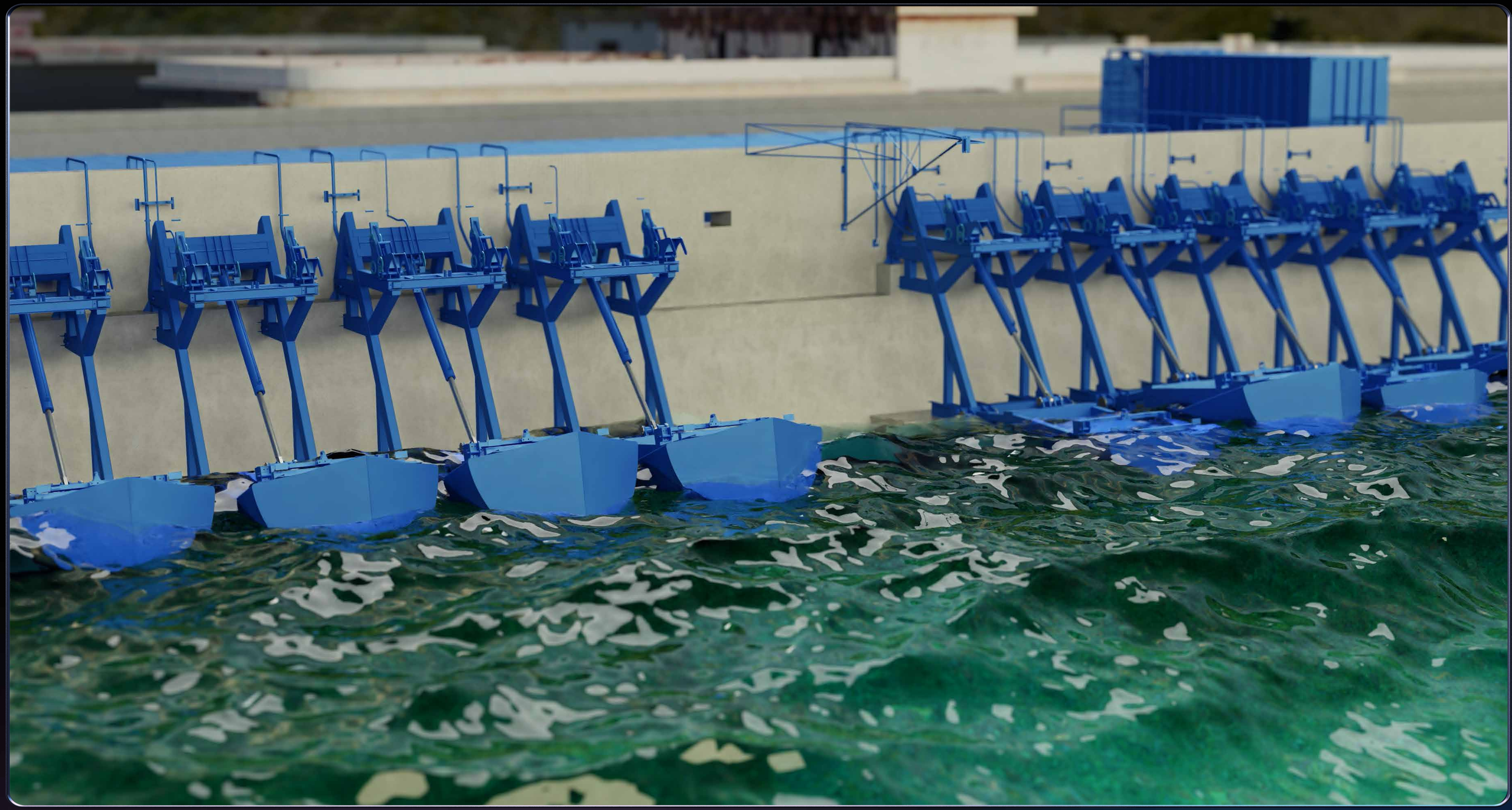

NVIDIA

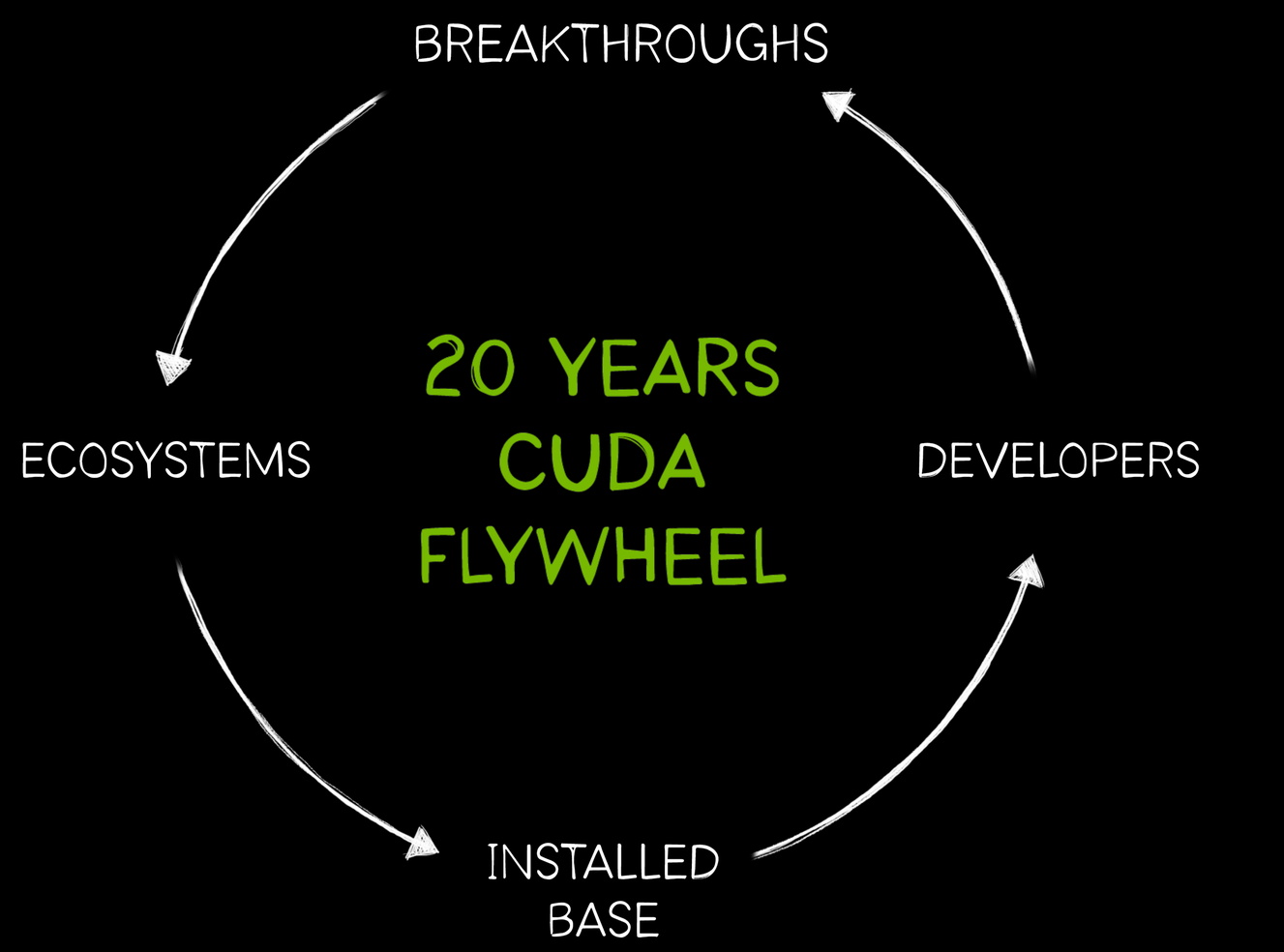


NVIDIA GTC

2026

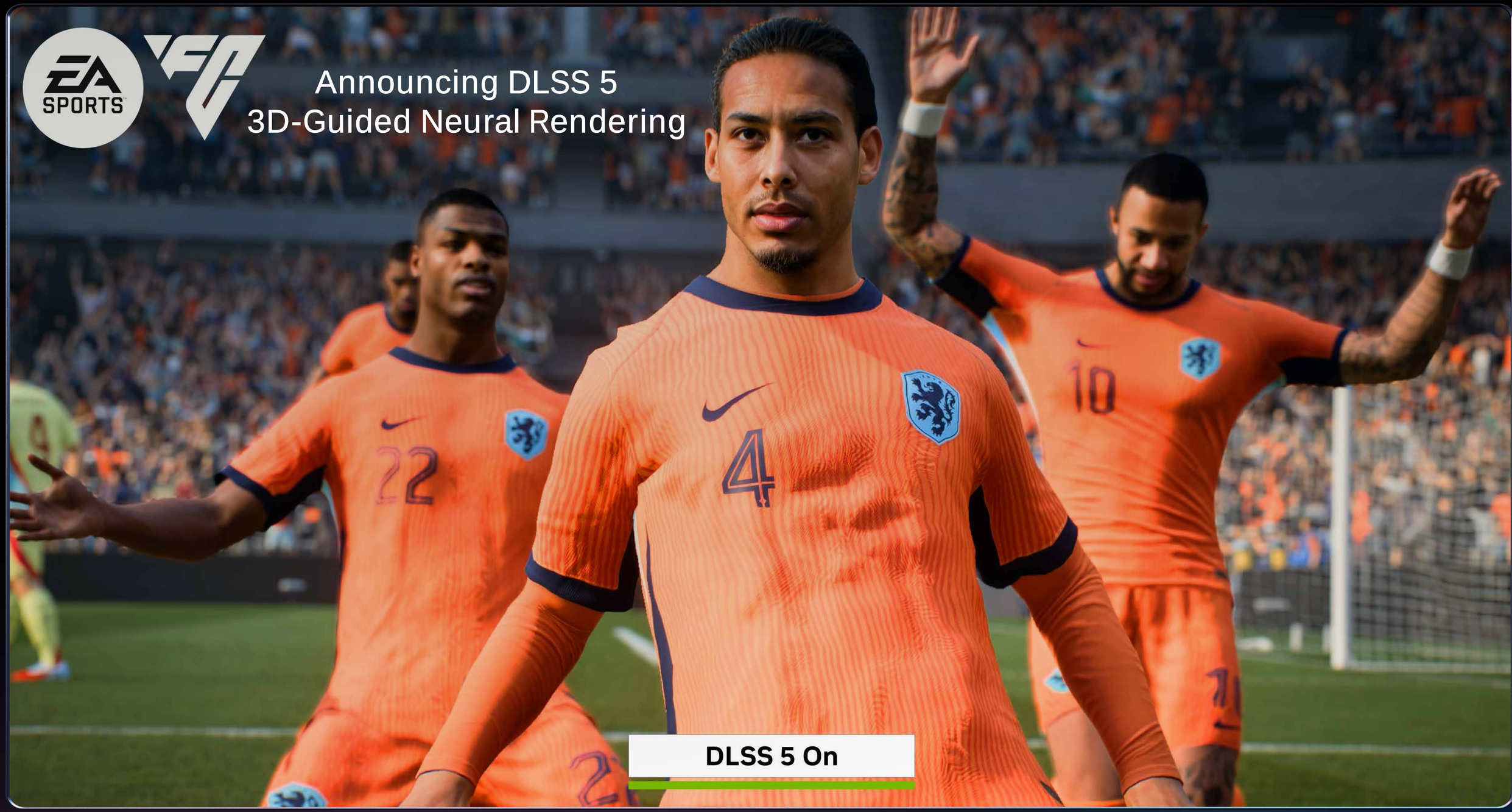

SARAH GUO
FOUNDER AND MANAGING PARTNER, CONVICTION
3RD PREGAME APPEARANCE
2X HOST
ALFRED LIN
PARTNER, SEQUOIA CAPITAL
2ND PREGAME APPEARANCE
2026 HOST DEBUT
GAVIN BAKER
FOUNDER, MANAGING PARTNER & CIO, ATREIDES MANAGEMENT
2ND PREGAME APPEARANCE
2026 HOST DEBUT
HARRISON CHASE
JACOMO CORBO
JOE CREED
MICHAEL DELL
ANIRUDH DEVGAN
MARK EDELSTONE
AIDAN GOMEZ
MICHAEL INTRATOR
AKI JAIN
TIFFANY JANZEN
1ST PREGAME APPEARANCE
2026 HOST DEBUT
ARTHUR MENSCH
DANIEL NADLER
DINESH NIRMAL
DEEPAK PATHAK
LIN QIAO
SAM RODRIQUES
ROBIN ROMBACH
ARAVIND SRINIVAS
PETER STEINBERGER
RAQUEL URTSAUN
VINCENT WEISSER
SAN JOSE | 2026
NVIDIA GTC
LIVE


ABB
abbvie
accenture
Adobe
AETHERFLUX
AHEAD
Ai2
AiVRES
Akamai
AT&T
Alibaba Cloud
ALTIMETER
aws
AMP COALITION
ANTHROPIC
ASRock Rack
ASUS
ATREIDES MANAGEMENT
AXIOM SPACE
baseten
biofy
Black Forest Labs
BLUE ORIGIN
BNY
BOSCH
bp
BYD
ByteDance
cadence
Caltech
Capital One
Carnegie Mellon University
CATERPILLAR
cisco
CLOUDIAN
cohere
Cognition
COMCAST
COMPAL
CONVICTION
CoolIT systems
CoreWeave
coupang
CROWDSTRIKE
Crusoe
CURSOR
DASSAULT SYSTEMES
databricks
ddn
deepinfra
DELL Technologies
Deloitte.
DELTA
Disney Research
EATON
Edison
EQUINIX
ETH zürich
Everpure
EY
Fireworks AI
firmus
Ford
FOXCONN
GE Healthcare
GE VERNOVA
GEELY
GEICO
gm
Georgia Institute of Technology
GIGABYTE
GMI
Google Cloud
HITACHI
Honeywell
hp
HPE
HYUNDAI
hyve solutions
IBM
ingrasys
instacart
Inventec
IQVIA
IREN
Jacobs
JR
JOHNS HOPKINS UNIVERSITY
Johnson&Johnson
KAIST
KLA
L'ORÉAL
Lambda
LangChain
Lenovo
Lilly
LOCKHEED MARTIN
Los Alamos NATIONAL LABORATORY
MEDBOT
Medtronic
Meta
micron
Microsoft
Mistral AI
MITAC Computing
Morgan Stanley
msi
Nasdaq
國立清華大學 NATIONAL TSING HUA UNIVERSITY
NEBIUS
NetApp
NISSAN MOTOR CORPORATION
NOKIA
NSCALE
NUTANIX
OpenAI
OpenClaw
OpenEvidence
ORACLE CLOUD Infrastructure
Palantir
paloalto
PayPal
PEGATRON
PEPSICO
perplexity
PNR
PHYSICS X
Pinterest
planet.
PNY
PRIME Intellect
PRINCETON UNIVERSITY
ptc
QCT
RADIANT
RAFAY
Red Hat
REDWIRE
Reflection
runway
salesforce
SAMSUNG
SJSU SAN JOSÉ STATE UNIVERSITY
SAP
Schneider Electric
서울대학교 SEOUL NATIONAL UNIVERSITY
SEQUOIA
servicenow
Shopee
shopify
SUPERMICRO
SIEMENS
SK hynix
SKILD AI
SONY
Stanford University
Starcloud
stryker
switch
synopsys
T-Mobile
TCS TATA CONSULTANCY SERVICES
TESLA
TACC
大阪大学 OSAKA UNIVERSITY
ThermoFisher SCIENTIFIC
THINKING MACHINES
together.ai
Tripadvisor
TSMC
Uber
UC Berkeley
UCLA
UNIVERSITY OF CAMBRIDGE
UNIVERSITY of WASHINGTON
VAST
VERTIV.
VOLKSWAGEN
VULTR
waabi
Walmart
WAYMO
wiwynn x wistron
WEKA
WELLS FARGO
World Labs
World Wide Technology
xAI
XIAOMI
YOTTA
Yum!
GTC
The Next Generation of AI Begins


BREAKTHROUGHS
ECOSYSTEMS
20 YEARS
CUDA
FLYWHEEL
DEVELOPERS
INSTALLED
BASE

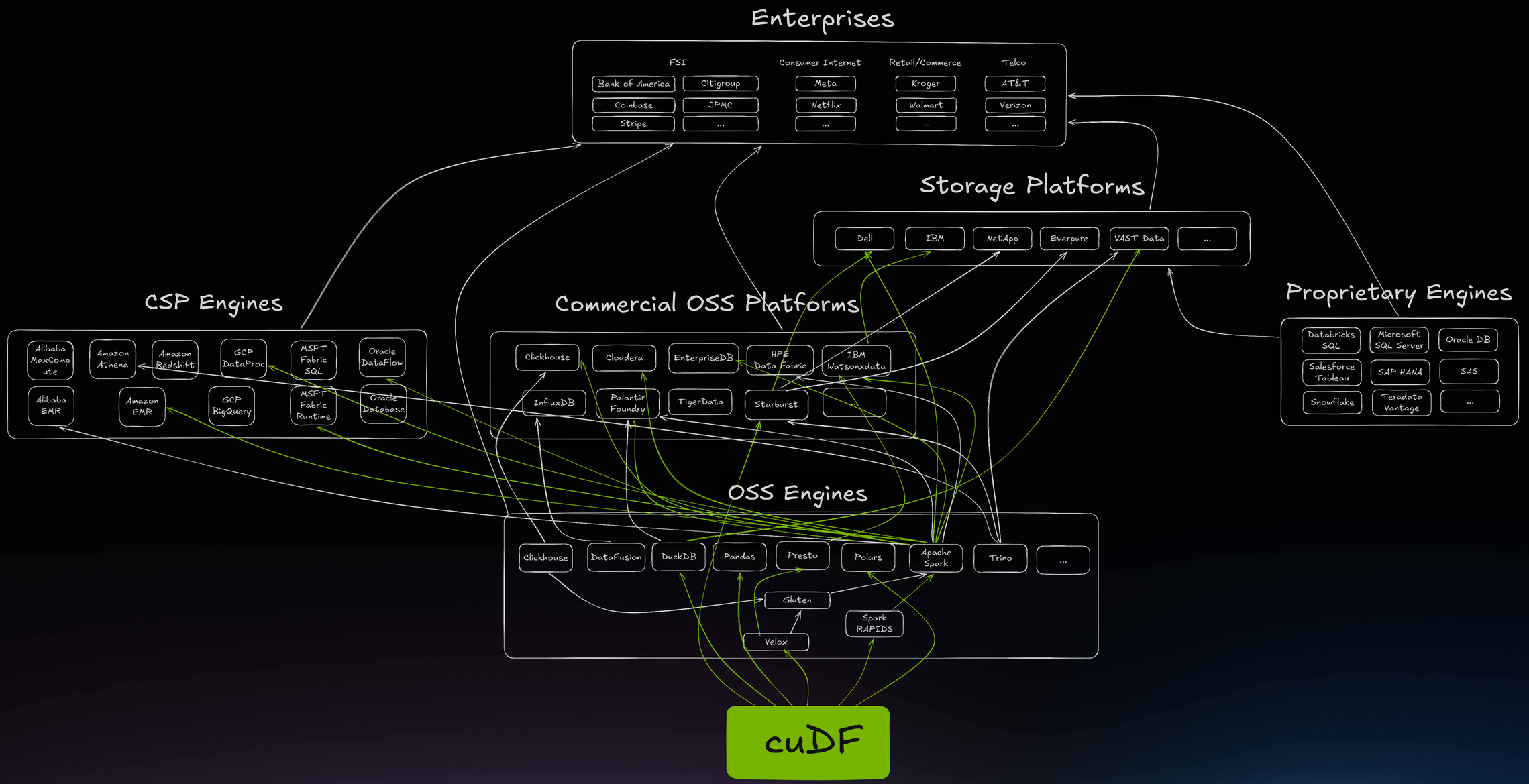
GeForce GTX 1080


PORTKEY GAMES
HOGWARTS
LEGACY
DLSS 5 On


Announcing DLSS 5
3D-Guided Neural Rendering
DLSS 5 On

Structured Data is the Ground Truth of AI

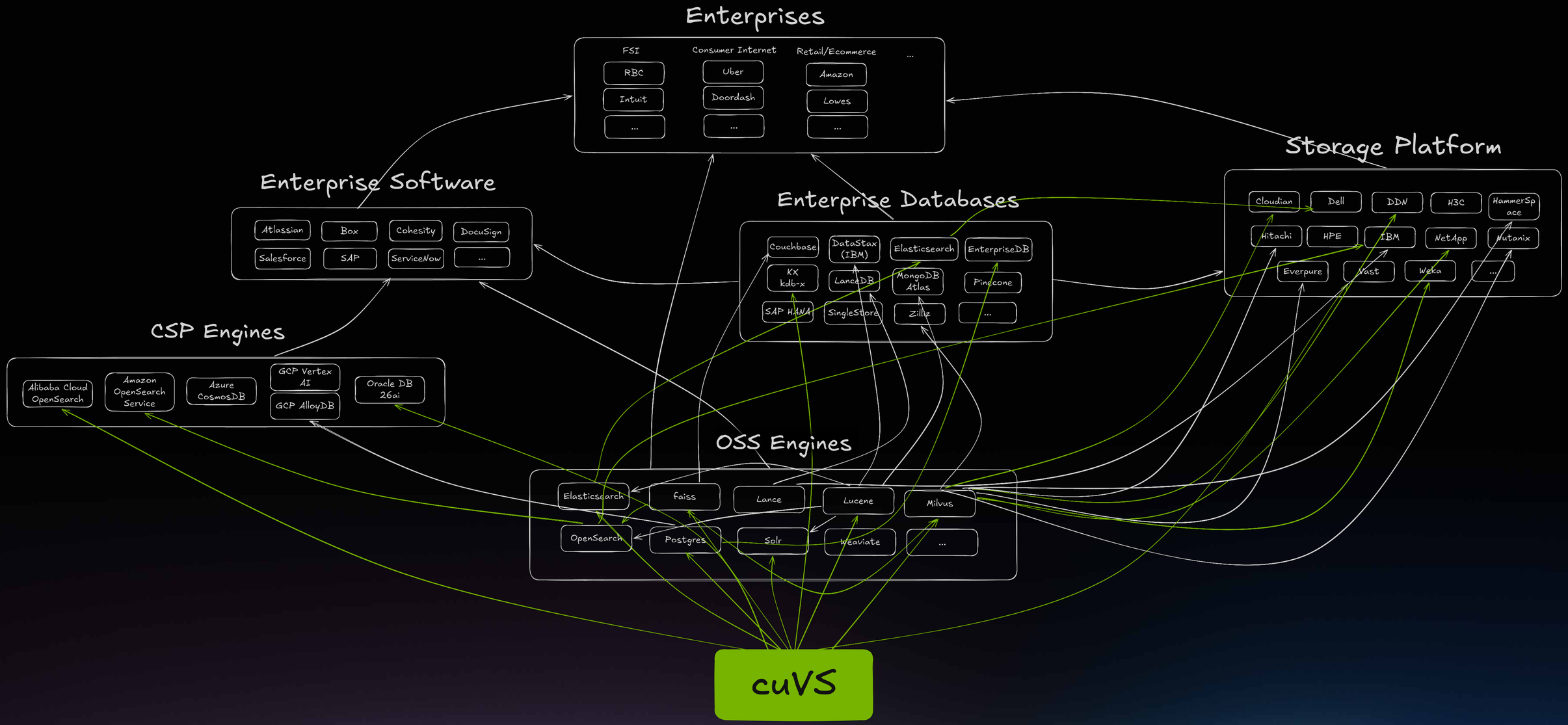
$120B Structured Data Ecosystem – The Ground Truth of Enterprises



Source: Gartner DBMS, 2025

Unstructured Data is the Context of AI

100’s of Zettabytes Per Year of Unstructured Data – Growing Exponentially

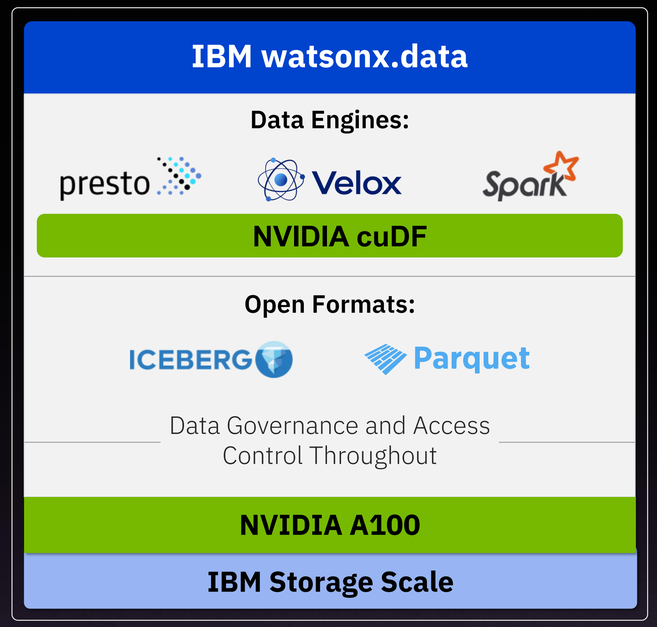


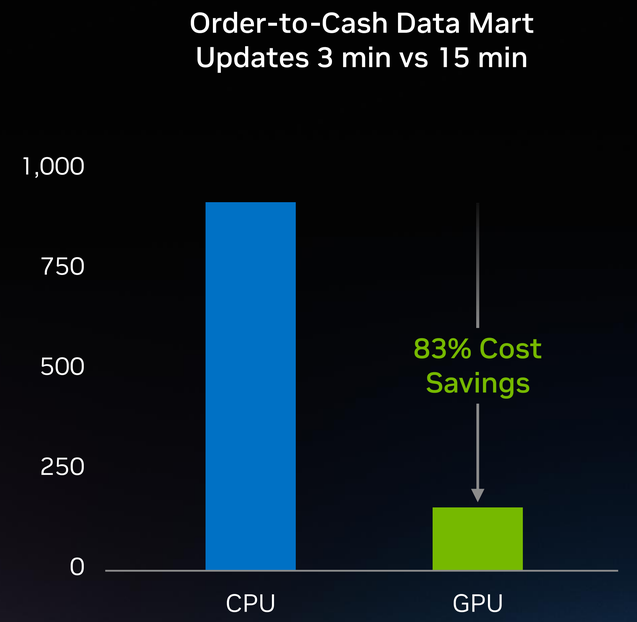
Source: IDC Global DataSphere, 2025

NVIDIA ❤ IBM ❤ Nestlé

*"Working with IBM and NVIDIA, early results show we can **refresh global operations data in minutes** at reduced cost—making a capability that can be turned into tangible business impact in areas such as manufacturing or warehousing."*

— Chris Wright, Chief Information and Digital Officer of Nestlé



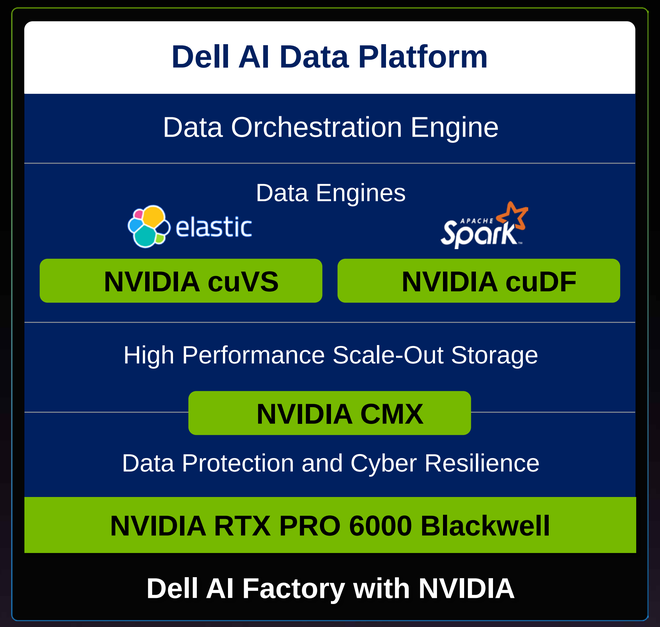


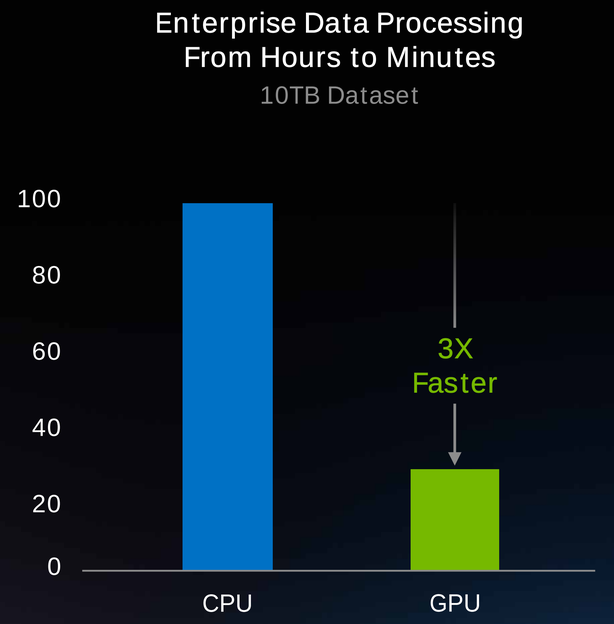

Nestlé

NVIDIA ❤ DELL Technologies ❤ NTT DATA

"Dell AI Data Platform with NVIDIA helps us build scalable and repeatable data pipelines that drive automation—processing massive data volumes in minutes instead of hours and delivering transformative value for our clients."

— Abhijit Dubey, CEO of NTT DATA

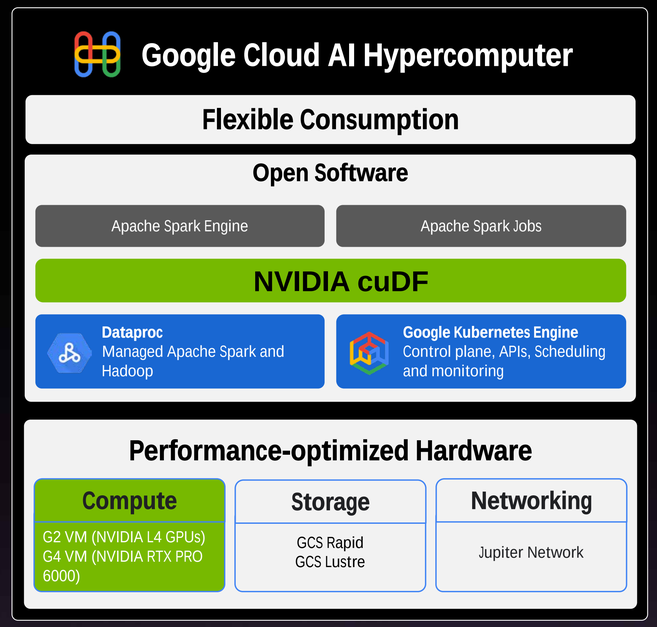


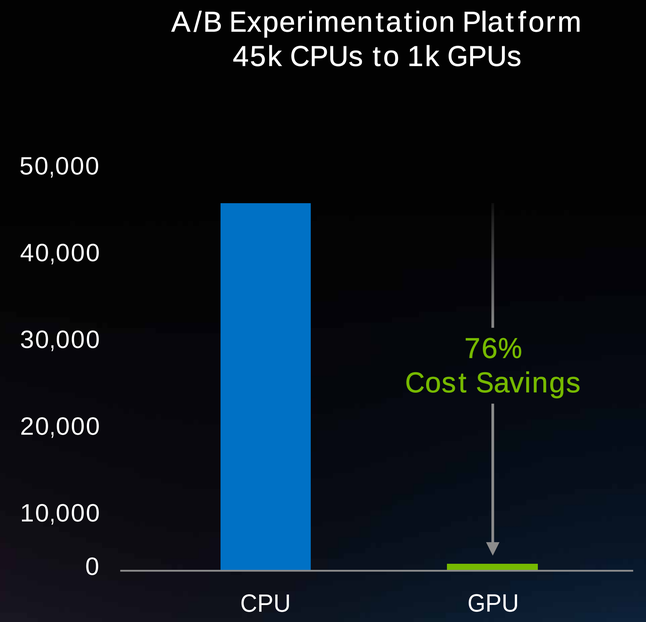


NVIDIA ❤ Google Cloud ❤

“Our collaboration with NVIDIA and Google Cloud helps us innovate faster for more than a billion Snapchatters worldwide. By lowering costs and scaling experiments across petabytes of data, we’re delivering AI-powered experiences more quickly and efficiently.”

— Saral Jain, CIO of Snap





NVIDIA ❤ Google Cloud

baseten · CROWDSTRIKE · gm · otto group · PUMA · Reve · salesforce · Schrödinger · SNAPCHAT

AI Optimized Infrastructure

- Liquid Cooling
- Storage
- Networking
- Google Distributed Cloud

AI Platforms

- Vertex AI
- Google Kubernetes Engine
- Cloud Run
- Google Compute Engine

Data Science and Analytics

- BigQuery
- Dataproc
- Dataflow
- Bigtable

Open Software, Frameworks

- JAX
- XLA
- vLLM

System Builders and OEMs

NVIDIA AI Platform

cuDNN · Magnum IO · cuOpt · cuVS · cuDF · Omniverse · TRT-LLM · Dynamo · NeMo · PhysicsNeMo · Cosmos · Nemotron

NVIDIA Hopper

NVIDIA RTX PRO Blackwell Server Edition

NVIDIA Blackwell

Confidential Computing

NVIDIA Rubin

NVIDIA ConnectX and BlueField Networking

NVIDIA Quantum IB, Spectrum-X Networking

NVIDIA ❤ aws

CROWDSTRIKE · databricks · HUMAIN · OpenAI · perplexity · Palantir · qodo · salesforce · snowflake

Amazon Braket · Amazon EKS · AWS Nitro Enclaves · AWS Batch · Amazon EMR · Amazon Bedrock · Amazon SageMaker AI · Amazon EC2 · Amazon EFA · AWS ParallelCluster

System Builders and OEMs

NVIDIA AI Platform

cuDNN · Magnum IO · cuOpt · cuVS · cuDF · Omniverse · TRT-LLM · Dynamo · NeMo · PhysicsNeMo · Cosmos · Nemotron

NVIDIA Hopper · NVIDIA RTX PRO Blackwell Server Edition · NVIDIA Blackwell · Confidential Computing · NVIDIA Rubin · NVIDIA ConnectX and BlueField Networking · NVIDIA Quantum IB, Spectrum-X Networking

NVIDIA ❤ Microsoft Azure

ANTHROP\C · BlackRock · BMW · KRONES · Rockwell Automation · SYNOPSYS

Microsoft Foundry · Windows 365 · Microsoft Fabric · Azure Container Apps · Azure Kubernetes Service · Foundry Local · Microsoft 365 Copilot · Microsoft Bing

System Builders and OEMs

NVIDIA AI Platform

cuDNN · Magnum IO · cuOpt · cuVS · cuDF · Omniverse · TRT-LLM · Dynamo · NeMo · PhysicsNeMo · Cosmos · Nemotron

NVIDIA Hopper · NVIDIA RTX PRO Blackwell Server Edition · NVIDIA Blackwell · Confidential Computing · NVIDIA Rubin · NVIDIA ConnectX and BlueField Networking · NVIDIA Quantum IB, Spectrum-X Networking

NVIDIA ❤ ORACLE

cohere | Fireworks AI | NEOSPACE | SIEMENS | sky | Uber | zoom

ORACLE CLOUD Infrastructure

- OCI Superclusters
- Oracle AI Database
- Oracle Autonomous AI Lakehouse
- Oracle AI Data Platform
- OCI AI Accelerator Packs
- OCI Generative AI
- OCI Data Science
- OCI OKE

System Builders and OEMs

NVIDIA AI Platform

cuDNN | Magnum IO | cuOpt | cuVS | cuDF | Omniverse | TRT-LLM | Dynamo | NeMo | PhysicsNeMo | Cosmos | Nemotron

- NVIDIA Hopper
- NVIDIA RTX PRO Blackwell Server Edition
- NVIDIA Blackwell
- Confidential Computing
- NVIDIA Rubin
- NVIDIA ConnectX and BlueField Networking
- NVIDIA Quantum IB, Spectrum-X Networking

NVIDIA ❤ CoreWeave

Canva · cohere · CURSOR · mercado libre · Mistral AI · Morgan Stanley · OpenAI · runway

W&B Models · W&B Training · W&B Inference · W&B Weave · CoreWeave Inference · CoreWeave Slurm on Kubernetes · CoreWeave Kubernetes Service · CoreWeave Compute · CoreWeave Storage · CoreWeave Mission Control

System Builders and OEMs

NVIDIA AI Platform

cuDNN · Magnum IO · cuOpt · cuVS · cuDF · Omniverse · TRT-LLM · Dynamo · NeMo · PhysicsNeMo · Cosmos · Nemotron


NVIDIA Hopper


NVIDIA RTX PRO Blackwell Server Edition


NVIDIA Blackwell


Confidential Computing

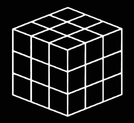

NVIDIA Rubin


NVIDIA ConnectX and BlueField Networking


NVIDIA Quantum IB, Spectrum-X Networking

NVIDIA ❤ Palantir ❤ DELL Technologies

accenture · CenterPoint Energy · Hertz · LOWE'S · TETON RIDGE


Rubix


Ontology


Pipeline Builder


Fusion


Agent Studio


Model Studio


AIP Logic


Model Catalog


Workshop


AI FDE

Dell

NVIDIA AI Platform

cuDNN · Magnum IO · cuOpt · cuVS · cuDF · Omniverse · TRT-LLM · Dynamo · NeMo · PhysicsNeMo · Cosmos · Nemotron


NVIDIA Hopper


NVIDIA RTX PRO Blackwell Server Edition


NVIDIA Blackwell


Confidential Computing


NVIDIA Rubin

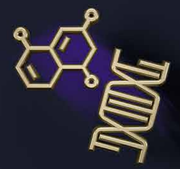

NVIDIA ConnectX and BlueField Networking


NVIDIA Quantum IB, Spectrum-X Networking

NVIDIA GTC

AUTOMOTIVE

FINANCIAL SERVICES



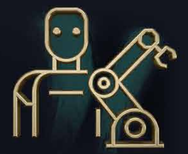
HEALTHCARE AND LIFE SCIENCES

INDUSTRIAL

MEDIA AND ENTERTAINMENT

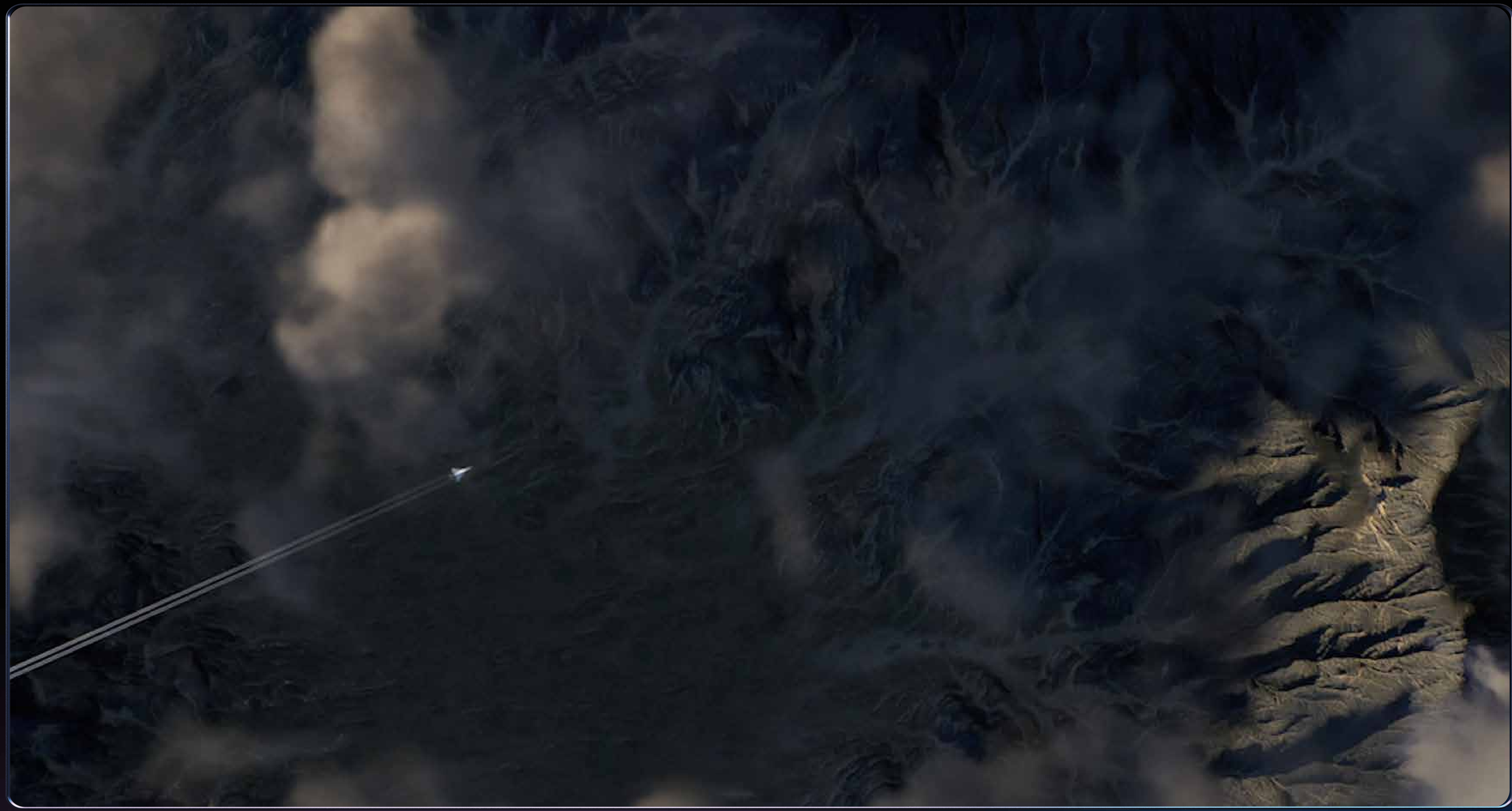
QUANTUM

RETAIL AND CPG



ROBOTICS

TELCO



NVIDIA ❤ AI Natives

AI for Auto
Aurora · momenta · nuro · pony.ai · waabi · WAYMO · WAYVE

AI for Customer Support
Decagon · IIElevenLabs · Fin · kore.ai · SIERRA · synthesia

AI for Engineering
BeyondMath · Emmi AI · NEURAL CONCEPT · luminary · PASTEUR SI · periodic labs · PHYSICS X · RICURSIVE INTELLIGENCE

AI for Healthcare
Arc Institute · biohub · Boltz · Chai Discovery · Edison · Isomorphic Labs · L I L A · Institute for Protein Design

AI for Robotics
AGIBOT · AGILE ROBOTS · FieldAI · FIGURE · Generalist · Physical Intelligence (π) · Robotics and AI Institute · SKILD AI · 自变量

AI for Search
brave · glean · perplexity · you.com

AI for Software Development
CodeRabbit · Cognition · CURSOR · FACTORY · greptile · Lovable · opencode · OpenHands · qodo · tabnine

DL Frameworks
PyTorch · JAX

Inference Frameworks
SGL · vLLM

Agent Frameworks/Protocols
A2A · AutoGen · crewai · LangChain · Model Context Protocol · OpenClaw

Frontier Model Builders
Ai2 · Advanced Machine Intelligence · ANTHROPIC · Black Forest Labs · cohere · Decart · deepseek · Gemma · humans& · inception · ineffable intelligence · KIMI · Magic · Midjourney · Mistral AI · OpenAI · poolside · Qwen · Reflection · runway · THINKING MACHINES · xAI

Model to Production
Anyscale · baseten · BigPanda · Braintrust · DataRobot · Distyl · Fireworks AI · H2O.ai · LogicMonitor · Modal · NUTANIX · OpenRouter · Red Hat · Replicate · scale · snowflake · together.ai

NVIDIA AI Platform

cuDNN · Magnum IO · cuOpt · cuVS · cuDF · Omniverse · TRT-LLM · Dynamo · NeMo · PhysicsNeMo · Cosmos · Nemotron


NVIDIA Hopper


NVIDIA RTX PRO Blackwell Server Edition


NVIDIA Blackwell


Confidential Computing

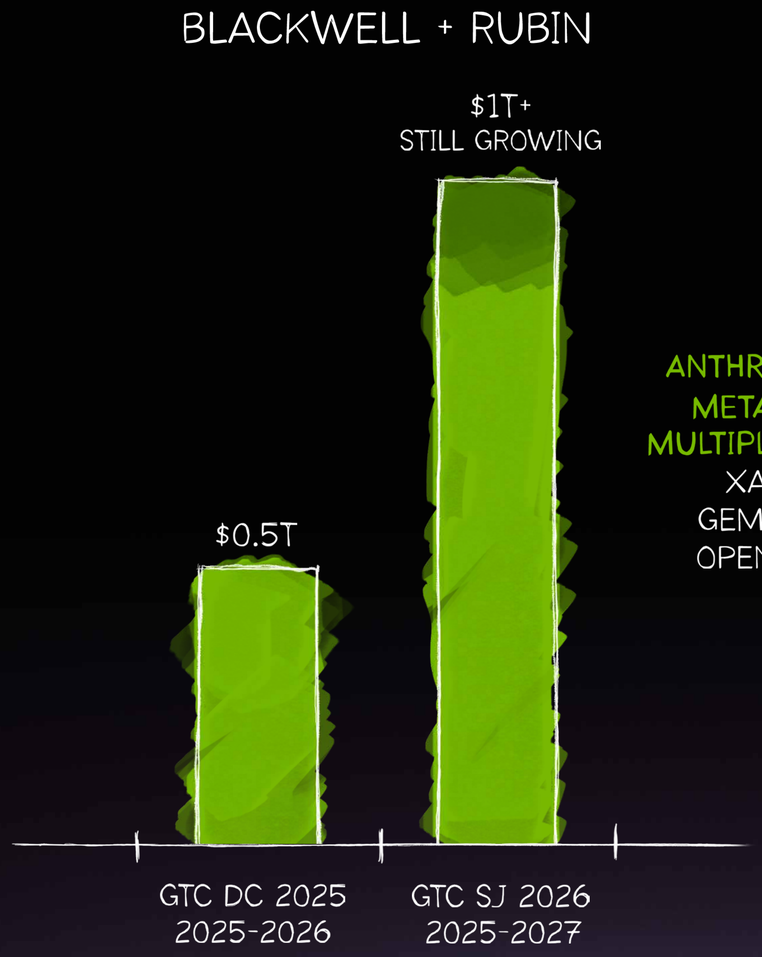

NVIDIA Rubin


NVIDIA ConnectX and BlueField Networking


NVIDIA Quantum IB, Spectrum-X Networking

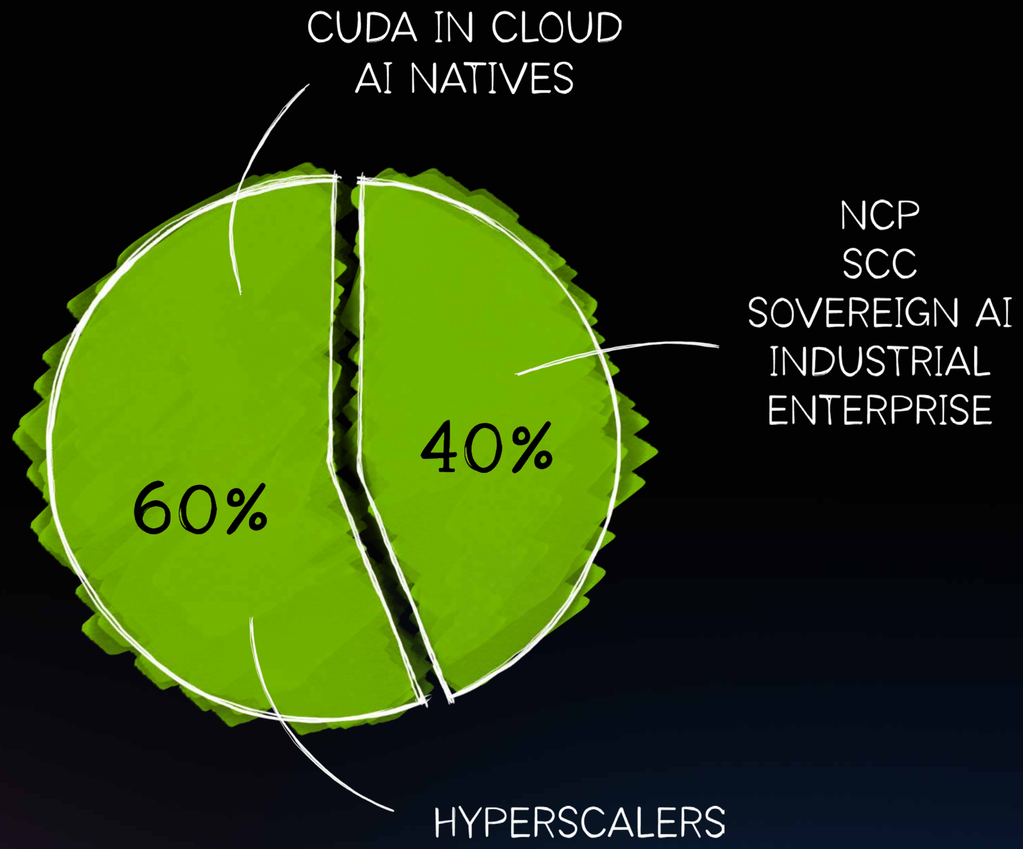
Inference Inflection Arrives
10,000X ChatGPT Compute

Inference Inflection Drives Strong Growth
NVIDIA Full-Stack Expanding AI to All Regions and Industries



ANTHROPIC
METASL
MULTIPLE OSS
XAI
GEMINI
OPENAI

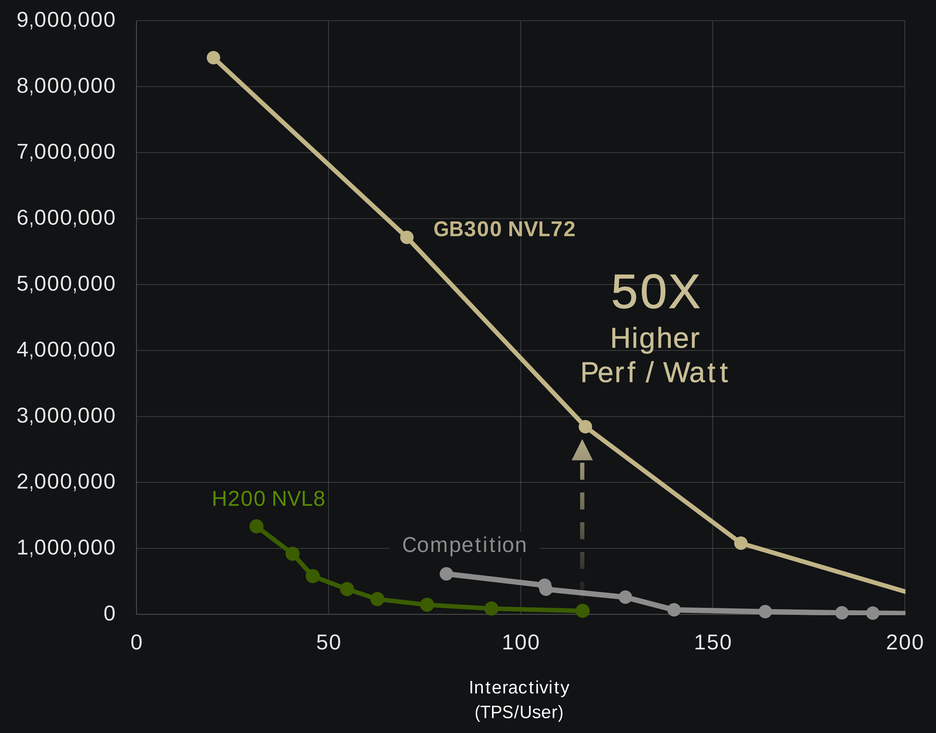
NEW SINCE
GTC DC 2025



NVIDIA Extreme Co-Design Revolutionized Token Cost

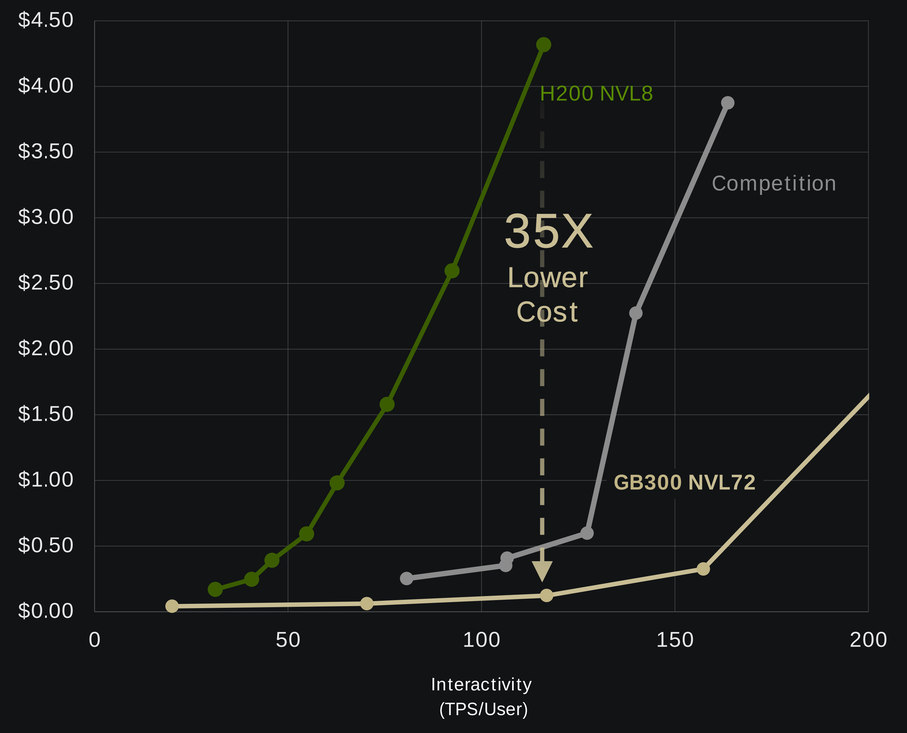
“GB NVL72 Inference King”

Tokens per Watt Drives Factory Revenue



InferenceX
by semianalysis

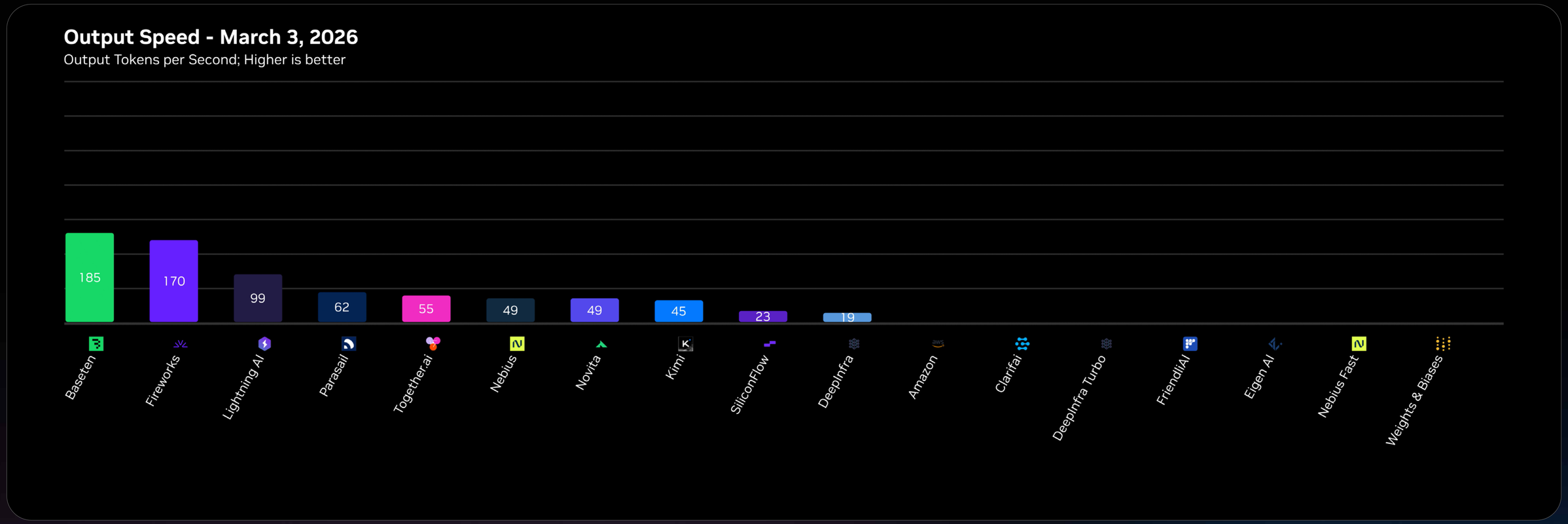
Performance Drives Token Cost



DeepSeek R1 0528 · FP4 · 1K/1K · Source: SemiAnalysis InferenceX

NVIDIA is the Global Standard for AI Inference at Scale

All Leading Inference End-Points Run on NVIDIA



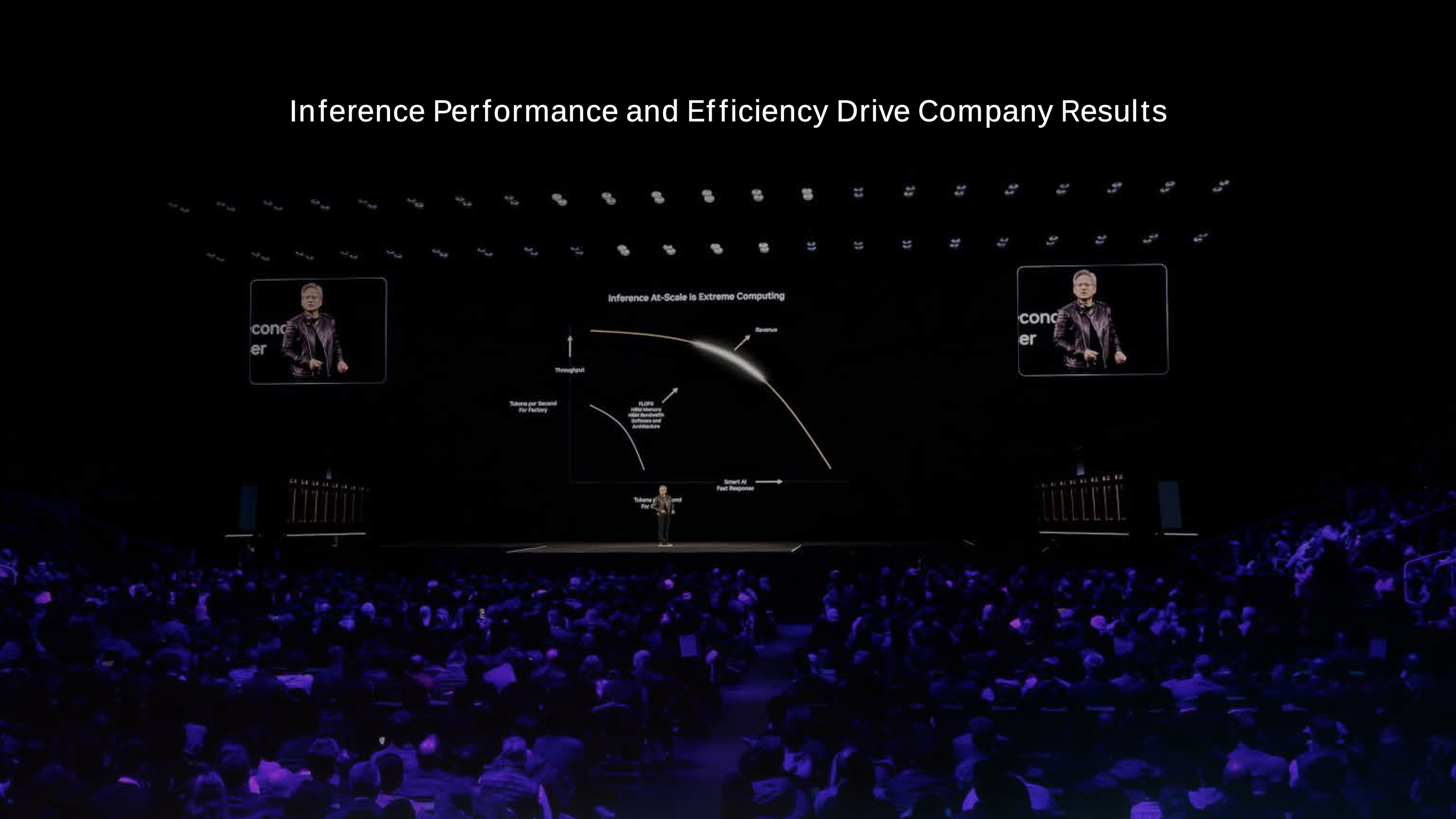
Kimi K2.5 Reasoning

AI Factories are the Industrial Infrastructure of the AI Era
Inference is the Workload
Tokens are the New Commodity
Compute is Revenue

Vera Rubin POD

AI FLOPS	60,000 PF
Memory Bandwidth	92 TB/s
All-to-All Bandwidth	260 TB/s
Scale-Out Radix	131,072

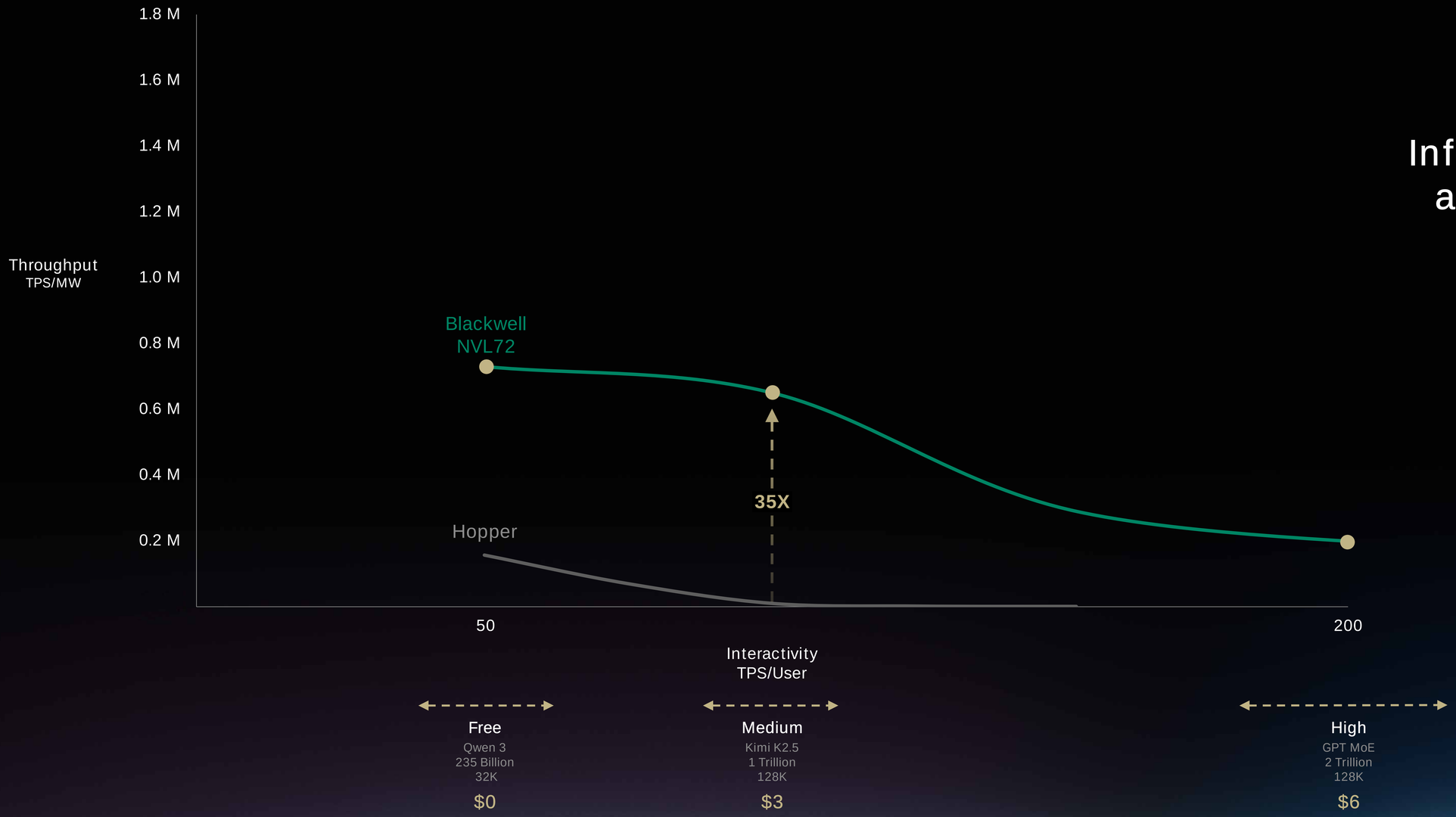
Inference Performance and Efficiency Drive Company Results



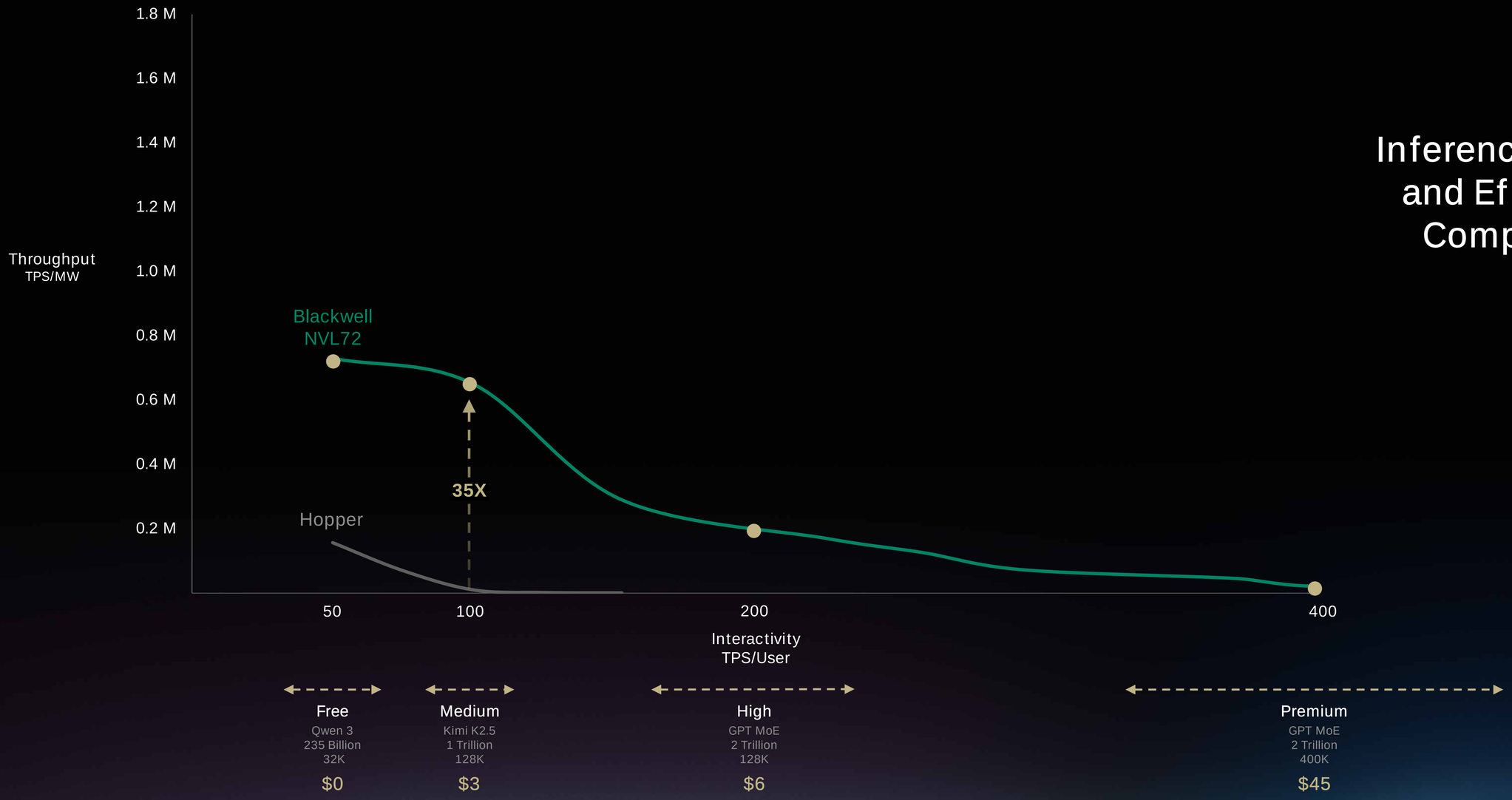
Inference Performance and Efficiency Drive Company Results

Inference Performance and Efficiency Drive Company Results

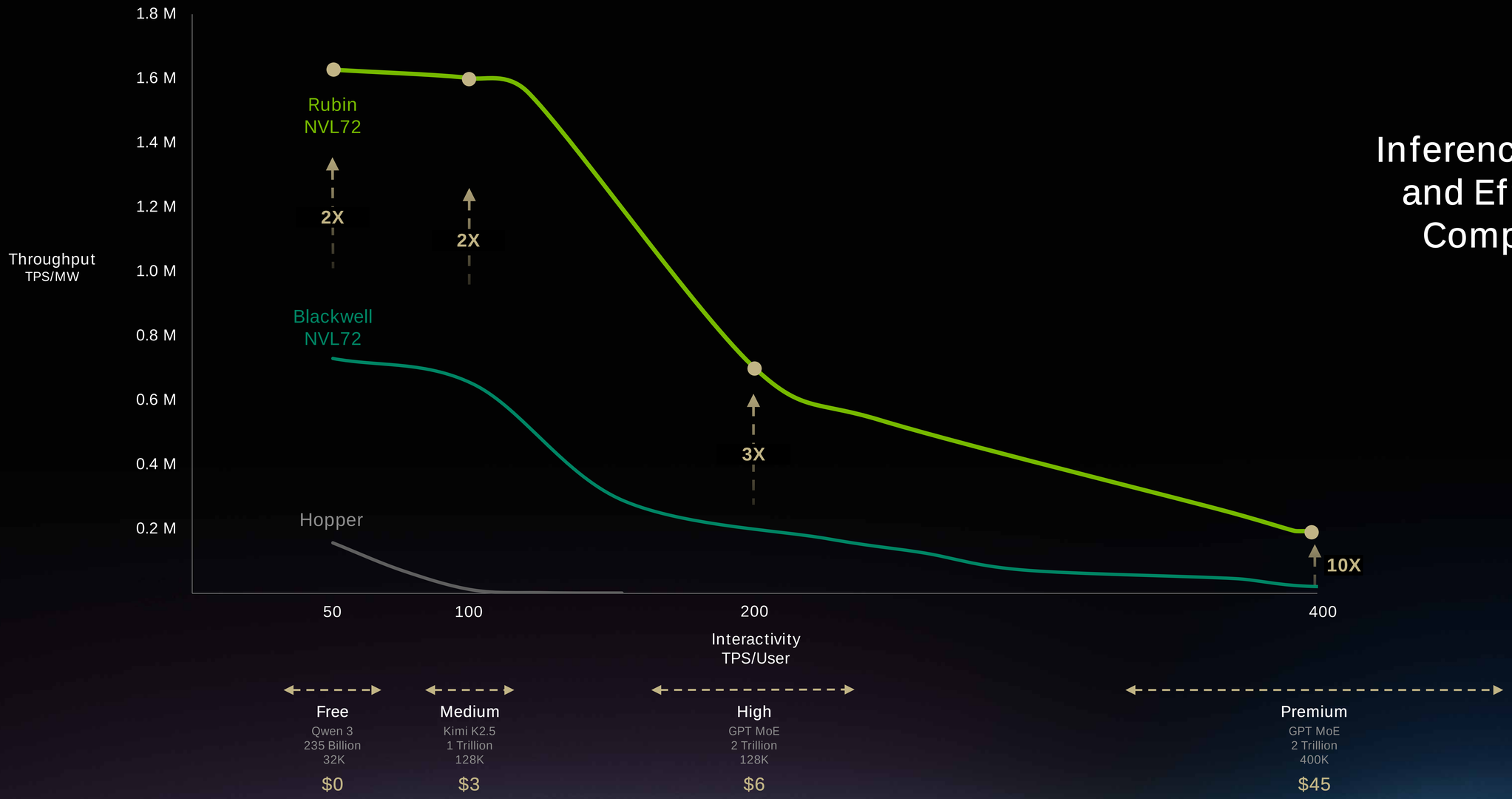
Inference Performance and Efficiency Drive Company Results



Inference Performance and Efficiency Drive Company Results

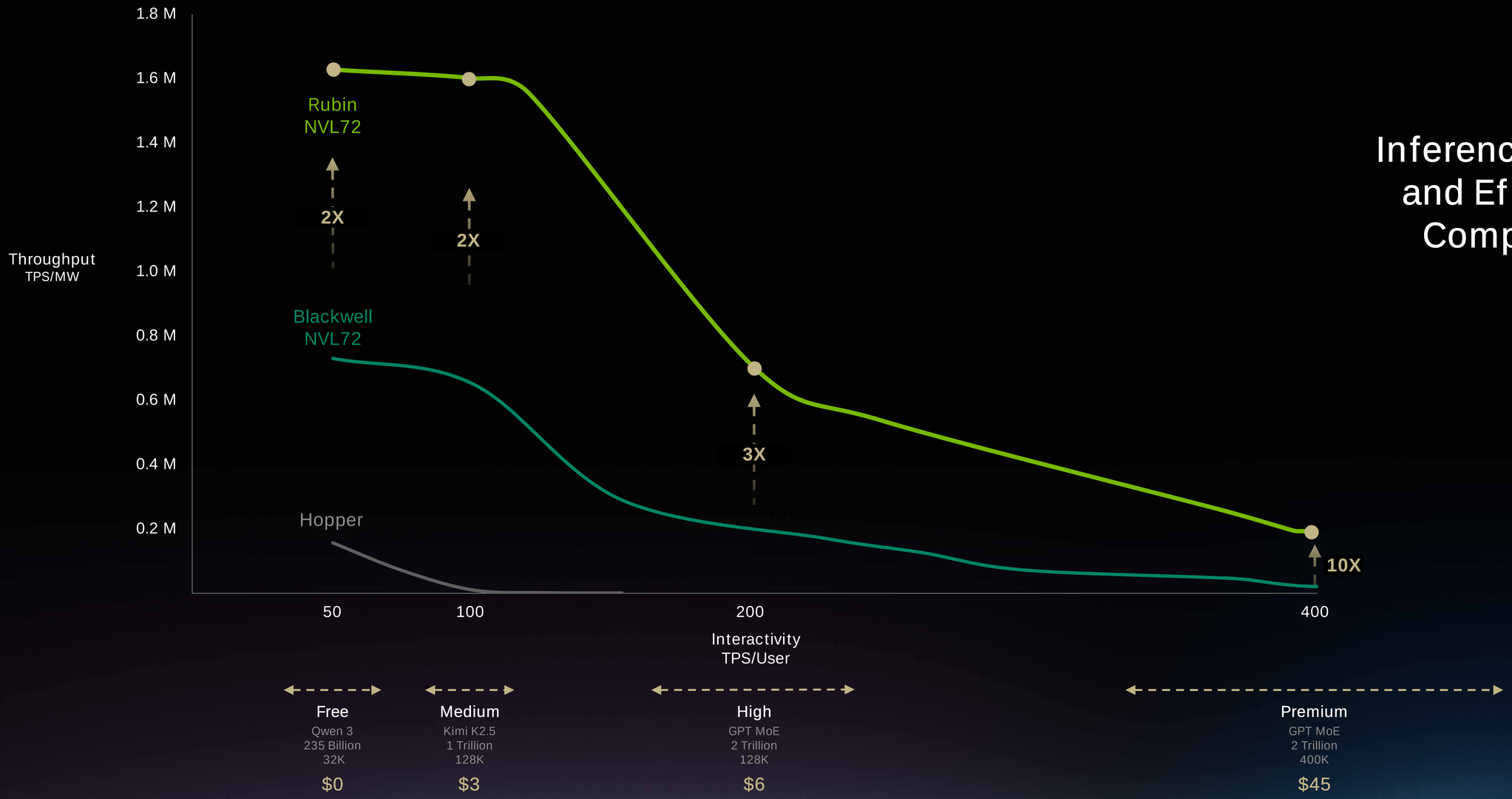


Inference Performance and Efficiency Drive Company Results

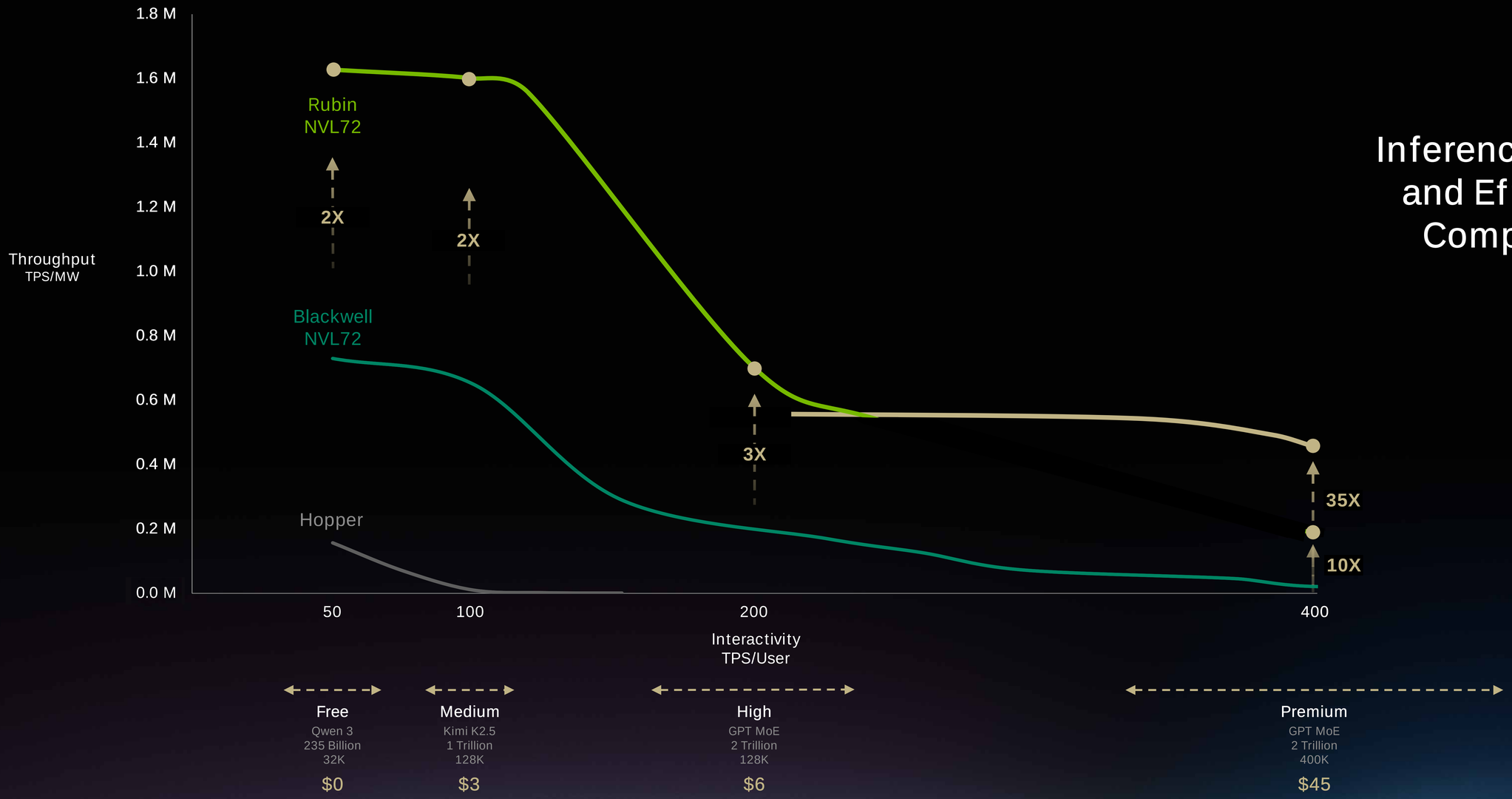


NVIDIA Vera Rubin Unlocks $150B Revenue Opportunity

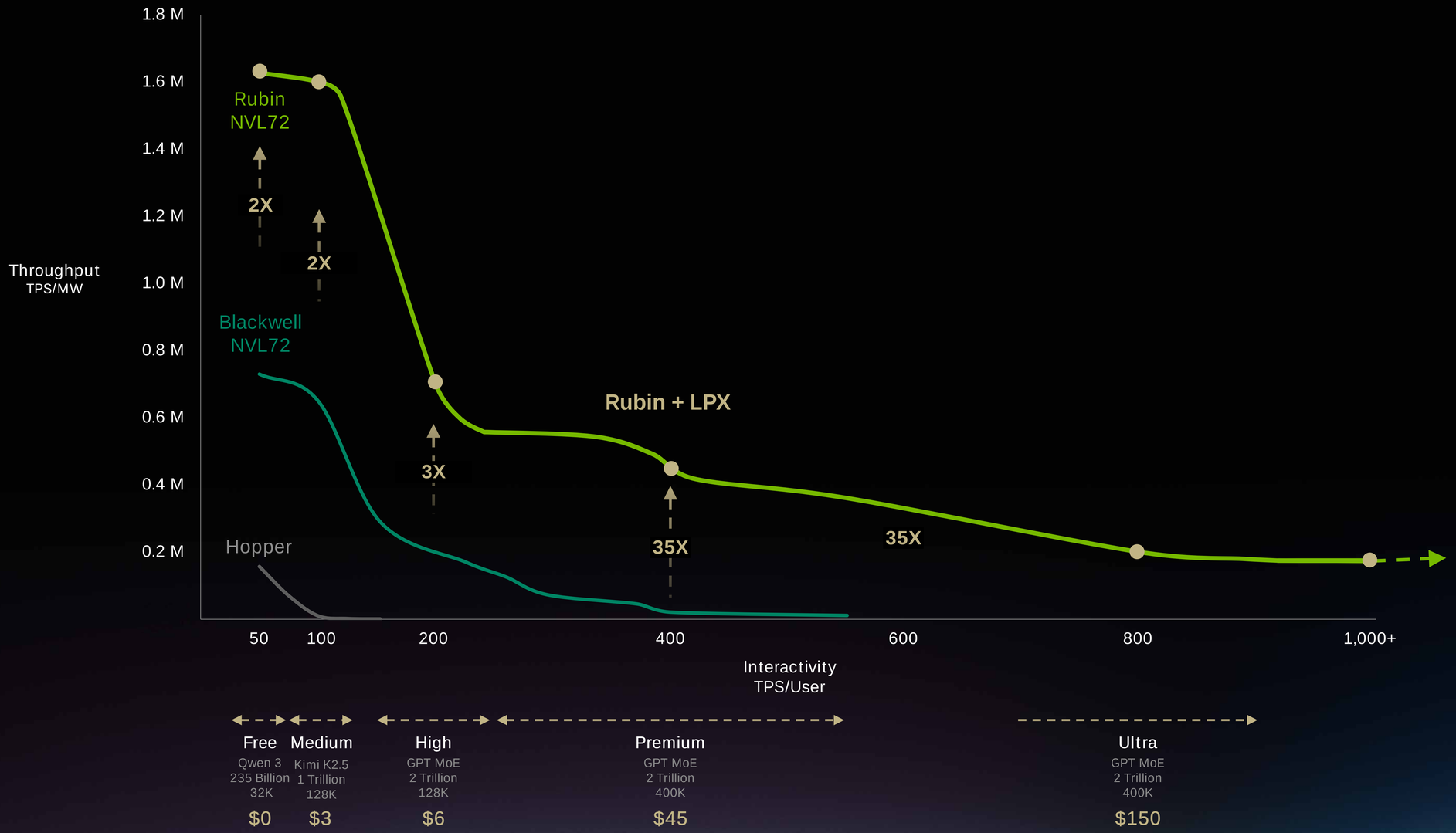
Inference Performance and Efficiency Drive Company Results



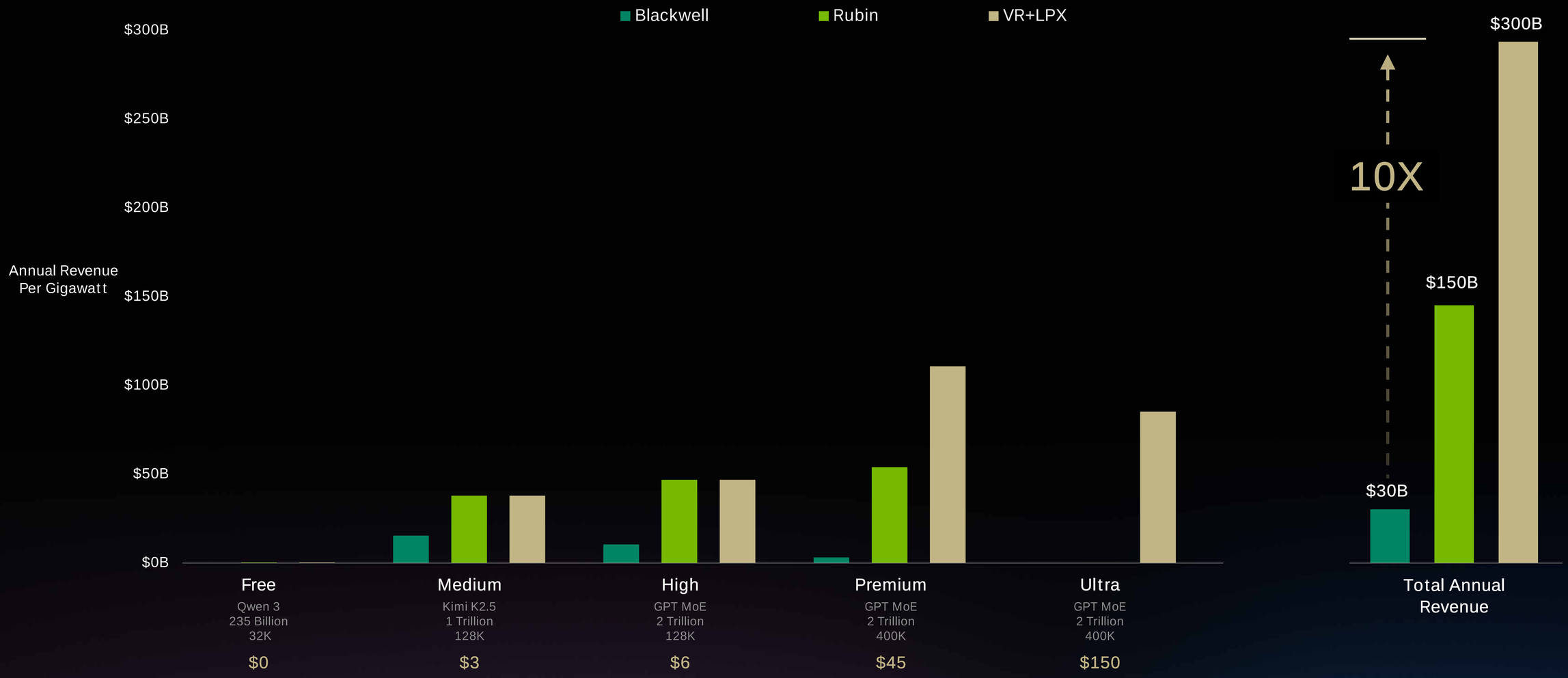
Inference Performance and Efficiency Drive Company Results



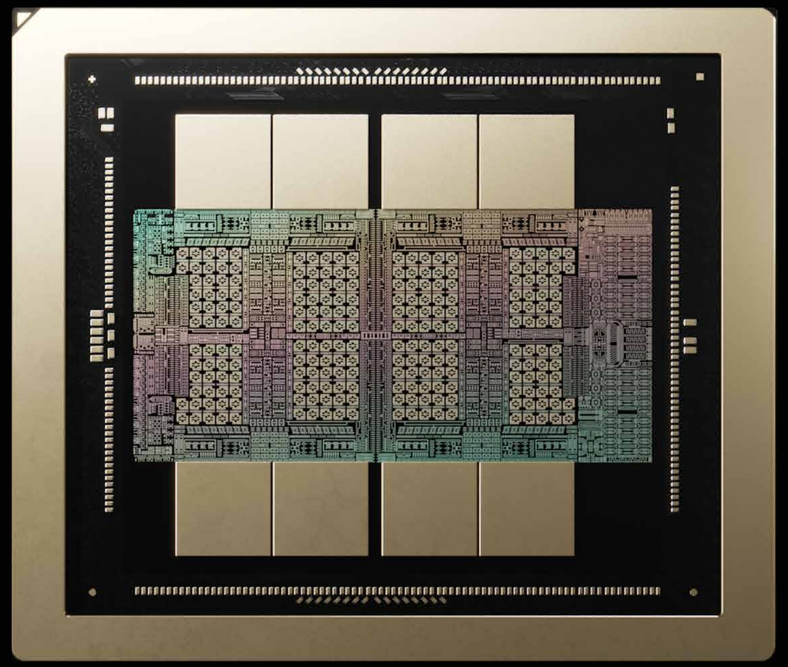
35X Performance Efficiency Drives Company Results



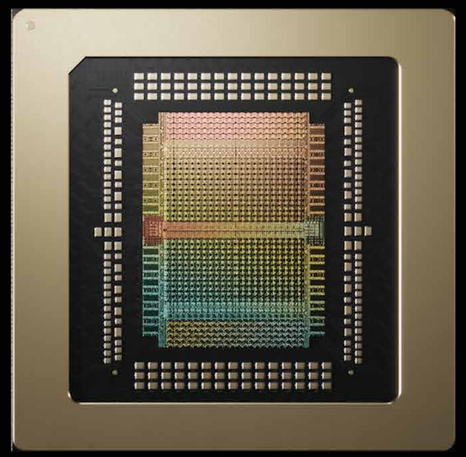
NVIDIA Vera Rubin + LPX Unlocks $300B Revenue Opportunity



Rubin GPU



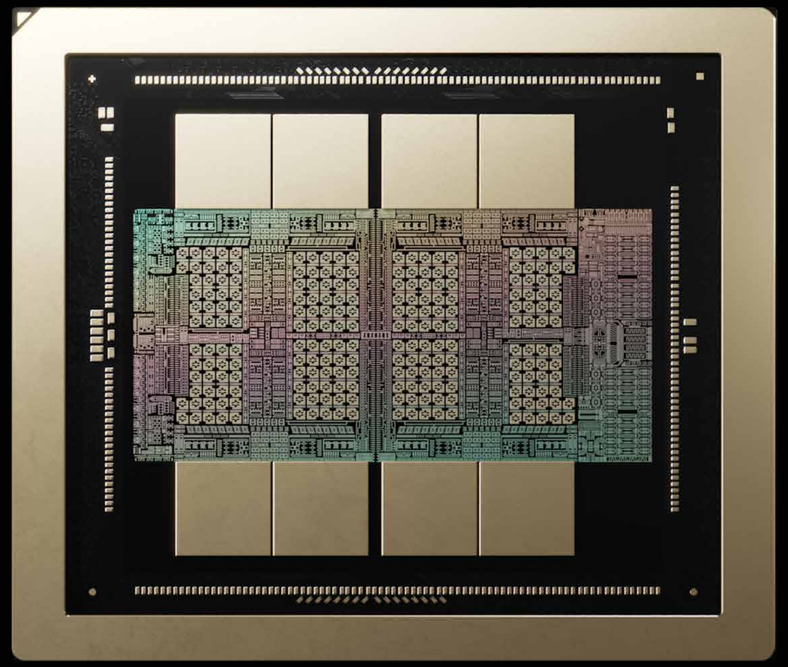
Groq 3 LPU



288 GB HBM4	500 MB SRAM
22 TB/s	150 TB/s SRAM Bandwidth
50 PFLOPs (NVFP4)	1.2 PFLOPs (FP8)
336B Transistors + 2.5T (HBM4)	98B Transistors

Uniting Processors of Extreme Performances

Rubin GPU



Groq 3 LPU

288 GB HBM4	4 GB SRAM	
22 TB/s	1,200 TB/s SRAM Bandwidth	55X
50 PFLOPs (NVFP4)	9.6 PFLOPs (FP8)	
336B Transistors + 2.5T (HBM4)	784B Transistors	

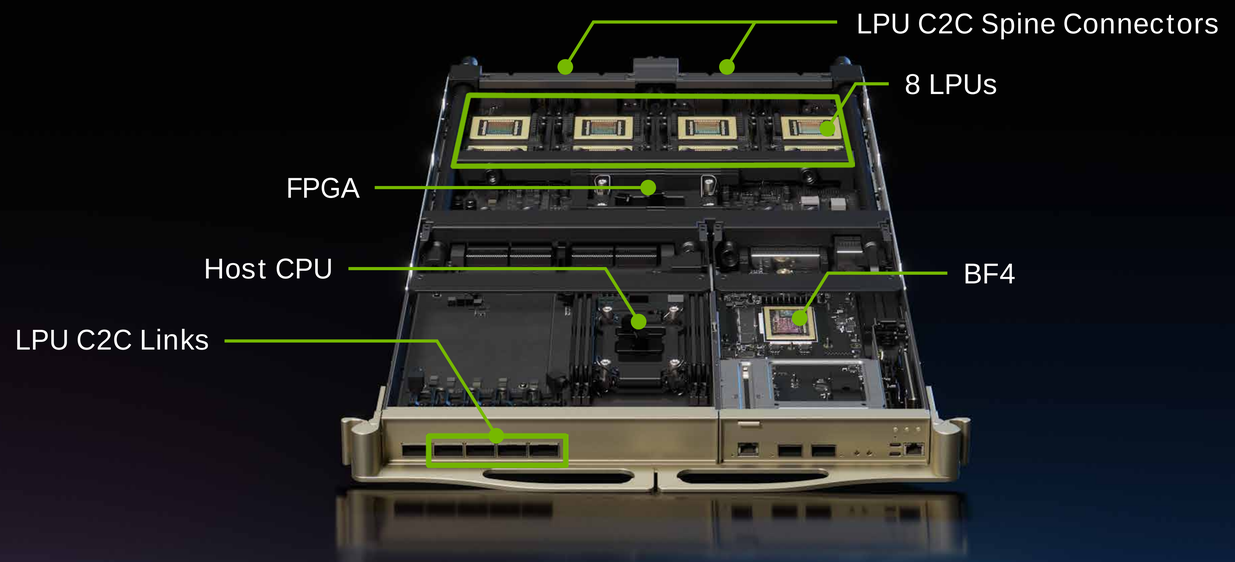
Uniting Processors of Extreme Performances

Uniting Processors of Extreme FLOPS and Bandwidth

Available 2H26

NVIDIA Groq 3 LPX

AI Inference Compute	**315 PFLOPS**
SRAM Capacity	**128 GB**
Memory Bandwidth	**40 PB/s**
Scale-Up Density	**256 Chips**
Scale-Up Bandwidth	**640 TB/s**



Announcing NVIDIA Vera Rubin NVL72 Launch Partners

10X Perf/W | 3.6 EF NVFP4 | 1.6 PB/s HBM4 | 260 TB/s NVLink6

AI Labs

ANTHROPIC · CURSOR · Meta · OpenAI · perplexity · Black Forest Labs · cohere · Harvey · Mistral AI · OpenEvidence · runway · THINKING MACHINES

Cloud

aws · Google Cloud · Microsoft Azure · ORACLE CLOUD Infrastructure · CoreWeave · Crusoe · Lambda · NEBIUS · NSCALE

OEMs and ODMs

DELL Technologies · HPE · Lenovo · SUPERMICRO · AIVRES · ASUS · FOXCONN · GIGABYTE · Inventec · PEGATRON · QCT · wistron · wiwynn

256 Vera CPUs | 300 TB/s LPDDR5X | ETL Spine | 6.5X Throughput

Announcing NVIDIA Vera CPU Launch Partners

Cloud

Meta
Oracle Cloud Infrastructure
CoreWeave
Crusoe
Lambda
Nebius
Nscale
together.ai

OEMs and ODMs

Cisco
Dell Technologies
HPE
Lenovo
Supermicro
AIVRES
ASRock Rack
ASUS
Compal
Foxconn
Gigabyte
hyve solutions
Inventec
MiTAC Computing
msi
Pegatron
QCT
Wistron
wiwynn

Announcing NVIDIA BlueField-4 STX Launch Partners

Cloud

CoreWeave · Crusoe · IREN · Lambda · Mistral AI · NEBIUS · Oracle Cloud Infrastructure · VULTR

OEMs

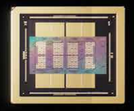
AIC · Cloudian · ddn · Dell Technologies · Everpure · HPE · Hitachi Vantara · IBM · MinIO · NetApp · Nutanix · QCT · Supermicro · VAST · WEKA

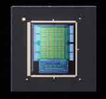
5x Tokens/sec | 50 Tb/s Networking BW | 16TB Shared Context / GPU

NVIDIA Vera Rubin

7 Chips – 5 Rack Systems
AI Factory for the Agentic AI Frontier

1 GW AI Factory	X86 + Hopper	Vera Rubin
# of GPUs	600K	300K
AI FLOPS	1.2 ZFLOPS	16 ZFLOPS
All-to-All Scale-up	7.2 TB/s	260 TB/s
Memory BW-per-Domain (GROQ SRAM)	2 EB/s	100 EB/s
Tokens per Second	2M	700M


Rubin GPU


Vera CPU


CX9


BF4


NVLink Switch


Spectrum CPO


Groq 3 LPU


Vera Rubin Compute Tray


NVLink Switch Tray


Vera Compute Tray


BF4 STX Server


Spectrum Switch


Groq 3 Compute Tray

NVIDIA Extreme Co-Design Delivering X-Factors Every Year
From Chips to Racks to AI Factories

NVIDIA DSX AI Factory Platform

Extreme Co-Design at Infrastructure Scale

Reference Designs, Best Practices, Methodologies

Rubin DSX Reference Designs

DSX Ecosystem

Omniverse DSX Blueprint

DSX Libraries, APIs, Software

DSX Max-Q

DSX Flex

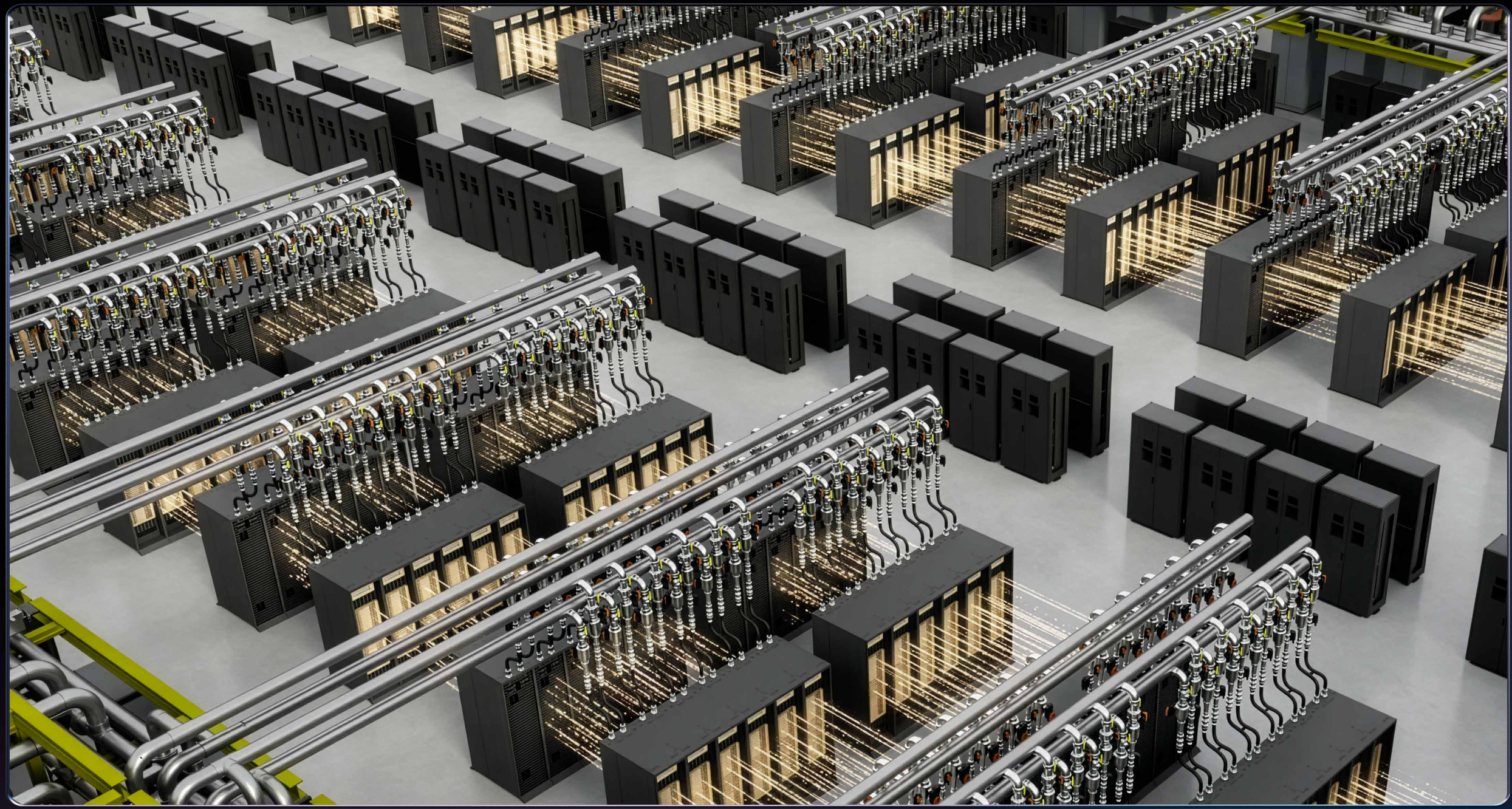
DSX Exchange

DSX Sim

Chips, Systems, Facilities

Power

Liquid Cooling



NVIDIA DSX AI Factory Platform

Accelerates Scalable, Energy-Efficient AI Factory Deployment

DSX Flex

CoreWeave · Crusoe · Digital Realty · emeraldai · Nebius · Nscale · switch

DSX Max-Q

CoreWeave · Crusoe · Digital Realty · emeraldai · Nebius · Nscale · Red Hat · run:ai · phaidra · switch

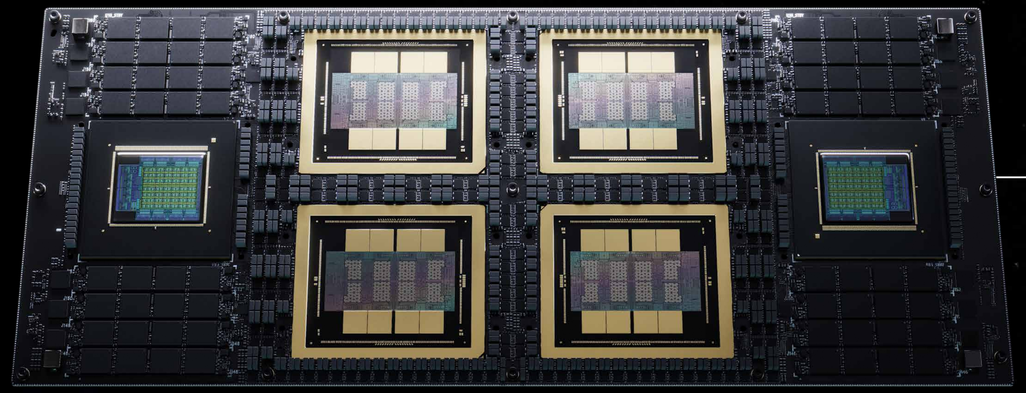
DSX Sim

ABB · Armada · Aveva · aviz · Bechtel · networks · cadence · Check Point · Cloudian · Dassault Systèmes · Dell Technologies · Eaton · etap · GE Vernova · hedgehog · Hitachi · Jacobs · Keysight · metalsoft · Mirantis · Mitsubishi Electric · netris · Nokia · Open Nebula · paloalto networks · ptc · Rafay · Red Hat · run:ai · Schneider Electric · Siemens · Siemens Energy · slurm · Spectro Cloud · Trane Technologies · TrendAI · VAST · vCluster · Vertiv · WEKA

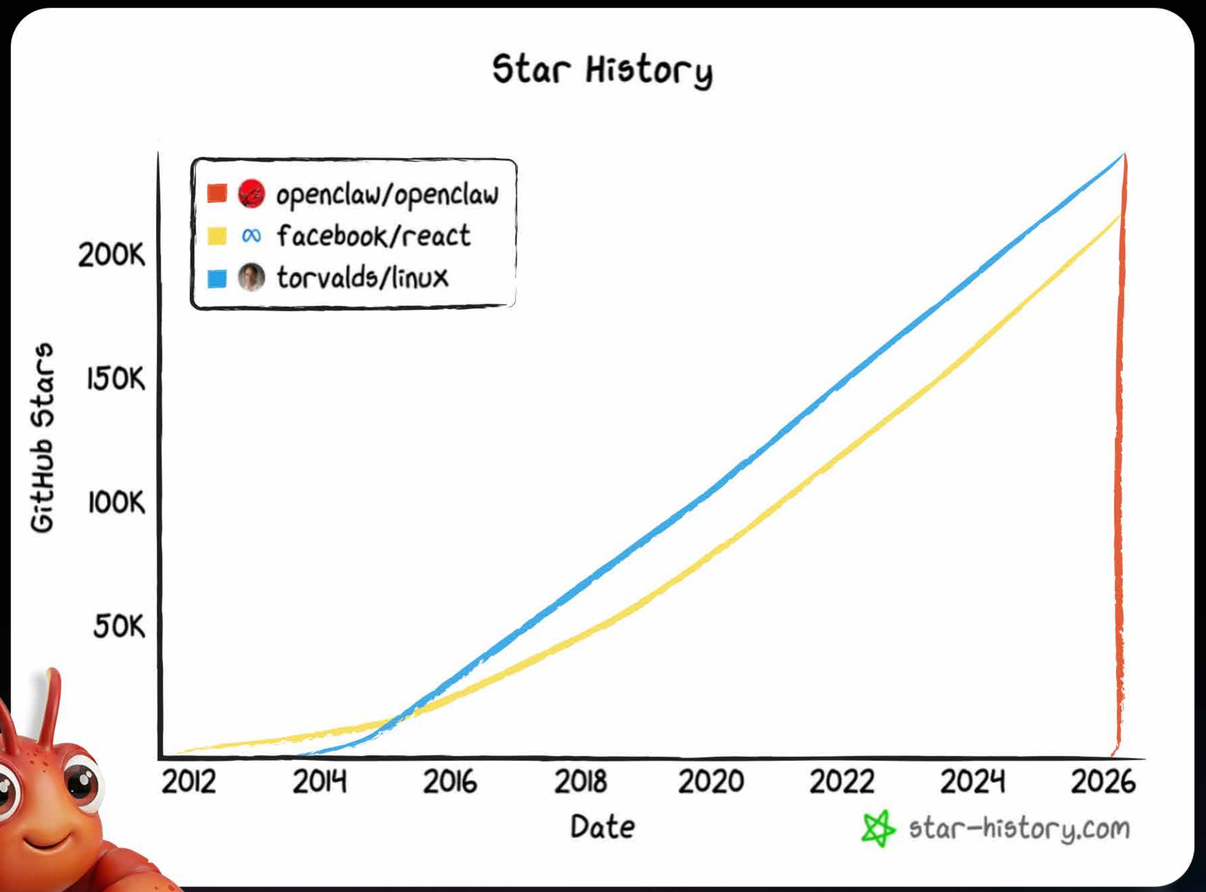
DSX Exchange

ABB · Armada · Aveva · aviz · networks · cadence · Check Point · Cloudian · CoreWeave · Crusoe · Digital Realty · Dell Technologies · Eaton · emeraldai · GE Vernova · hedgehog · Hitachi · Keysight · metalsoft · Mirantis · Mitsubishi Electric · Nebius · netris · Nokia · Nscale · Open Nebula · paloalto networks · phaidra · Rafay · Red Hat · run:ai · Schneider Electric · Siemens Energy · slurm · Spectro Cloud · switch · Trane Technologies · TrendAI · VAST · vCluster · Vertiv · WEKA

Announcing NVIDIA Space-1 Vera Rubin Module



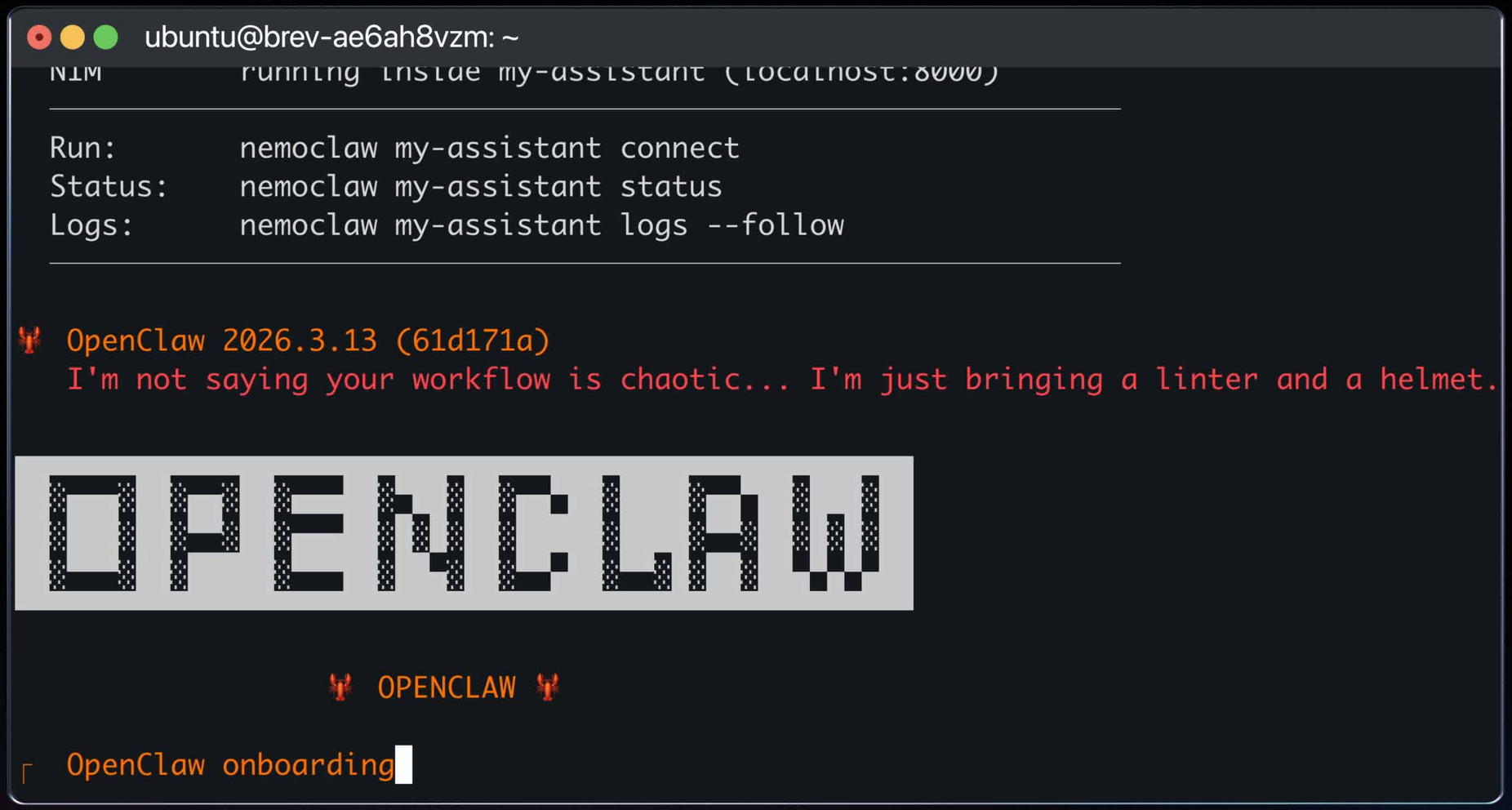
Inference Inflection Arrives



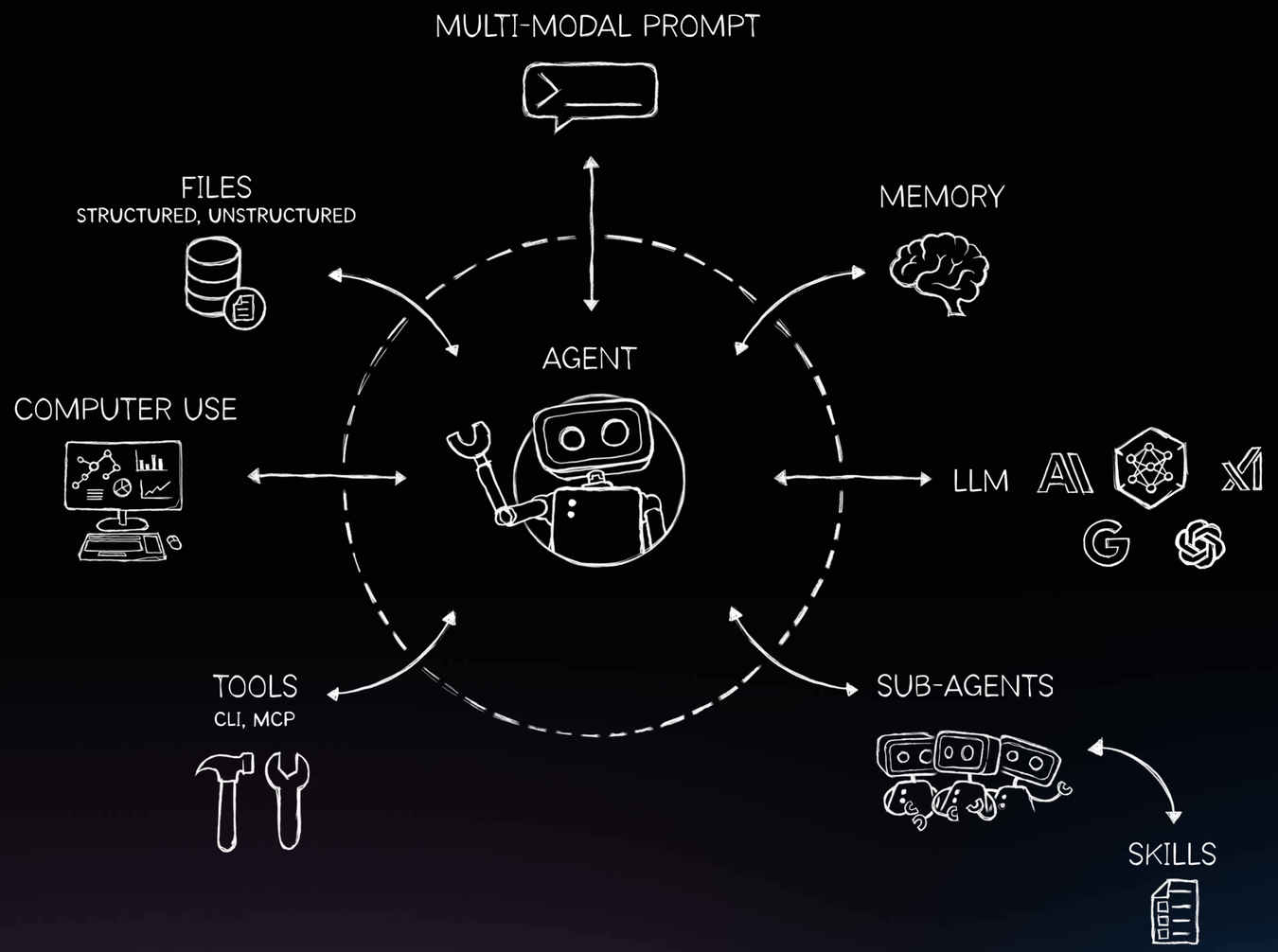
Announcing NVIDIA NemoClaw for OpenClaw

ubuntu@brev-ae6ah8vzm: ~

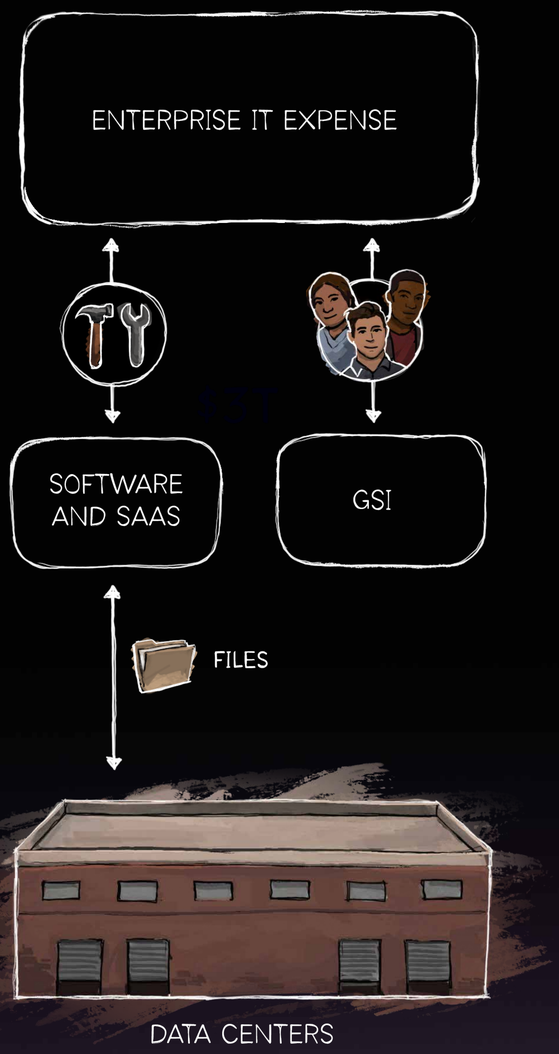
```
➜  ~ curl -fsSL https://nvidia.com/nemoclaw.sh | bash
➜  ~ nemoclaw onboard
```


ubuntu@brev-ae6ah8vzm: ~
NIM running inside my-assistant (localhost:8000)
Run: nemoclaw my-assistant connect
Status: nemoclaw my-assistant status
Logs: nemoclaw my-assistant logs --follow
OpenClaw 2026.3.13 (61d171a)
I'm not saying your workflow is chaotic... I'm just bringing a linter and a helmet.
OPENCLAW
OPENCLAW
OpenClaw onboarding

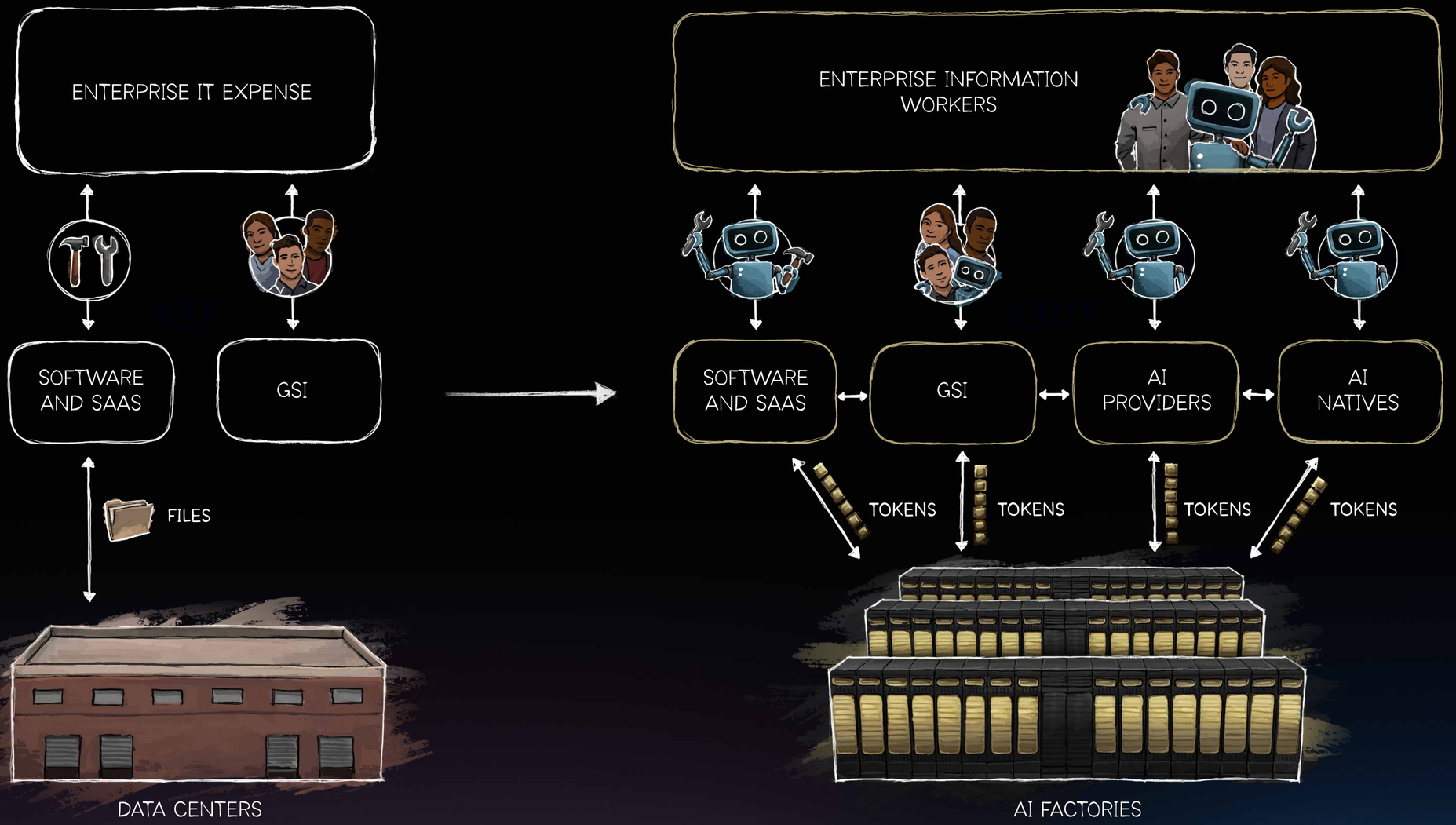
Agents – A New Computing Platform



Enterprise IT Renaissance from SaaS to Agent-as-a-Service

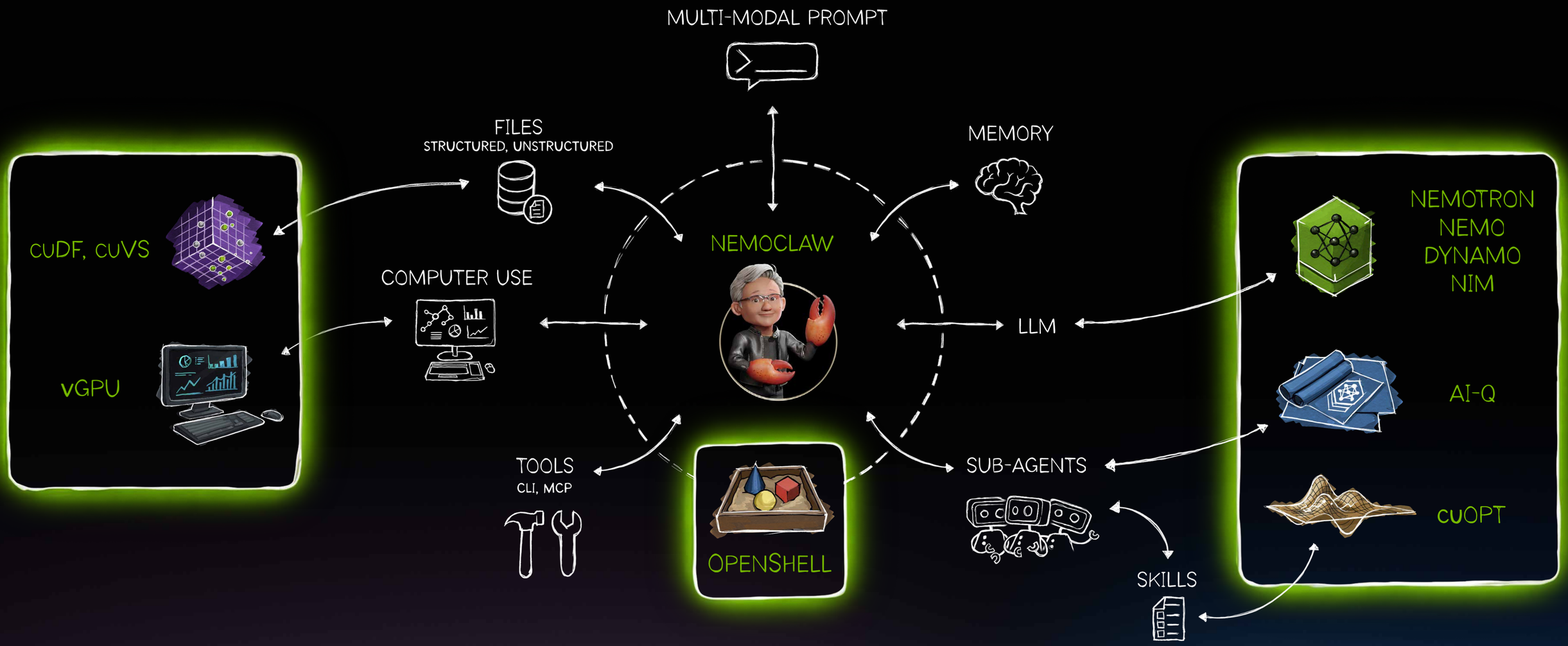


Enterprise IT Renaissance from SaaS to Agent-as-a-Service

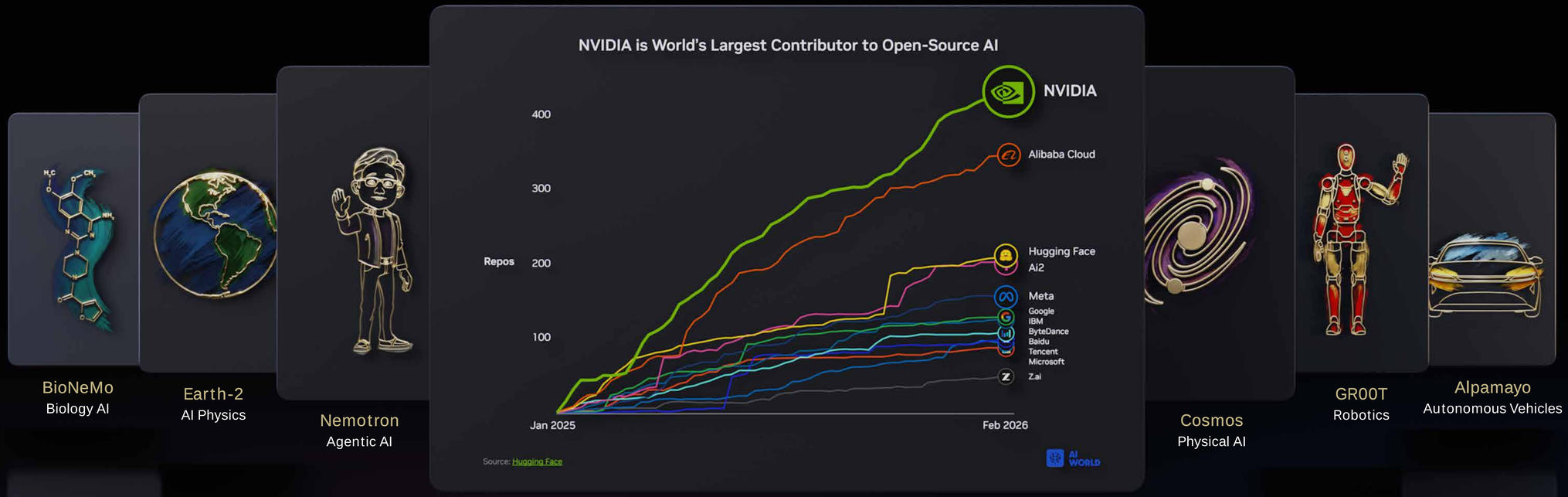


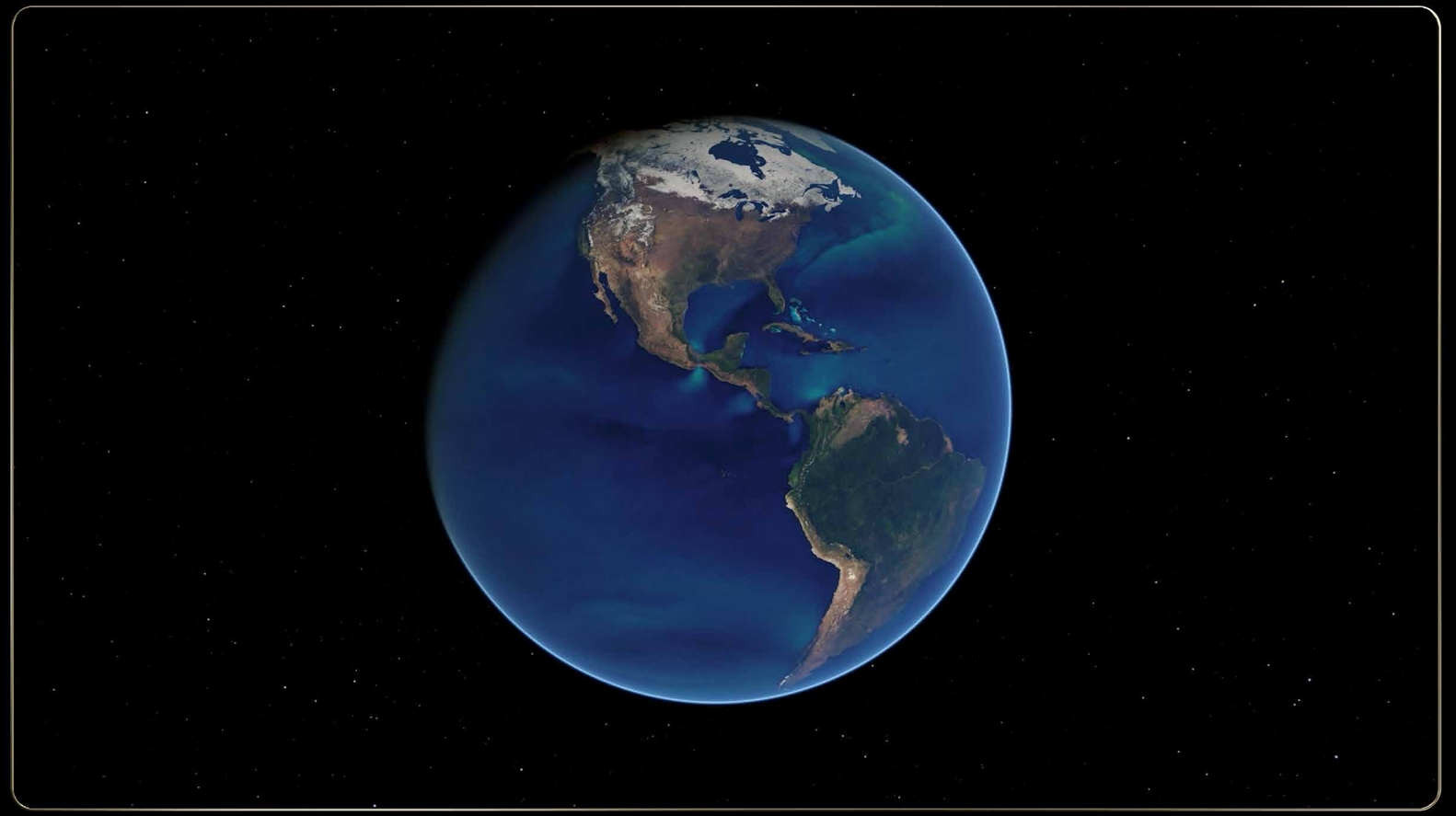
Announcing NVIDIA NemoClaw Reference OpenClaw

NVIDIA Agent Toolkit for Building Specialized Agents



NVIDIA Leaderboard Topping Open Models

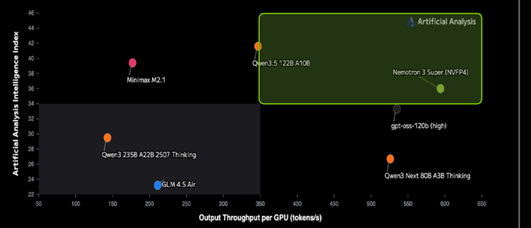




Earth-2 Medium Range NEW

AA Intelligence vs Efficiency

Nemotron 3 Super



PinchBench

Nemotron 3 Super

DeepResearch Bench II

Nemotron 3 Super (AI-Q)

1	**nvidia-aiq (Nemotron 3, Opus 4.6)** NVIDIA	**54.58**
2	**Xiaoyi DeepResearch Evo** Huawei, 2026	**49.34**
3	**Xiaoyi DeepResearch** Huawei, 2026	**46.98**
4	**OpenAI-GPT-o3 Deep Research** OpenAI, 2025	**45.40**
5	**Gemini-3-Pro Deep Research** Google, 2025	**44.60**
6	**Gemini-2.5-Pro Deep Research** Google, 2024	**41.98**
	[illegible]	**40.99**

ViDoRe, MMTEB, BRIGHT

Nemotron RAG

1	nemotron-colembed-vl-8b-v2	16722	8.7	4096
2	tomoro-colqwen3-embed-8b	16724	8.0	320
3	nemotron-colembed-vl-4b-v2	9206	4.8	2560
4	Ops-Colqwen3-4B	9206	4.8	2560
5	tomoro-colqwen3-embed-4b	8466	4.0	320
6	llama-nemotron-colembed-vl-3b-v2	8403	4.4	3072
7	SauerkrautLM-ColQwen3-8b-v0.1	15536	8.1	128
8	jina-embeddings-v4	7500	3.9	2048
9	colnomic-embed-multimodal-7b	14400	7.0	128
10	llama-nemoretriever-colembed-3b-v1	8403	4.4	3072
11	SauerkrautLM-ColQwen3-4b-v0.1	8465	4.4	128

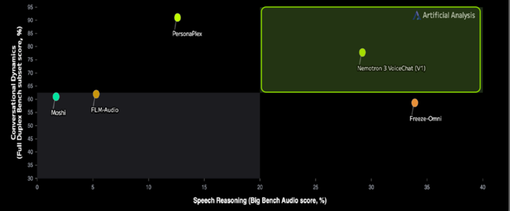
OCRBench v2

NemotronVL

5	Nemotron Nano V2 VL	-	Yes	12B	61.2	67.6	54.5
6	Gemini-2.5-Pro	-	No	-	59.3	70.9	45.8
7	GLM-4.6V-Flash	-	No	-	59.0	75.3	52.5
8	Llama-3.1-Nemotron-Nano-VL-8B-V1	-	Yes	8B	56.4	62.9	61.3
9	GPT5-2025-08-07	-	No	-	55.5	69.3	28.2
10	Ovis2.5-8B	Arxiv 2025	Yes	8B	54.1	63.6	31.2
11	Gemini1.5-Pro	Arxiv 2024	No	-	51.6	59.1	41.2
12	GPT-4.2	-	No	-	50.5	57.5	29.3

AA Conversational Dynamics vs Reasoning

Nemotron VoiceChat



Physical AI Bench

Cosmos Predict

Cosmos-Predict2.5-14B	81.0	83.8	78.1	94.2	67.8	79.9	87.7
Cosmos-Predict2.5-2B	81.0	84.0	77.9	94.1	66.1	80.8	87.8
Wan2.2-I2V-A14B	80.6	84.1	77.2	93.2	66.3	81.7	89.2
Wan2.2-TI2V-5B	80.4	83.4	77.4	93.1	65.2	79.3	88.4
Cosmos-Predict2-14B-Video2World	80.0	84.3	75.8	93.1	67.1	80.3	86.9
Wan2.1-I2V-14B-720P	79.8	82.7	76.8	90.6	66.9	80.1	89.7
Cosmos-Predict2-2B-Video2World	79.6	83.9	75.2	92.0	66.1	80.6	86.0
MAGI-1-24B	78.5	80.5	76.5	90.6	61.8	73.5	84.1
CogVideoX1.5-5B-I2V	78.3	80.1	76.6	89.1	59.7	73.0	84.4

Physical Reasoning from Video

Cosmos Reason

Cosmos-Reason2-8B	58.14	47.19	59.14
Cosmos-Reason1-7B	59.88	41.31	48.17
V-JEPA 2	56.4	44.5	44.89
GPT-4o	53.19	32.5	50.95
Qwen2.5-VL	49.12	36.7	49.05
Gemini 2.5 Flash	56.1	-	61.66
Perception Language Model (PLM)	-	39.7	50.06
InternVL2.5	-	39.9	47.54

LIBERO

Cosmos Policy

Cosmos Policy	98.1	100.0	98.2	97.6	98.5
Diffusion Policy	78.3	92.5	68.3	50.5	72.4
Dita	97.4	94.8	93.2	83.6	92.3
π0	96.8	98.8	95.8	85.2	94.2
UVA	--	--	--	90.0	--
UniVLA	96.5	96.8	95.6	92.0	95.2
π0.5	98.8	98.2	98.0	92.4	96.9
Video Policy	--	--	--	94.0	--
OpenVLA-OFT	97.6	98.4	97.9	94.5	97.1
CoRVLA	98.6	98.8	96.6	95.4	97.4

MolmoSpaces

GR00T N

▸	1	DreamZero DROID	NVIDIA GEAR	52.06 ±3.11
▸	3	$\pi_{0.5}$ DROID	Physical Intelligence	36.40 ±3.03
▸	3	CAP-C + Gemini-ER-1.5	NYU & Berkeley	33.80 ±2.99
▸	5	π_0-FAST DROID	Physical Intelligence	22.00 ±2.67
▸	6	CAP-B + Gemini-ER-1.5	NYU & Berkeley	16.30 ±2.42

RoboArena

GR00T N

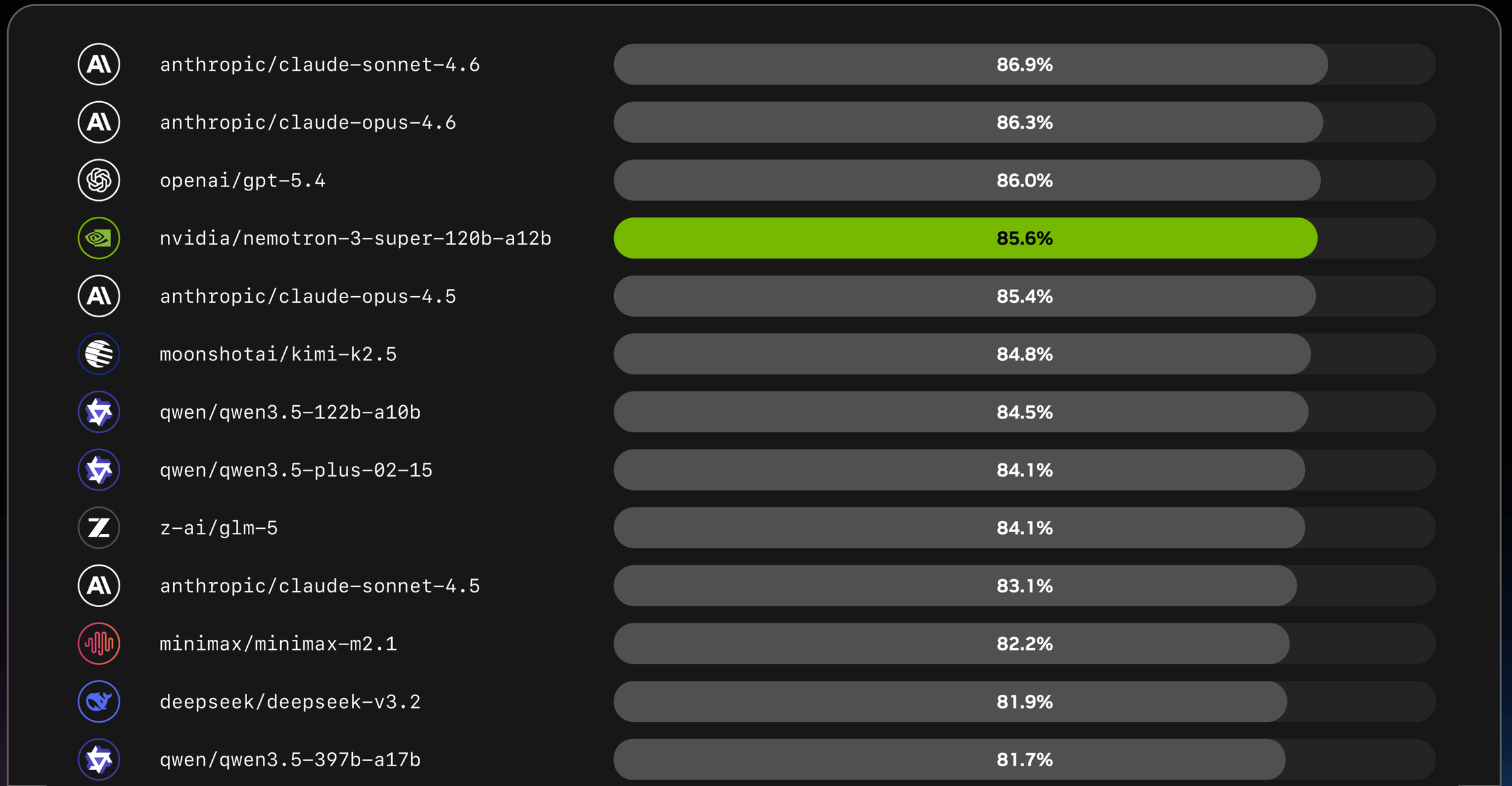
1	DreamZero X	1752	40.4	168
2	pi05_droid X	1621	27.2	615
3	pi0_fast_droid X	1595	25.8	797
4	floppy_bagel_195k_temp1p0	1588	147.1	8
5	floppy_bagel_525k_temp1p0	1560	74.4	33
6	paligemma_vq_droid	1559	25.9	814
7	paligemma_diffusion_droid	1548	25.9	807
8	paligemma_fast_specialist_droid	1546	24.7	1,007
9	paligemma_fast_droid	1536	25.1	1,007

LingoQA Waymo E2E

Alpamayo

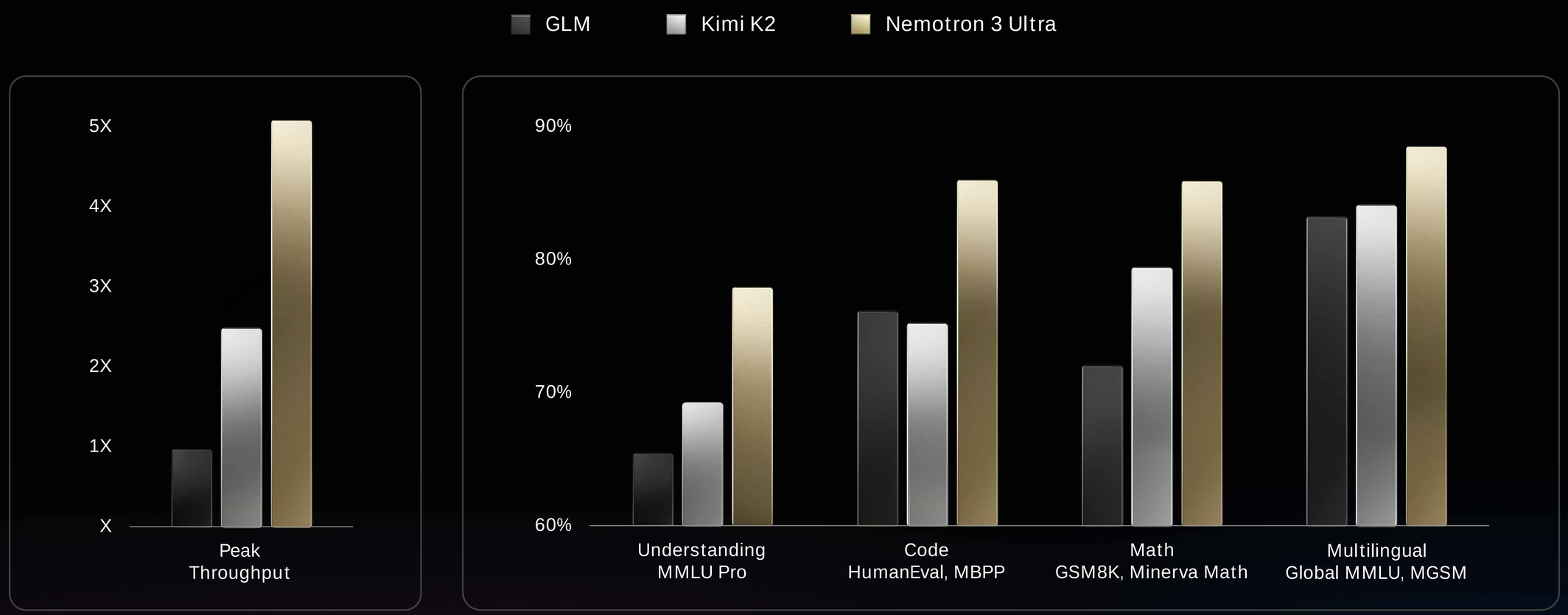
1	**Alpamayo 1.5**	**71.6**
2	MiMo-Embodied	69.9
3	RoboTron-Drive	69.2
4	Cosmos-Reason1	66.2
5	ACE-Brain-0	65.8
6	Gemini 2.5 Pro	64.1
7	Qwen2.5-VL	62.2
8	LingoQA paper model (fine-tuned 5-frame)	60.8
9	Alpamayo 1	60.4
10	GPT-4V	59.6
11	Vlaser	59.6
12	[illegible]	59

NVIDIA Nemotron 3 Super – Best Open Model for OpenClaw



NVIDIA Nemotron 3 Ultra – Best Open Base Model

5X Efficiency and Highest Reasoning Accuracy on NVIDIA GB200 NVL72



The World Building Regional AI With NVIDIA Nemotron

Announcing Global AI Leaders Join NVIDIA Nemotron Coalition to Advance Open Frontier Models

NVIDIA

❤️❤️❤️

Black Forest Labs

CURSOR

LangChain

Mistral AI

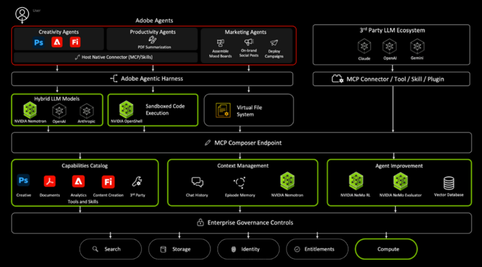
perplexity

Reflection

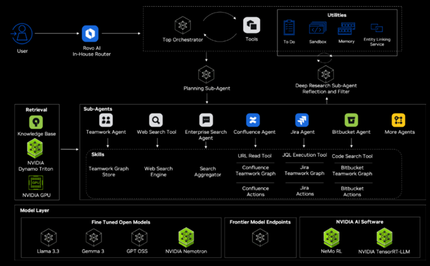
sarvam

THINKING MACHINES

Adobe

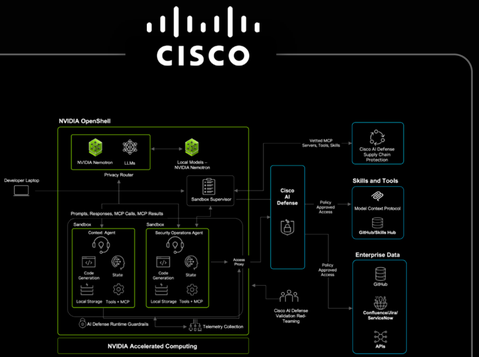


ATLASSIAN

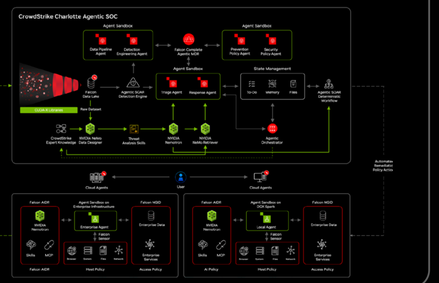


cādence

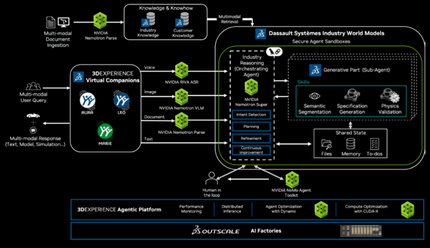
cisco



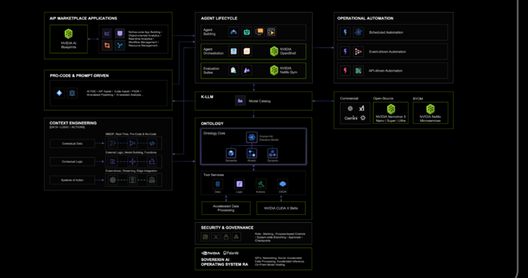
CROWDSTRIKE



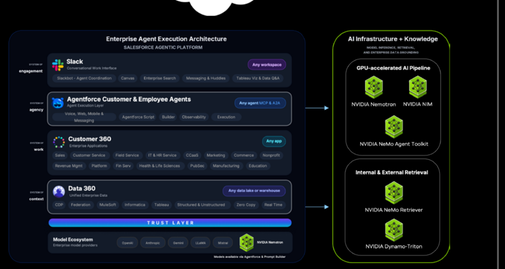
DASSAULT SYSTEMES



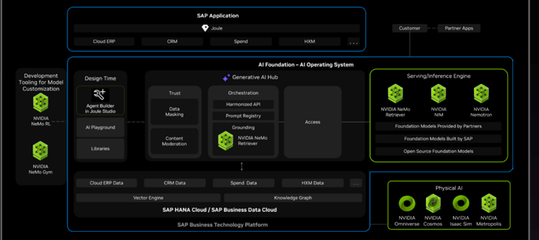
Palantir



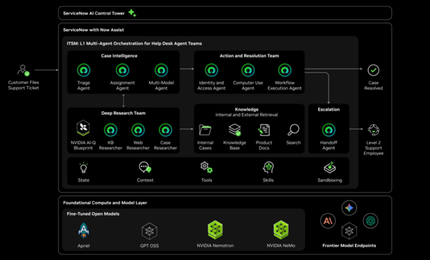
salesforce



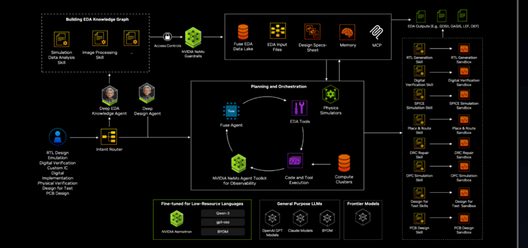
SAP



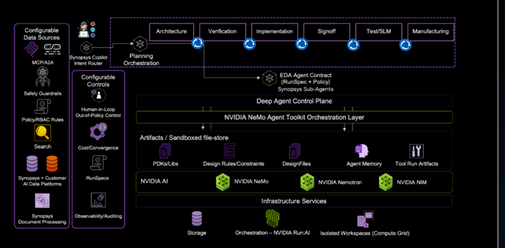
servicenow



SIEMENS

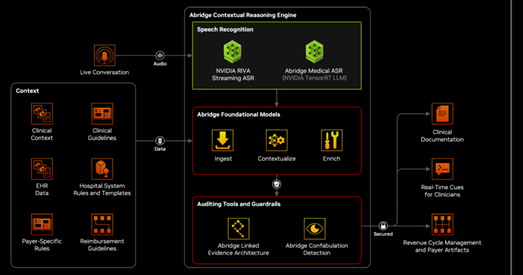


SYNOPSYS®

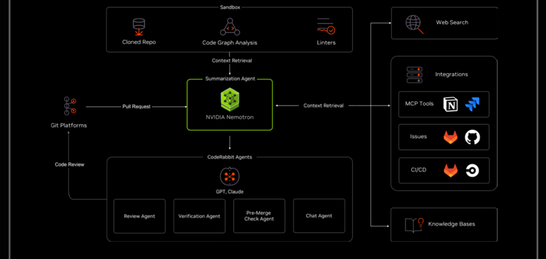


Alibaba · Automation Anywhere · aws · baseten · ByteDance · Cloudian · crewai · Cognition · CoreWeave · Dell Technologies · ddn · Everpure · Factory · Google Cloud · GCore · glean · Honeywell · harbor · Harmonic · HPE · Hugging Face · IBM · LangChain · LMCache · llm-d · Meituan · NEBIUS · NetApp · Ollama · Oracle Cloud Infrastructure · OpenEnv · openCode · OpenRouter · OpenHands · Prime Intellect · PyTorch · SoftBank · Tencent Cloud · together.ai · unsloth · VAST · VULTR · WEKA

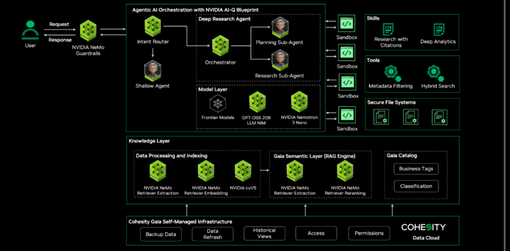
ABRIDGE



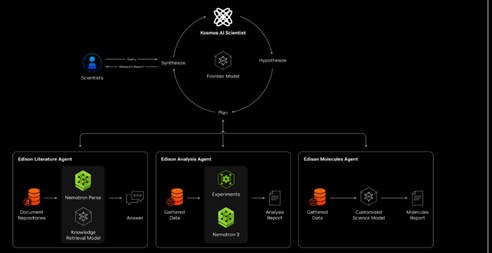
CodeRabbit



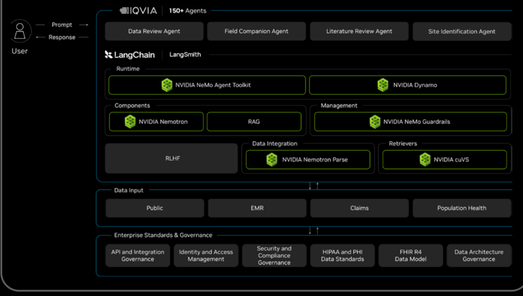
COHESITY



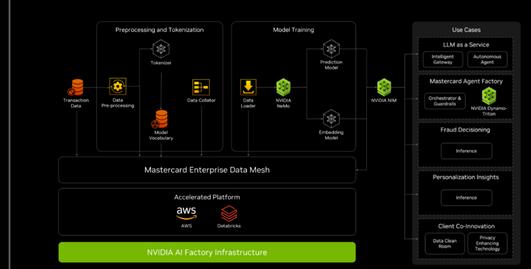
Edison



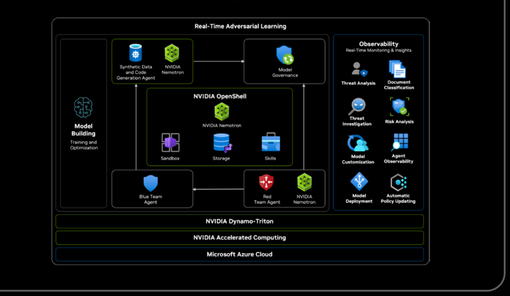
IQVIA



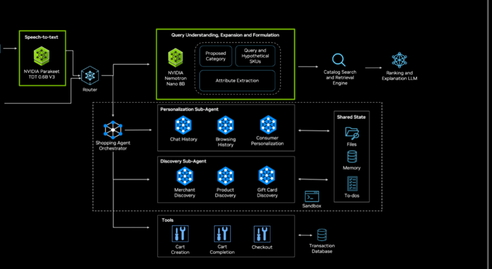
mastercard



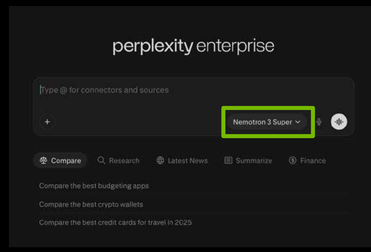
Microsoft Security NEXT AI



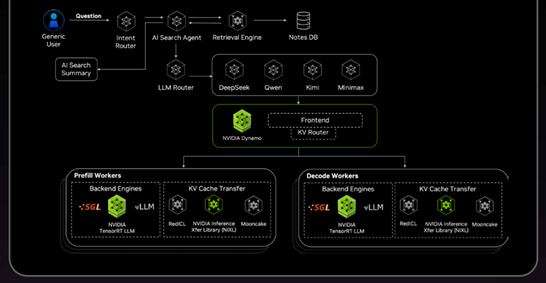
PayPal



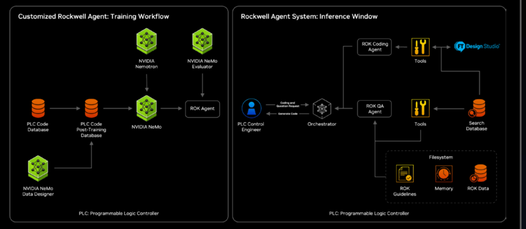
perplexity



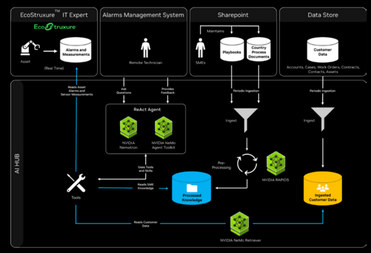
小红书



Rockwell Automation



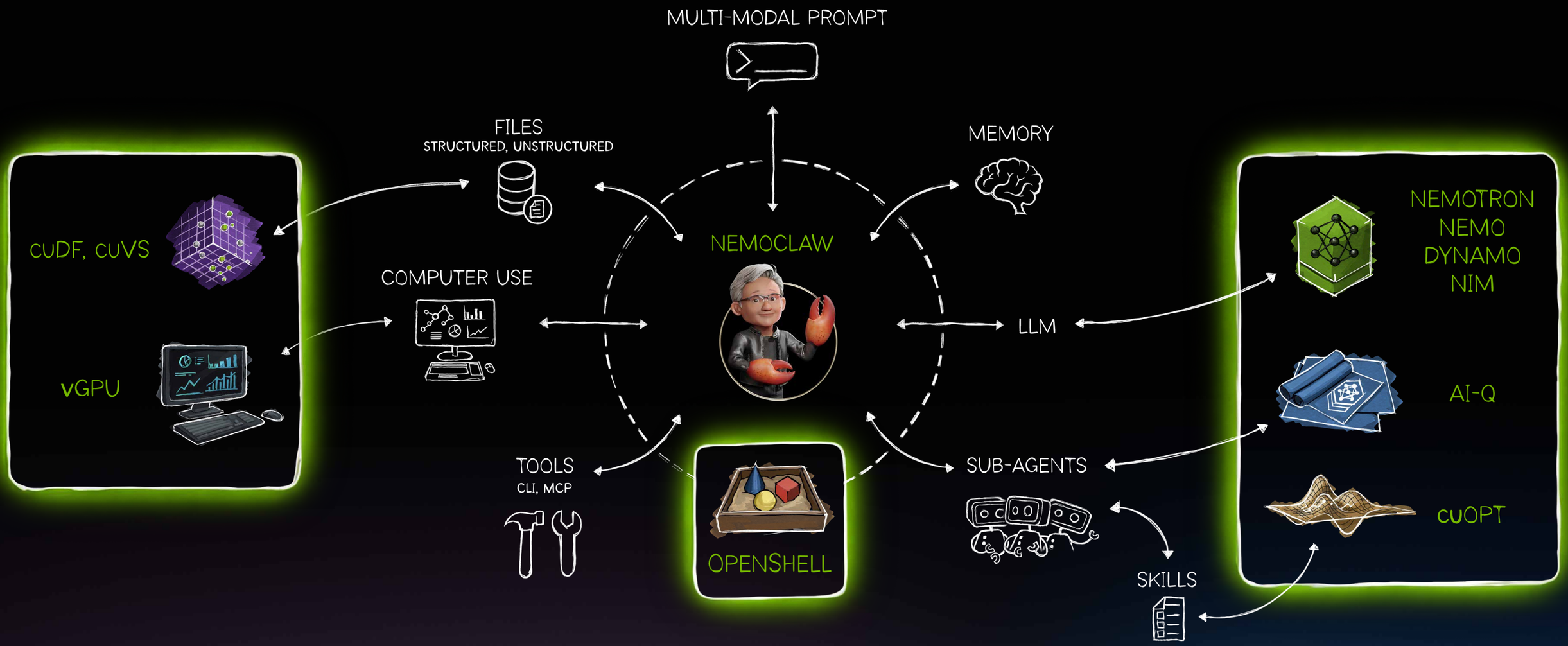
Schneider Electric



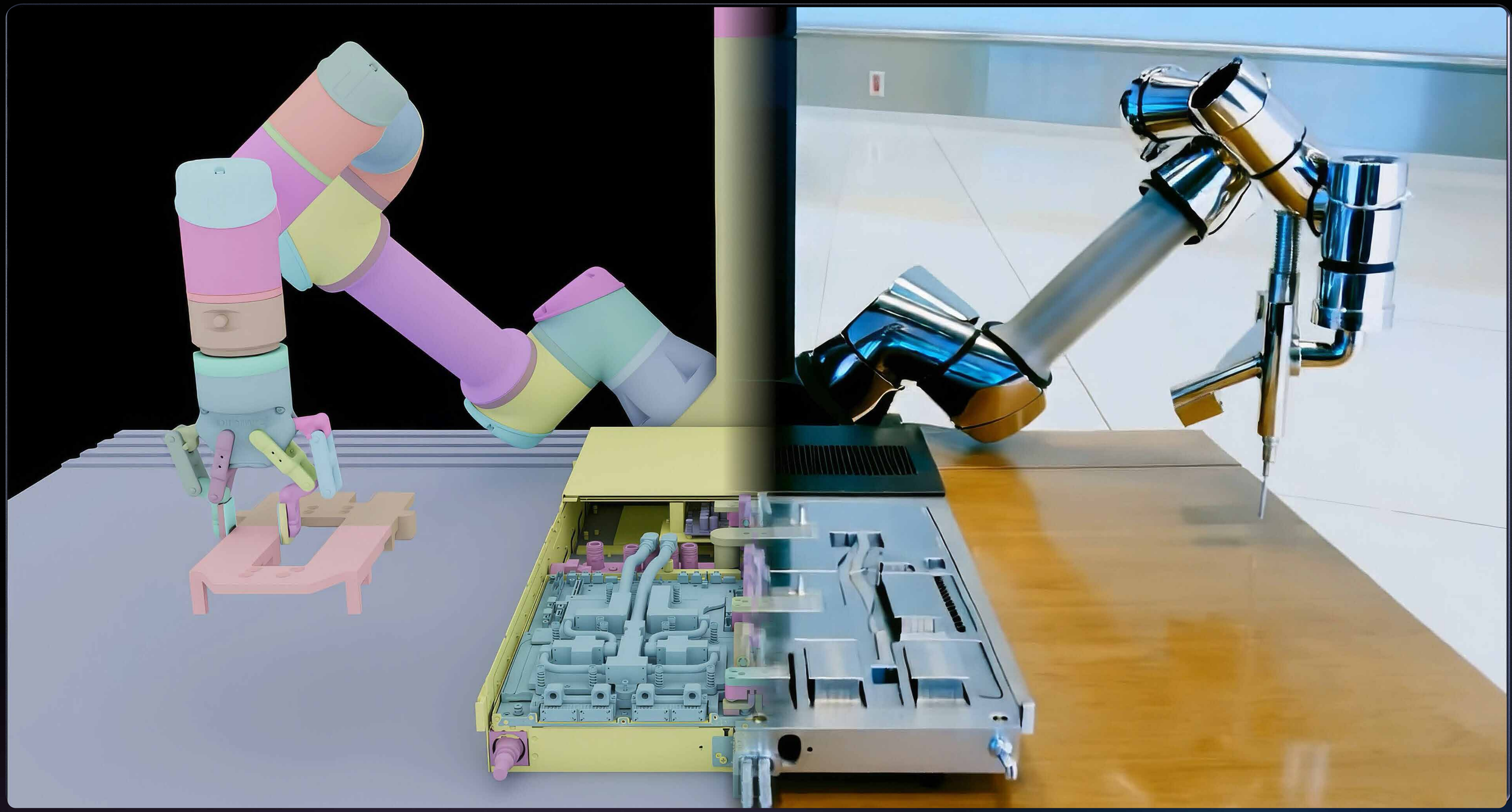
Alibaba · Automation Anywhere · aws · baseten · ByteDance · Cloudian · crewai · Cognition · CoreWeave · Dell Technologies · ddn · Everpure · Factory · Google Cloud
GCORE · glean · Honeywell · harbor · Harmonic · HPE · Hugging Face · IBM · LangChain · LMCache · llm-d · Meituan · NEBIUS · NetApp · Ollama · Oracle Cloud Infrastructure
OpenEnv · openCode · OpenRouter · OpenHands · Prime Intellect · PyTorch · SoftBank · Tencent Cloud · together.ai · unsloth · VAST · VULTR · WEKA

Announcing NVIDIA NemoClaw Reference OpenClaw

NVIDIA Agent Toolkit for Building Specialized Agents




Caterpillar
Mentee
Persona AI
NEURA Robotics
Franka Robotics
Agile Robots
AgiBot
RobotERA
T-Mobile
Zipline
ABB Robotics
NVIDIA
Boston Dynamics
Figure
Foxconn Fii
Agility
1X
LG Electronics
XPENG
Fauna Robotics
Hexagon Robotics
AgiBot
Linde
Universal Robots
Humanoid
Diligent
Mercedes-Benz
Enchanted Tools
Disney Research
Serve Robotics
Pollen Robotics
Idealworks




FROZEN


NVIDIA