Sovereign AI: Europe’s Path to AI Factories




www.beyond.pl
www.cra.cz
NVIDIA
22 Praha
Roche
SIEMENS
Mercedes-Benz

AI Factories Build the Intelligence Economy

Turn power and data into intelligence at industrial scale



AI factories turn power and data into intelligence

Standardization makes deployment faster and more repeatable

Faster deployment brings intelligence online sooner

NVIDIA Enterprise Reference Architectures

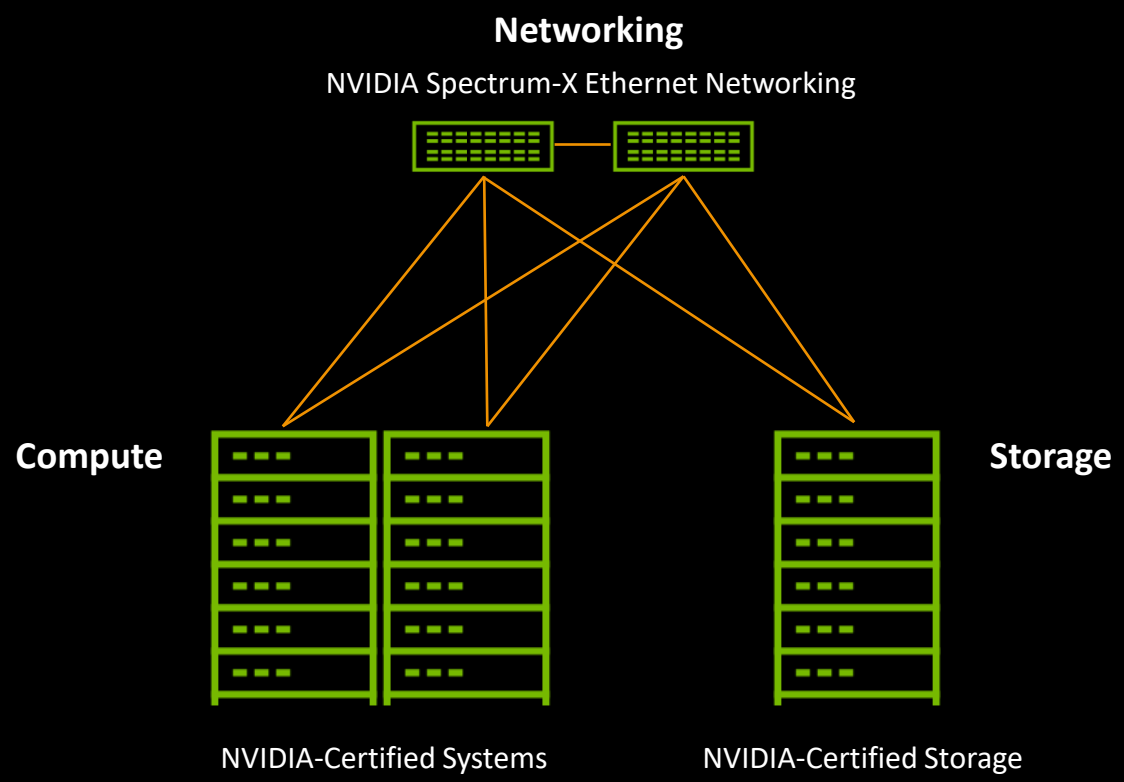
Comprehensive design recommendations for building high-performance, scalable data center infrastructure

NVIDIA-Certified Systems: Optimized Scale-Up & Scale-Out Configurations

High-Performance AI Networking: NVIDIA Spectrum-X Ethernet and Bluefield Networking Platforms

Multiple discrete design points: from 4 up to 128 nodes ranging up to 1024 GPUs

Companion Deployment Documentation: Spanning infrastructure Software, Networking Configuration, Observability, AI Workload Sizing and Deployment Guides



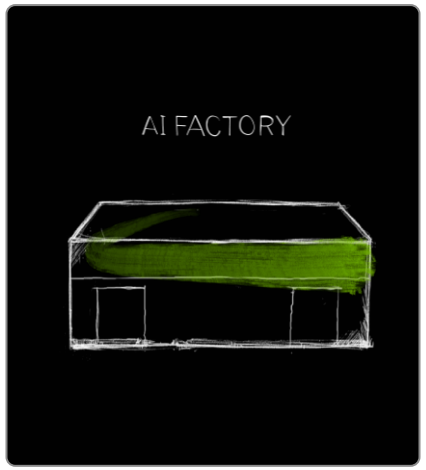
Time to value Scalability Manageability Security

A New Industrial Revolution - Tokens are the New Currency

AI Factories: Converting Data and Electricity into Profit-Generating Assets

AEC | Auto | Consumer Internet | Energy | FSI | Healthcare and Life Sciences | Higher Education and Research | HPC / Supercomputing | Manufacturing and Industrials | Media and Entertainment | Public Sector (US) | Retail and CPG | Robotics | Smart Cities and Spaces | Telco

$100T



Levers Impacting Infrastructure Economics

How tokens are valued, supplied, consumed, and monetized

Token Utility

How Valuable?

- Model complexity
- Context length
- Interactivity



Token Demand

How Variable?

- Base demand
- Workload multipliers
- Operational factors



Token Supply

How Many?

- Infrastructure cost
- Throughput per Watt (higher is better)
- Token Latency (lower is better)
- Token Cost



Token Monetization

How Much?

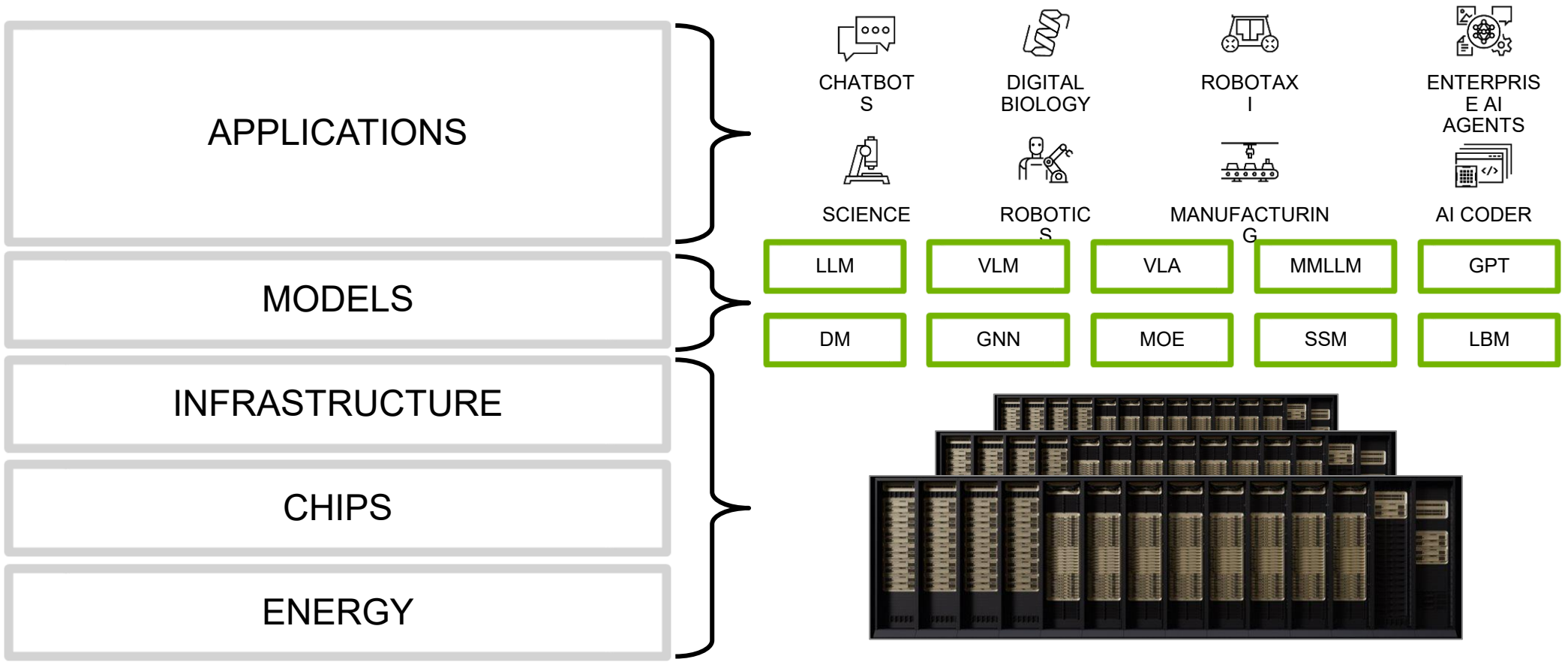
- AI business models
- Token pricing
- Business efficiency
- Business transformation



NVIDIA

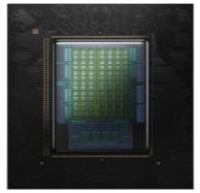
Token Efficiency Requires Vertical Optimization

AI as multi-layer infrastructure driving global job creation and economic growth

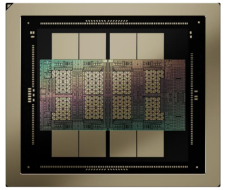


Six New Chips – One AI Supercomputer

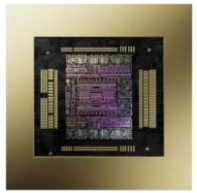
Vera Rubin Platform



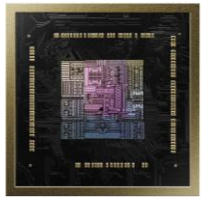
Vera



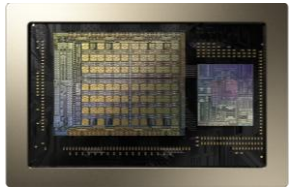
Rubin



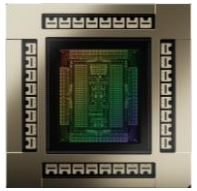
NVLink6 Switch



CX9



BF4

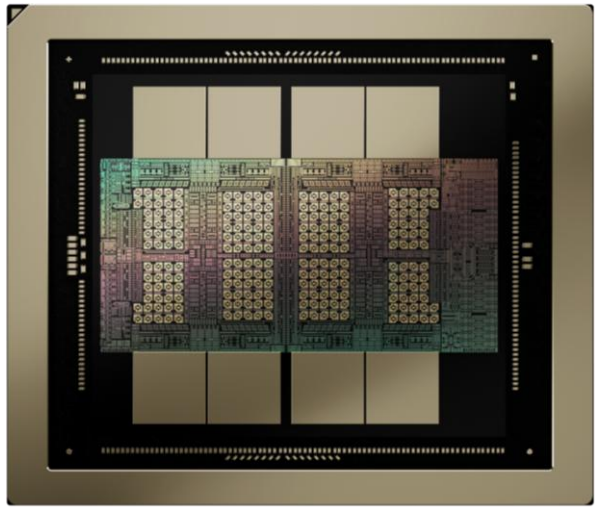


Spectrum-X
102.4T CPO



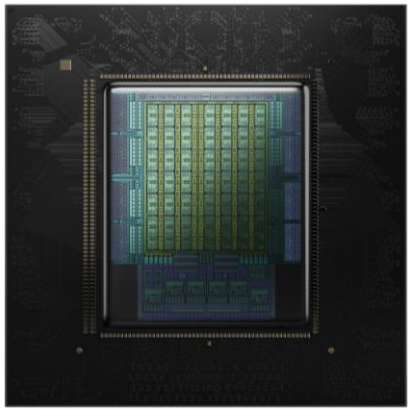
Vera Rubin DGX
SuperPod

Agentic AI & Reinforcement Learning Require Optimised CPUs



Tokens
Code and Queries

Evaluations
Results and Data



Model / Agent
Rubin GPU

Reinforcement Learning Training
Agent Inference

Sandbox Execution
Vera CPU

Tool Calls, Simulation & Tests
Search, Query, Transform
API and Browser

NVIDIA Vertically Integrated Full Stack for AI Factories

Simplifies AI Factories with Full Stack Solutions for Agentic AI, Physical AI, and HPC

Applications

Agentic AI



Supply Chain | Customer Service | Marketing | Video Analytics

Physical AI



Robotics | Digital Twin

HPC



CUDA-X Libraries

Models

NVIDIA Enterprise AI Factory Validated Design

Models

Security

Observability

AI Ops

NVIDIA AI Enterprise

NVIDIA NeMo | NVIDIA Omniverse

NVIDIA Run:ai

Confidential Computing

Data Platform

Data Connectors

Repository

Infrastructure

Chips

Energy

NVIDIA Enterprise Reference Architecture

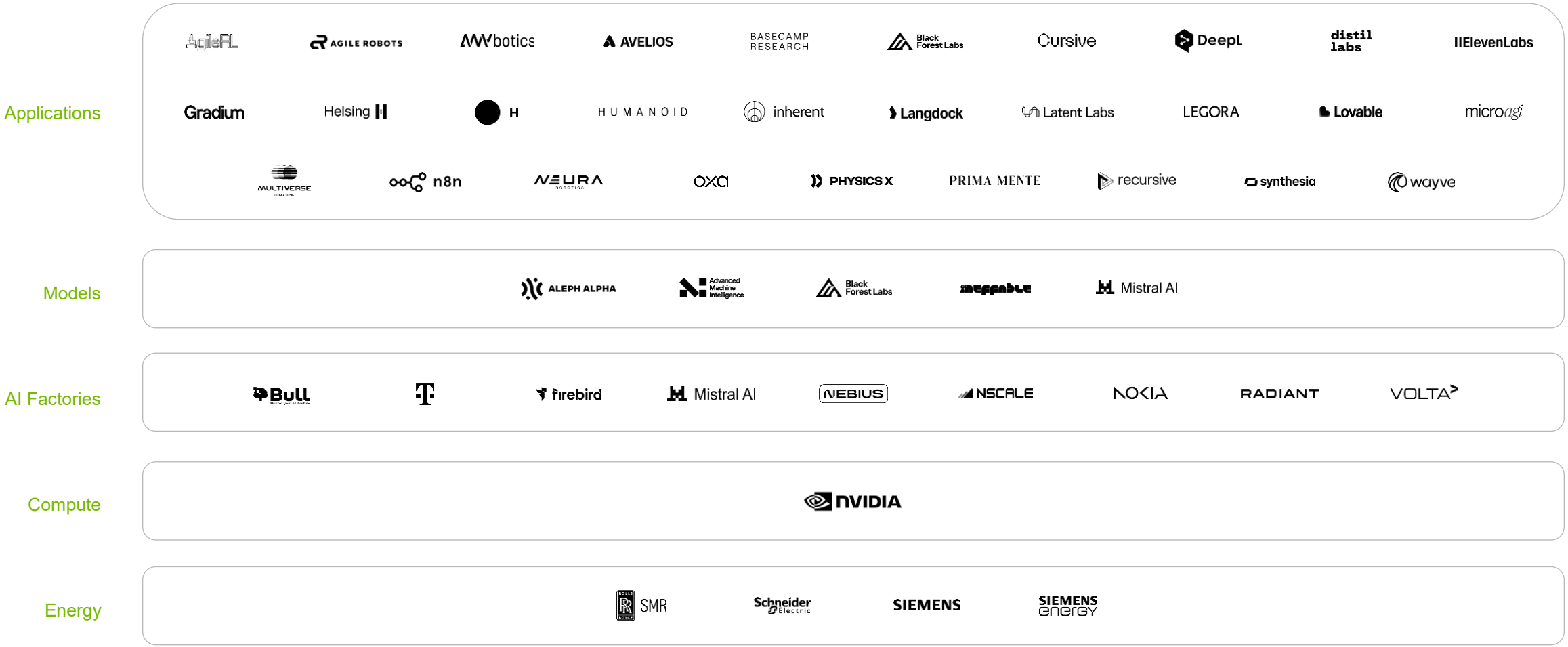
Kubernetes

NVIDIA-Certified Systems | NVIDIA Networking | NVIDIA-Certified Storage

Mechanical | Electrical | Power

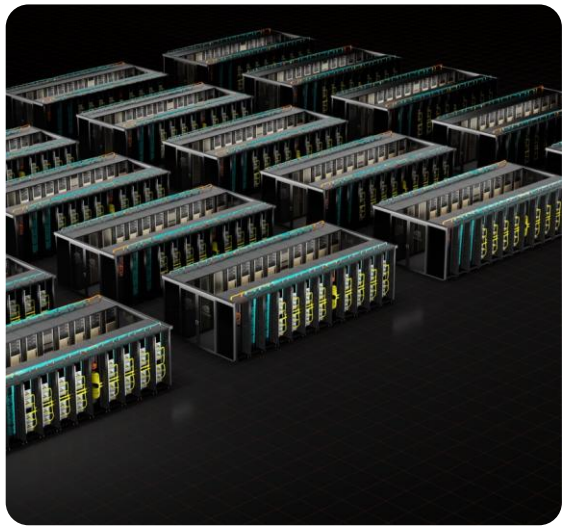
NVIDIA

Building Europe’s AI Economy



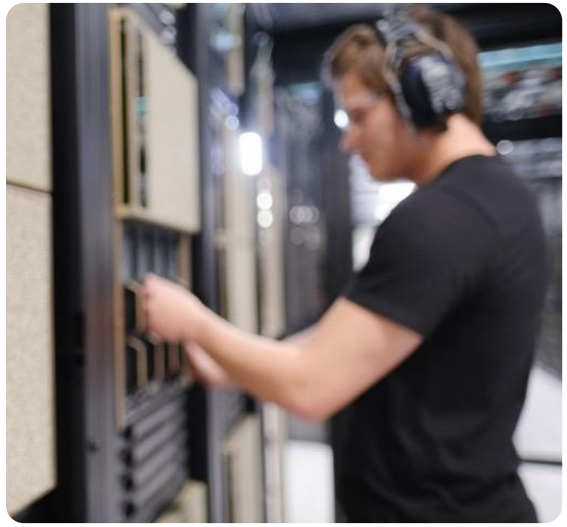
AI Nations Framework

NVIDIA powers every AI at every level



Infrastructure

- World-class AI DC HW
- World's #1 AI Platform
- Sovereign LLM expertise
- Platforms: AI, Sim, HPC, Robotics, AV, ..
- Inference Microservices (NIMs)



Expertise

- NVIS Operations Center
- University Teaching Kits
- Deep domain knowledge
- Inception Innovation Center



Ecosystem

- 20K Inception startups
- Worlds leading ISVs on AI
- Global GSIs + AI blueprints
- AI Nations expertise



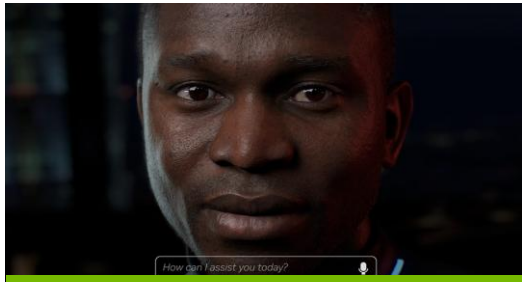
Programs

- AI Summits & Workshops, Hackathons, Bootcamps
- Deep Learning Institute
- Global AI marketing

NVIDIA Blueprints: A foundation for Sovereign AI Applications

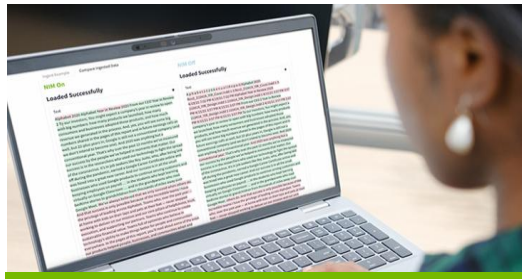
Available on build.nvidia.com

Digital Humans
for Customer Service



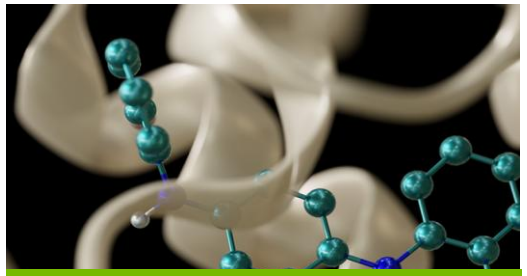
NVIDIA AI Blueprint

Multimodal PDF Data Extraction
for Enterprise RAG



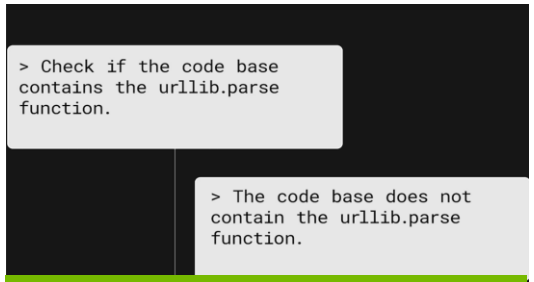
NVIDIA AI Blueprint

Generative Virtual Screening
for Drug Discovery



NVIDIA BioNeMo Blueprint

Vulnerability Analysis
for Container Security



NVIDIA AI Blueprint

AI Virtual Assistants
for Customer Service



NVIDIA AI Blueprint

Visual AI Agent
for Video Search and Summarization



NVIDIA AI Blueprint

3D Conditioning for
Precise Visual Generative AI



NVIDIA Omniverse Blueprint

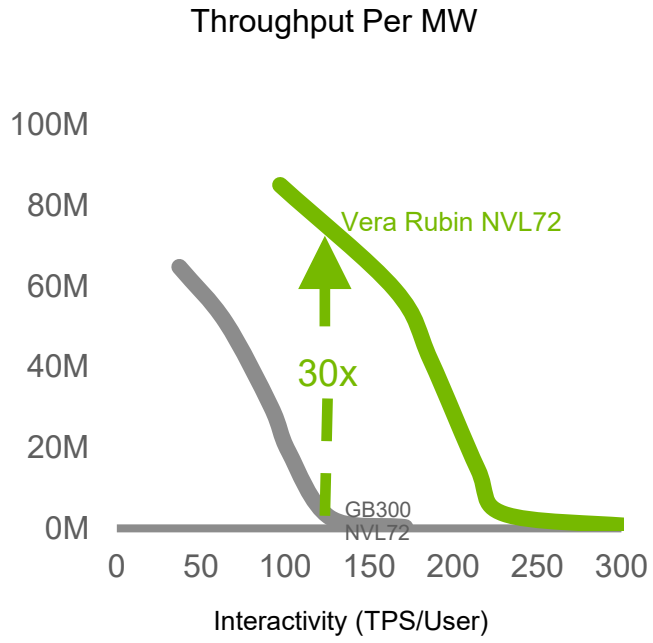
Digital Twin for
Interactive Fluid Simulation



NVIDIA Omniverse Blueprint

NVIDIA AI Factories Are Productive, Durable, and Fungible

A new investable asset class



Productive
Extreme codesign maximizes ROI

Volta
10 years old, still serving
4x $/hour than first customer

Ampere
6 years old
New contracts with **3-year** commits

NVL72
Contracts with **7-year** commits
Option to 10 years

Durable
Appreciating lifelong value

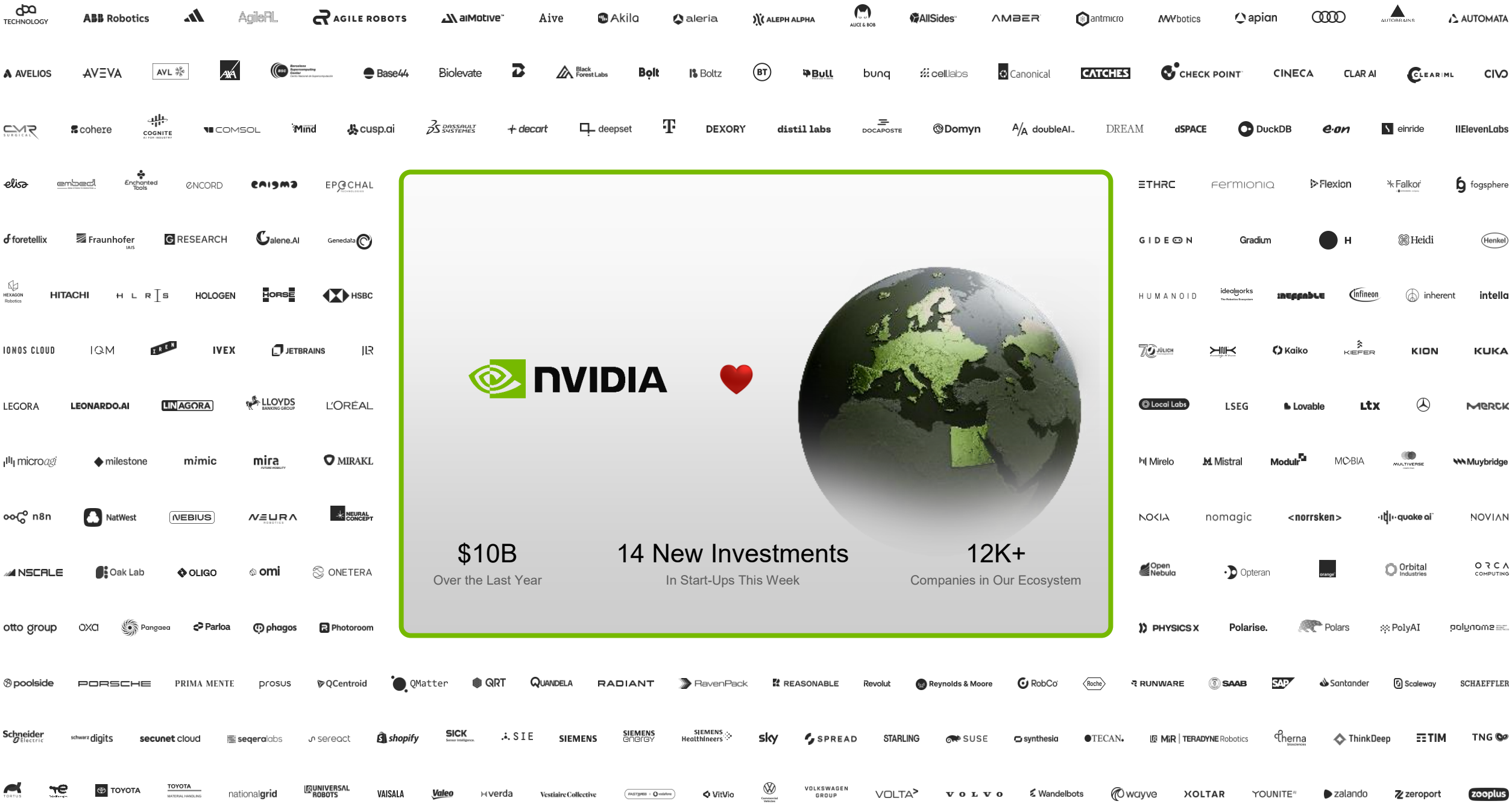
Every computing workload—gaming, rendering, scientific computing, data processing, AI

Every closed
and open model

Every hyperscaler, sovereign cloud, enterprise on prem, robot, space

Fungible
One platform, everywhere, many offtakers, any workload

NVIDIA


NVIDIA
$10B
Over the Last Year
14 New Investments
In Start-Ups This Week
12K+
Companies in Our Ecosystem

Traditional Data Center Construction Delays AI Capacity

Complex on-site integration and skilled-labor shortages extend schedules and delay time to first token

288,000
Projected construction-worker shortfall by 2027

30-40%
Of project labor hours are electrical

600,000
On-site labor hours for one conventional facility

~300
Skilled workers on-site at peak build out phase

Labor constraints extend schedules while increasing cost and risk, constraining how quickly new compute facilities can come online

NVIDIA

The NVIDIA DSX Full-Stack AI Factory Platform



CUDA-X Libraries

cuSPARSE
NIXL
CCCL
cuDNN
NVTX
cuBLAS
NVSHMEM
CUDA Python
NCCL
NVRTC
TensorRT-LLM
cuTILE Python
Megatron Core
CuTe DSL
DOCA
Transformer Engine
CUTLASS
Dynamo

Platform Libraries



Open Models

Vera Rubin NVL72

Rubin

Vera

BlueField-4

ConnectX-9



NVLink 6 Switch



Vera BlueField-4 Compute

BlueField-4



Spectrum-6 Switch

Spectrum-6 Switch



Groq 3 LPX

Groq 3



Vera BlueField-4 STX



Vera RTX Compute

Vera

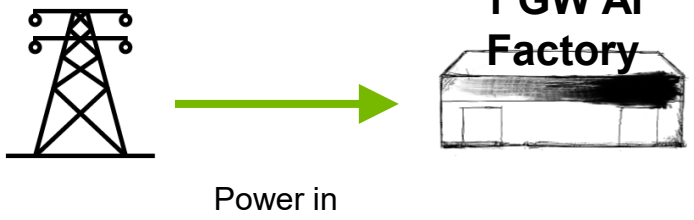


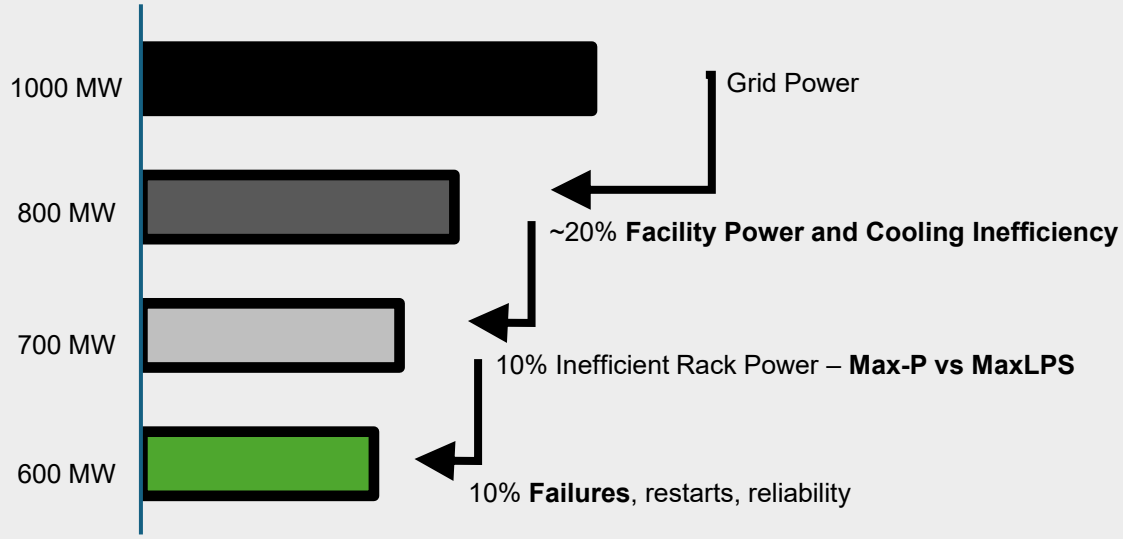


NVIDIA DSX

Objective is to maximize AI tokens per watt of available energy

Facility, rack, downtime inefficiencies cause only a fraction of grid power to contribute to AI output





60-70% for AI Output

- Stranded Power
- Fixed and Static, Conservative Power / Cooling Configurations
- Overprovisioned for Failures and Extremes

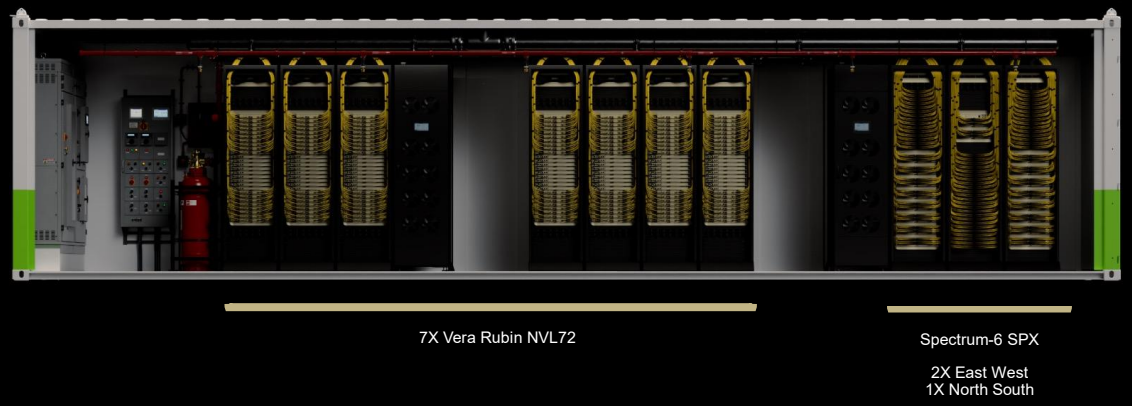
NVIDIA DSX POD

NVIDIA reference design for containerized AI factories

Factory integration accelerates time to first token
Vera Rubin infrastructure is installed, cabled, piped, and tested in the factory, reducing on-site work

On-site deployment can shrink from months to days
Connect containers to prepared site power, cooling, and networking, then complete bring-up and acceptance testing

Designed by NVIDIA, delivered by partners
NVIDIA engineers and validates the design, partners customize, manufacture, and support solutions



NVIDIA DSX POD Compute Container

THE FASTEST PATH TO PRODUCTIVE INTELLIGENCE AT SCALE

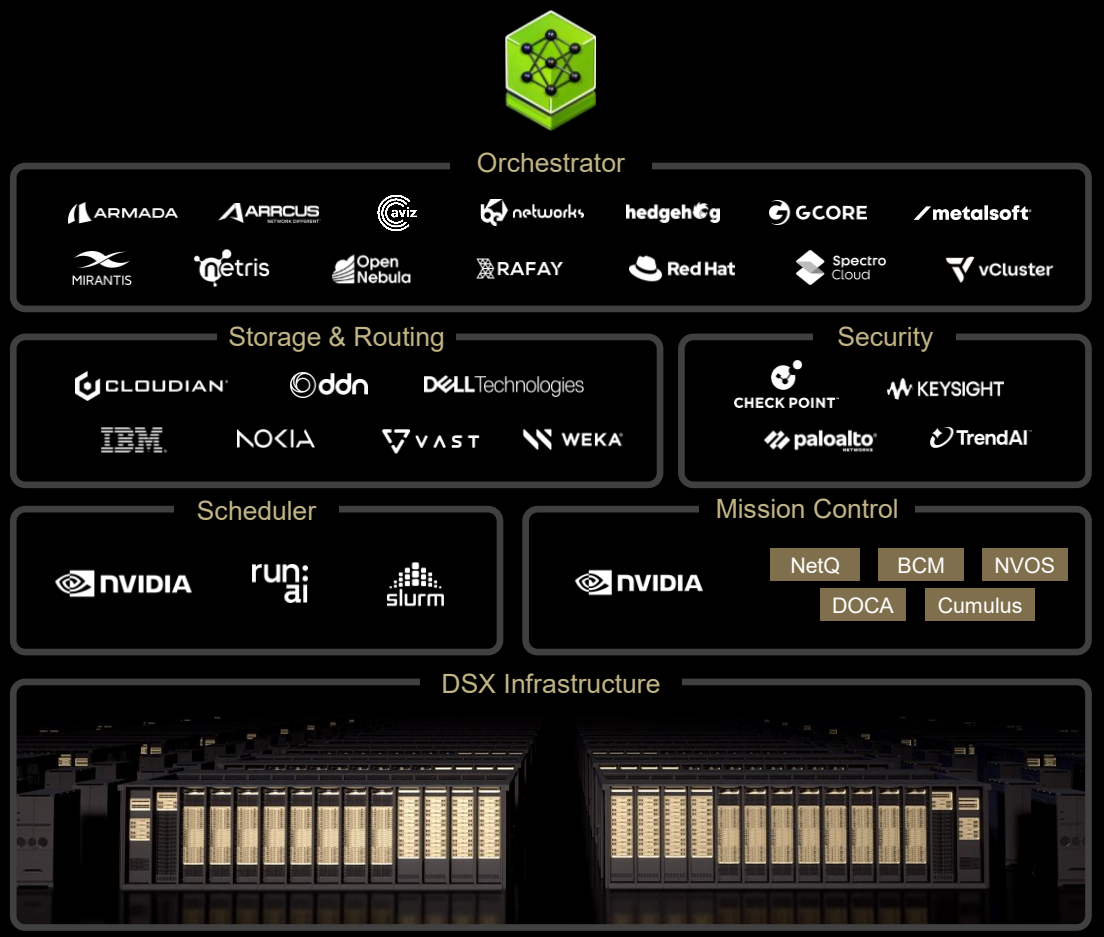
NVIDIA DSX Air

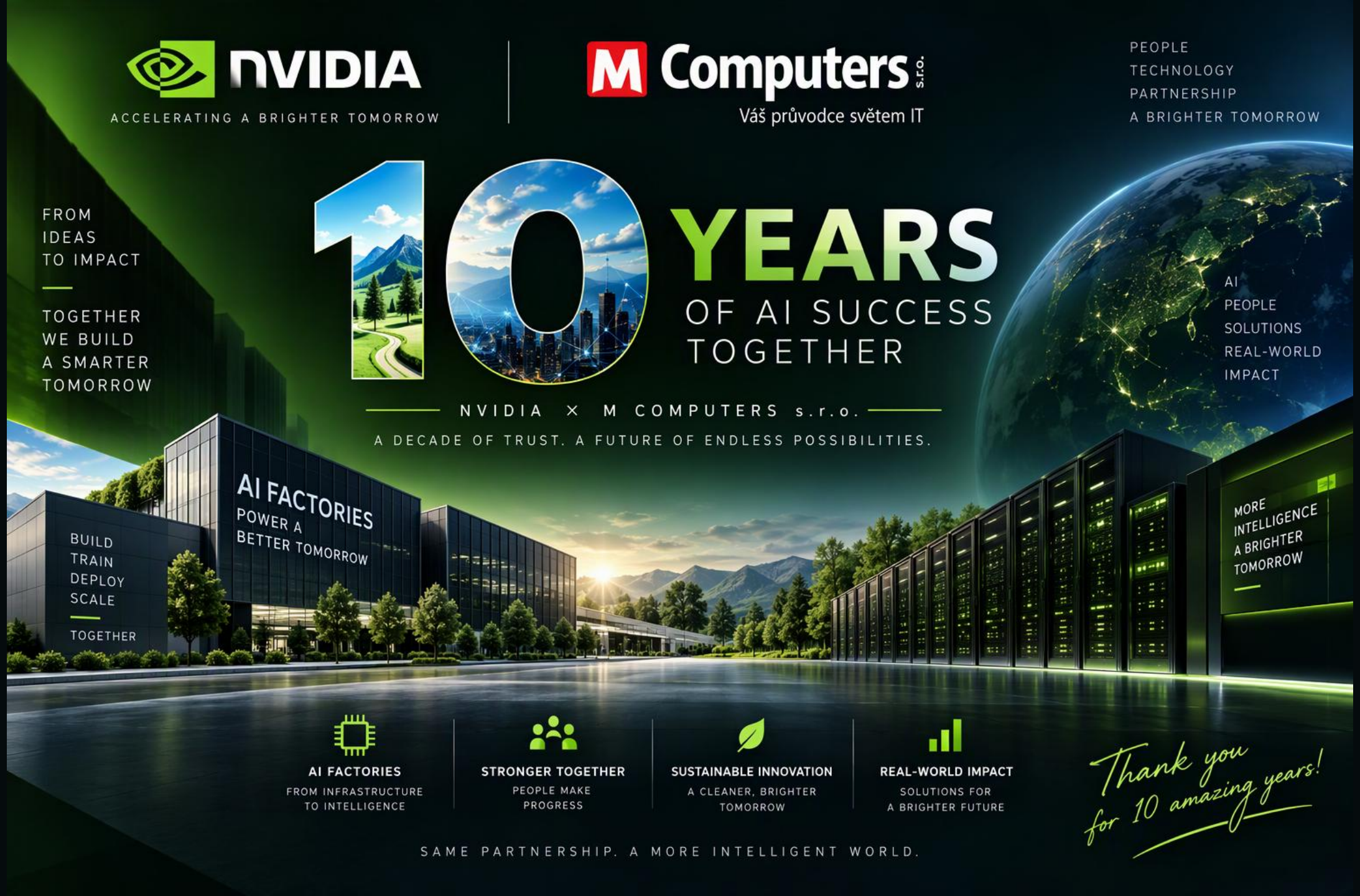
AI Factory Design and Validation Platform

Deploying AI Factories in Days Instead of Weeks

from

6 Months to 1 Week	3 Months to 1 Week	2 Weeks to 1 Day
Infrastructure Validation	Software Bring-Up	Deployment Time




NVIDIA
ACCELERATING A BRIGHTER TOMORROW
M Computers s.r.o.
Váš průvodce světem IT
PEOPLE
TECHNOLOGY
PARTNERSHIP
A BRIGHTER TOMORROW
FROM
IDEAS
TO IMPACT
TOGETHER
WE BUILD
A SMARTER
TOMORROW
10 YEARS
OF AI SUCCESS
TOGETHER
AI
PEOPLE
SOLUTIONS
REAL-WORLD
IMPACT
NVIDIA × M COMPUTERS s.r.o.
A DECADE OF TRUST. A FUTURE OF ENDLESS POSSIBILITIES.
AI FACTORIES
POWER A
BETTER TOMORROW
BUILD
TRAIN
DEPLOY
SCALE
TOGETHER
MORE
INTELLIGENCE
A BRIGHTER
TOMORROW
AI FACTORIES
FROM INFRASTRUCTURE
TO INTELLIGENCE
STRONGER TOGETHER
PEOPLE MAKE
PROGRESS
SUSTAINABLE INNOVATION
A CLEANER, BRIGHTER
TOMORROW
REAL-WORLD IMPACT
SOLUTIONS FOR
A BRIGHTER FUTURE
Thank you
for 10 amazing years!
SAME PARTNERSHIP. A MORE INTELLIGENT WORLD.


NVIDIA