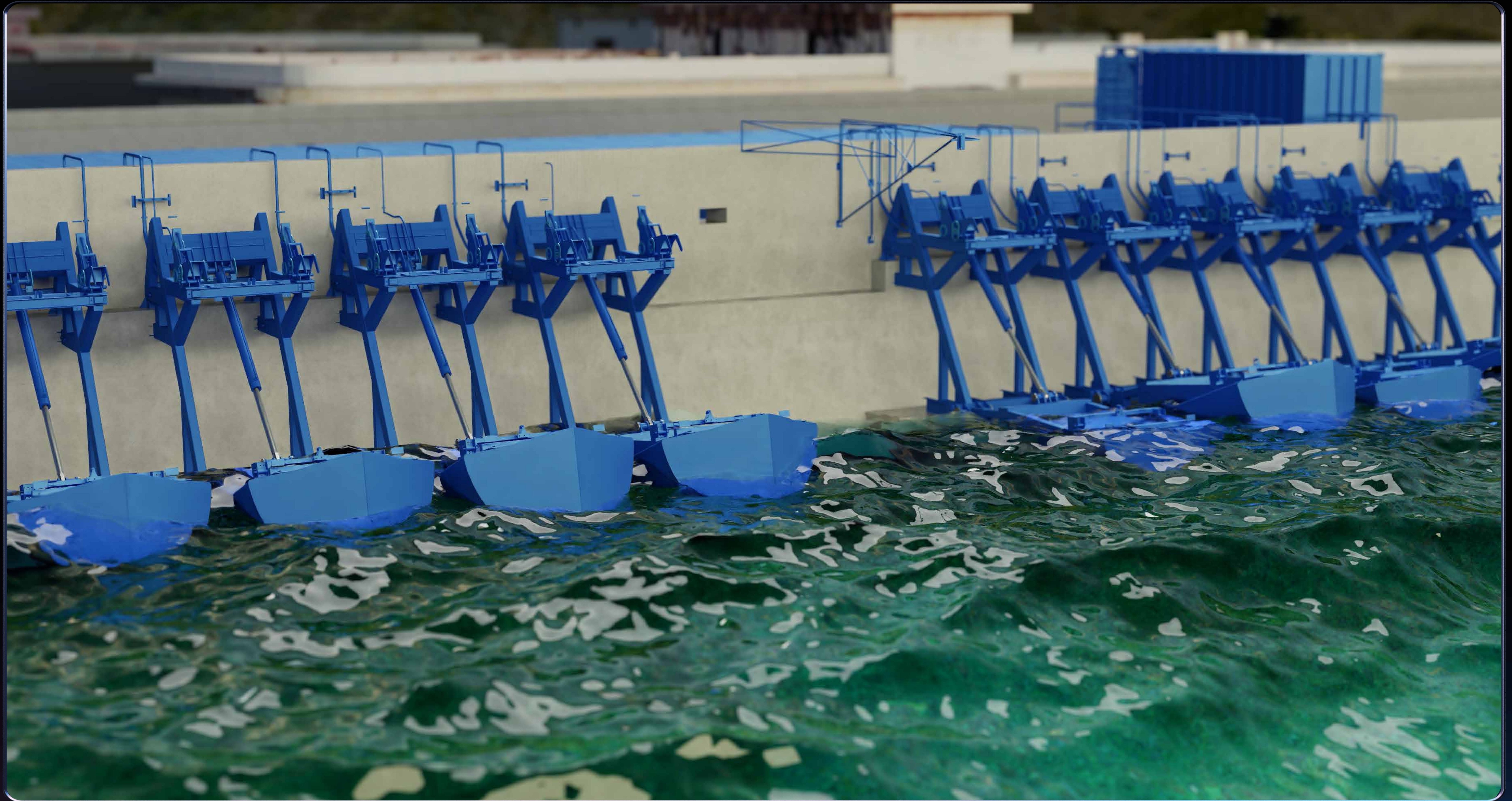


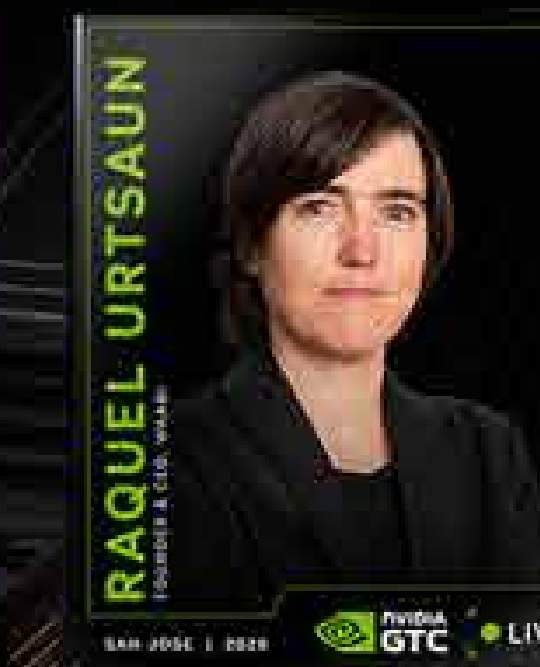
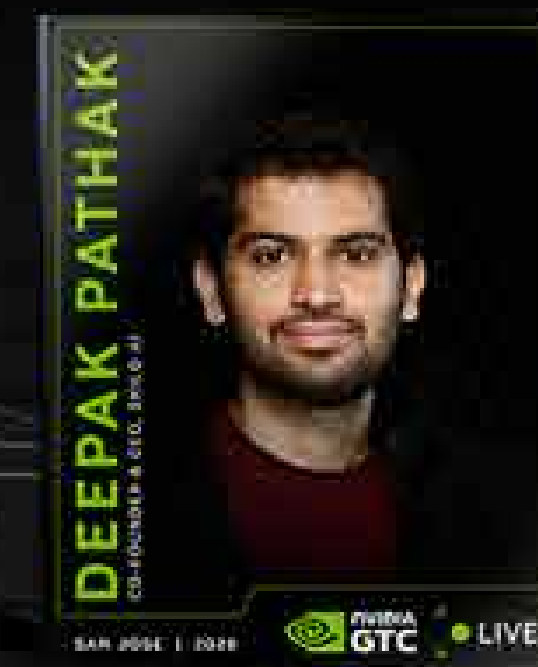
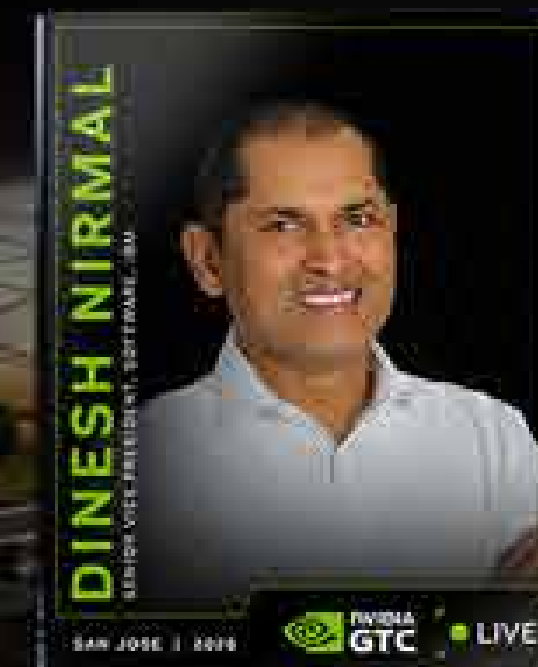
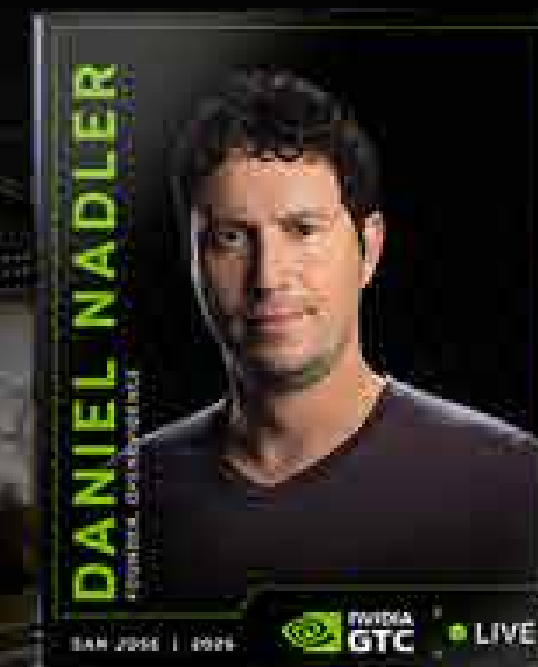
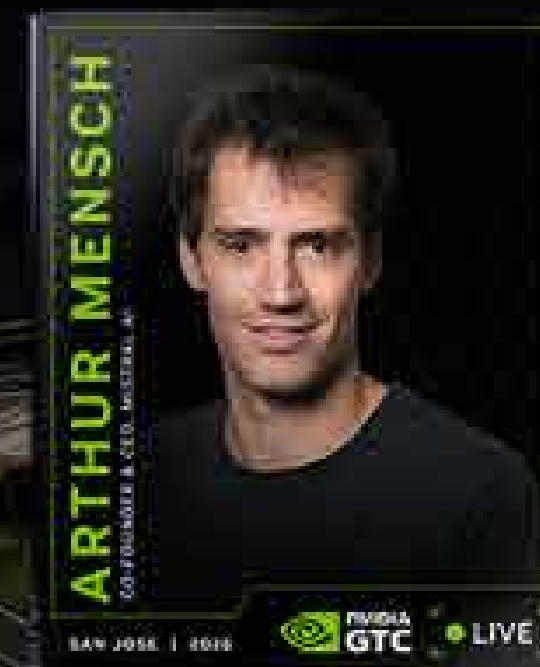
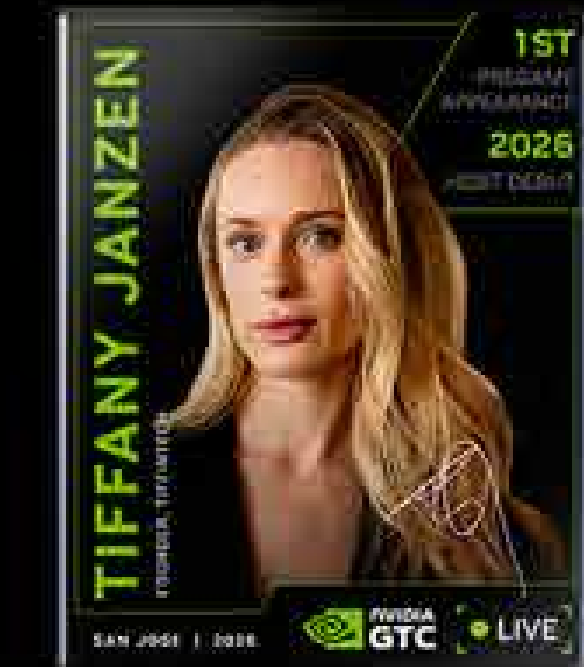
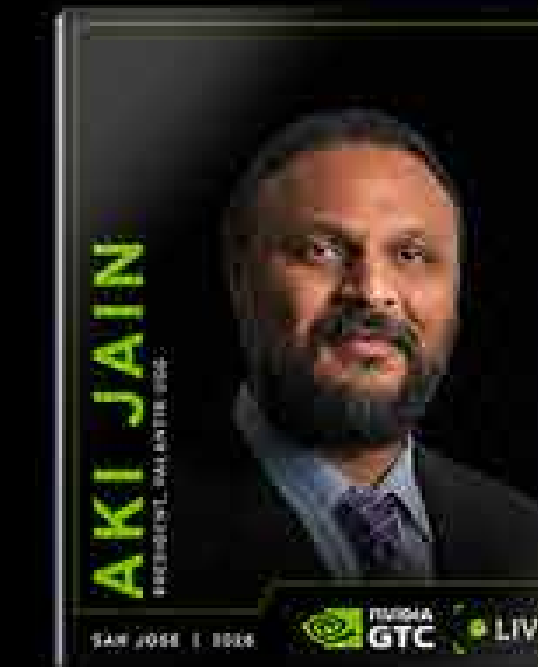
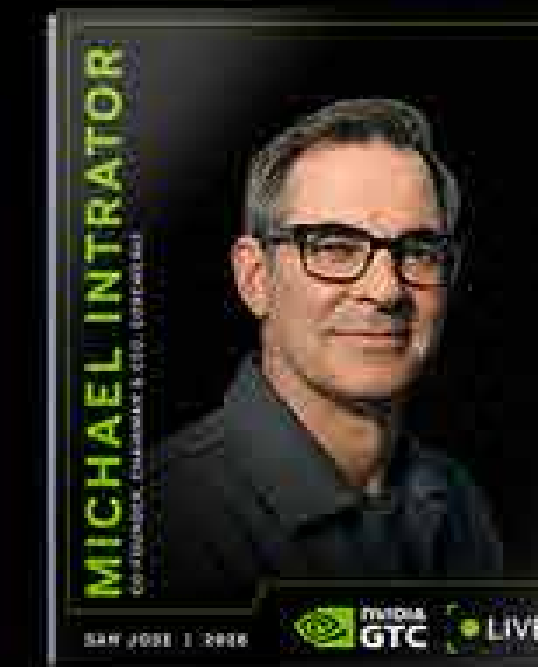
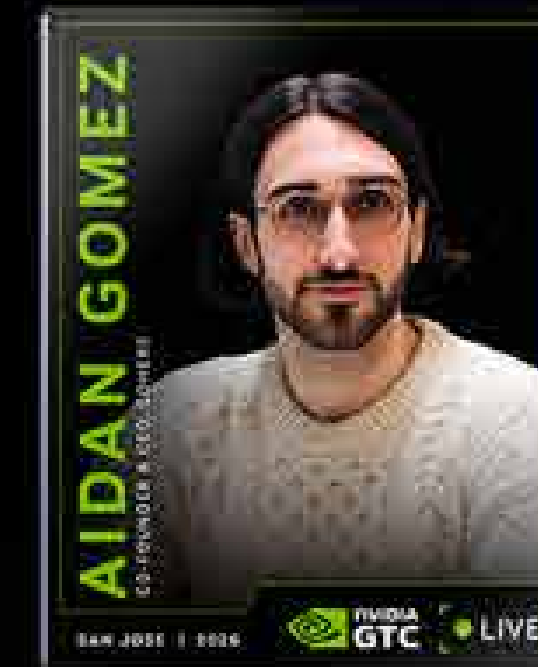
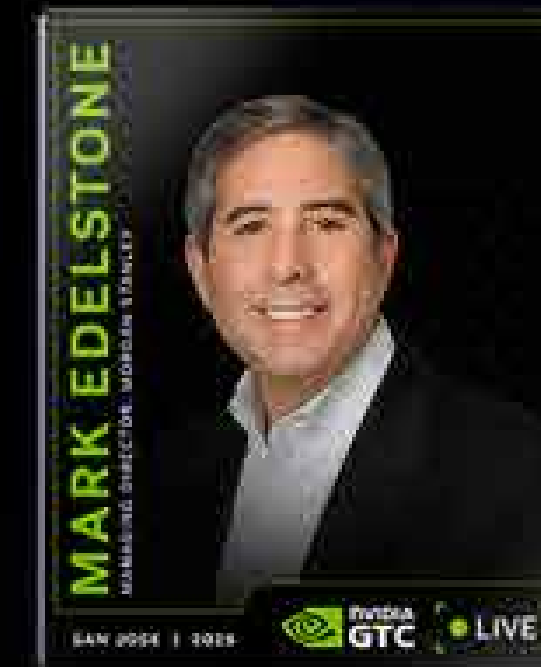
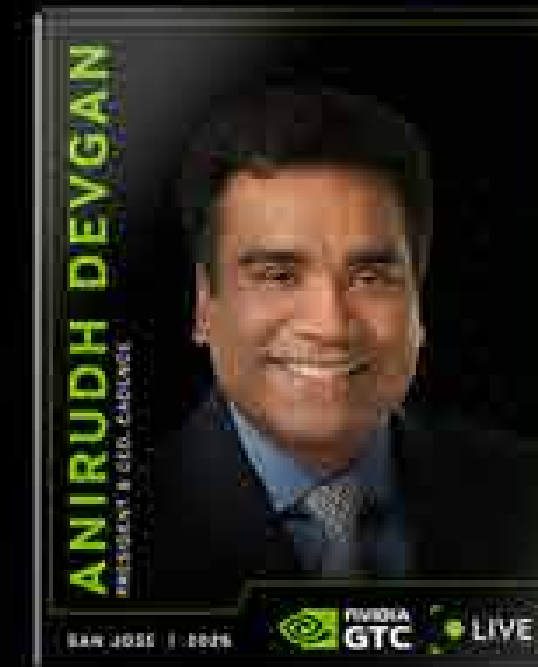
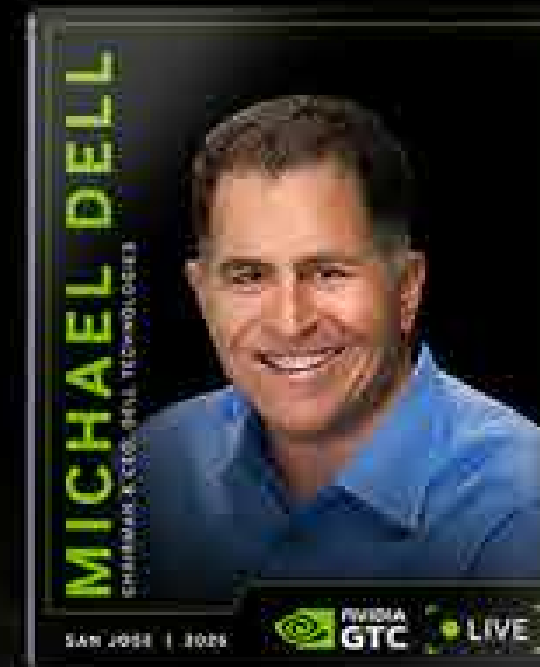
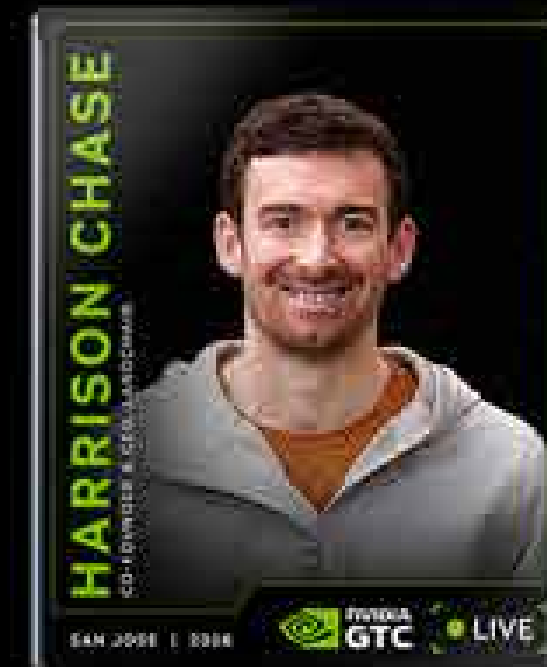
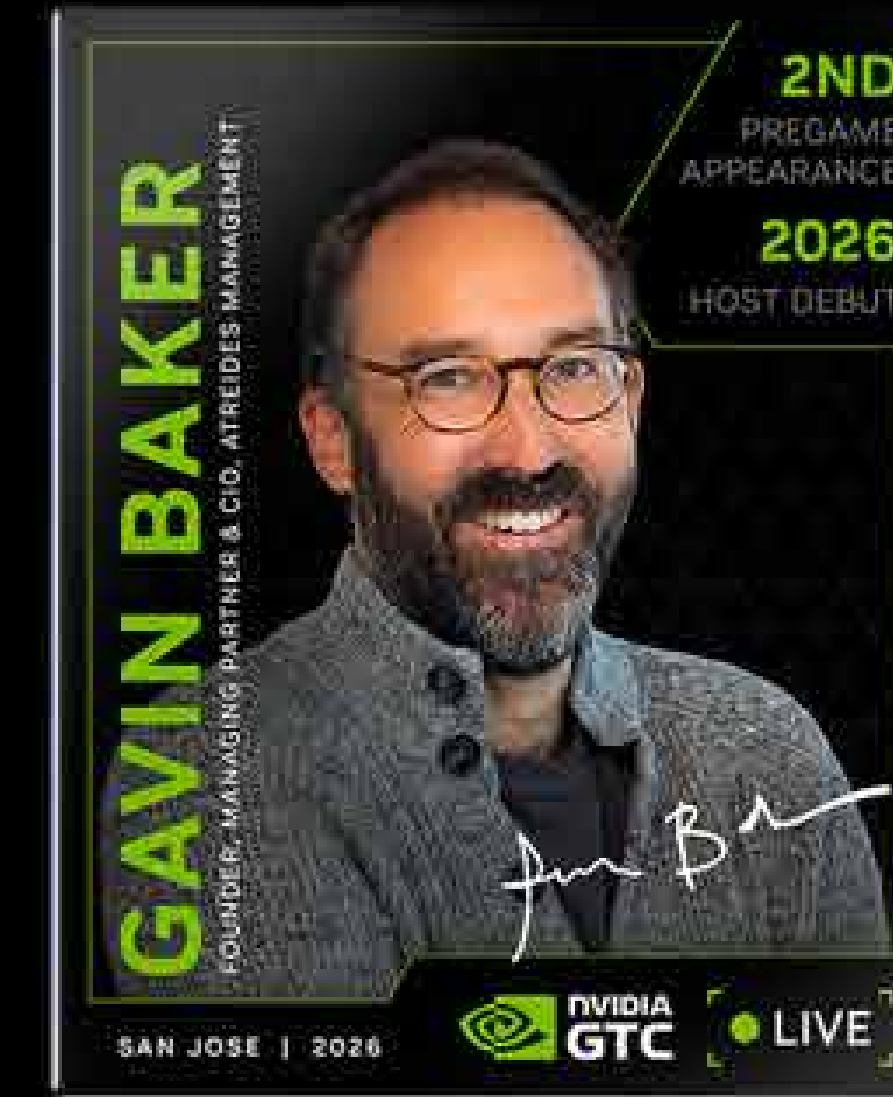
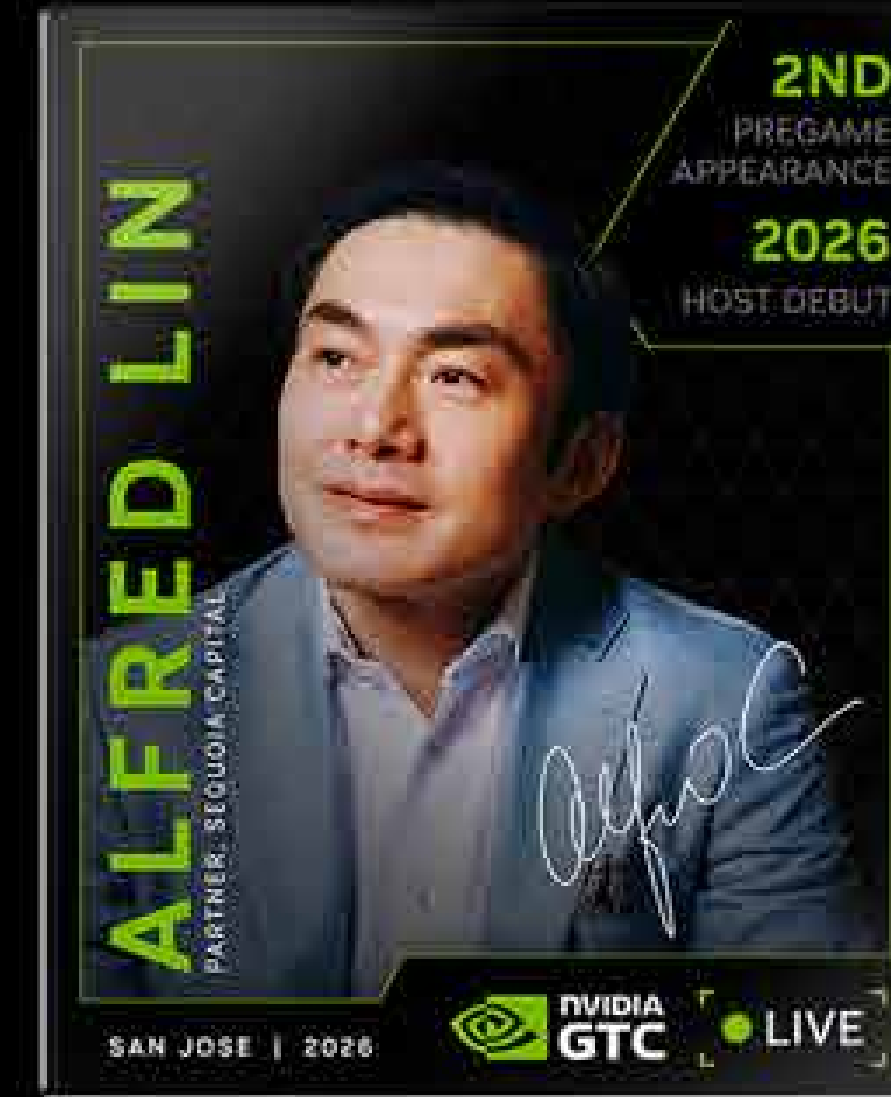






nVIDIA
GTC
2026







ABB

abbvie

accenture

Adobe

AETHERFLUX

AHEAD

Ai2

AiVRES

Akamai

AT&T

Alibaba Cloud

ALTIMETER

aws

AMP COALITION

ANTHROPAC

Apple

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

JLR

JOHNS HOPKINS UNIVERSITY

Johnson & Johnson

KAIST

KLA

L'ORÉAL

Lambda

LangChain

Lenovo

Lilly

LOCKHEED MARTIN

Los Alamos NATIONAL LABORATORY

MEDBOT 微创机器人

Medtronic

Mercedes

Meta

micron

Microsoft

Mistral AI

MIT

MITAC MITAC Computing

Morgan Stanley

msi

Nasdaq

NATIONAL TSING HUA UNIVERSITY

NEBIUS

NetApp

NISSAN MOTOR CORPORATION

NOKIA

NSCALE

NUTANIX

OpenAI

OpenClaw

OpenEvidence

ORACLE CLOUD Infrastructure

Palantir

paloalto NETWORKS

PayPal

PEGATRON

PEPSICO

perplexity

PG&E

PHYSICS X

Pinterest

planet.

PNY

PRIME Intellect

PRINCETON UNIVERSITY

ptc

QCT

RADIANT

RAFAY

Red Hat

REDWIRE

Reflection

runway

salesforce

SAMSUNG

SJSU SAN JOSE 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

VW

VULTR

waabi

Walmart

WAYMO

wiwynn x wlstron

WEKA

WELLS FARGO

World Labs

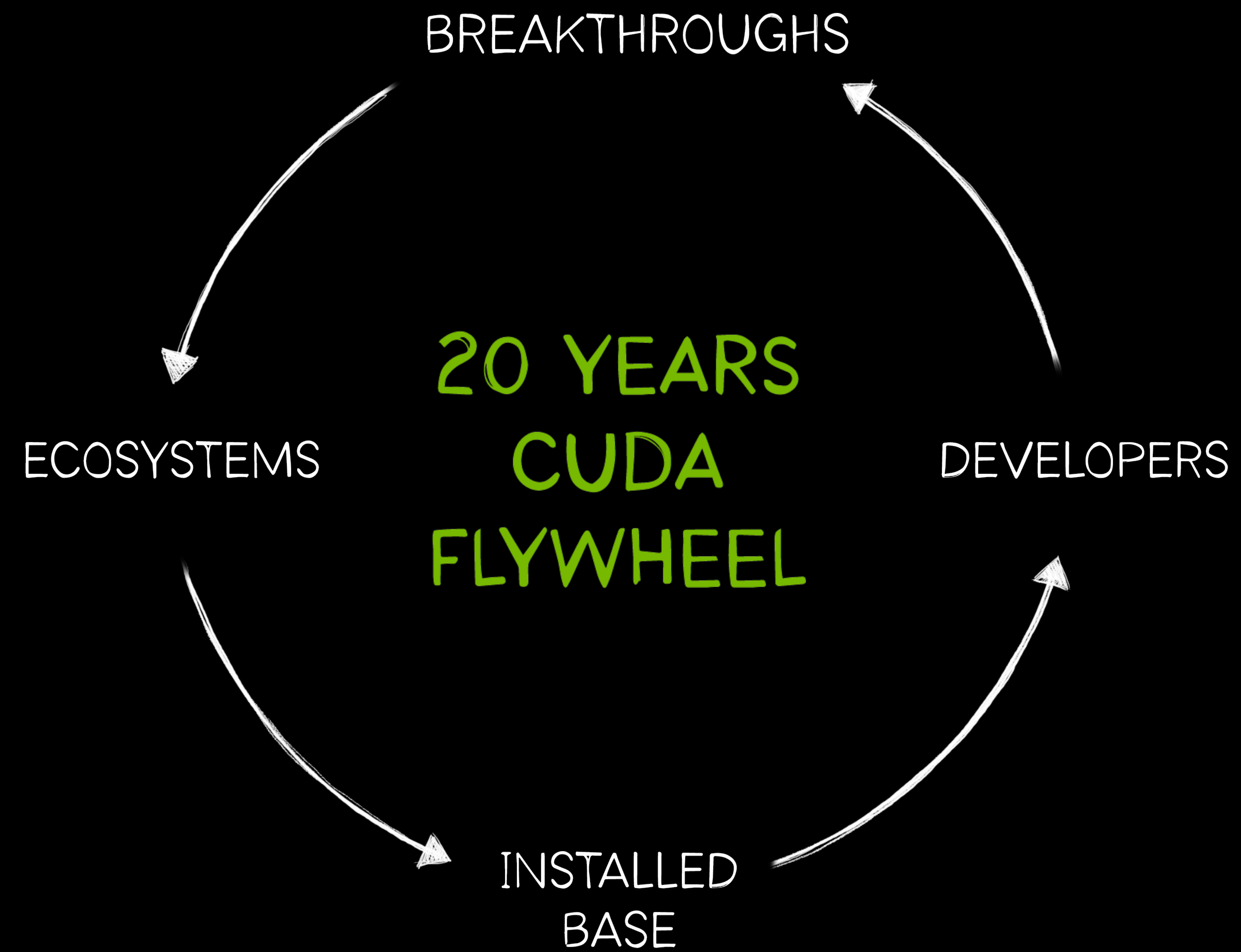
World Wide Technology

XI

xiaomi

YOTTA

Yum!





GeForce GTX 1080

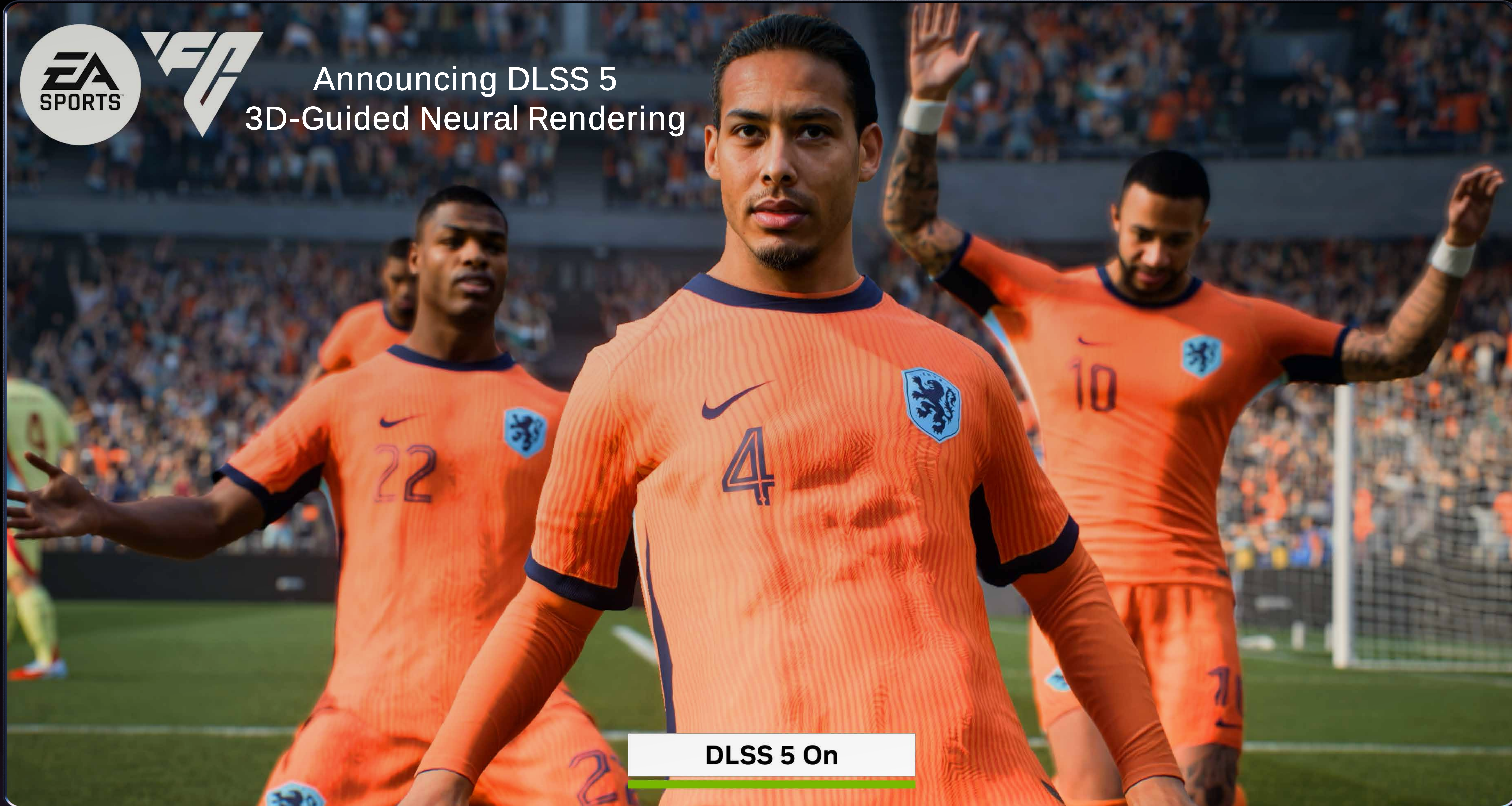
PORTKEY GAMES

HOGWARTS LEGACY

DLSS 5 On



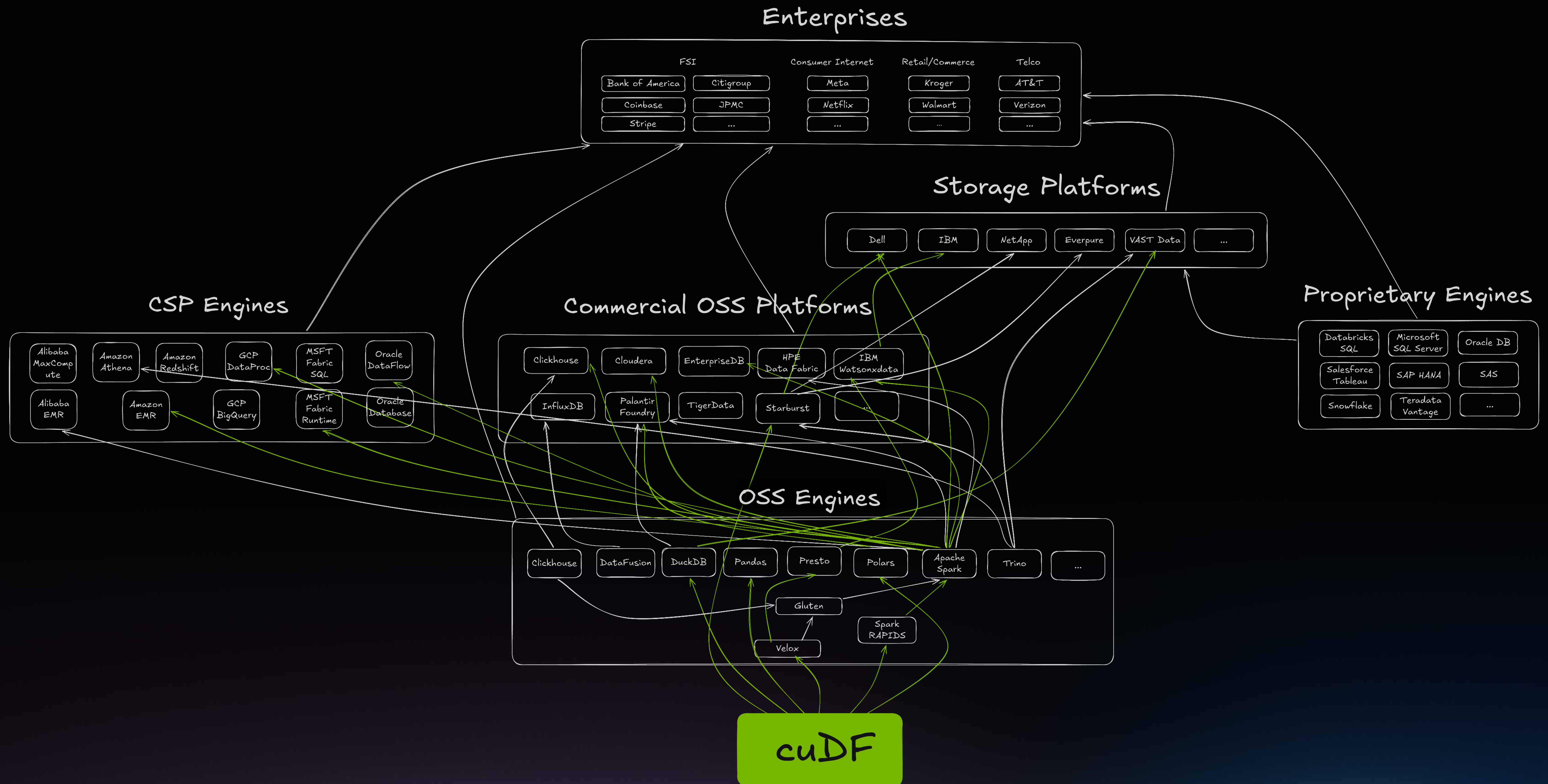
Announcing DLSS 5 3D-Guided Neural Rendering



DLSS 5 On

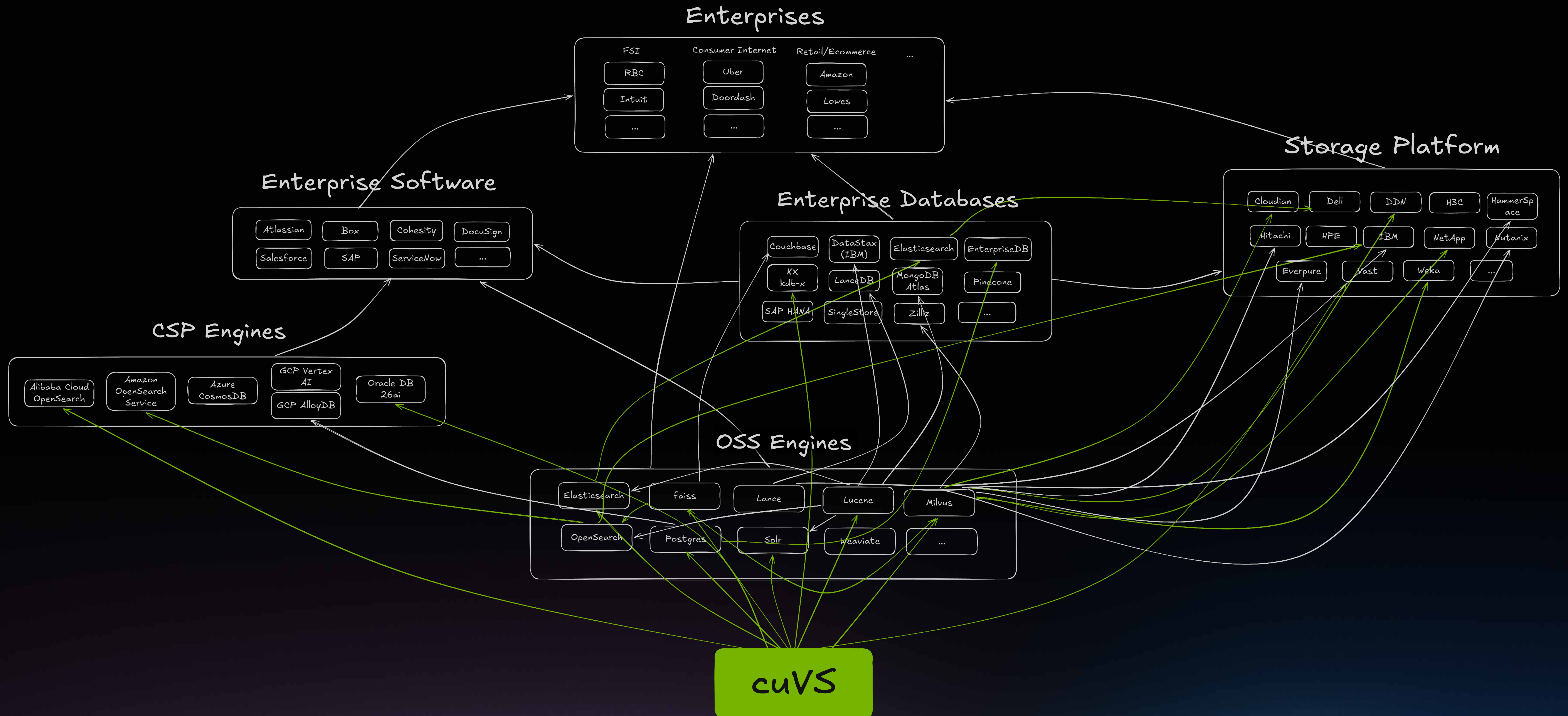
Structured Data is the Ground Truth of AI

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



Unstructured Data is the Context of AI

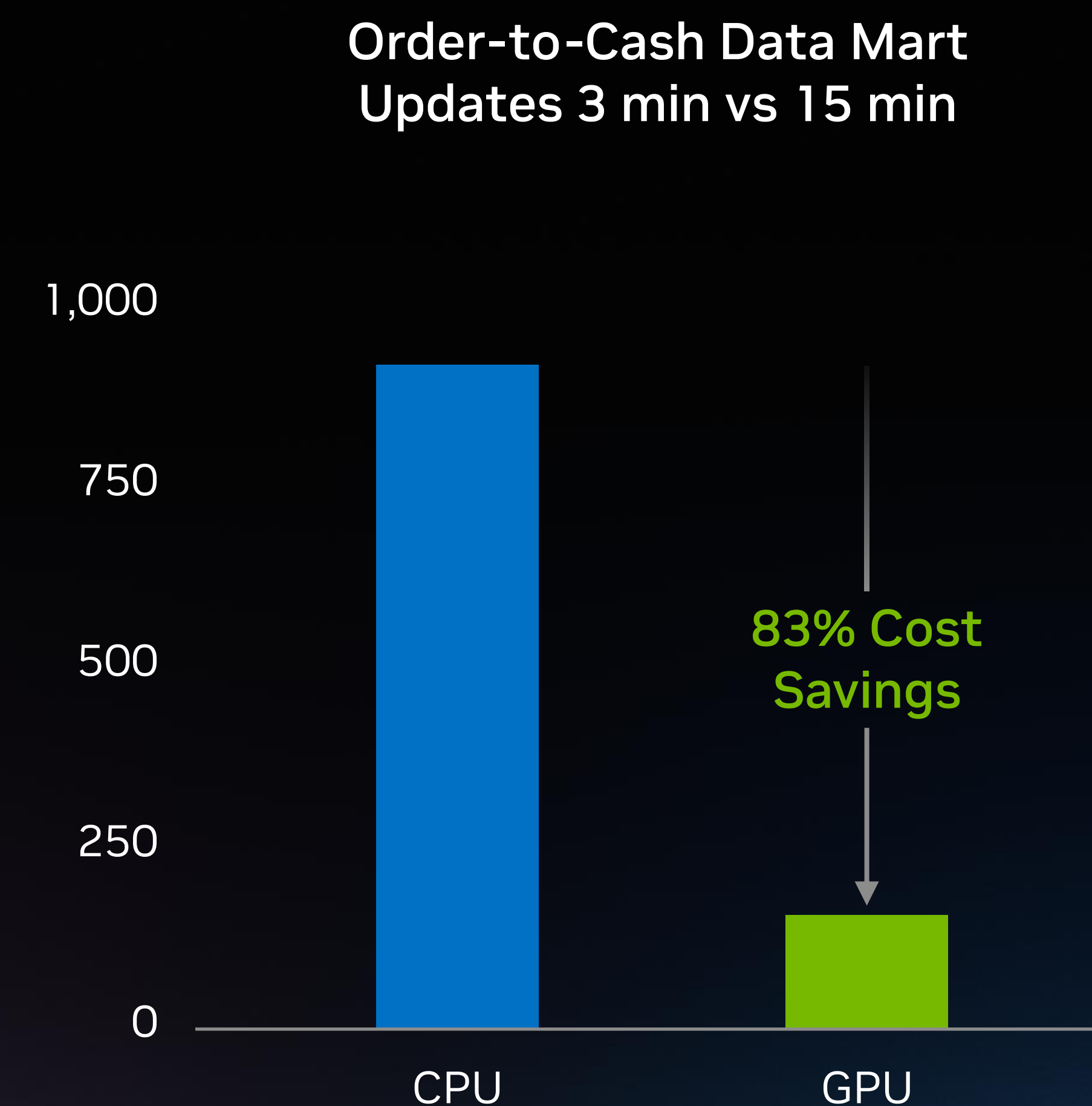
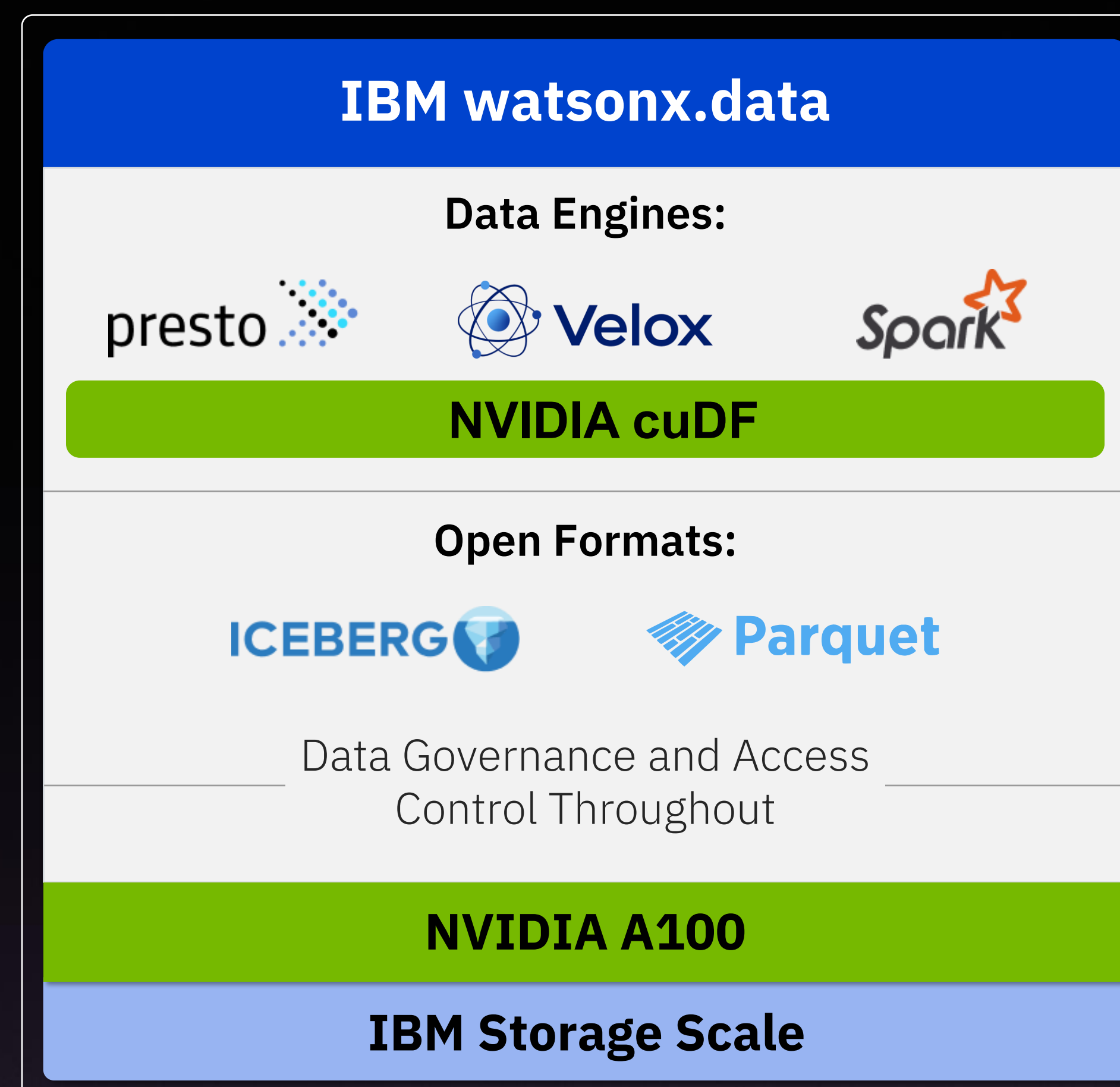
100's of Zettabytes Per Year of Unstructured Data – Growing Exponentially





*“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

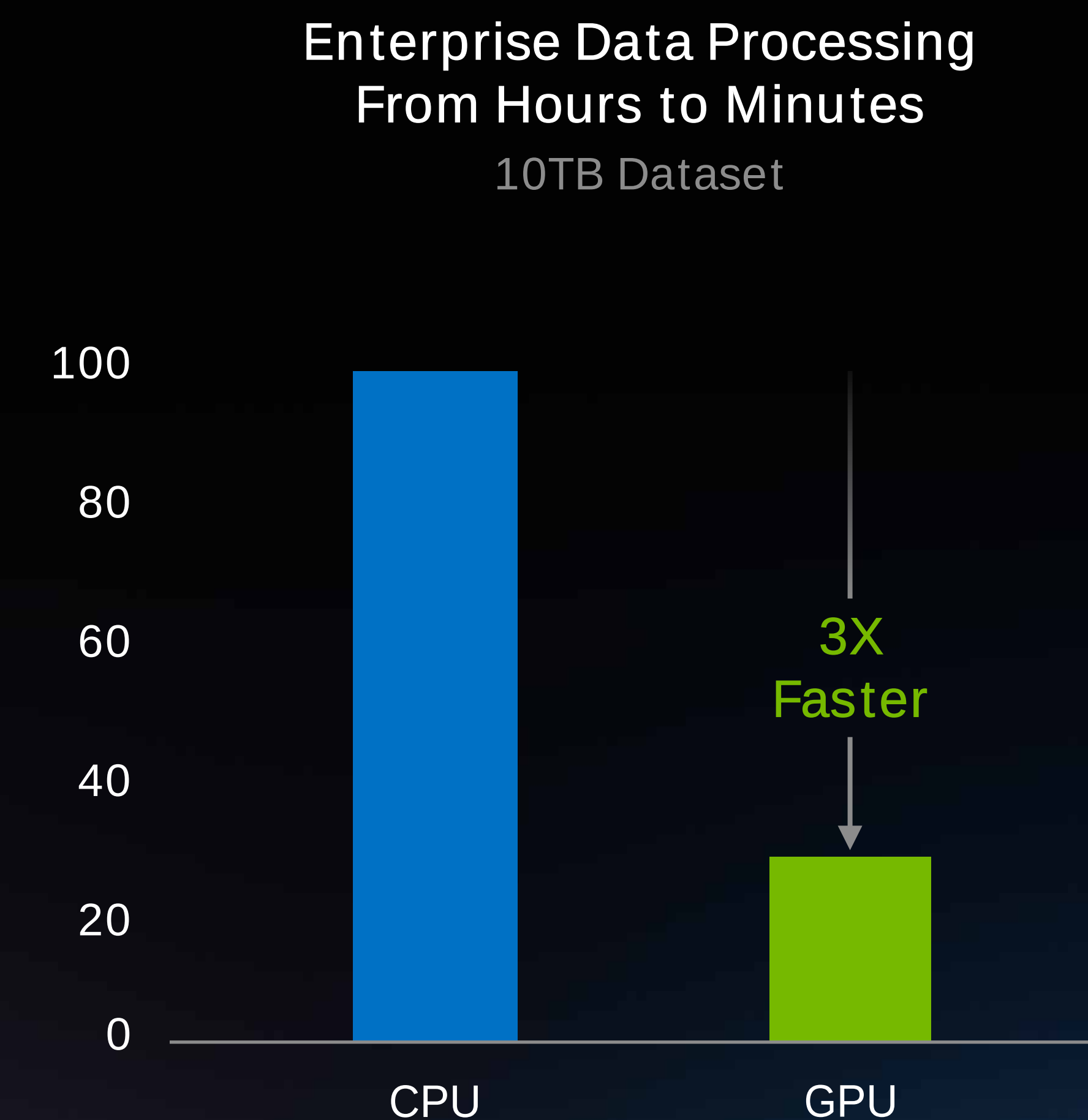
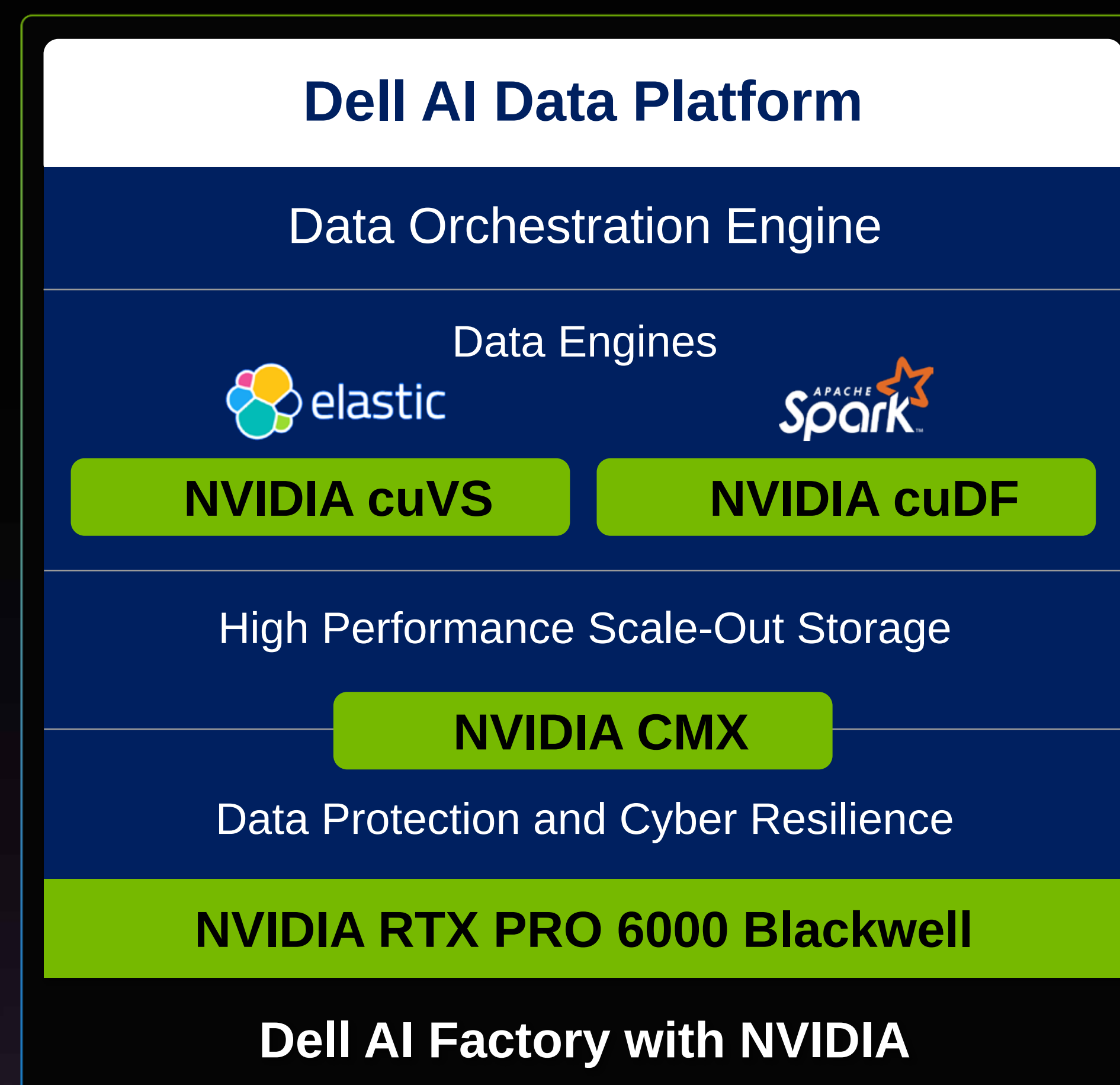






*“Dell AI Data Platform with NVIDIA helps us build scalable and repeatable data pipelines that drive automation—processing massive data volumes in **minutes instead of hours** and delivering transformative value for our clients.”*

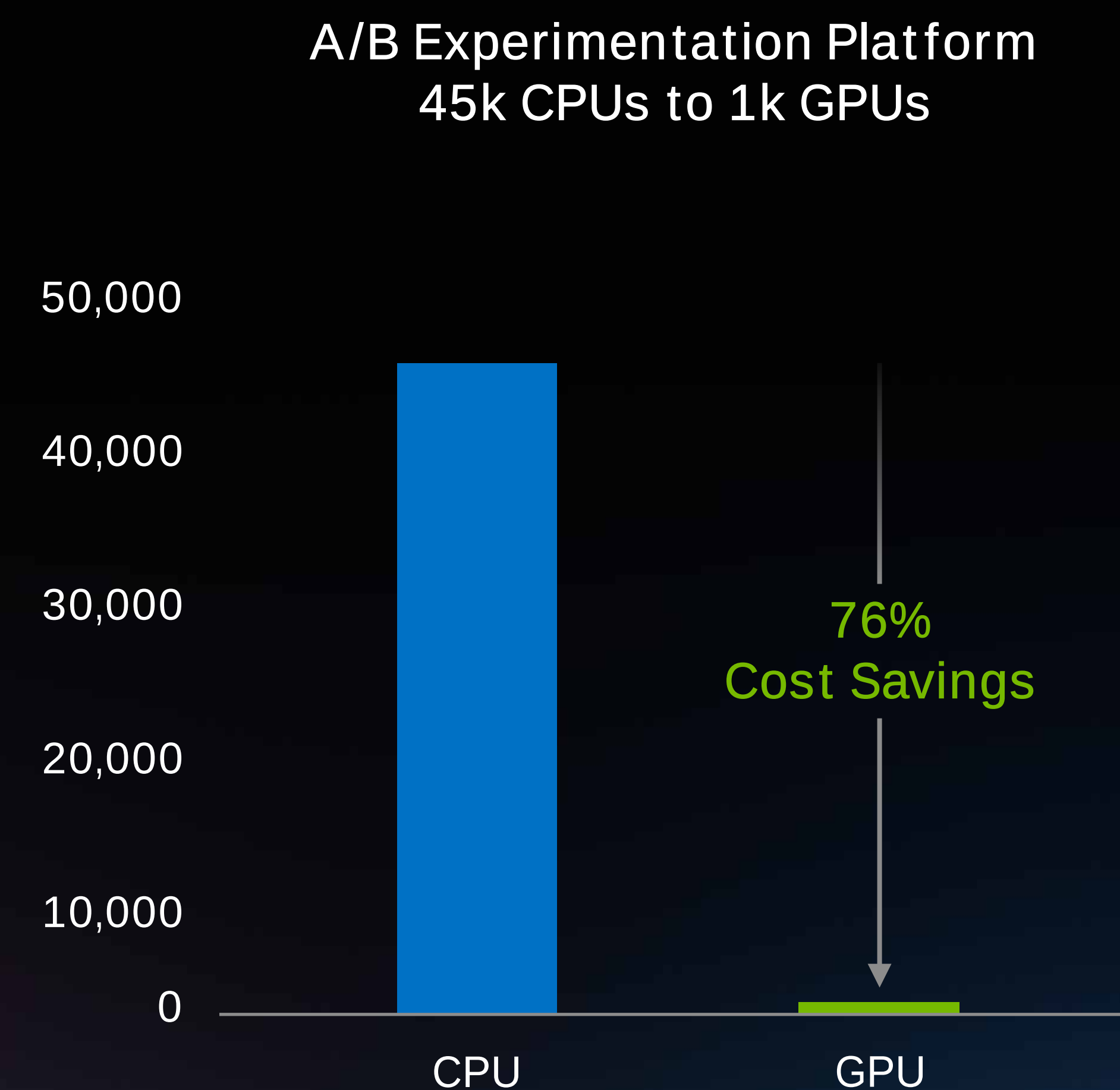
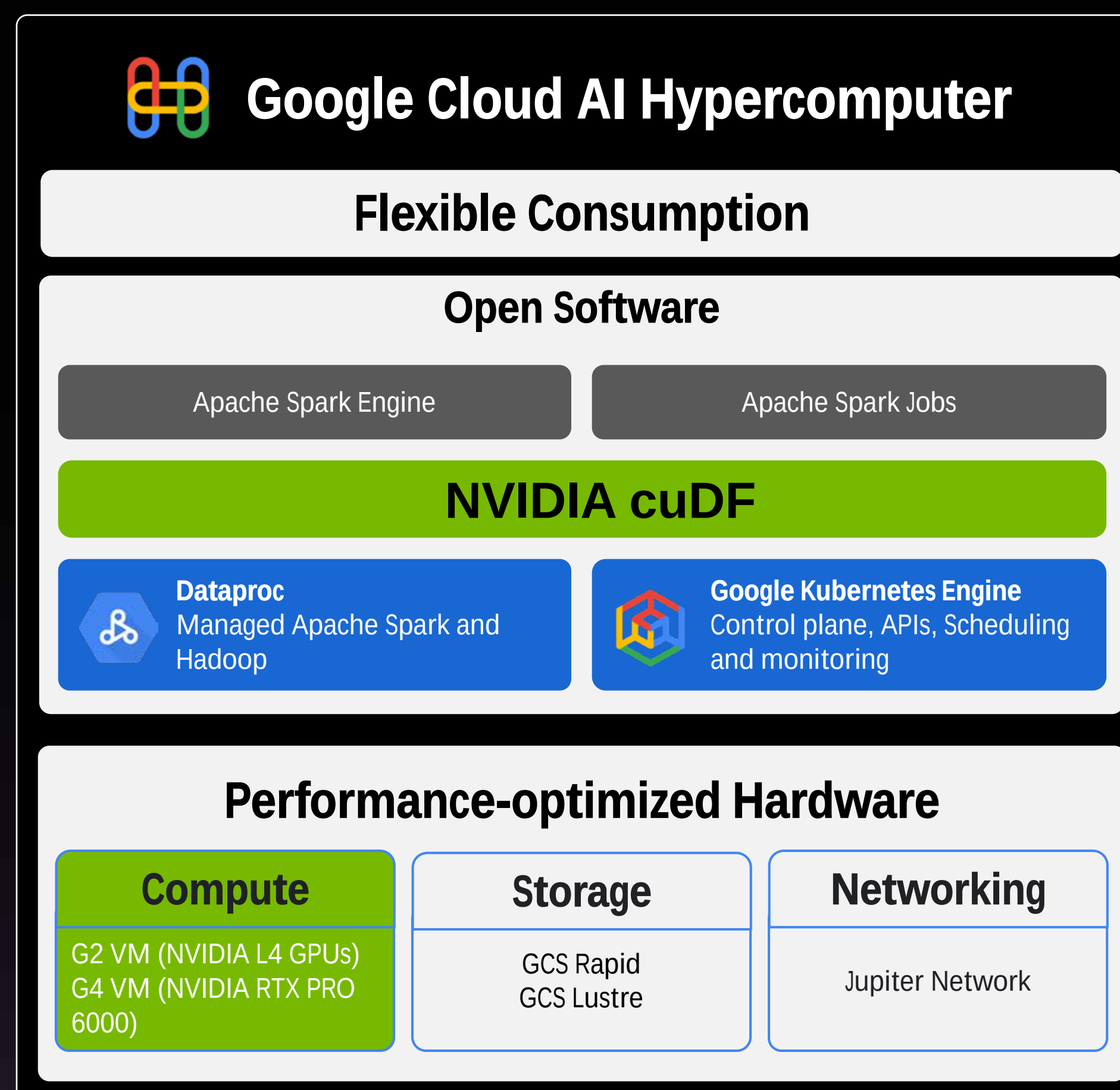
— Abhijit Dubey, CEO of NTT DATA





*“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





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



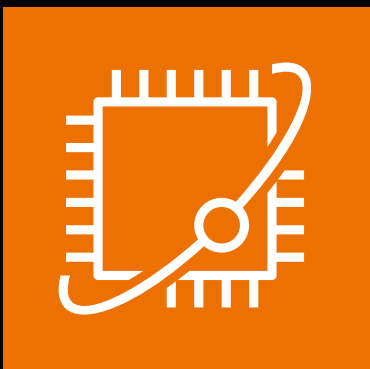
NVIDIA Rubin



NVIDIA ConnectX and BlueField Networking



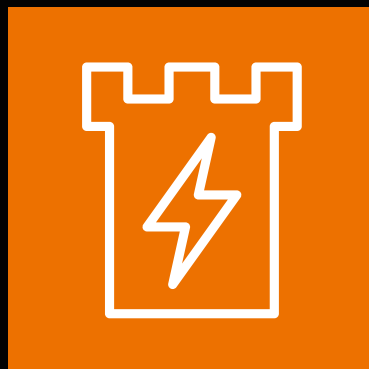
NVIDIA Quantum IB, Spectrum-X Networking



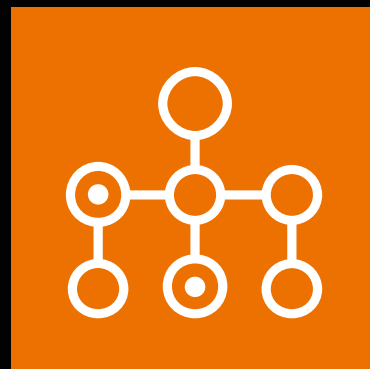
Amazon
Braket



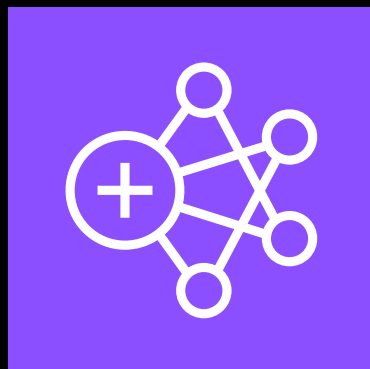
Amazon
EKS



AWS
Nitro Enclaves



AWS
Batch



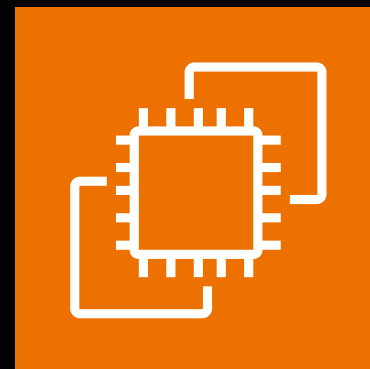
Amazon
EMR



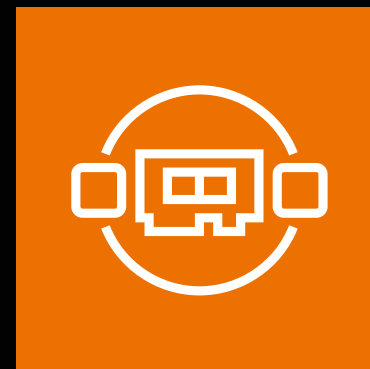
Amazon
Bedrock



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



ANTHROPIC

BlackRock



KRONES



SYNOPSYS



Microsoft Foundry



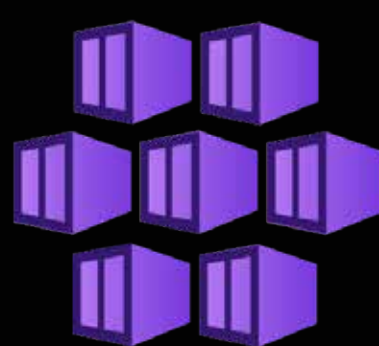
Windows 365



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



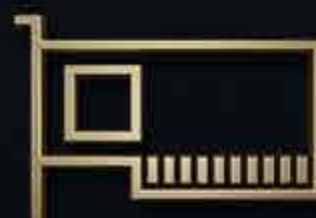
NVIDIA Blackwell



Confidential Computing



NVIDIA Rubin



NVIDIA ConnectX and BlueField Networking



NVIDIA Quantum IB, Spectrum-X Networking



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

AI

AI

AI

AI

AI

AI

AI

AI



Canva

cohere

CURSOR



Mistral AI

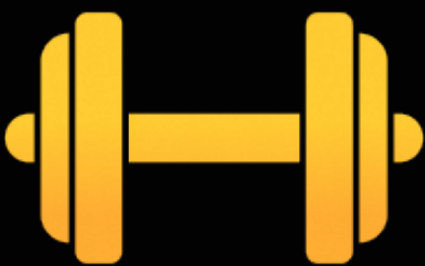
Morgan Stanley

OpenAI

runway



W&B
Models



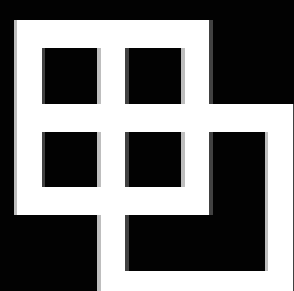
W&B
Training



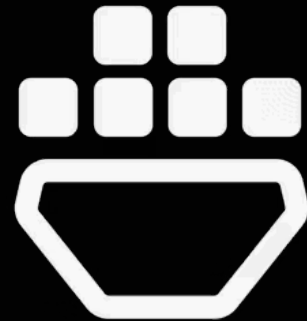
W&B
Inference



W&B
Weave



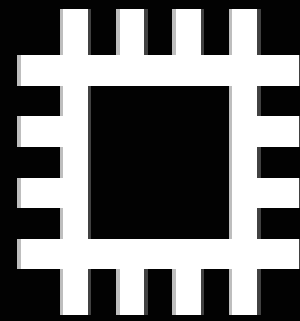
CoreWeave
Inference



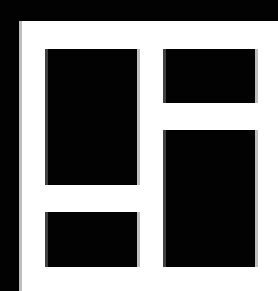
CoreWeave
Slurm on
Kubernetes



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



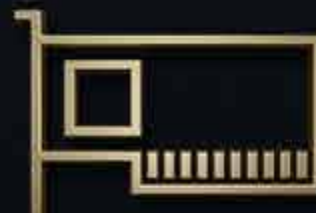
NVIDIA
Blackwell



Confidential
Computing



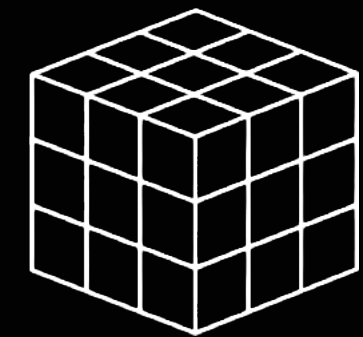
NVIDIA
Rubin



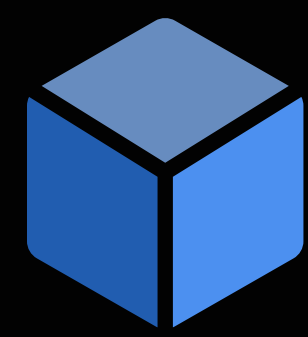
NVIDIA ConnectX
and BlueField Networking



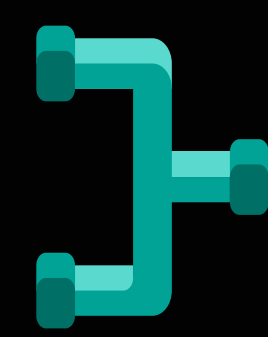
NVIDIA Quantum IB,
Spectrum-X Networking



Rubix



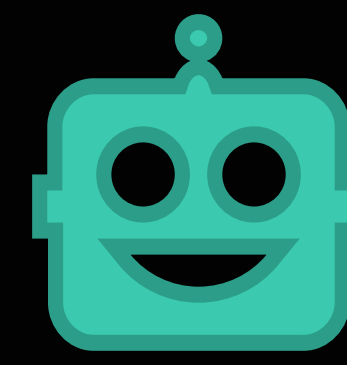
Ontology



Pipeline Builder



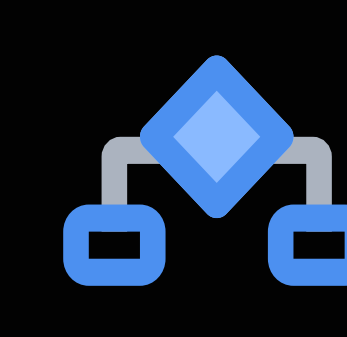
Fusion



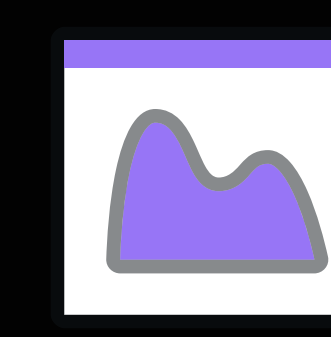
Agent Studio



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



AUTOMOTIVE



FINANCIAL
SERVICES



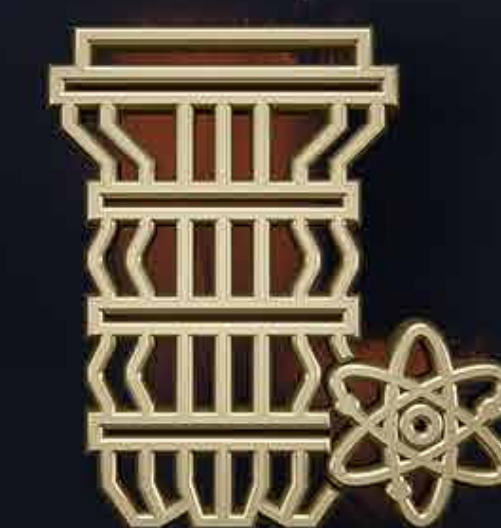
HEALTHCARE AND
LIFE SCIENCES



INDUSTRIAL



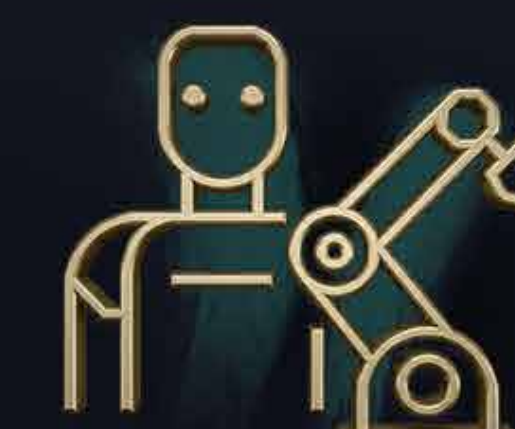
MEDIA AND
ENTERTAINMENT



QUANTUM



RETAIL
AND CPG



ROBOTICS



TELCO



 AI Natives

Aurora

momenta

nuro

pony

waabi

 WAYMO

 WAYVE

AI for Auto

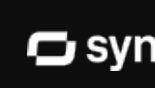
Decagon

IIElevenLabs

Fin

kore.ai

 SIERRA

 synthesia

AI for Customer Support

BeyondMath

Emmi AI

 NEURAL CONCEPT

 luminary

 PASTEUR ISI

 periodic labs

 PHYSICS X

 RECURSIVE INTELLIGENCE

AI for Engineering

Arc Institute

bijohub

Boltz

 Chai Discovery

 Edison

 Isomorphic Labs

L I L A

 INSTITUTE FOR Protein Design

AI for Healthcare

AGIBOT

AGILE ROBOTS

FieldAI

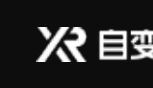
 FIGURE

 Generalist

Physical Intelligence (π)

 rai Robotics and AI Institute

 SKILD AI

 变量

AI for Robotics

brave

glean

 perplexity

 you.com

AI for Search

CodeRabbit

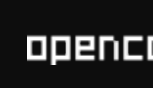
Cognition

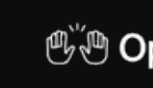
CURSOR

 FACTORY

 greptile

Lovable

 opencode

 OpenHands

 qodo

 tabnine

AI for Software Development

PyTorch



DL Frameworks

 A2A

AutoGen

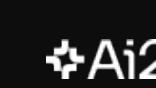
 crewai

 LangChain

 Model Context Protocol

 OpenCLAW

Agent Frameworks/Protocols

 A12



ANTHROPIC

 Black Forest Labs

cohere

Decart

 deepseek

 Gemma

humans&

 inception

 ineffable intelligence

KIMI

Magic

 Midjourney

 Mistral AI

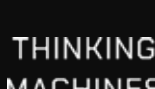
OpenAI

 poolside

 Qwen

 Reflection

runway

 THINKING MACHINES

 X

Frontier Model Builders

Anyscale

baseten

BigPanda

Braintrust



DataRobot

 Distyl

 Fireworks AI

H2O.ai

 LogicMonitor

 Modal

NUTANIX

 OpenRouter

 Red Hat

 replicate

scale

 snowflake

 together.ai

Model to Production

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

BLACKWELL + RUBIN

\$1T+
STILL GROWING

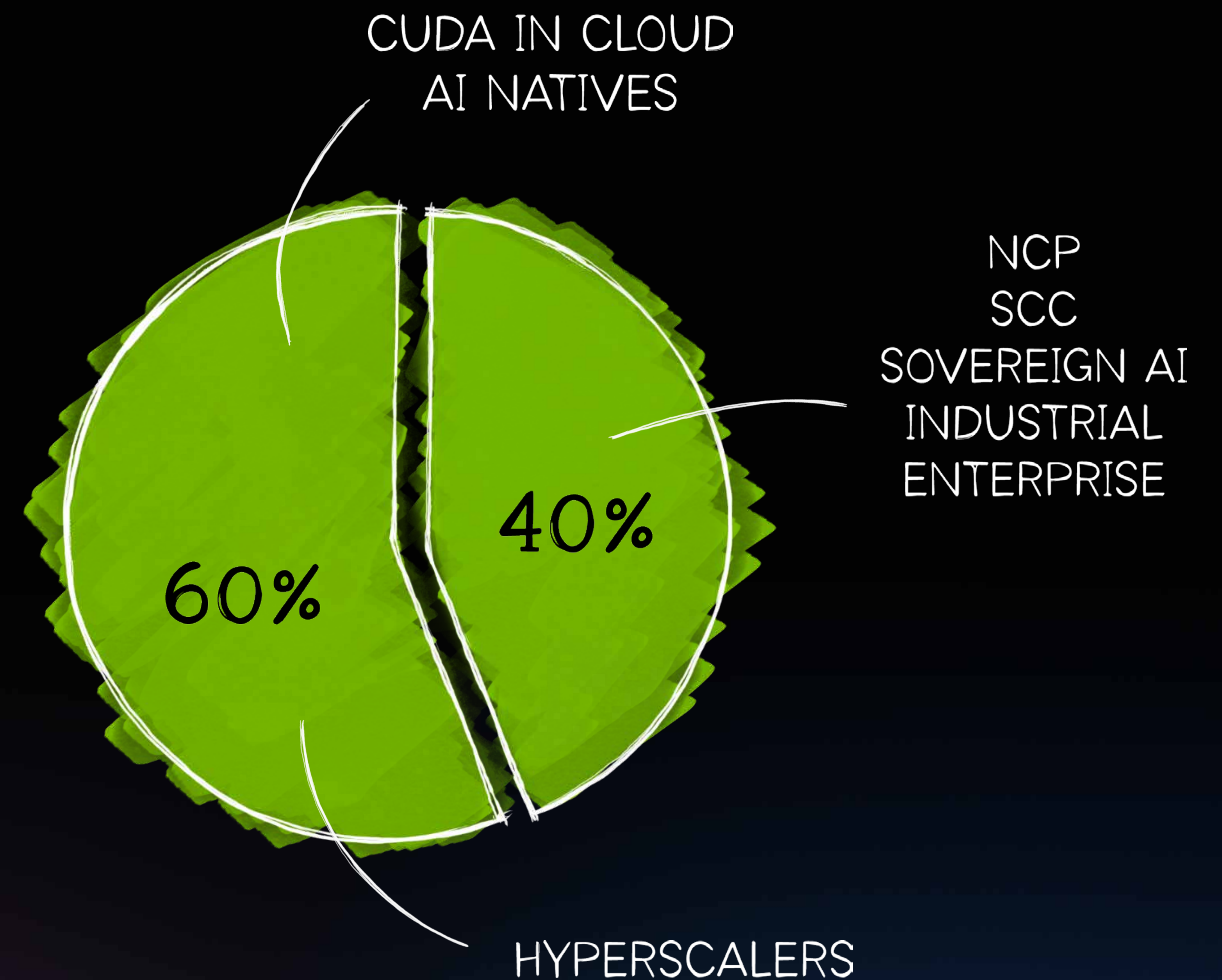
\$0.5T

GTC DC 2025
2025-2026

GTC SJ 2026
2025-2027

ANTHROPIC
METASL
MULTIPLE OSS
XAI
GEMINI
OPENAI

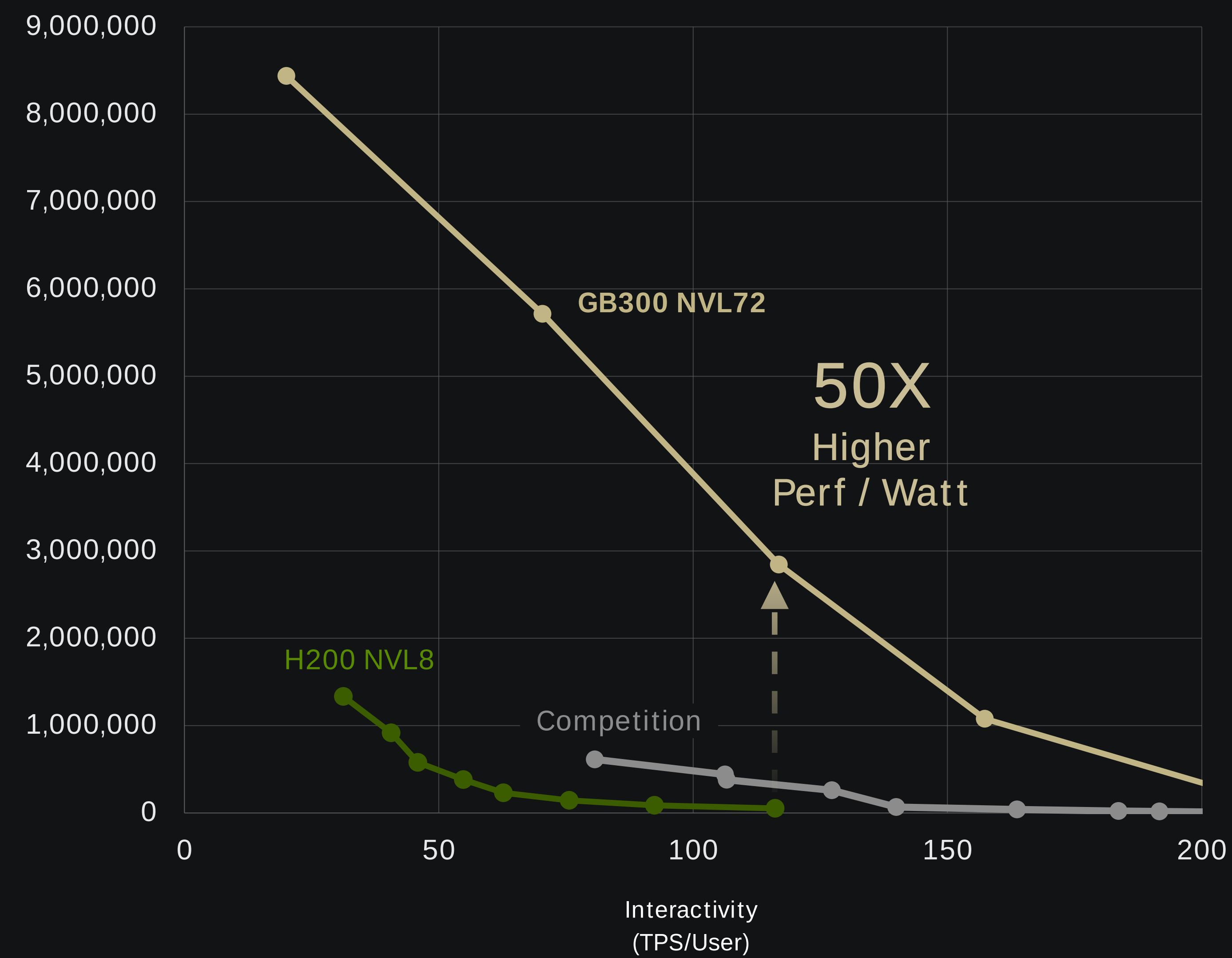
NEW SINCE
GTC DC 2025



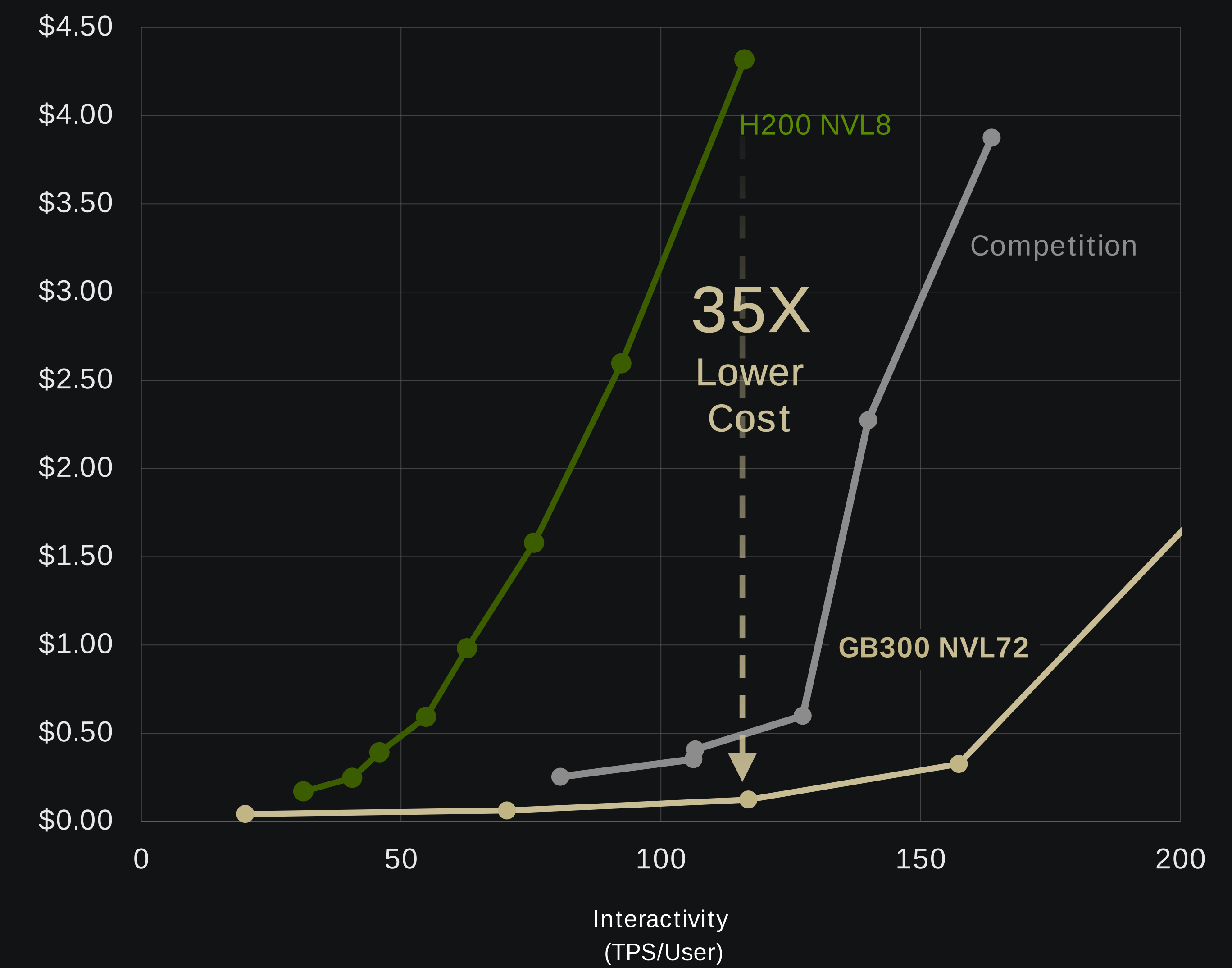
NVIDIA Extreme Co-Design Revolutionized Token Cost

“GB NVL72 Inference King”

Tokens per Watt Drives Factory Revenue



Performance Drives Token Cost

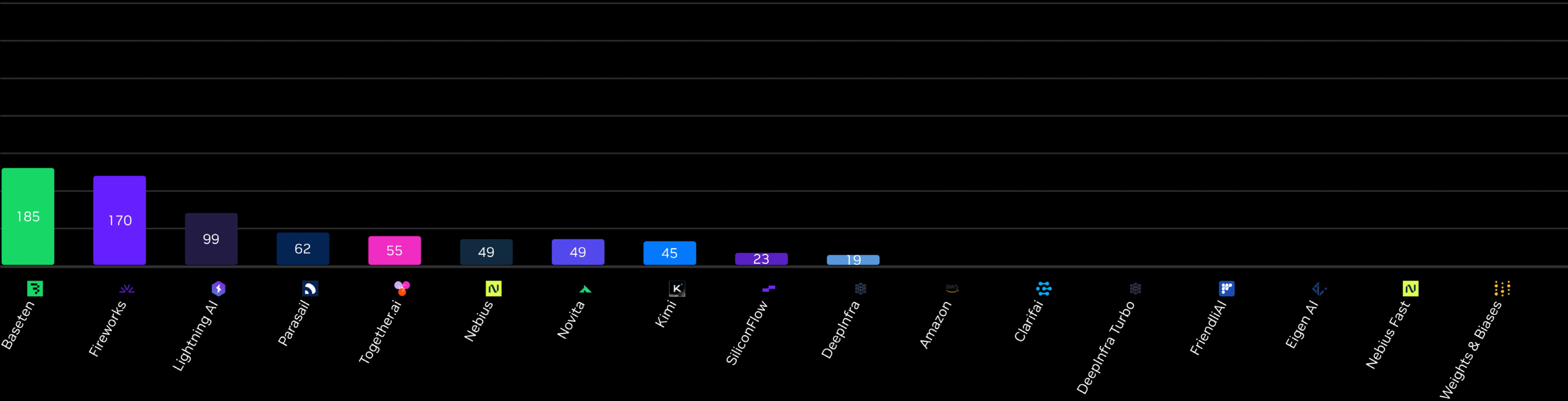


NVIDIA is the Global Standard for AI Inference at Scale

All Leading Inference End-Points Run on NVIDIA

Output Speed - March 3, 2026

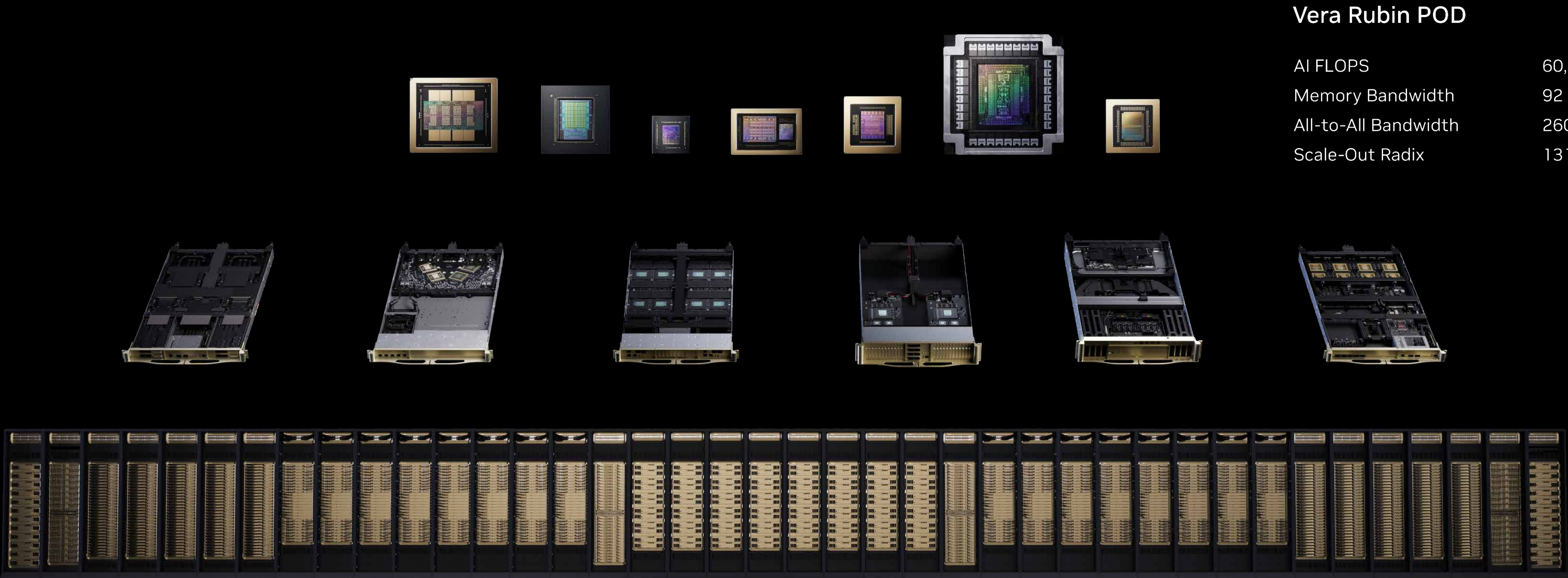
Output Tokens per Second; Higher is better



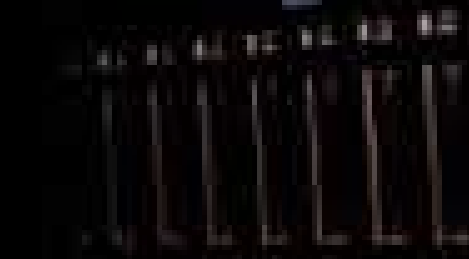
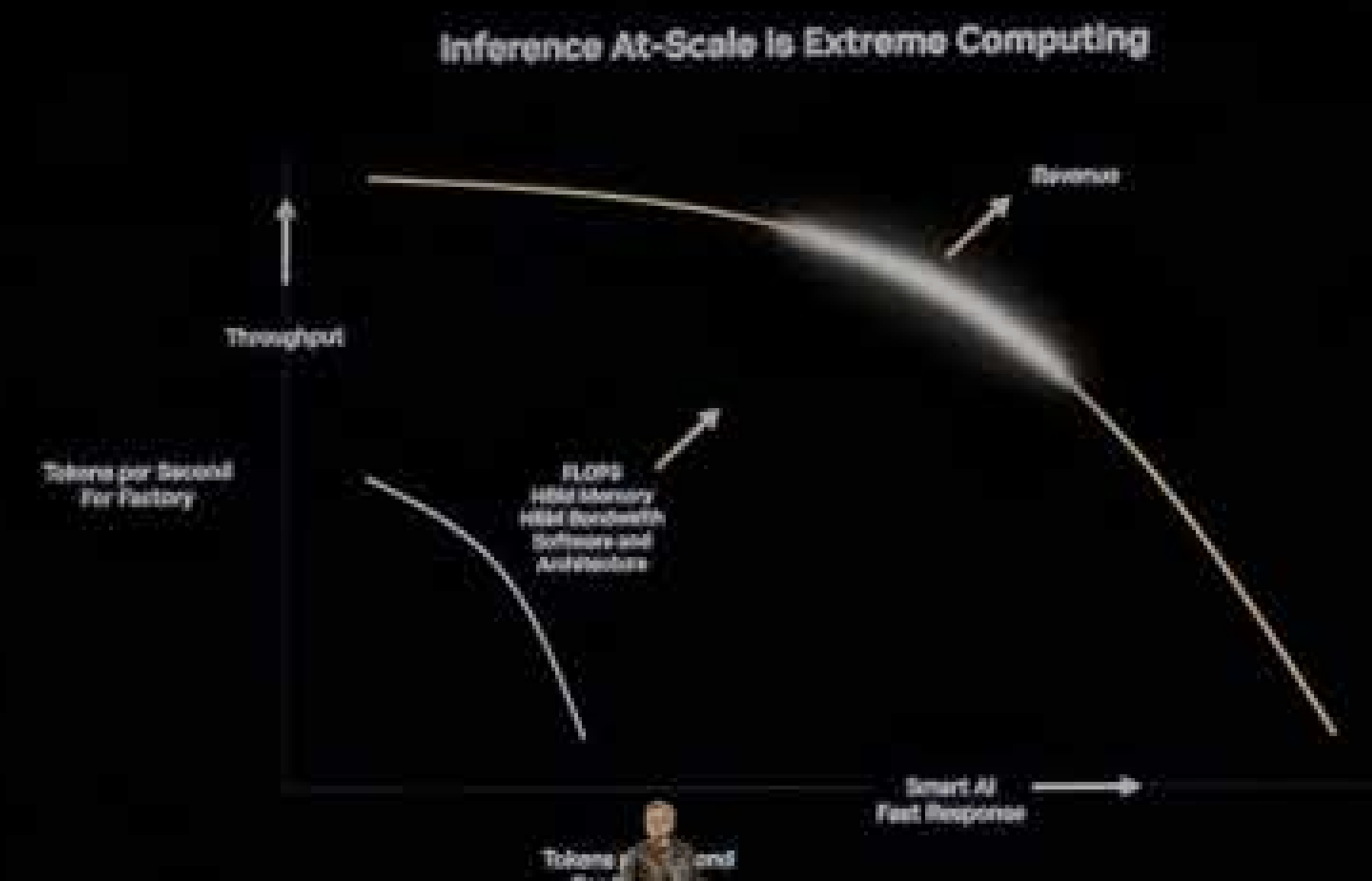
Kimi K2.5 Reasoning

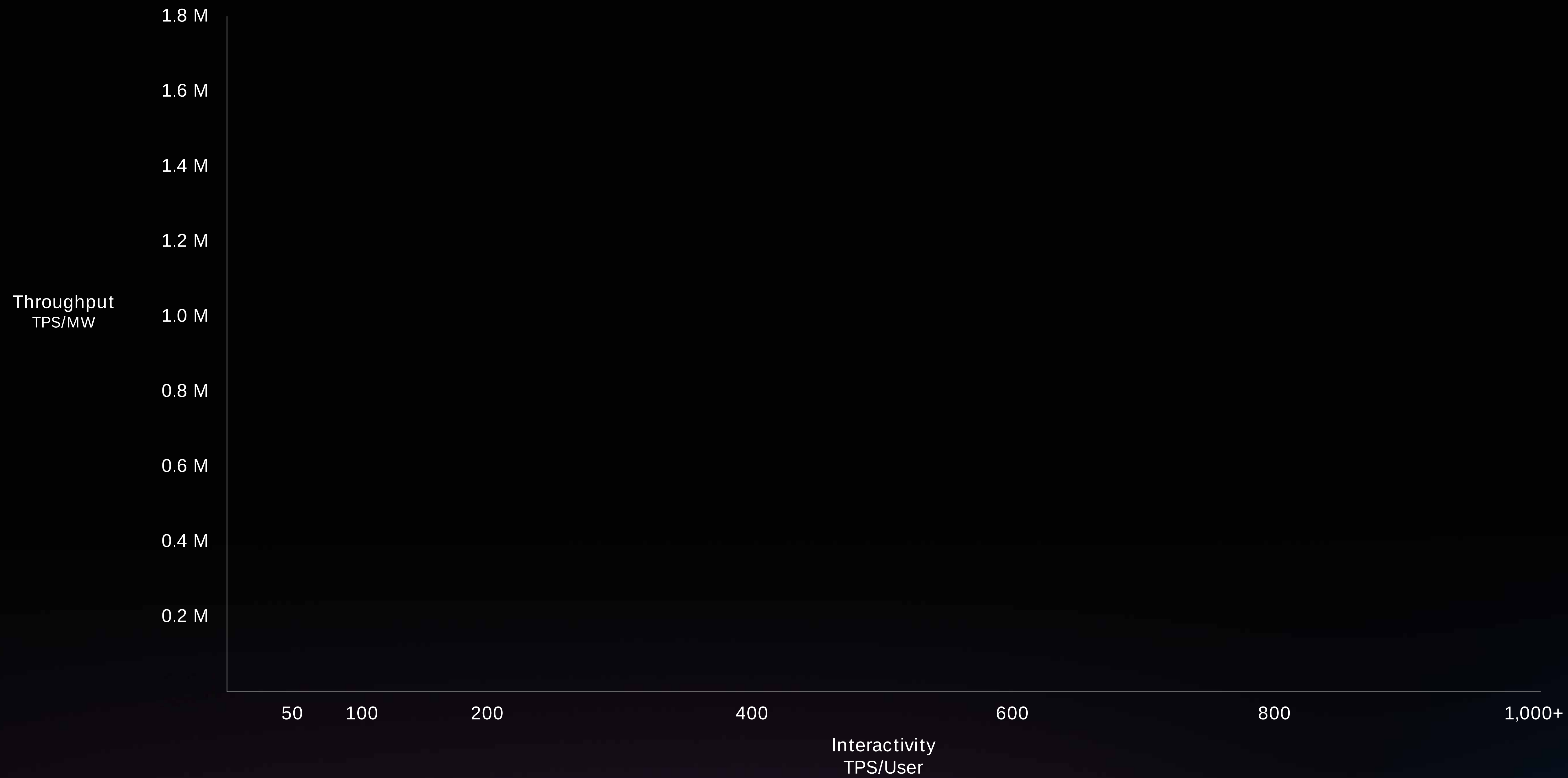
AI Factories are the Industrial Infrastructure of the AI Era
Inference is the Workload
Tokens are the New Commodity
Compute is Revenue





Inference Performance and Efficiency Drive Company Results





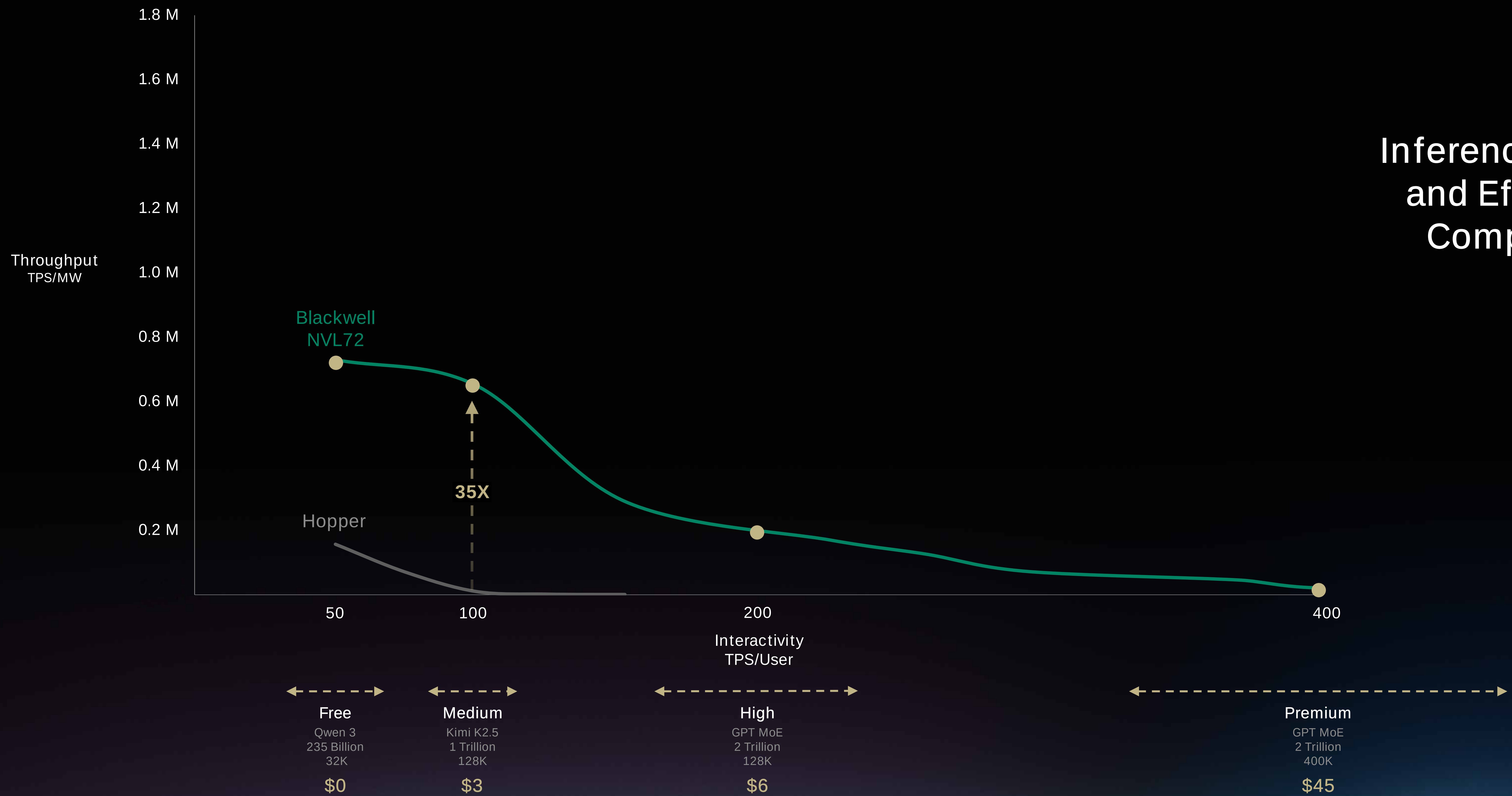
Inference Performance and Efficiency Drive Company Results



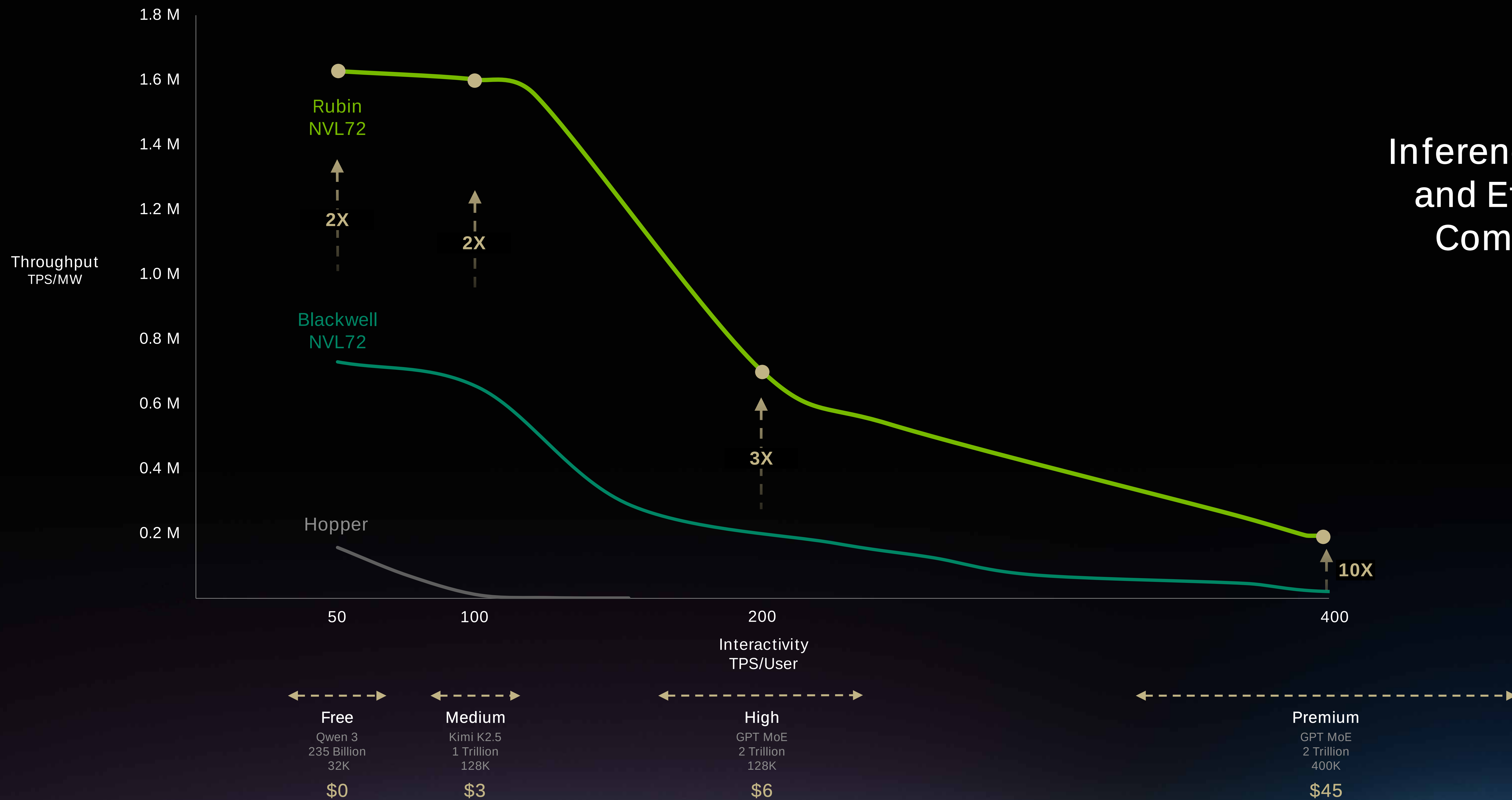
Inference Performance
and Efficiency Drive
Company Results



Inference Performance and Efficiency Drive Company Results



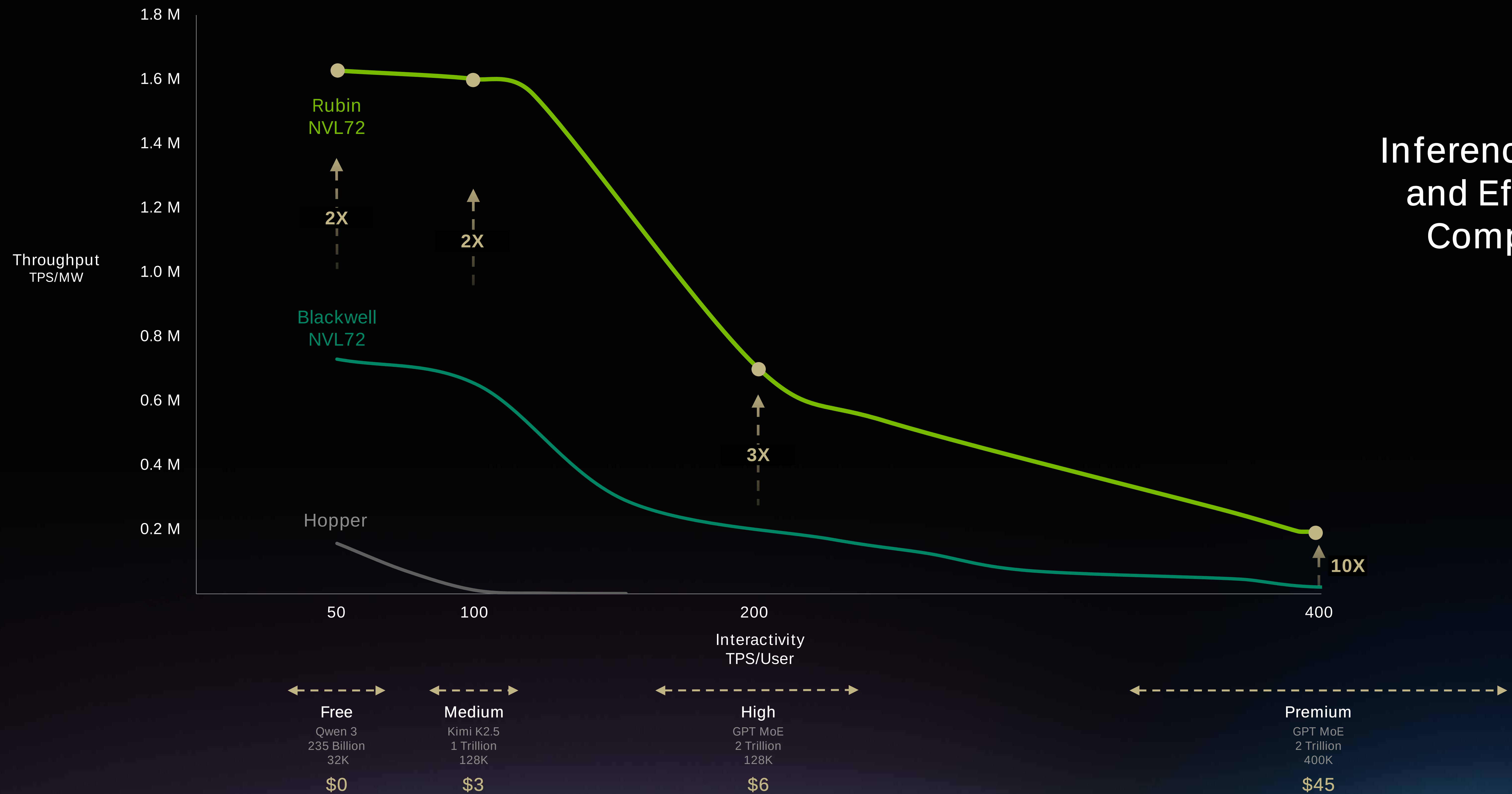
Inference Performance and Efficiency Drive Company Results



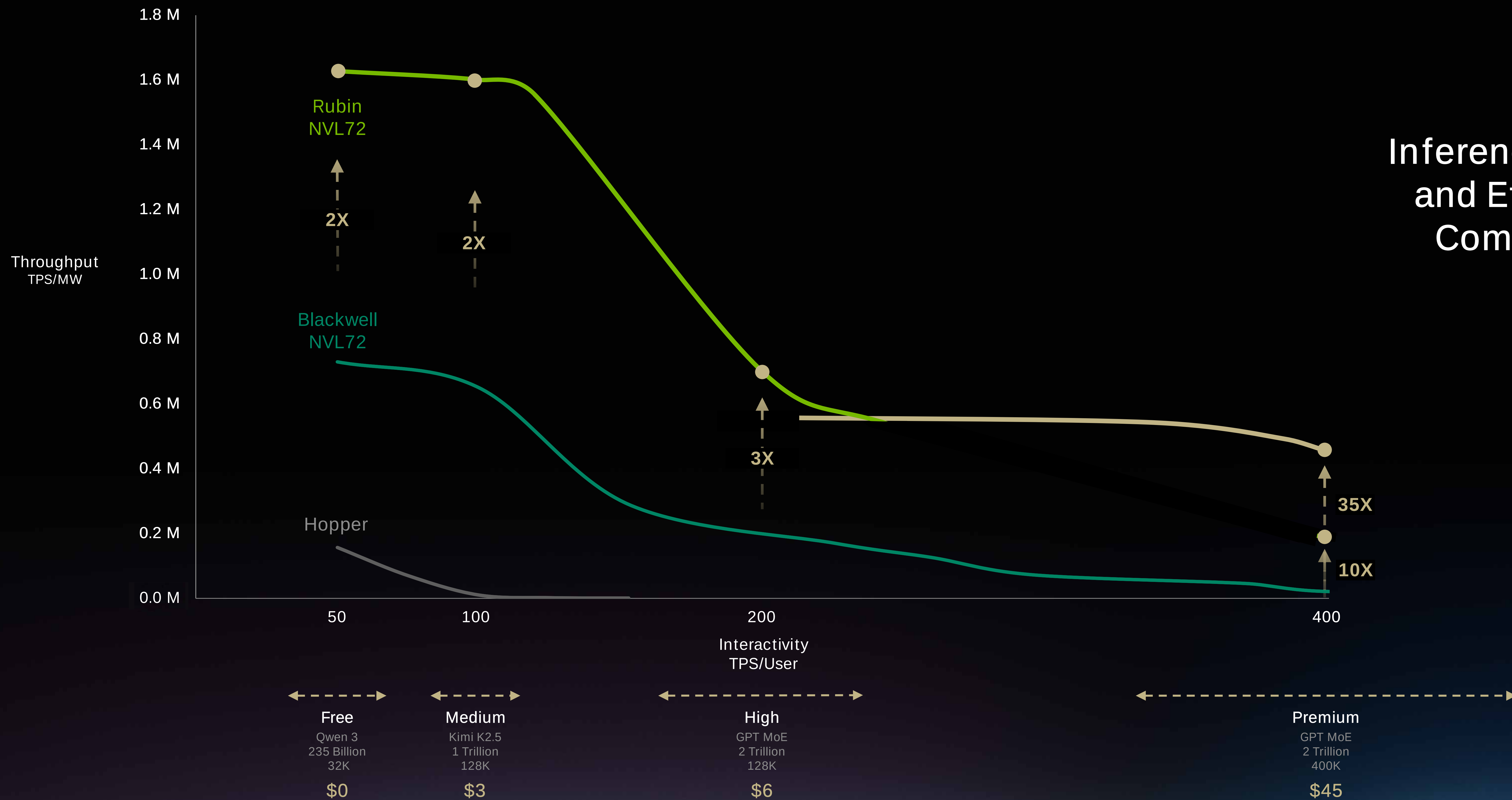
Inference Performance
and Efficiency Drive
Company Results

NVIDIA Vera Rubin Unlocks \$150B Revenue Opportunity

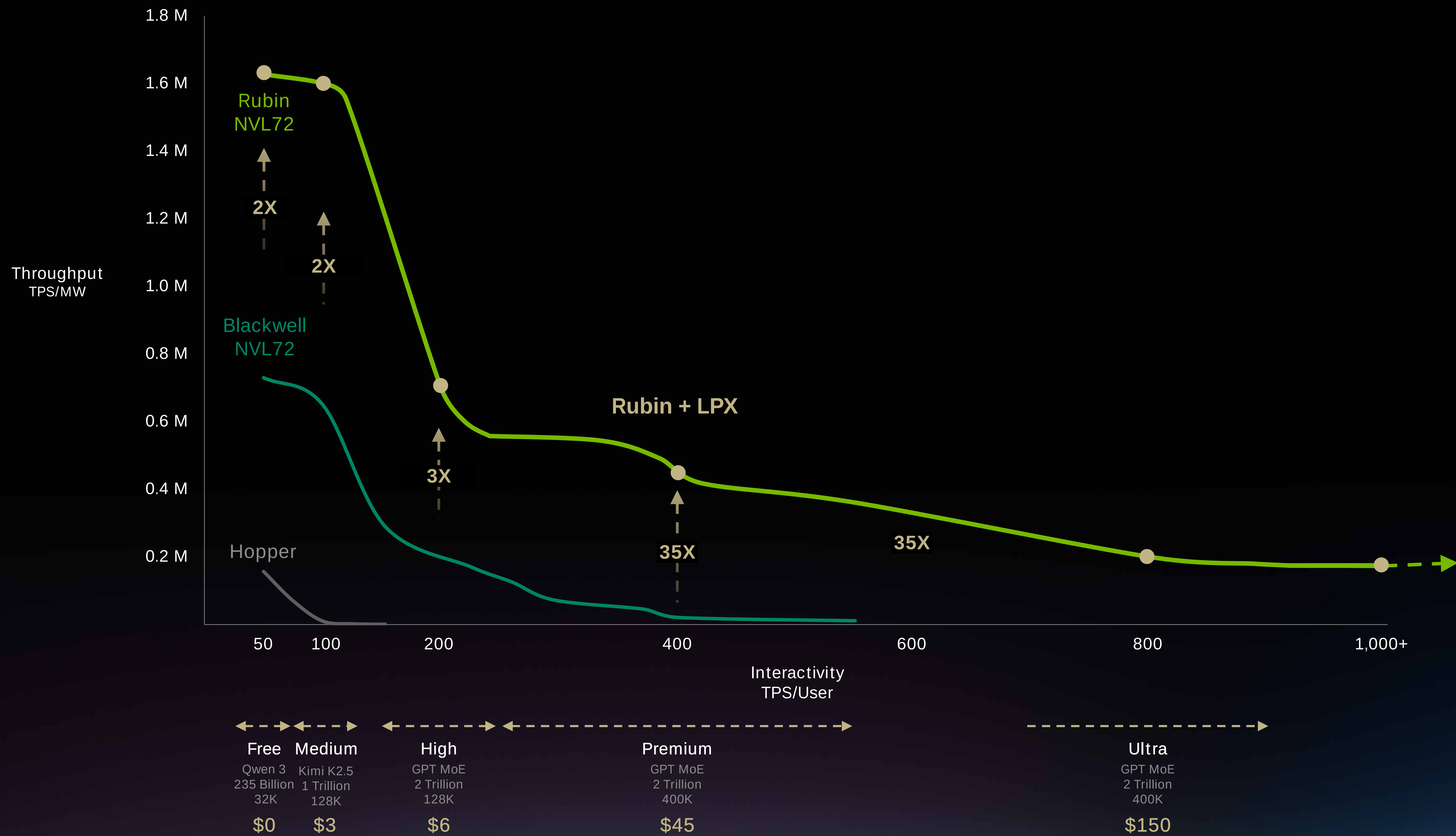




Inference Performance
and Efficiency Drive
Company Results

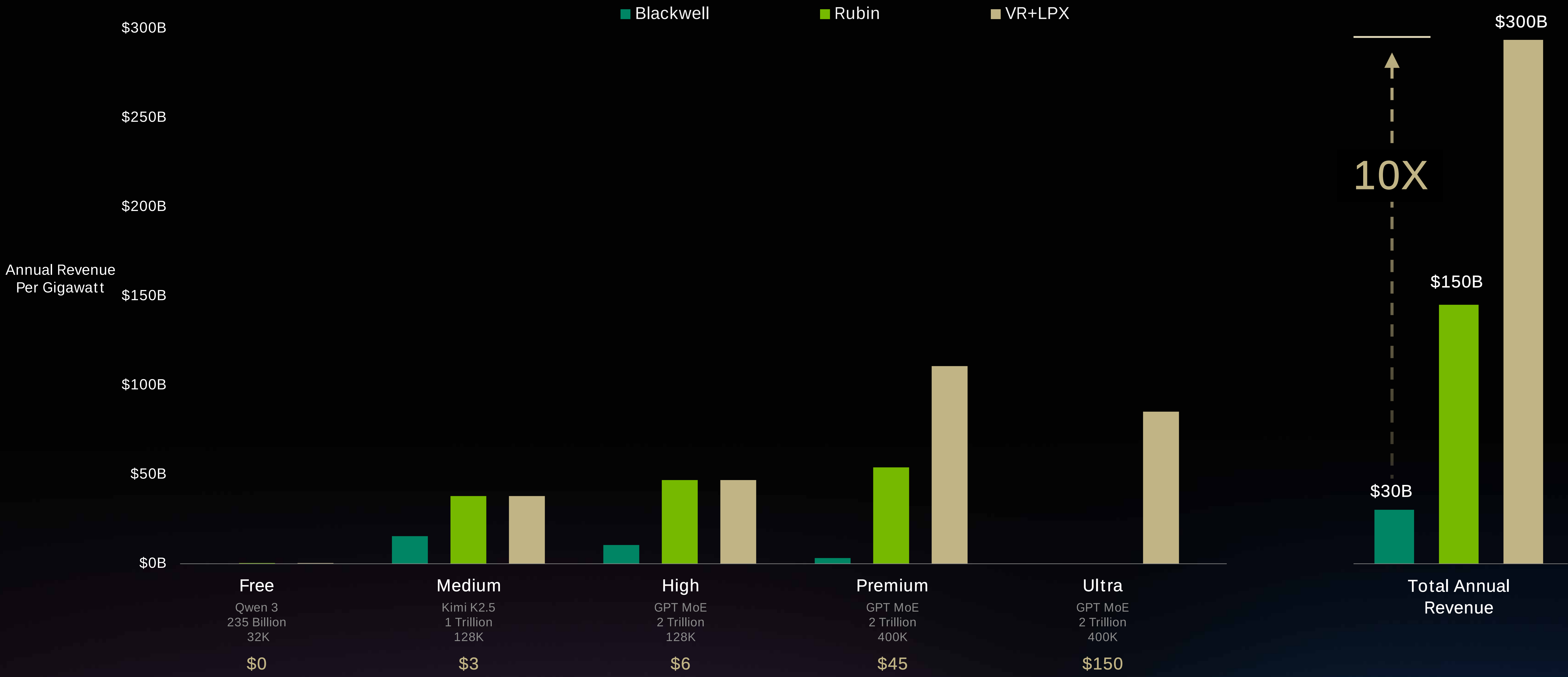


Inference Performance and Efficiency Drive Company Results

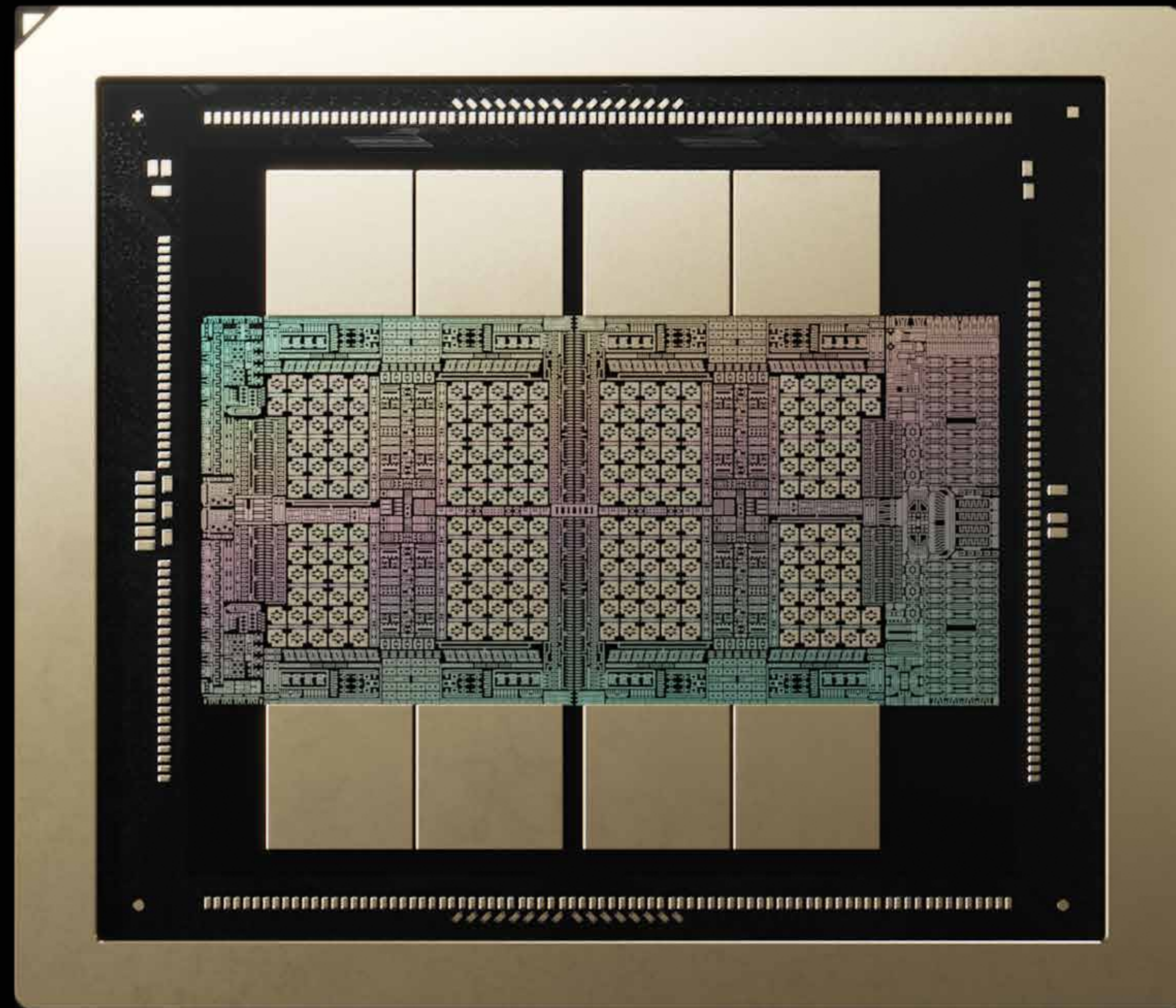


35X Performance
Efficiency Drives
Company Results

NVIDIA Vera Rubin + LPX Unlocks \$300B Revenue Opportunity



Rubin GPU



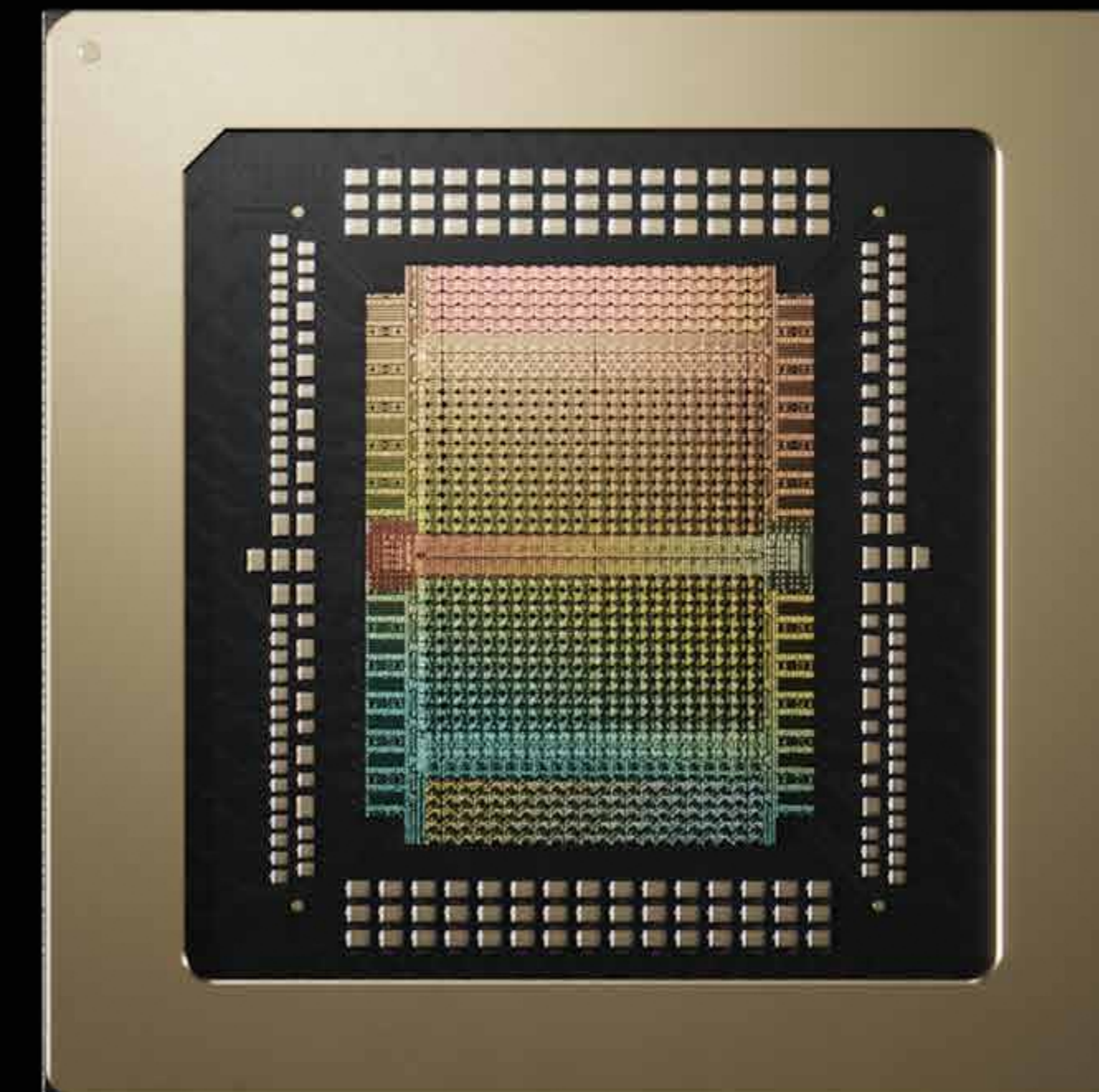
288 GB HBM4

22 TB/s

50 PFLOPs (NVFP4)

336B Transistors
+ 2.5T (HBM4)

Groq 3 LPU



500 MB SRAM

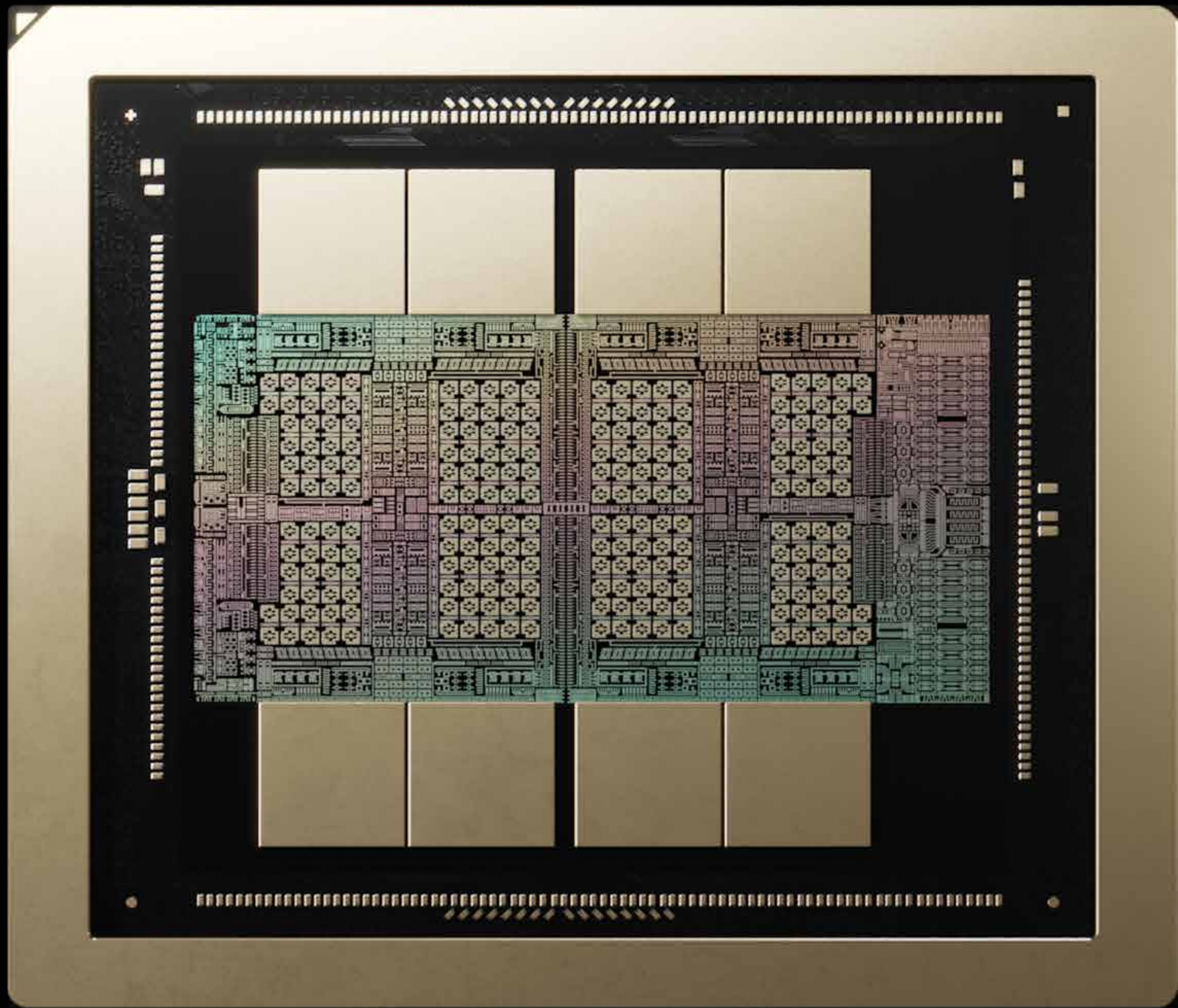
150 TB/s SRAM Bandwidth

1.2 PFLOPs (FP8)

98B Transistors

Uniting Processors of Extreme Performances

Rubin GPU



288 GB HBM4

22 TB/s

50 PFLOPs (NVFP4)

336B Transistors
+ 2.5T (HBM4)

Groq 3 LPU



4 GB SRAM

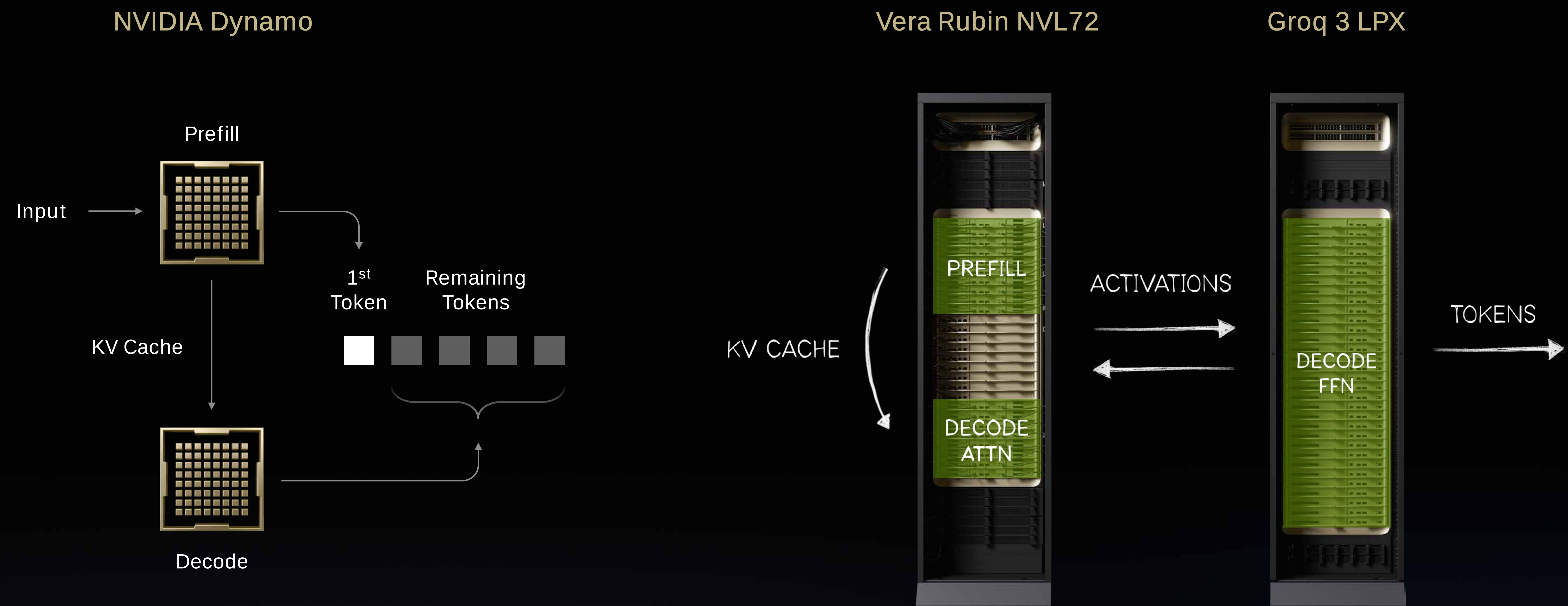
1,200 TB/s SRAM Bandwidth

9.6 PFLOPs (FP8)

784B Transistors

55X

Uniting Processors of Extreme Performances



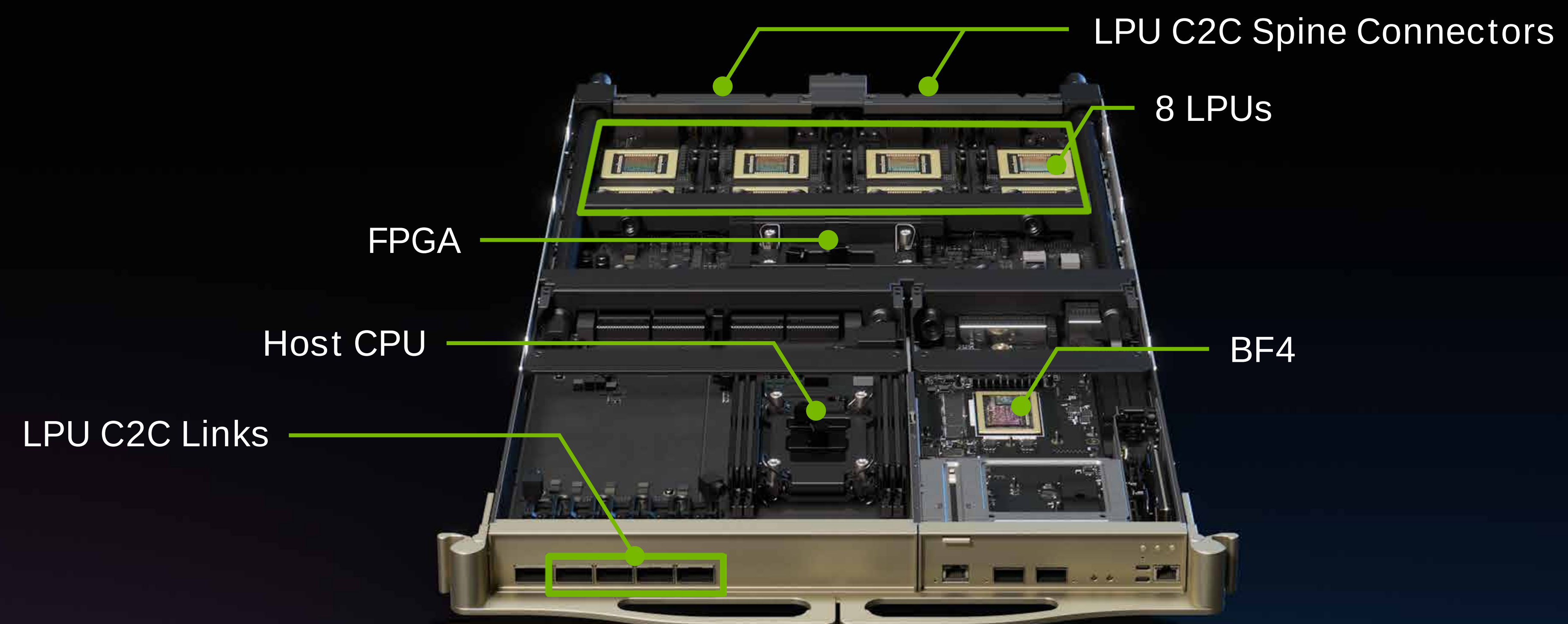
Uniting Processors of Extreme FLOPS and Bandwidth



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

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

wlstron

wiwynn

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



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
MITAC Computing

msi

PEGATRON

QCT

wistron

wiwynn

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



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

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)

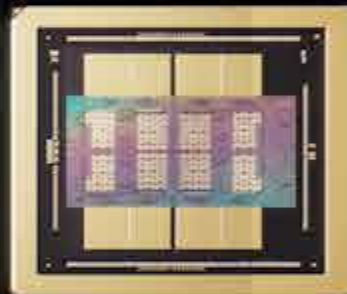
2 EB/s

100 EB/s

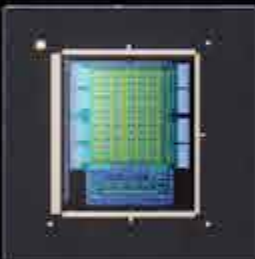
Tokens per Second

2M

700M



Rubin GPU



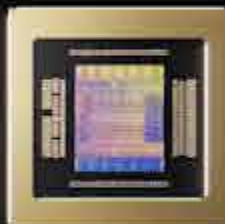
Vera CPU



CX9



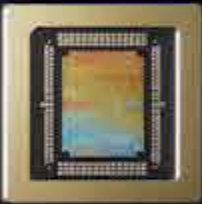
BF4



NVLink Switch



Spectrum CPO



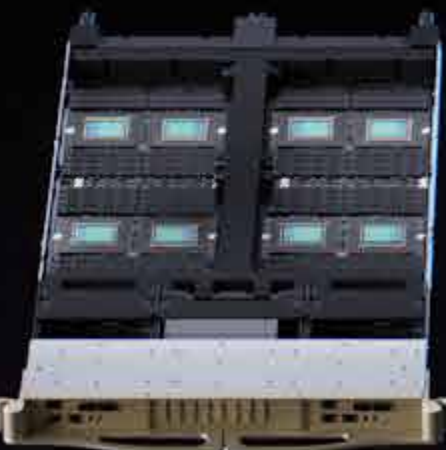
Groq 3 LPU



Vera Rubin Compute Tray



NVLink Switch Tray



Vera Compute Tray



BF4 STX Server



Spectrum Switch



Groq 3 Compute Tray



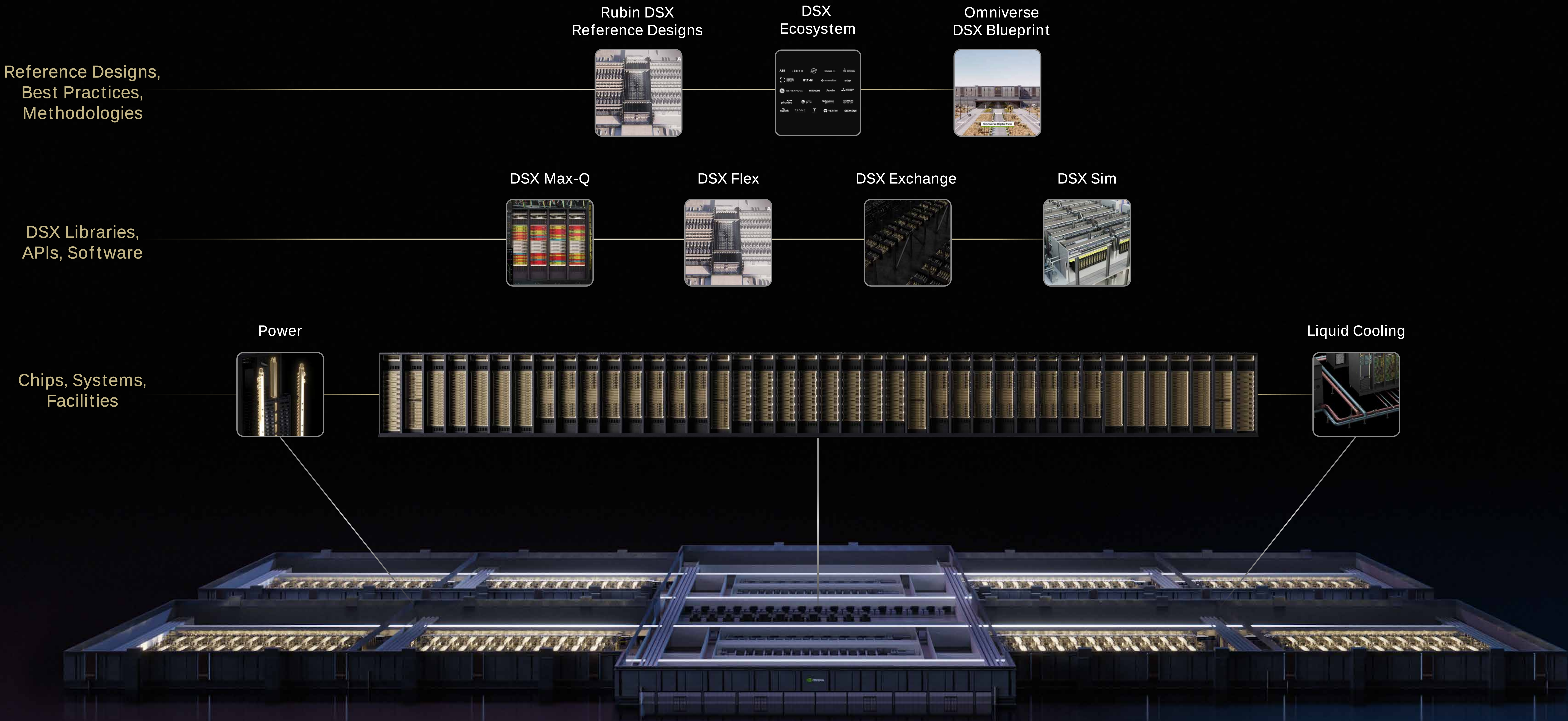
NVIDIA Extreme Co-Design Delivering X-Factors Every Year

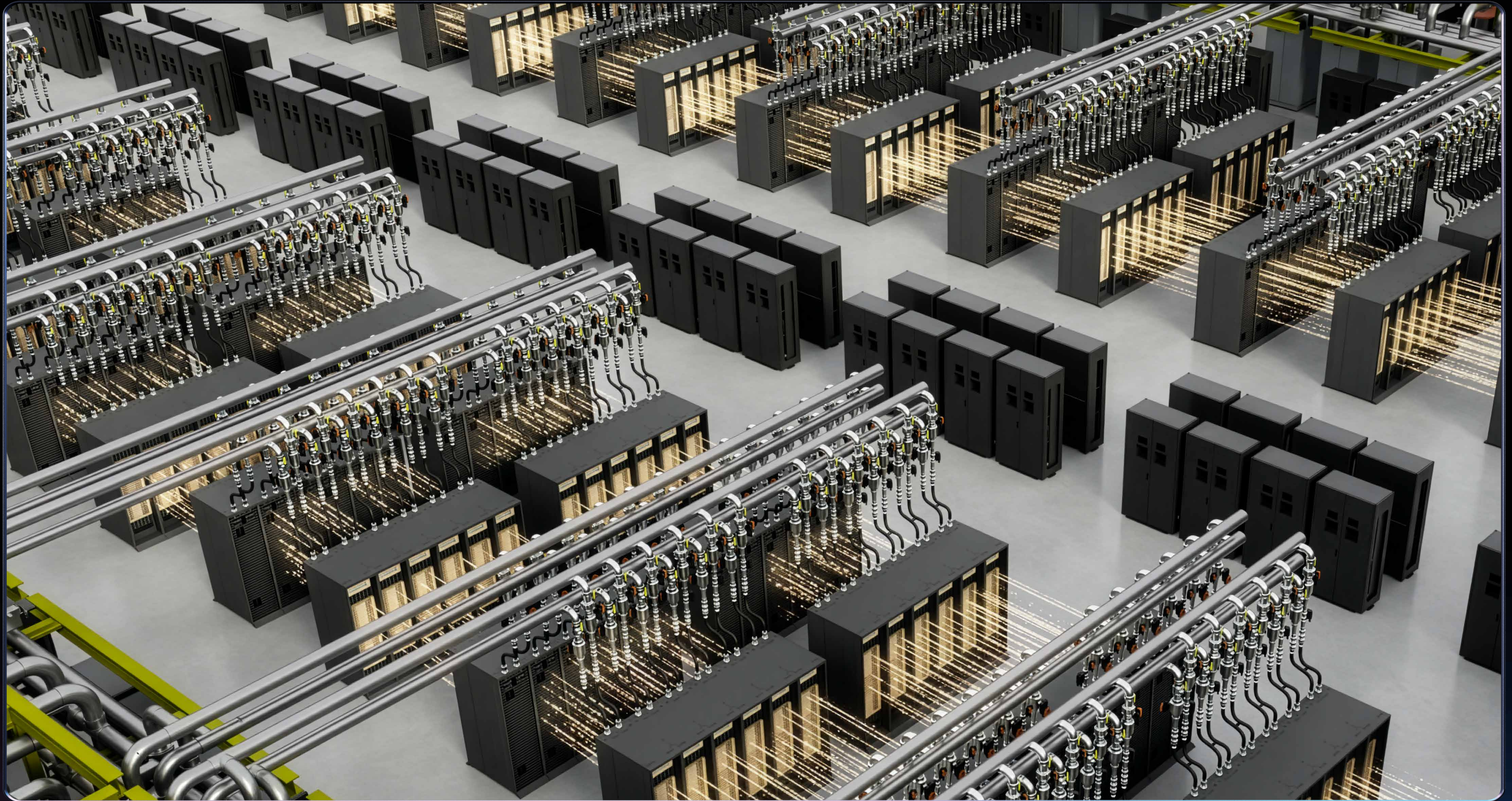
From Chips to Racks to AI Factories



NVIDIA DSX AI Factory Platform

Extreme Co-Design at Infrastructure Scale





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

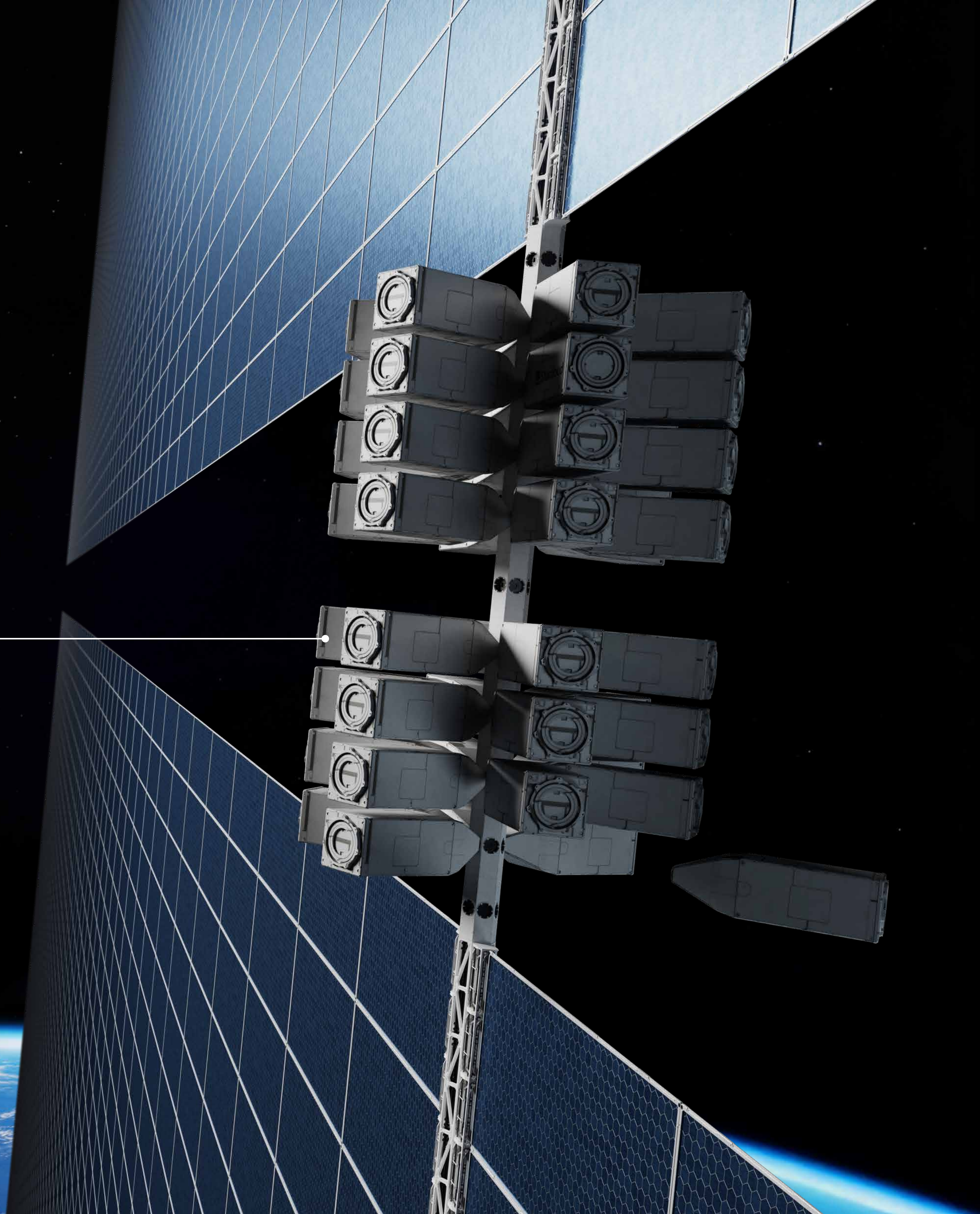
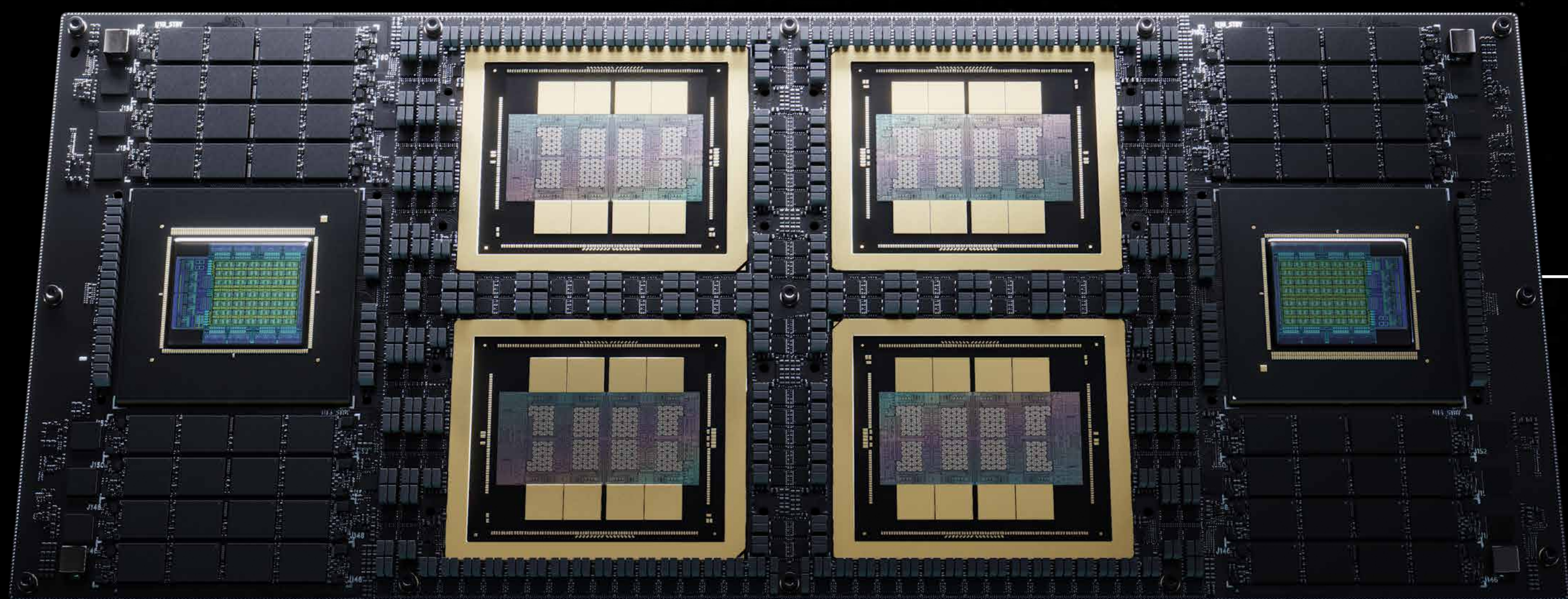
ABB ARMADA AVEVA aviz BECHTEL networks cadence CHECK POINT
CLOUDIAN DASSAULT SYSTEMES DELL Technologies E.T.N etap GE VERNOVA hedgehog HITACHI
Jacobs KEYSIGHT metalsoft MIRANTIS MITSUBISHI ELECTRIC netris NOKIA Open Nebula
paloalto ptc RAFAY Red Hat run:ai Schneider Electric SIEMENS SIEMENS energy
slurm Spectro Cloud TRANE TECHNOLOGIES TrendAI VAST vCluster VERTIV WEKA

DSX Exchange

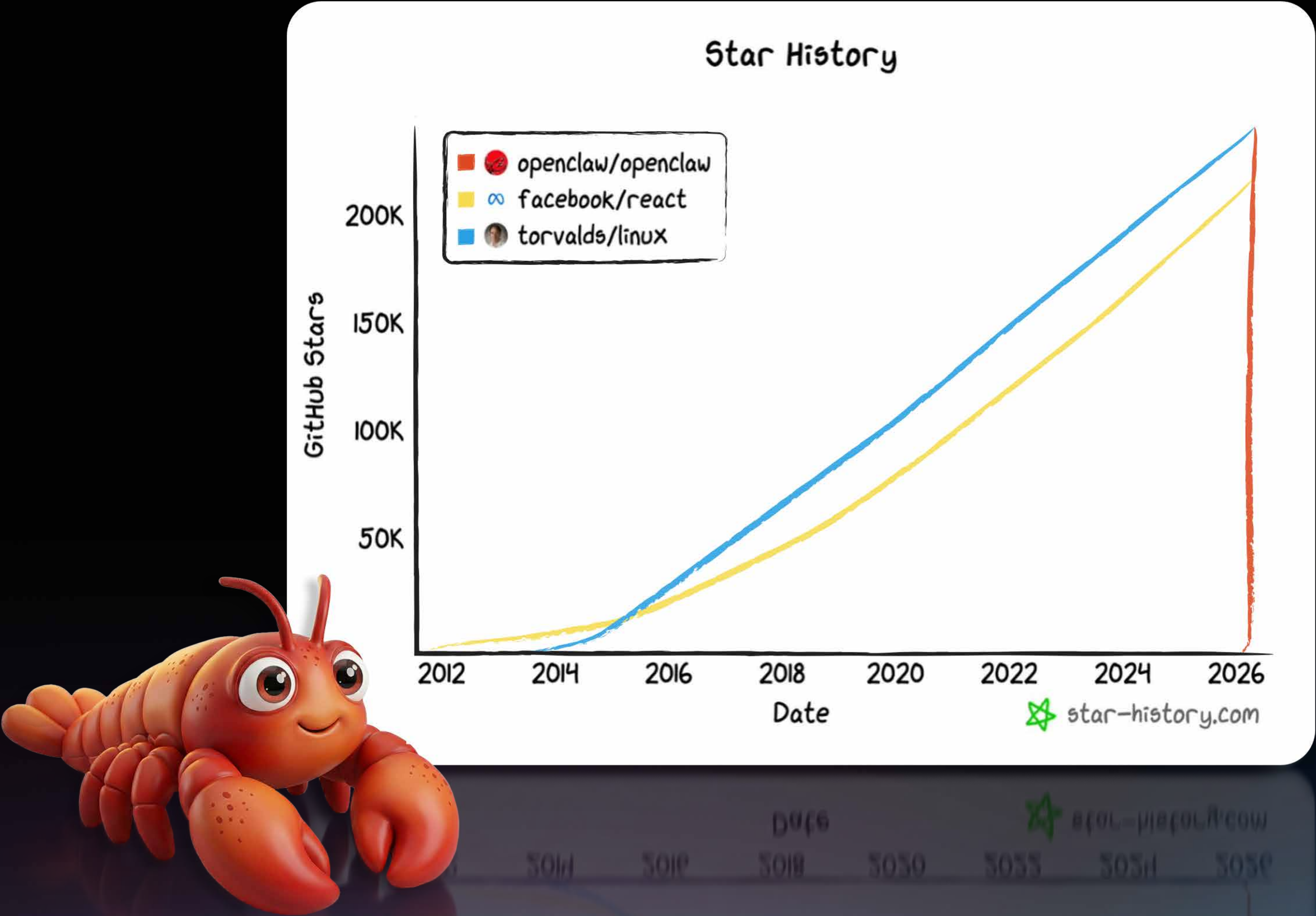
ABB ARMADA AVEVA aviz networks cadence CHECK POINT CLOUDIAN
CoreWeave Crusoe DIGITAL REALTY DELL Technologies E.T.N emeraldai GE VERNOVA hedgehog
HITACHI KEYSIGHT metalsoft MIRANTIS MITSUBISHI ELECTRIC NEBIUS netris NOKIA
NSCALE Open Nebula paloalto phaidra RAFAY Red Hat run:ai Schneider Electric SIEMENS 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/nemoclaws.sh | bash  
→ ~ nemoclaws 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)

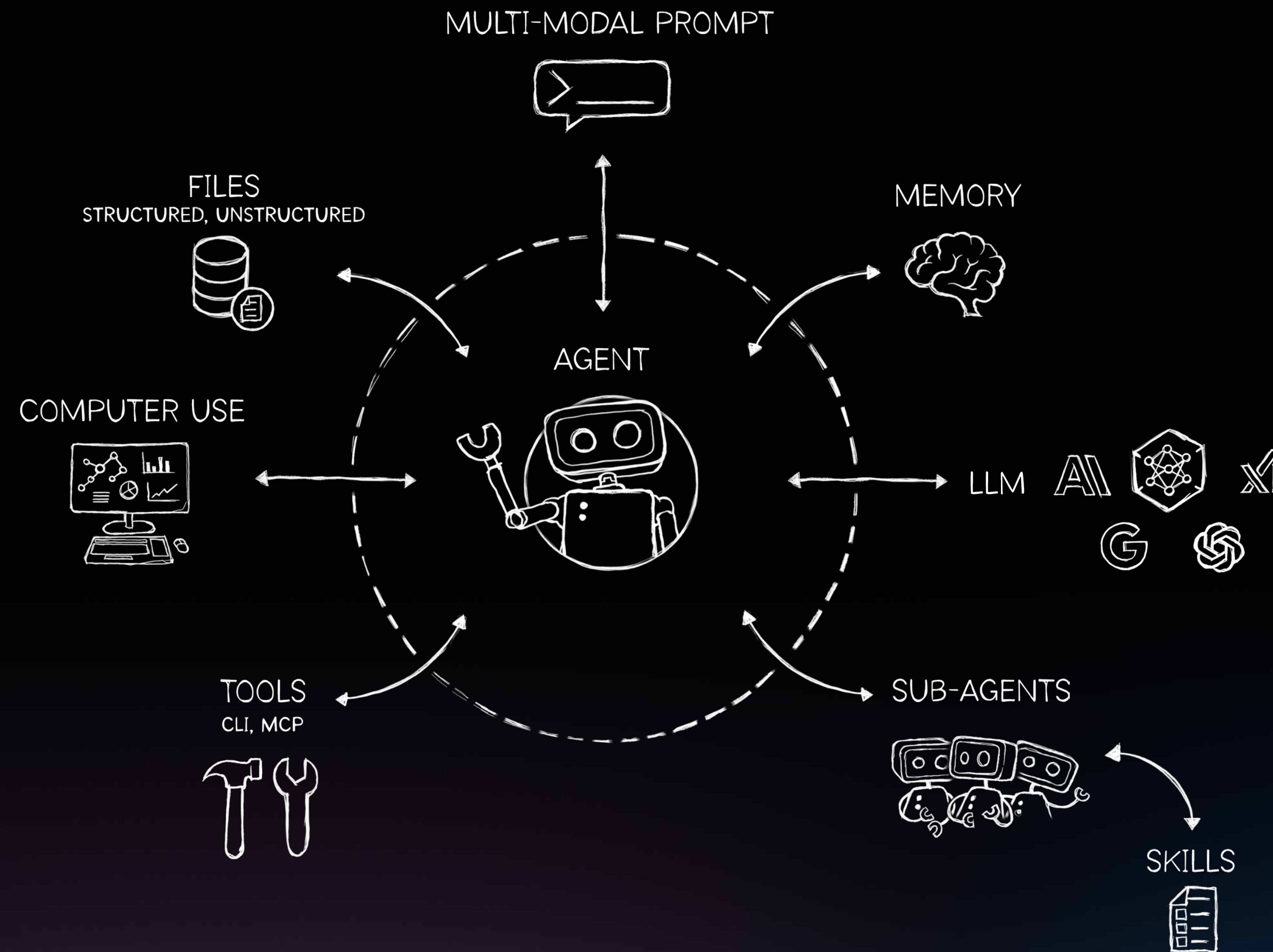
I'm not saying your workflow is chaotic... I'm just bringing a linter and a helmet.

OPENCLAW

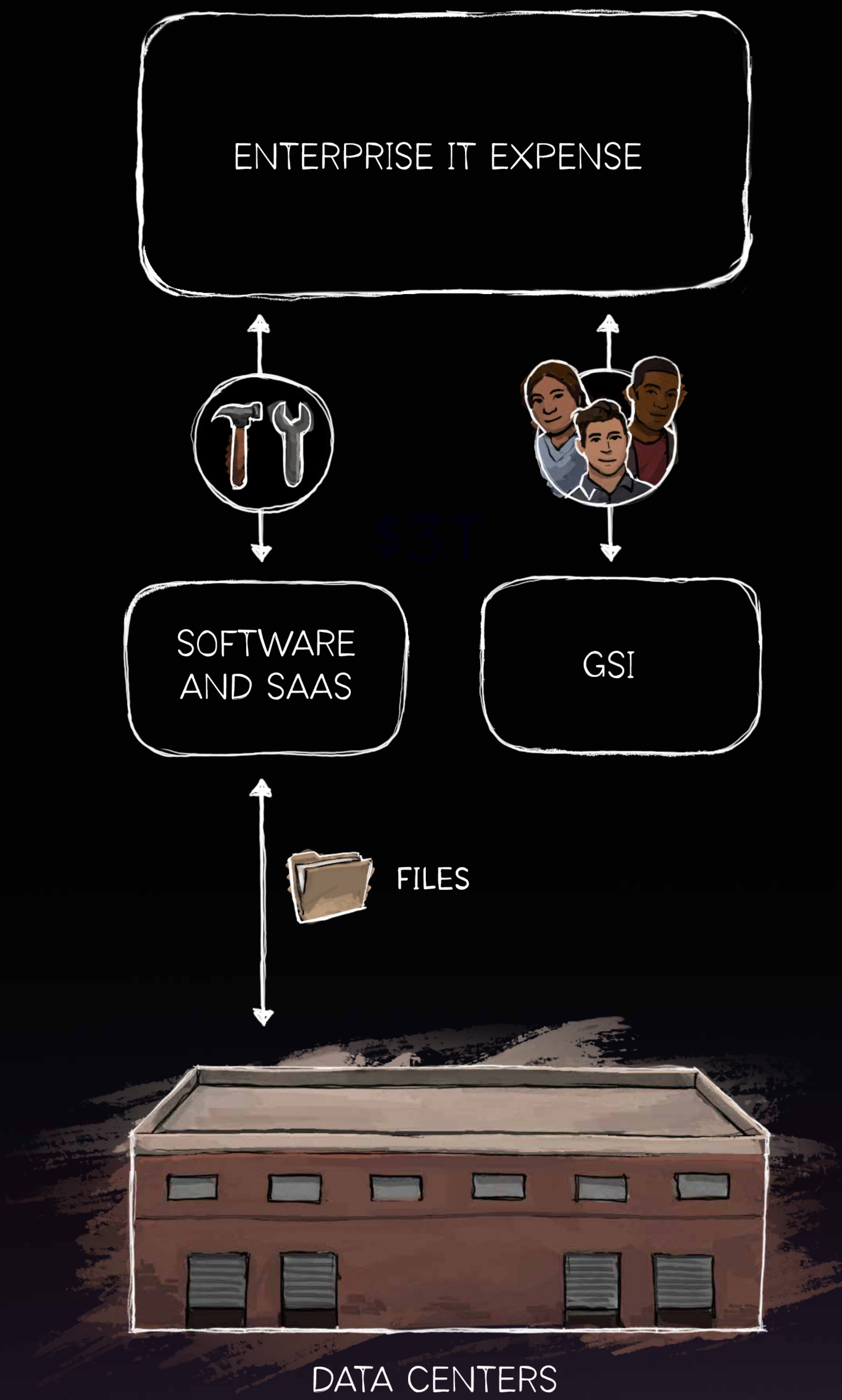
🦀 OPENCLAW 🦀

└─ OpenClaw onboarding

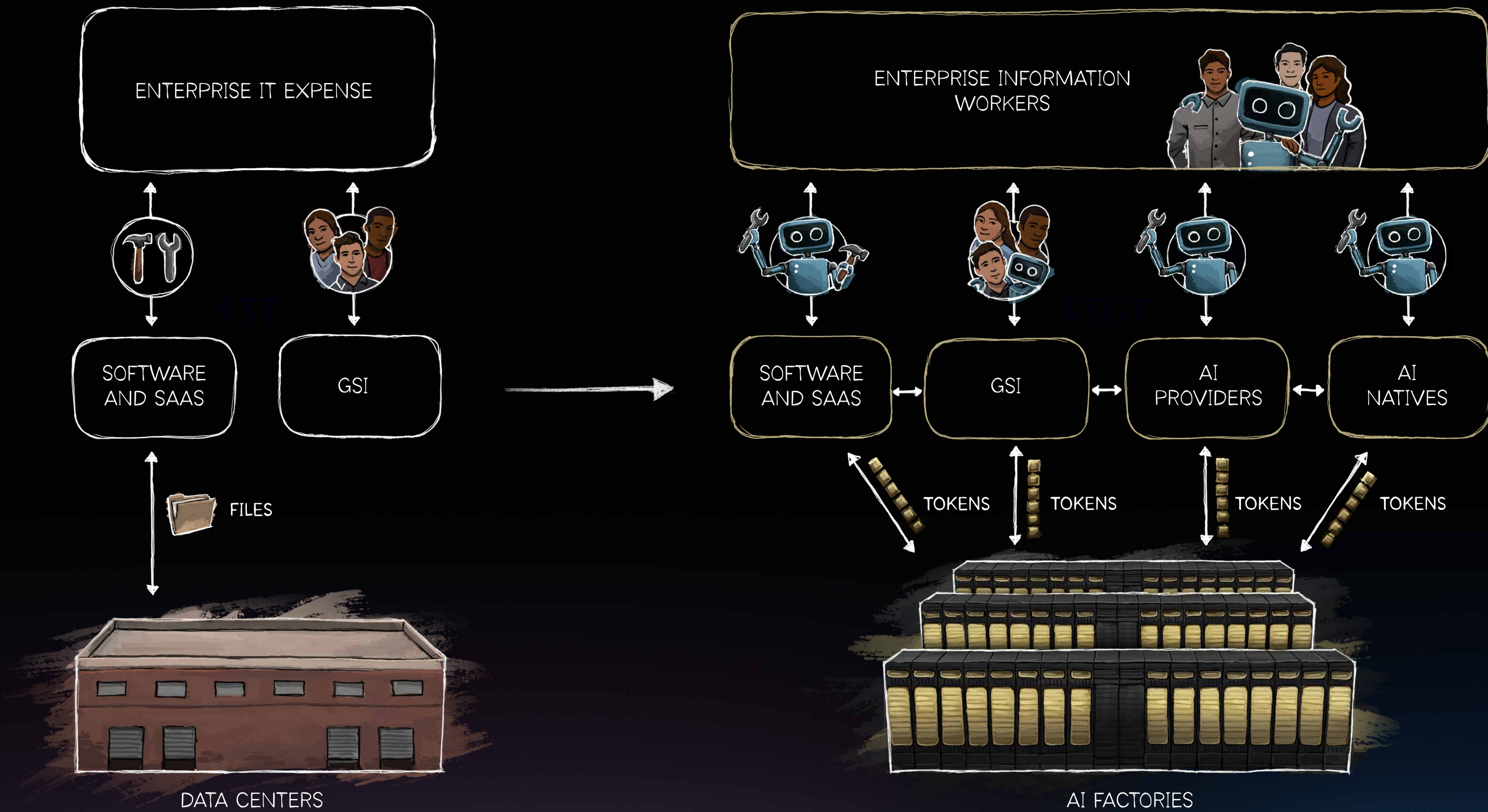
Agents – A New Computing Platform



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

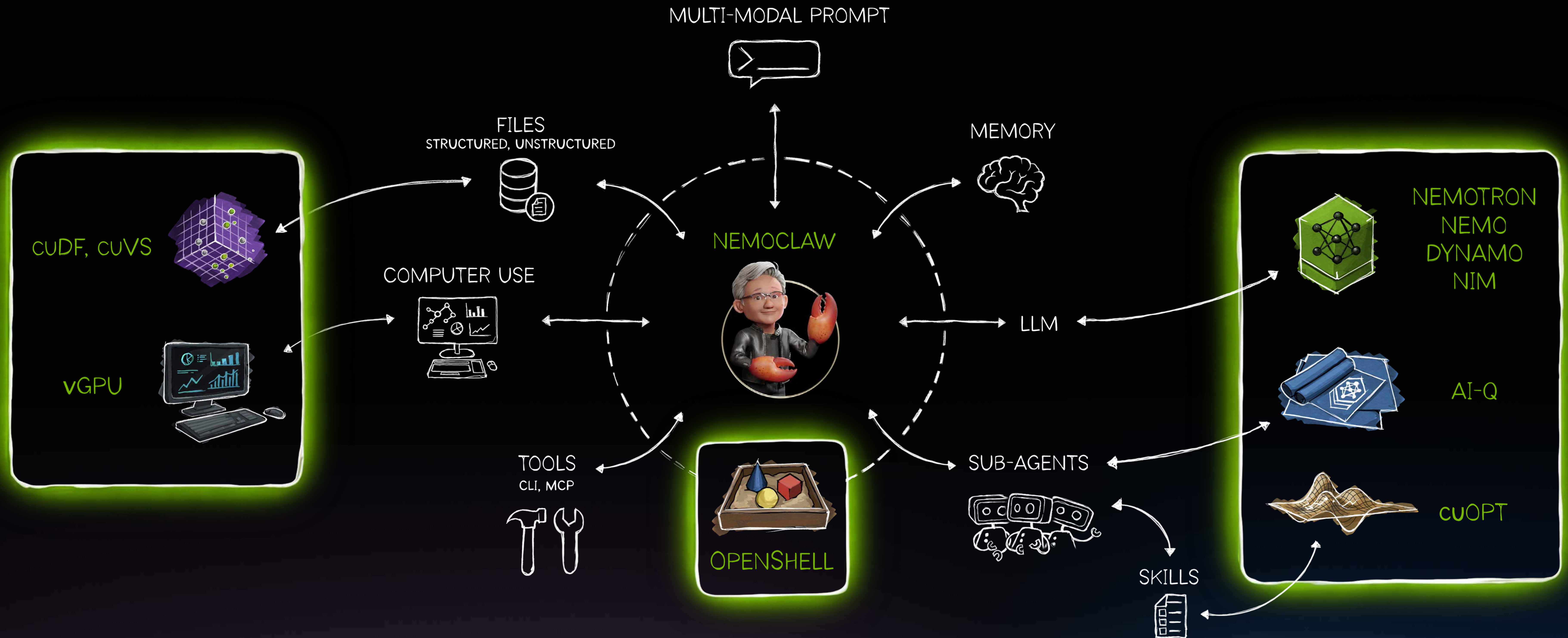


Enterprise IT Renaissance from SaaS to Agent-as-a-Service



Announcing NVIDIA NemoClaw Reference OpenClaw

NVIDIA Agent Toolkit for Building Specialized Agents



NVIDIA Leaderboard Topping Open Models



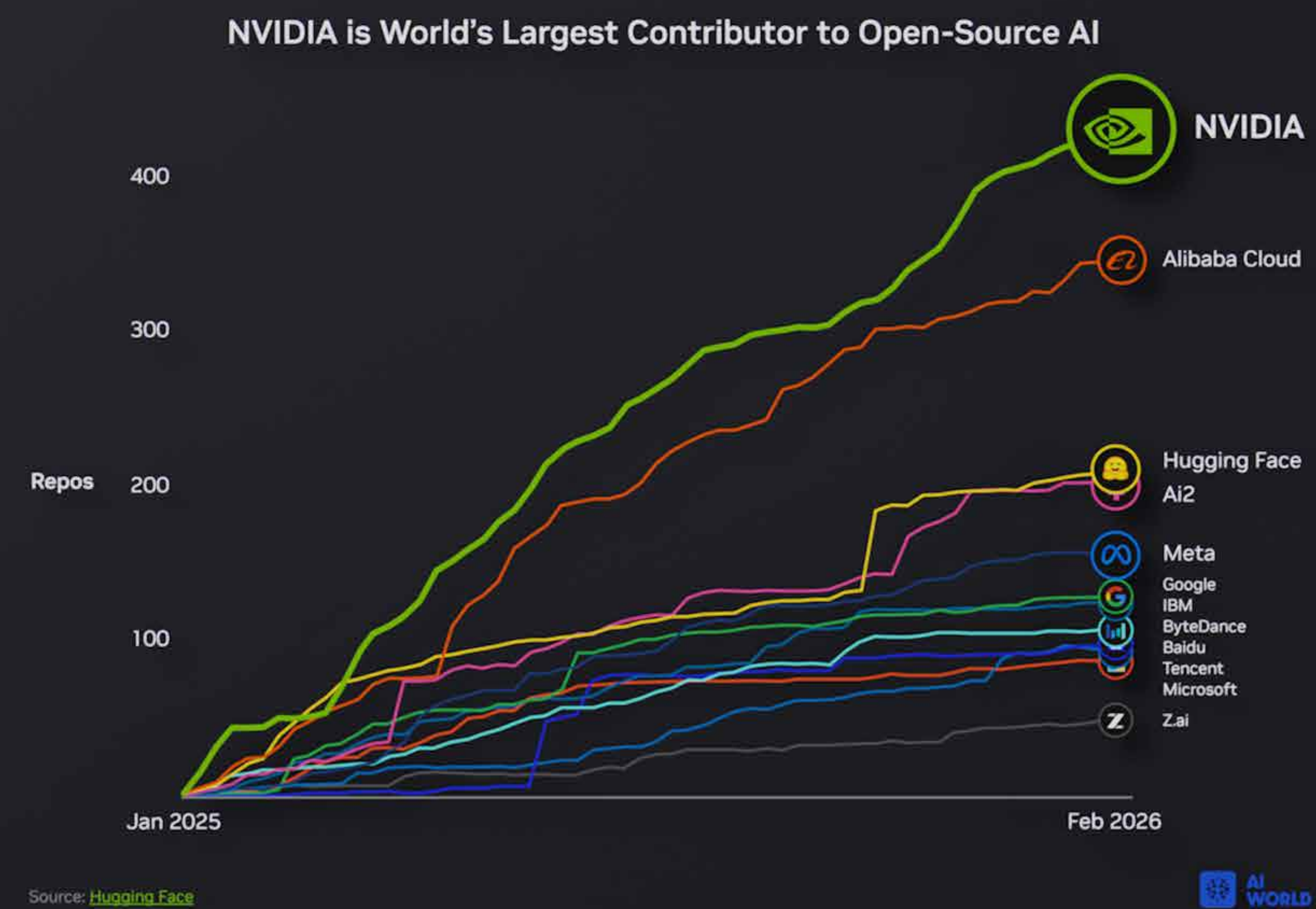
BioNeMo
Biology AI



Earth-2
AI Physics



Nemotron
Agentic AI



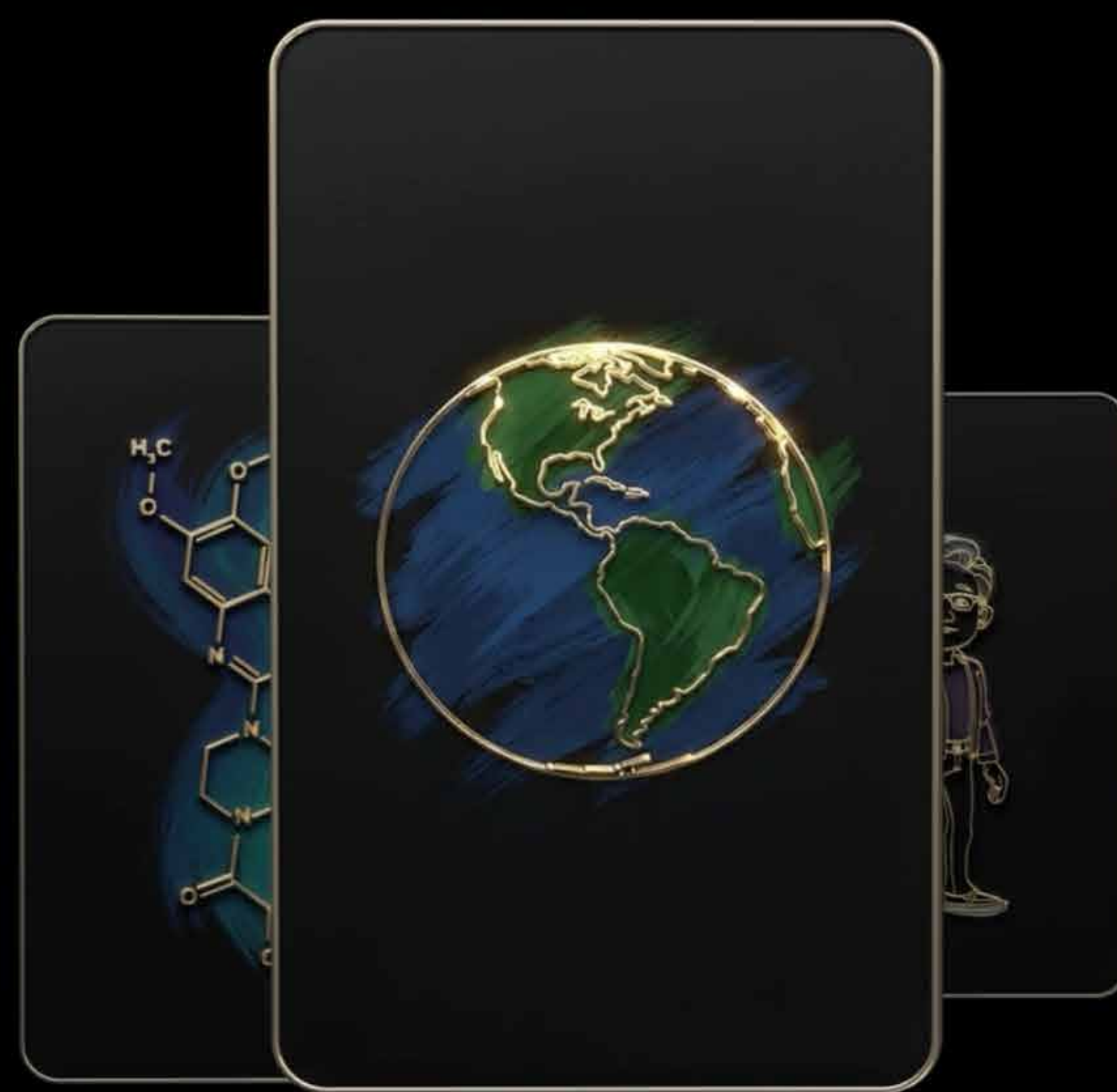
Cosmos
Physical AI



GR00T
Robotics



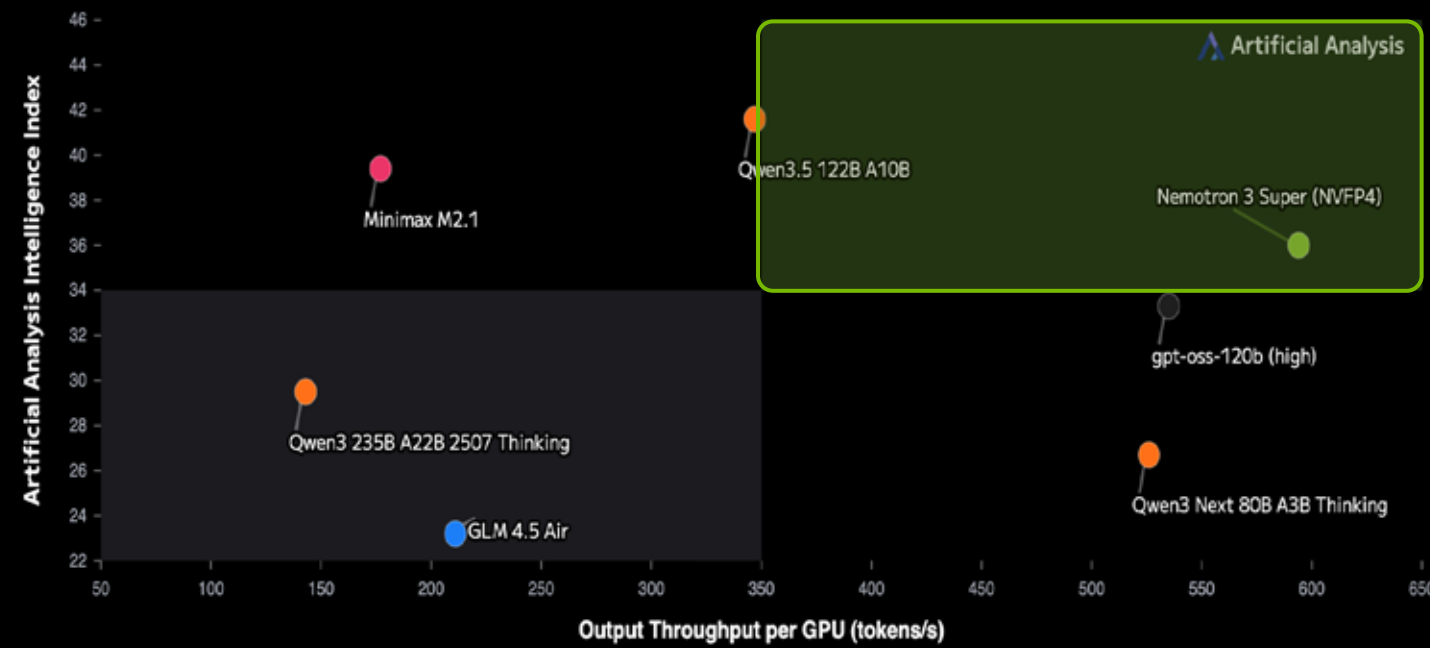
Alpamayo
Autonomous Vehicles



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)	54.58
2	Xiaoyi DeepResearch Evo	49.34
3	Xiaoyi DeepResearch	46.98
4	OpenAI- GPT-o3 Deep Research	45.48
5	Gemini-3-Pro Deep Research	44.68
6	Gemini-2.5-Pro Deep Research	41.98
	DeepSeek Deep Research	40.99

ViDoRe, MMTEB, BRIGHT

Nemotron RAG

1	nemotron-colembd-v1-8b-v2	16722	8.7	4096
2	tomoro-colqwen3-embed-8b	16724	8.0	320
3	nemotron-colembd-v1-4b-v2	9286	4.8	2560
4	Qns-Colqwen3-4B	9286	4.8	2560
5	tomoro-colqwen3-embed-4b	8466	4.0	320
6	llama-nemotron-colembd-v1-3b-v2	8483	4.4	3872
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-colembd-3b-v1	8483	4.4	3872
11	SauerkrautLM-ColQwen3-4b-v0.1	8465	4.4	128

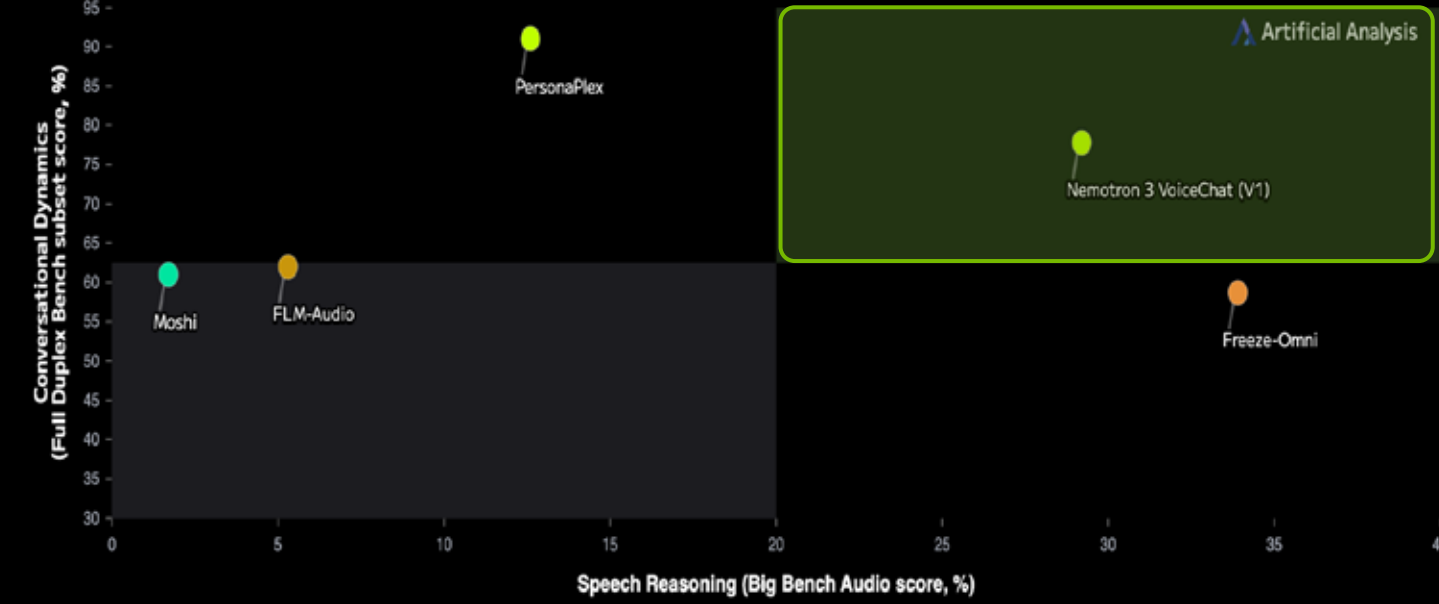
OCRBench v2

NemotronVL

5	Nemotron Nano V2 VL	-	Yes	12B	61.2	67.6	54.5
6	Gemini-2.5-Pro	-	No	-	59.3	70.9	45.6
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-5.2	-	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-T2V-A14B	80.6	84.1	77.2	93.2	66.3	81.7	89.2
Wan2.2-T2V-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-T2V-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-T2V	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
π_0	96.8	98.8	95.8	85.2	94.2
UVA	--	--	--	90.0	--
UniVLA	96.5	96.8	95.6	92.0	95.2
$\pi_0.5$	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
CogVLA	98.6	98.8	96.6	95.4	97.4

MolmoSpaces

GR00T N

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

RoboArena

GR00T N

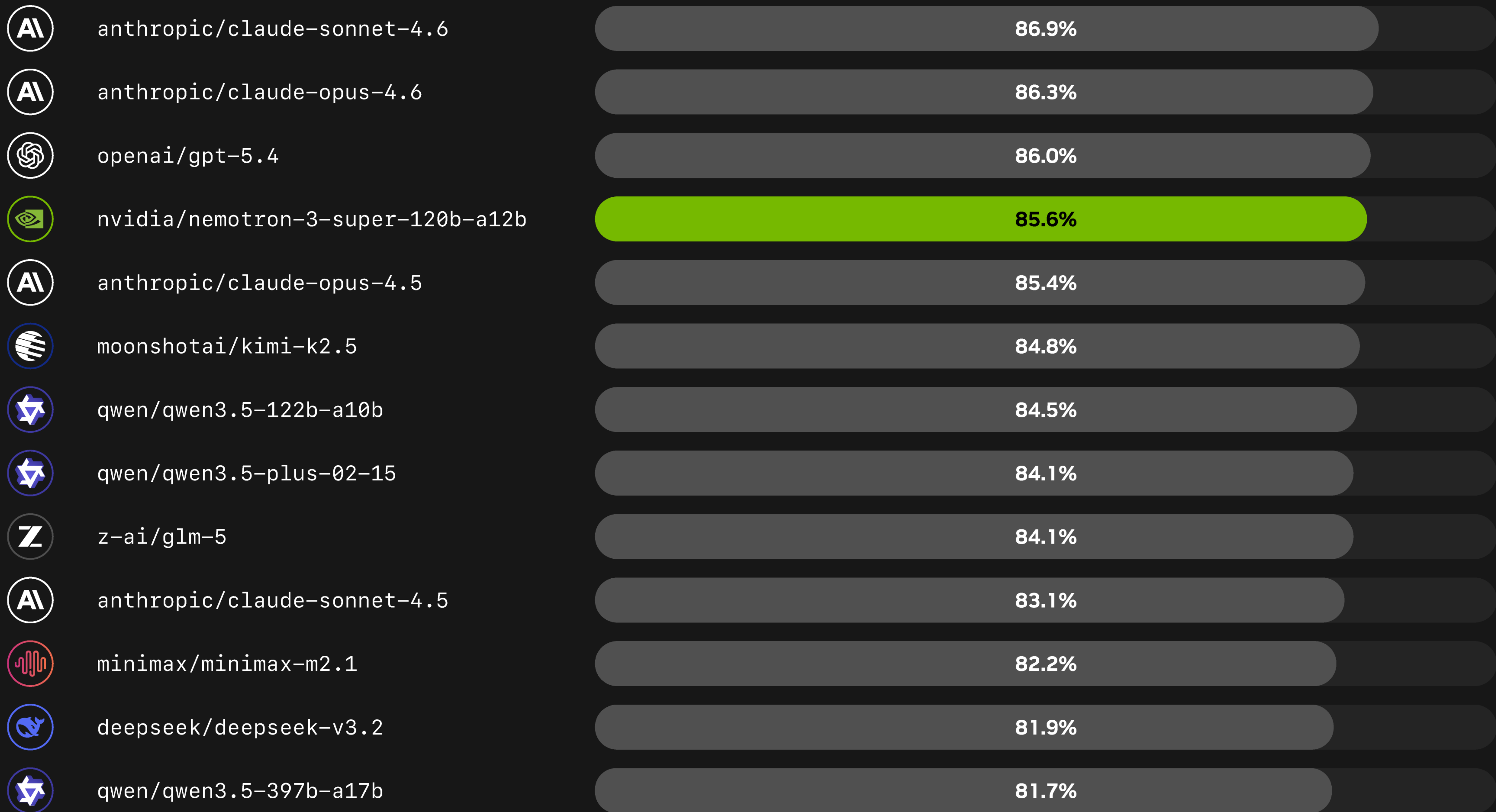
1	DreamZero X	1752	40.4	168
2	pi05_droid X	1621	27.2	615
3	pi0_fast_droid X	1595	25.8	797
4	floppy_bagel_195k_temp1p0	1588	147.1	8
5	floppy_bagel_525k_temp1p0	1560	74.4	33
6	paligemma_vq_droid	1559	25.9	814
7	paligemma_diffusion_droid	1548	25.9	807
8	paligemma_fast_specialist_droid	1546	24.7	1,007
9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

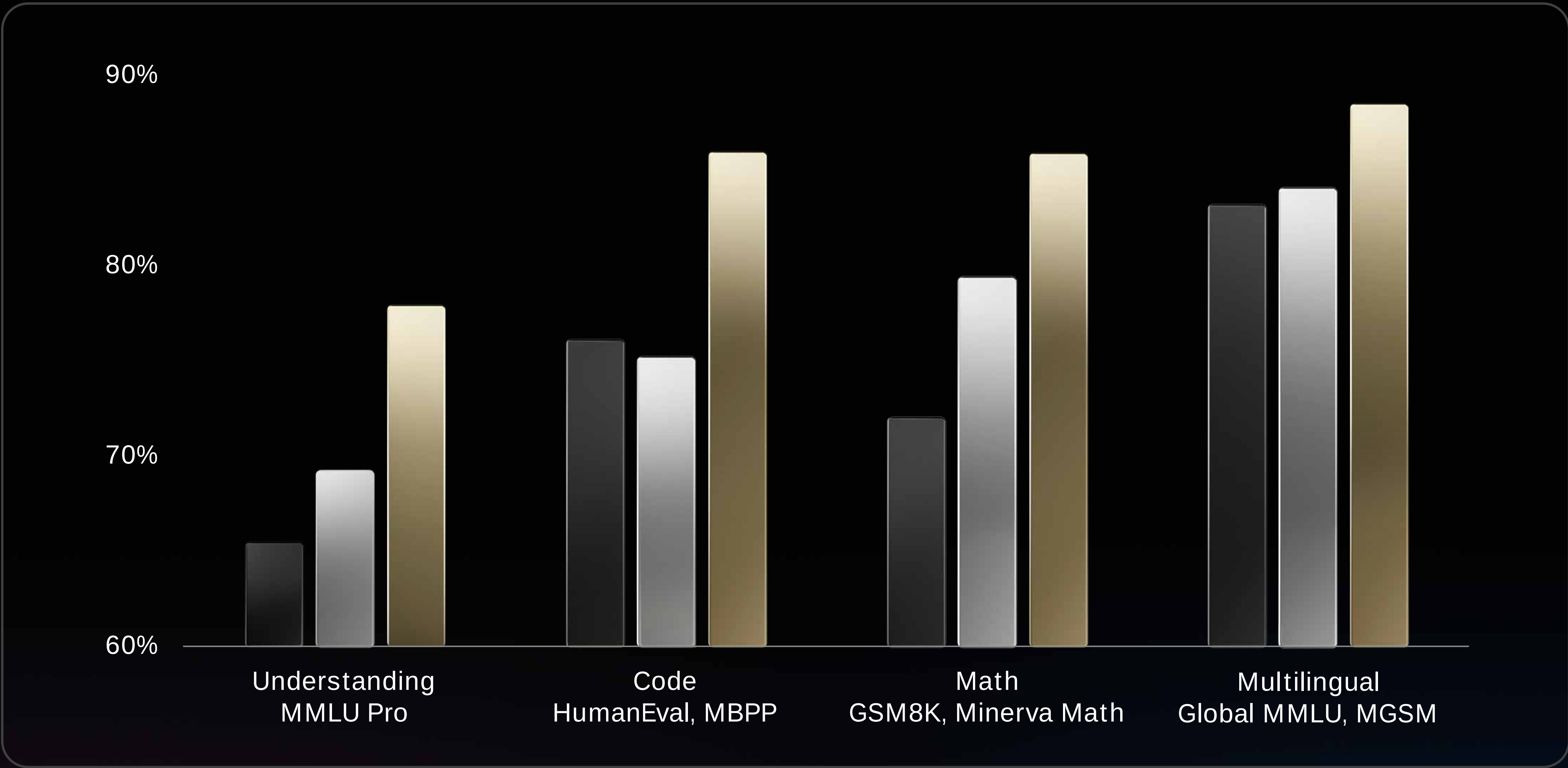
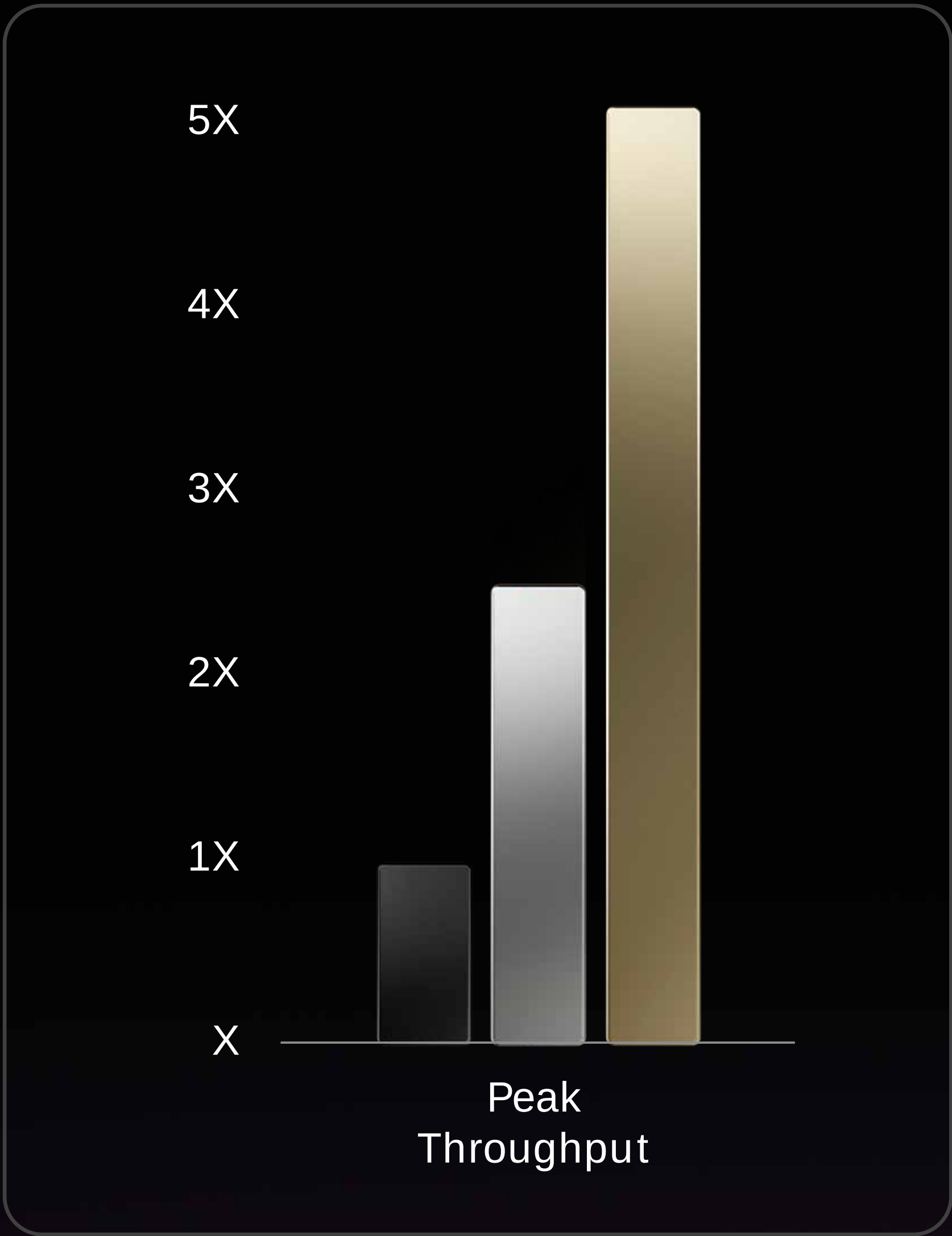
NVIDIA Nemotron 3 Super – Best Open Model for OpenClaw



NVIDIA Nemotron 3 Ultra – Best Open Base Model

5X Efficiency and Highest Reasoning Accuracy on NVIDIA GB200 NVL72

GLM Kimi K2 Nemotron 3 Ultra

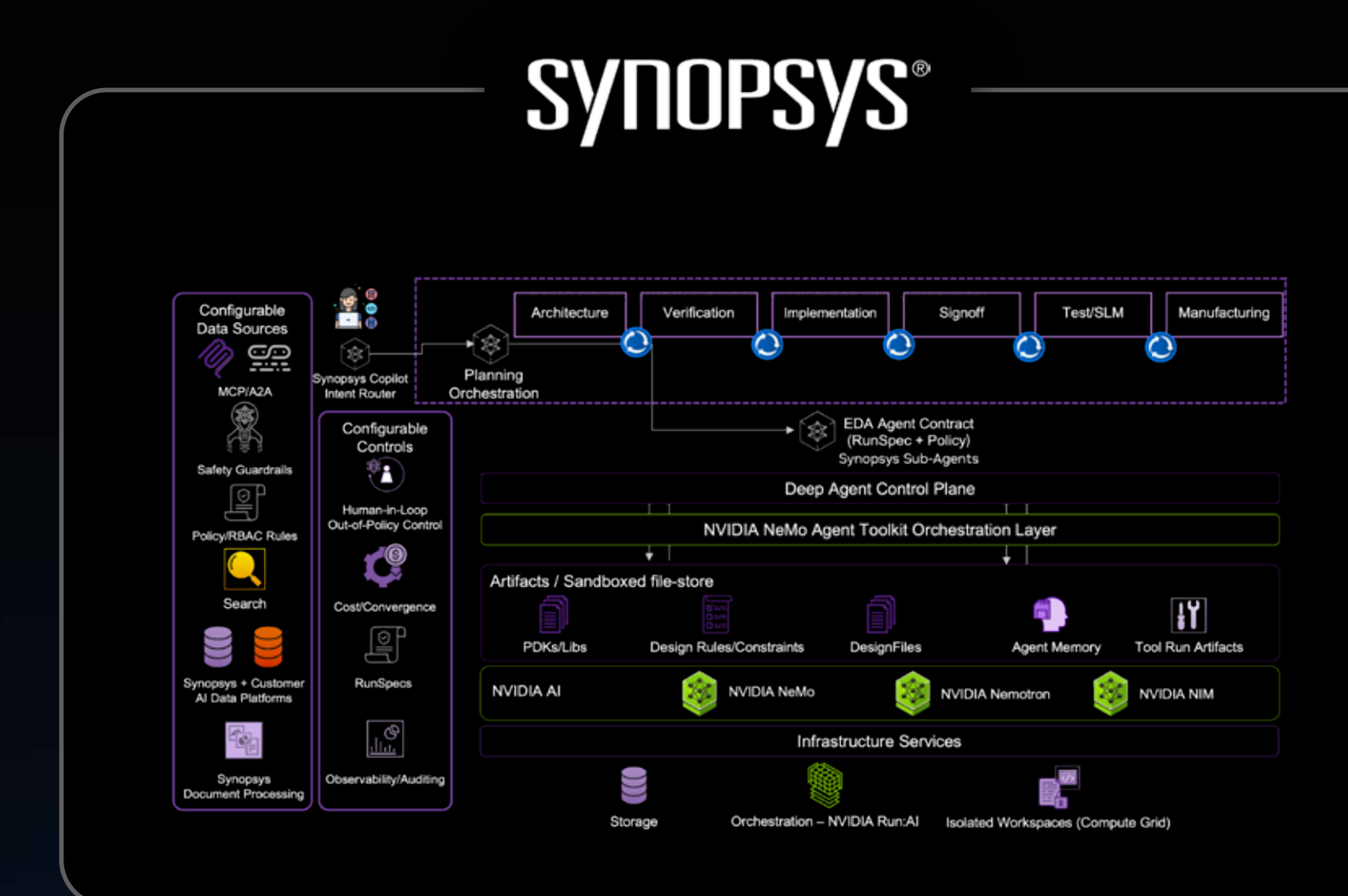
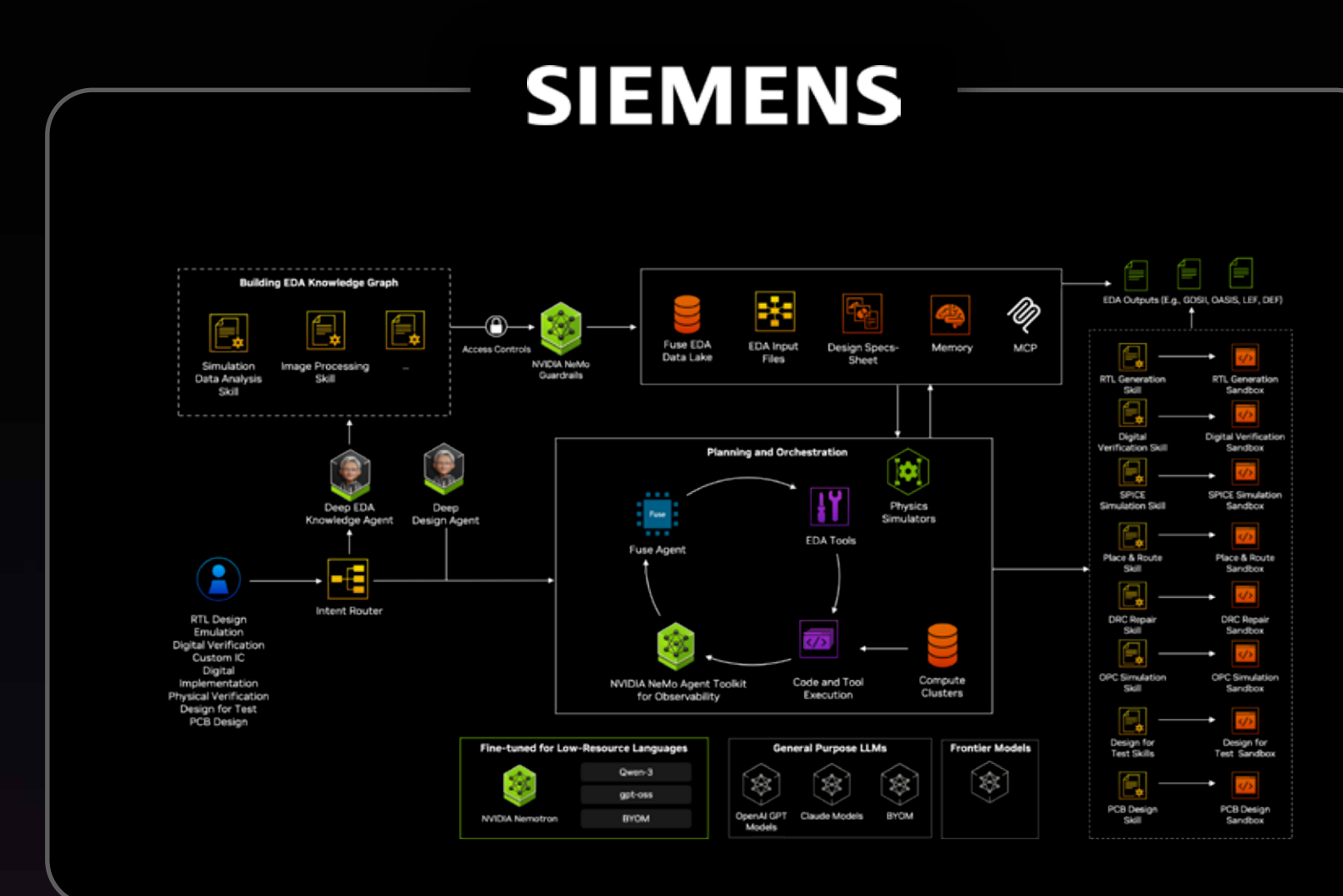
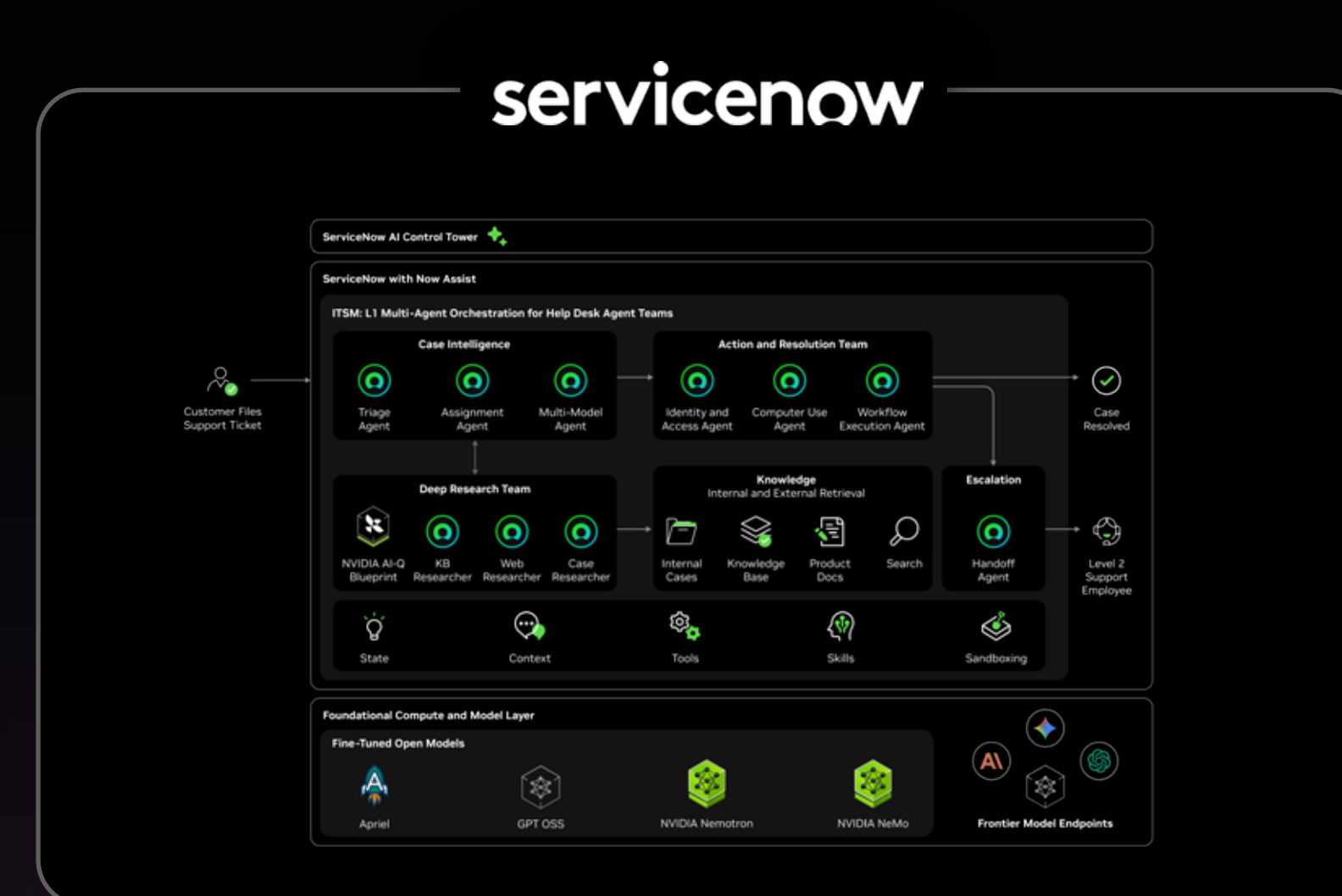
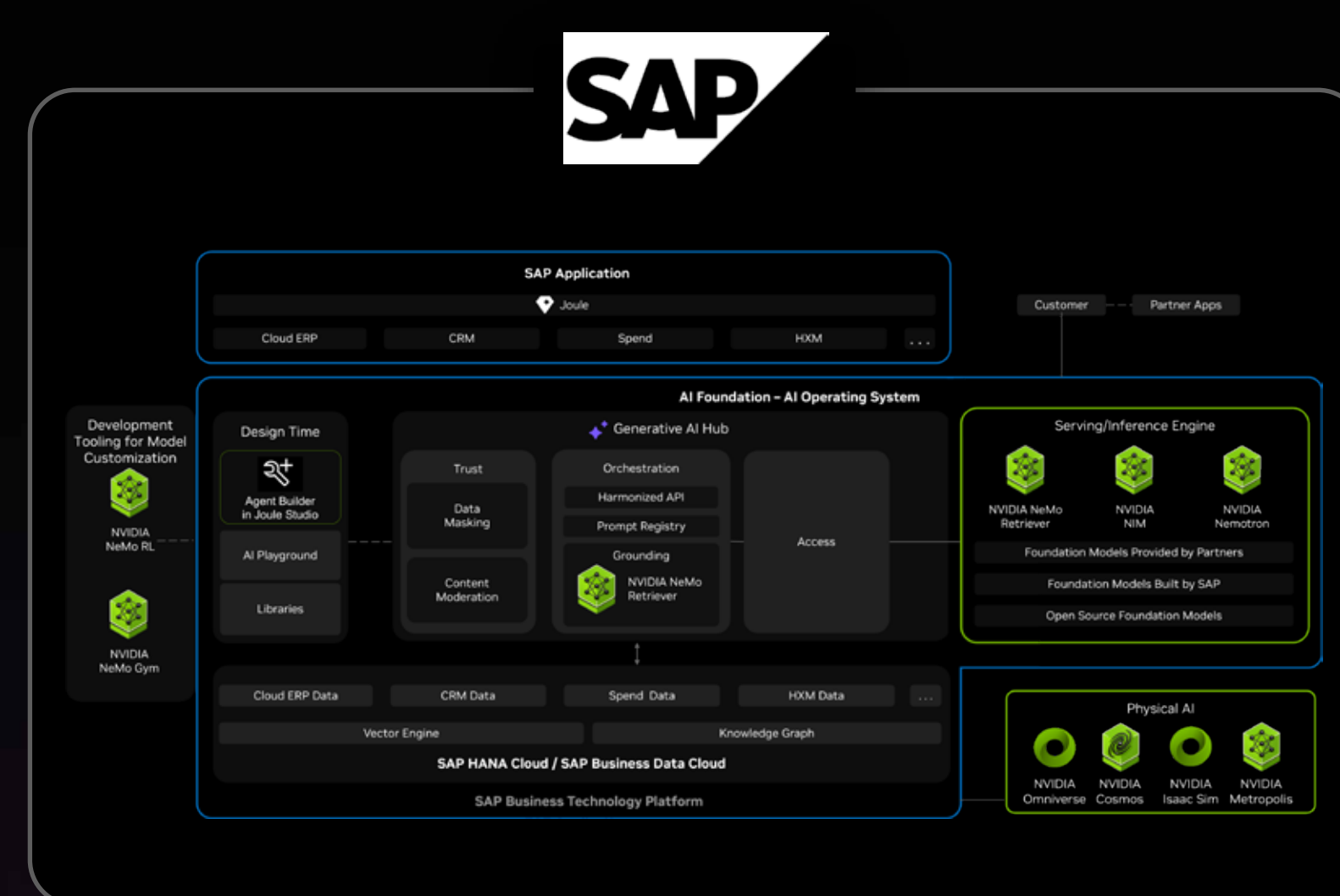
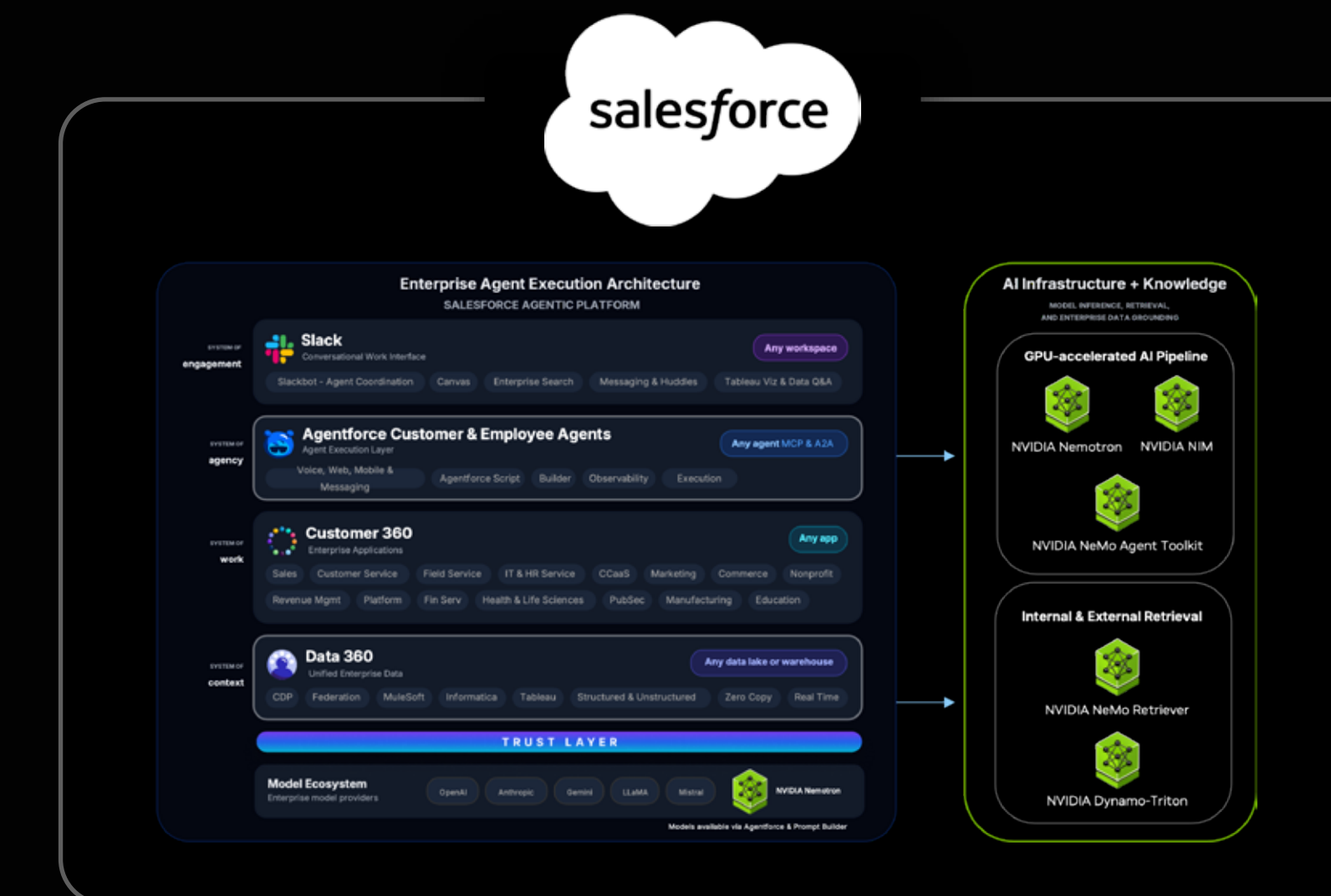
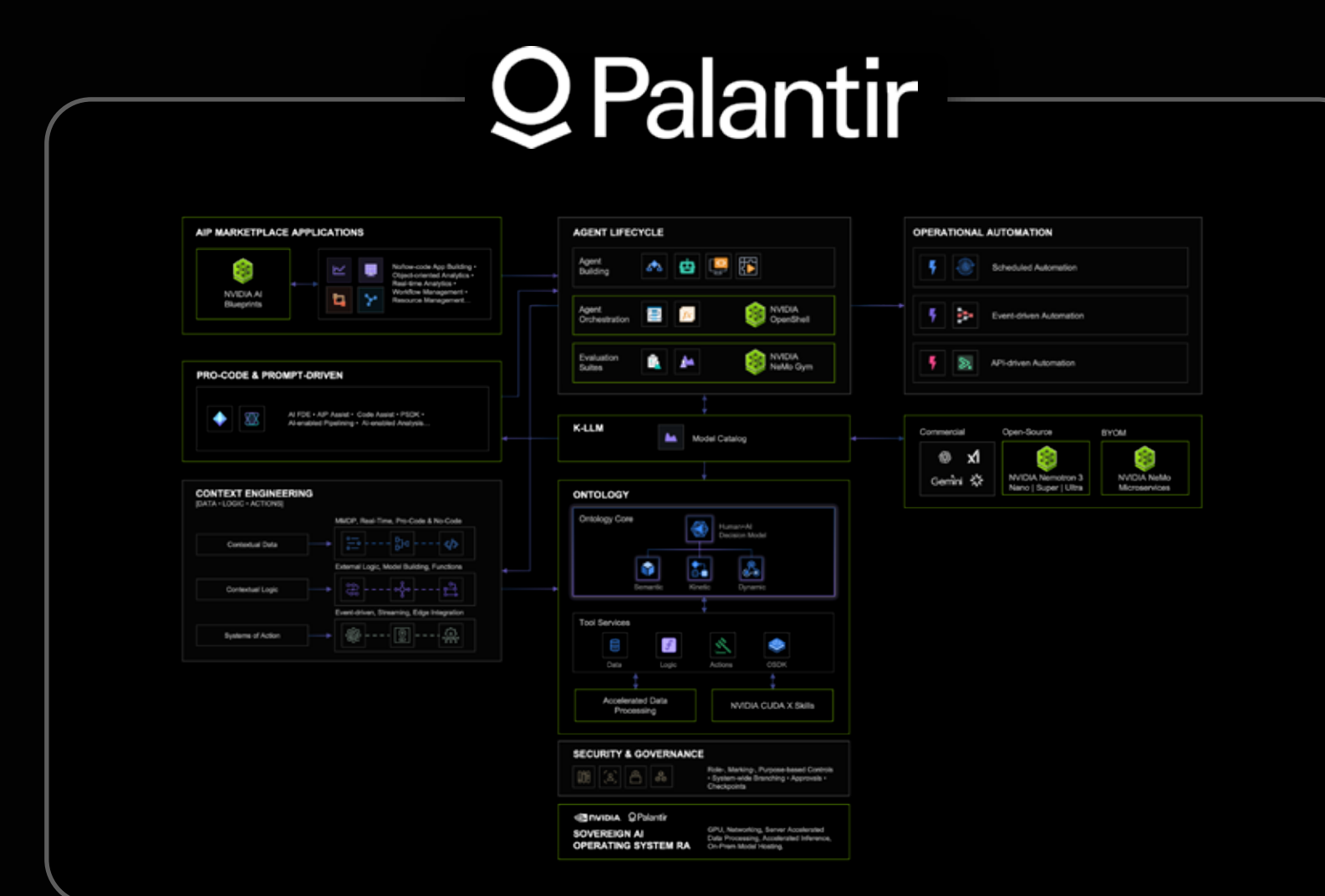
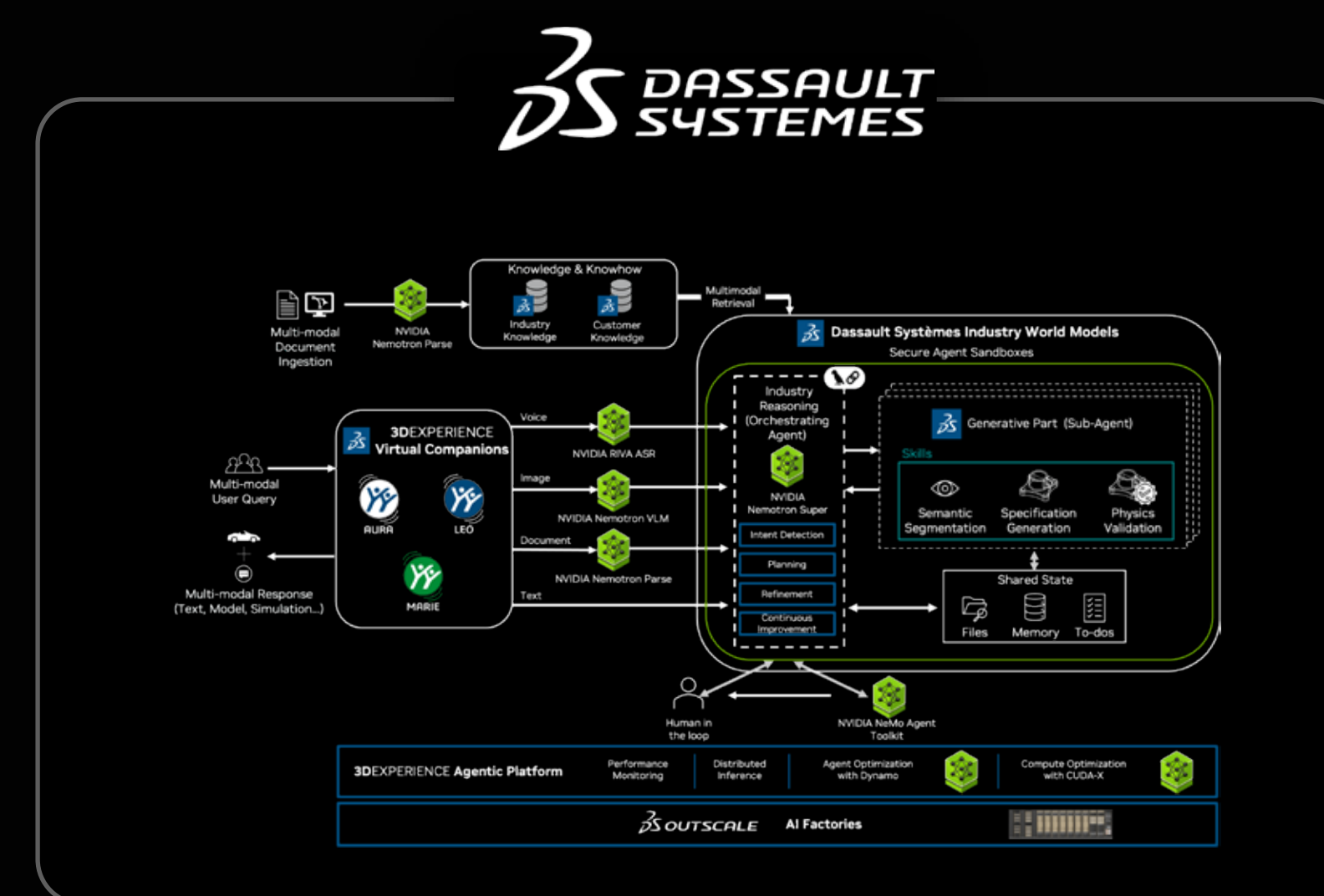
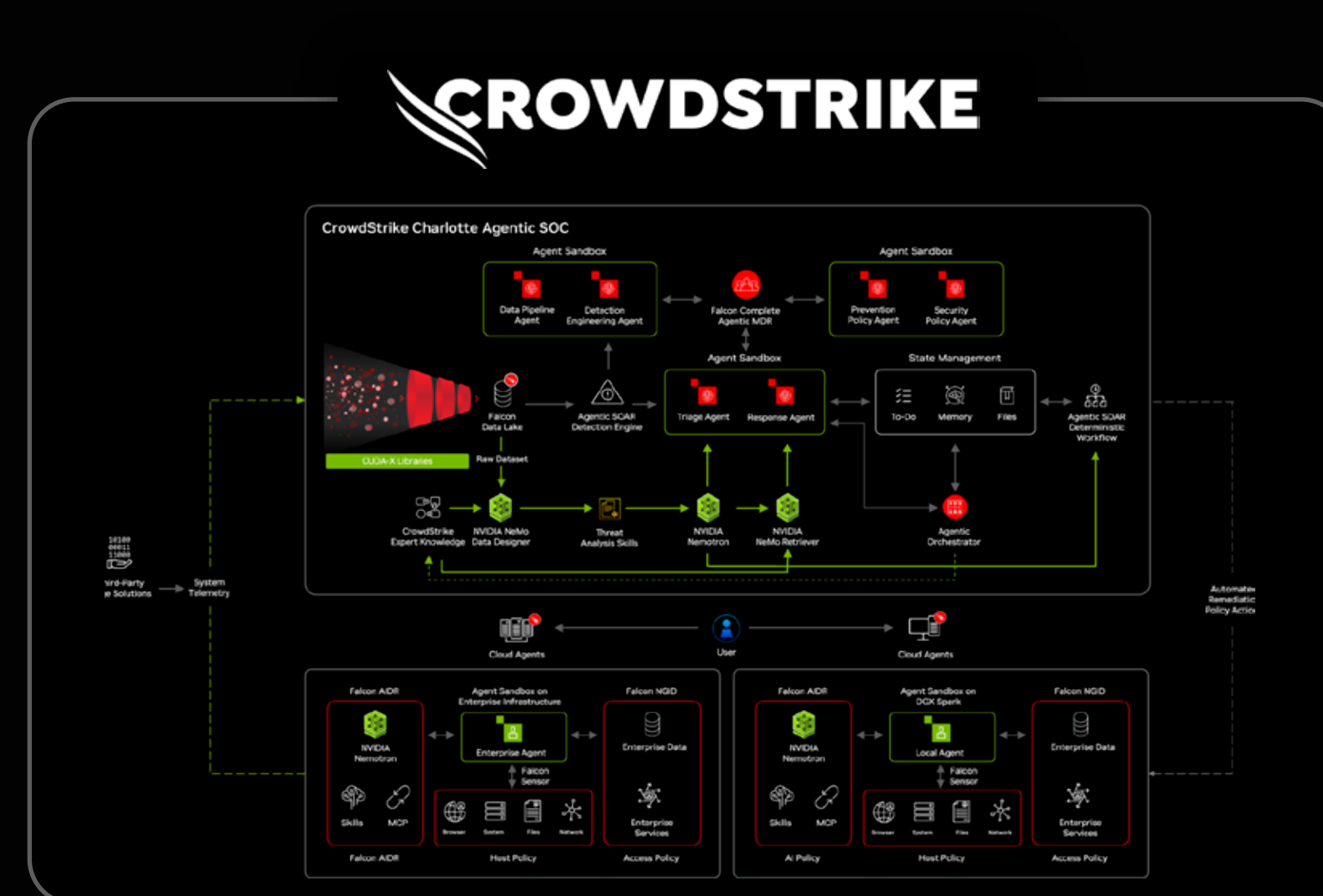
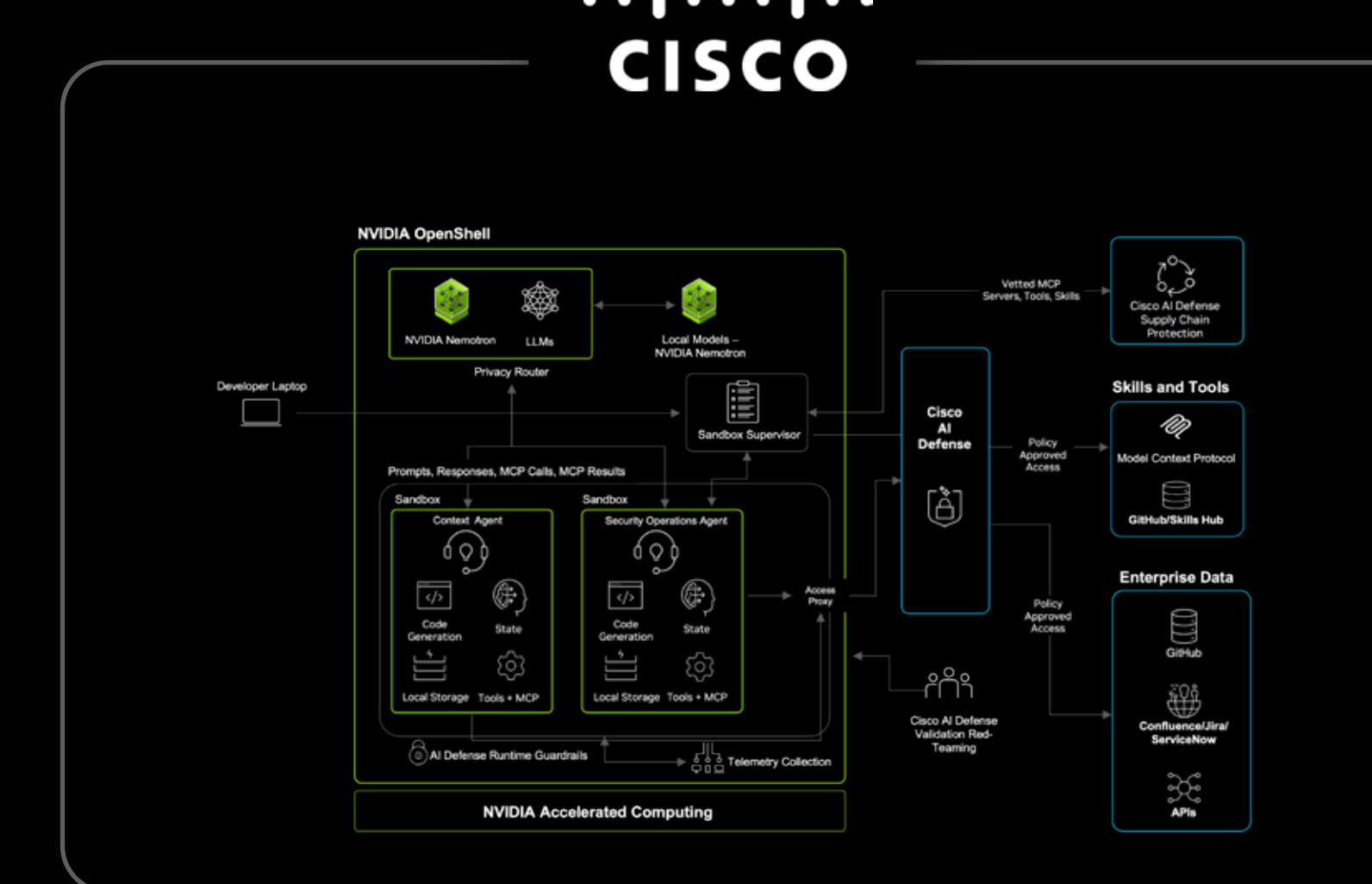
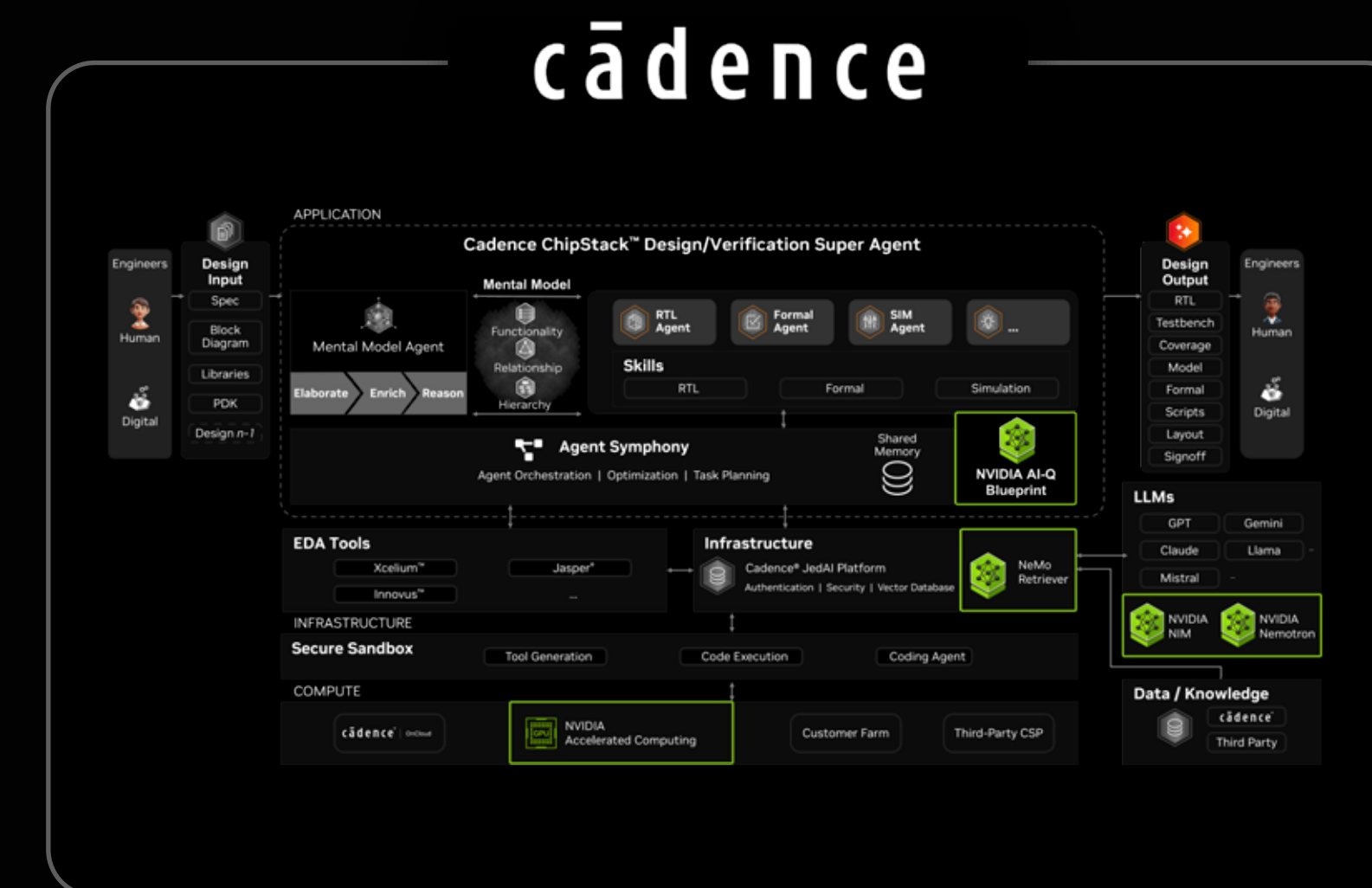
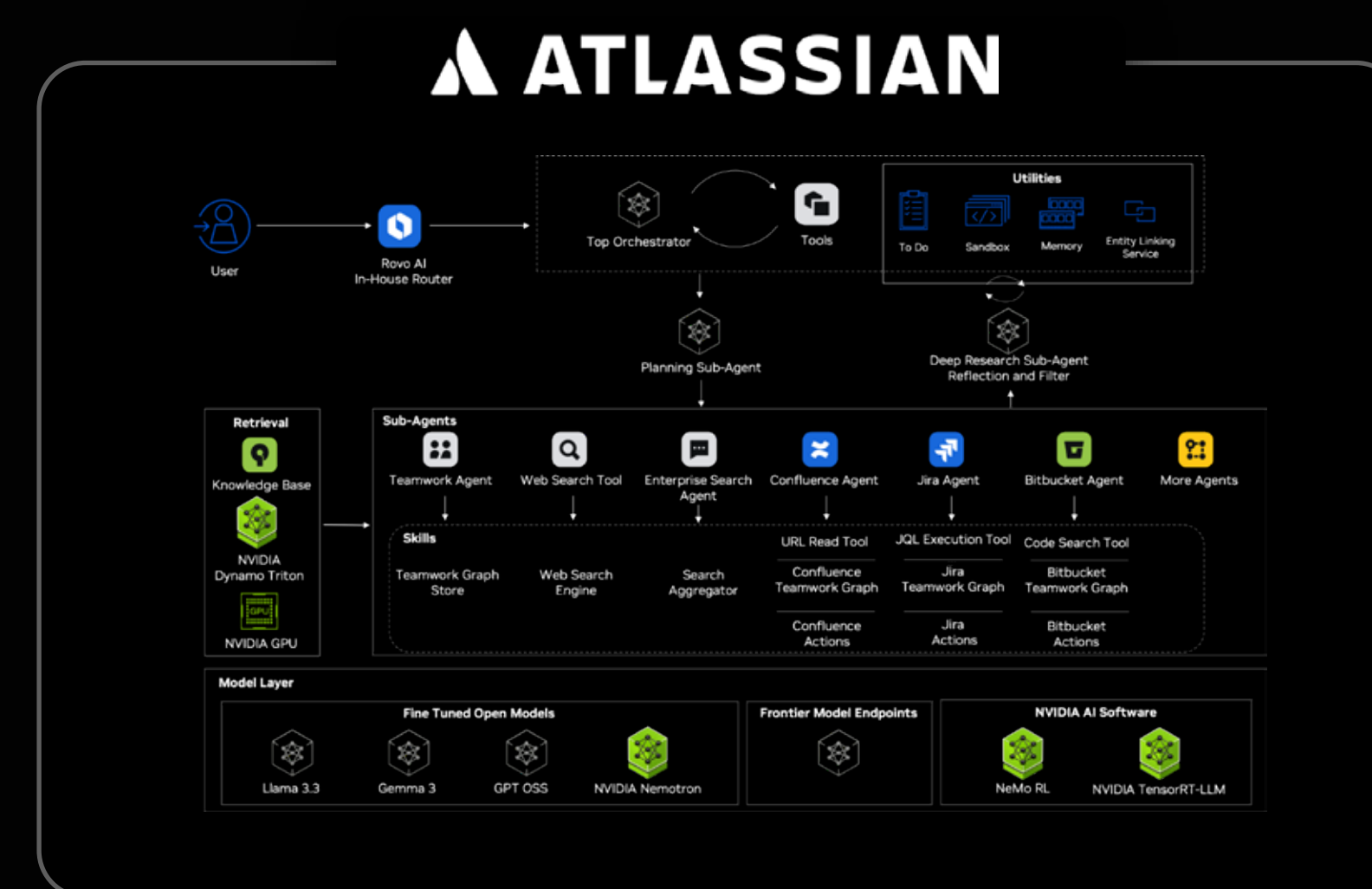
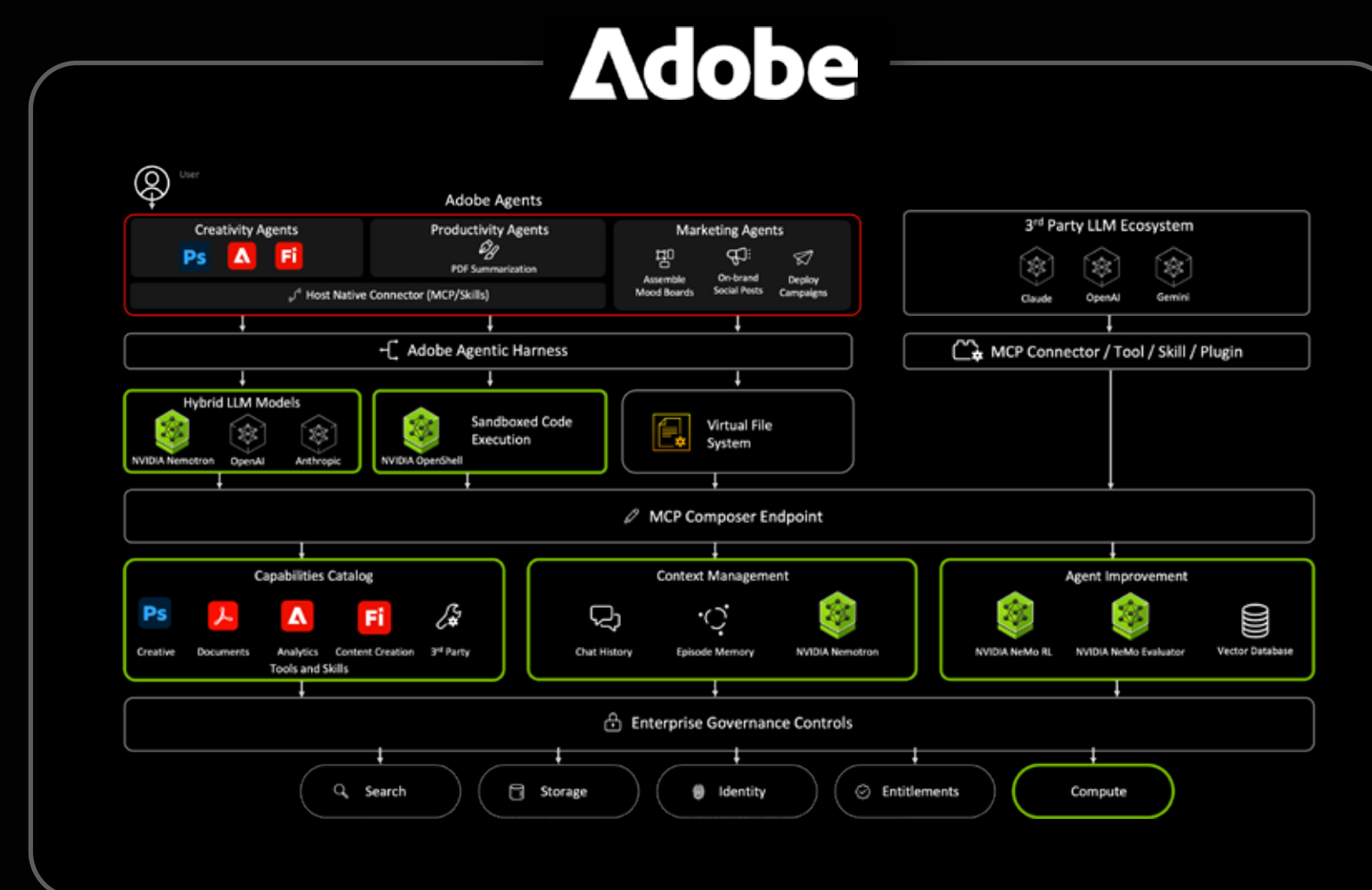


The World Building Regional AI With NVIDIA Nemotron

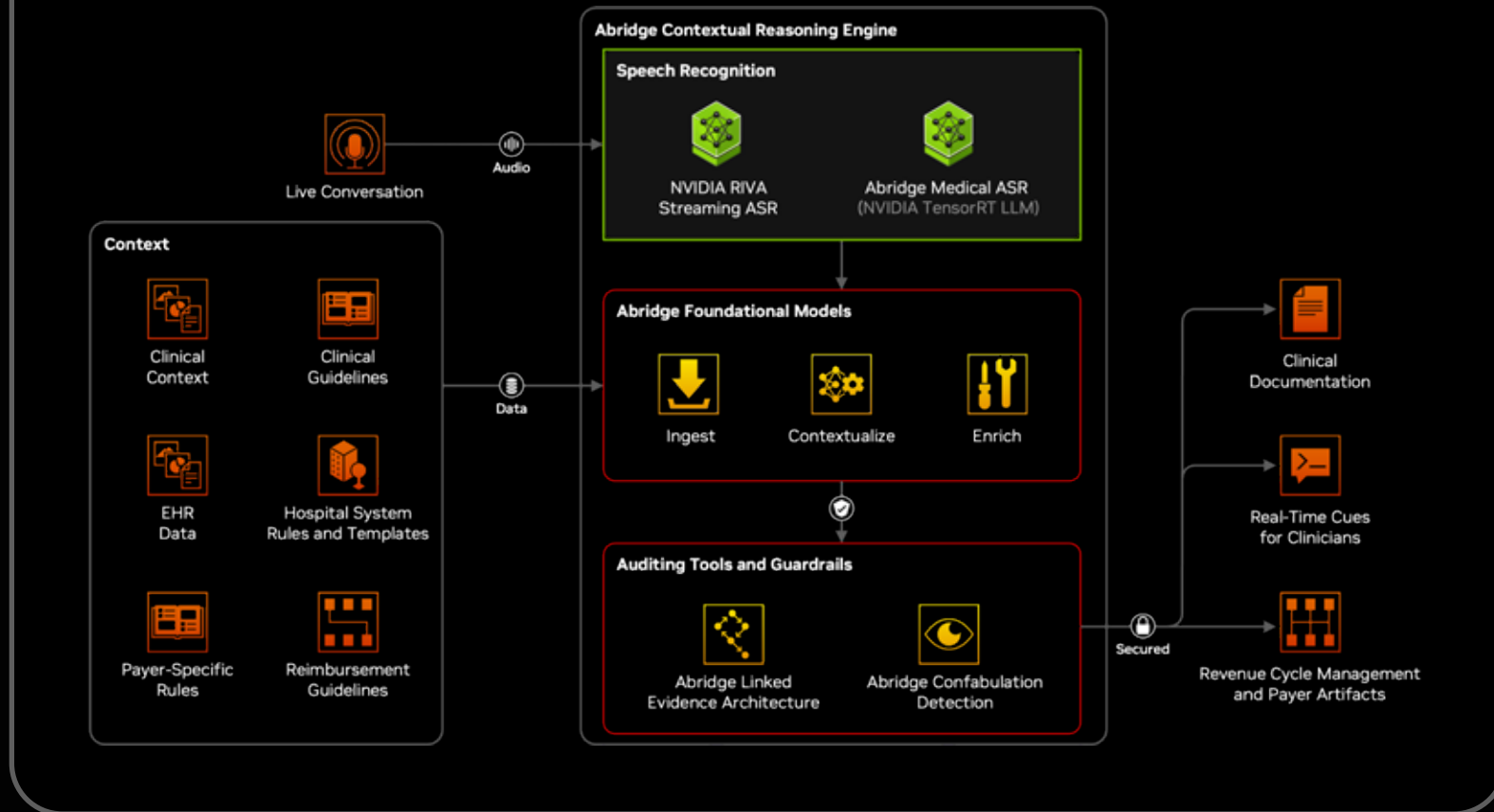


Announcing Global AI Leaders Join NVIDIA Nemotron Coalition to Advance Open Frontier Models

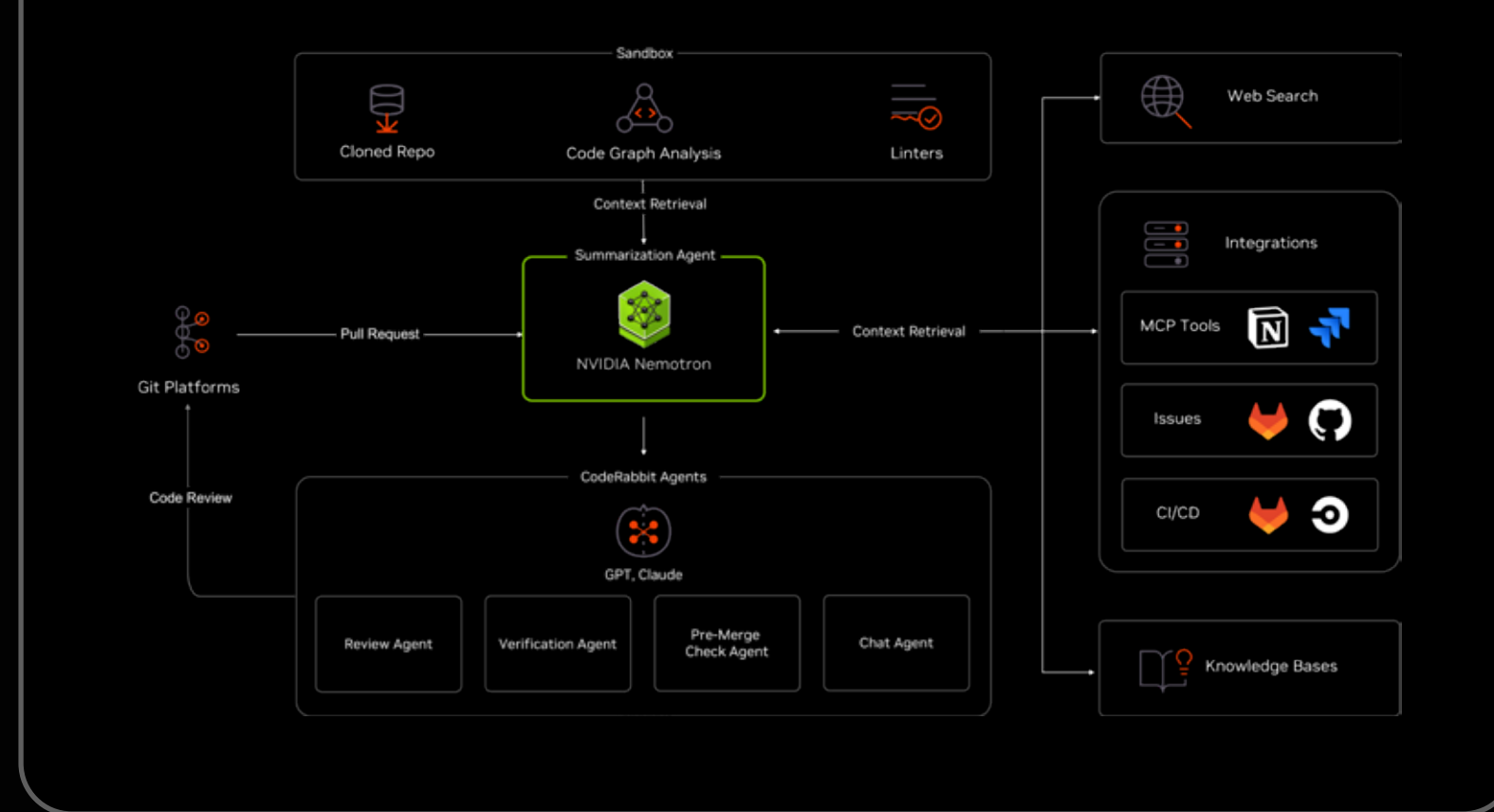




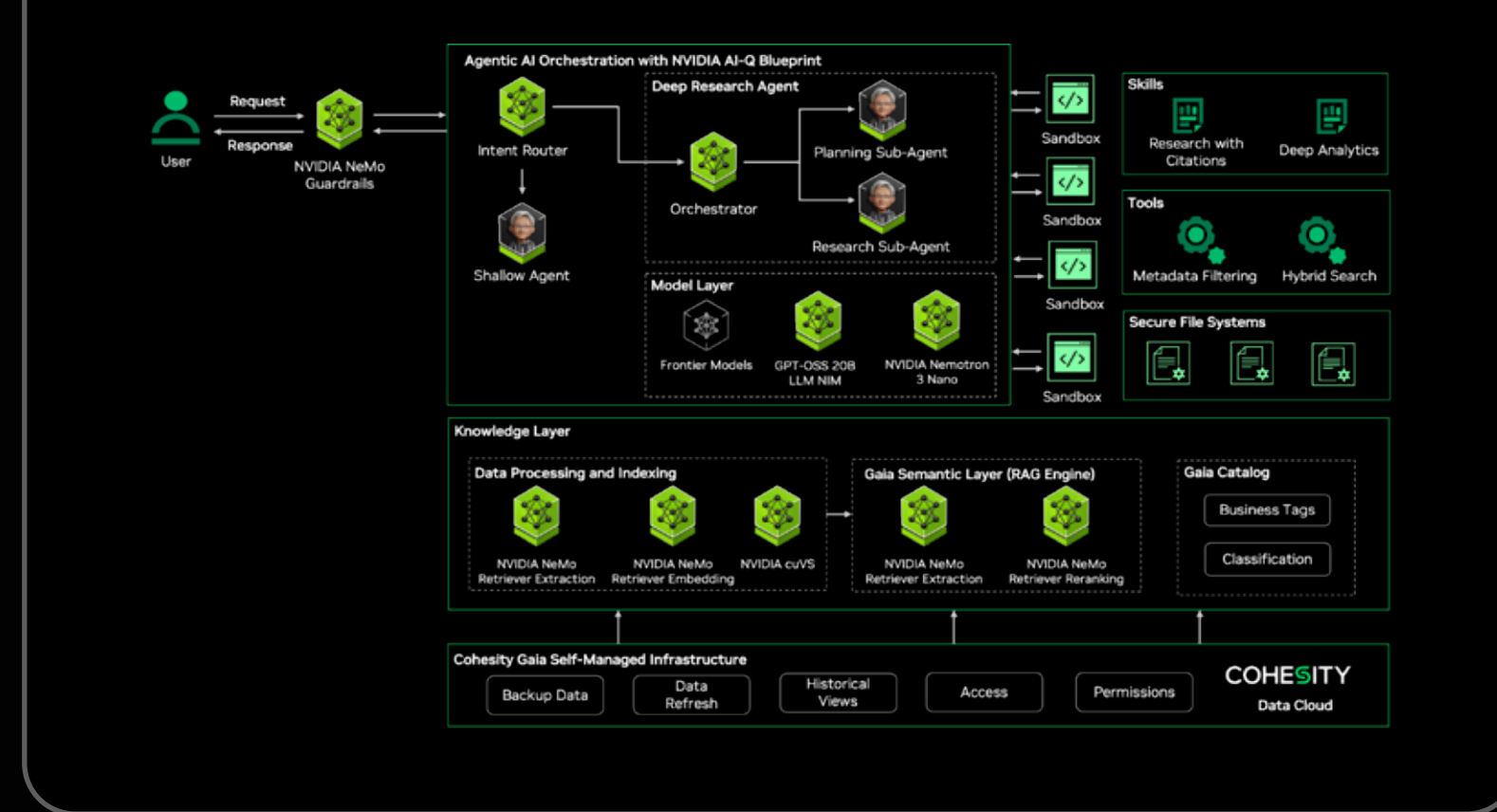
ABRIDGE



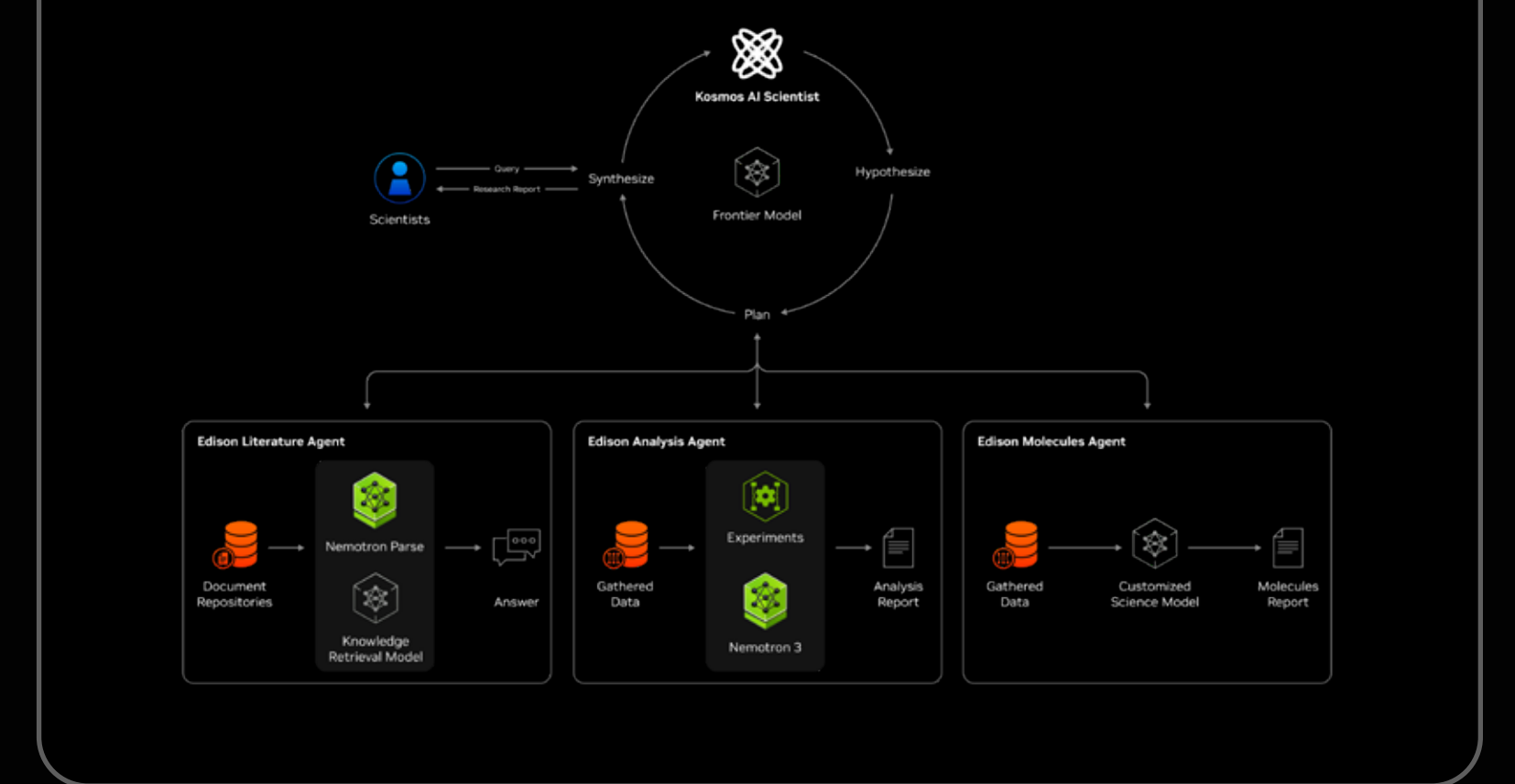
CodeRabbit



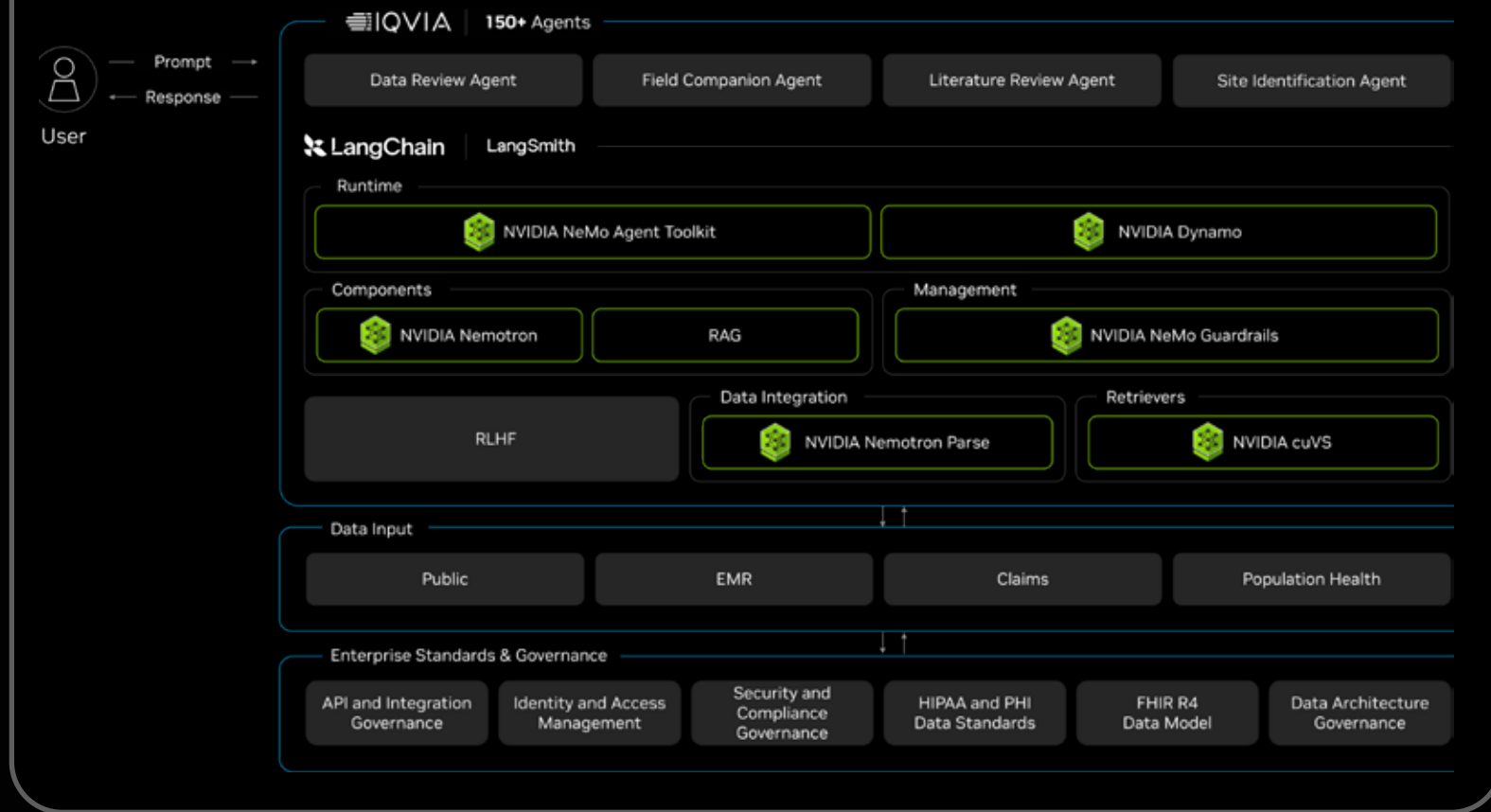
COHESITY



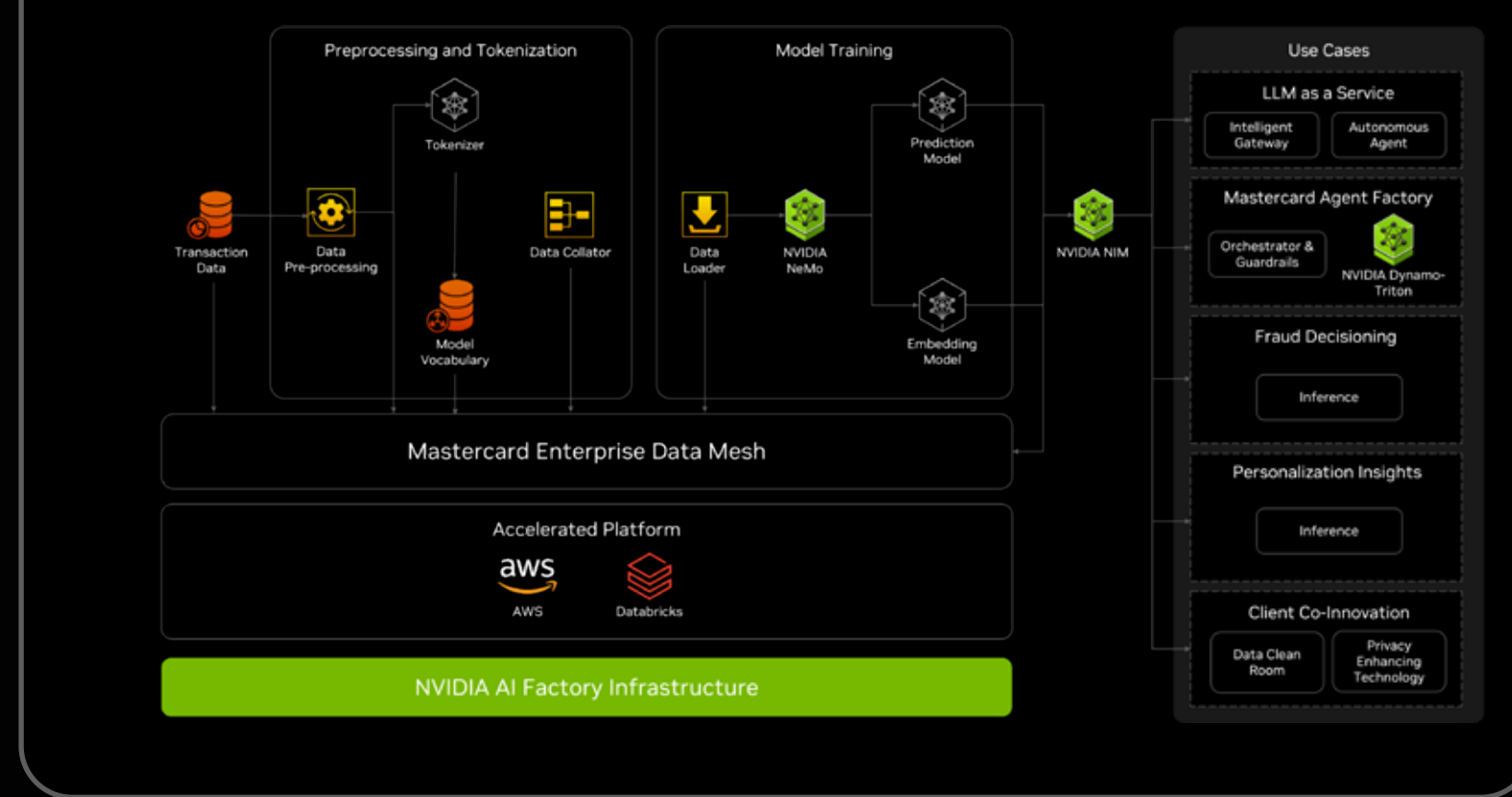
Edison



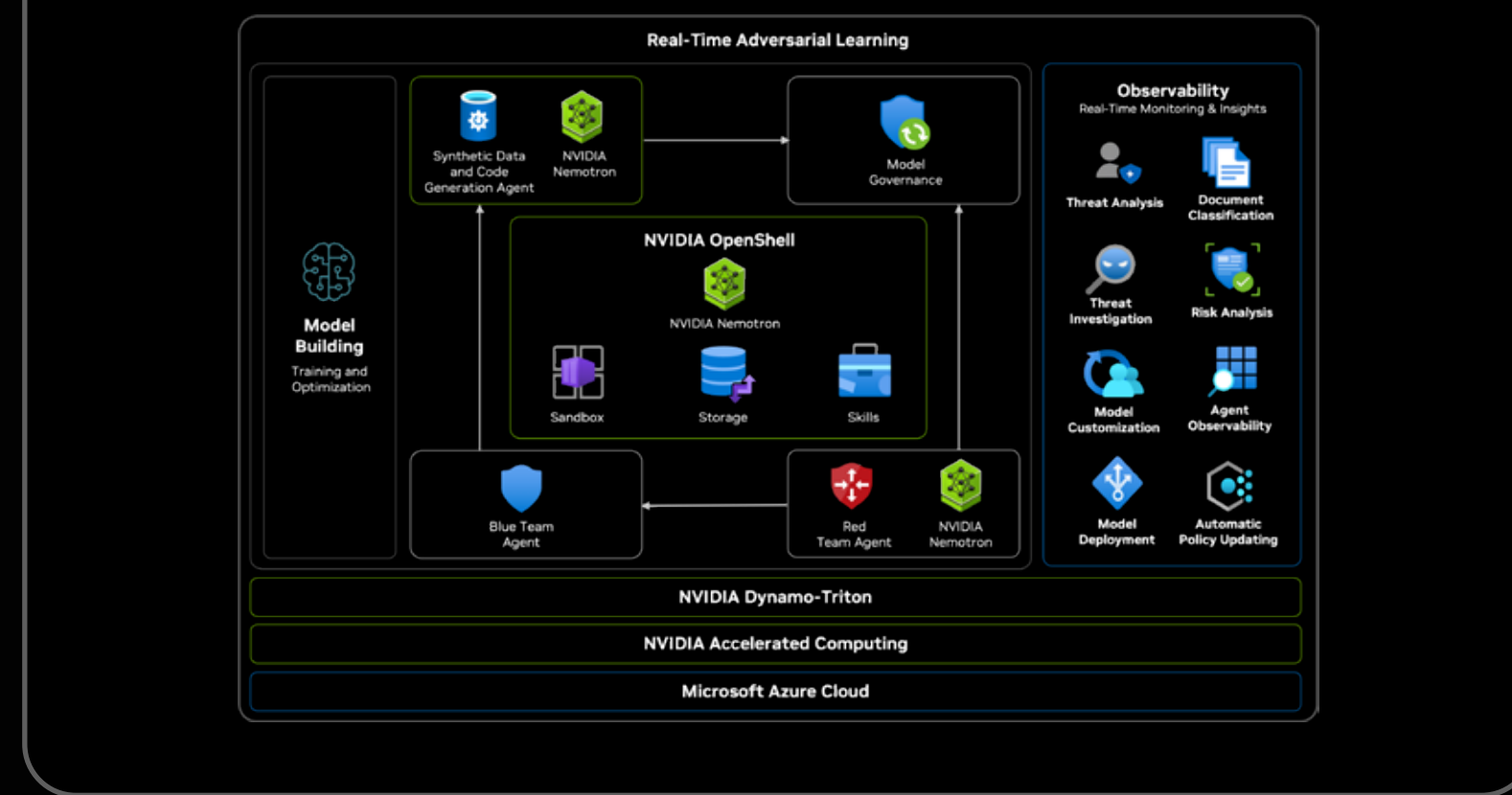
IQVIA



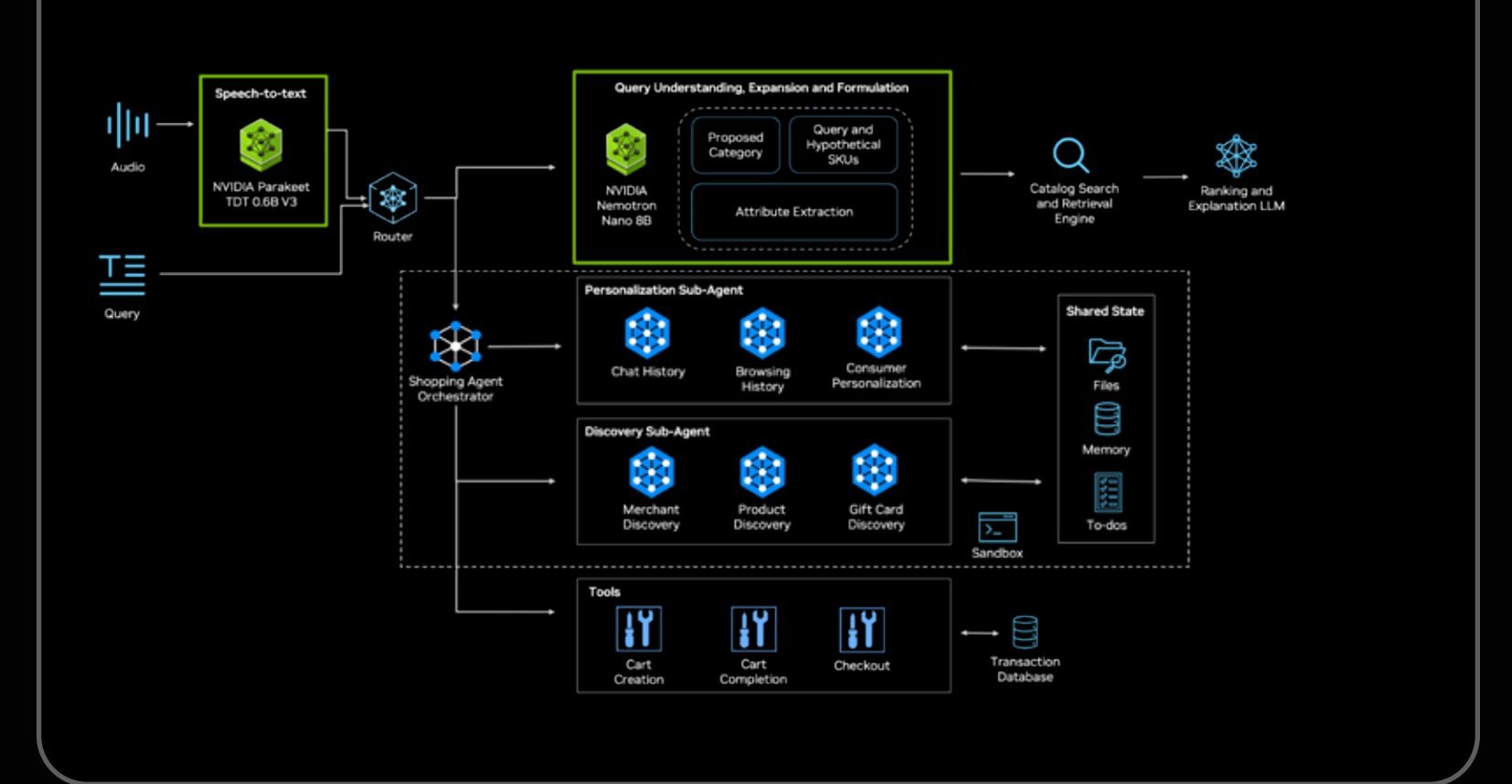
mastercard



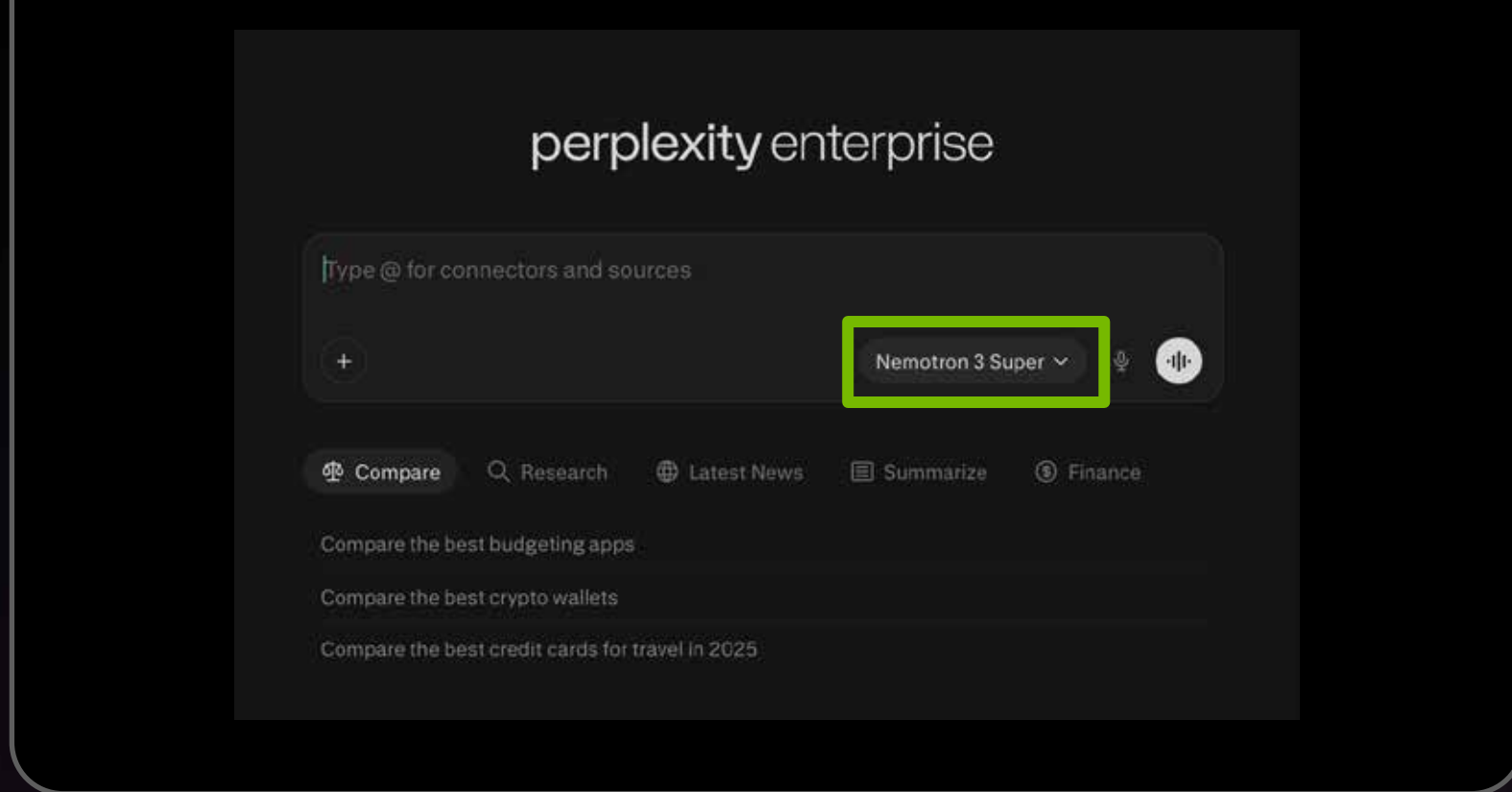
Microsoft Security NEXT AI



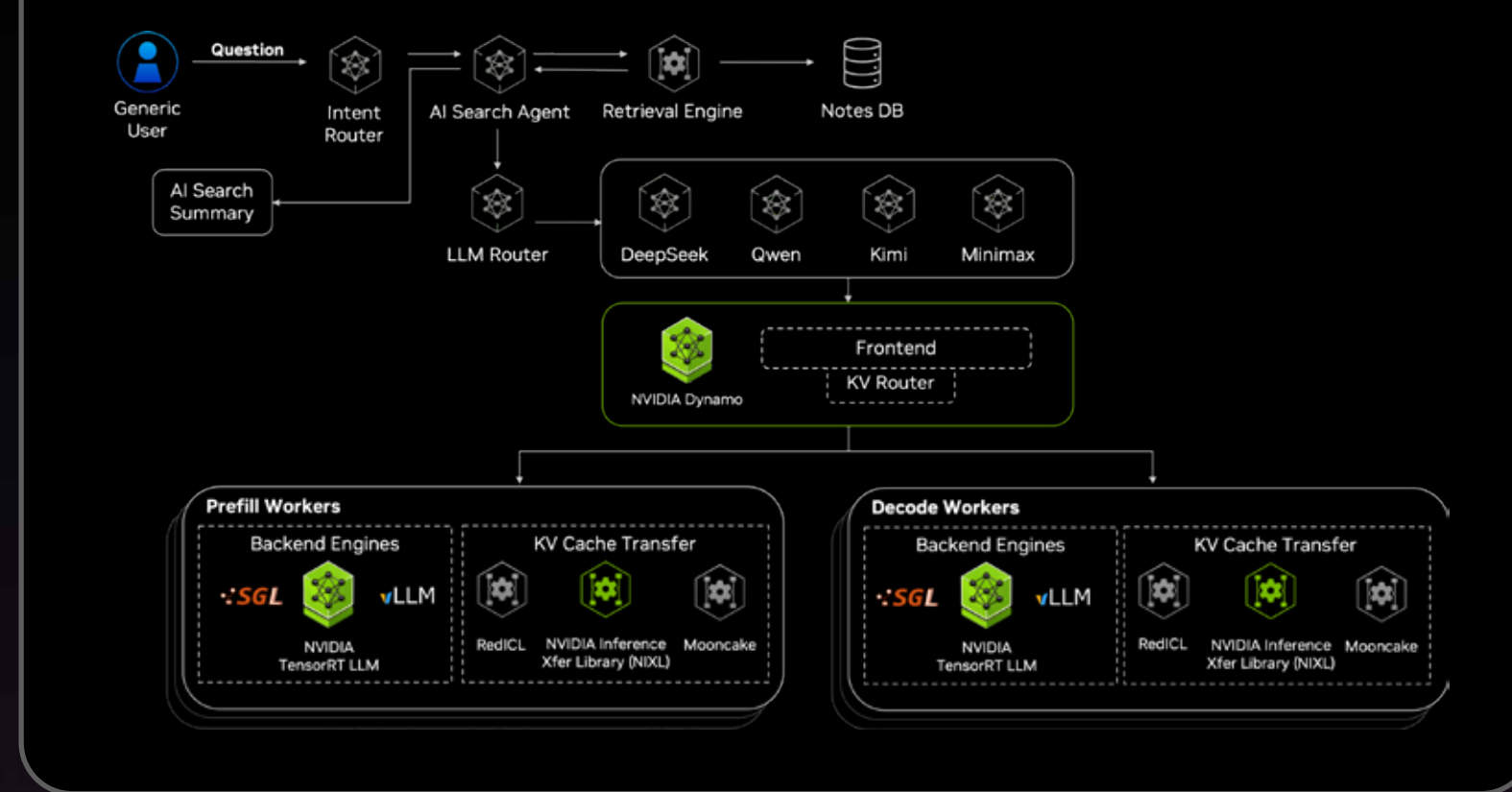
PayPal



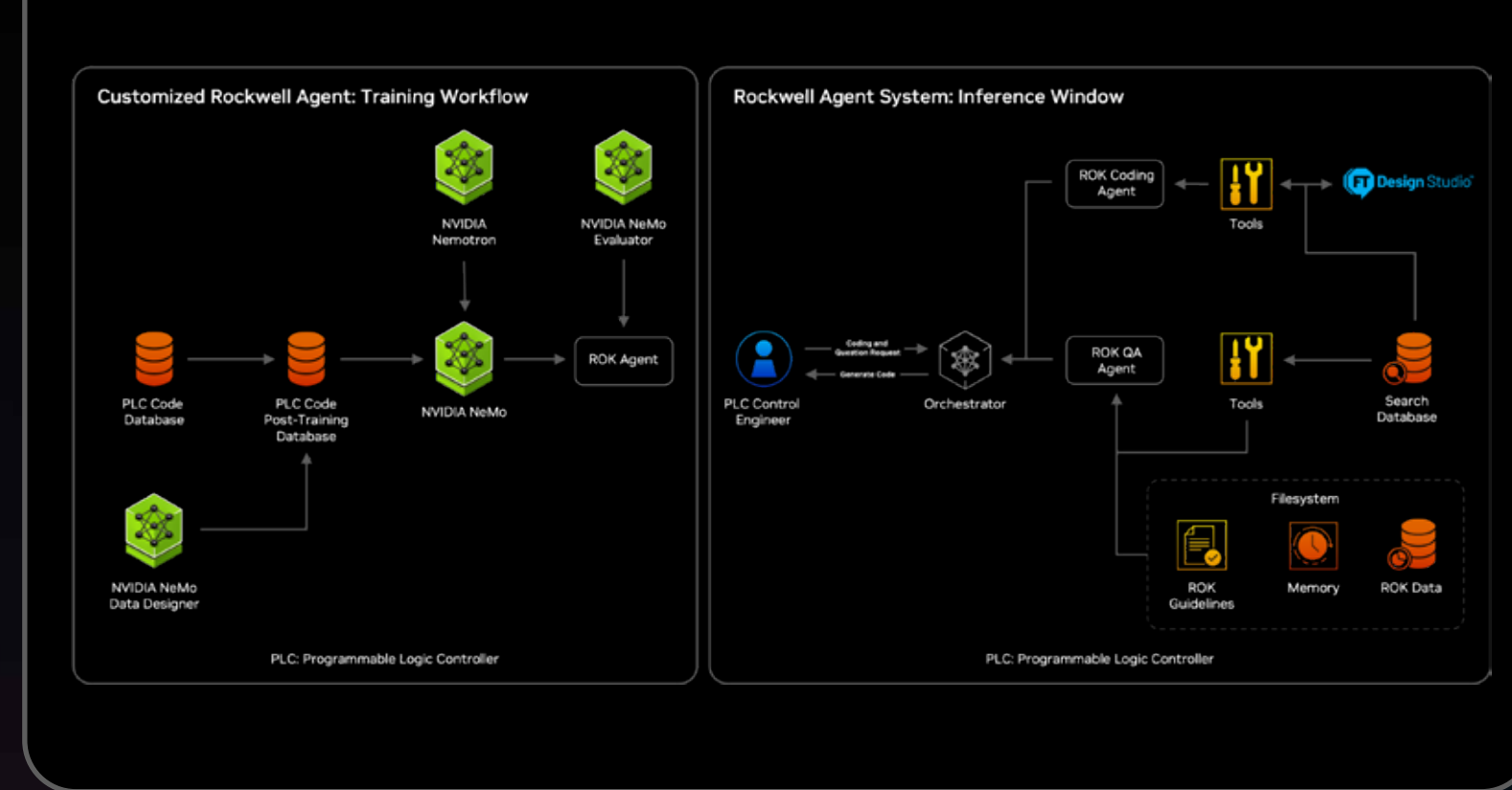
perplexity



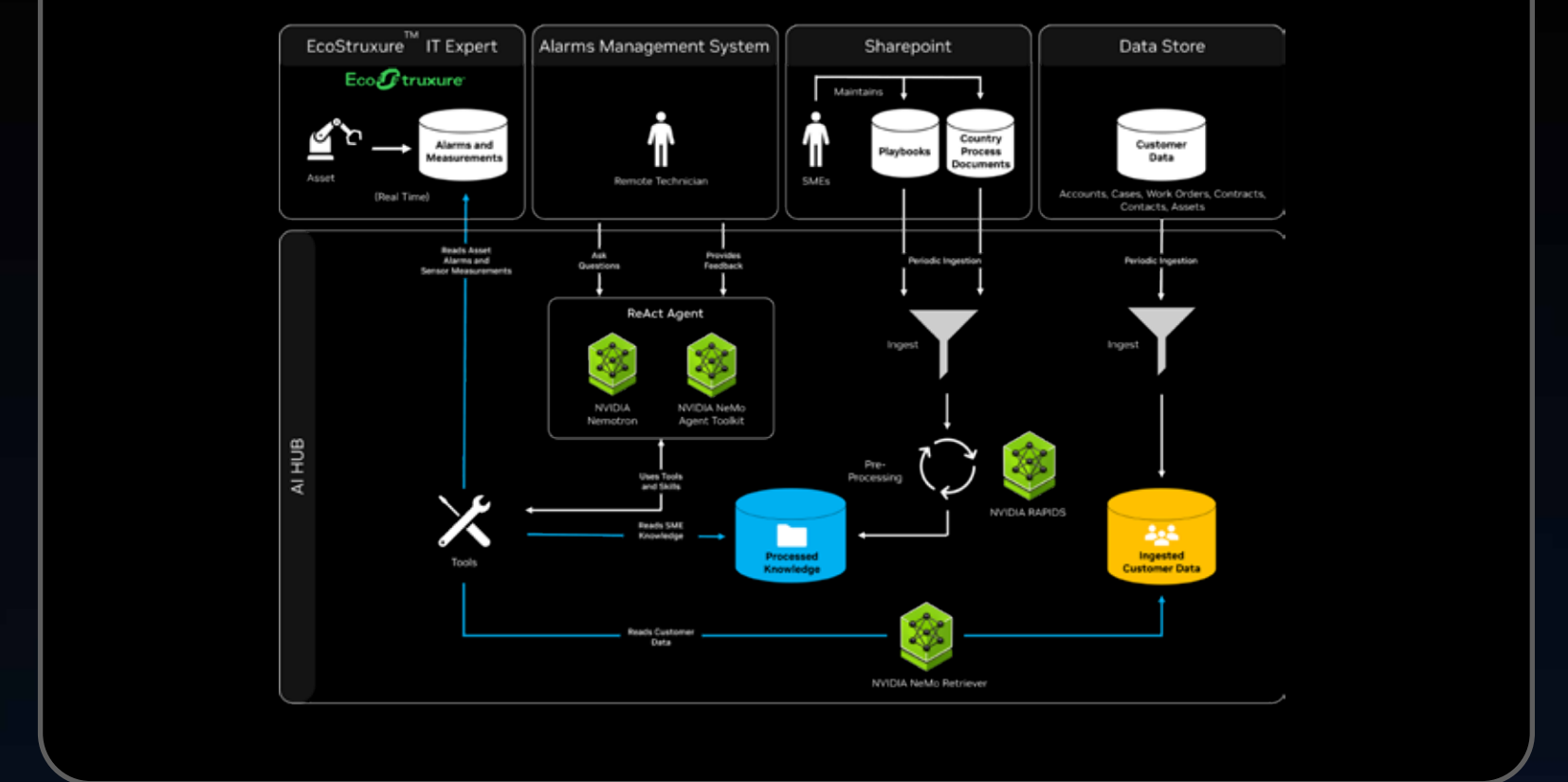
小红书



Rockwell Automation



Schneider Electric



Alibaba

Automation ANYWHERE

AWS

baseten

ByteDance

CLOUDIAN

crewai

Cognition

CoreWeave

DELL Technologies

ddn

Everpure

FACTORY

Google Cloud

G-CORE

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

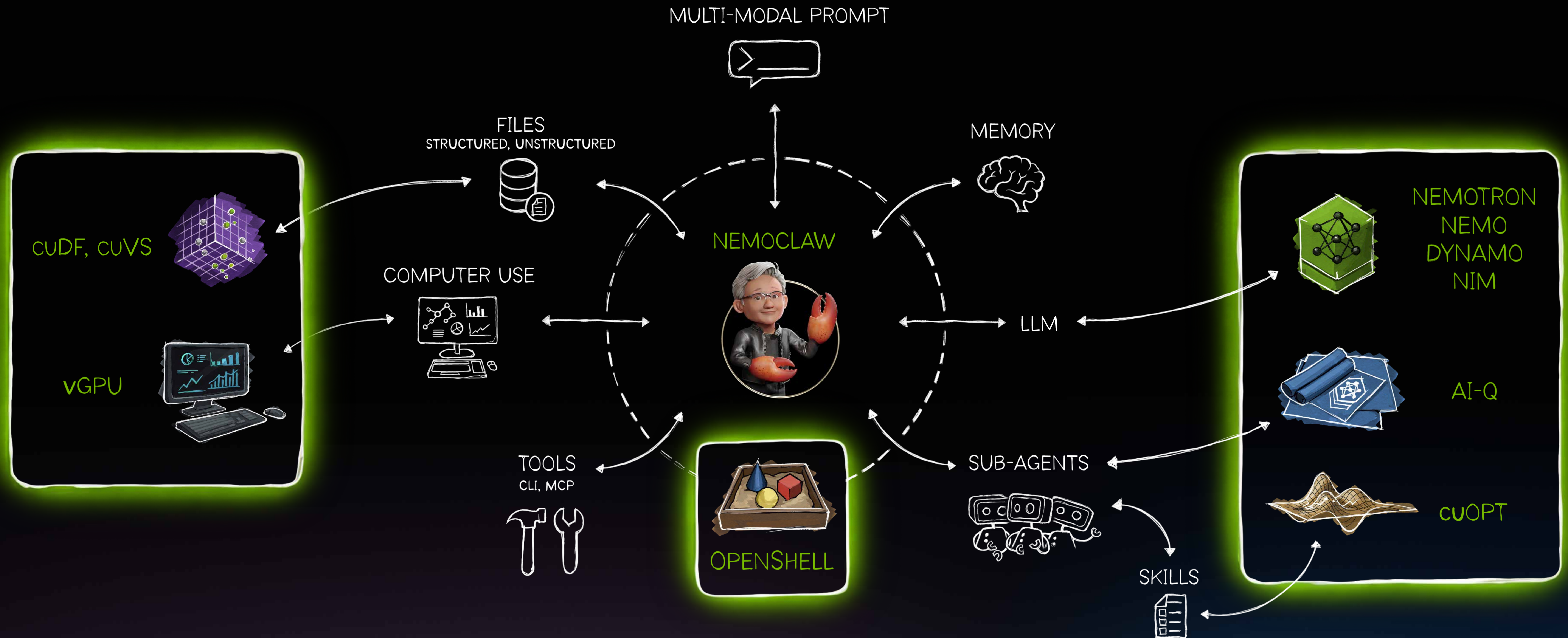
VAST

VULTR

WEKA

Announcing NVIDIA NemoClaw Reference OpenClaw

NVIDIA Agent Toolkit for Building Specialized Agents





Caterpillar

Mentee

Persona AI

NEURA Robotics

Franka Robotics

Agile Robots

AgilBot

RobotERA

T-Mobile

Zipline

NVIDIA

ABB Robotics

FAN

Linde

Universal Robots

Humanoid

Diligent

XPENG

Boston Dynamics

Figure

Fauna Robotics

Foxconn Fil

Agility

TX

LG Electronics

Hexagon Robotics

AgilBot

Mercedes-Benz

Enchanted Tools

Pollen Robotics

Idealworks

Serve Robotics

Disney Research

