NVIDIA

GTC 2024

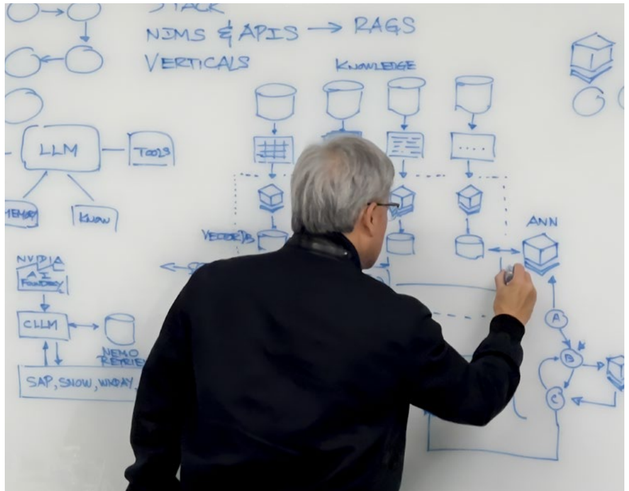
Highlights


NVIDIA
GTC
SAP center


Welcome to the GTC Experience


NVIDIA
GTC
SAN JOSE MCENERY CONVENTION CENTER

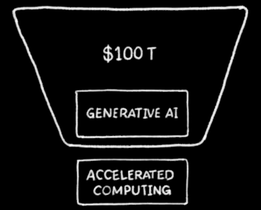

NVIDIA GTC
DELL Technologies


NIMS & APIS → RAGS
VERTICALS
KNOWLEDGE
LLM
TOOLS
ANN
SAP, SNOW, WORKDAY

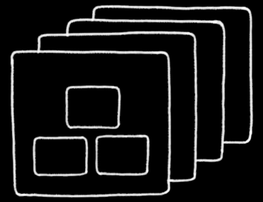

SAMSUNG
VOLVO
Expo Hall Theater
HEXAGON
RENTEX
MUST REMAIN UPRIGHT


“GTC is a front-row seat to what’s happening in AI.”
Bloomberg

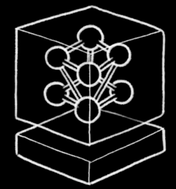
A New Industrial Revolution



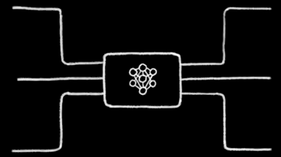
NEW INDUSTRY



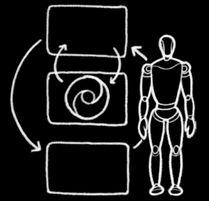
BLACKWELL PLATFORM



NIMs



NEMO AND NVIDIA AI FOUNDRY

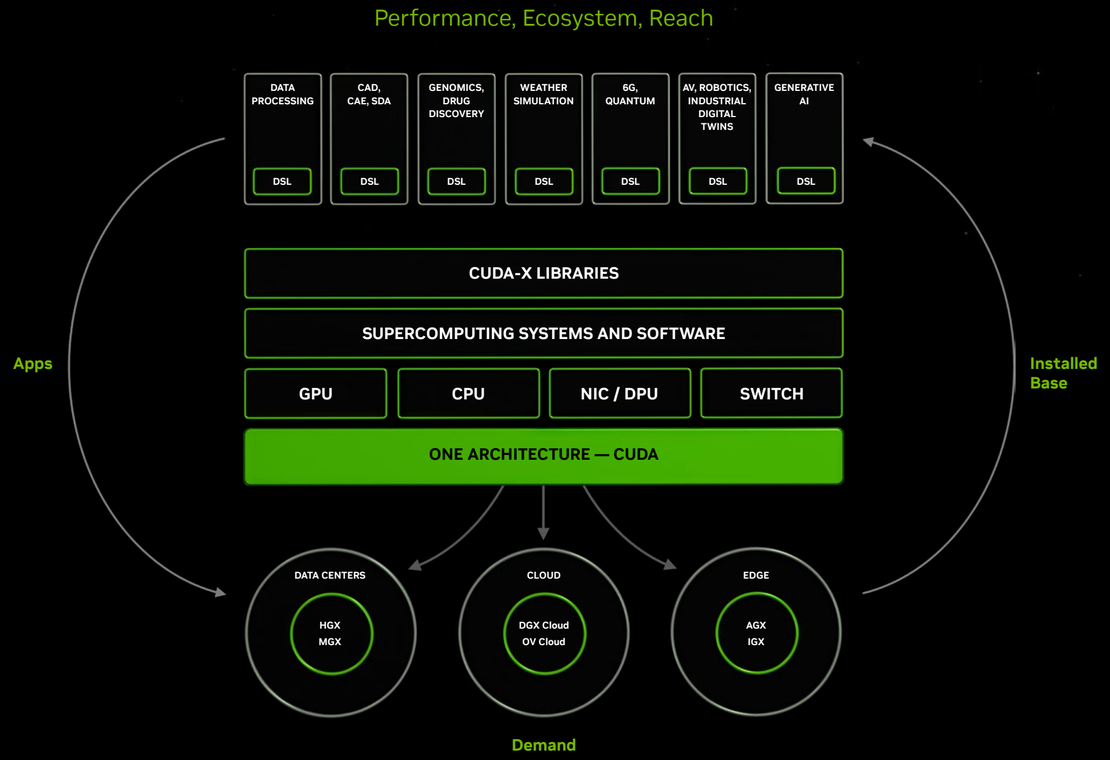


OMNIVERSE AND ISAAC ROBOTICS

“NVIDIA’s moat is its software and ecosystem.”

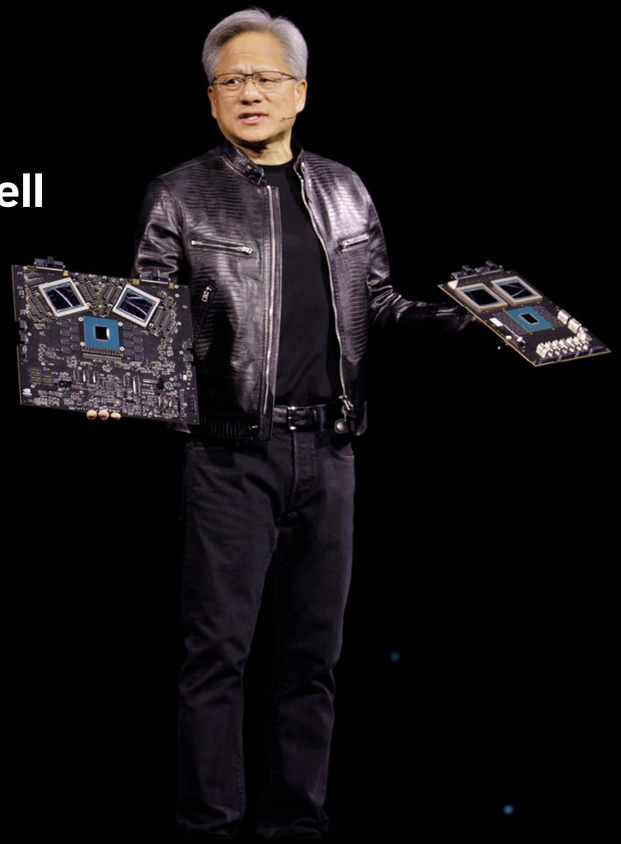
The Edge Singapore

The global NVIDIA ecosystem now approaches 5 million developers—capping a record-setting year for new developers. Forty thousand companies have worked with NVIDIA. There are now more than 3,300 GPU-accelerated applications. More than 1,600 generative AI companies are building on NVIDIA.



“NVIDIA’s new Blackwell chip is key to the next stage of AI.”

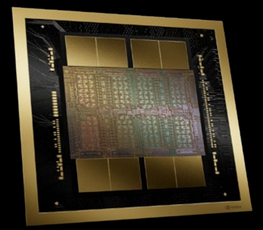
Bloomberg

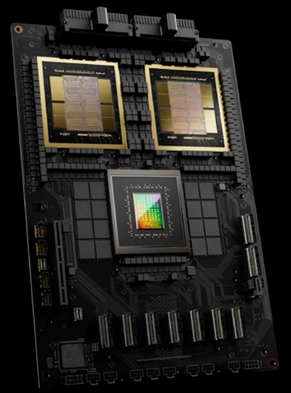


The NVIDIA Blackwell platform has arrived—enabling organizations everywhere to build and run real-time generative AI on trillion-parameter large language models at up to 25X less cost and energy consumption than its predecessor.

“Meet Blackwell, the new GPU for the AI era.”

Engineering.com





NVIDIA Blackwell Platform Arrives to Power a New Era of Computing

March 18, 2024

GTC—Powering a new era of computing, NVIDIA today announced that the NVIDIA Blackwell platform has arrived — enabling organizations everywhere to build and run real-time generative AI on trillion-parameter large language models at up to 25x less cost and energy consumption than its predecessor.

The Blackwell GPU architecture features six transformative technologies for accelerated computing, which will help unlock breakthroughs in data processing, engineering simulation, electronic design automation, computer-aided drug design, quantum computing and generative AI — all emerging industry opportunities for NVIDIA.

“For three decades we’ve pursued accelerated computing, with the goal of enabling transformative breakthroughs like deep learning and AI,” said **Jensen Huang, founder and CEO of NVIDIA.** “Generative AI is the defining technology of our time. Blackwell is the engine to power this new industrial revolution. Working with the most dynamic companies in the world, we will realize the promise of AI for every industry.”

Among the many organizations expected to adopt Blackwell are Amazon Web Services, Dell Technologies, Google, Meta, Microsoft, OpenAI, Oracle, Tesla and xAI.

Sundar Pichai, CEO of Alphabet and Google: “Scaling services like Search and Gmail to billions of users has taught us a lot about managing compute infrastructure. As we enter the AI platform shift, we continue to invest deeply in infrastructure for our own products and services, and for our Cloud customers. We are fortunate to have a longstanding partnership with NVIDIA, and look forward to bringing the breakthrough capabilities of the Blackwell GPU to our Cloud customers and teams across Google, including Google DeepMind, to accelerate future discoveries.”

Andy Jassy, president and CEO of Amazon: “Our deep collaboration with NVIDIA goes back more than 13 years, when we launched the world’s first GPU cloud instance on AWS. Today we offer the widest range of GPU solutions available anywhere in the cloud, supporting the world’s most technologically advanced accelerated workloads. It’s why the new NVIDIA Blackwell GPU will run so well on AWS and the reason that NVIDIA chose AWS to co-develop Project Ceiba, combining NVIDIA’s next-generation Grace Blackwell Superchips with the AWS Nitro System’s advanced virtualization and ultra-fast Elastic Fabric Adapter networking, for NVIDIA’s own AI research and development. Through this joint effort between AWS and NVIDIA engineers, we’re continuing to innovate together to make AWS the best place for anyone to run NVIDIA GPUs in the cloud.”

Michael Dell, founder and CEO of Dell Technologies: “Generative AI is critical to creating smarter, more reliable and efficient systems. Dell Technologies and NVIDIA are working together to shape the future of technology. With the launch of Blackwell, we will continue to deliver the next-generation of accelerated products and services to our customers, providing them with the tools they need to drive innovation across industries.”

Demis Hassabis, cofounder and CEO of Google DeepMind: “The transformative potential of AI is incredible, and it will help us solve some of the world’s most important scientific problems. Blackwell’s breakthrough technological capabilities will provide the critical compute needed to help the world’s brightest minds chart new scientific discoveries.”

Mark Zuckerberg, founder and CEO of Meta: “AI already powers everything from our large language models to our content recommendations, ads, and safety systems, and it’s only going to get more important in the future. We’re looking forward to using NVIDIA’s Blackwell to help train our open-source Llama models and build the next generation of Meta AI and consumer products.”

Satya Nadella, executive chairman and CEO of Microsoft: “We are committed to offering our customers the most advanced infrastructure to power their AI workloads. By bringing the GB200 Grace Blackwell processor to our datacenters globally, we are building on our long-standing history of optimizing NVIDIA GPUs for our cloud, as we make the promise of AI real for organizations everywhere.”

Sam Altman, CEO of OpenAI: “Blackwell offers massive performance leaps, and will accelerate our ability to deliver leading-edge models. We’re excited to continue working with NVIDIA to enhance AI compute.”

Larry Ellison, chairman and CTO of Oracle: “Oracle’s close collaboration with NVIDIA will enable qualitative and quantitative breakthroughs in AI, machine learning and data analytics. In order for customers to uncover more actionable insights, an even more powerful engine like Blackwell is needed, which is purpose-built for accelerated computing and generative AI.”

Elon Musk, CEO of Tesla and xAI: “There is currently nothing better than NVIDIA hardware for AI.”

Named in honor of David Harold Blackwell — a mathematician who specialized in game theory and statistics, and the first Black scholar inducted into the National Academy of Sciences — the new architecture succeeds the NVIDIA Hopper™ architecture, launched two years ago.

s to Fuel Accelerated Computing and Generative AI

ary technologies, which together enable AI training and real-time LLM inference for models scaling up to 10 e:

ful Chip — Packed with 208 billion transistors, Blackwell-architecture GPUs are manufactured using a MC process with two-reticle limit GPU dies connected by 10 TB/second chip-to-chip link into a single, unified

Transformer Engine — Fueled by new micro-tensor scaling support and NVIDIA’s advanced dynamic range hms integrated into NVIDIA TensorRT™-LLM and NeMo Megatron frameworks, Blackwell will support double del sizes with new 4-bit floating point AI inference capabilities.

/Link — To accelerate performance for multitrillion-parameter and mixture-of-experts AI models, the latest VLink® delivers groundbreaking 1.8TB/s bidirectional throughput per GPU, ensuring seamless high-speed ng up to 576 GPUs for the most complex LLMs.

ell-powered GPUs include a dedicated engine for reliability, availability and serviceability. Additionally, the adds capabilities at the chip level to utilize AI-based preventative maintenance to run diagnostics and forecast maximizes system uptime and improves resiliency for massive-scale AI deployments to run uninterrupted for s at a time and to reduce operating costs.

d confidential computing capabilities protect AI models and customer data without compromising performance, native interface encryption protocols, which are critical for privacy-sensitive industries like healthcare and

ine — A dedicated decompression engine supports the latest formats, accelerating database queries to deliver nce in-data analytics and data science. In the coming years, data processing, on which companies spend tens nnually, will be increasingly GPU-accelerated.

Blackwell Superchip connects two NVIDIA B200 Tensor Core GPUs to the NVIDIA Grace CPU over a 900GB/s hip-to-chip interconnect.

nance, GB200-powered systems can be connected with the NVIDIA Quantum-X800 InfiniBand and Spectrum ns, also announced today, which deliver advanced networking at speeds up to 800Gb/s.

onent of the NVIDIA GB200 NVL72, a multi-node, liquid-cooled, rack-scale system for the most loads. It combines 36 Grace Blackwell Superchips, which include 72 Blackwell GPUs and 36 Grace CPUs generation NVLink. Additionally, GB200 NVL72 includes NVIDIA BlueField®-3 data processing units to enable on, composable storage, zero-trust security and GPU compute elasticity in hyperscale AI clouds. The GB200 0x performance increase compared to the same number of NVIDIA H100 Tensor Core GPUs for LLM reduces cost and energy consumption by up to 25x.

gle GPU with 1.4 exaflops of AI performance and 30TB of fast memory, and is a building block for the newest

00, a server board that links eight B200 GPUs through NVLink to support x86-based generative AI platforms. orking speeds up to 400Gb/s through the NVIDIA Quantum-2 InfiniBand and Spectrum-X Ethernet networking

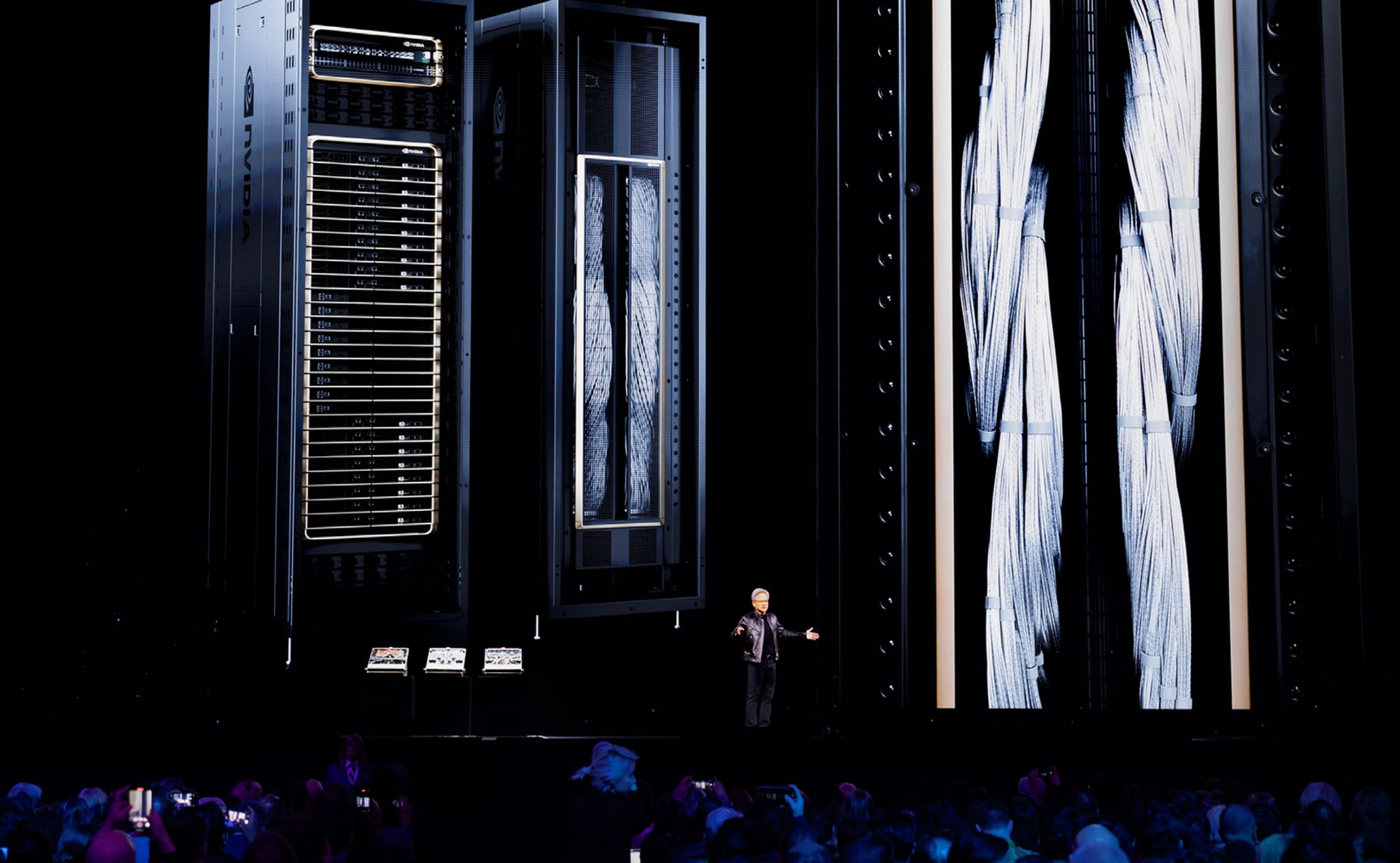
Blackwell Partners

will be available from partners starting later this year.

soft Azure and Oracle Cloud Infrastructure will be among the first cloud service providers to offer ces, as well as NVIDIA Cloud Partner program companies Applied Digital, CoreWeave, Crusoe, IBM Cloud and ds will also provide Blackwell-based cloud services and infrastructure, including Indosat Ooredoo Hutchinson, acle EU Sovereign Cloud, the Oracle US, UK, and Australian Government Clouds, Scaleway, Singtel, Northern Yotta Data Services’ Shakti Cloud and YTL Power International.

le on NVIDIA DGX™ Cloud, an AI platform co-engineered with leading cloud service providers that gives dicated access to the infrastructure and software needed to build and deploy advanced generative AI models. racle Cloud Infrastructure plan to host new NVIDIA Grace Blackwell-based instances later this year.

rd Enterprise, Lenovo and Supermicro are expected to deliver a wide range of servers based on Blackwell products, as are Aivres, ASRock Rack, ASUS, Eviden, Foxconn, GIGABYTE, Inventec, Pegatron, QCT, Wistron, Wiwynn and ZT Systems.



“NVIDIA moved the ball forward with the latest iteration of its speedy NVLink technology.”

HPCWire

To accelerate performance for multitrillion-parameter and mixture-of-experts AI models, the latest iteration of NVIDIA NVLink delivers groundbreaking 1.8TB/s bidirectional throughout per GPU. This ensures seamless high-speed communication among up to 576 GPUs for the most complex LLMs.

“The new AI platform could be a game changer.”

Barron’s

Generative AI is the defining technology of our time. Blackwell is the engine to power this new industrial revolution. Working with the most dynamic companies in the world, we’ll realize the promise of AI for every industry.

HGX B100

NVLink Switch

GB200 Superchip
Compute Node

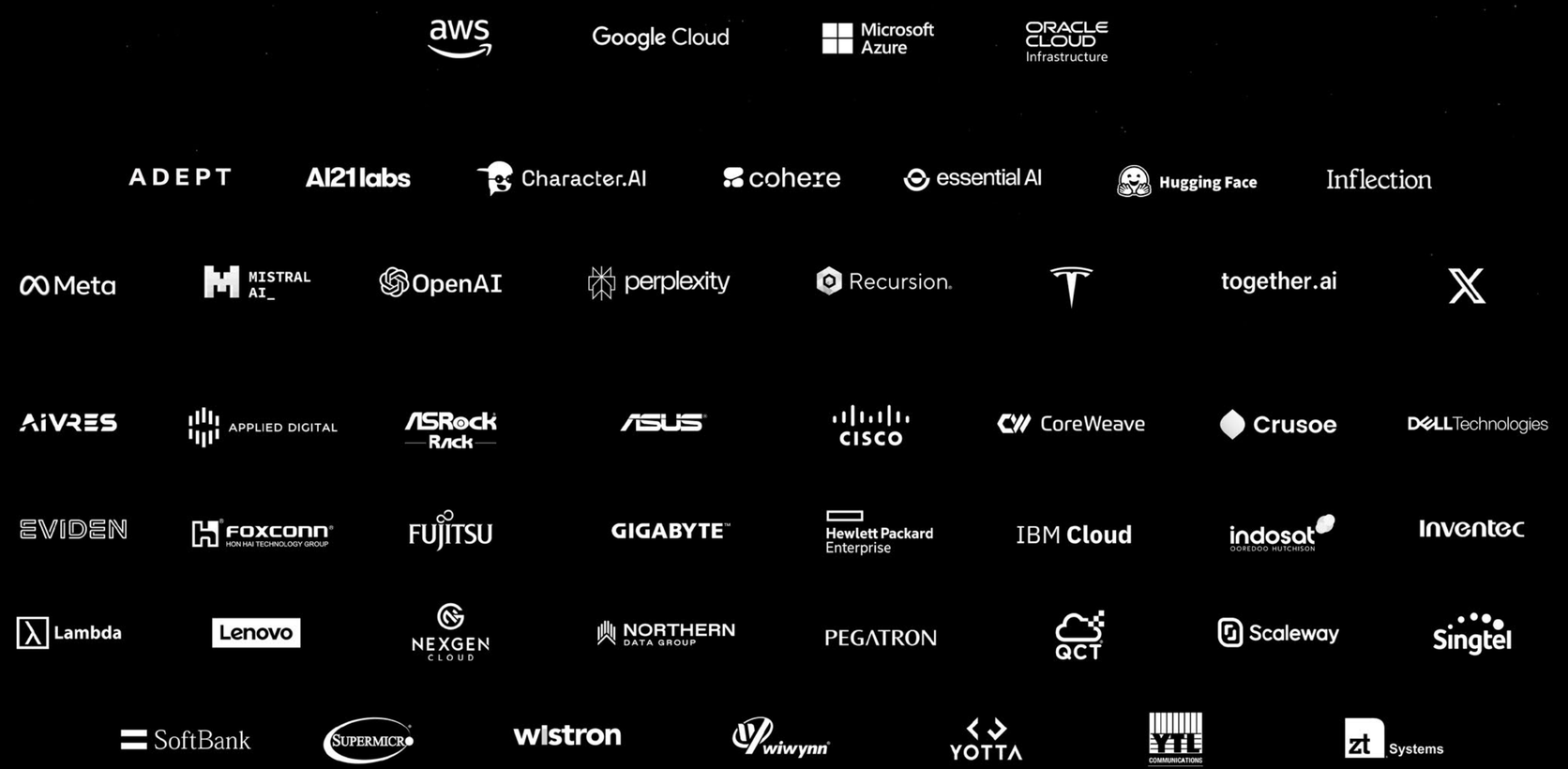
Quantum X800 Switch
ConnectX-8 SuperNIC

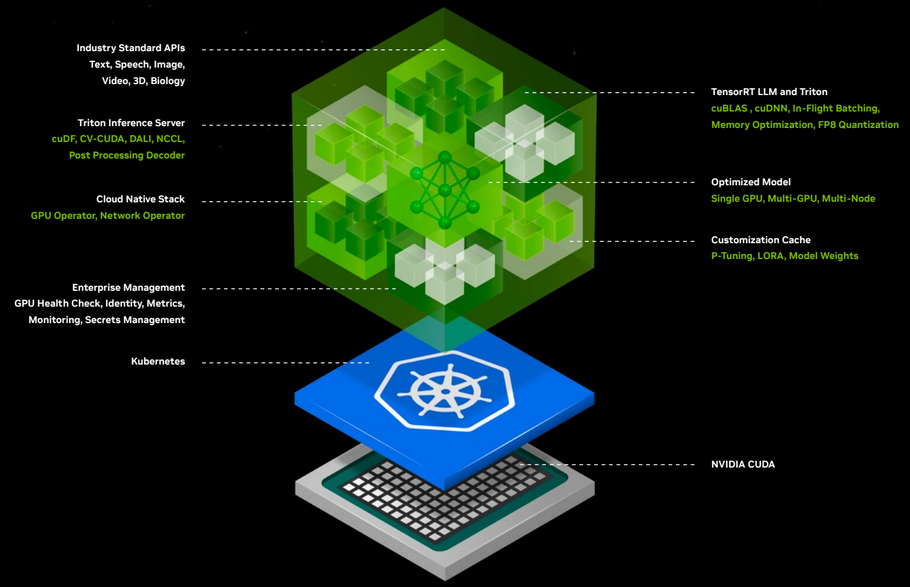
Spectrum X800 Switch
BlueField-3 SuperNIC

“Widespread adoption anticipated.”

VentureBeat

Among the many organizations expected to adopt Blackwell are Amazon Web Services, Dell Technologies, Google, Meta, Microsoft, OpenAI, Oracle, Tesla, and xAI.





100's of Millions of CUDA GPUs Installed Base

“NVIDIA launches NIM to make it smoother to deploy AI models into production.”

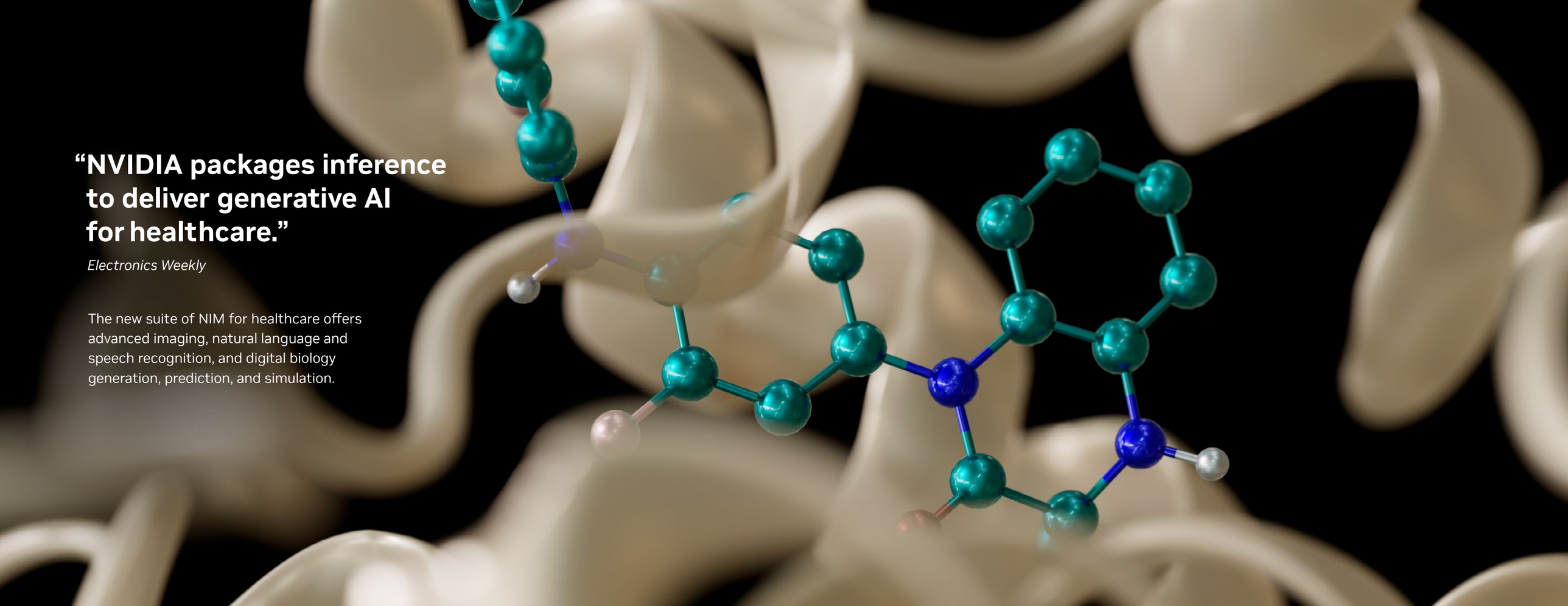
TechCrunch

NVIDIA Inference Microservices are a new way to package and deliver AI software. The curated selection of microservices adds a new layer to NVIDIA's full-stack computing platform. This layer connects the AI ecosystem of model developers, platform providers, and enterprises with a standardized path to run custom AI models.

“NVIDIA has virtually recreated the entire planet—and now it wants to use its digital twin to crack weather forecasting for good.”

TechRadar

To help combat the $140 billion in economic losses due to extreme weather brought on by climate change, we announced the Earth-2 digital twin cloud platform for simulating and visualizing weather and climate at unprecedented scale.



“NVIDIA packages inference to deliver generative AI for healthcare.”

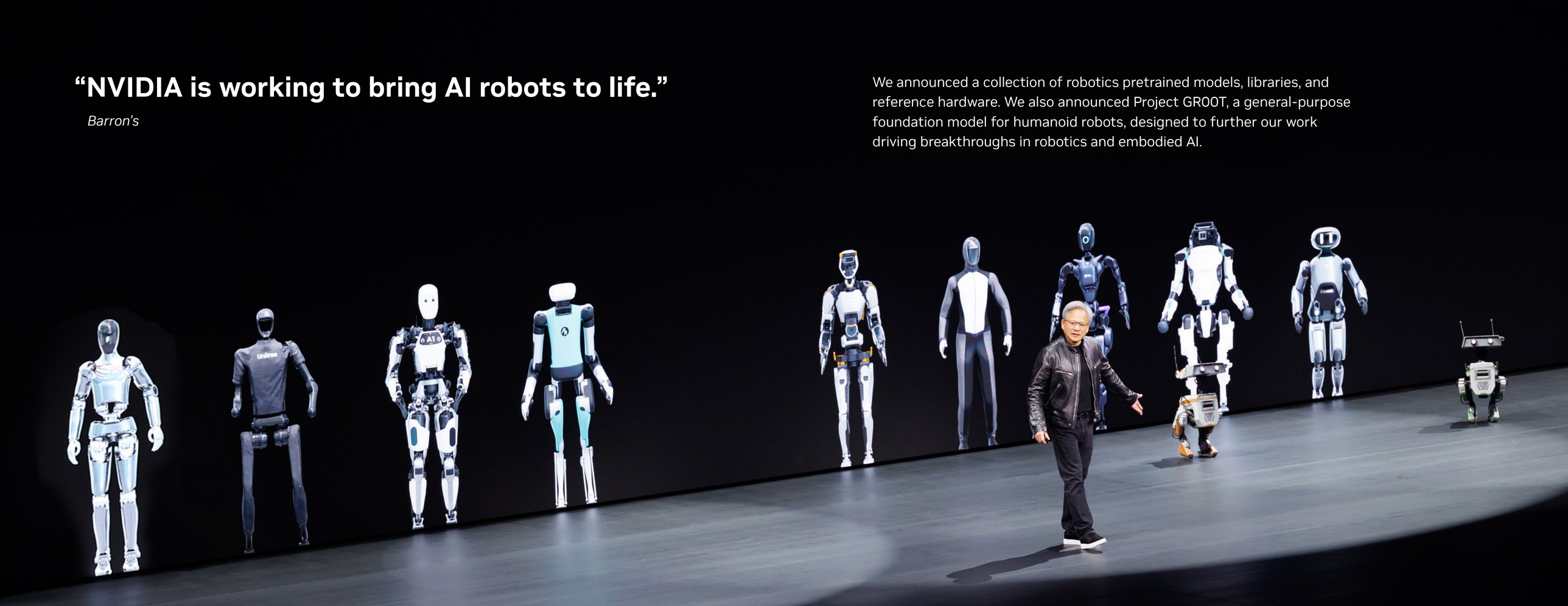
Electronics Weekly

The new suite of NIM for healthcare offers advanced imaging, natural language and speech recognition, and digital biology generation, prediction, and simulation.

“NVIDIA is working to bring AI robots to life.”

Barron’s

We announced a collection of robotics pretrained models, libraries, and reference hardware. We also announced Project GR00T, a general-purpose foundation model for humanoid robots, designed to further our work driving breakthroughs in robotics and embodied AI.

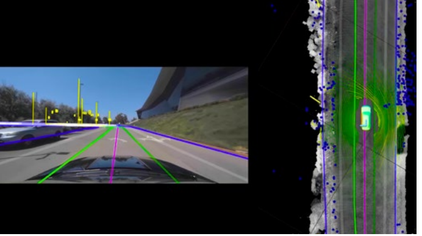
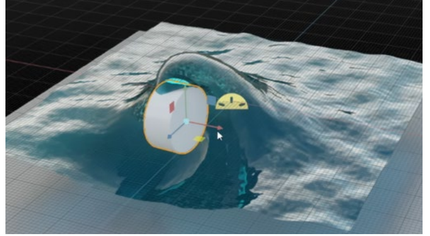


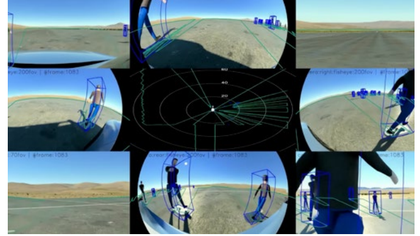
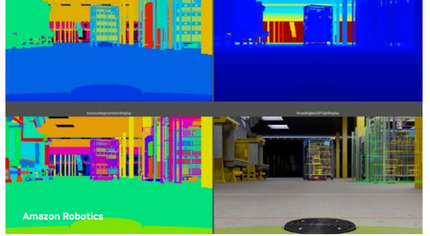
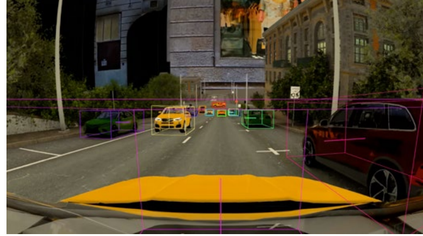
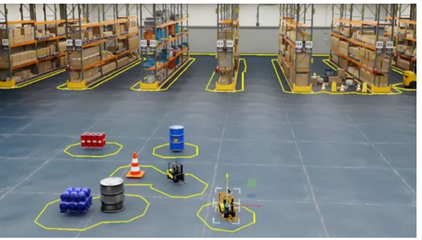
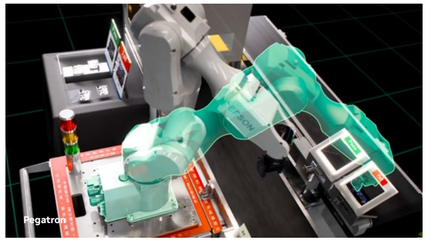
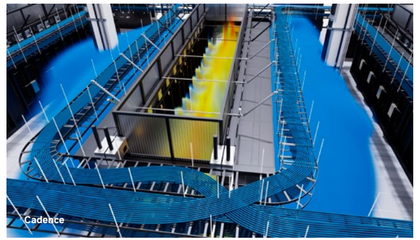
“NVIDIA Omniverse Cloud APIs will elevate digital twins for a new industrial revolution.”

VentureBeat

NVIDIA Omniverse Cloud will be available as APIs, extending the reach of the world’s leading platform for creating industrial digital twin applications and workflows across the entire ecosystem of software makers.

Siemens

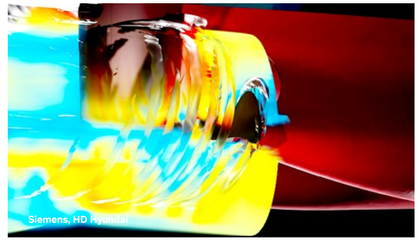



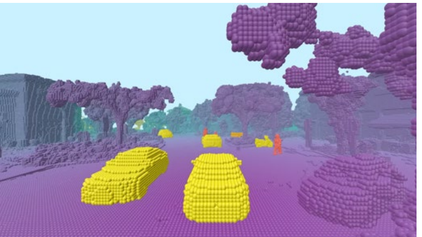


Amazon Robotics



Pegatron

Cadence

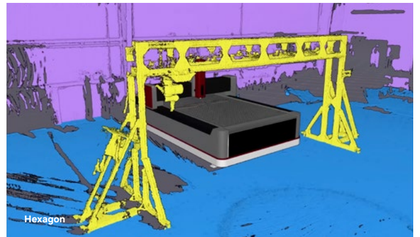
Ansys

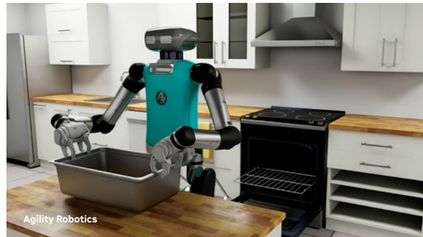

Amazon Robotics

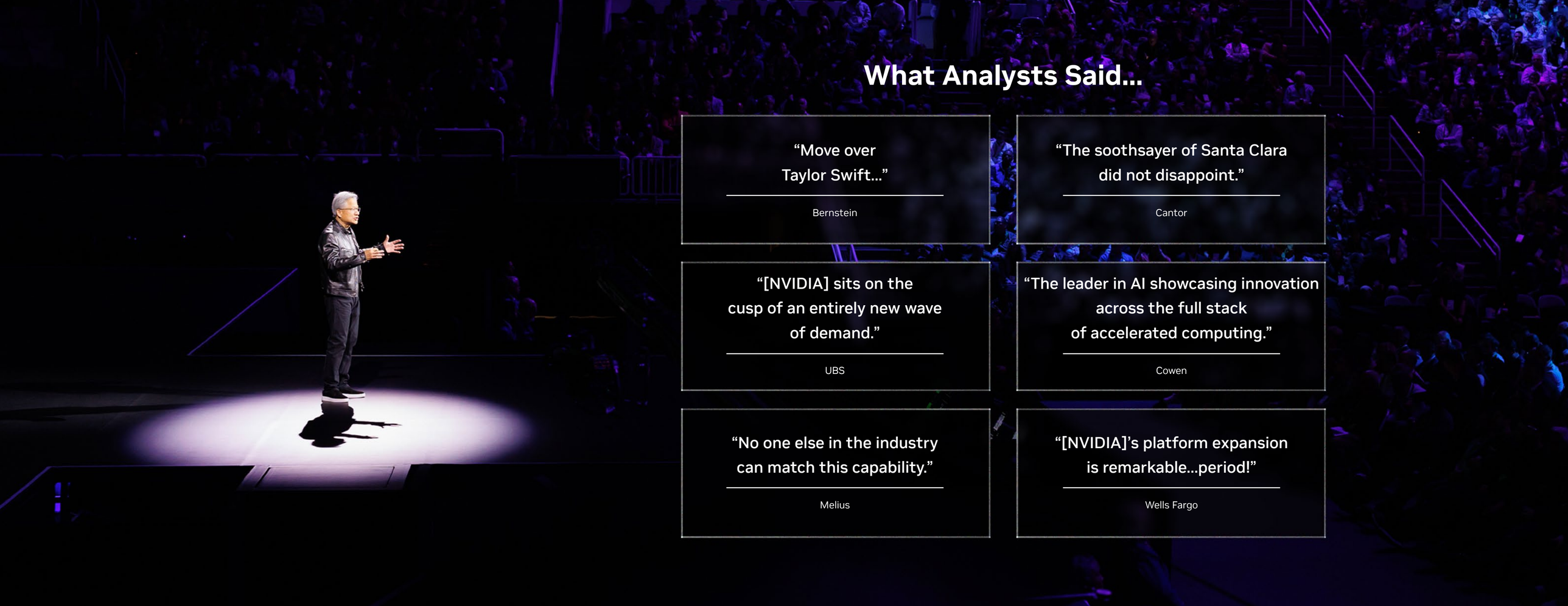


Techman Robot

Siemens, HD Hyundai


Hexagon



Agility Robotics



What Analysts Said...

"Move over Taylor Swift..."

Bernstein

"The soothsayer of Santa Clara did not disappoint."

Cantor

"[NVIDIA] sits on the cusp of an entirely new wave of demand."

UBS

"The leader in AI showcasing innovation across the full stack of accelerated computing."

Cowen

"No one else in the industry can match this capability."

Melius

"[NVIDIA]'s platform expansion is remarkable...period!"

Wells Fargo

What Press Said...

"NVIDIA's CEO painted a vision of AI turbocharging computing power."

The New York Times

"NVIDIA expects to win outsized chunk of data center spending."

Bloomberg

"Event was a daunting reminder of the speed at which NVIDIA is moving."

Financial Times

"With Blackwell GPUs, AI gets cheaper and easier, competing with NVIDIA gets harder."

The Next Platform

"The NVIDIA frenzy over artificial intelligence has come to this: AI Woodstock."

The Wall Street Journal

"GTC can serve as a preview of where the entire field is going."

Fast Company


NVIDIA
GTC
The Conference for the Era of AI
SAN JOSE McENERY CONVENTION CENTER
#GTC24


"GTC exists to inspire the world on the art-of-the-possible."
NVIDIA GTC
Explore and Share
@nvidia #gtc24

300,000
Registrations

19,000
In-Person Attendees

1,100
Sessions

33M
Keynote Views

29,000
Press Articles

300
Partner Sponsorships

“GTC: Spearheading AI and accelerated computing innovation.”

siliconANGLE

Transforming AI Panel: the authors of the seminal research paper—*Attention Is All You Need*—that introduced the transformer neural network architecture came together at GTC.


NVIDIA
GTC

“NVIDIA’s GTC conference, not surprisingly, was an absolute whirlwind.”

VentureBeat

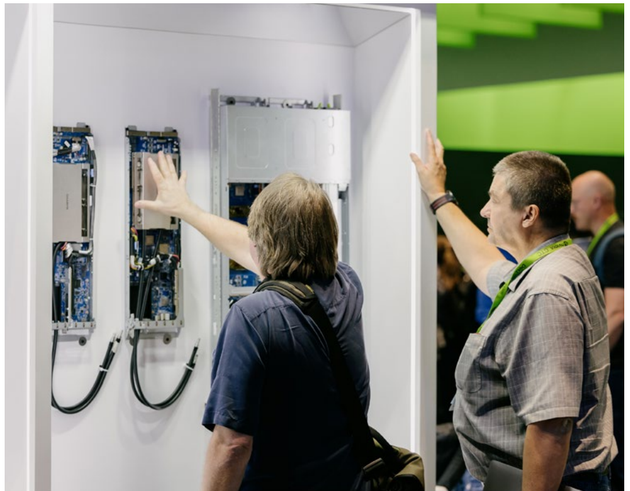
“GTC 2024 was the single most important event in the history of the technology industry.”

SiliconANGLE theCUBE


GIGABYTE
GIGABYTE
NVIDIA
GIGABYTE


NVIDIA
NVIDIA
NVIDIA
NVIDIA
NVIDIA


QCT




“NVIDIA’s conference captured the industry’s attention.”
Yahoo Finance


ASUS
IN SEARCH OF INCREDIBLE
Generative AI
Pioneering AI supercomputing
NVIDIA


SUPERMICRO
GB200 NVL72


VULTR
Lenovo


WeRide Robobus


VULTR

“The Generative AI Future Is Now.”

HPCWire

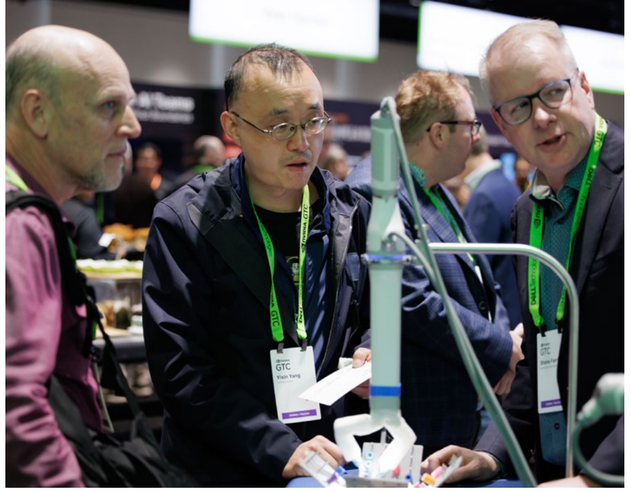

Inchained Tools


VOLVO
EX90
Designed to be the safest
Volvo car ever made


NVIDIA
GTC
BAR MENU
COCKTAILS
VIRGIN MOCKTAILS


GIGABYTE
VULTR
VULTR
VULTR
THANK YOU
Lenovo
ASUS
SAMSUNG
Expo Hall Theater
Innovation Bravery
backend


FOXCONN
GIGABYTE


GTC

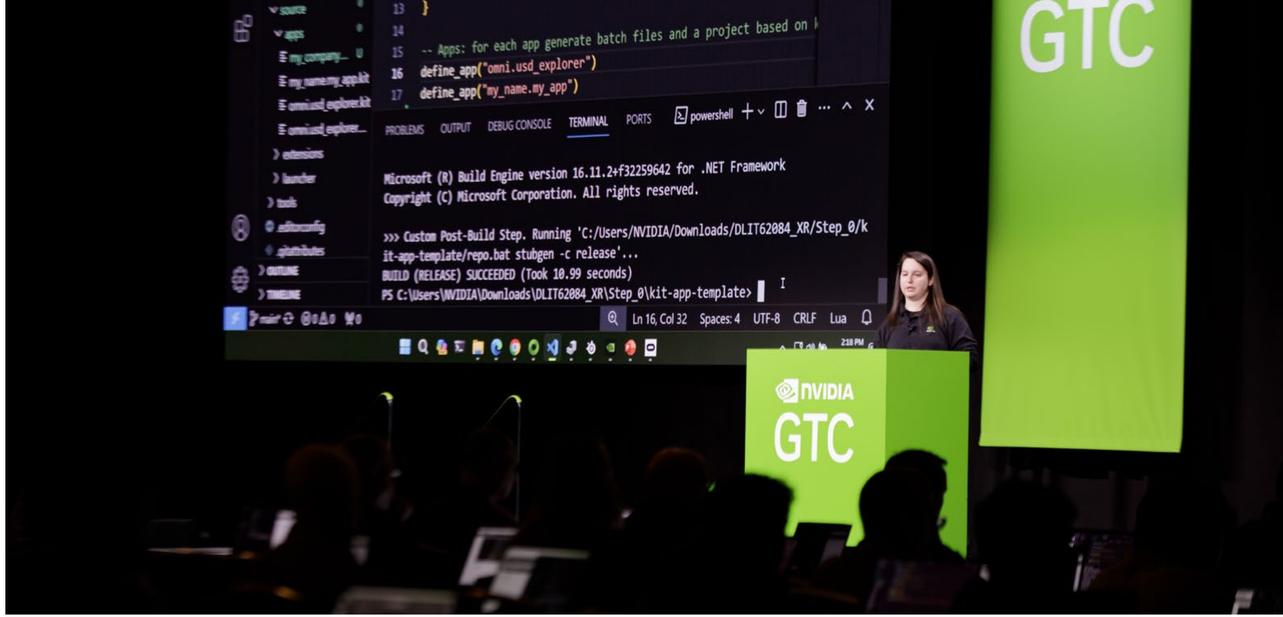

APPTRONIK

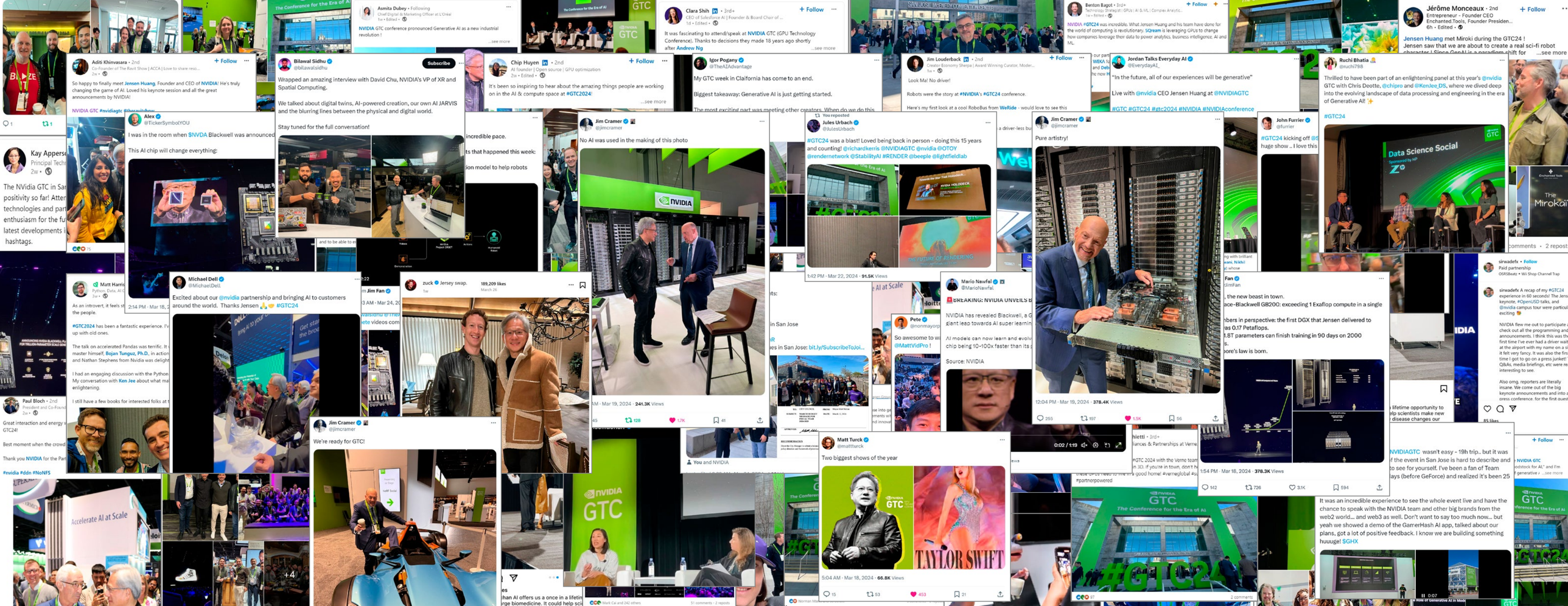

nuro
Nuro R3
AI-First Auto


NVIDIA
GTC
aws
aws
Proud Sponsor
DATA GROUP


Generative AI

Engagement with brilliant minds was evident, from ‘Connect with Experts’ sessions to watching industry leaders speak, and even casual encounters in the networking lounge.

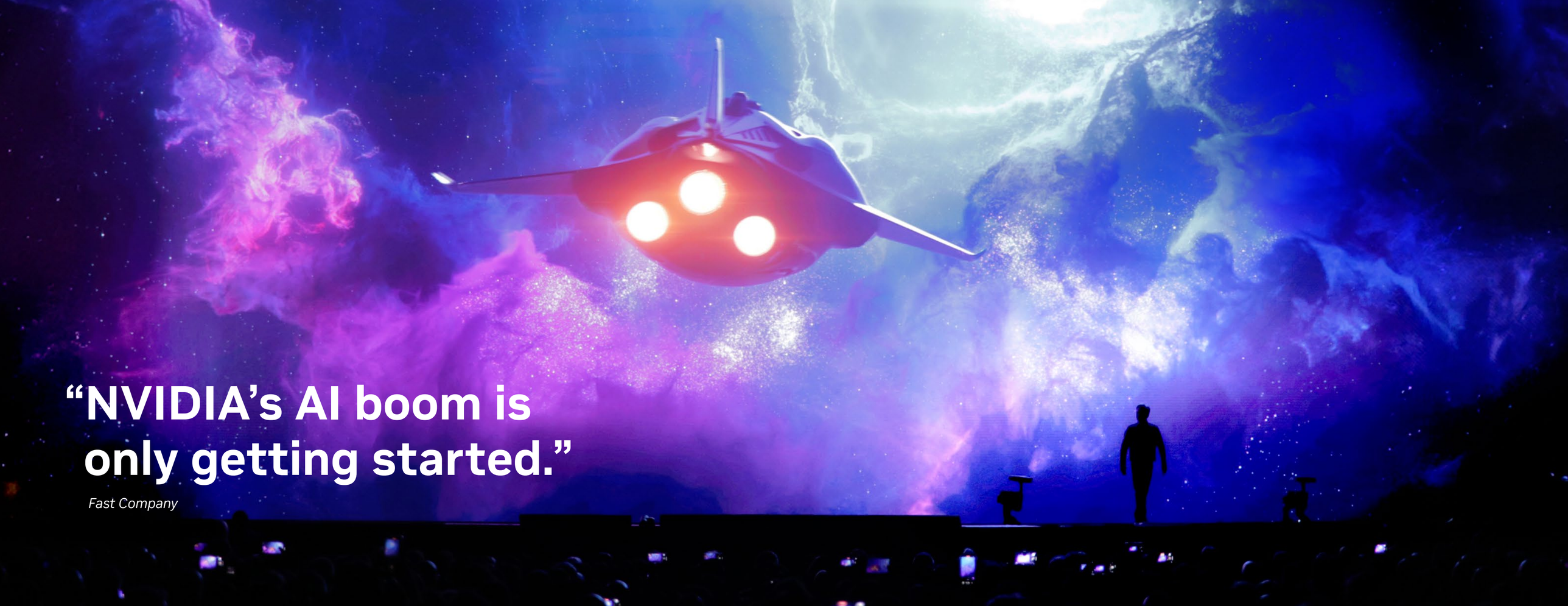



Asmita Dubey • Following
Chief Digital & Marketing Officer at L'Oréal
1w • Edited •
NVIDIA GTC conference pronounced Generative AI as a new industrial revolution !
...see more
Clara Shih • 3rd+
CEO of Salesforce AI | Founder & Board Chair of ...
1d • Edited •
Follow
It was fascinating to attend/speak at NVIDIA GTC (GPU Technology Conference). Thanks to decisions they made 18 years ago shortly after Andrew Ng
...see more
Benton Bagot • 3rd+
Technology Strategist | GPUs | AI & ML | Complex Analytic...
1w • Edited •
Follow
NVIDIA #GTC24 was incredible. What Jensen Huang and his team have done for the world of computing is revolutionary. SQream is leveraging GPUs to change how companies leverage their data to power analytics, business intelligence, AI and ML.
Jérôme Monceaux • 2nd
Entrepreneur - Founder CEO Enchanted.Tools, Founder Presiden...
6h • Edited •
Follow
Jensen Huang met Miroki during the GTC24 !
Jensen saw that we are about to create a real sci-fi robot character. I Since GenAI is a paradigm shift for
...see more
Aditi Khinvasara • 2nd
Co-Founder of The Ravit Show | ACCA | Love to share reso...
2w •
Follow
So happy to finally meet Jensen Huang, Founder and CEO of NVIDIA! He's truly changing the game of AI. Loved his keynote session and all the great announcements by NVIDIA!
NVIDIA GTC #nvidiagtc #huangsjacket
Bilawal Sidhu
@bilawalsidhu
Subscribe
Wrapped an amazing interview with David Chu, NVIDIA's VP of XR and Spatial Computing.
We talked about digital twins, AI-powered creation, our own AI JARVIS and the blurring lines between the physical and digital world.
Stay tuned for the full conversation!
Chip Huyen • 2nd
AI founder | Open source | GPU optimization
2w • Edited •
Follow
It's been so inspiring to hear about the amazing things people are working on in the AI & compute space at #GTC2024!
...see more
Igor Pogany
@TheAIAdvantage
My GTC week in California has come to an end.
Biggest takeaway: Generative AI is just getting started.
The most exciting part was meeting other creators. When do we do this
Jim Louderback • 2nd
Creator Economy Sherpa | Award Winning Curator, Moder...
1w •
Follow
Look Ma! No driver!
Robots were the story at #NVIDIA's #GTC24 conference.
Here's my first look at a cool RoboBus from WeRide - would love to see this
Jordan Talks Everyday AI
@EverydayAI_
"In the future, all of our experiences will be generative"
Live with @nvidia CEO Jensen Huang at @NVIDIAGTC
#GTC #GTC24 #gtc2024 #NVIDIA #NVIDIAconference
Ruchi Bhatia
@ruchi798
Thrilled to have been part of an enlightening panel at this year's @nvidia GTC with Chris Deotte, @chiprо and @KenJee_DS, where we dived deep into the evolving landscape of data processing and engineering in the era of Generative AI! ✨
#GTC24
Data Science Social
Powered by HP
comments • 2 repost
Kay Appers
Principal Tech
2w •
The NVidia GTC in San
positivity so far! Atter
technologies and part
enthusiasm for the fu
latest developments i
hashtags.
Alex
@TickerSymbolYOU
I was in the room when $NVDA Blackwell was announced
This AI chip will change everything:
incredible pace.
ts that happened this week:
ion model to help robots
Jim Cramer
@jimcramer
No AI was used in the making of this photo
NVIDIA
M · Mar 19, 2024 · 241.3K Views
128
1.7K
41
You reposted
Jules Urbach
@JulesUrbach
#GTC24 was a blast! Loved being back in person - doing this 15 years and counting! @richardkerris @NVIDIAGTC @nvidia @OTOY @rendernetwork @StabilityAI #RENDER @beeple @lightfieldlab
1:42 PM · Mar 22, 2024 · 91.5K Views
Jim Cramer
@jimcramer
Pure artistry!
12:04 PM · Mar 19, 2024 · 378.4K Views
255
197
1.5K
56
John Furrier
@furrier
#GTC24 kicking off @S
huge show .. I love this
Matt Harris
Python, Data, AI C
2w •
As an introvert, it feels s
the people.
#GTC2024 has been a fantastic experience. I'v
up with old ones.
The talk on accelerated Pandas was terrific. It
master himself, Bojan Tunguz, Ph.D., in action
and Nathan Stephens from Nvidia was delight
I had an engaging discussion with the Python
My conversation with Ken Jee about what ma
enlightening.
I still have a few books for interested folks at t
Michael Dell
@MichaelDell
Excited about our @nvidia partnership and bringing AI to customers around the world. Thanks Jensen 🙏 #GTC24
2:14 PM · Mar 18, 2
zuck
Jersey swap.
1w
189,209 likes
March 26
Jim Fan
3 AM · Mar 24, 20
videos com
Mario Nawfal
@MarioNawfal
BREAKING: NVIDIA UNVEILS B
NVIDIA has revealed Blackwell, a G
giant leap towards AI super learnin
AI models can now learn and evolv
chip being 10-100x faster than its
Source: NVIDIA
0:02 / 1:19
Pete
@nonmayorpe
So awesome to w
@MattVidPro !
in San Jose
es in San Jose: bit.ly/SubscribeToJoi...
You and NVIDIA
Jim Fan
@JimFan
, the new beast in town.
ace-Blackwell GB200: exceeding 1 Exaflop compute in a single
bers in perspective: the first DGX that Jensen delivered to
as 0.17 Petaflops.
.8T parameters can finish training in 90 days on 2000
s.
oore's law is born.
1:54 PM · Mar 18, 2024 · 378.3K Views
142
726
3.1K
594
lifetime opportunity to
elp scientists make new
disease changes our
85 likes
siwadefx • Follow
Paid partnership
OSRSBeatz • Wi Shop Channel Trap
siwadefx A recap of my #GTC24 experience in 60 seconds! The Jensen keynote, #OpenUSD talks, and @nvidia campus tour were particularly exciting
NVIDIA flew me out to participate and check out all the programming and announcements. I think this was the first time I've ever had a driver wait at the airport with my name on a sign. It felt very fancy. It was also the first time I got to go on a press junket! Q&As, media briefings, etc were all interesting to see.
Also omg, reporters are literally insane. We come out of the big keynote announcements and into a press conference for the first and
Paul Bloch • 2nd
President and Co-Found
2w •
Great interaction and energy
GTC24!
Best moment when the crowd
Thank you NVIDIA for the Part
#nvidia #ddn #NoNFS
Jim Cramer
@jimcramer
We're ready for GTC!
Matt Turck
@mattturck
Two biggest shows of the year
NVIDIA
GTC
TAYLOR SWIFT
5:04 AM · Mar 18, 2024 · 66.8K Views
15
53
453
21
Accelerate AI at Scale
+4
NVIDIA
GTC
Mark Cai and 242 others
51 comments • 2 reposts
#partnerpowered
NVIDIA
GTC
The Conference for the Era of AI
#GTC24
97
2 comments
NVIDIAGTC wasn't easy - 19h trip.. but it was
f the event in San Jose is hard to describe and
o see for yourself. I've been a fan of Team
ays (before GeForce) and realized it's been 25
It was an incredible experience to see the whole event live and have the chance to speak with the NVIDIA team and other big brands from the web2 world... and web3 as well. Don't want to say too much now... but yeah we showed a demo of the GamerHash AI app, talked about our plans, got a lot of positive feedback. I know we are building something huuuge! $GHX
0:07
Role of Generative AI in Knowl
Follow
#NVIDIA GTC
odstock for AL" and I'm
generative ♪
...see more

“For three decades we’ve pursued accelerated computing, with the goal of enabling transformative breakthroughs like deep learning and AI. Generative AI is the defining technology of our time.

Blackwell is the engine to power this new industrial revolution. Working with the most dynamic companies in the world, we will realize the promise of AI for every industry.”

Jensen Huang



“NVIDIA’s AI boom is only getting started.”

Fast Company