



Science-fiction > fact > deploy

ALISON B. LOWNDES | SENIOR SCIENTIST | GLOBAL AI

[@alisonblowndes](https://twitter.com/alisonblowndes)

A41109

What to expect in 2023

<https://blogs.nvidia.com/blog/2022/12/13/2023-ai-predictions/>

- large-scale digital twins bringing million-x acceleration to insight & optimisation of manufacturing efficiency, streamlining of operations & **safety**.
- generalist AI (not generative) able to continuously learn & solve open-ended tasks with LLMs, RL & foundation models.
- unified AI pipelines across hybrid cloud/on-prem infra for a single, connected experience balancing costs against strategic objectives.
- generative AI (genAI), LLMs & recommender systems (Merlin) transforming production applications & fuelling advances in personalized customer service*.
- coupling LLMs with the ability to describe biology & chemistry in highly capable AI models for drug discovery & potential therapies.
- photorealistic rendering and accurate physics modelling combined with the ability to simulate - in parallel - millions of instances of a robot or AV while genAI creates highly realistic 3D simulation scenarios & synthetic data to further accelerate training of robust systems that have encountered & learnt from any possible scenario.
 - increasingly robust robots means more ability to be deployed with & learn from human interaction - learning to safely operate & understand the deployment environment (various workplaces, indoor and outdoors).
- USD enables connection to Omniverse by any tool, viewer & browser seamlessly & consistently. Companies will save on waste, cost & time (increase operational efficiencies).
- Combining AI and physics will continue to improve prediction and provide real-time updates as climate conditions change. Also enable more frequent and more accurate forecasting, ensuring fast & accurate supply chains & route optimization (cuOpt).
- AI-based risk evaluation in real-time to improve the value firms deliver to their investors & customers & deliver operational resiliency.
- Accelerated transition to energy-efficient, green computing with lower total cost of ownership (TCO), reduced carbon footprint, harnessing custom combinations of CPU, GPU & DPU.
 - electric utility companies deploy edge AI to improve operational efficiency, enhance functional safety, increase accuracy of load and demand forecasting, and accelerate the adoption of renewable energy. AI at the edge will increase grid resiliency, while reducing energy waste and cost. Better planning for renewable energy expenditures, predict generation capacity and prepare for severe weather events.
- Bespoke zero-trust AI cybersecurity systems will discover more threats & prevent more breaches, bringing increased transparency as risk of being caught grows.



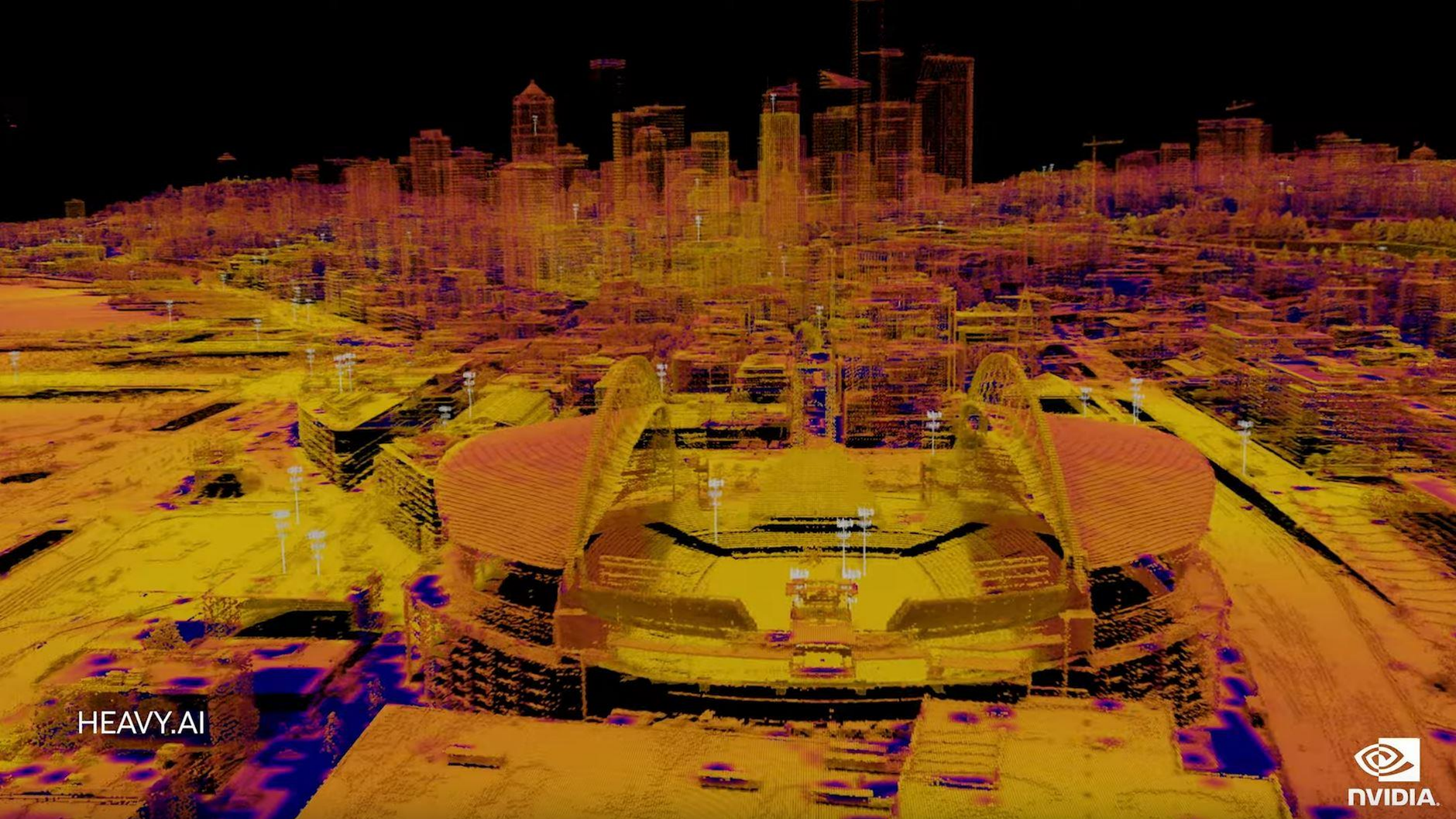
Automatic door

Automatic door

Food For Now

Doritos
Doritositos
Doritos
Doritos
Doritos





HEAVY.AI



NVIDIA CUOPT

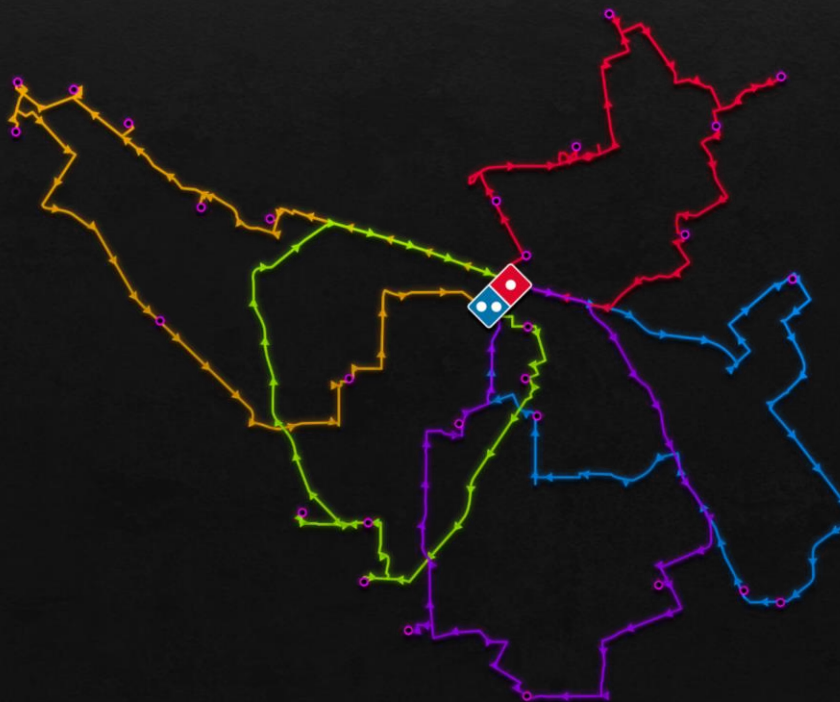
Re-Optimize Logistics and Supply Chain in Real-Time

Accelerated Solver for Vehicle Route, Warehouse Picking, Fleet-Mix Optimization

Massively Parallel Algorithm Generates Thousands of Solution Candidates and Refinements

Dynamic Rerouting Reduces Travel Time – Save Billions for a \$10 Trillion Logistics Industry

Available Now
developer.nvidia.com/cuopt-logistics-optimization



World Record Accuracy

2.96% Gap on Gehring and Homberger

Scalable to 1,000s of Locations

3 Seconds vs 5 Minutes

to Route 1,000 Packages



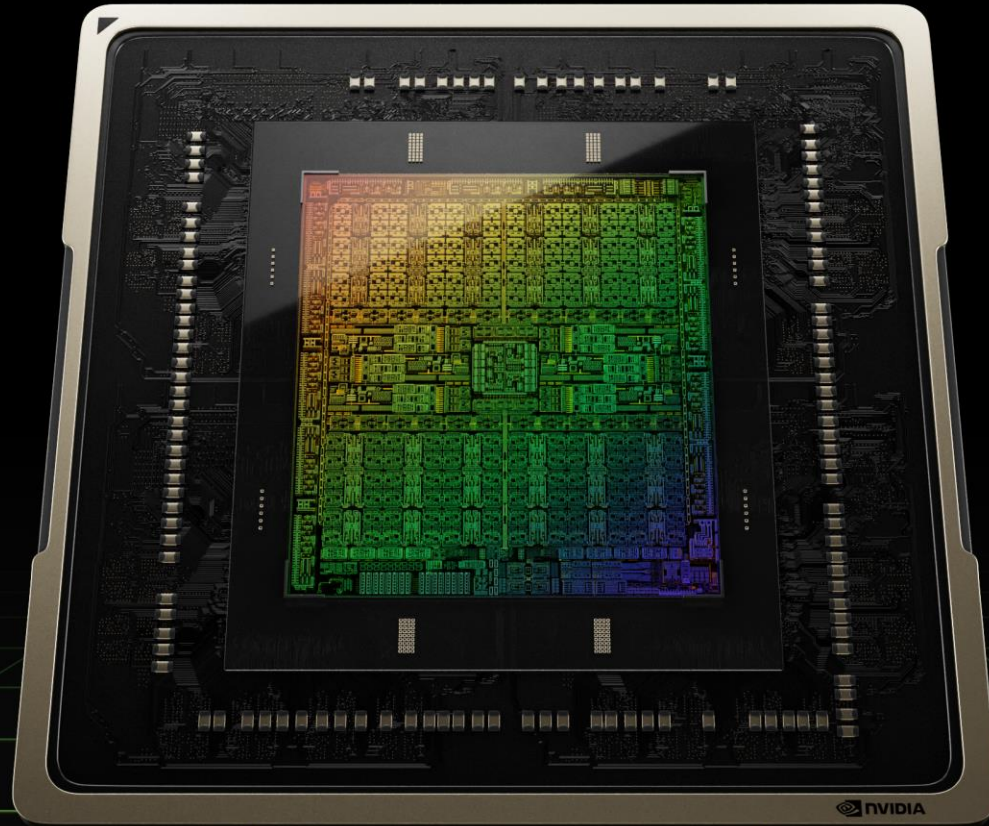
NVIDIA AI

NVIDIA BASE COMMAND

NVIDIA FLEET COMMAND

NVIDIA ADA LOVELACE

76 Billion Transistors | TSMC 4N Process | Micron G6X Memory



Shaders

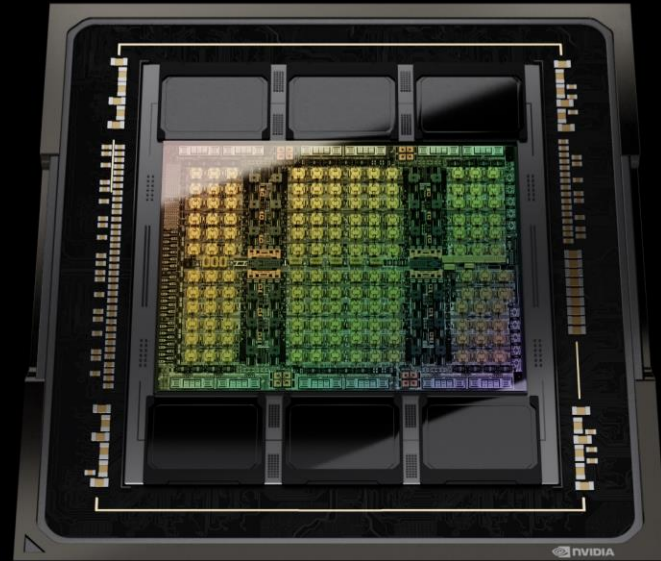
New Streaming Multiprocessor
90 Shader TFLOPs
2X Power Efficiency

Ray Tracing

3rd-Gen RT Cores
200 RT TFLOPs
2X Ray-Triangle Intersection

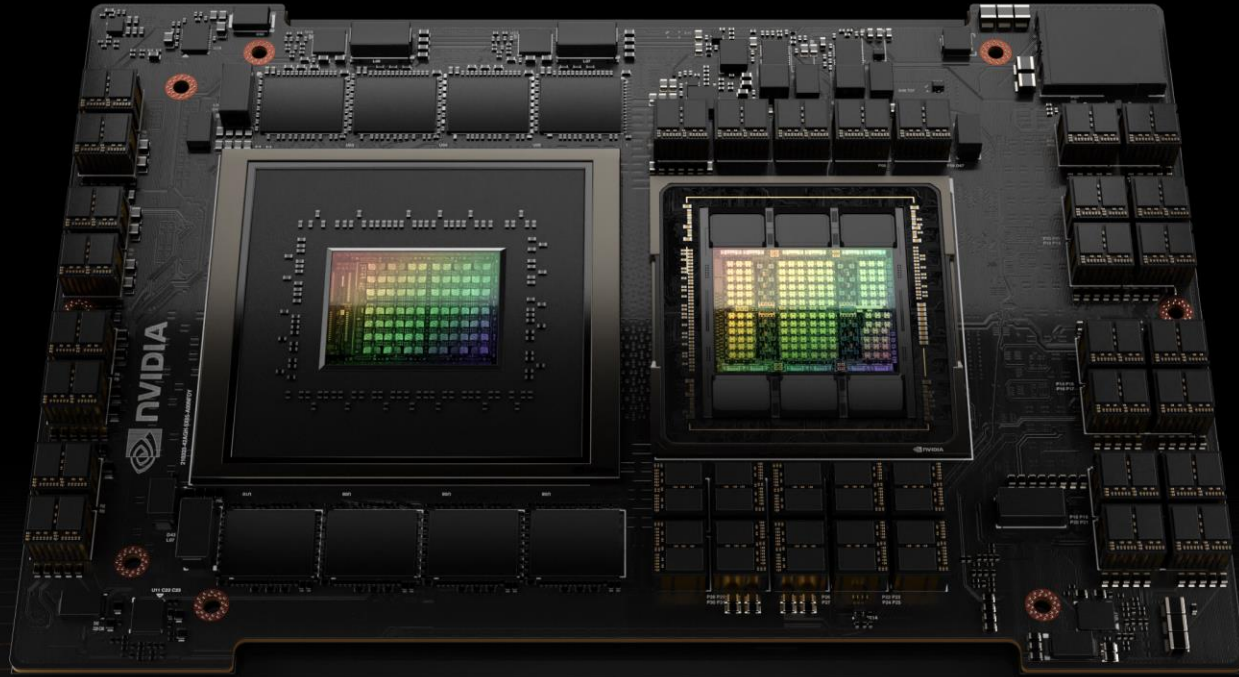
Deep Learning

4th-Gen Tensor Cores
1,400 Tensor TFLOPs
Optical Flow Accelerator



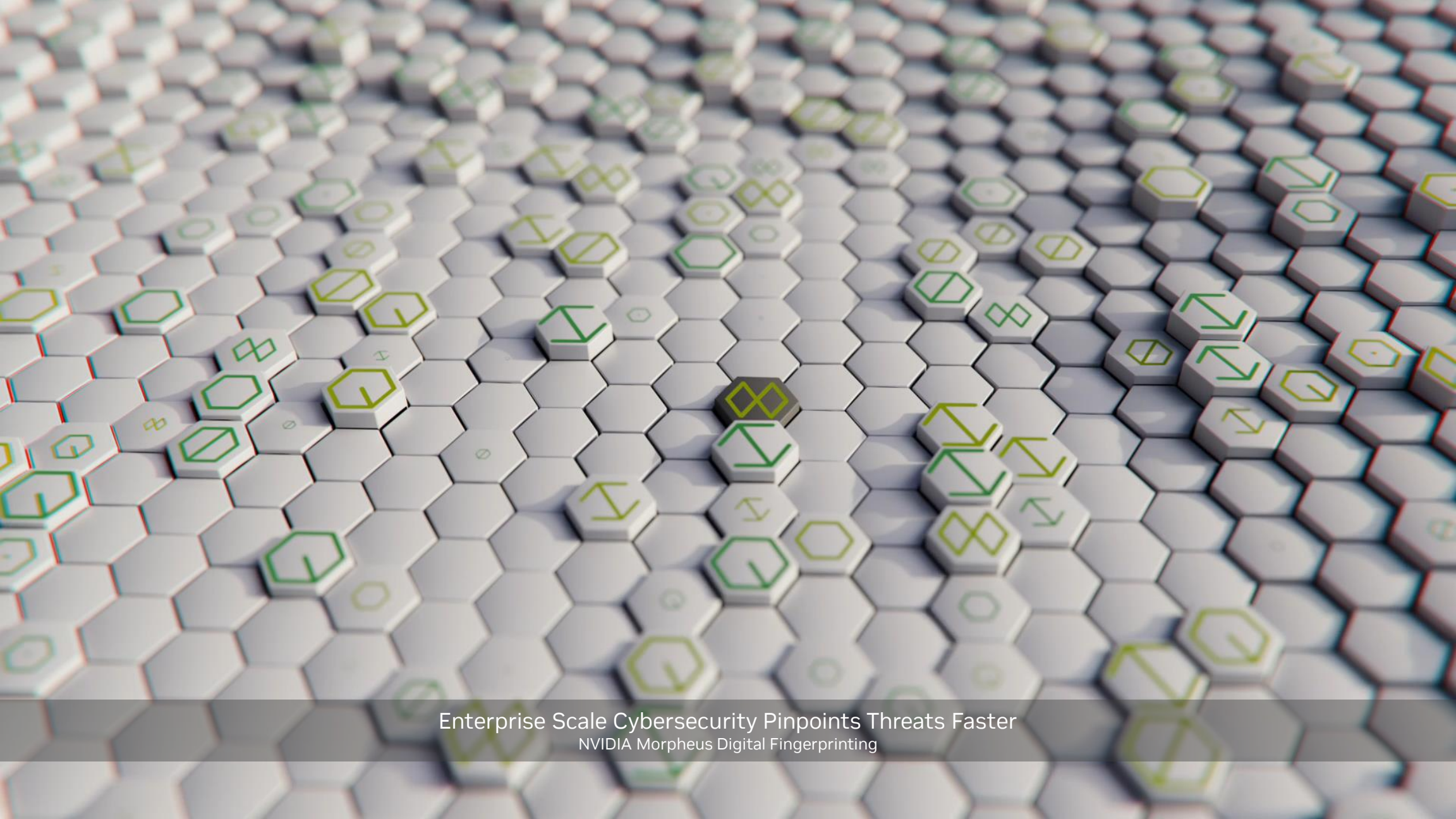
NVIDIA H100

80 Billion Transistors
TSMC 4N Process
4.9 TB/s Bandwidth



ANNOUNCING
NVIDIA GRACE HOPPER

Grace Hopper Superchip
Densest NVIDIA Accelerated Computing System
New NVLink Chip-to-Chip Coherent Inference
900 GB/s



Enterprise Scale Cybersecurity Pinpoints Threats Faster
NVIDIA Morpheus Digital Fingerprinting

ANNOUNCING NVIDIA MORPHEUS

Accelerated AI Platform for Next Gen SIEM

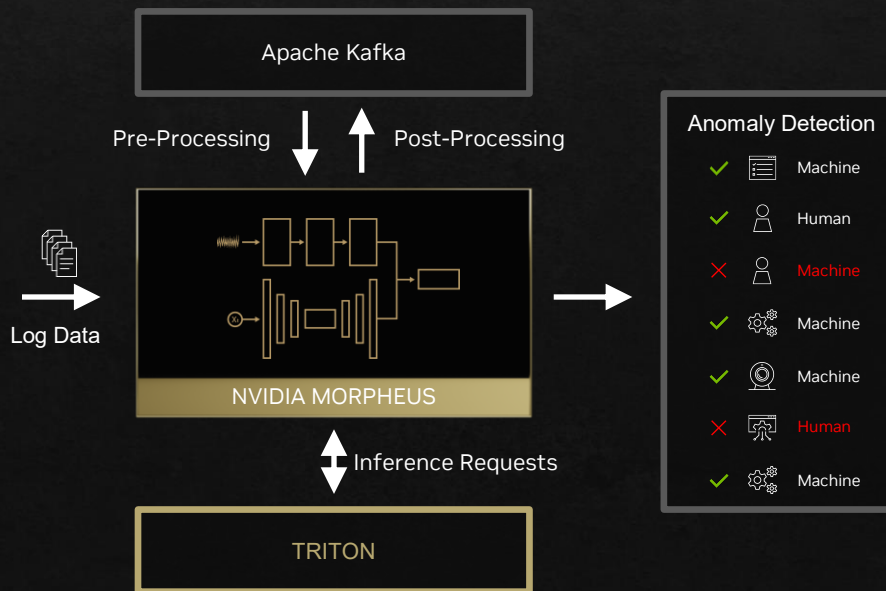
Built on NVIDIA RAPIDS and NVIDIA AI

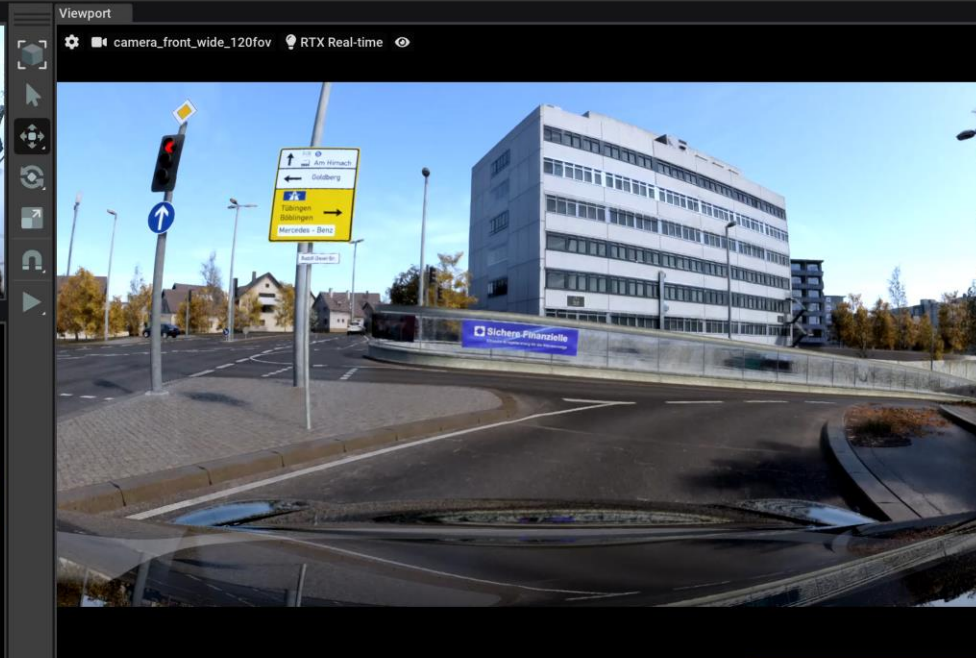
600X Faster Data Processing – Monitor Every User and Machine-Generated Data for Anomalous Behavior

Detect Anomalies with 10s of Millions of AI Models in Real-Time

Pre-Trained Models for User Activity Fingerprinting and Phishing Detection

Early Access 2 Available Now
nvidia.com/morpheus



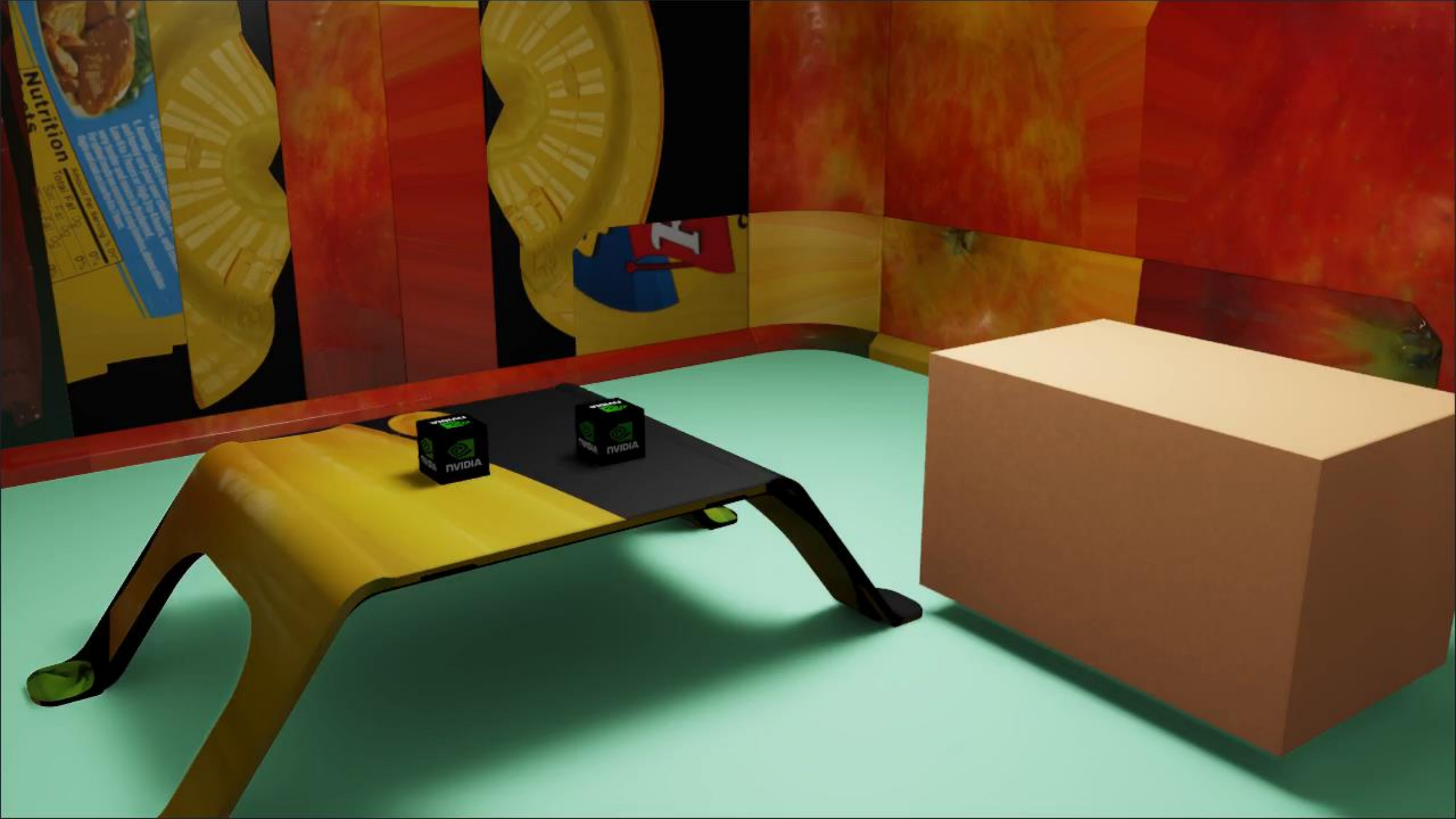


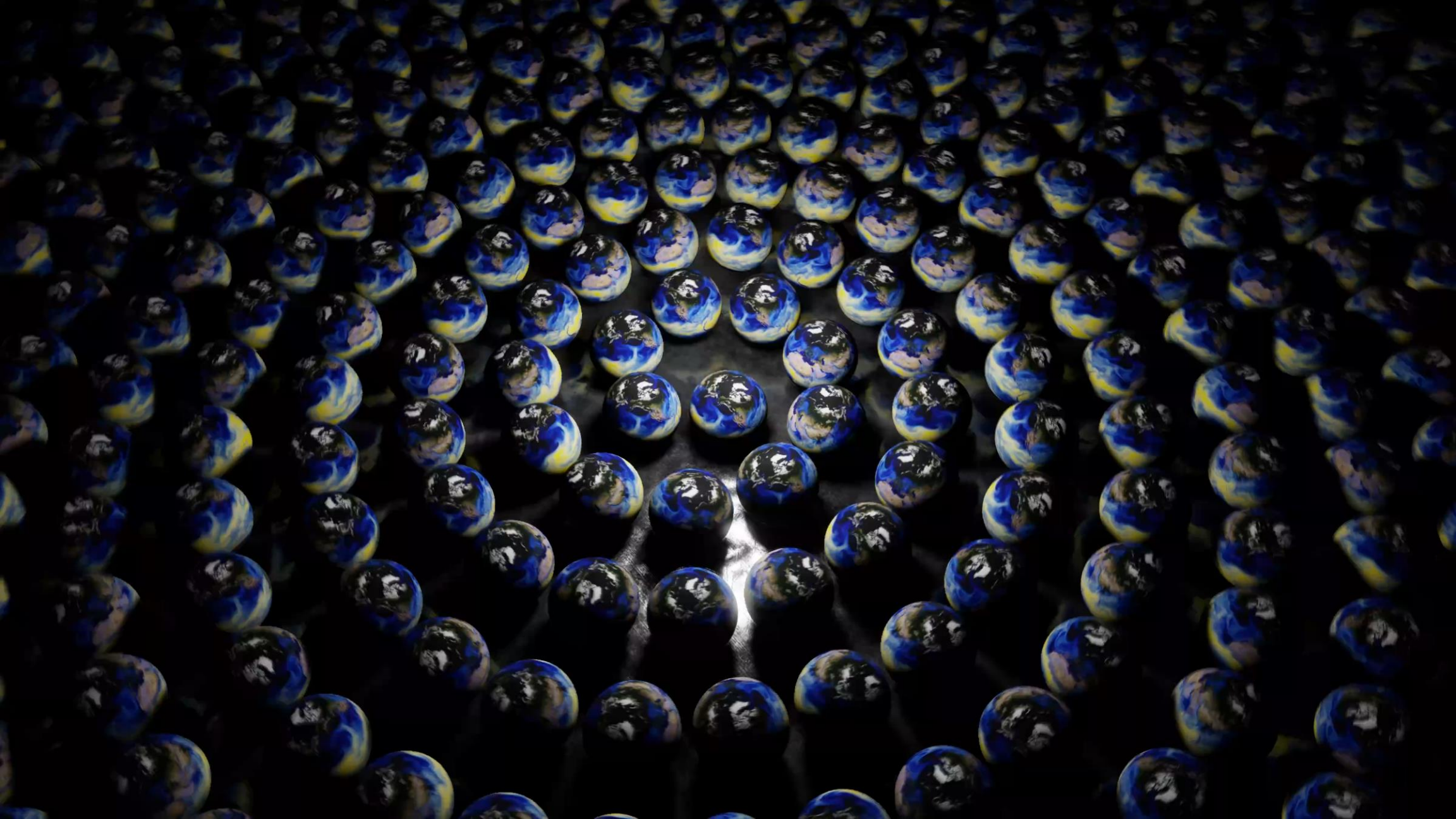
A DRIVE AROUND THE NEIGHBOURHOOD

THE METAVERSE

A Blending of Physical and Virtual Worlds







RTX - Real-Time Perspective

2017 August 24, 12:00:00 +0000




Stage Layer Render Settings Extensions

+ modulus All

NVIDIA COMMUNITY

v22.10.0 USER

MODULUS FOURCASTNET SCENARI...

 Simulation
modulus_scenario_fcn
v22.10.0 USER

MODULUS EPGA SCENARIO

Property Modulus

▼ Modulus FourCastNet

Weather Dataset HARVEY

Forecast Length 12

Ensemble

Ensemble Size 10

Ensemble Noise 0.02

FourCastNet Inference

▶ Dataset Information

▼ Visualization Parameters

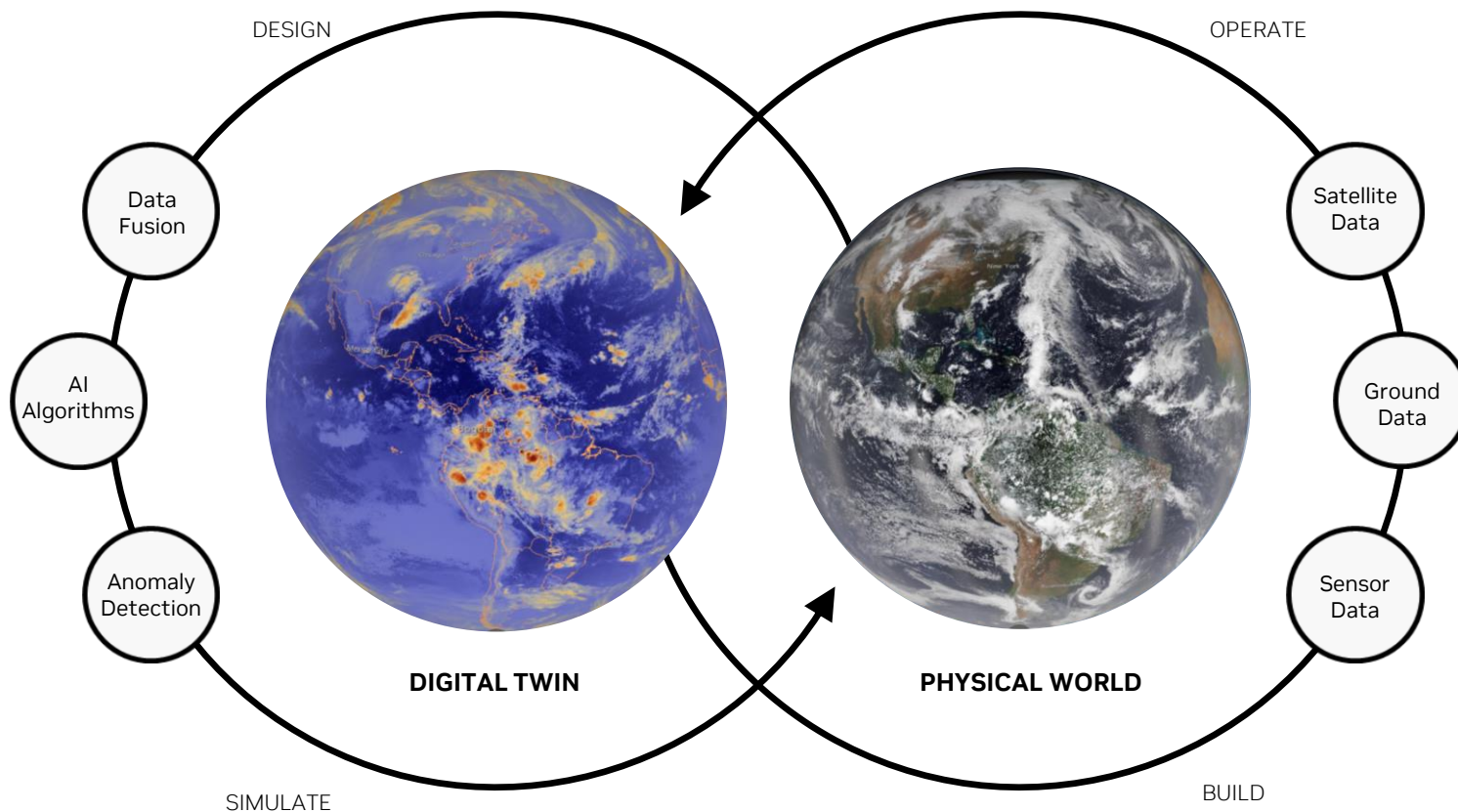
Content NVIDIA Assets Asset Stores (beta) Samples Environments Materials Console

0 4 8 12 16 20 24 28 32 36 40 44 48 52 56 60 64 68 72 76 80 84 88 92 96 100

0 100 100 24.00 FPS Auto

Lockheed Martin and NVIDIA Selected by NOAA for Earth Observation

Building a Climate Research Data Pipeline with NVIDIA Omniverse



2016

2017

2018

2019

2020

2021

2022

“NVIDIA is proud to have been one of ‘the Mavericks’ at the inception of the Frontier Development Laboratory. Our headquarters are named Endeavor and Voyager – mighty ships for our journey to the stars. We are delighted to support FDL as we push the frontiers for the human race.”

JEN-HSUN HUANG
PRESIDENT & CEO, NVIDIA CORPORATION



intel



kx

AIRBUS

Google Cloud

ELEMENT AI



MIT Portugal

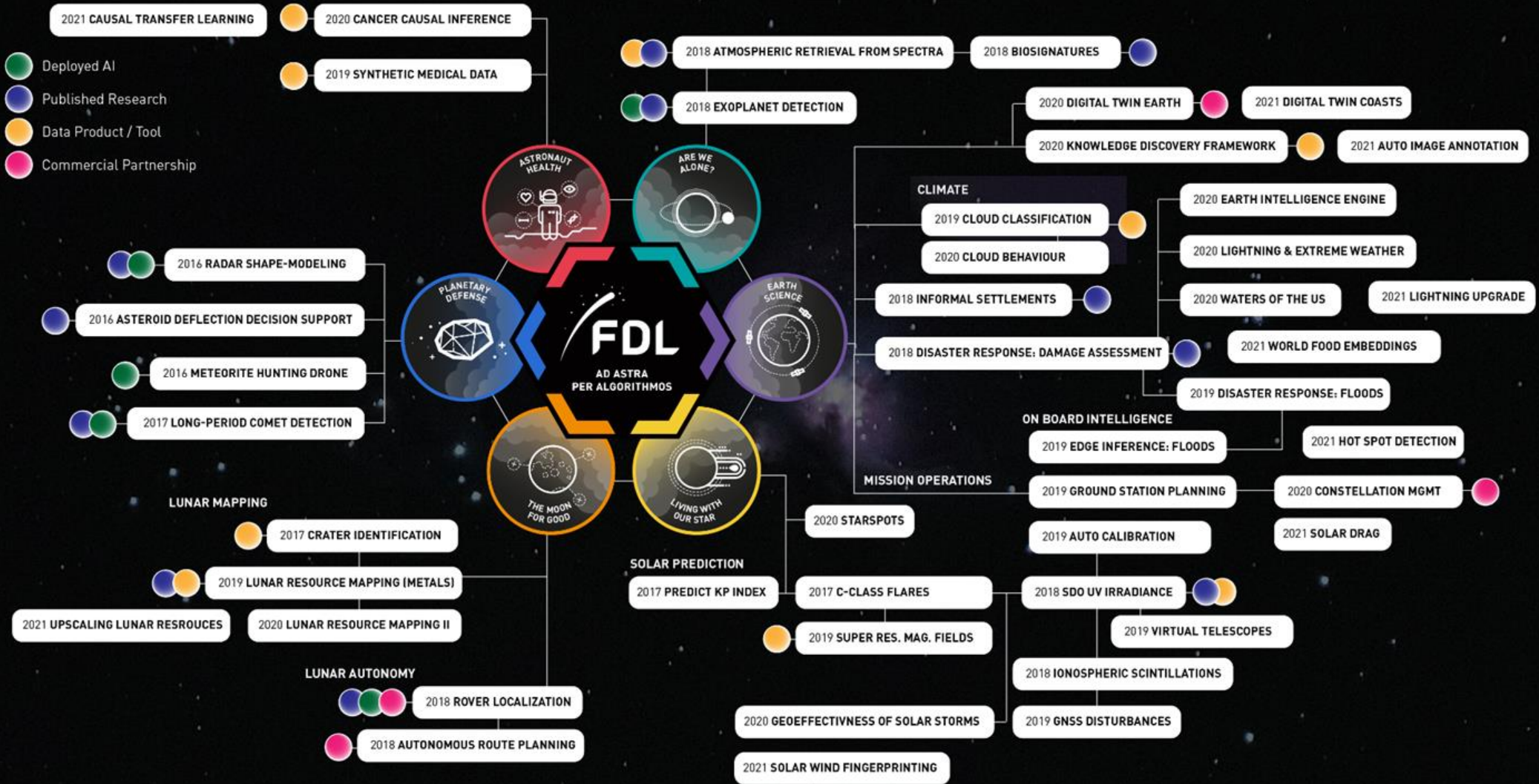


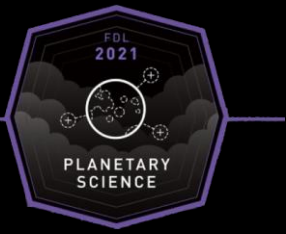
SCAN

Microsoft



AI PORTFOLIO





Lunar surface imagery enhancement



Fig. 1: NASA's Lunar Reconnaissance Orbiter



Fig. 2: Apollo XVI landing site

Next Level of AI GPGPU in Space Applications

Aitech's S-A1760 Venus™: most powerful and smallest space AI GPGPU in small form factor (SFF). Suitable for the next gen of short duration spaceflight, NEO and LEO.



Aitech S-A1760 Venus™ Radiation-characterized Space AI GPGPU

Based on NVIDIA® Jetson™ TX2i SoM



KEY PRODUCT INSIGHTS

NVIDIA Jetson TX2i SoM, 1 TFLOPS, 60 GFLOP/W	Linux OS pre-installed
Pascal™ architecture, GPU with 256 CUDA® cores	Dimensions: TBD Weight: TBD
2 Dual-core ARM® CPU + Cortex®-A57 Quad-core ARM CPU	11-36 VDC input voltage range, 8-10W under typical CUDA load (20W fully utilized)
8 GB LPDDR4, 128-bit interface, TX2i @ 1,600 MHz w/ECC	EMI/RFI per MIL-STD-461, vibration and shock per VITA 47
I/O: Gigabit Ethernet, DVI/HDMI out, CANbus, USB 2.0, discretes, UART Serial, composite in, SDI In	TID >1.5 krad (Si), watchdog mitigated, no more than one Type-2 SEFI per 14 days at ISS orbit

Note: Aitech has characterized the TX2i module in proton irradiation with final characterization to be done for updates at the box-level characterization later in 2021.



Get Technical Training at GTC

Enhance your skills with expert-led online workshops designed for developers of all types—from data scientists to students, researchers to instructors. These workshops are ideal for large enterprises to start-ups in AI, HPC, the metaverse, and beyond.

Here is a list of **recommended training for energy sector professionals:**

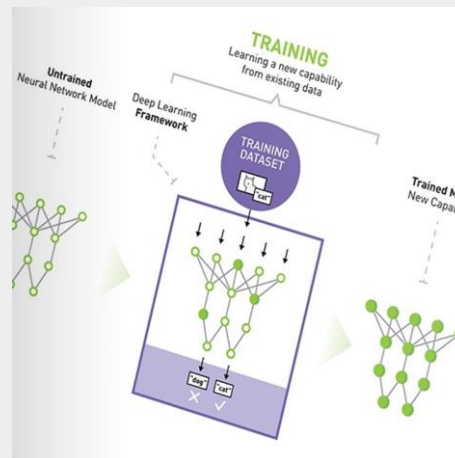
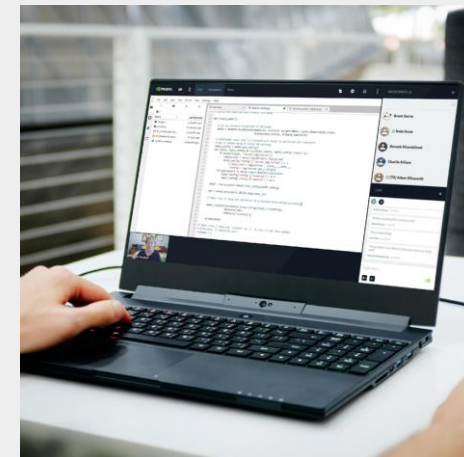
Labs: [Portable Acceleration of HPC Applications using ISO C++ — Part 1: Fundamentals*](#) | [Portable Acceleration of HPC Applications using ISO C++ — Part 2: Multi-GPU Applications*](#) | [Using Machine Learning for Anomaly Detection and Predictive Maintenance Scenarios*](#) | [Accelerated AI Logistics and Route Optimization 101*](#) | [Focused on Industrial Usability: Building Simulation Tools for Manufacturing Automation*](#)

Workshops: [Applications of AI for Anomaly Detection*](#) | [Applications of AI for Predictive Maintenance*](#) | [Building Transformer-Based Natural Language Processing Applications*](#) | [Computer Vision for Industrial Inspection*](#) | [Fundamentals of Accelerated Computing with CUDA Python*](#) | [Fundamentals of Accelerated Data Science*](#)

Register today for full-day workshops or free two-hour training labs.

GTC workshops start at just \$149 per seat (standard \$500). We also offer a discounted price of \$99 per seat for group purchases of 5+ seats. **Register your team now.**

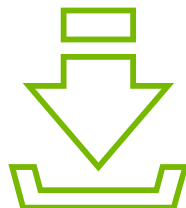
March 20-23 | www.nvidia.com/gtc/training



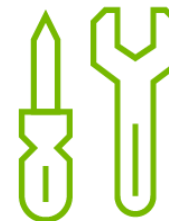
SEE YOU IN OMNIVERSE



EXPLORE OMNIVERSE ENTERPRISE



GET ACCESS TO A FREE TRIAL



DEVELOP ON OMNIVERSE



DOCUMENTATION
docs.omniverse.nvidia.com



**TUTORIALS AND
WEBINARS**
omniverse.nvidia.com/tutorials



FORUMS
omniverse.nvidia.com/forums



DISCORD
discord.gg/nvidiaomniverse

The Conference for the Era of AI and the Metaverse

March 20-23, 2023

[nvidia.com/gtc/keynote](https://www.nvidia.com/gtc/keynote)

