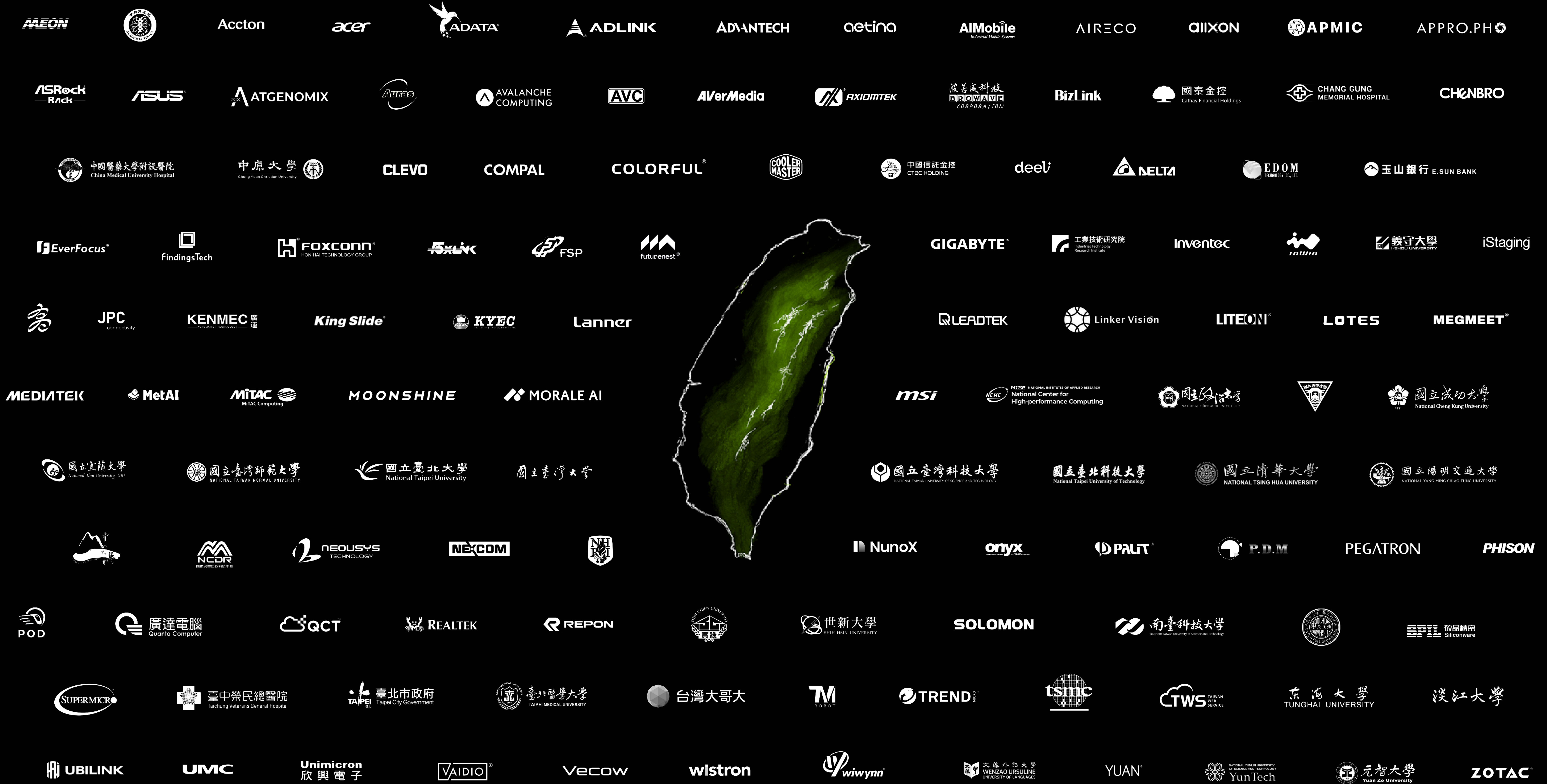


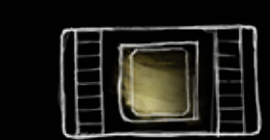




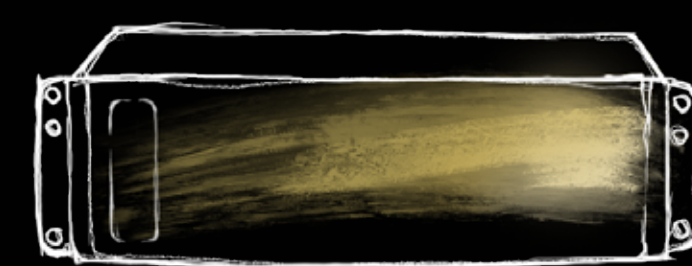




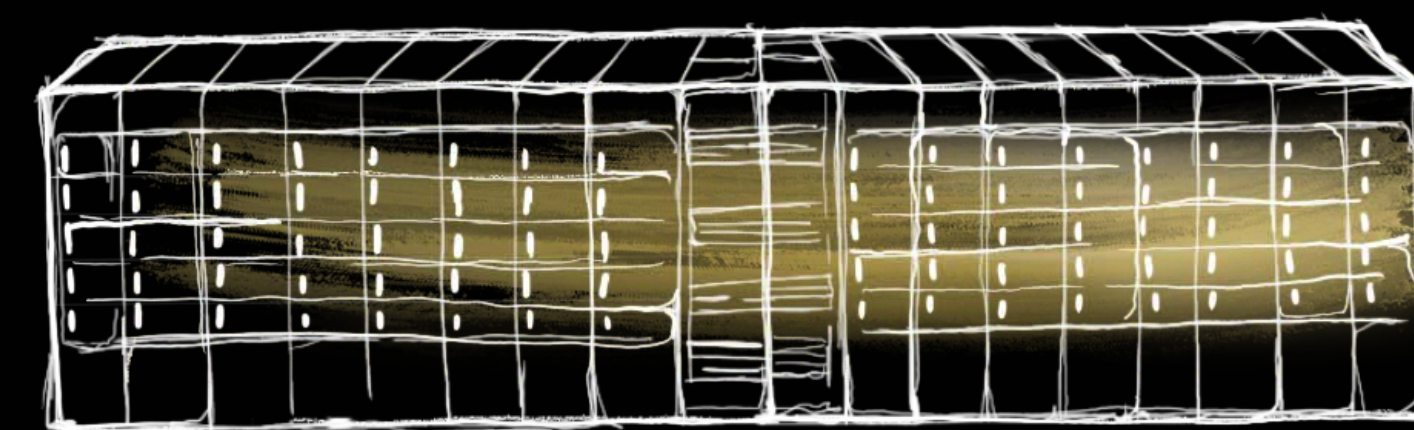
# NVIDIA Transformation From GPU Chip to AI Infrastructure



CHIP



SYSTEM

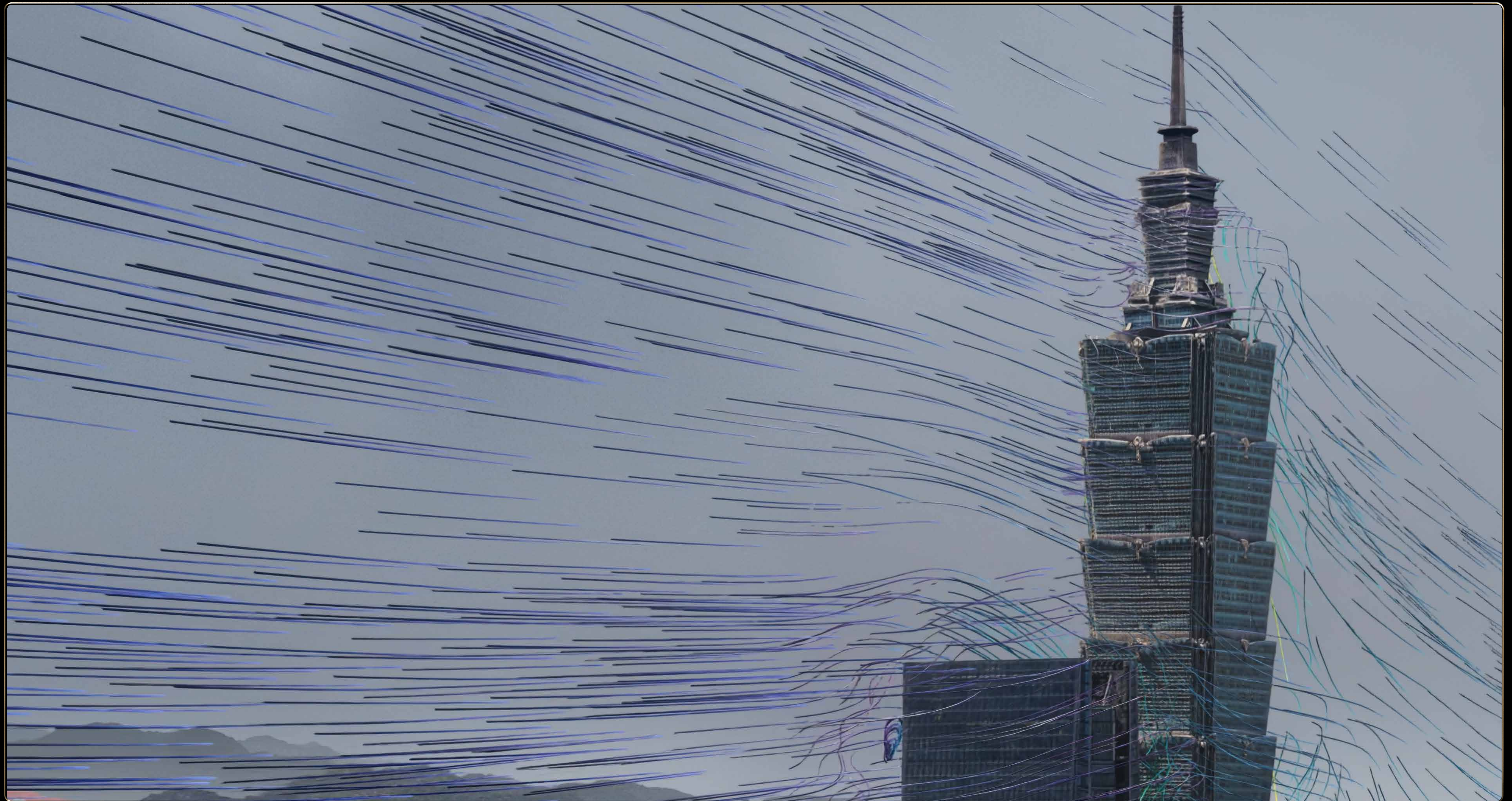


DATA CENTER



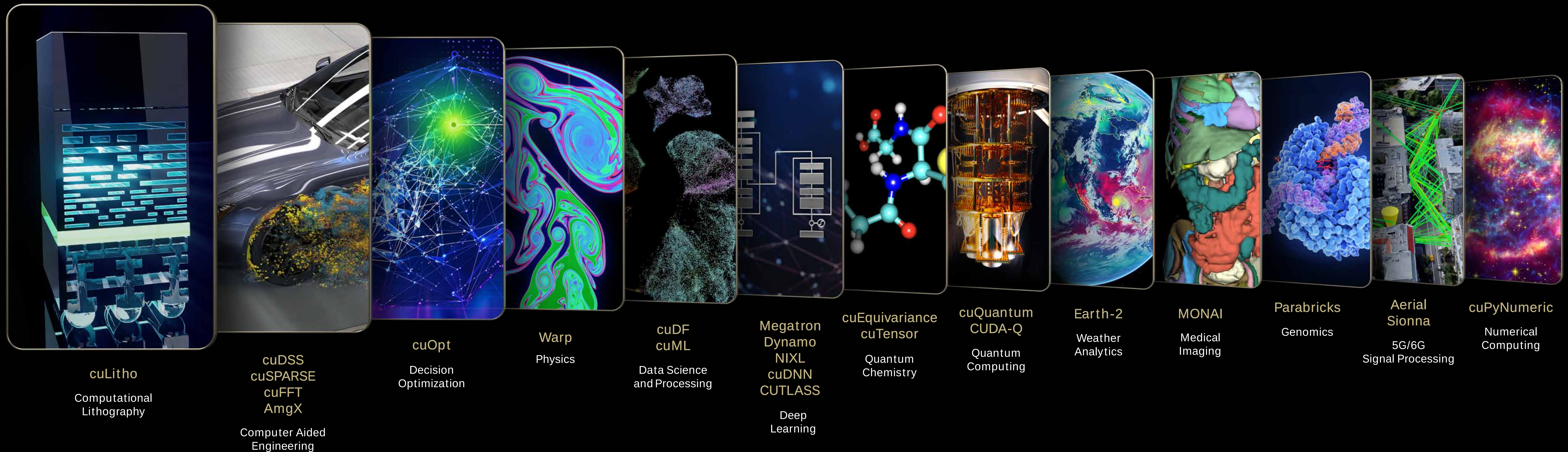
INFRASTRUCTURE





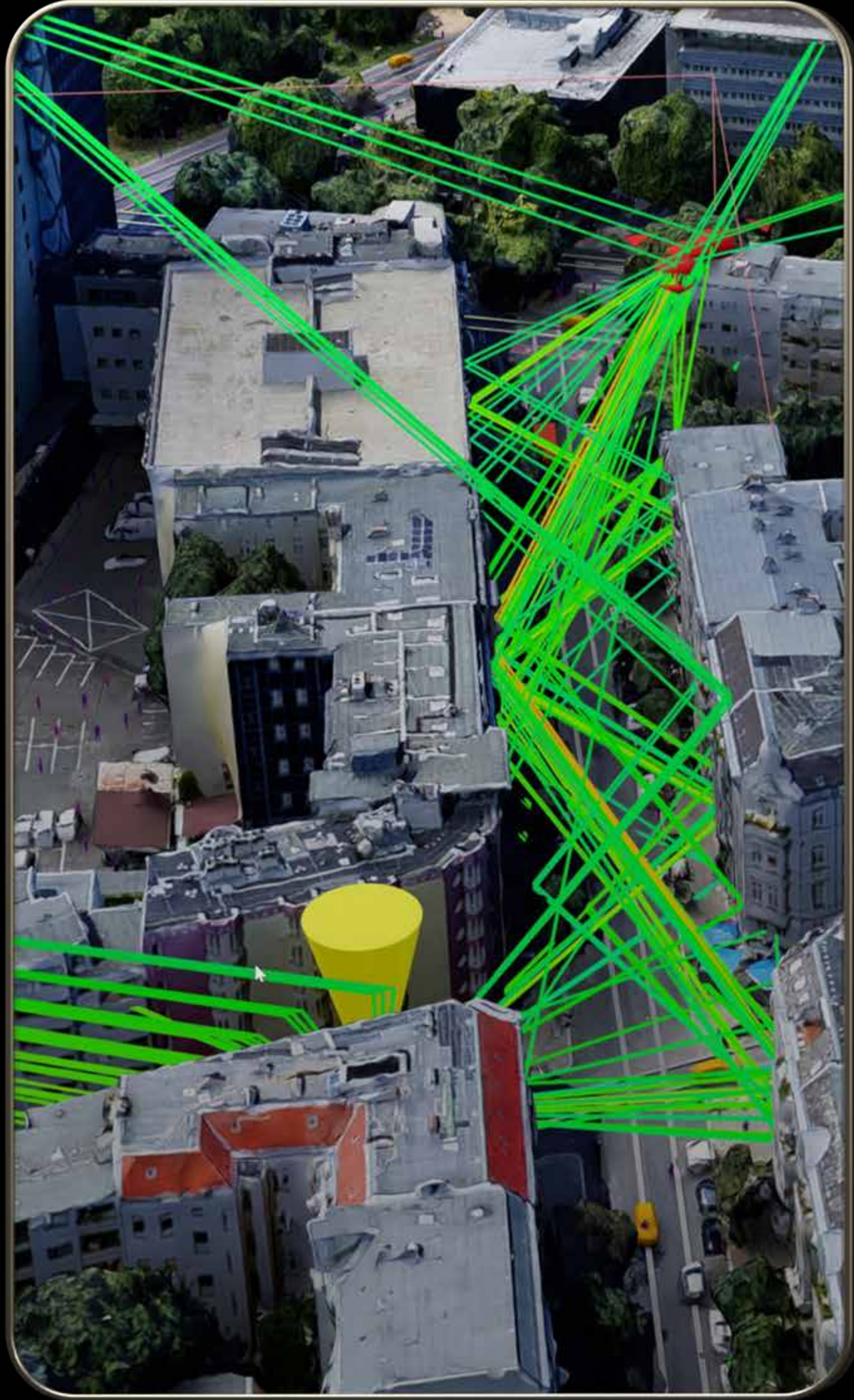


# Accelerated Computing Begins With CUDA-X Libraries





# NVIDIA Building 6G on AI



Aerial  
Sionna  
5G/6G  
Signal Processing



NVIDIA Grace CPU  
NVIDIA PCIe GPU  
NVIDIA ConnectX-7  
Compact AI-RAN

SoftBank

T Mobile

indosat  
OOREDOO HUTCHISON

vodafone

NOKIA

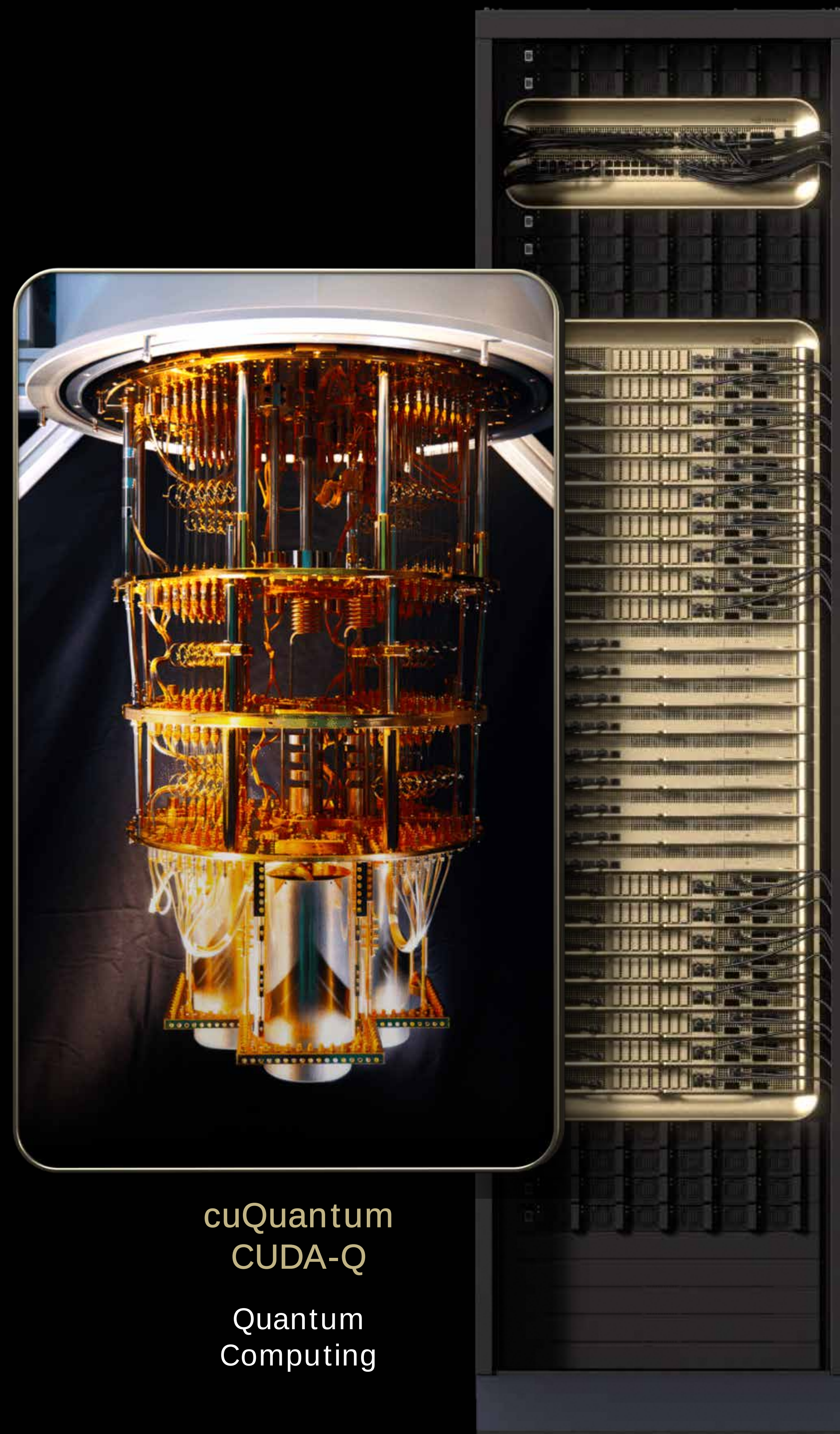
SAMSUNG

KYOCERA

FUJITSU

CISCO





cuQuantum  
CUDA-Q  
Quantum  
Computing

GB200 NVL72

## NVIDIA Building Hybrid Quantum-GPU Computing



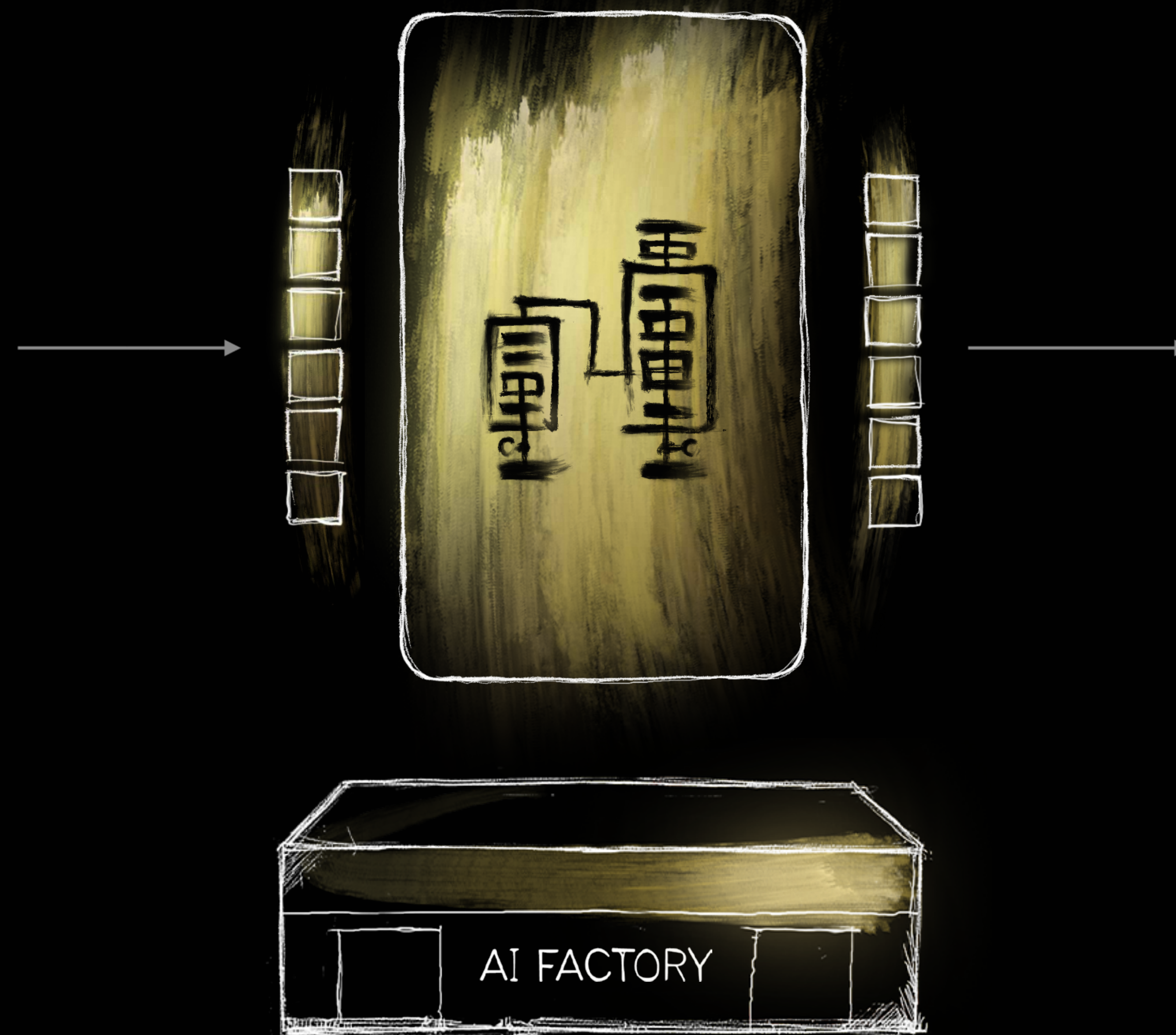


# Agentic AI

## “Understand-Think-Act” is 100-1000X More Compute Than One-Shot

你是我在台北的行程規劃師。

請從台北流行音樂中心出發，規劃一個三小時（18:30–21:30）的行程，捷運加步行，造訪台北五大夜市中的至少三個，品席六種不同的台灣特色街頭小吃，其中一種需來自米其林必比登或米其林入選的攤位。請列出詳細的時間表，包括抵達時間、攤位名稱、菜品、捷運路線與站名、步行距離、路線說明和每個資訊的出處。最後，請以約50個字摘要此行程的亮點。



### Reasoning

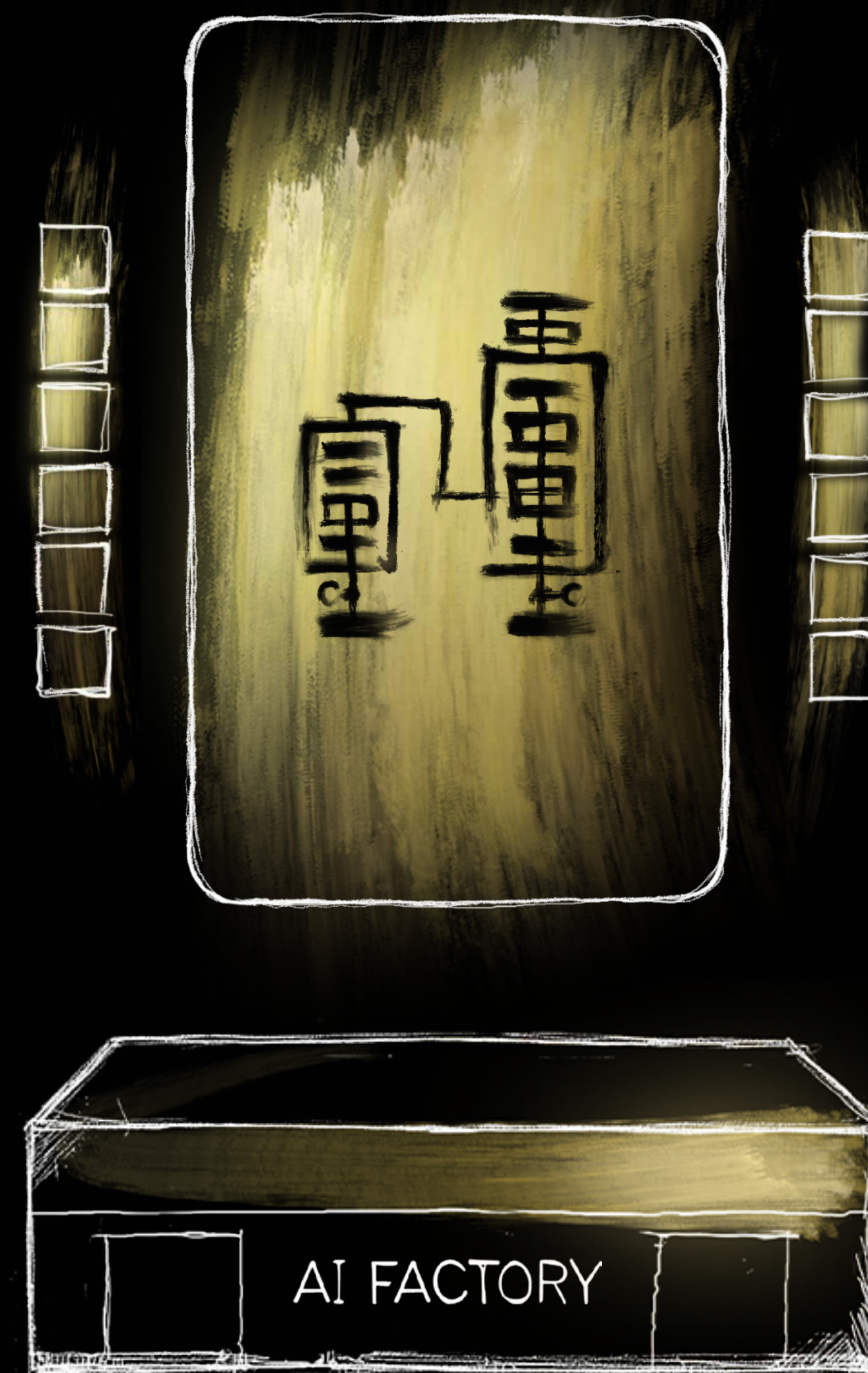
Okay, let's tackle this. The user needs three search queries to help plan the sections of their Taiwan night market food tour report. The tour starts at Taipei Music Center, let's say it takes the MRT to Raohe, then Ningxia, and ends at Shilin by 21:30. Each stop has specific iconic foods. The report sections probably break down each market visit, the foods, timing, etc. The queries should help answer questions in the report's sections. Let's look at the plan's components: locations (markets), specific foods, timing, transportation (MRT), and ensuring the



# Physical AI

## The Foundation of Robotic Revolution

生成三個詳細且多樣化的場景，描述在接下來的五秒內可能出現在攝影機視野中的事物。



Scenario 1

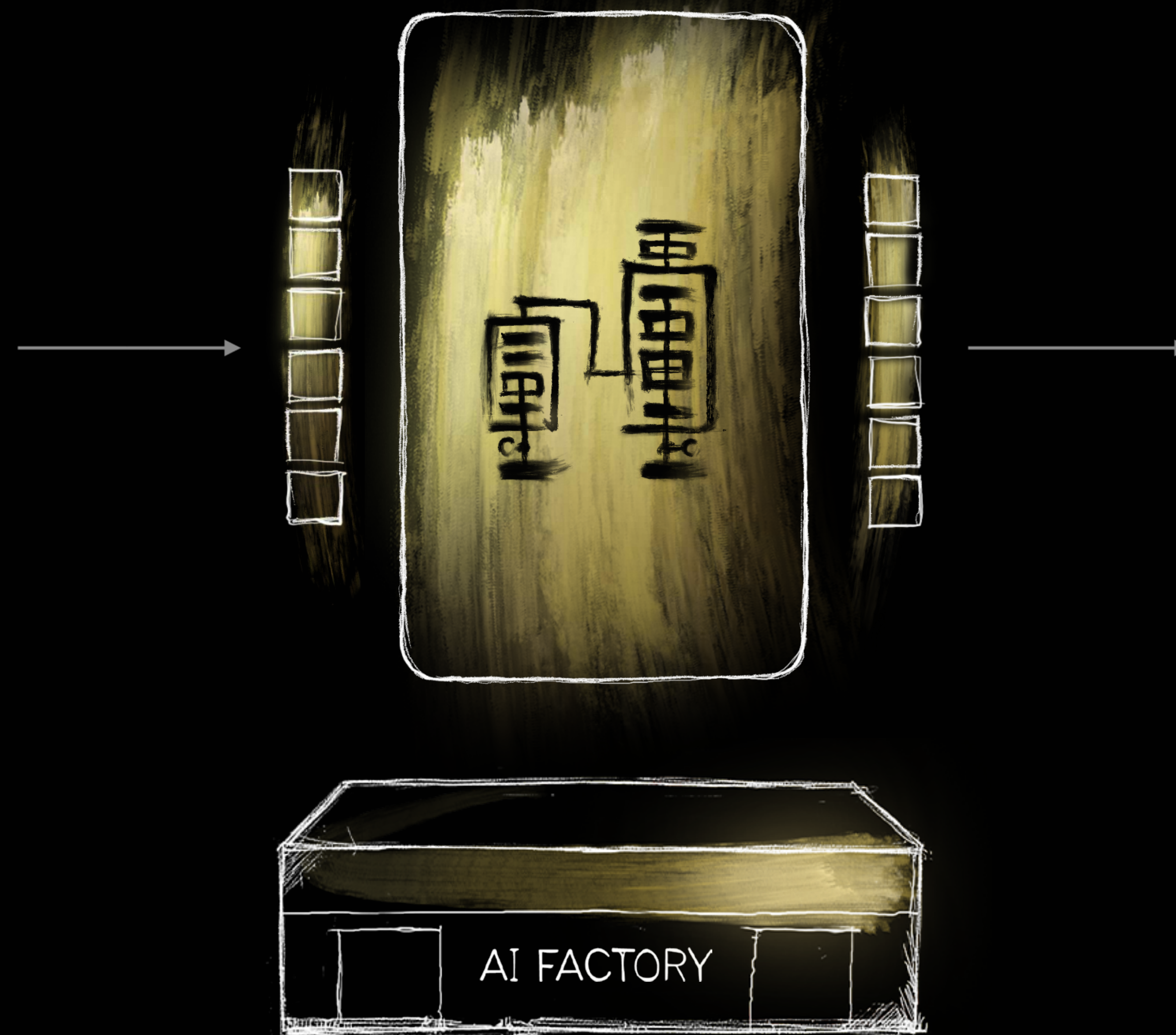
A shadow cast by a tree on the right briefly obscures a **pedestrian** stepping out from between packed cars, causing the ego driver to slow further.



# Physical AI

## The Foundation of Robotic Revolution

生成三個詳細且多樣化的場景，描述在接下來的五秒內可能出現在攝影機視野中的事物。



Scenario 2

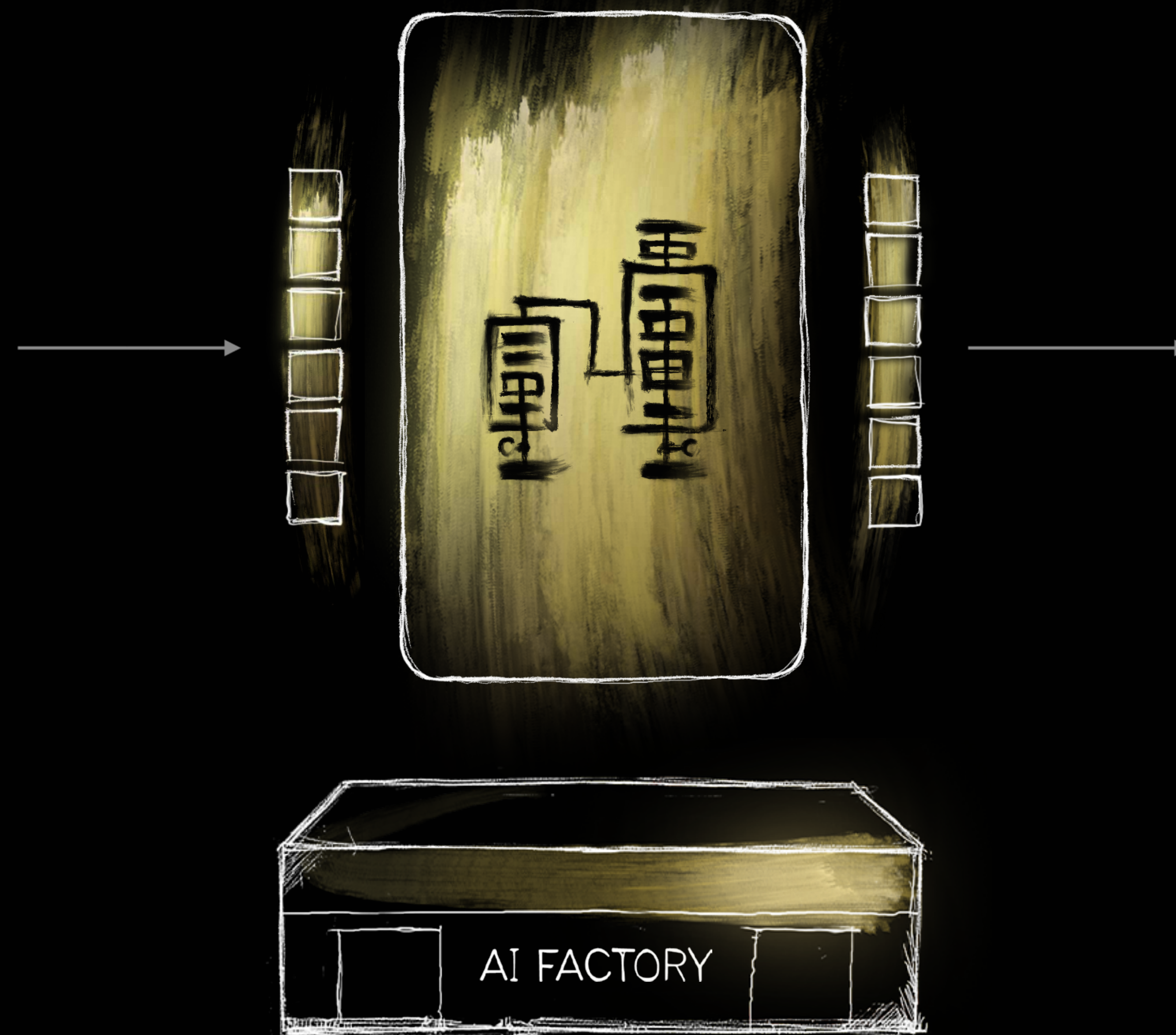
A car is stopped on a quiet suburban road, with a clear blue sky overhead and lush green trees lining the streets. A pigeon enters the frame from the right, flying low over the street with graceful, fluid movements. It gracefully glides past the parked cars, then lands softly on one of the vehicles. The scene captures a serene moment, with the pigeon's flight adding a touch of life to the tranquil neighborhood setting. A wide-angle shot from inside the car captures both the bird and the surrounding environment.



# Physical AI

## The Foundation of Robotic Revolution

生成三個詳細且多樣化的場景，描述在接下來的五秒內可能出現在攝影機視野中的事物。



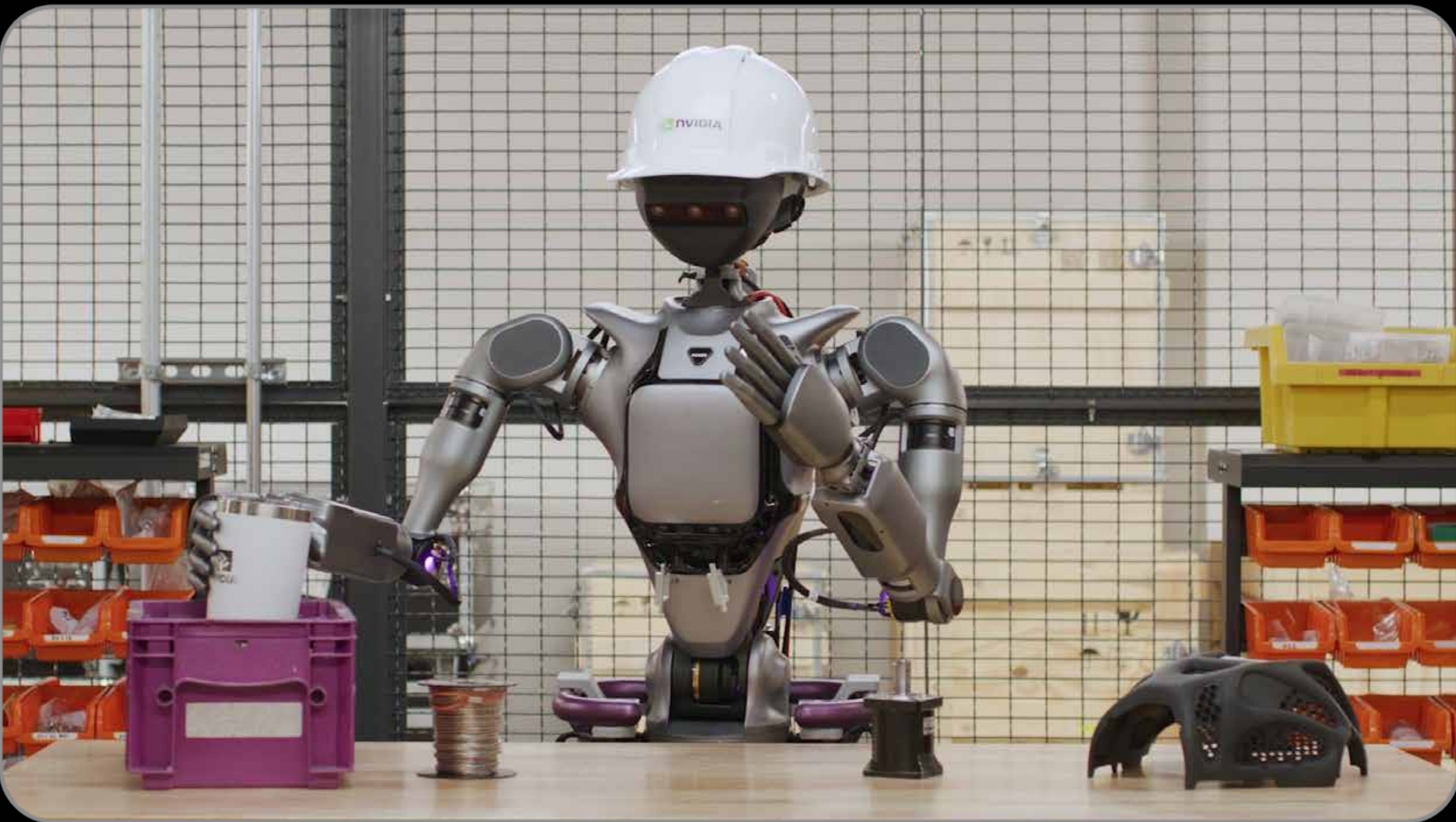
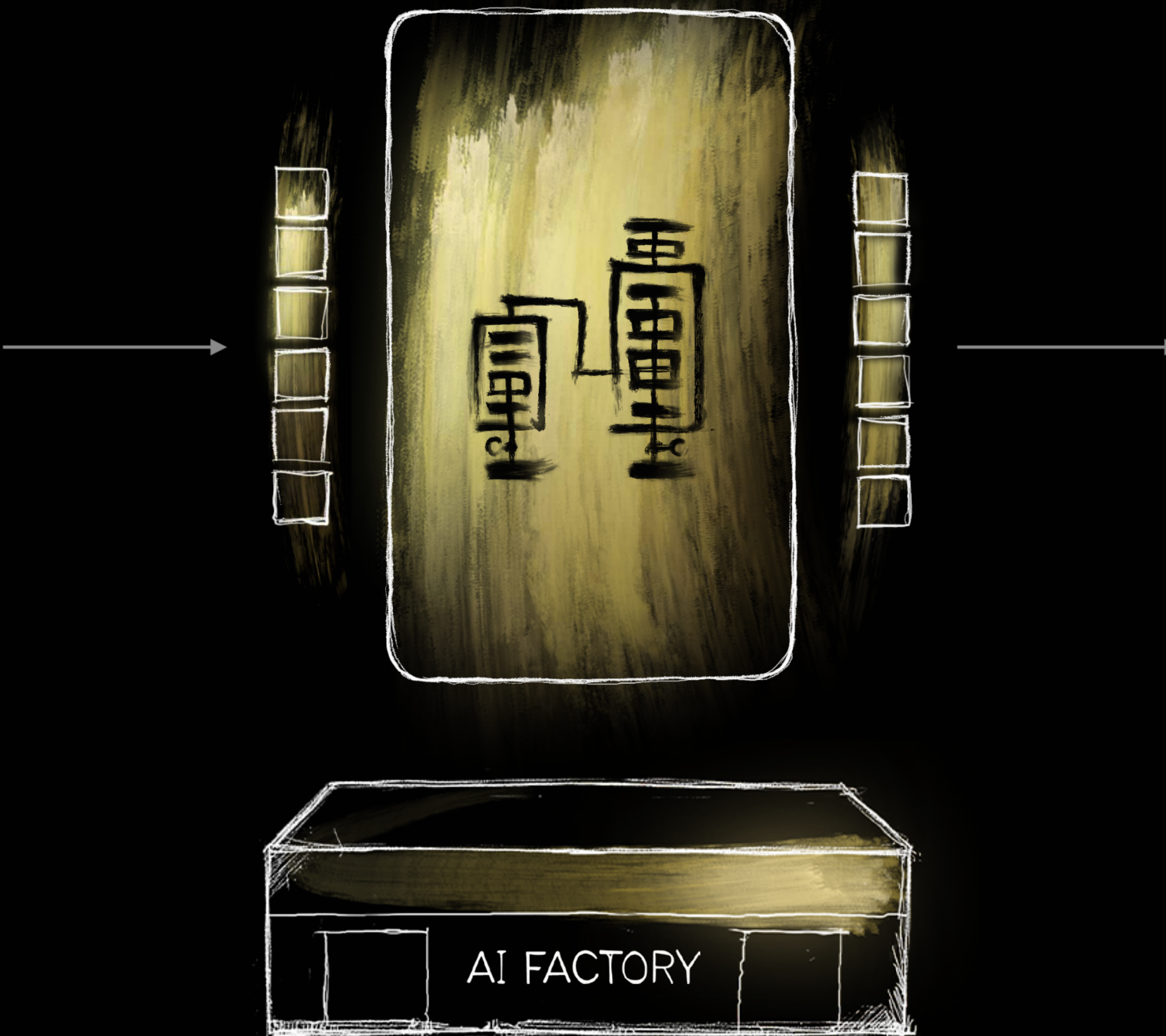
Scenario 3

The car is stopping on the road. **Dogs** barking and darting into the street from a nearby yard.



# General Robotics Opens the Next Trillion-Dollar Industry

拿起由 NVIDIA 製造的物體，並輕輕地將其放入抽屜中。  
在開始之前，解釋你將採取的每個步驟。



## Reasoning

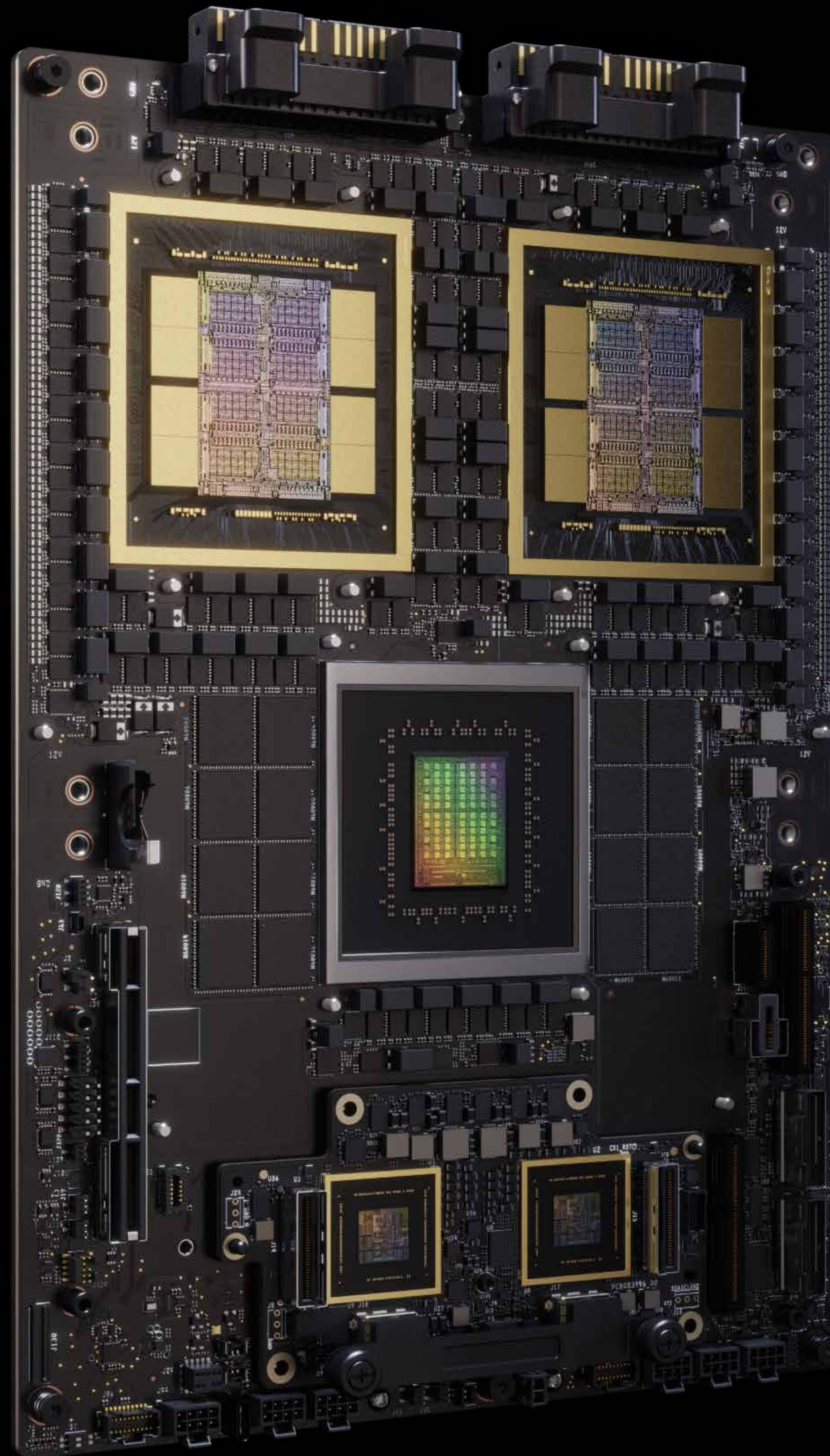
Okay, let's break down the problem. The task is to pick up the mug on the table. First, I need to confirm that the "mug" shown is the same as the "water bottle" shown in the video. The user might be using "mug" interchangeably with "water bottle." Let me check the video again. The objects on the table are: Robot Arm 1 (left), Robot Arm 2 (right), a spray bottle (left center), a water bottle (center), a black lid (right of the water bottle), and a purple container (far right). The instruction is to pick up the mug. Since there's no explicit mention of a mug, but there's a water



GB300 NVL72

Grace Blackwell NVL72  
“A Thinking Machine”





## Grace Blackwell Ultra Superchip

40 PFLOPS TE FP4

1.5X GB200

576 GB at 16 TB/s HBM3e

1.5X GB200

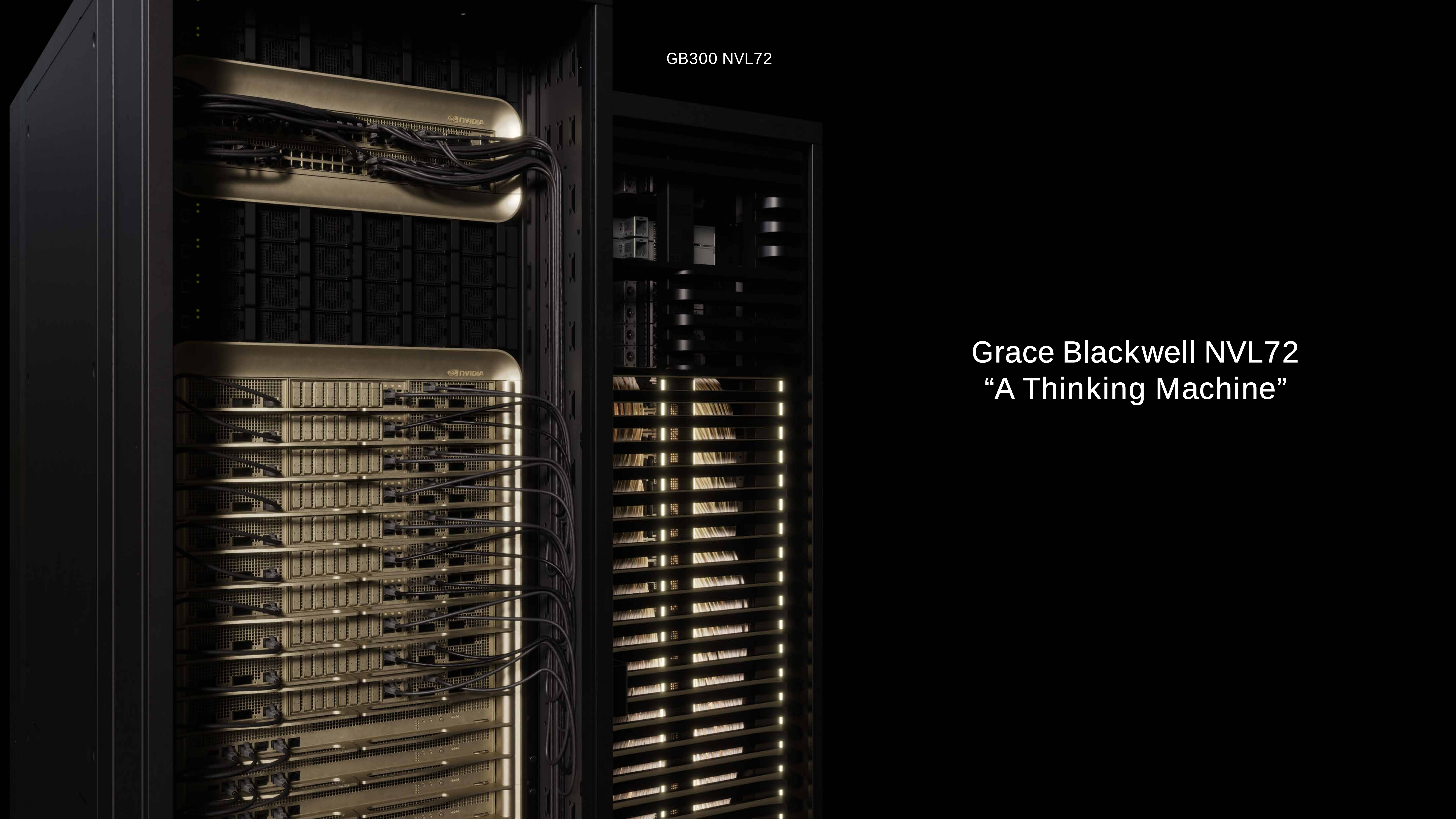
800 Gb/s CX8

2X GB200



GB300 NVL72

Grace Blackwell NVL72  
“A Thinking Machine”





CoreWeave  
220 MW  
USA, Scaling Globally



Courtesy of CoreWeave



Oracle Cloud - Supercluster  
260 MW  
Indonesia, Scaling Globally



Courtesy of Oracle



xAI - Colossus  
150 MW  
Memphis, Tennessee



Courtesy of xAI



Stargate  
200 MW, Scaling to 1.2 GW  
Abilene, Texas



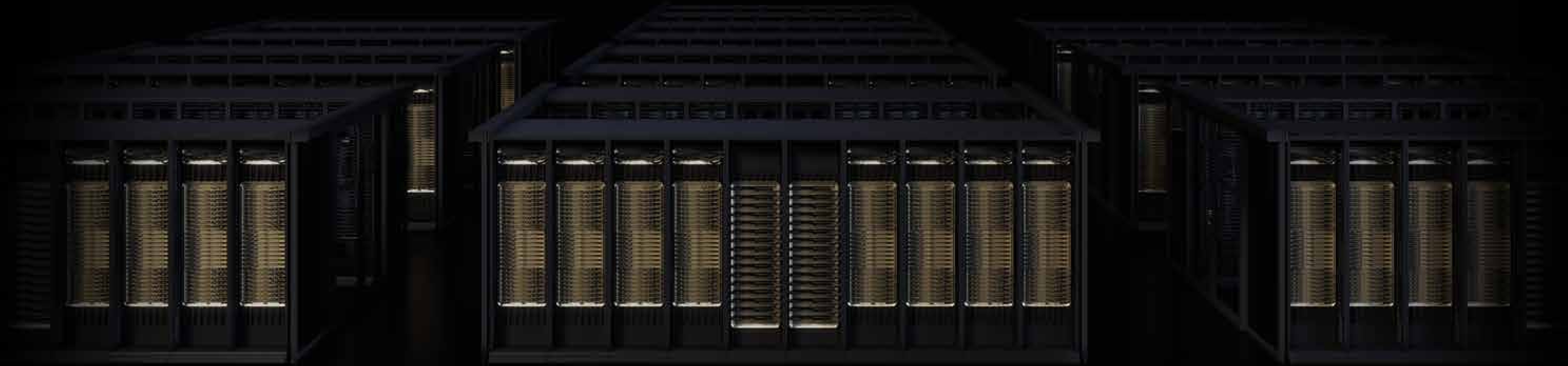
Courtesy of Crusoe







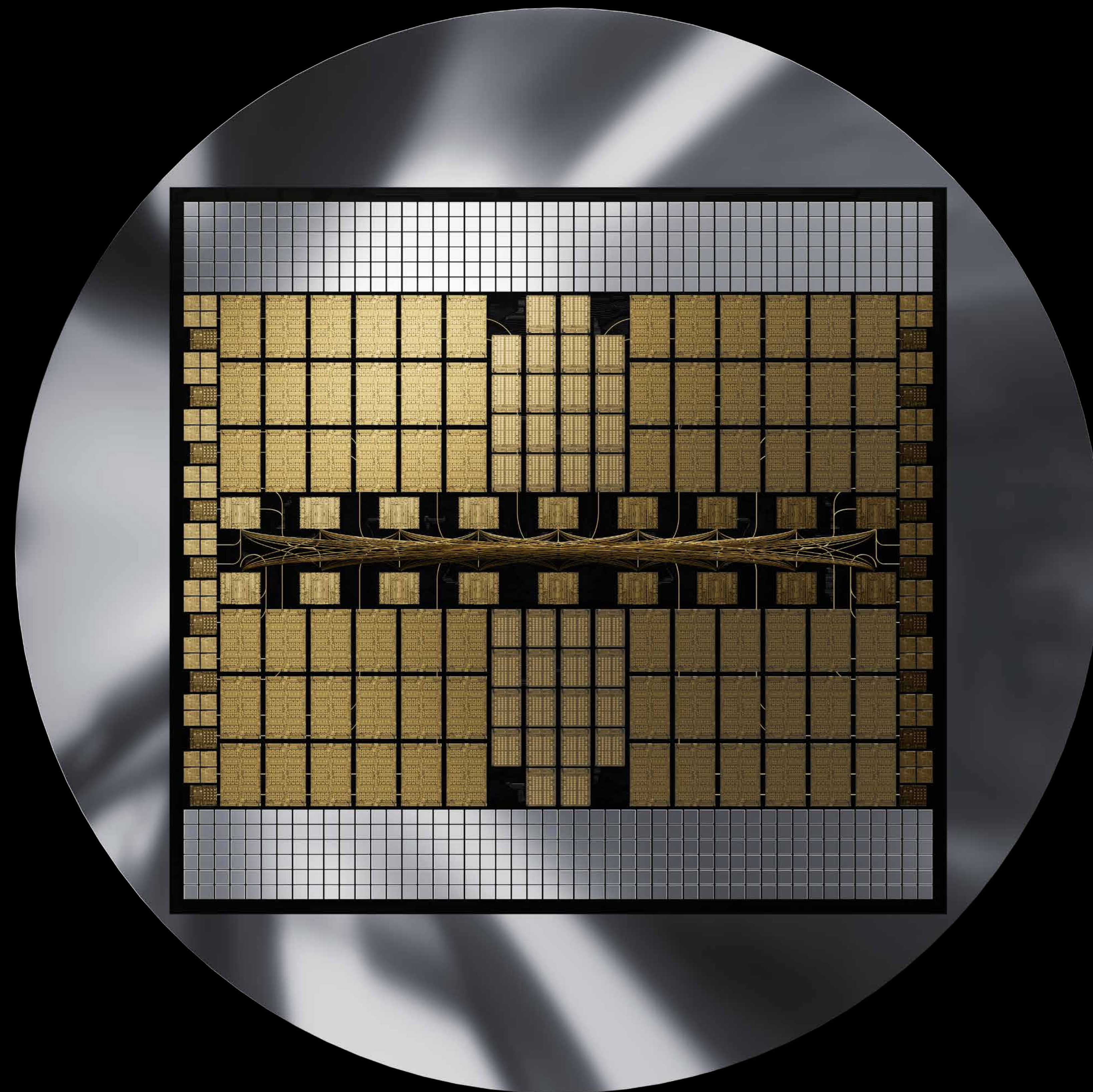
# Announcing Foxconn, Taiwan, TSMC and NVIDIA Build AI Infrastructure for Taiwan





# Grace Blackwell NVL72

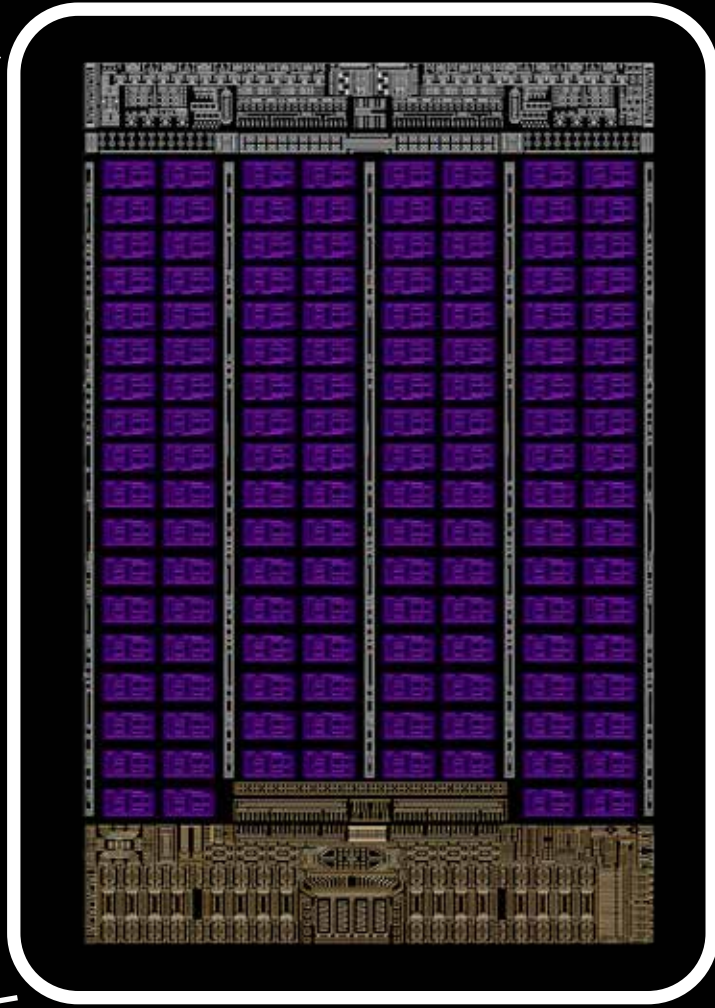
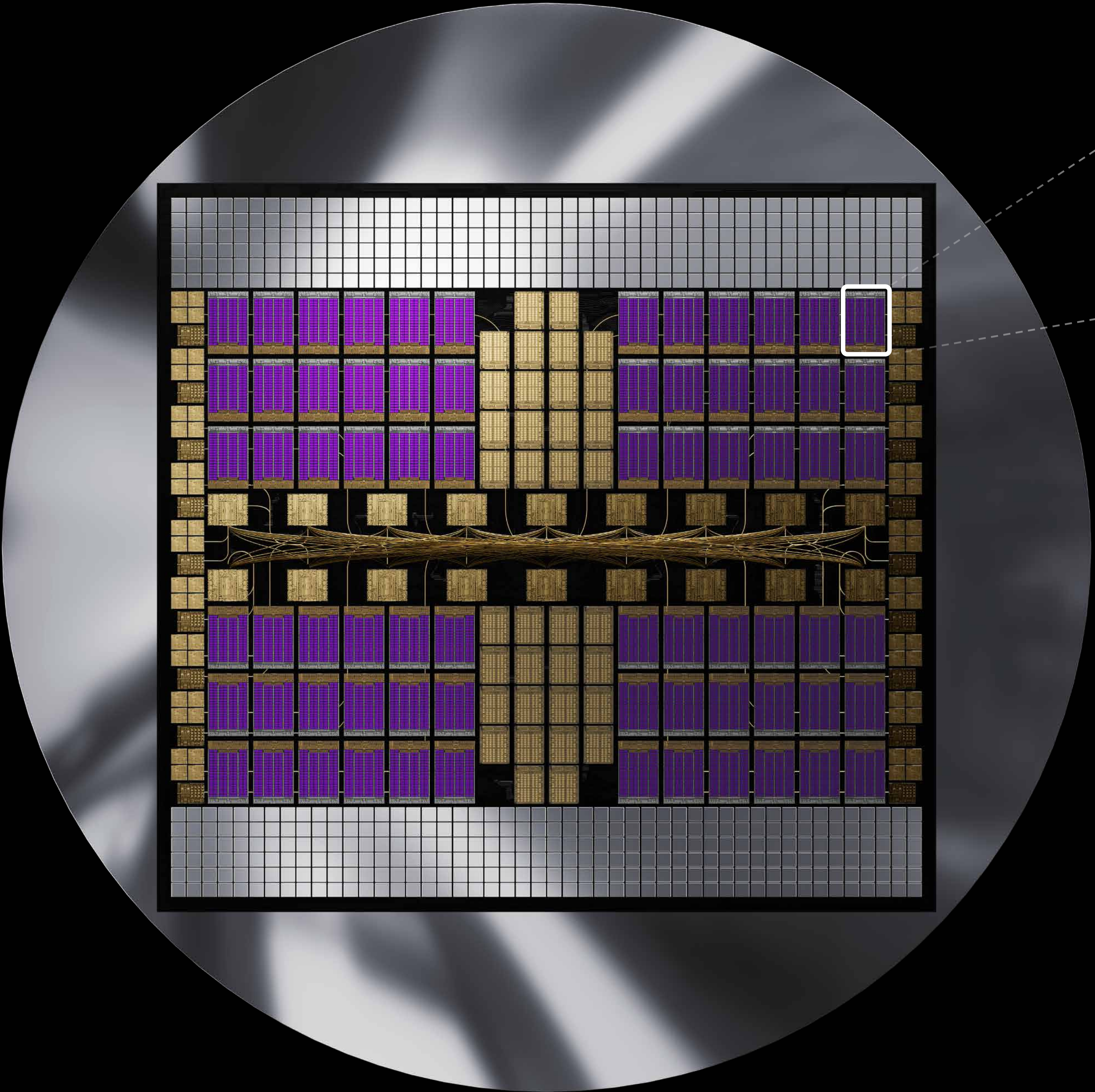
## “One Giant GPU”





# Announcing NVLink Fusion

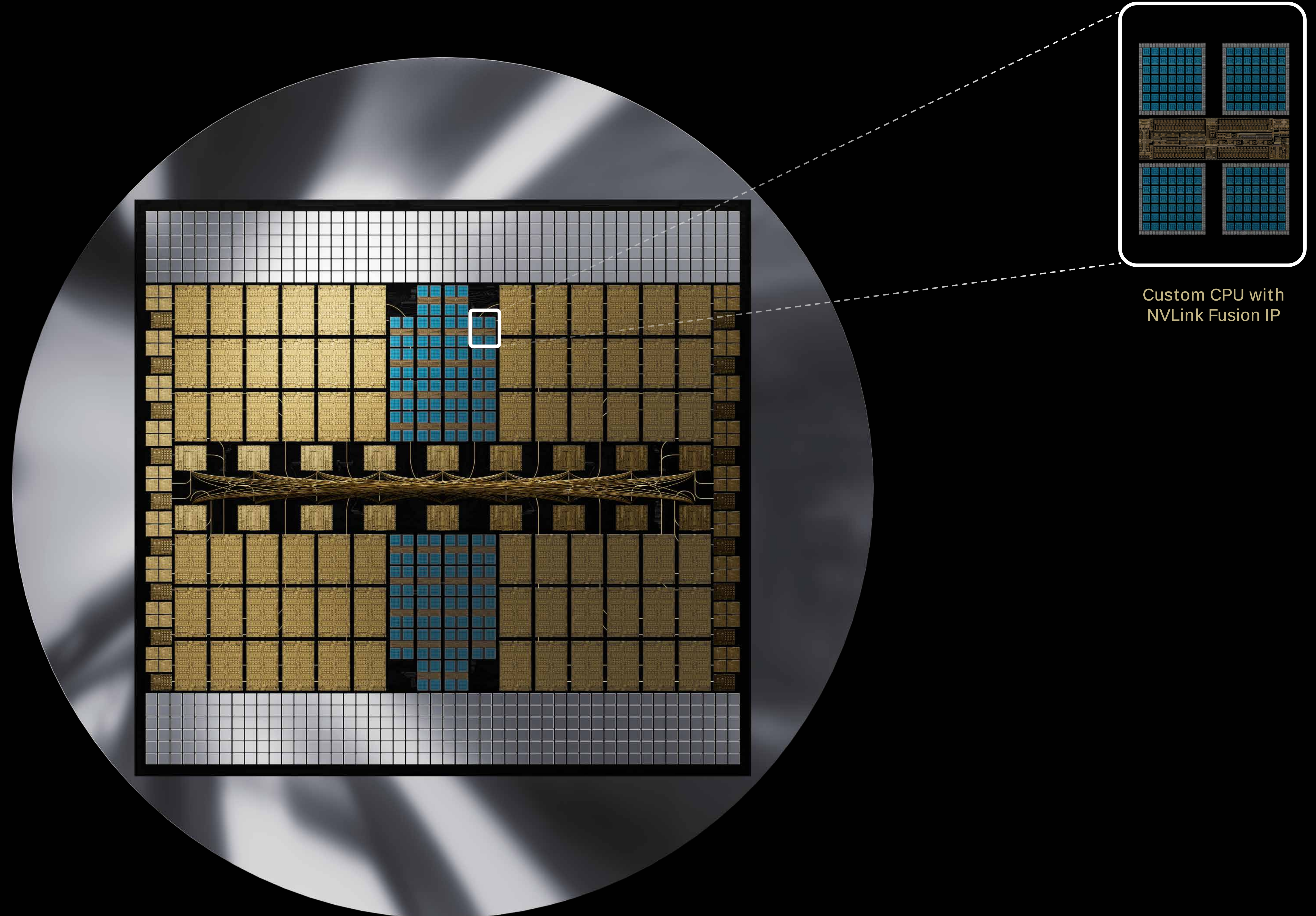
## NVIDIA AI Factory Plus Custom ASIC



Custom ASIC with  
NVLink Fusion IP



# Announcing NVLink Fusion NVIDIA AI Factory Plus Custom CPU





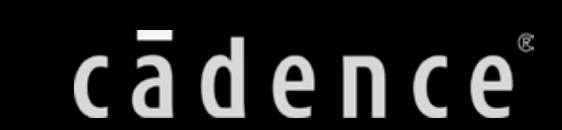


## NVIDIA NVLink Fusion Semi-Custom AI Infrastructure

NVIDIA Plus ASIC Hybrid AI Infrastructure

Proven Scale-Up and Scale-Out Ecosystem

Single Scalable Hardware Architecture

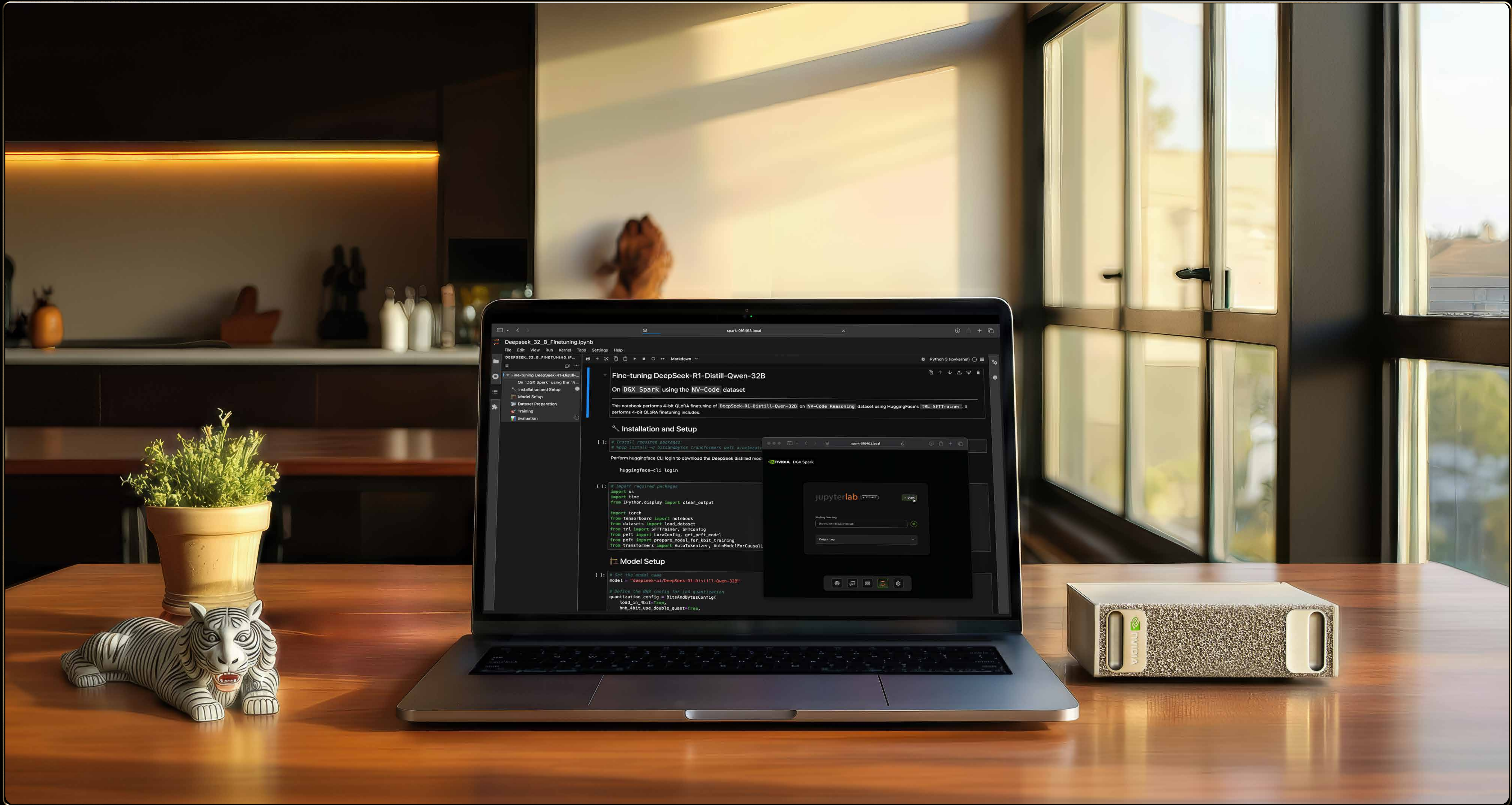




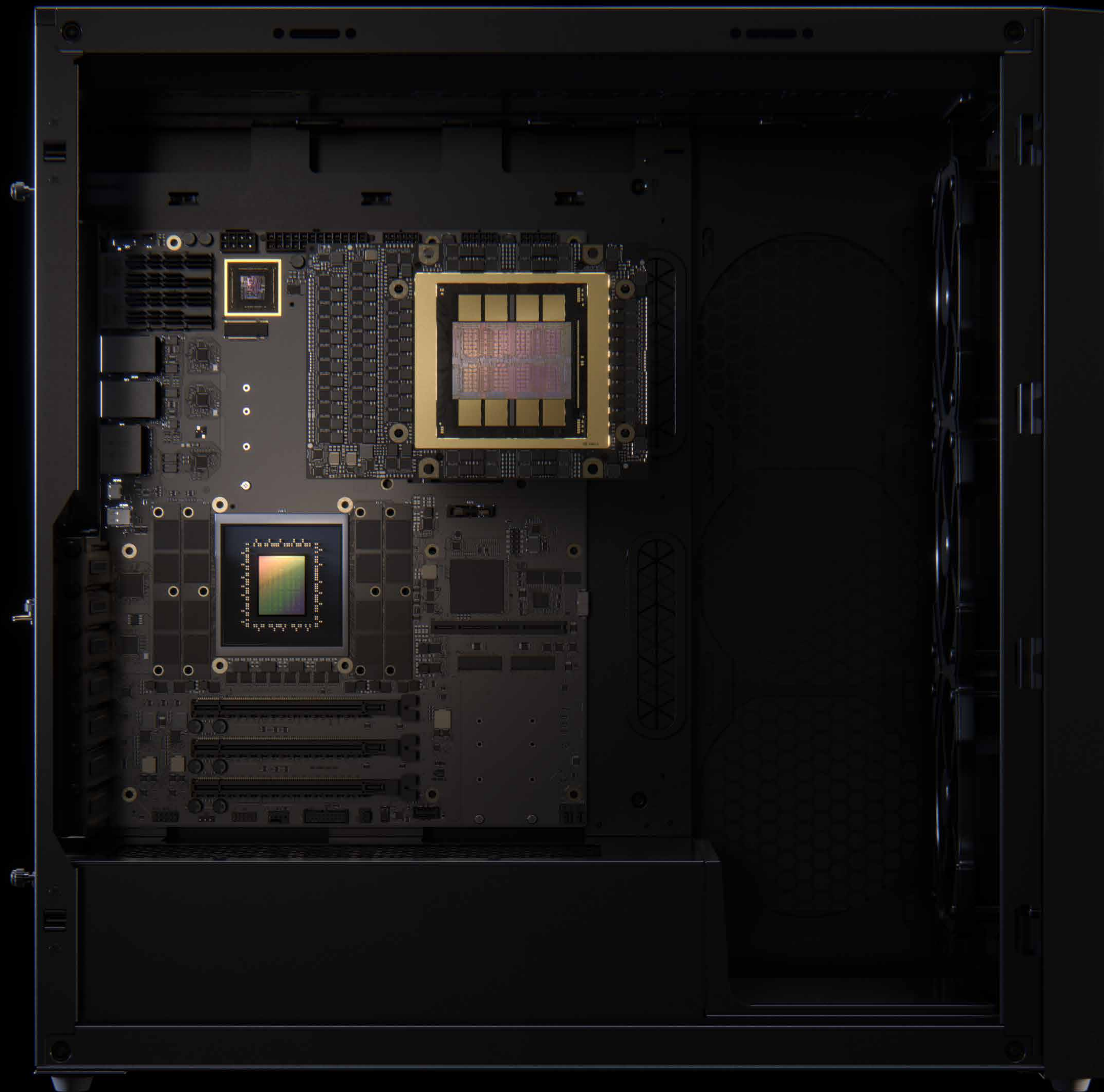
# One Architecture – from Cloud AI, Enterprise AI, Personal AI, to Edge AI











## Announcing DGX Station

### 1 Trillion Parameter Models Personal AI Machine

20 PFLOPS AI Performance  
GB300 Grace Blackwell Ultra Superchip

800 GB Unified System Memory  
Up to 1T Parameter Models

ConnectX-8 SuperNIC  
800 Gb/s Network Connection

ASUS

BOX

DELL Technologies

GIGABYTE



Lambda

msi





# NVIDIA Enterprise AI Platform Accelerates Trillion-Dollar Global IT Investment

Agentic AI

accenture

Infosys

Deloitte.

tcs

DATA CONSULTANCY SERVICES

EY

wipro



AIOPS

\$1T  
Enterprise IT and Industrial Automation  
Software Industry

AI Infrastructure

cisco

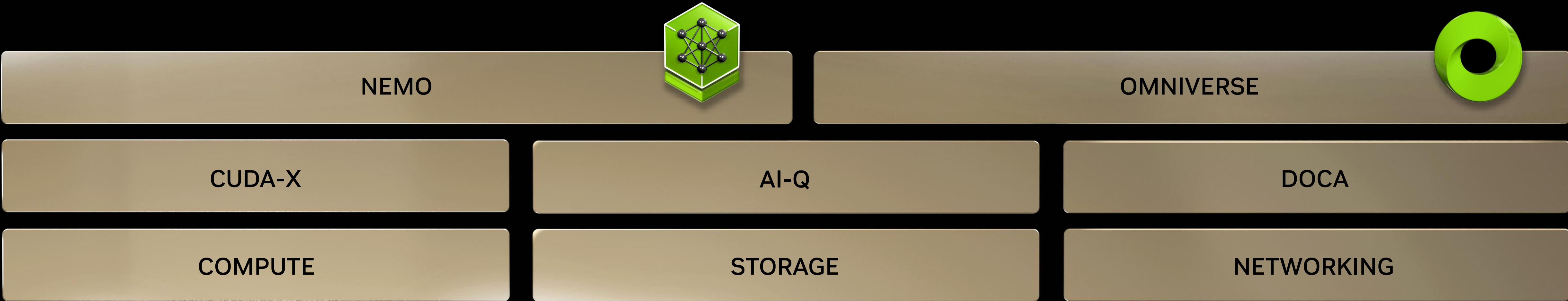
DellTechnologies

Hewlett Packard Enterprise

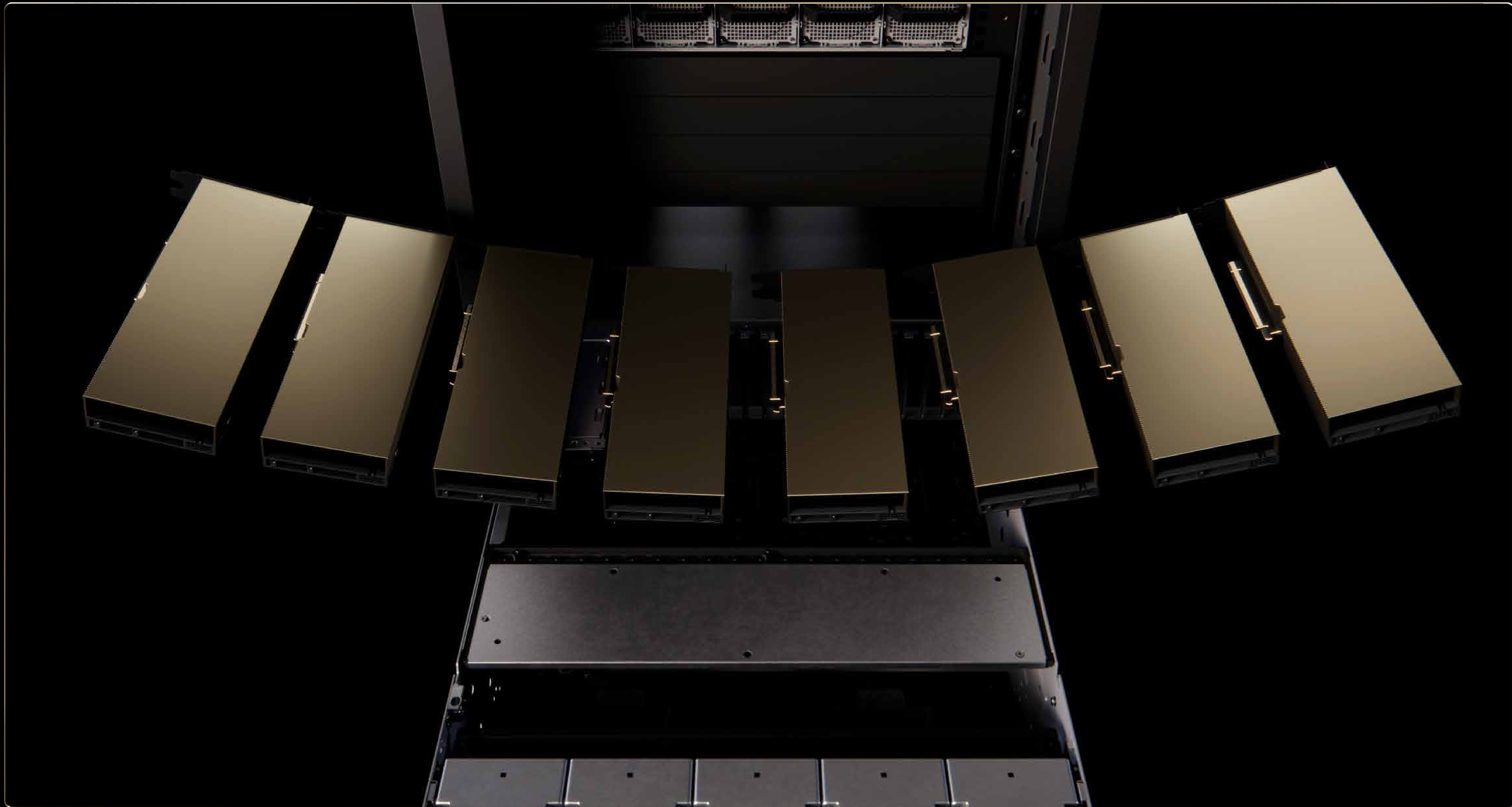
Hitachi Vantara

Lenovo

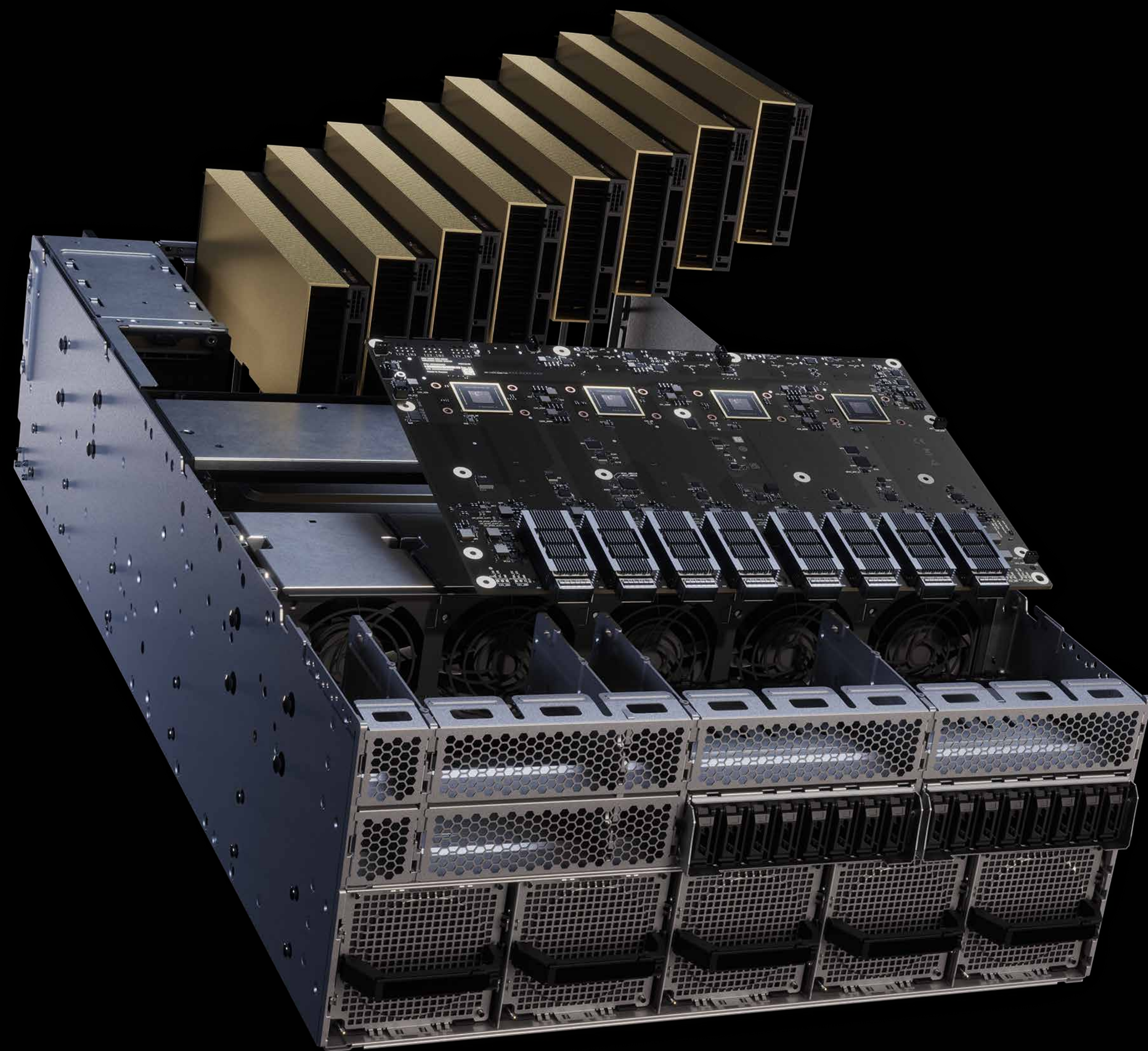
NetApp











## Announcing NVIDIA RTX PRO Server

### Enterprise AI and Industrial AI Computing Platform

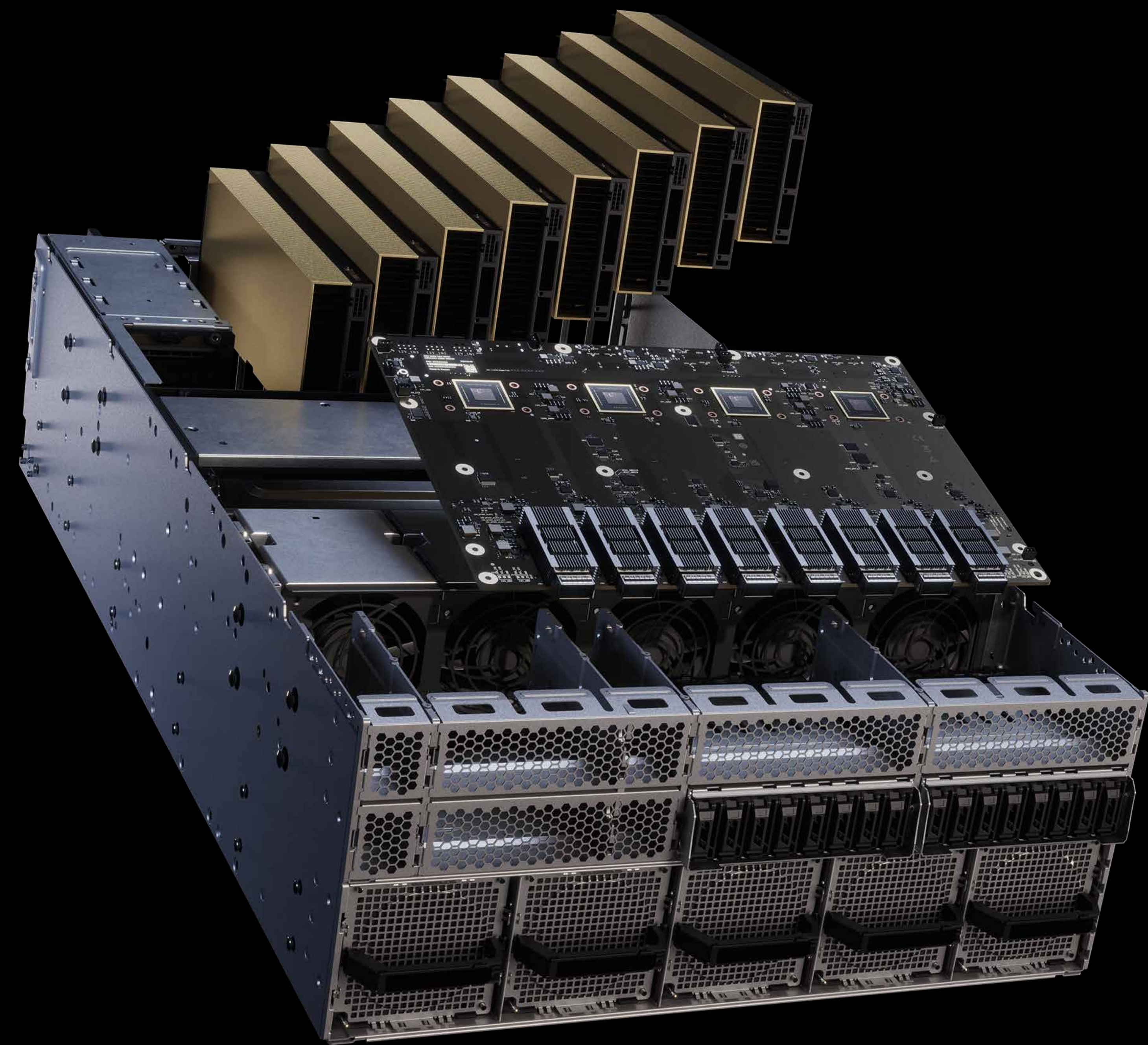
**30 PFLOPS FP4 AI Performance**  
8 x RTX PRO 6000 Blackwell Server Edition GPUs

**3 PFLOPS RTX Graphics Performance**  
Real-Time Omniverse Digital Twin Simulation

**800 GB GPU Memory**  
13 TB/s Bandwidth

**ConnectX-8 SuperNIC Switch**  
800 GB/s All-to-All Bandwidth



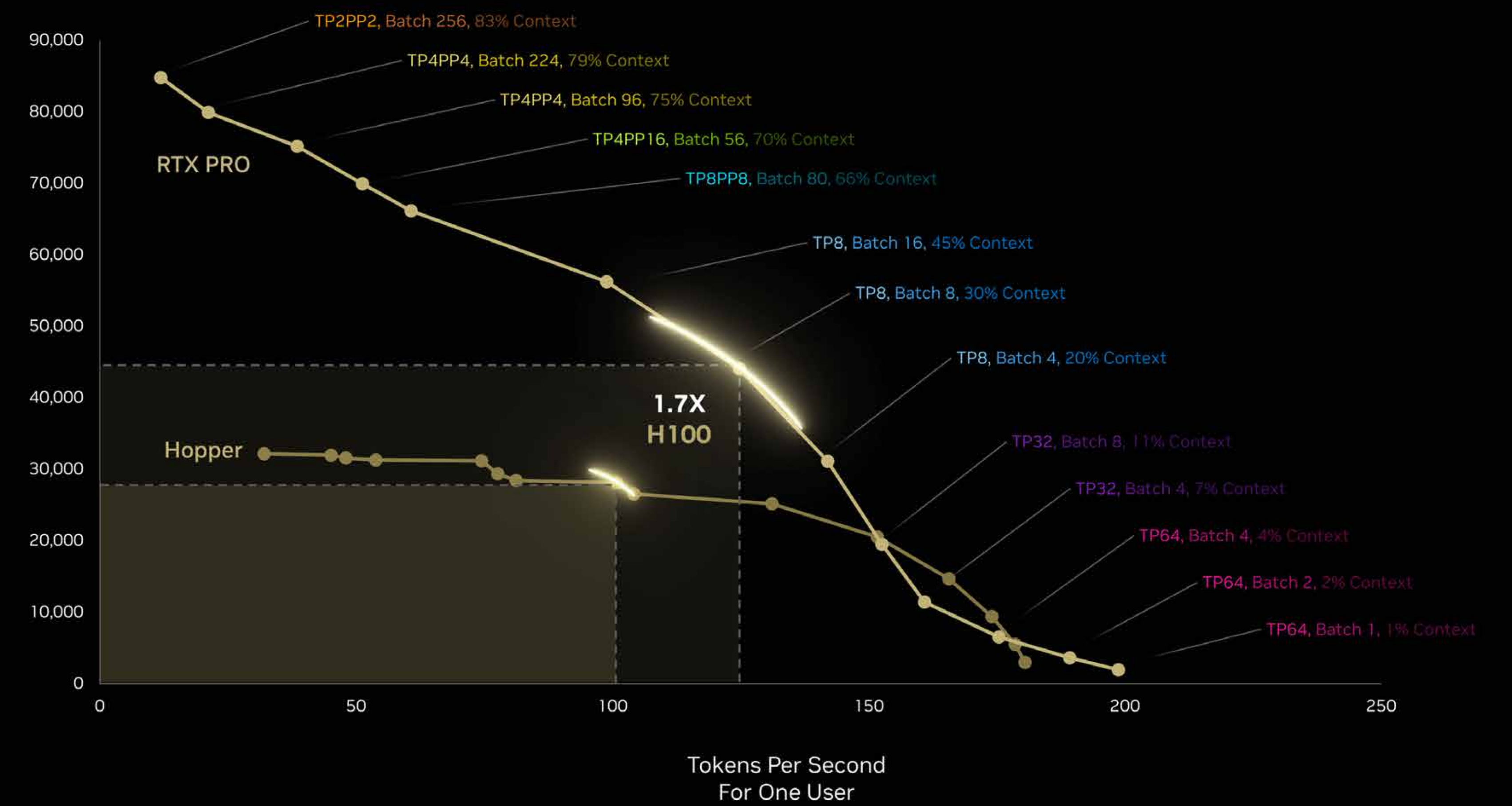


# Announcing NVIDIA RTX PRO Server

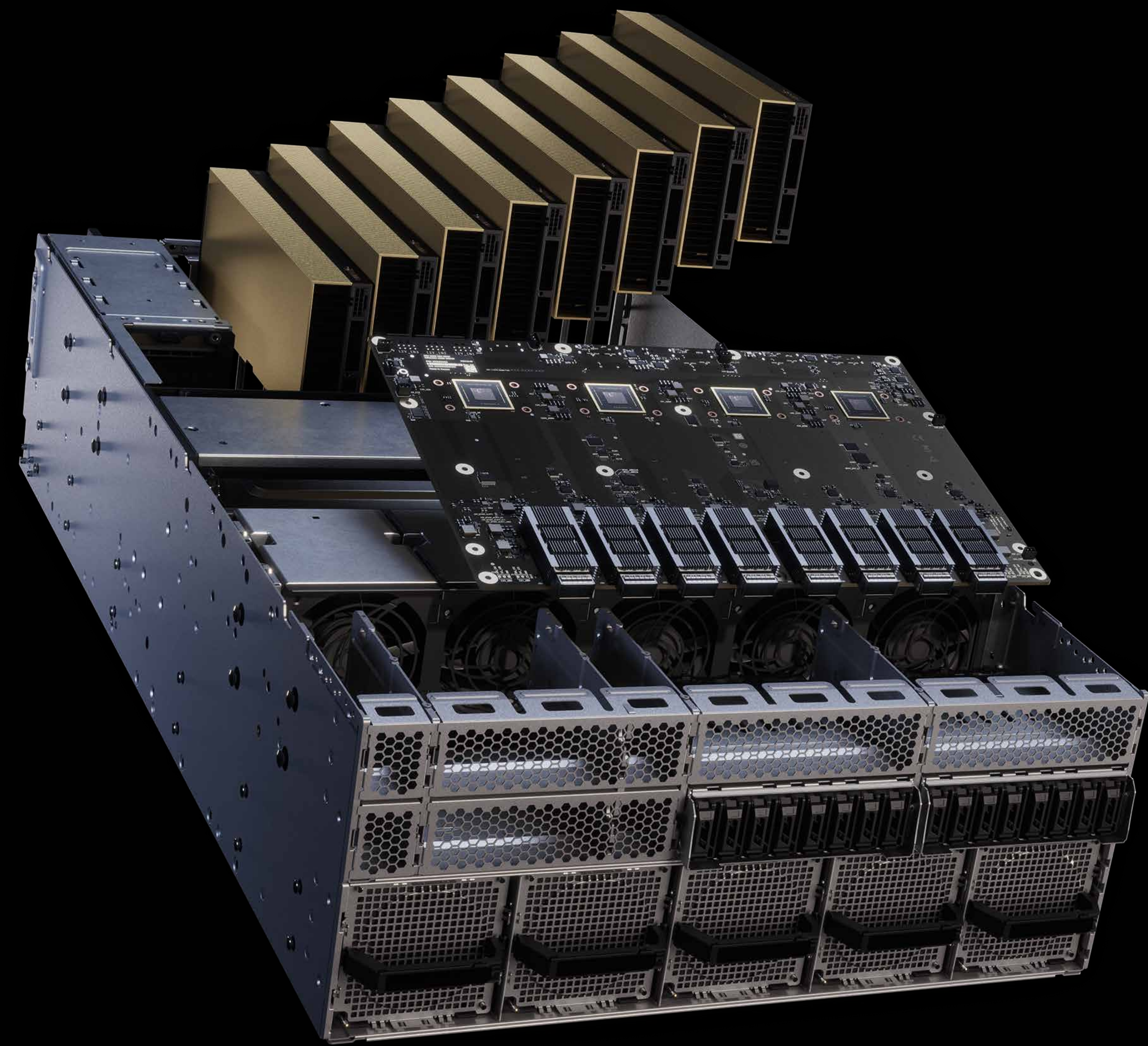
## Enterprise AI and Industrial AI Computing Platform

Tokens Per Second  
Per \$5M

Llama 70B 8K ISL / 256 OSL





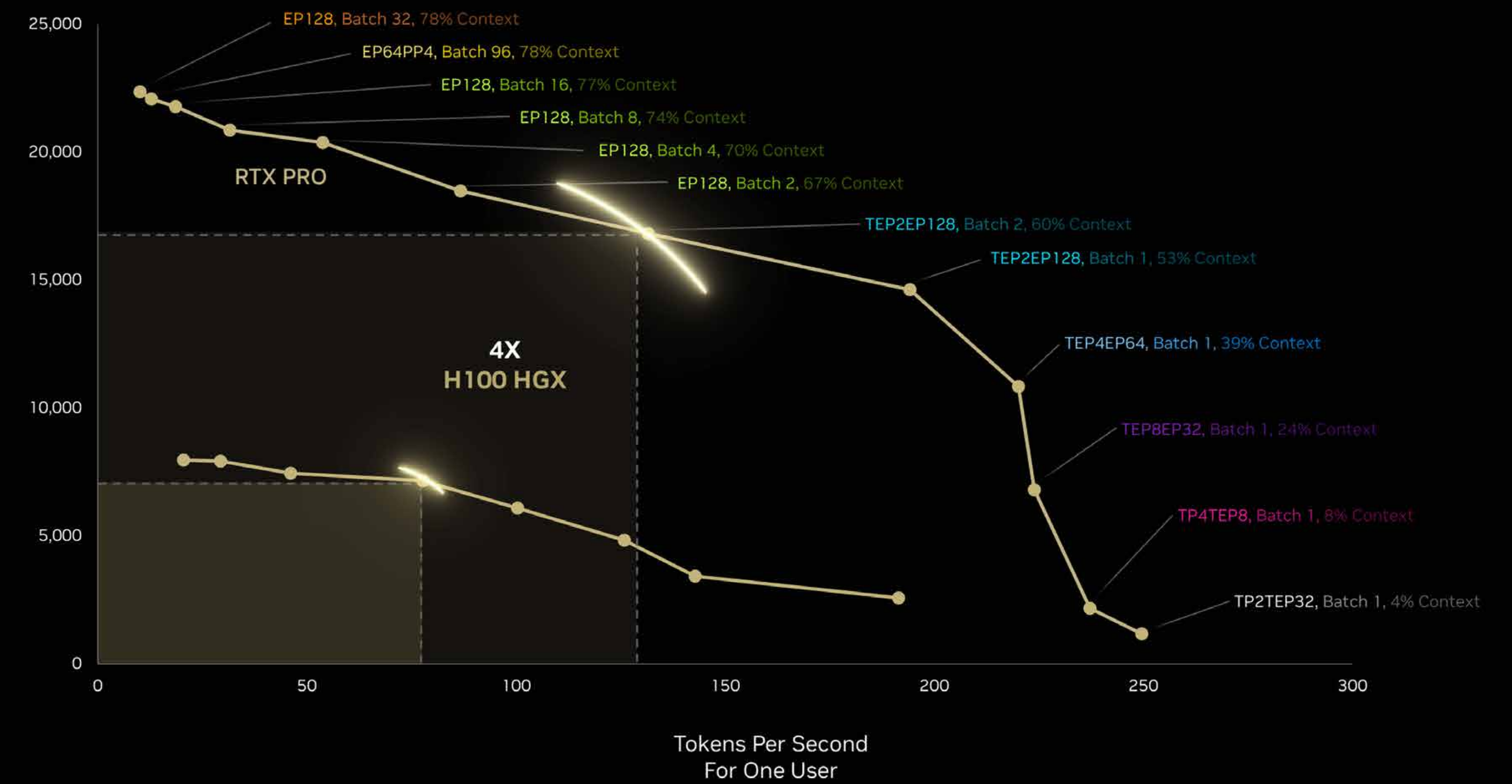


# Announcing NVIDIA RTX PRO Server

## Enterprise AI and Industrial AI Computing Platform

DeepSeek R1 128K ISL / 4K OSL

Tokens Per Second  
Per \$5M







# Announcing NVIDIA RTX PRO Server

## Enterprise AI and Industrial AI Computing Platform

Available Now



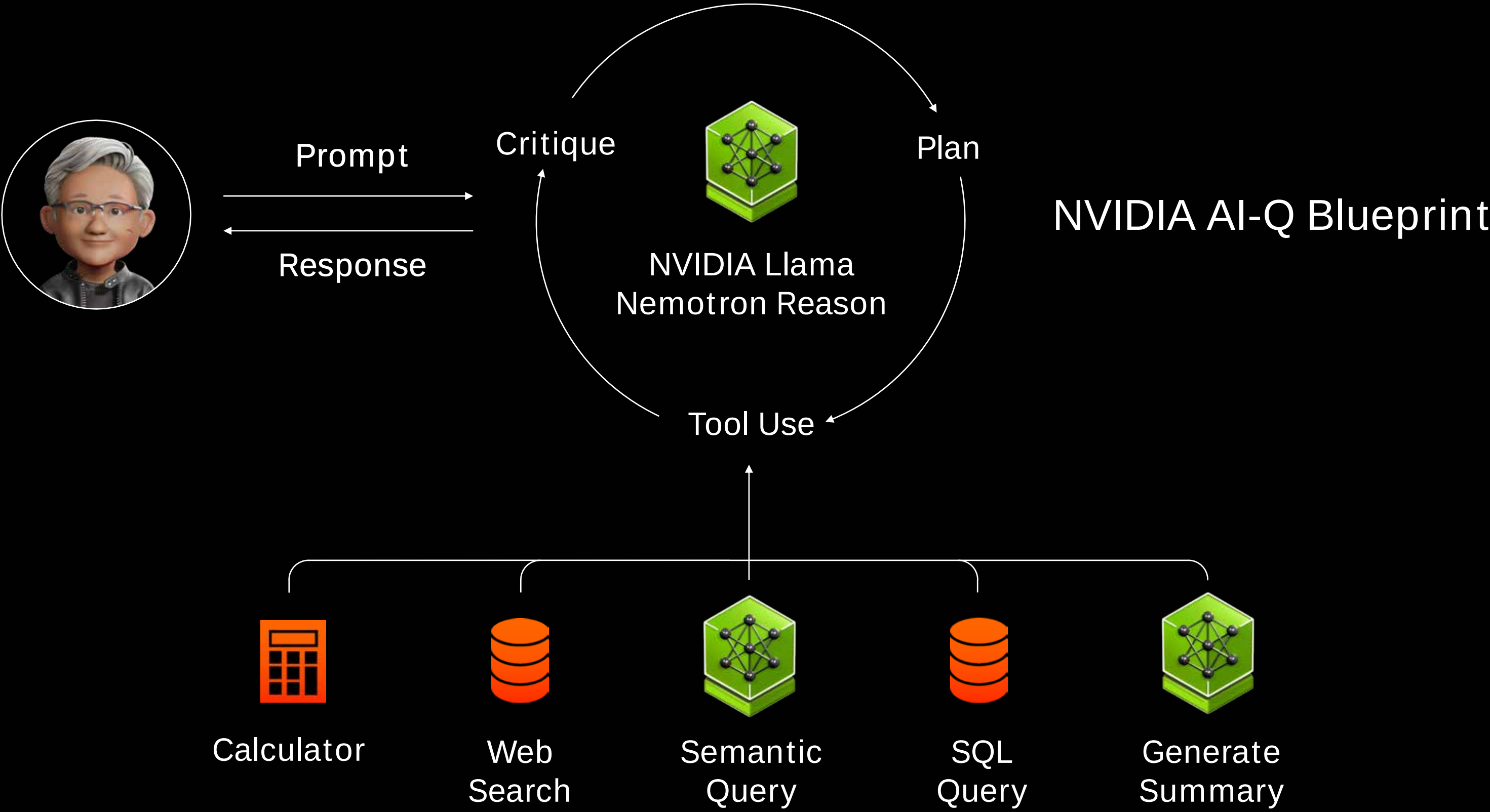
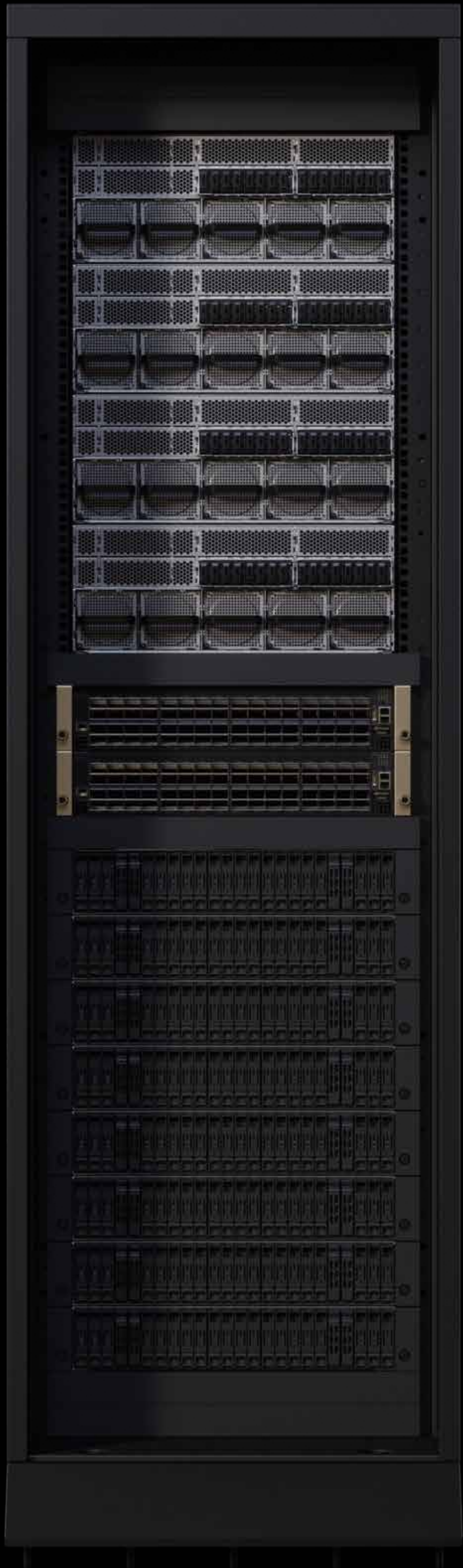


# NVIDIA AI Data Platform

**Compute**  
RTX PRO 6000 Blackwell Server Edition  
NVIDIA BlueField

**Networking**  
NVIDIA Spectrum-X

**Storage**  
NVIDIA BlueField





# NVIDIA Top Leaderboard Open Models and Datasets

# 15X Faster

## Faster Data Extraction NeMo Retriever



## Open Reasoning Models

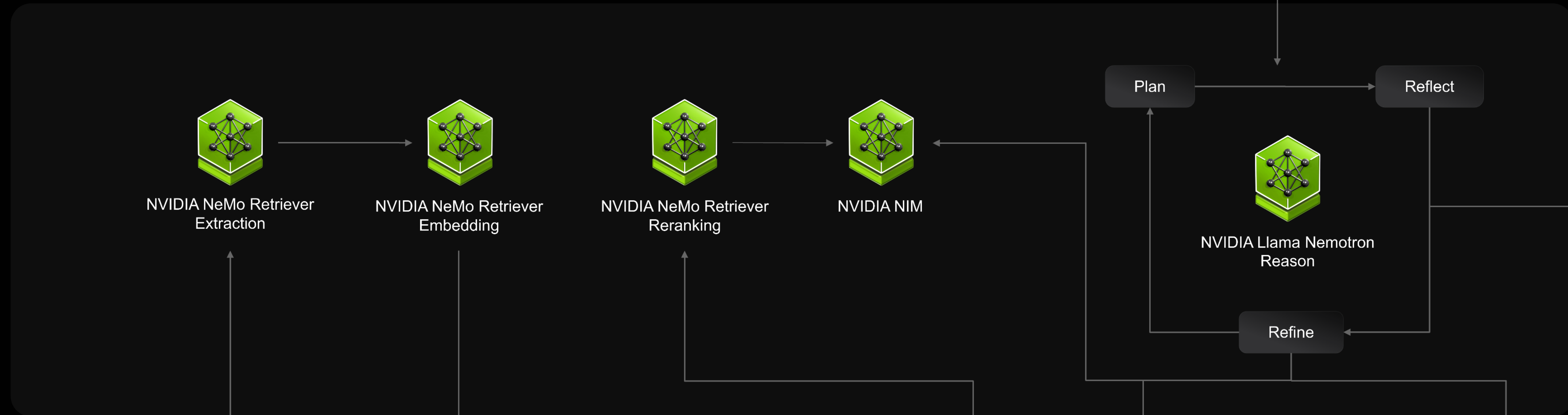
# 50% Greater

## Higher Retrieval Accuracy





 DataEngine



Images



Documents



Videos



Chat



Email



Vectors



Data Catalog



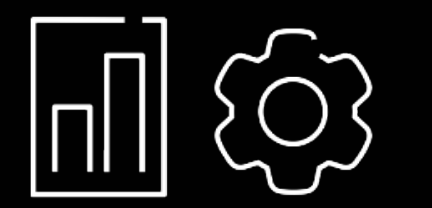
EDW



Internet

 DataStore

 DataBase



Report / Actions



Build a plan for the top sales opportunities in APAC.



Review Thinking Process



Outline



CRM Retrieval



Multimodal Data Retrieval



Build Action Plan



Reply to your agent...

Start New Chat



Create Action Plan

Step 4

Prompt:

Key rules that apply: provide data-driven, actionable guidance for new APAC sales reps; use evidence from actual deals and testimonials; align recommendations with known GTM strategy focus (AI, enterprise storage, cloud, performance/scale, etc.); respect privacy, provide transparent sources. We have results on closed-won APAC deals and a set of global testimonials, but couldn't access the GTM doc for APAC priorities. Will construct the action plan using observed deal patterns and testimonials relevant for the region/industry, and will highlight that GTM strategy was not directly accessible.

Thoughts: - Key rules that apply: provide data-driven, actionable guidance for new APAC sales reps; use evidence from actual deals and testimonials; align recommendations with known GTM strategy focus (AI, enterprise storage, cloud, performance/scale, etc.); respect privacy, provide transparent sources. We have results on closed-won APAC deals and a set of global testimonials, but couldn't access the GTM doc for APAC priorities. Will construct the action plan using observed deal patterns and testimonials relevant for the region/industry, and will highlight that GTM strategy was not directly accessible.



Client Applications



Agent



Prompt



Identity &  
Access  
Management



Response



NVIDIA Llama  
Nemotron  
Reasoning

DELL AI Data Platform With NVIDIA

Data Platform Services



Storage Access Control  
& Identity Management



Federated SQL Engine



AI Data Pipelines



Data Processing  
Engine



Unstructured Data  
Engine



Web Search



External Data  
Sources  
Structured &  
Unstructured

RAG Pipeline



NVIDIA NeMo  
Retriever Reranking



Index  
NVIDIA cuVS



NVIDIA NeMo  
Retriever Embedding



NVIDIA NeMo  
Retriever Extraction

Storage Platform

Project Lightning | PowerScale | ObjectScale | PowerProtect



Vector Data



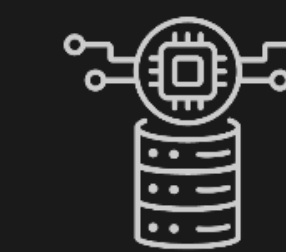
Iceberg-based  
S3 Tables



Semi-structured &  
Unstructured Data



Tensor Data



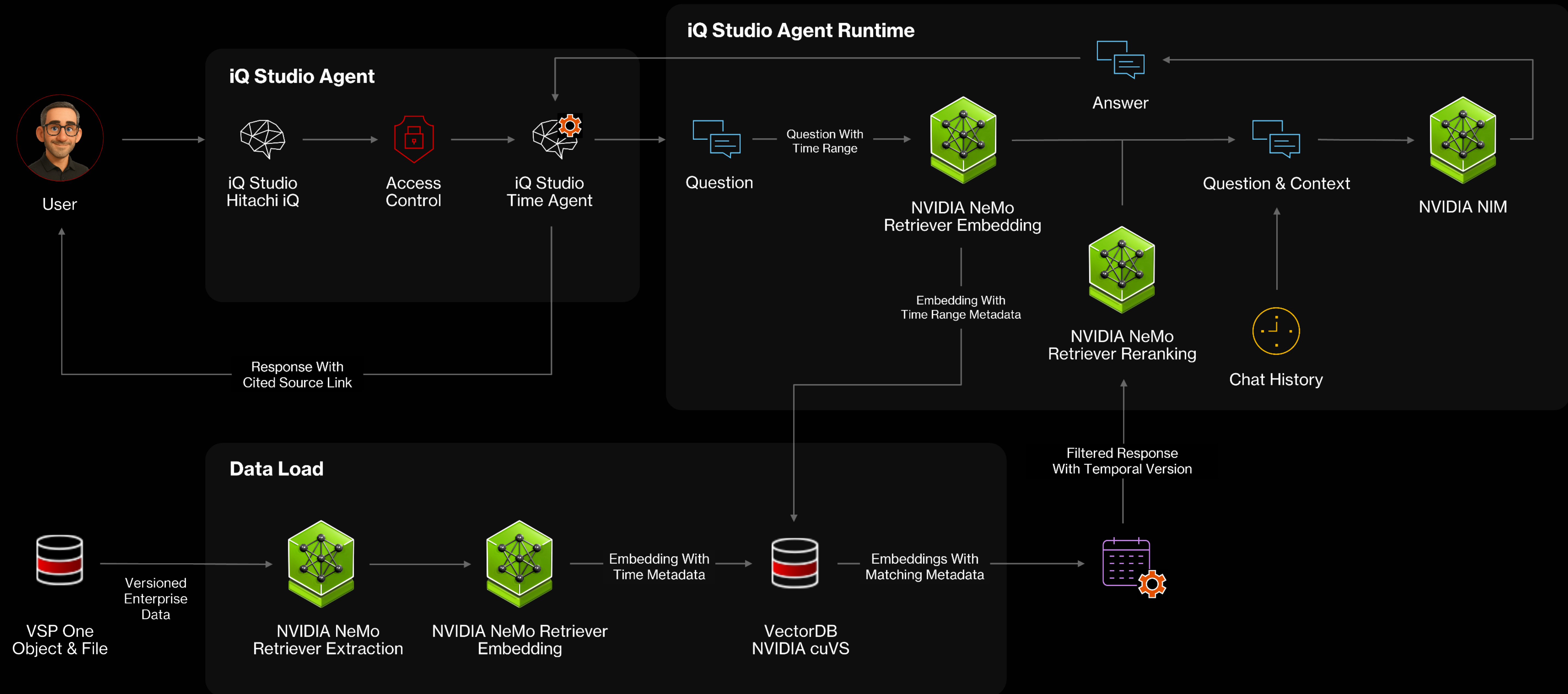
PowerScale  
Metadata IQ

Dataset Interface  
& Access Control

Data Resiliency, Protection, Deduplication









IBM



PDF



S3/NFS



IBM Storage  
Scale + AFM



Kafka

Triggers



CAS  
Runtime

Data



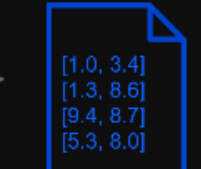
Extract



Embed



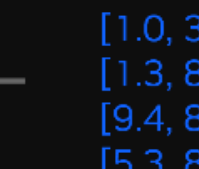
Vectorize



Index

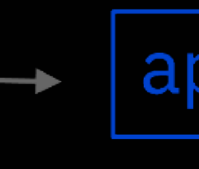


Vector  
Database



Vectorized  
Re-rank

[1.0, 3.4]  
[1.3, 8.6]  
[9.4, 8.7]  
[5.3, 8.0]



CAS Search  
API

Refine



NVIDIA Llama  
Nemotron  
Reason

Reflect

Generate



NVIDIA NIM



Answer



Query

Plan



Generate



NVIDIA NIM



Answer



PDF

Data

Metadata



Enterprise Metadata and  
Role-Based Access Control

Top-K

Access  
Control

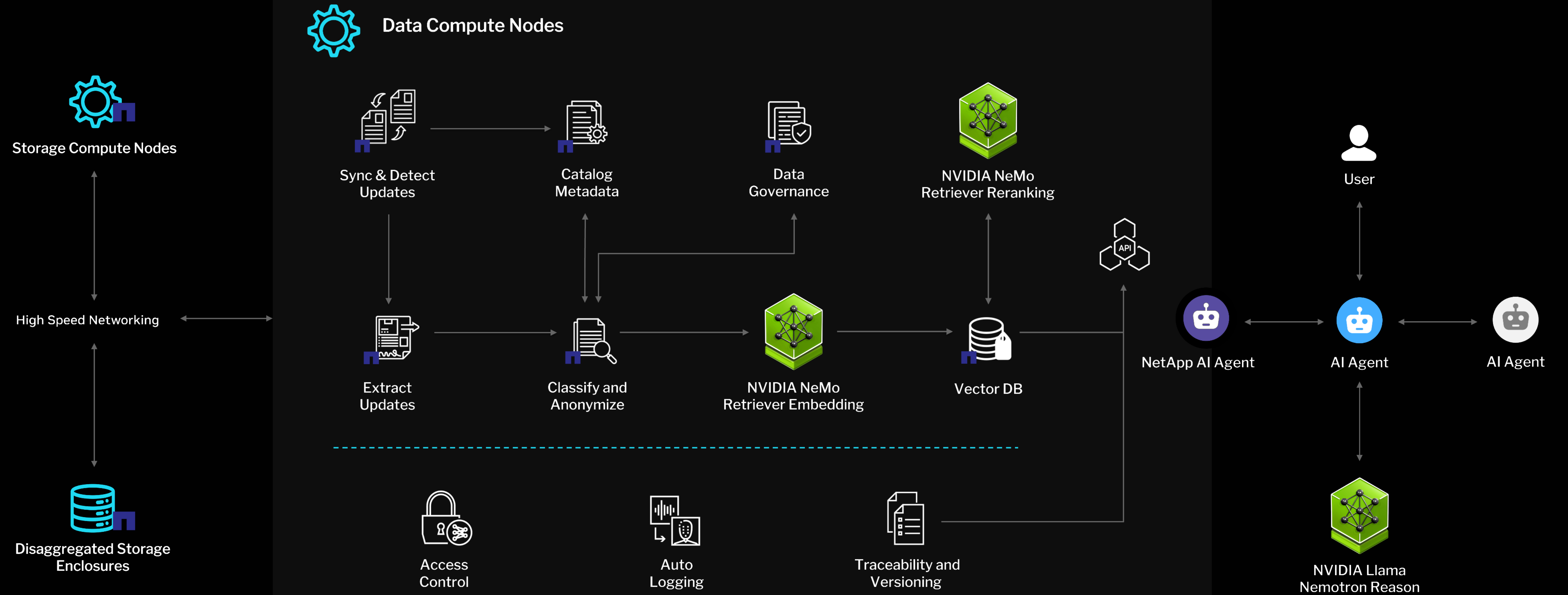


NVIDIA GPU Operator



NVIDIA Network Operator







# NVIDIA Enterprise AI Platform Accelerates Trillion-Dollar Global IT Investment

Agentic AI

accenture

Infosys

Deloitte.

tcs

Deloitte.

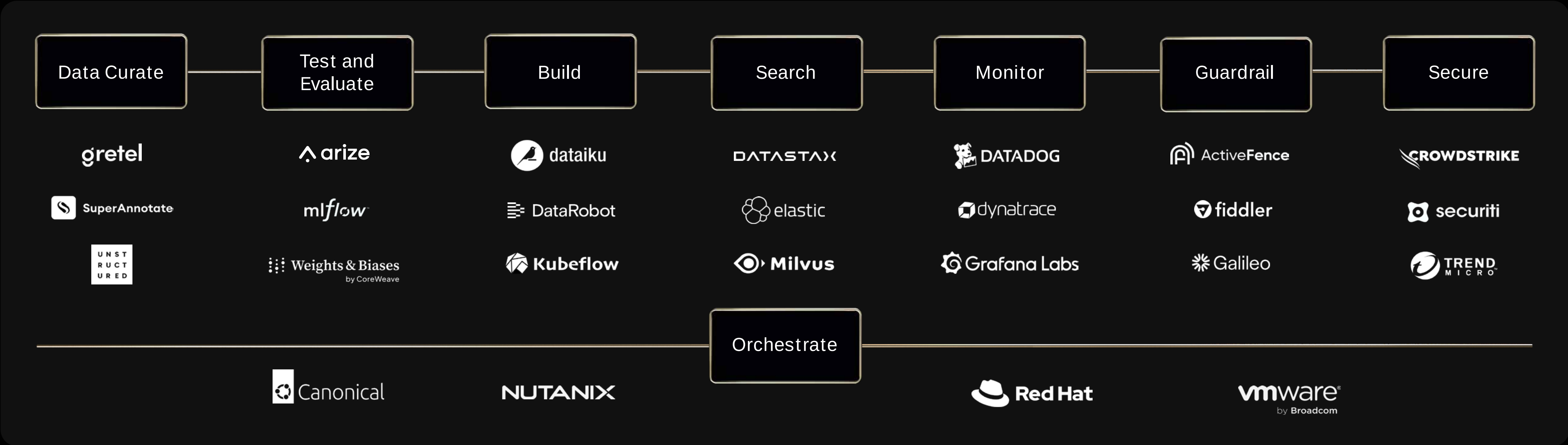
Wipro

EY

TATA CONSULTANCY SERVICES



AIOPS



AI Infrastructure

cisco

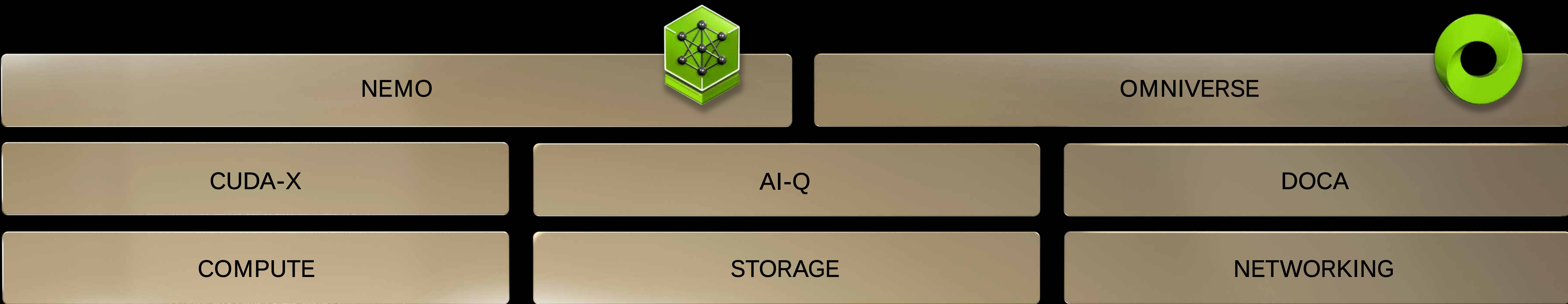
Hitachi Vantara

DELL Technologies

Lenovo

Hewlett Packard Enterprise

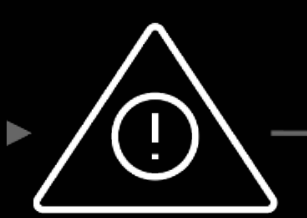
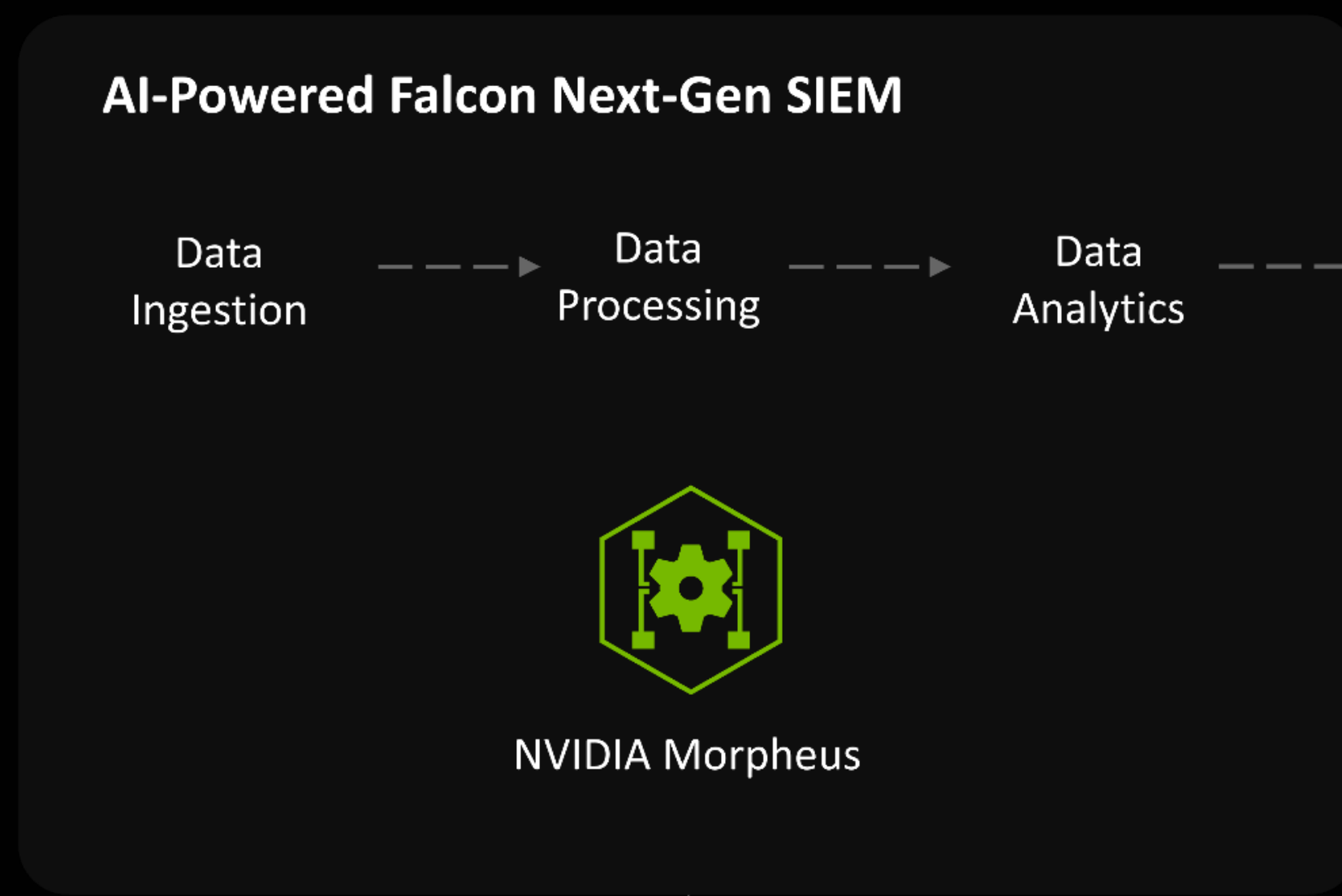
NetApp



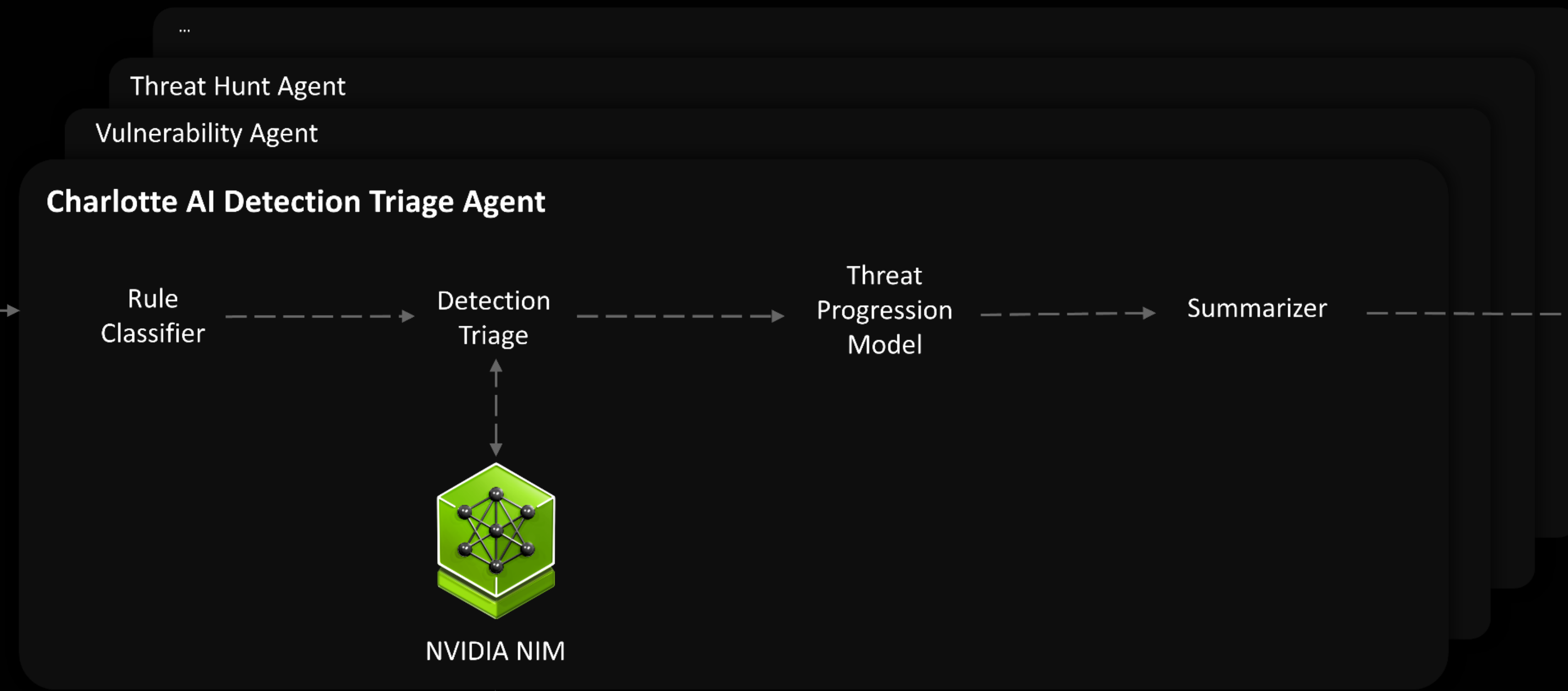




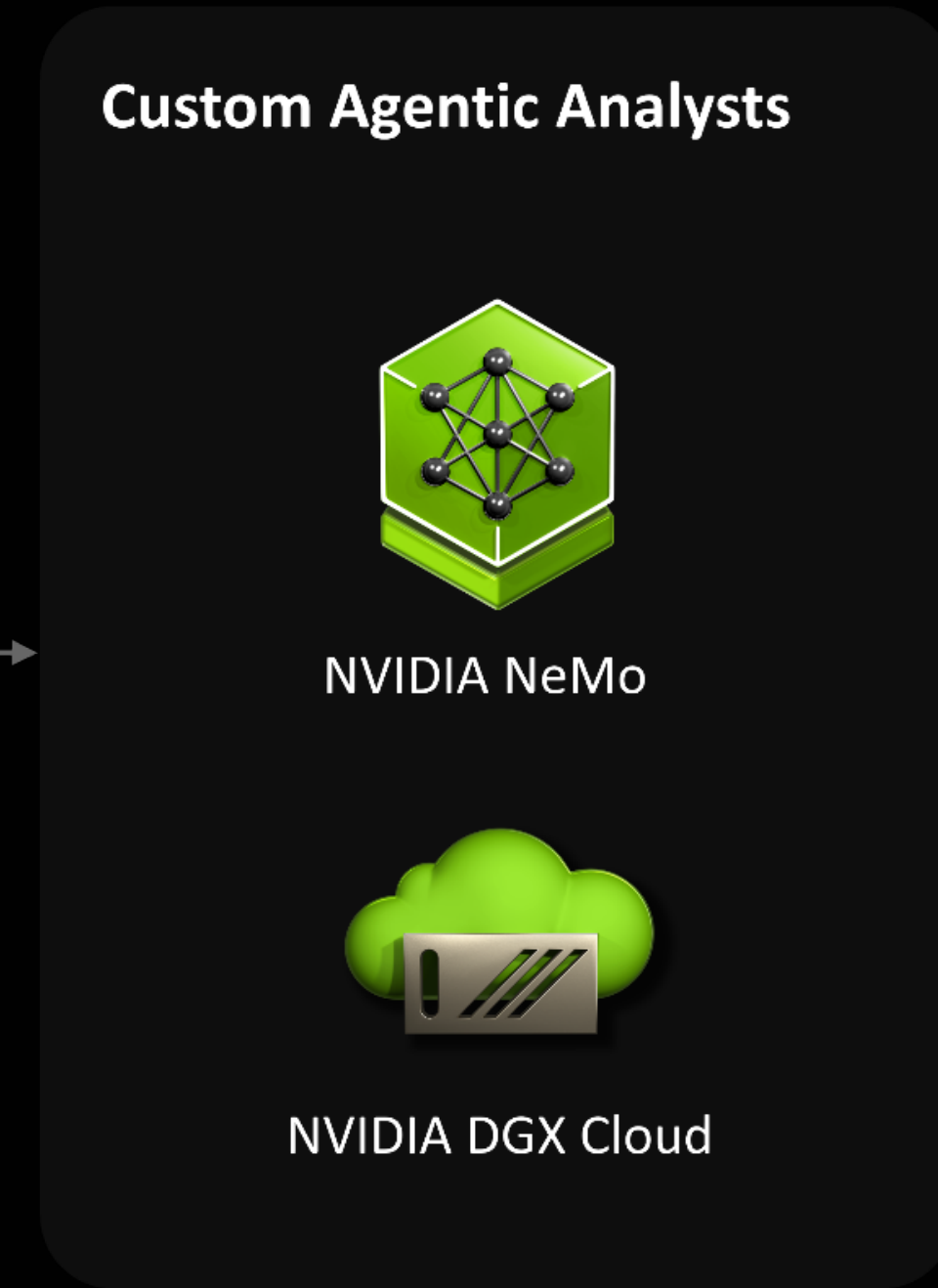
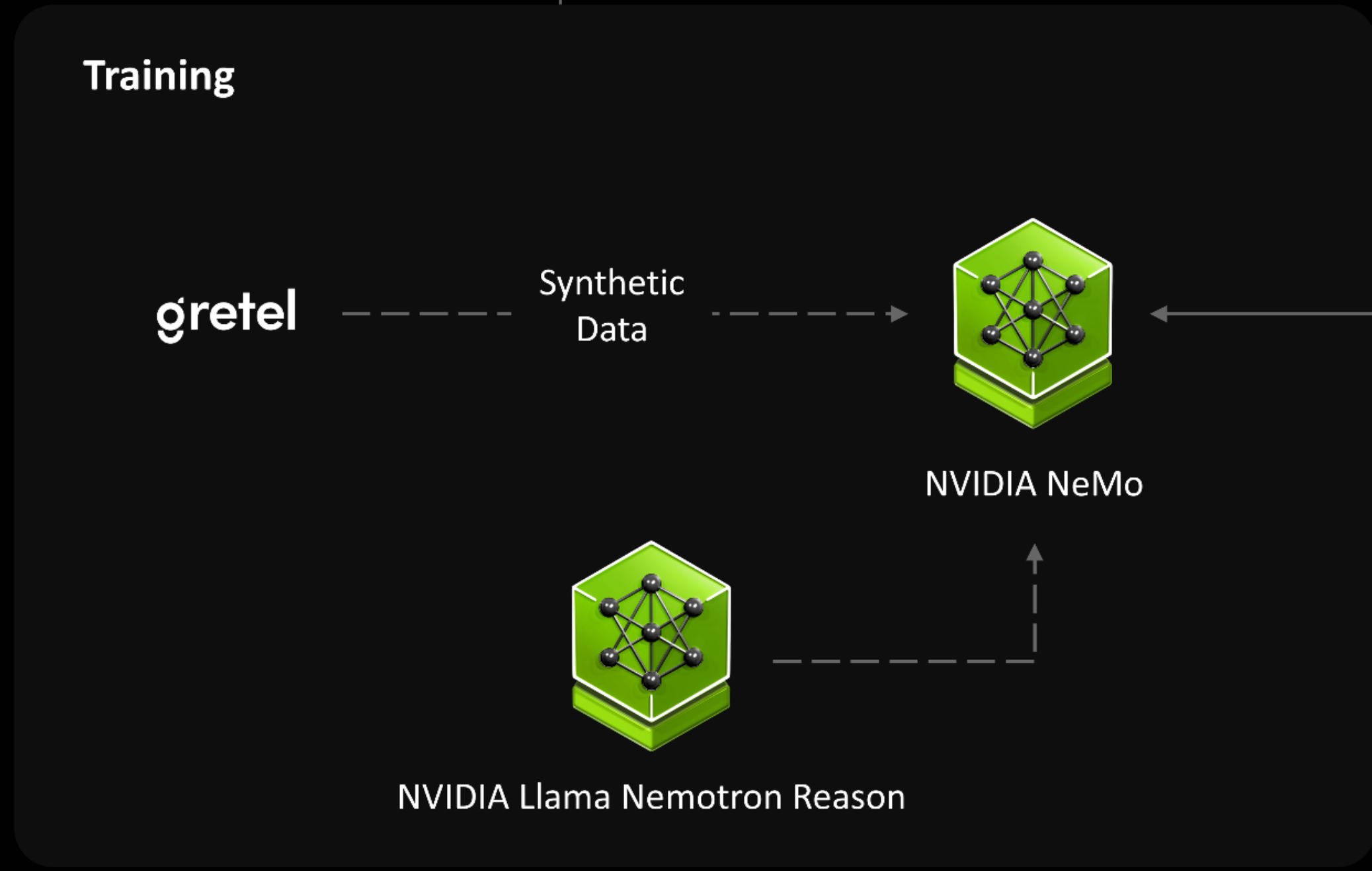
 Always-on Agentic Analysts



Threat  
Detection  
Documents

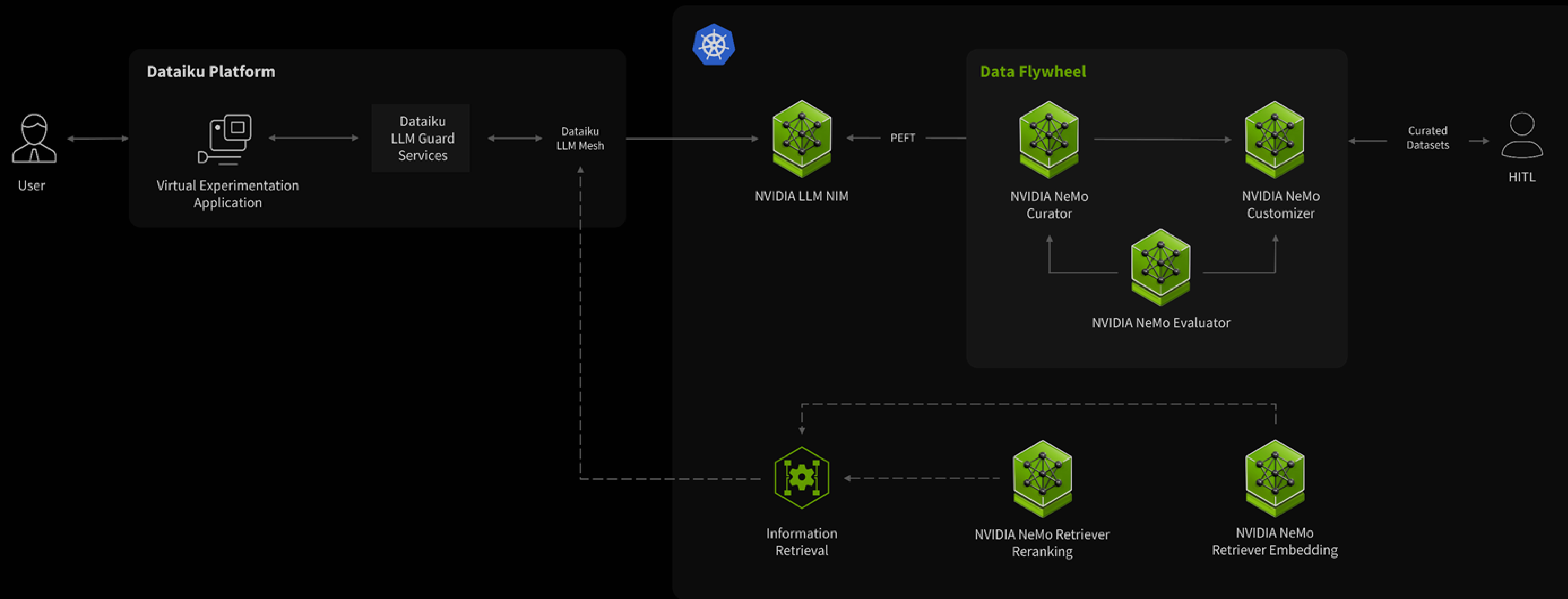


Triage Detection  
Summary

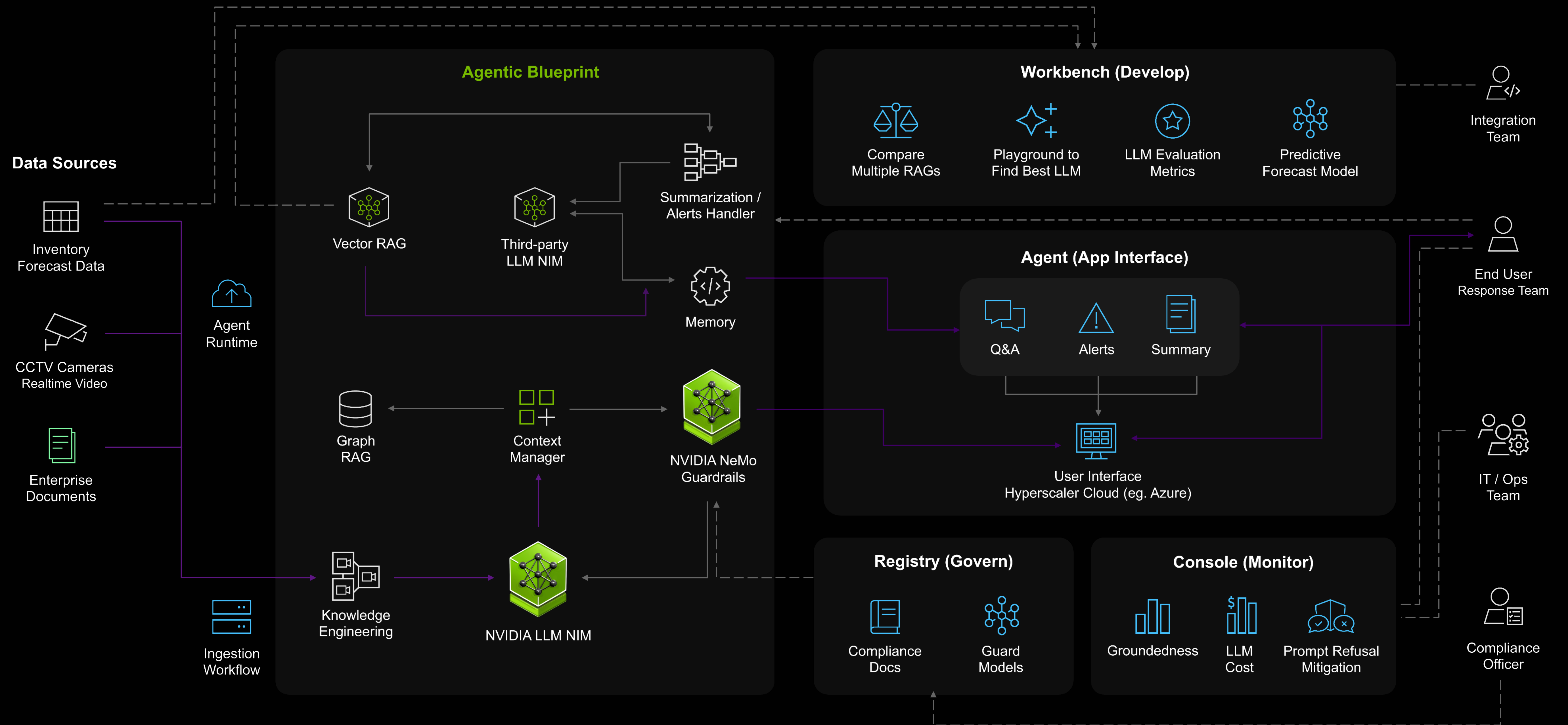


Human Analyst



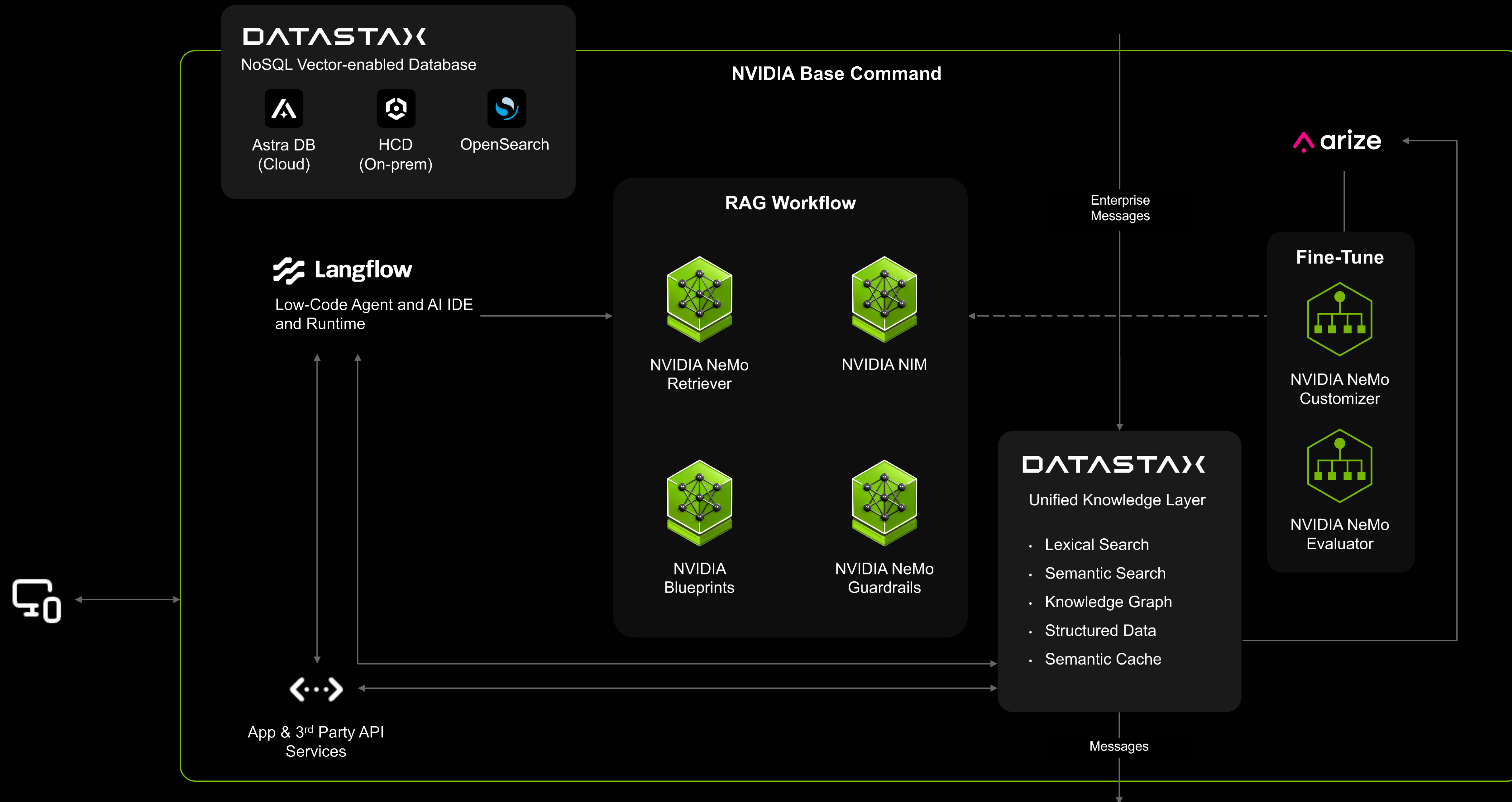








# DATASTAX







Multimodal Enterprise  
Data



NVIDIA NeMo  
Retriever Extraction



NVIDIA NeMo Retriever  
Embedding

Ingestion

Retrieval



Response



NVIDIA NeMo  
Guardrails



LLM



NVIDIA NeMo  
Retriever Reranking

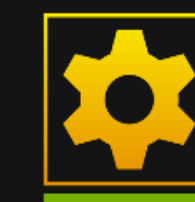


Elasticsearch Vector  
Database

NVIDIA cuVS



Search  
(CPU & GPU)



Indexing  
(GPU)

Query



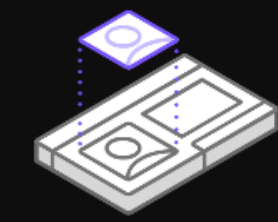
NVIDIA NeMo  
Guardrails



NVIDIA NeMo  
Retriever Embedding



# NUTANIX



Role Based  
Access Controls



Day 2  
Operations



Enterprise  
LLM Repository



AI User



RAG App  
/ Agent

## Inference Endpoint Service



NVIDIA NeMo  
Guardrails



NVIDIA NeMo  
Retriever Reranking

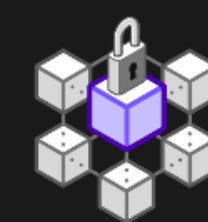


NVIDIA NeMo  
Retriever Embedding



NVIDIA Llama Nemotron  
Reason

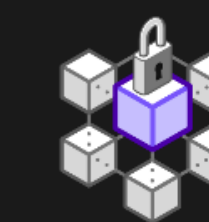
**RAG Inference  
Endpoint Service**



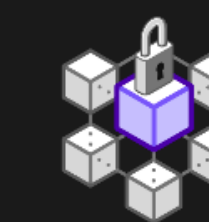
Guardrail  
Endpoint



Embedding  
Endpoint



Reranking  
Endpoint



Reasoning  
Endpoint



NVIDIA GPU Operator

Nutanix Kubernetes Platform (NKP)



NVIDIA Network Operator





NVIDIA NIM

Agentic App  
Development



TensorRT-LLM

vLLM

sGLang

Prompts and  
Responses



Optimized Profiles



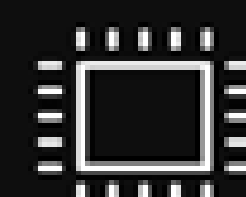
Model Monitoring & Observability



Metrics



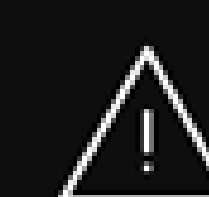
Telemetry



Accelerator Profiles



Logging



Alerting



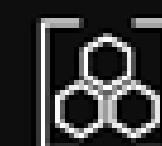
NVIDIA GPU Operator



NVIDIA Network Operator



NVIDIA NIM Operator



Enterprise Kubernetes



Enterprise Linux



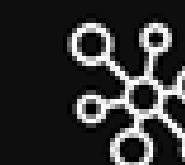
Container & VM Management



DevOps Tooling



Security

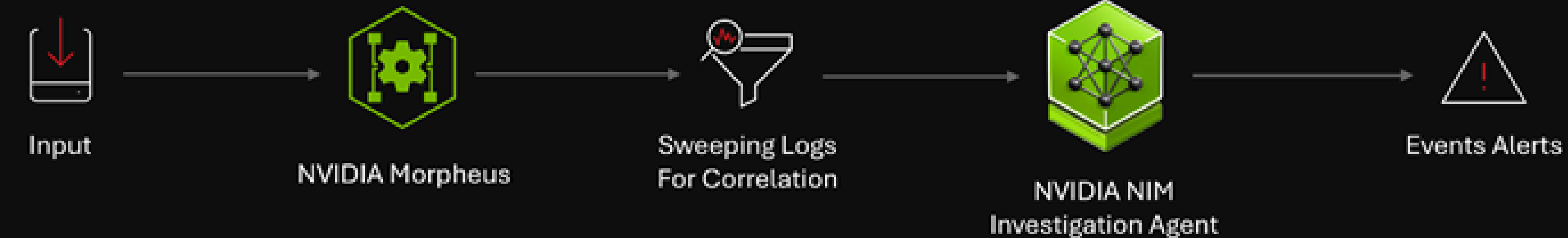


Ecosystem

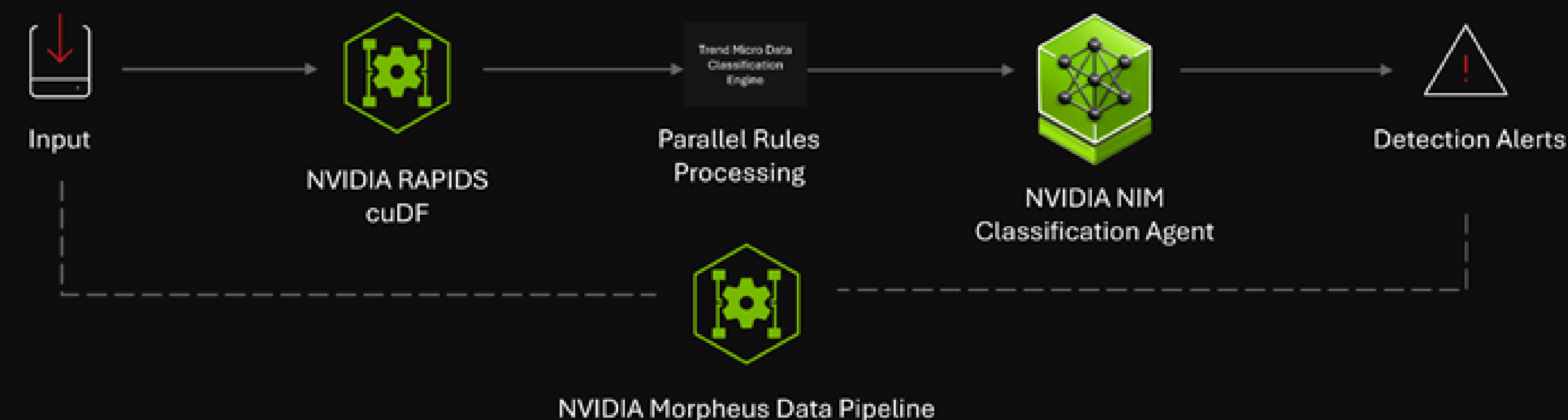


## Trend Vision One™

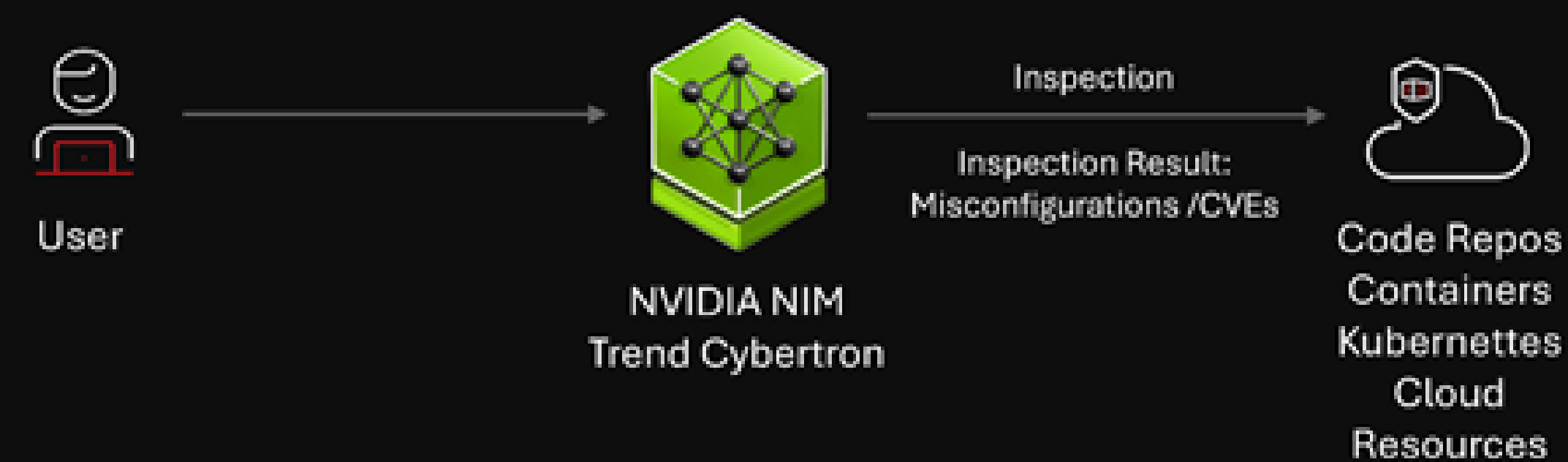
Extended Detection and Response (XDR)



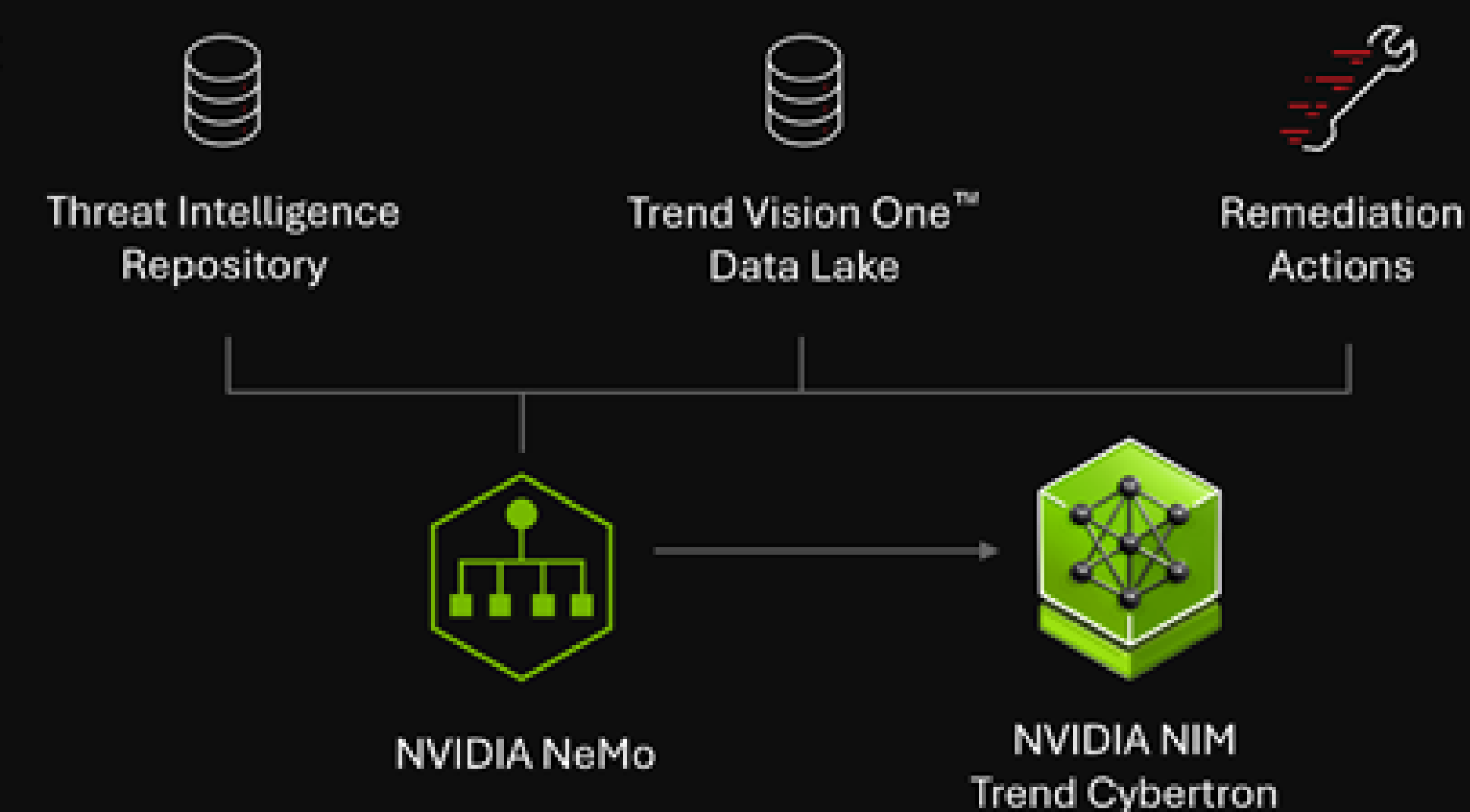
Data Security Posture Management



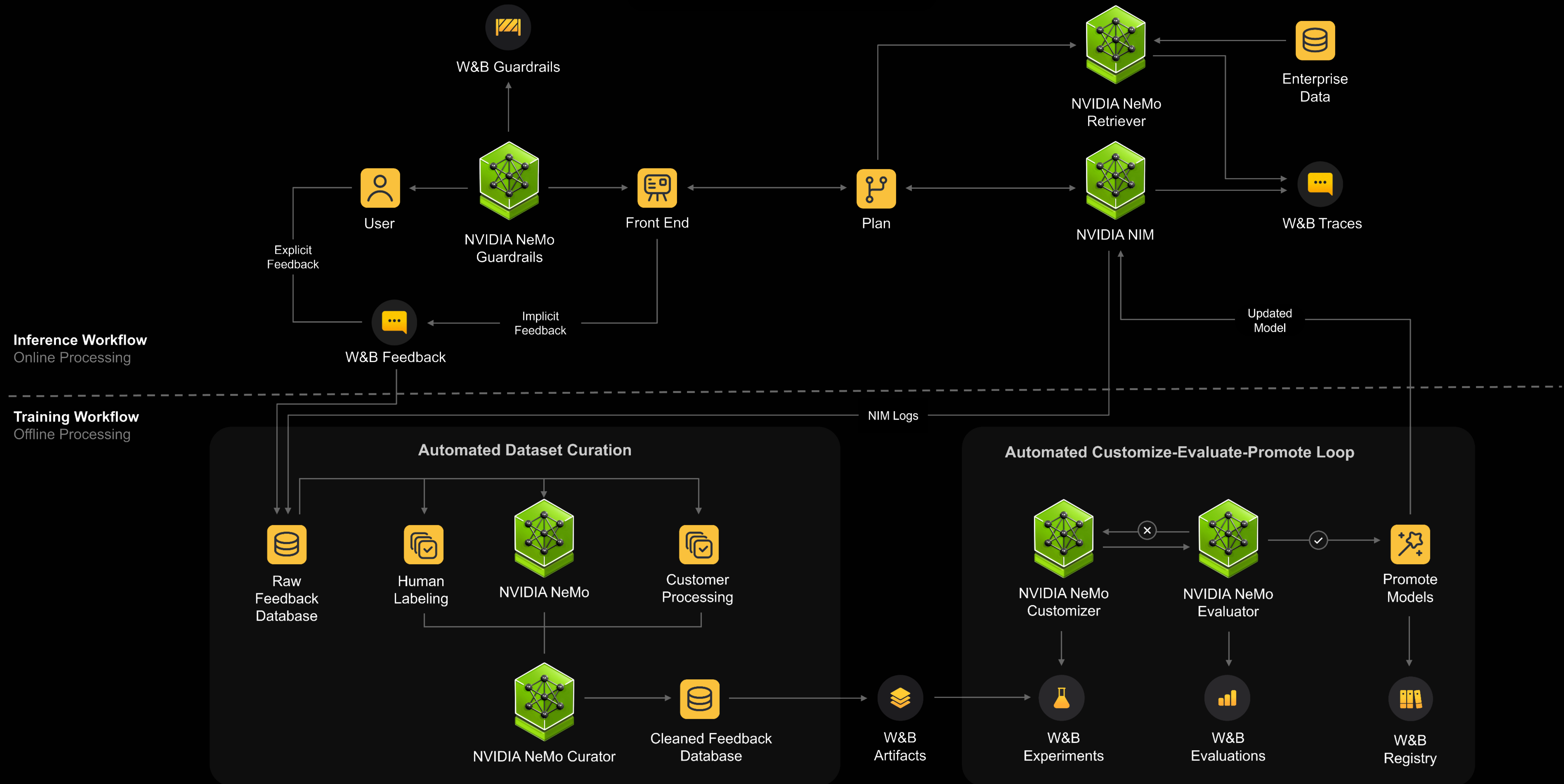
## Vulnerability Management



## Training









# NVIDIA Enterprise AI Platform Accelerates Trillion-Dollar Global IT Investment

Agentic AI

accenture

Infosys

Deloitte.

tcs

Deloitte.

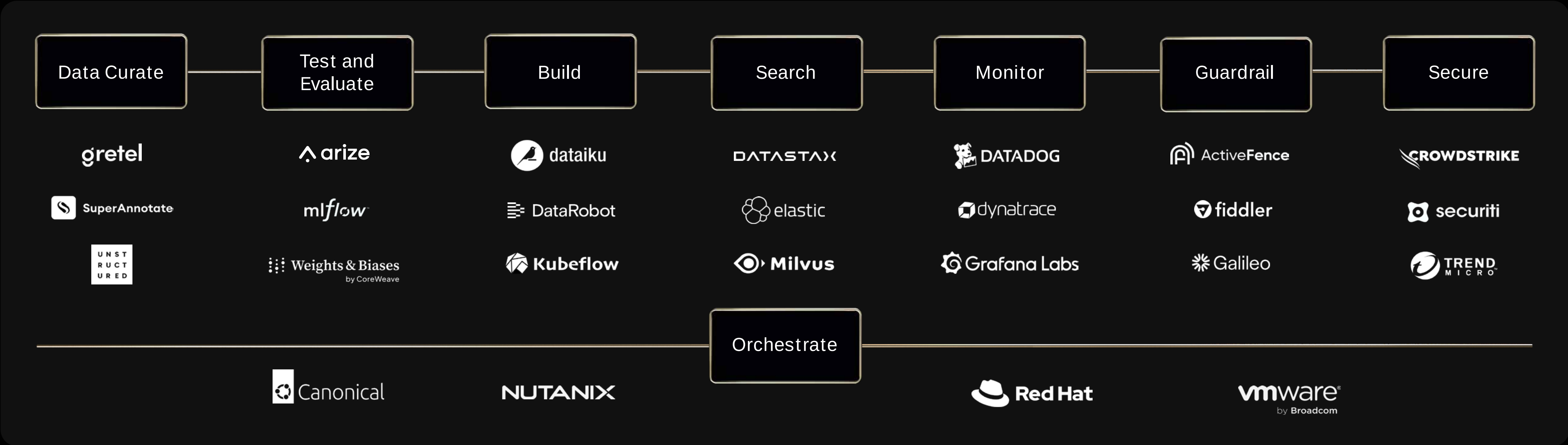
Wipro

EY

TATA CONSULTANCY SERVICES



AIOPS



AI Infrastructure

cisco

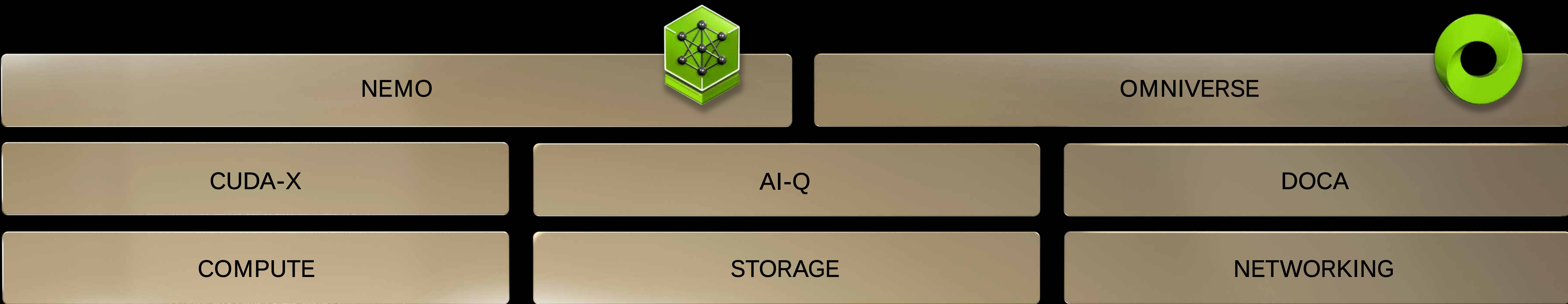
Hitachi Vantara

DELL Technologies

Lenovo

Hewlett Packard Enterprise

NetApp

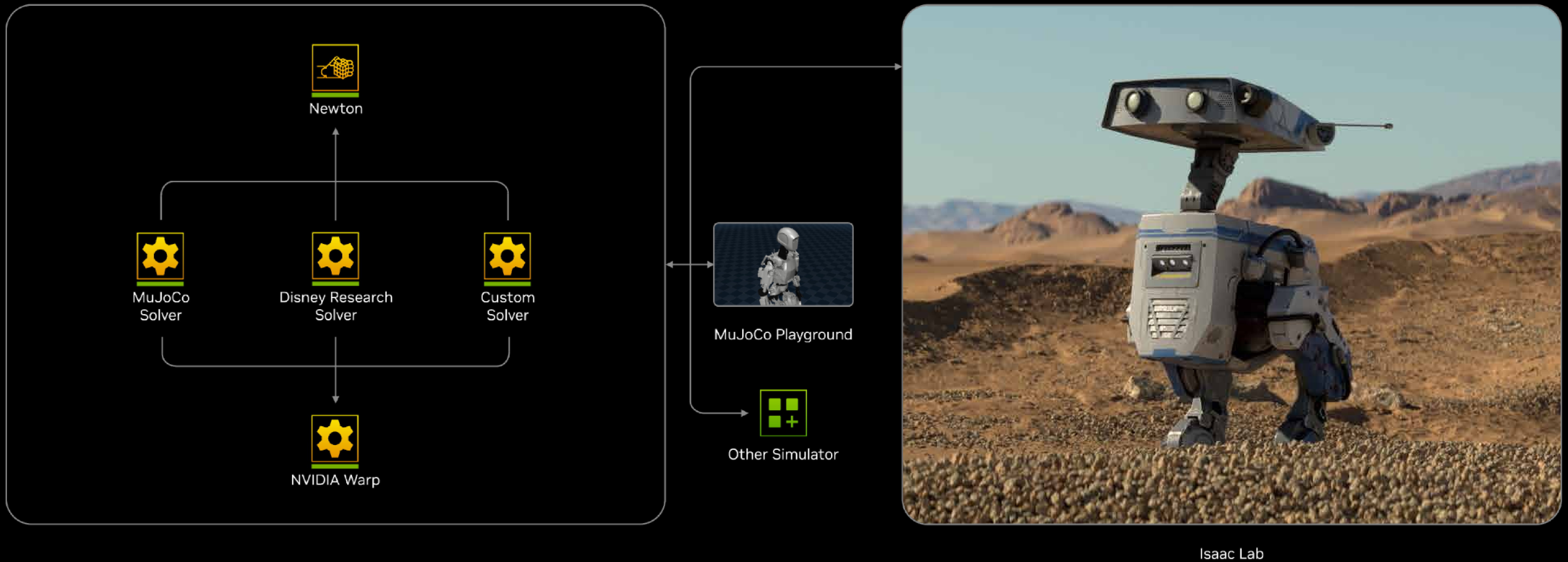




Disney Research

 NVIDIA

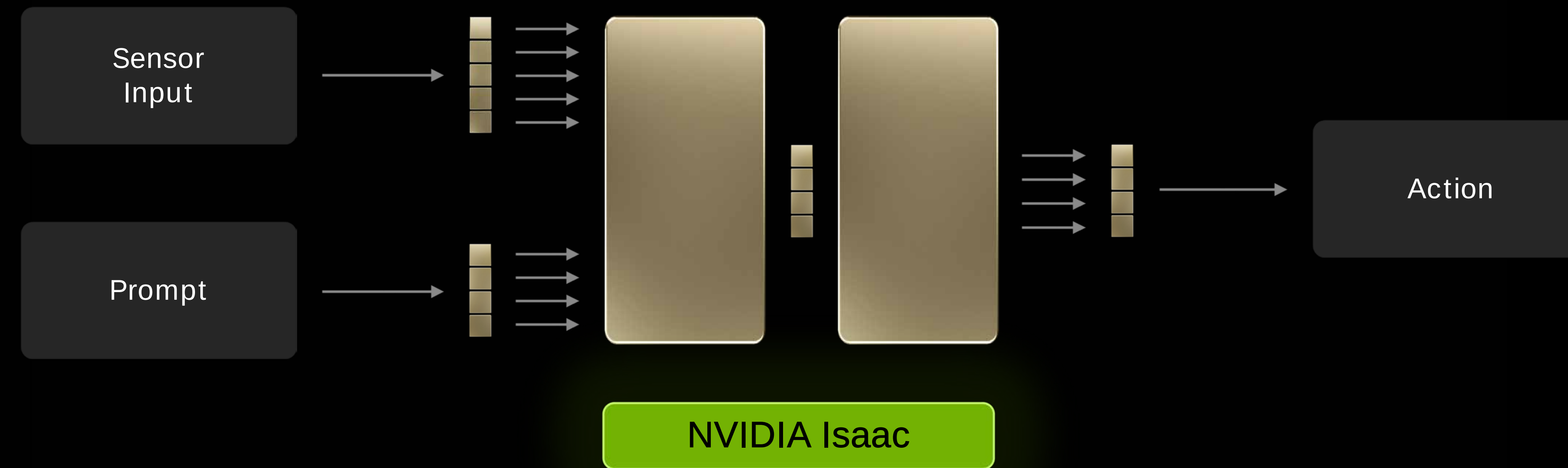
 Google DeepMind





# NVIDIA Isaac GR00T

## Open Humanoid Robot Development Platform

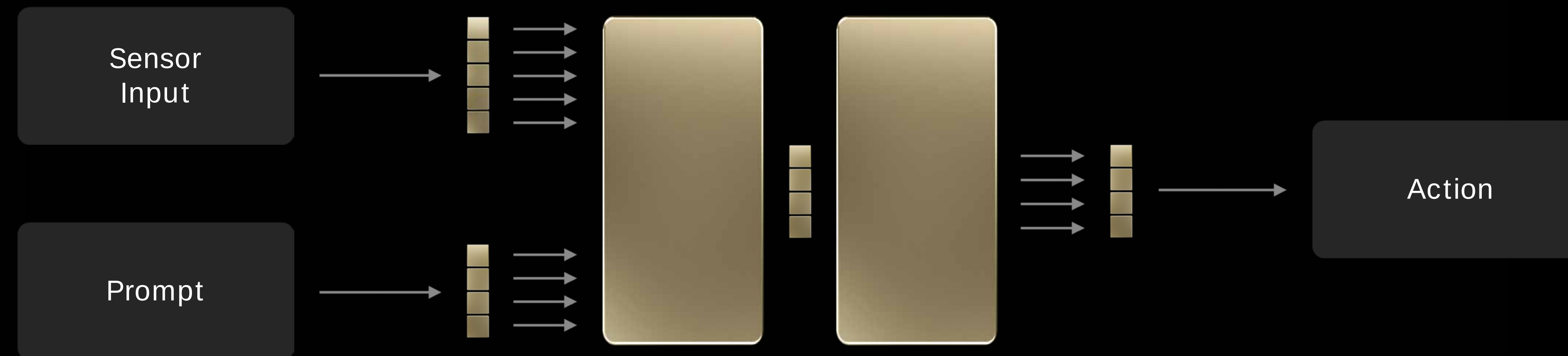


Jetson Thor



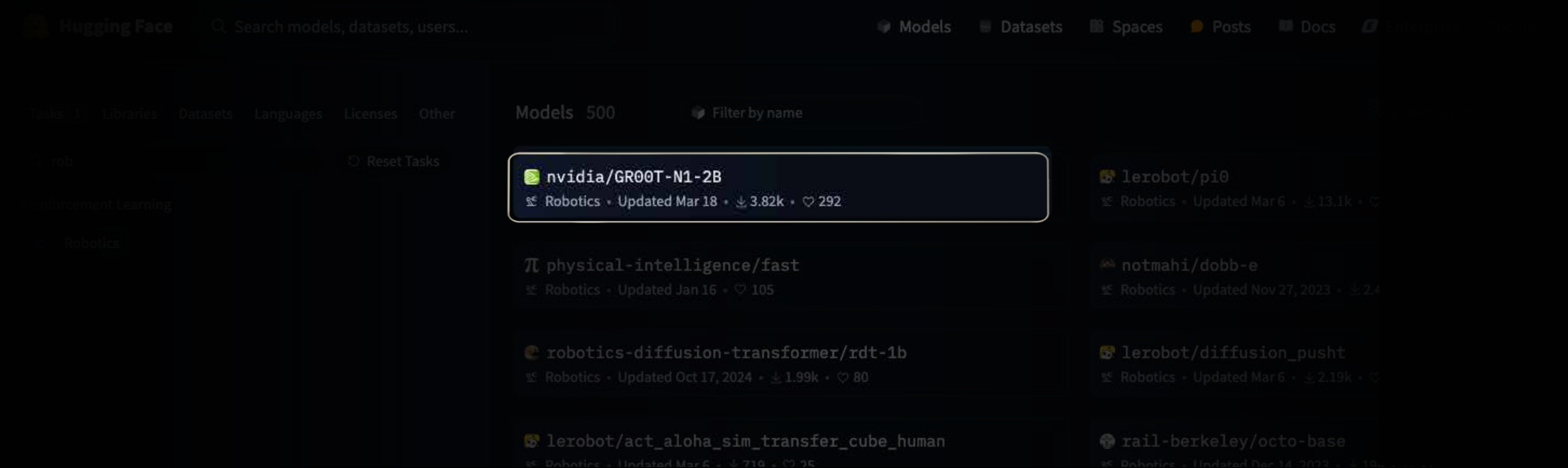
# Announcing NVIDIA Isaac GR00T N1.5

## Open Humanoid Robot Foundation Model



6K  
Model Downloads

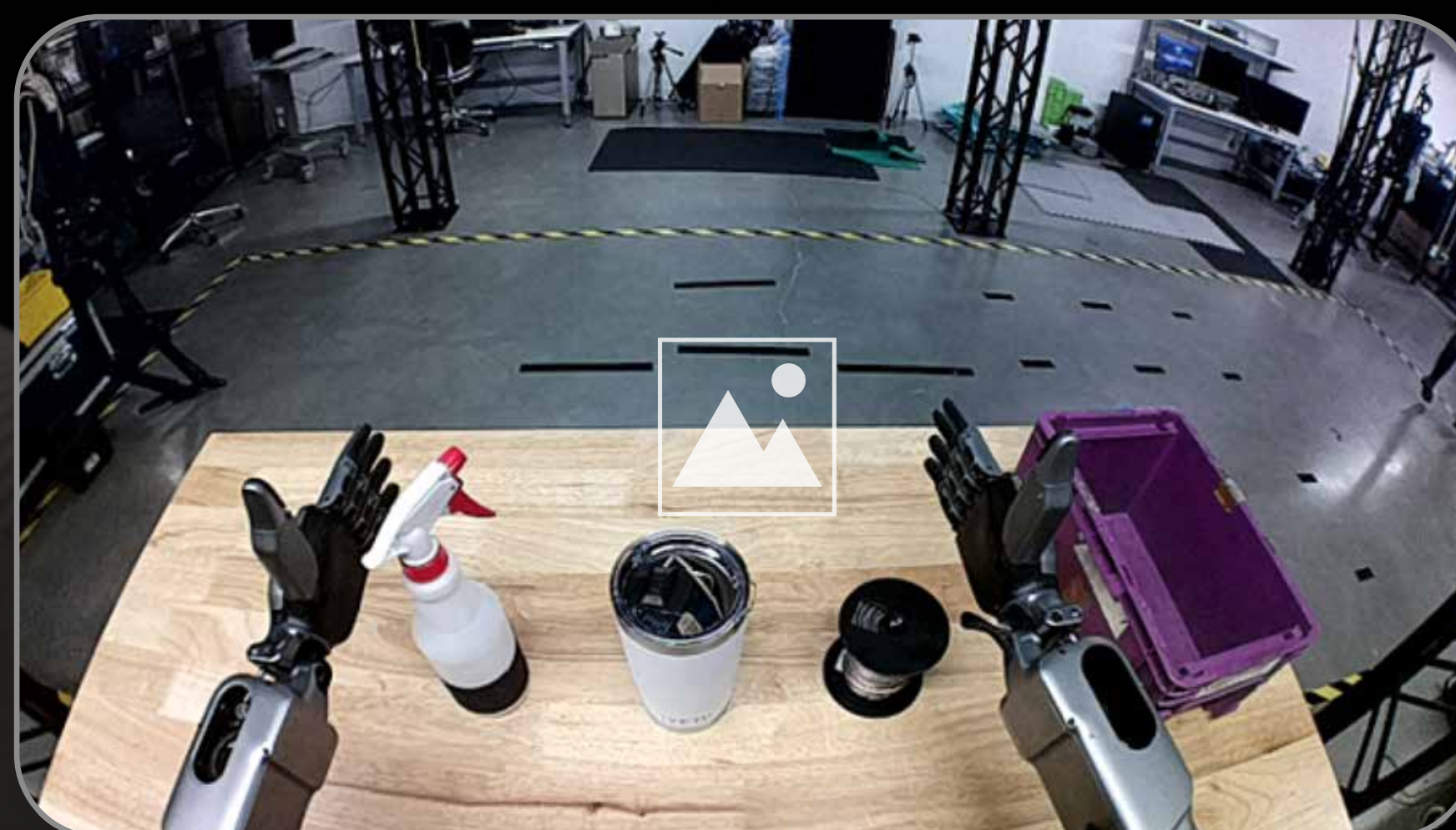
3.6K  
GitHub Stars





# Announcing NVIDIA Isaac GR00T-Dreams

## A Small Human Demonstration Team Can Do the Work of Thousands



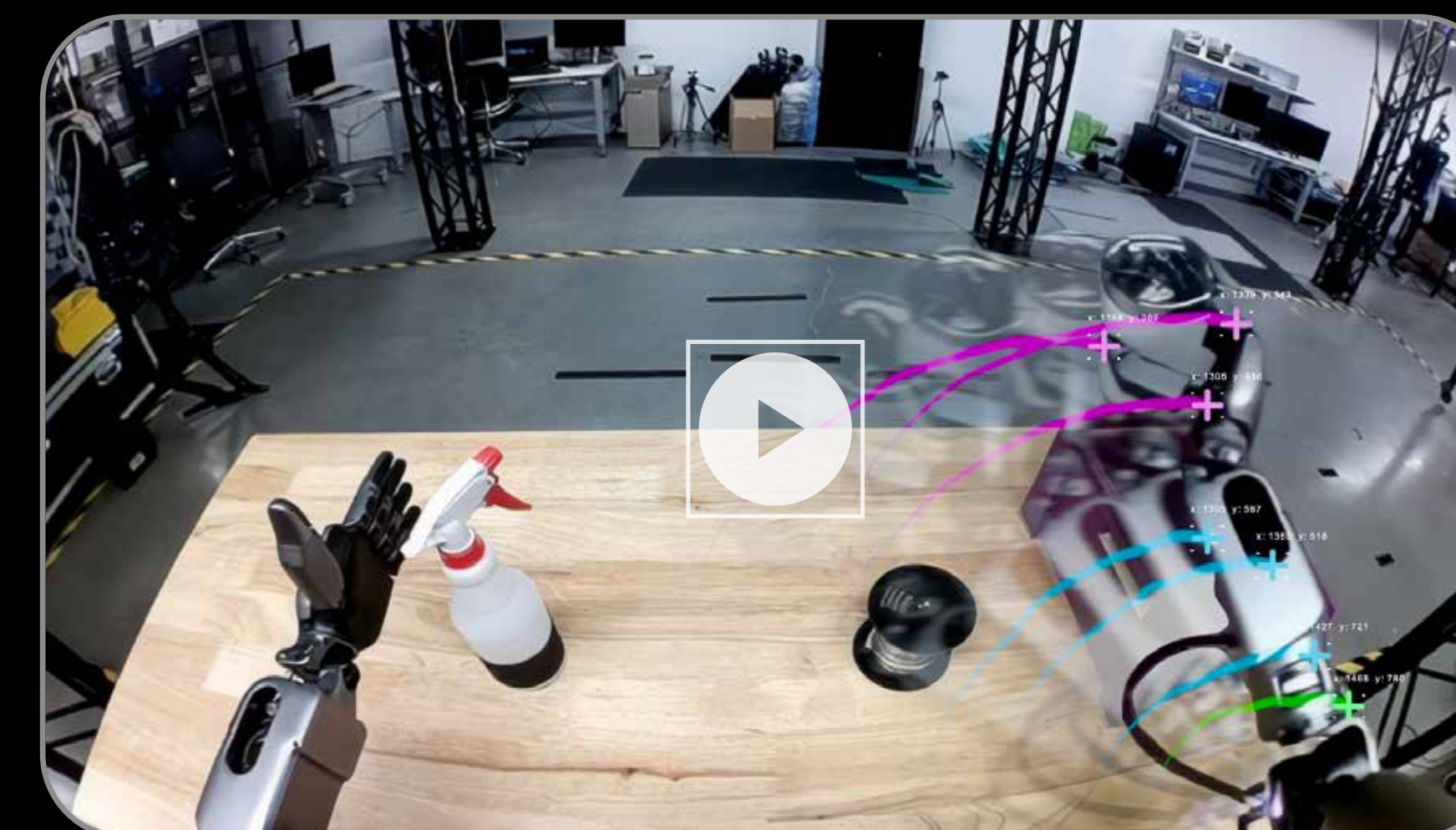
Pick up the object on the table that is used for drinking water



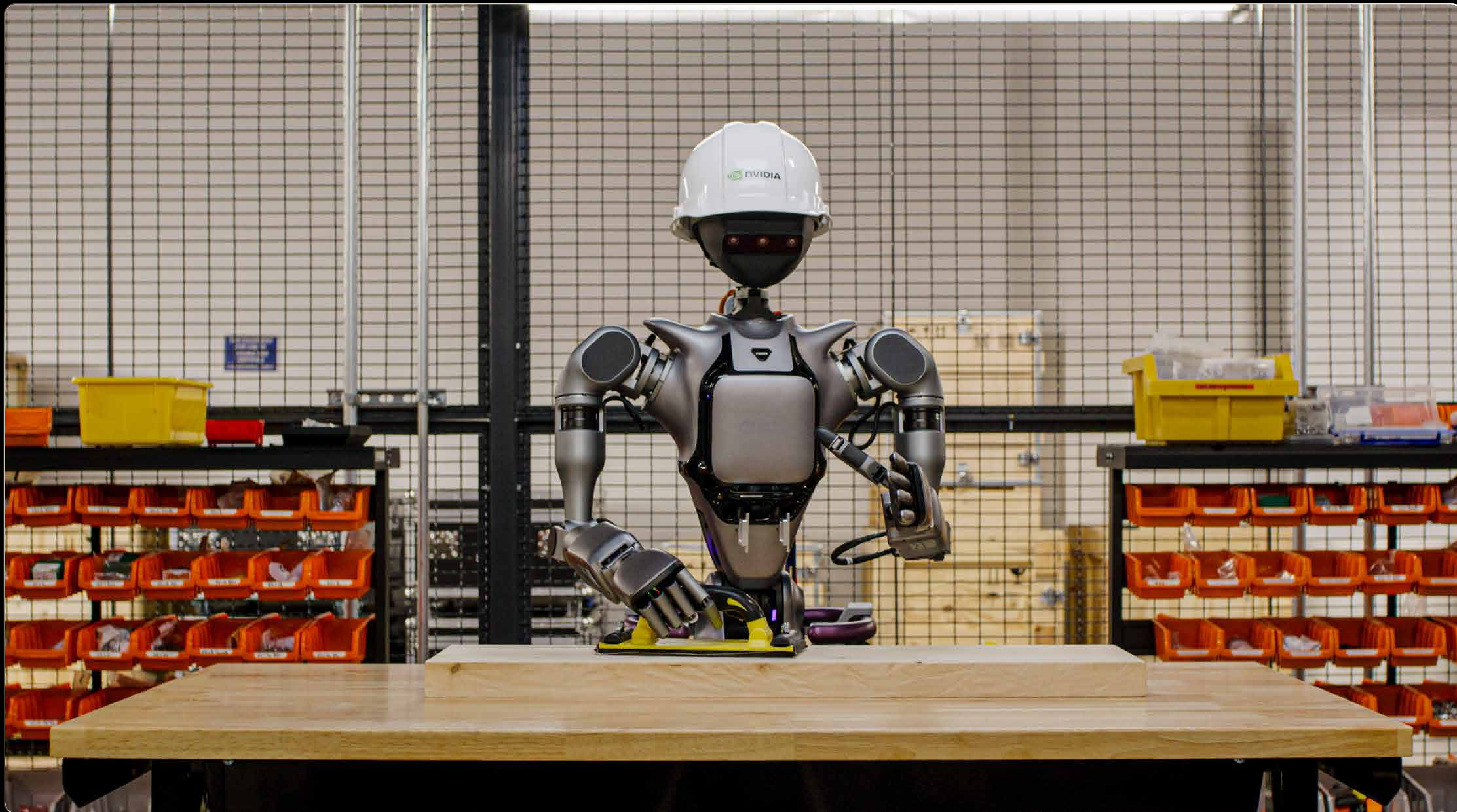
NVIDIA  
Cosmos



Action  
Labeling











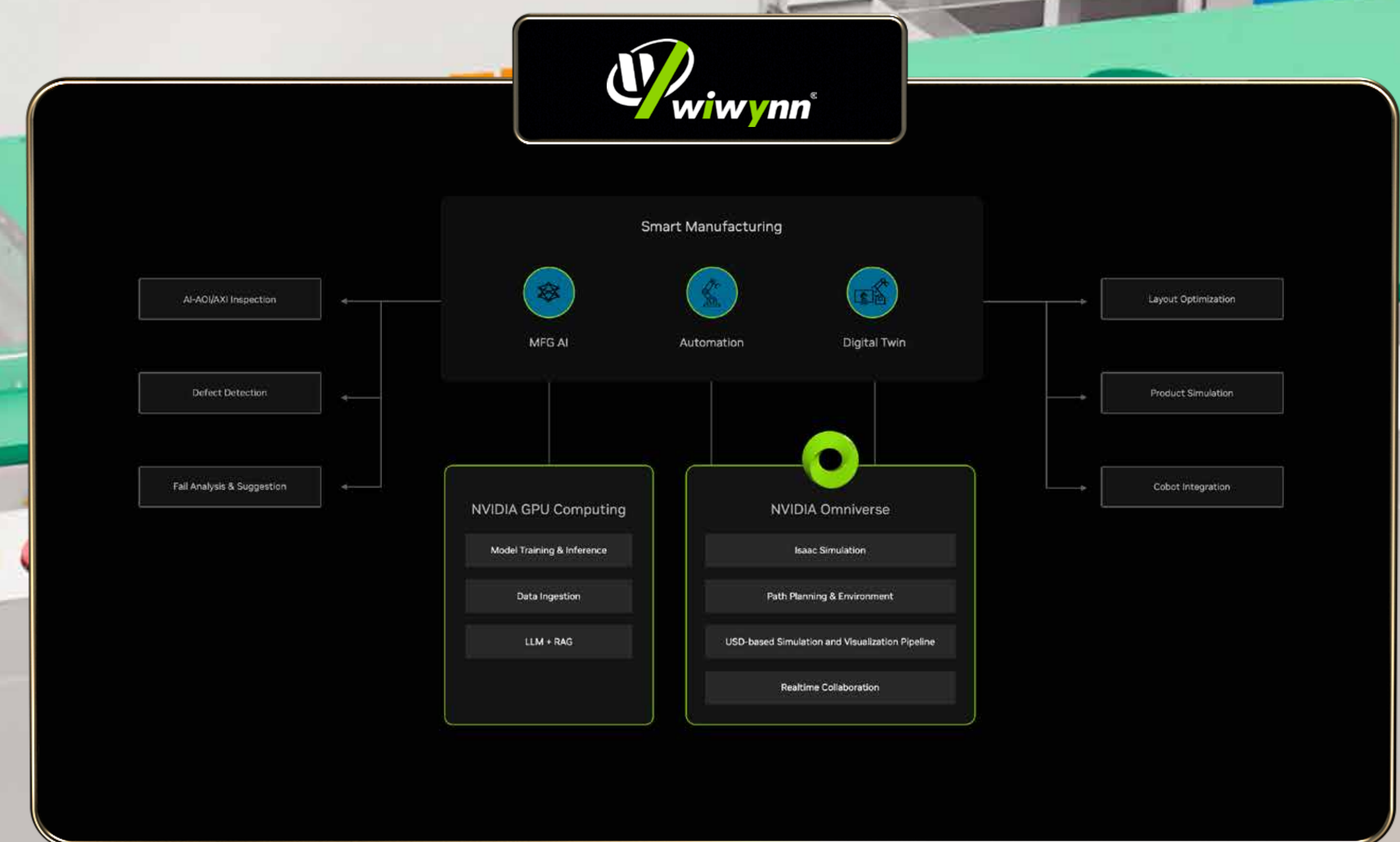
Omniverse Digital Twin







Omniverse Digital Twin







Omniverse Digital Twin







| Industry                   | SaaS                                                                                                                                                                                                                                                                                                                                                                                                   | PaaS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | IoT                                                                                                                                                                                                                                                                                                                                                                                                                |
|----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <br>Electronic Information | <div>Digitization and Intelligence</div> <div><br/>Manufacturing Operations Management (MOM)</div> <div><br/>Intelligent Park</div> <div><br/>Industrial AI/VR</div> <div><br/>Industrial Security</div> <div><br/>Geometry Twin</div> <div><br/>Automation Simulation</div> <div><br/>Humanoid Robot</div> <div><br/>CPU + Multi-Axis</div> <div><br/>VISA/MTMC</div> <div><br/>Enterprise Twin</div> | <div>NVIDIA Omniverse Platform</div> <div>USD, RTX, NuNet, Connect</div> <div><br/>USD Composer</div> <div><br/>Isaac Sim</div> <div><br/>CuMotion</div> <div><br/>MultiAxis</div> <div><br/>MTMC</div> <div><br/>Web Viewer</div> <div>General Intelligence Large Model</div> <div>General Large Model</div> <div>Industry Large Model</div> <div>Third-Party Large Models</div> <div>Model</div> <div>Design</div> <div>GPUs</div> <div>Platform Tools</div> <div>IoT Platform</div> <div>Control Platform</div> <div>Low-code Platform</div> <div>3D Web Editing</div> <div>AR/VR Editing</div> | <div>Robotization</div> <div>Autonomous Mobile Robots</div> <div><br/>Conveying / Composite / Forklift / Lateral / Top LPT / Camera Transmission</div> <div>Automation</div> <div>Integration and Design</div> <div><br/>Automated Production Line Design</div> <div><br/>Intelligent Logistics Planning</div> <div><br/>Semiconductor Packaging and Testing</div> <div><br/>Intelligent Temperature Control</div> |

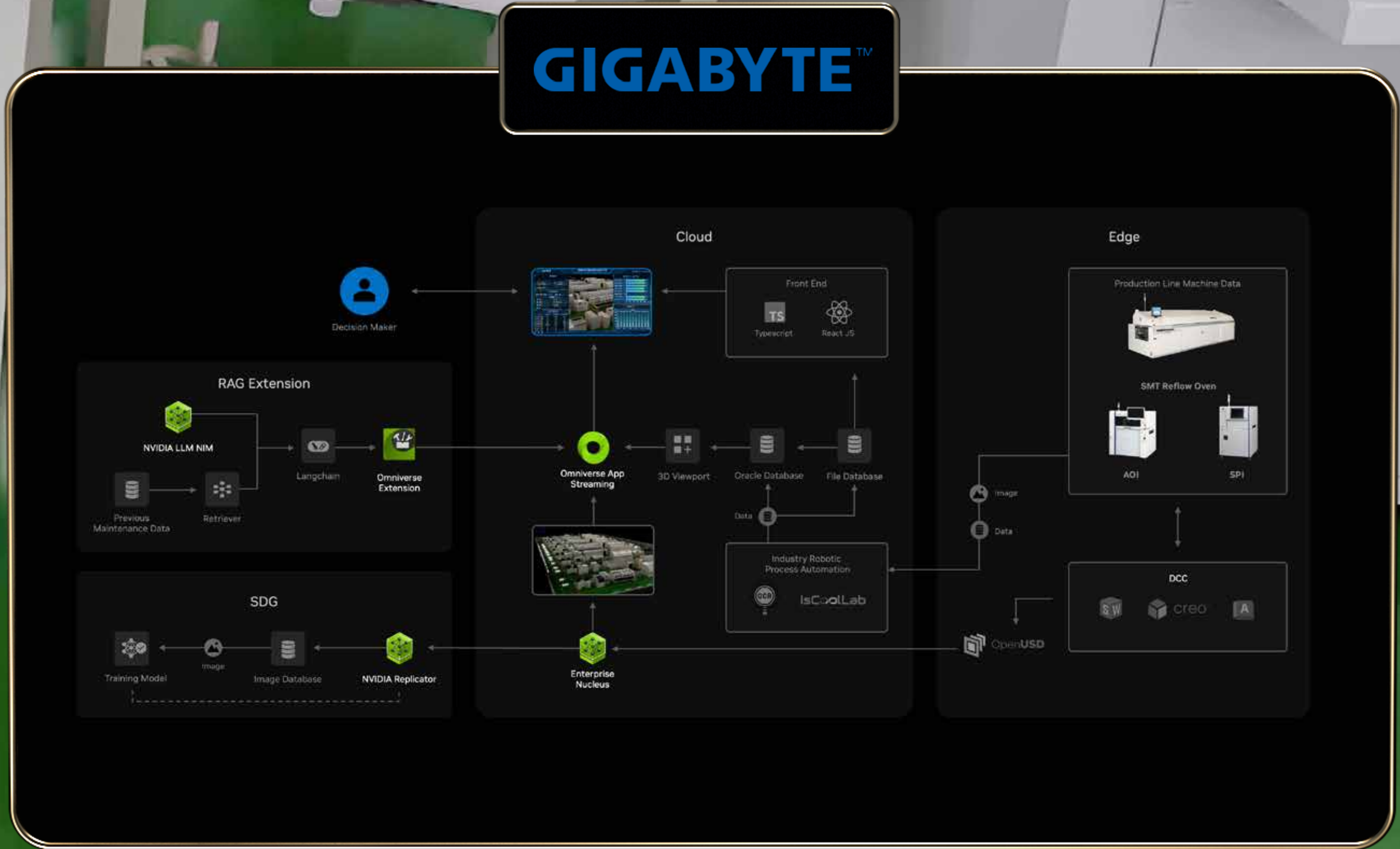




GIGABYTE

GPX-C  
FLUX

GIGABYTE™



Omniverse Digital Twin





Omniverse Digital Twin





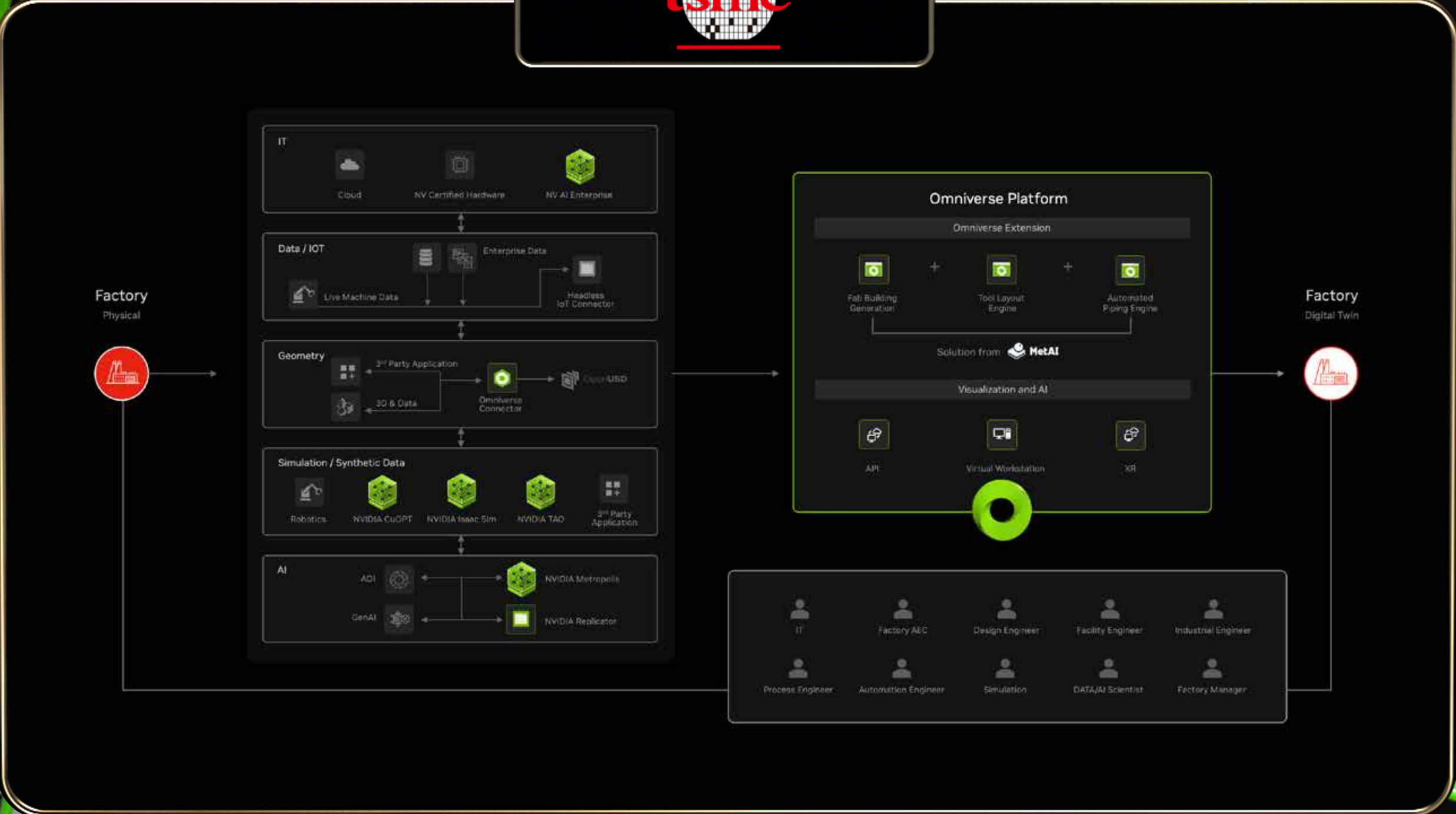
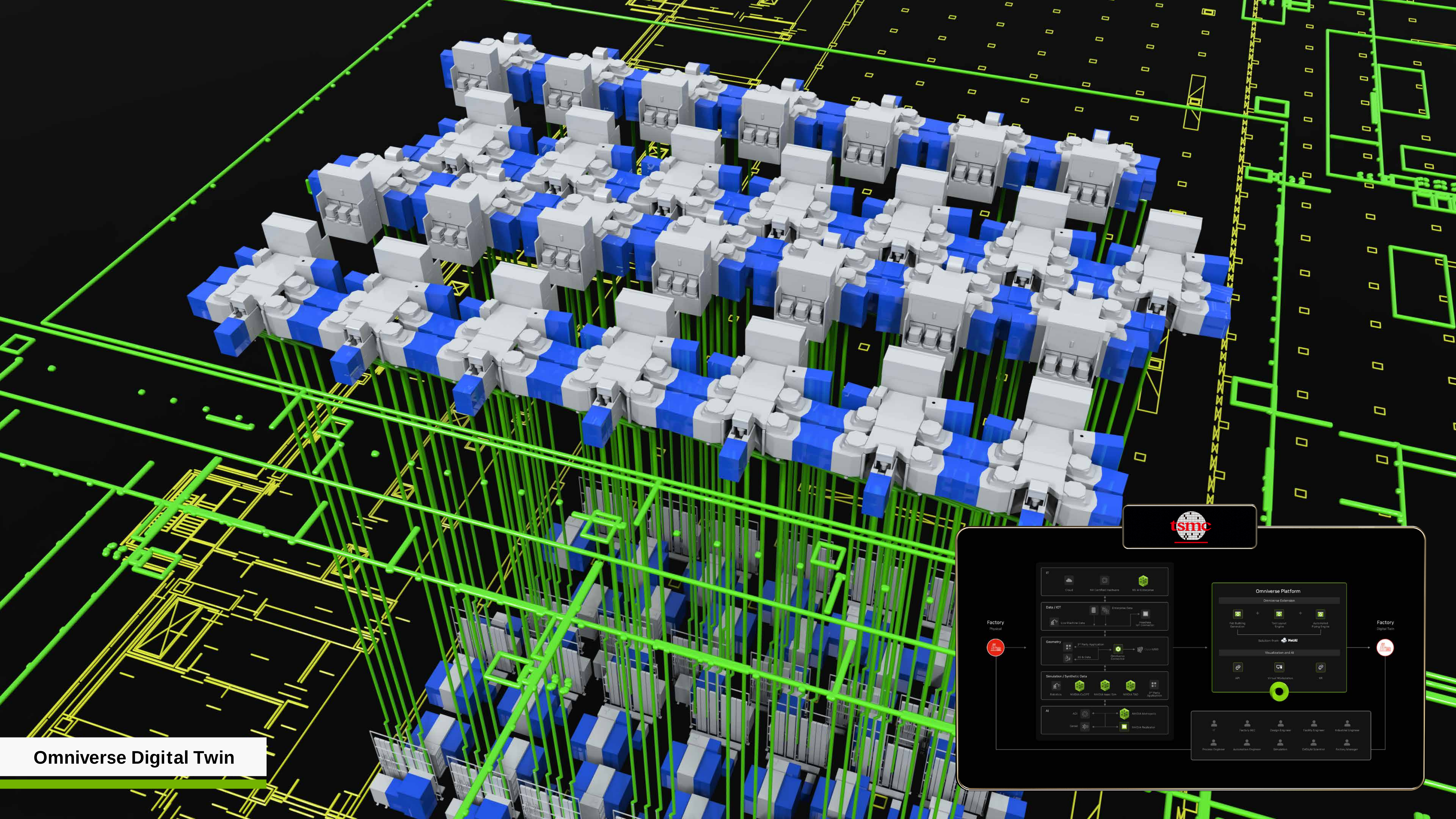
Module Screen

wistron



Omniverse Digital Twin





Omniverse Digital Twin



cā dence

Crusoe

DELTA

etap

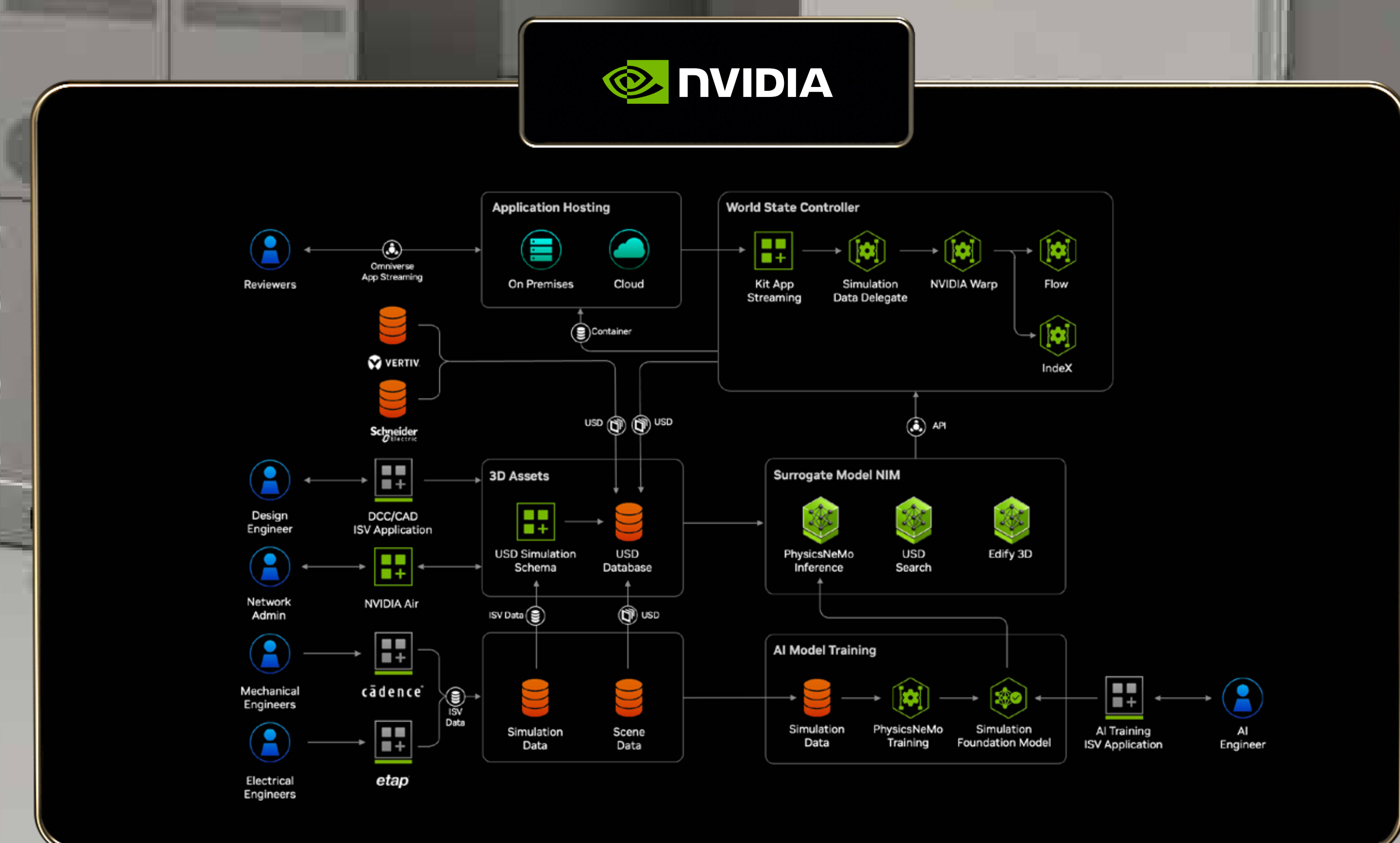
motivair  
by Schneider Electric

Schneider  
Electric

SIEMENS

VERTIV

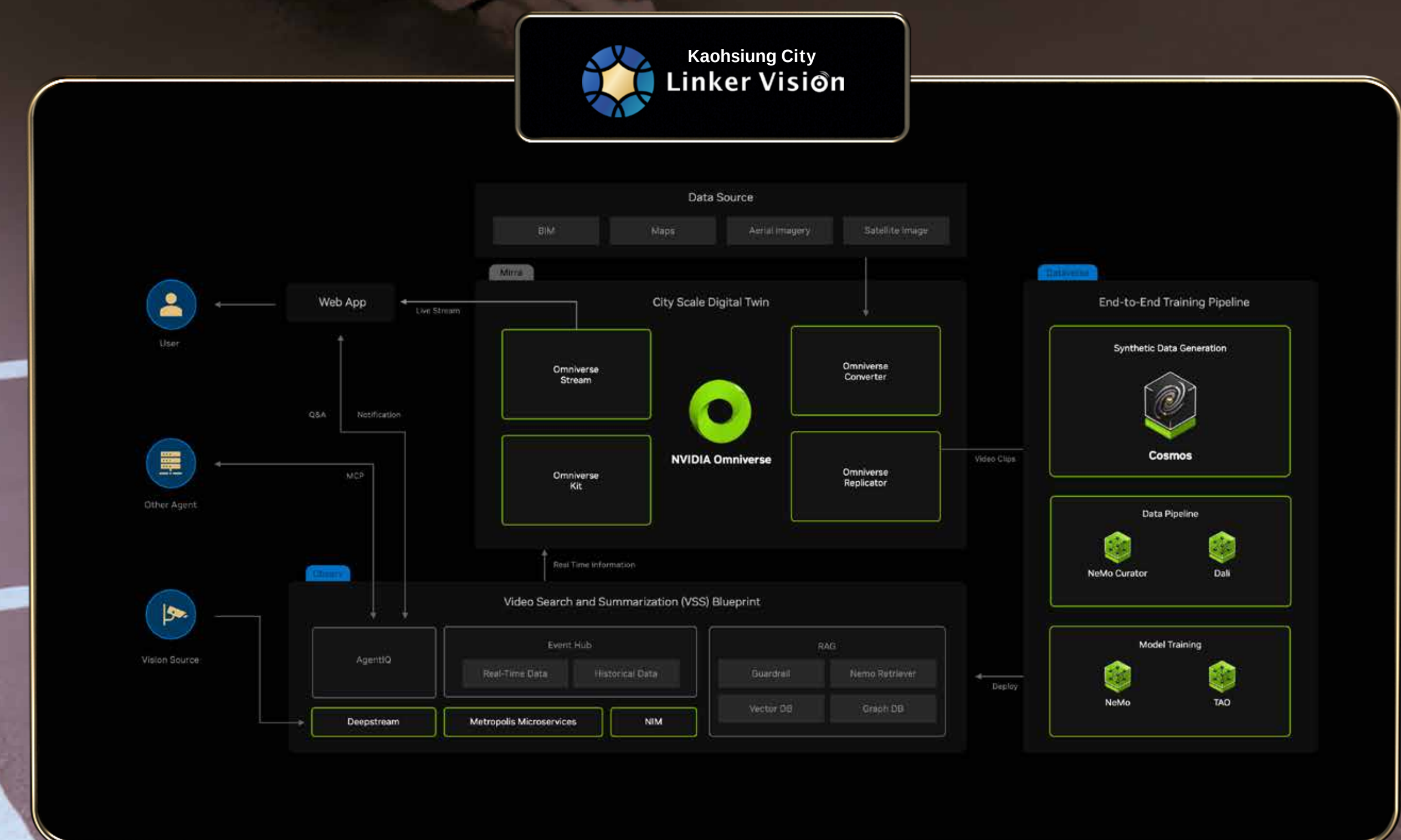
## Omniverse Digital Twin



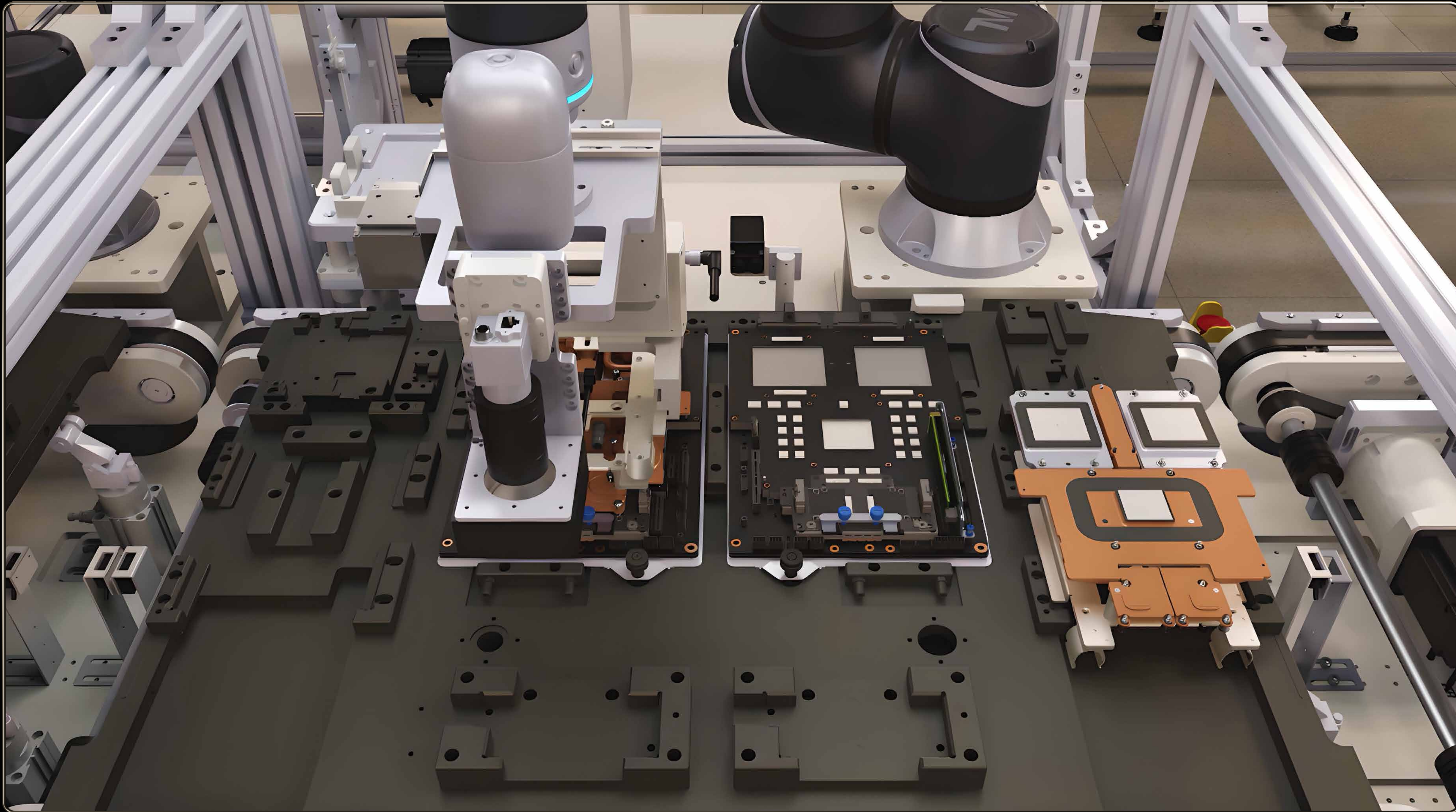




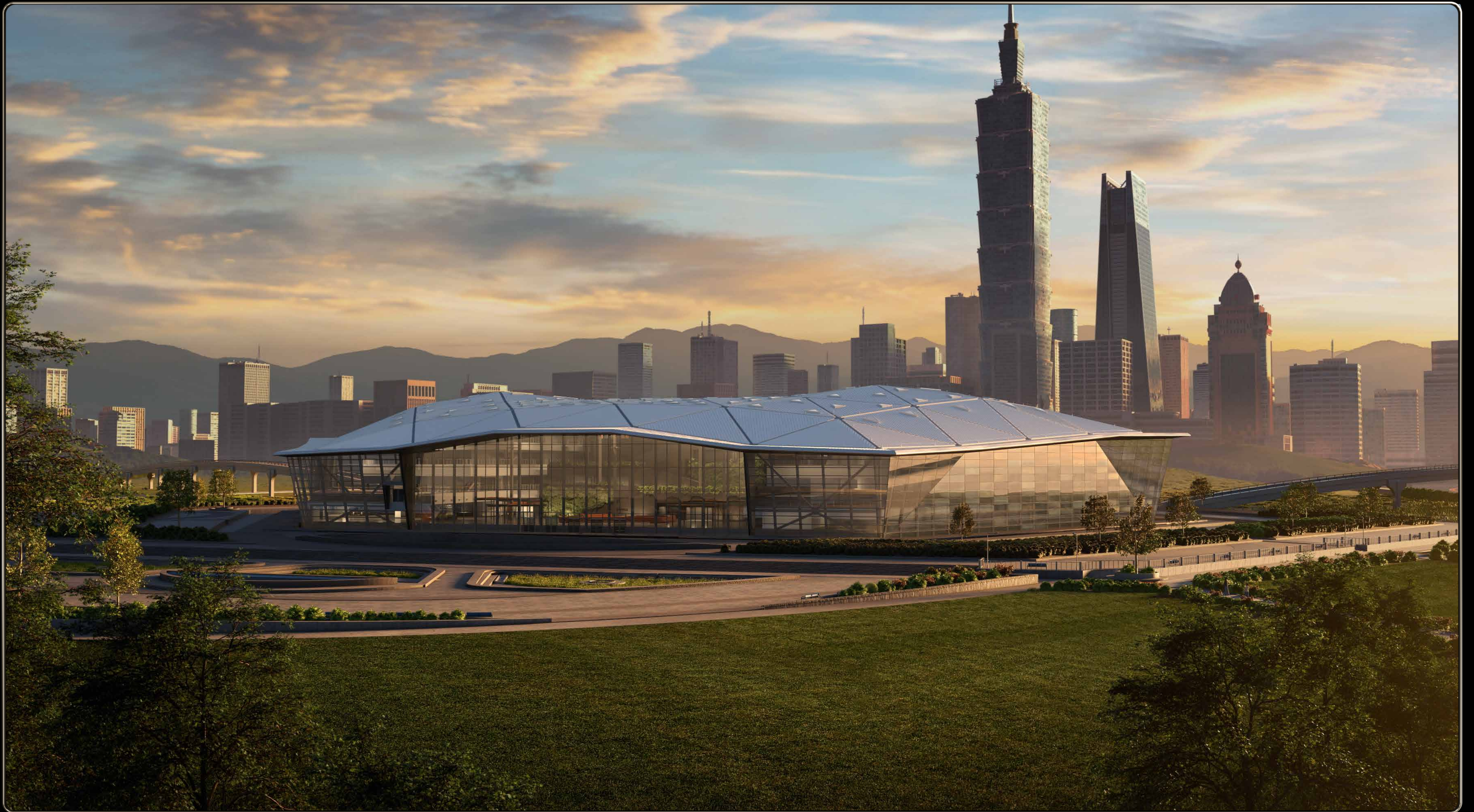
## Omniverse Digital Twin













NVIDIA

AI Factories, Agents, Robots | Amazing Partners | One Architecture

acer

ADVANTECH

ASRock  
Rock

ASUS

BOX

CISCO

COMPAL

ddn

DELLTechnologies

FOXCONN  
HON HAI TECHNOLOGY GROUP

GIGABYTE

H3C

Hewlett Packard  
Enterprise

Hitachi Vantara

hp

IBM

ingrasys

Inventec

Lambda

Lenovo

MITAC  
MITAC Computing

msi

NetApp

NUTANIX

PEGATRON

PURESTORAGE

QCT

R A Z E R

SUPERMICR

VAST

WEKA

wlstron

wiwynn



