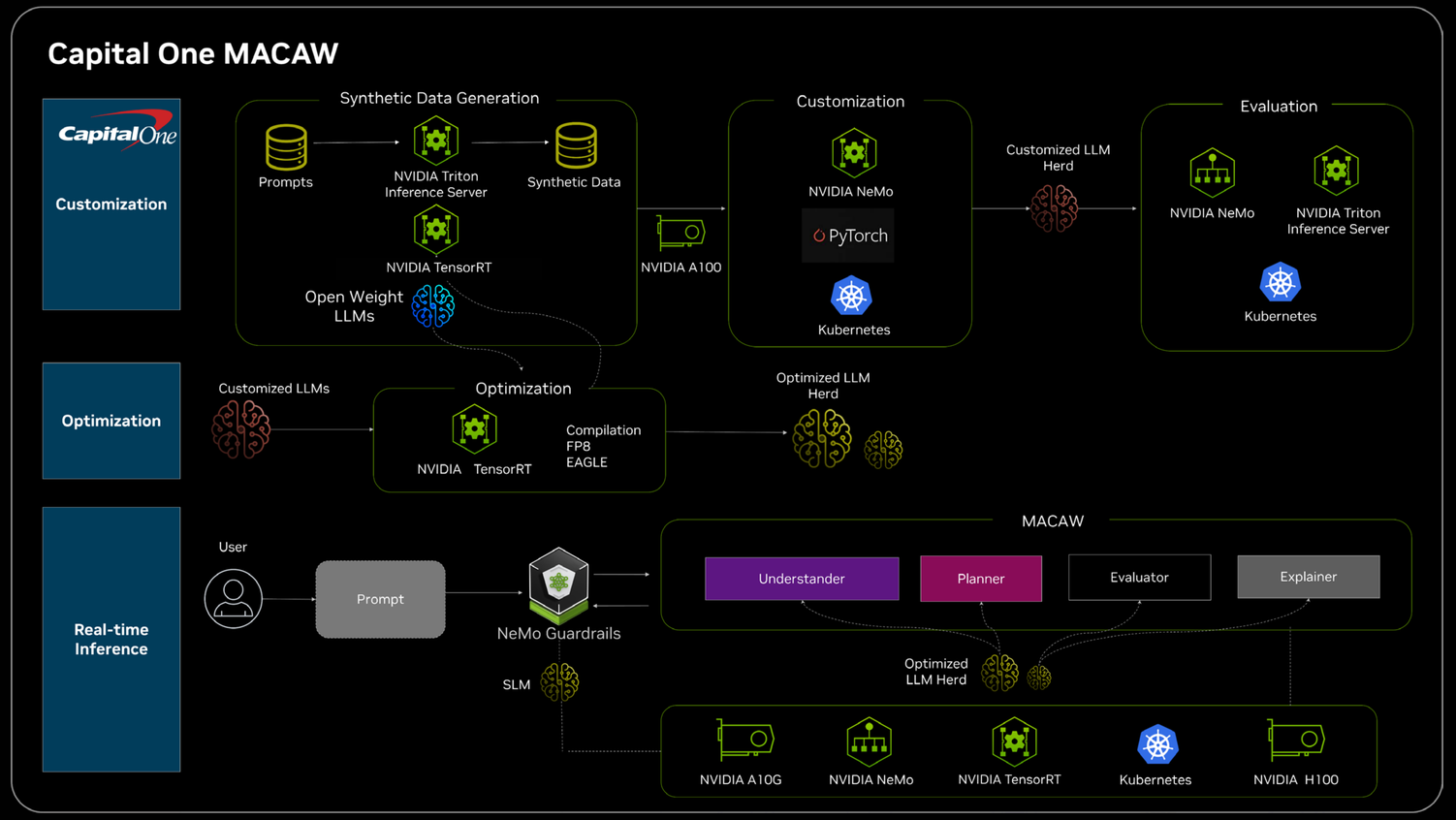
BlackRock's Aladdin Platform



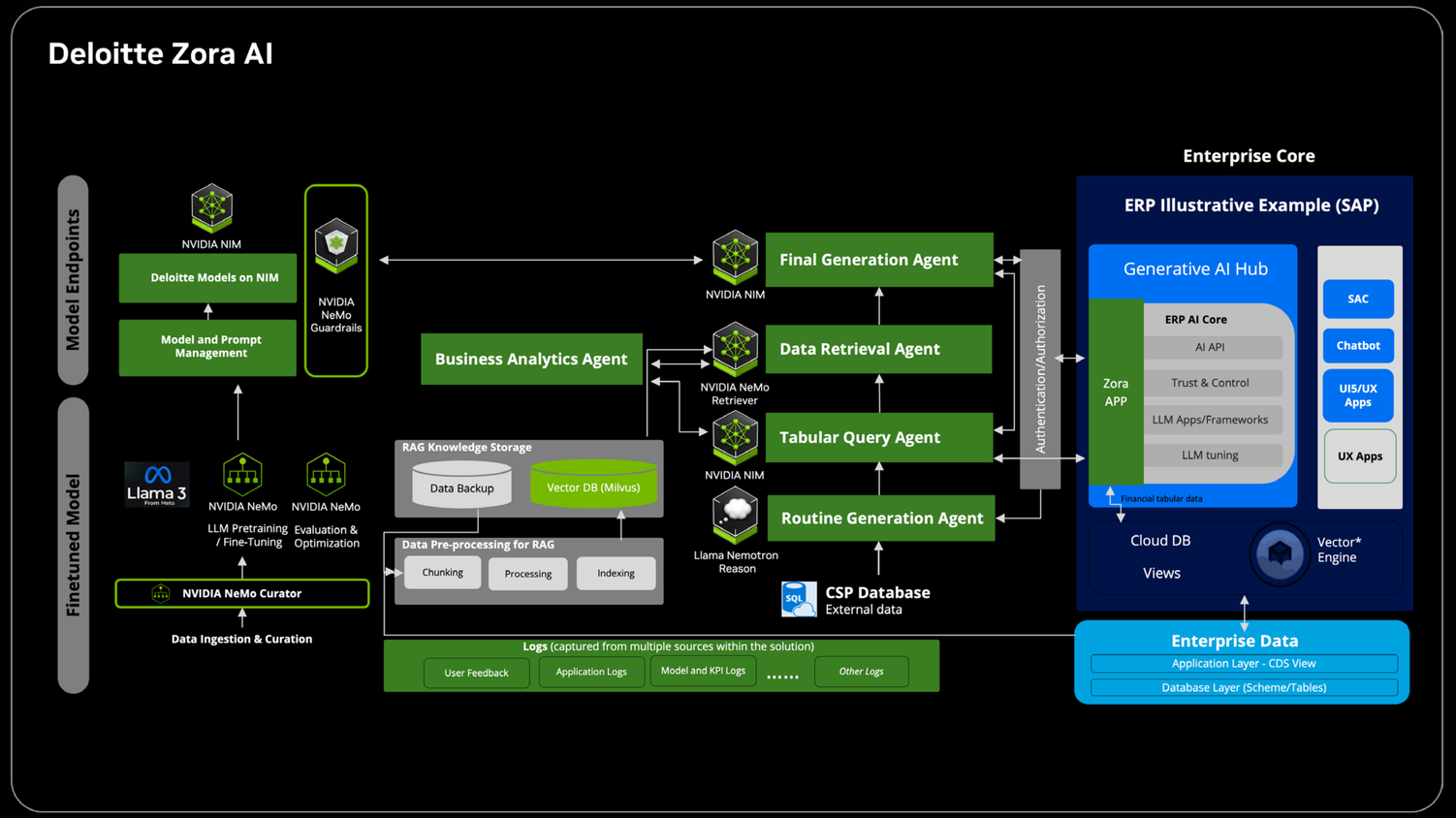
cadence®

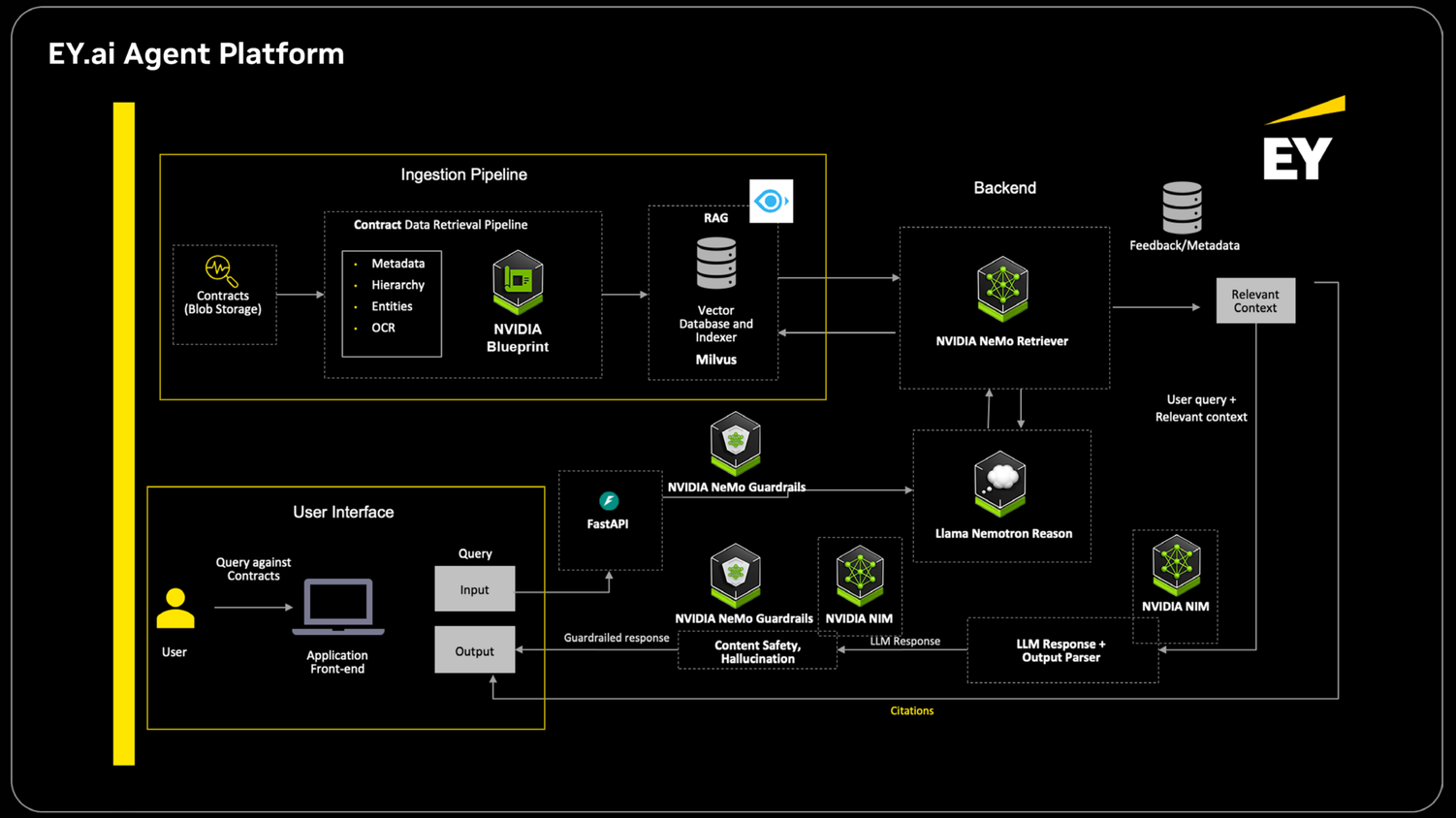

Cadence® JedAI Platform
Cadence Agentic AI
Digital
Custom
Verification
PCB/3D-IC
Multiphysics
Data Center
User Apps
Cadence
3rd Party
Cadence Tools as Agents
Innovus™
Spectre™
Jasper®
Xcelium™
• • •
Model Refinement And Embedding
Design
Post-silicon
Manufacturing
Cadence
3rd Party
Cadence® JedAI Platform
ANALYTICS
• Search
• Visualize
• Debug
• Diff
Data
Inference
Refinement
Embedding
Prompt
INGESTION API's
SECURITY
DATA LAKE
Retrieval Augmented Generation
NVIDIA NeMo Retriever
...
Customer RAG
3rd Party RAG
Foundation Models
NVIDIA NIM
Llama Nemotron Reason
...
Customer LLM
3rd Party LLM
COMPUTE
cadence OnCloud
NVIDIA
3rd Party CSP
Customer Farm

Capital One

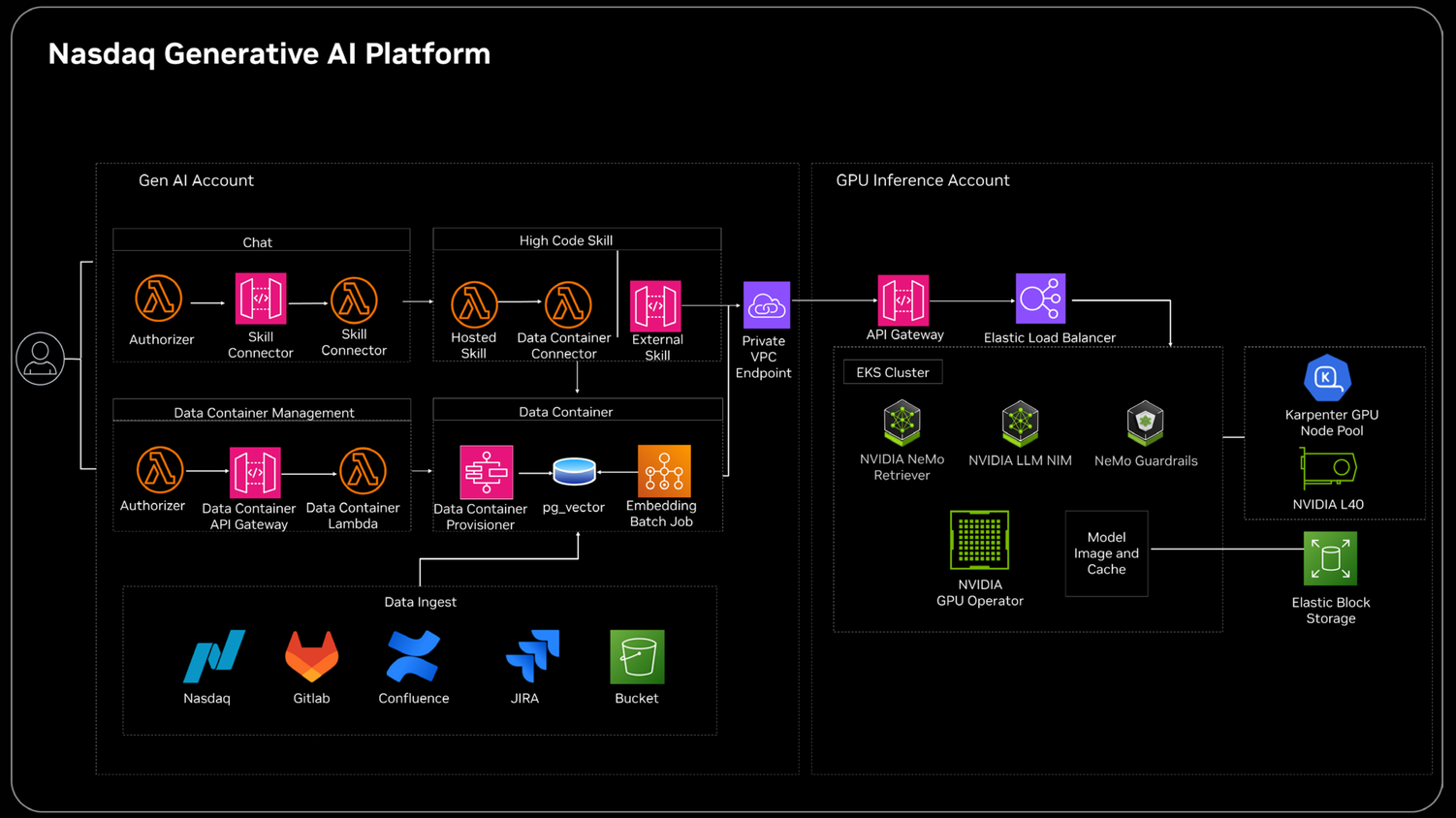

Capital One MACAW
Customization
Synthetic Data Generation
Prompts
NVIDIA Triton Inference Server
Synthetic Data
NVIDIA TensorRT
Open Weight LLMs
NVIDIA A100
Customization
NVIDIA NeMo
PyTorch
Kubernetes
Customized LLM Herd
Evaluation
NVIDIA NeMo
NVIDIA Triton Inference Server
Kubernetes
Optimization
Customized LLMs
Optimization
NVIDIA TensorRT
Compilation
FP8
EAGLE
Optimized LLM Herd
Real-time Inference
User
Prompt
NeMo Guardrails
SLM
MACAW
Understander
Planner
Evaluator
Explainer
Optimized LLM Herd
NVIDIA A10G
NVIDIA NeMo
NVIDIA TensorRT
Kubernetes
NVIDIA H100

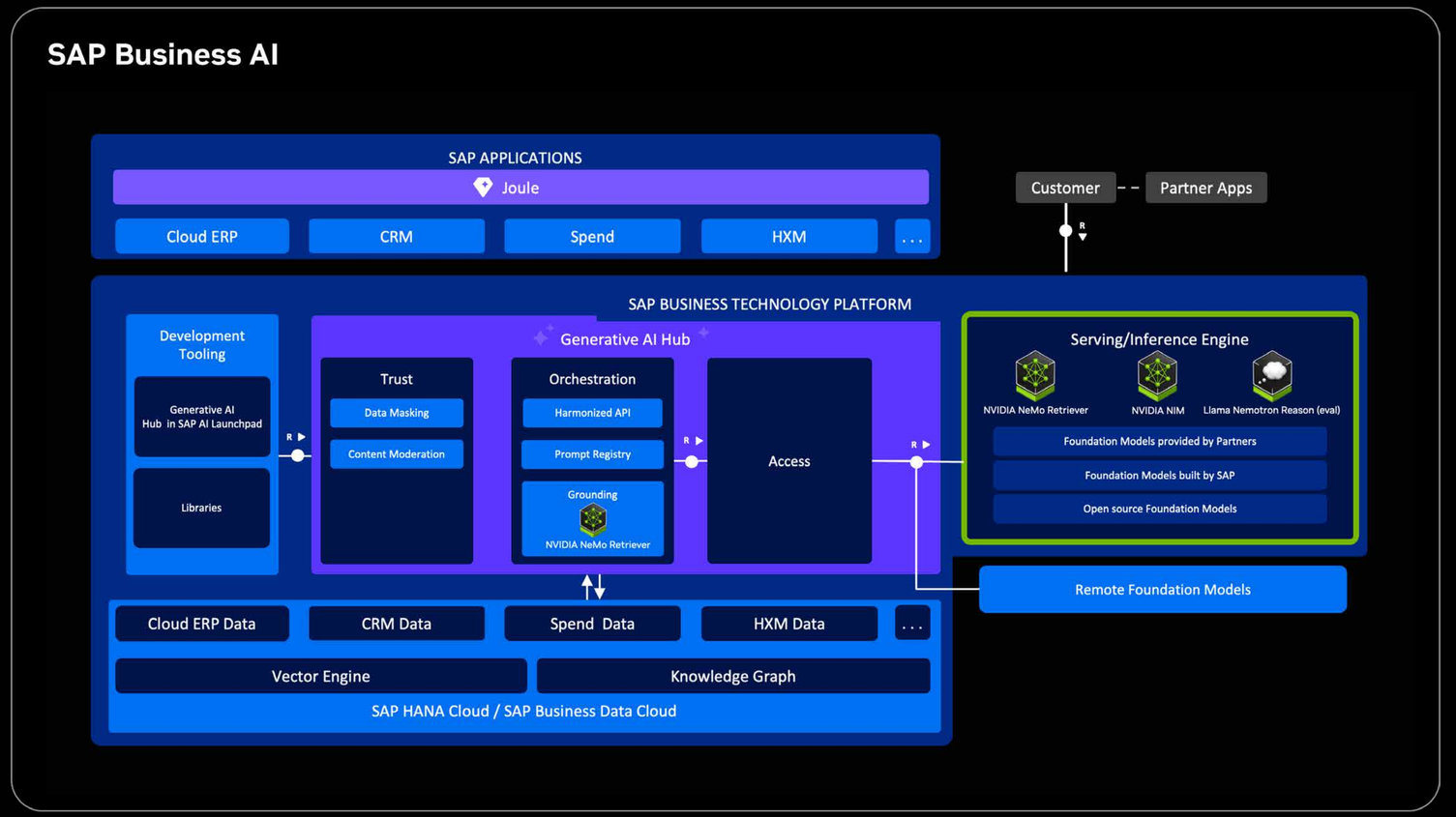
Deloitte.


Deloitte Zora AI
Model Endpoints
Finetuned Model
NVIDIA NIM
Deloitte Models on NIM
Model and Prompt Management
NVIDIA NeMo Guardrails
Llama 3
NVIDIA NeMo
LLM Pretraining / Fine-Tuning
NVIDIA NeMo
Evaluation & Optimization
NVIDIA NeMo Curator
Data Ingestion & Curation
Business Analytics Agent
RAG Knowledge Storage
Data Backup
Vector DB (Milvus)
Data Pre-processing for RAG
Chunking
Processing
Indexing
NVIDIA NIM
NVIDIA NeMo Retriever
NVIDIA NIM
Llama Nemotron Reason
Final Generation Agent
Data Retrieval Agent
Tabular Query Agent
Routine Generation Agent
CSP Database
External data
Authentication/Authorization
Enterprise Core
ERP Illustrative Example (SAP)
Generative AI Hub
Zora APP
ERP AI Core
AI API
Trust & Control
LLM Apps/Frameworks
LLM tuning
Financial tabular data
SAC
Chatbot
UI5/UX Apps
UX Apps
Cloud DB
Views
Vector* Engine
Enterprise Data
Application Layer - CDS View
Database Layer (Scheme/Tables)
Logs (captured from multiple sources within the solution)
User Feedback
Application Logs
Model and KPI Logs
.....
Other Logs

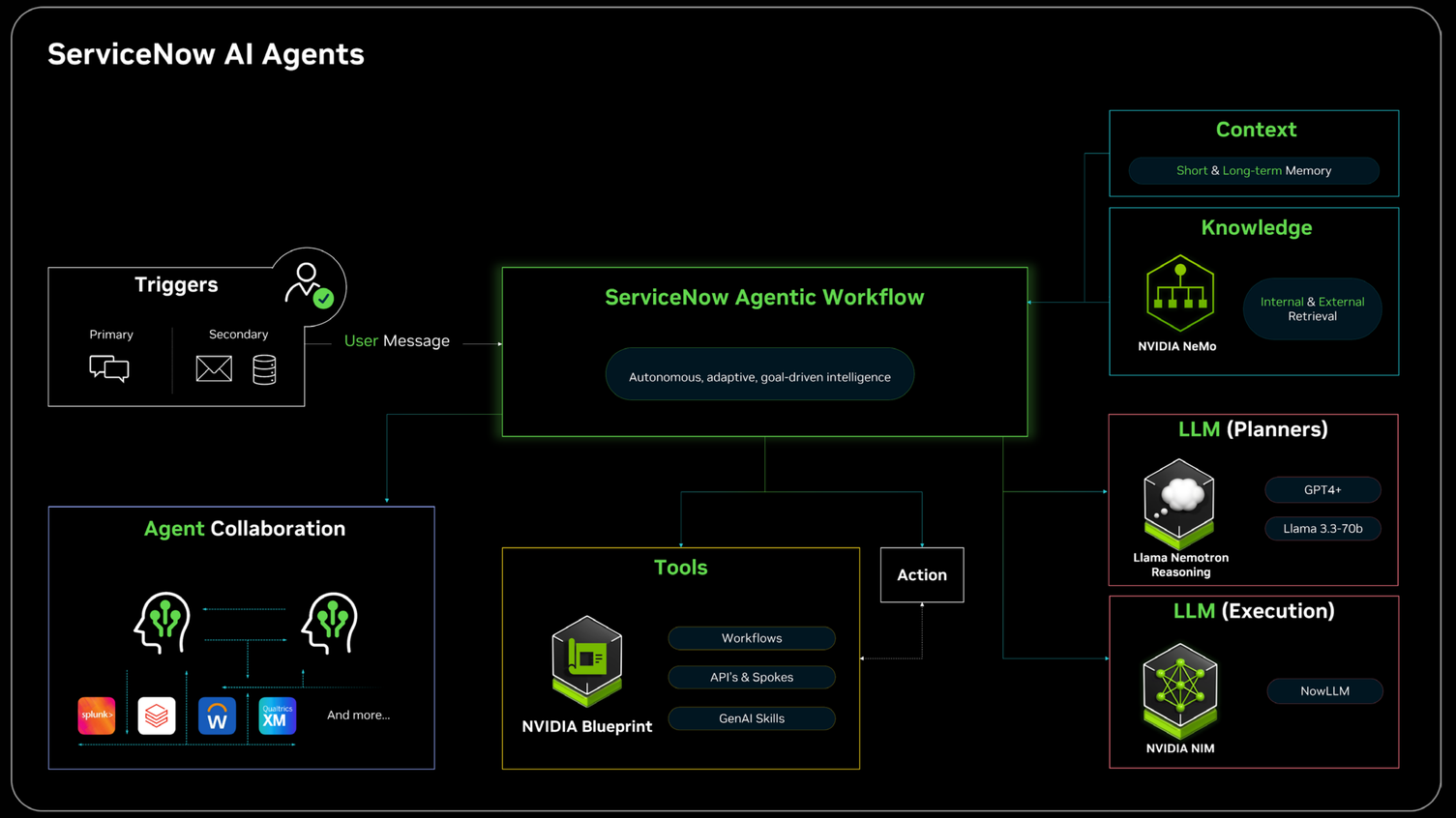

EY.ai Agent Platform
Ingestion Pipeline
Contracts (Blob Storage)
Contract Data Retrieval Pipeline
Metadata
Hierarchy
Entities
OCR
NVIDIA Blueprint
RAG
Vector Database and Indexer
Milvus
Backend
NVIDIA NeMo Retriever
Feedback/Metadata
Relevant Context
User query + Relevant context
NVIDIA NeMo Guardrails
Llama Nemotron Reason
FastAPI
User Interface
Query against Contracts
User
Application Front-end
Query
Input
Output
NVIDIA NeMo Guardrails
NVIDIA NIM
NVIDIA NIM
Guardrailed response
Content Safety, Hallucination
LLM Response
LLM Response + Output Parser
Citations

Nasdaq

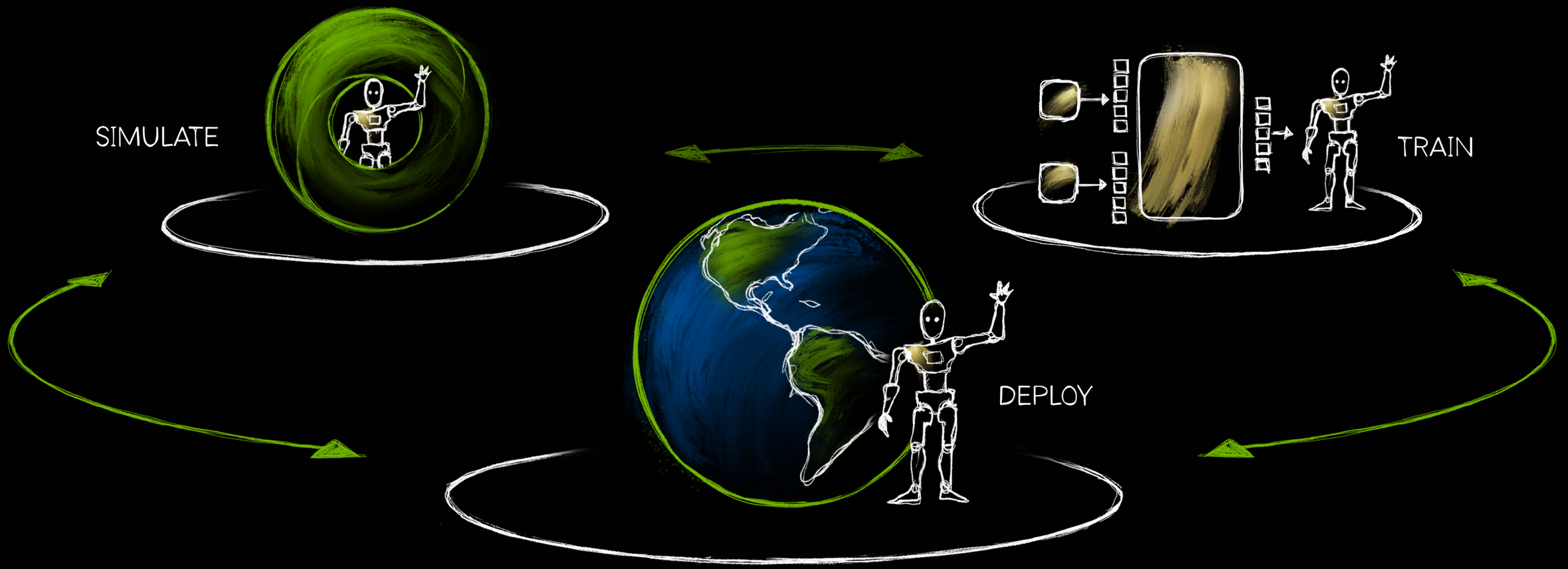

Nasdaq Generative AI Platform
Gen AI Account
Chat
Authorizer
Skill Connector
Skill Connector
High Code Skill
Hosted Skill
Data Container Connector
External Skill
Private VPC Endpoint
Data Container Management
Authorizer
Data Container API Gateway
Data Container Lambda
Data Container
Data Container Provisioner
pg_vector
Embedding Batch Job
Data Ingest
Nasdaq
Gitlab
Confluence
JIRA
Bucket
GPU Inference Account
API Gateway
Elastic Load Balancer
EKS Cluster
NVIDIA NeMo Retriever
NVIDIA LLM NIM
NeMo Guardrails
NVIDIA GPU Operator
Model Image and Cache
Karpenter GPU Node Pool
NVIDIA L40
Elastic Block Storage

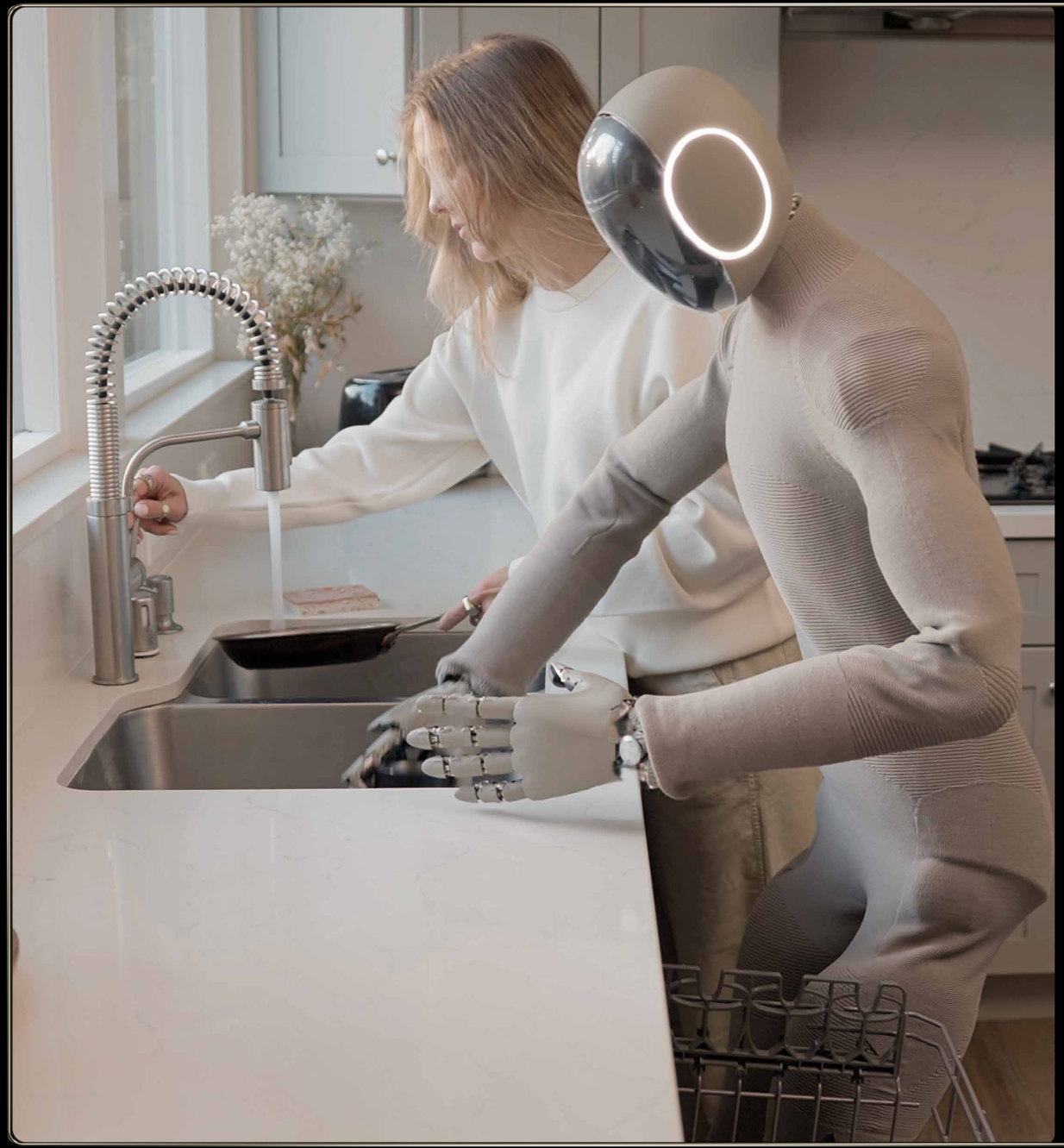

SAP Business AI
SAP APPLICATIONS
Joule
Cloud ERP
CRM
Spend
HXM
. . .
Customer
Partner Apps
R
SAP BUSINESS TECHNOLOGY PLATFORM
Development Tooling
Generative AI Hub in SAP AI Launchpad
Libraries
R
Generative AI Hub
Trust
Data Masking
Content Moderation
Orchestration
Harmonized API
Prompt Registry
Grounding
NVIDIA NeMo Retriever
R
Access
R
Serving/Inference Engine
NVIDIA NeMo Retriever
NVIDIA NIM
Llama Nemotron Reason (eval)
Foundation Models provided by Partners
Foundation Models built by SAP
Open source Foundation Models
Remote Foundation Models
Cloud ERP Data
CRM Data
Spend Data
HXM Data
. . .
Vector Engine
Knowledge Graph
SAP HANA Cloud / SAP Business Data Cloud

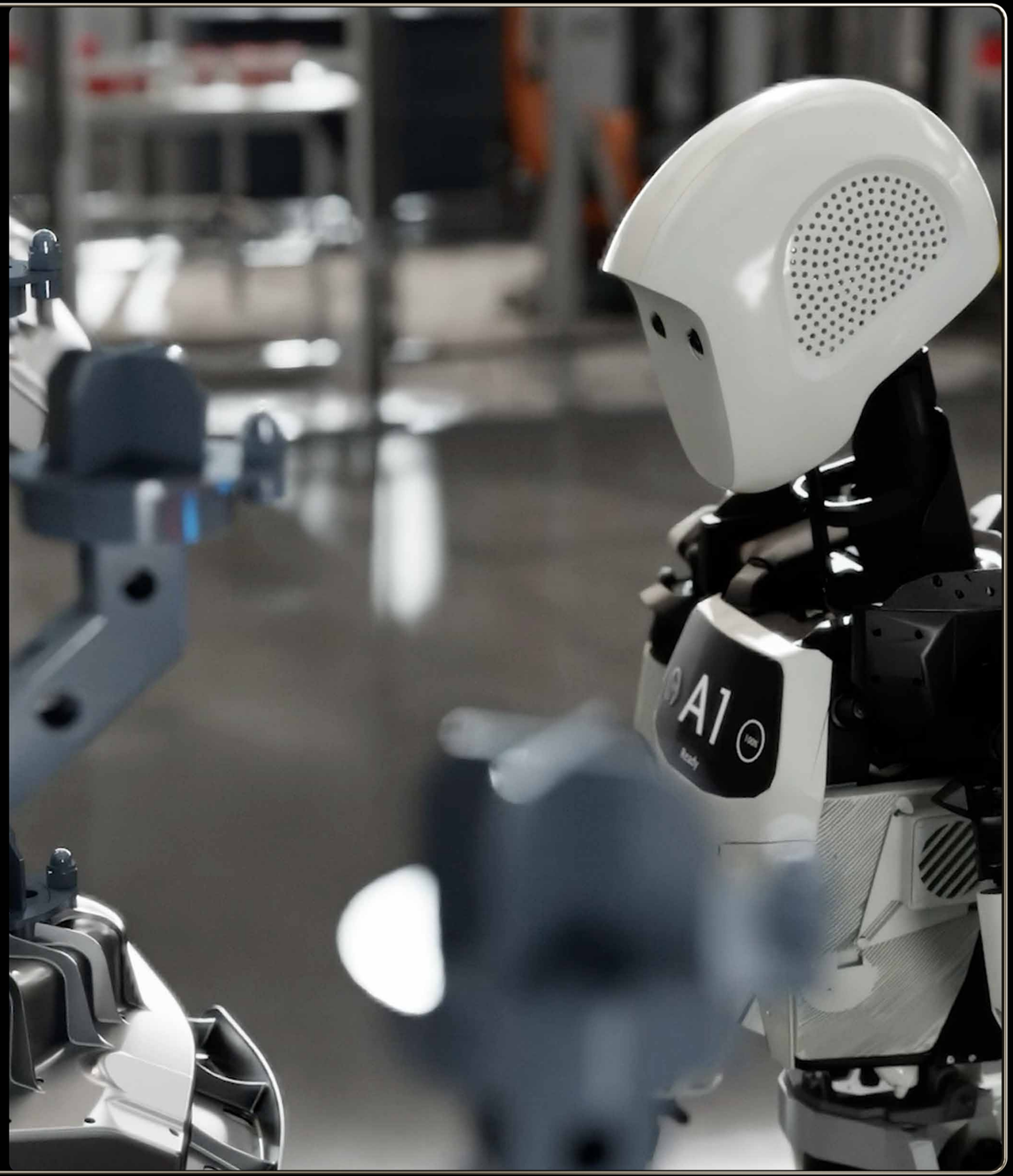
servicenow®


ServiceNow AI Agents
Triggers
Primary
Secondary
User Message
ServiceNow Agentic Workflow
Autonomous, adaptive, goal-driven intelligence
Context
Short & Long-term Memory
Knowledge
NVIDIA NeMo
Internal & External Retrieval
LLM (Planners)
Llama Nemotron Reasoning
GPT4+
Llama 3.3-70b
Agent Collaboration
And more...
Tools
NVIDIA Blueprint
Workflows
API's & Spokes
GenAI Skills
Action
LLM (Execution)
NVIDIA NIM
NowLLM

ROBOTS – THE NEXT MULTI-TRILLION DOLLAR INDUSTRY






A1

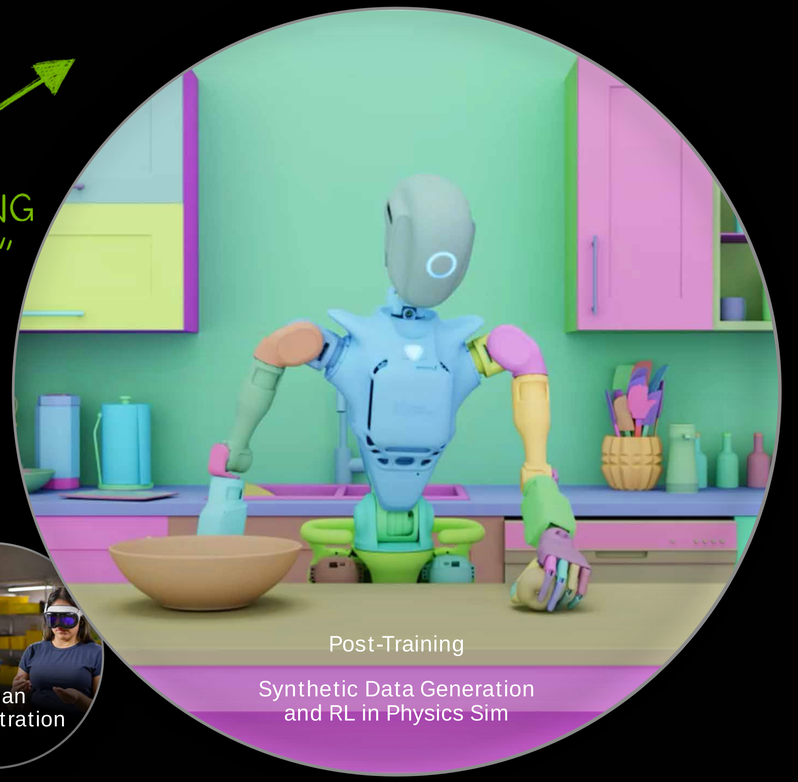
FROM ONE TO THREE SCALING LAWS

"INTELLIGENCE"

TEST-TIME SCALING
"LONG THINKING"

POST-TRAINING SCALING

PRE-TRAINING SCALING



PERCEPTION AI → GENERATIVE AI → AGENTIC AI → PHYSICAL AI →

NVIDIA Omniverse With Cosmos

Physical AI Digital Twin Operating System

NVIDIA Omniverse

Cosmos

Disney Research NVIDIA Google DeepMind

Announcing NVIDIA Isaac GR00T N1
Humanoid Foundation Model

Pick up the industrial object and place in yellow bin.

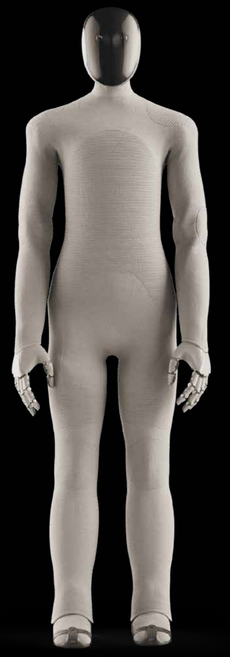
Sensor Tokens

Text Tokens

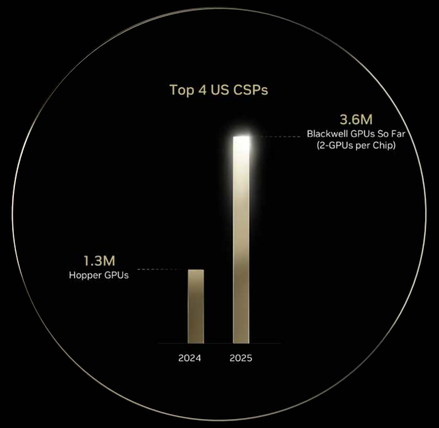
Vision Language Model

Diffusion Transformer Model

Action Tokens

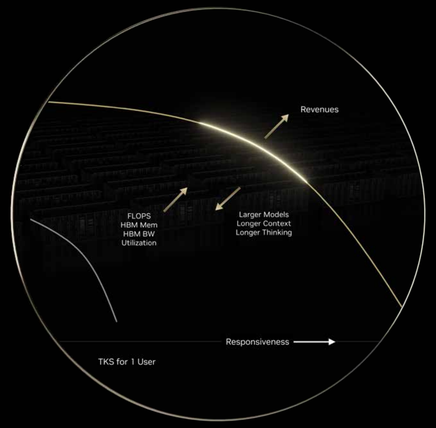


NVIDIA GTC 2025



BLACKWELL IN
FULL PRODUCTION

$1T Computing
Inflection Point

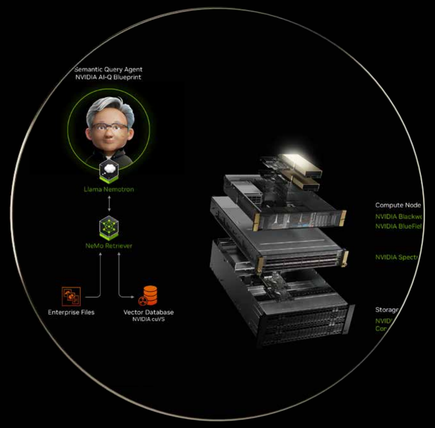


BLACKWELL NVL72 AND
DYNAMO 40X FOR INFERENCE

Reasoning 100X One-Shot
Blackwell 40X Hopper

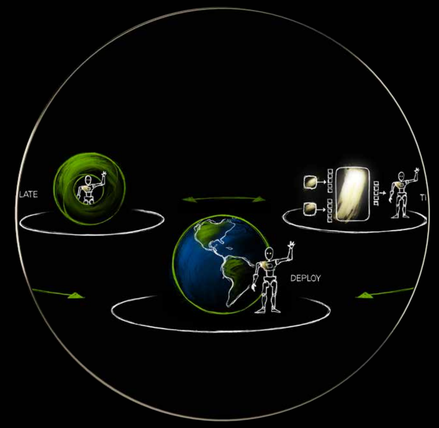
VERA RUBIN
NVIDIA PHOTONICS
AI INFRASTRUCTURE
FOR AI CLOUDS

Annual Rhythm for the
World to Build-Out
AI Infrastructure



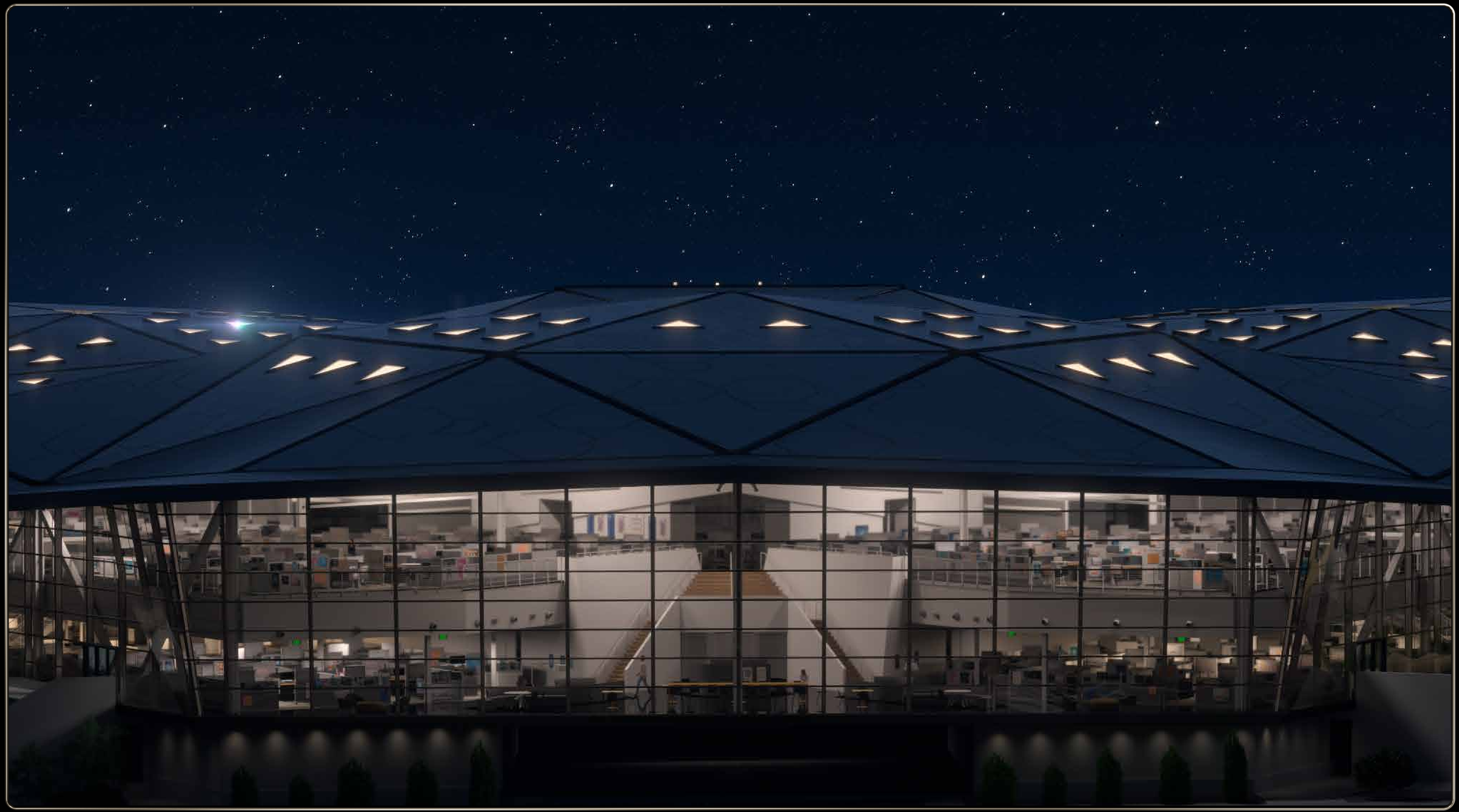
AI INFRASTRUCTURE
FOR $500B ENTERPRISE IT

New Compute, Networking,
Storage, Software



AI INFRASTRUCTURE
FOR ROBOTS

Physical AI For
$50T Industrial and Robotics




NVIDIA