AI, Gen AI and the Metaverse: A New Era of Possibilities and Challenges.

Valerio Rizzo, PhD | EMEA Head of AI & Metaverse SME

Workshop sul Calcolo nell'I.N.F.N. Palau (Sassari) 20 - 24 maggio 2024

Lenovo

About Me

Valerio Rizzo, PhD

| EMEA HEAD of AI & Metaverse SME

| Lenovo (Italy) Srl | ISG

| **email:** vrizzo@lenovo.com

Technology

- Machine Learning/Deep Learning Hardware and Software Infrastructure
- Digital Twin / Metaverse Hardware and Software
- AI Vertical and Horizontal use case applications

Location

- HO: Based in Sicily, Italy
- HQ: Lenovo (Italy) Srl, Via S. Bovio, 3, 20054 Segrate MI

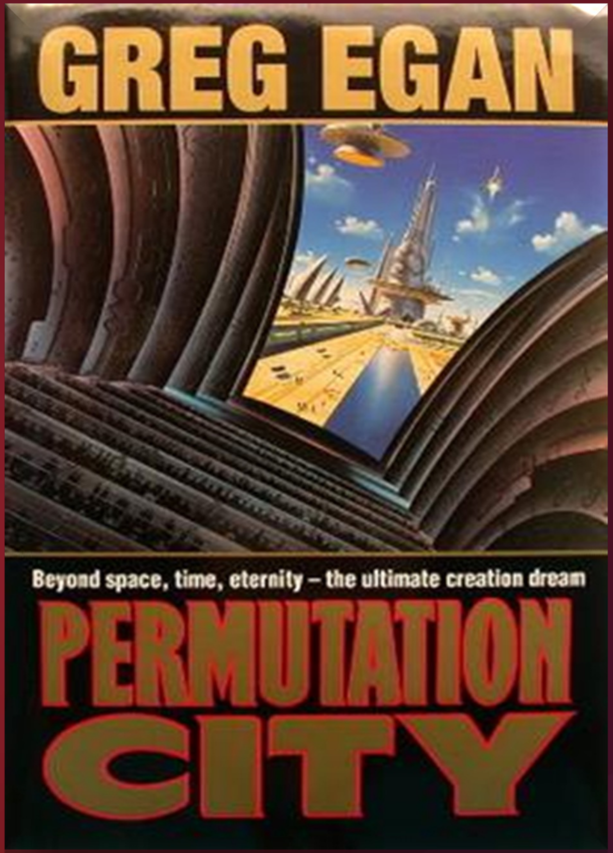
Professional and Educational Background

- PhD in Neuroscience and Neurophysiology
- Researcher, Lecturer, Reviewer and Associate Editor in neuroscience and neurophysiology
- Extensive professional experience in Immersive technology applied to pre-clinical research and M&E

Passion

- Climbing, Boxing, Trekking, Yoga Nidra
- Avid Book Reader and Movie Watcher
- Photogrammetry, VR Game Dev, Coding
- VR Game Player

From Fiction to Science



ARE YOU LIVING IN A COMPUTER SIMULATION?

BY NICK BOSTROM

[Published in *Philosophical Quarterly* (2003) Vol. 53, No. 211, pp. 243-255. (First version: 2001)]

This paper argues that *at least one* of the following propositions is true: (1) the human species is very likely to go extinct before reaching a "posthuman" stage; (2) any posthuman civilization is extremely unlikely to run a significant number of simulations of their evolutionary history (or variations thereof); (3) we are almost certainly living in a computer simulation. It follows that the belief that there is a significant chance that we will one day become posthumans who run ancestor-simulations is false, unless we are currently living in a simulation. A number of other consequences of this result are also discussed.

I. INTRODUCTION

Many works of science fiction as well as some forecasts by serious technologists and futurologists predict that enormous amounts of computing power will be available in the future. Let us suppose for a moment that these predictions are correct. One thing that later generations might do with their super-powerful computers is run detailed simulations of their forebears or of people like their forebears. Because their computers would be so powerful, they could run a great many such simulations. Suppose that these simulated people are conscious (as they would be if the simulations were sufficiently fine-grained and if a certain quite widely accepted position in the philosophy of mind is correct). Then it could be the case that the vast majority of minds like ours do not belong to the original race but rather to people simulated by the advanced descendants of an original race. It is then possible to argue that, if this were the case, we would be rational to think that we are likely among the simulated minds rather than among the original biological ones. Therefore, if we don't think that we are currently living in a computer simulation, we are not entitled to believe that we will have descendants who will run lots of such simulations of their forebears. That is the basic idea. The rest of this paper will spell it out more carefully.

Apart form the interest this thesis may hold for those who are engaged in futuristic speculation, there are also more purely theoretical rewards. The argument provides a stimulus for formulating some methodological and metaphysical questions, and it suggests naturalistic analogies to certain traditional religious conceptions, which some may find amusing or thought-provoking.

The structure of the paper is as follows. First, we formulate an assumption that we need to import from the philosophy of mind in order to get the argument started. Second,

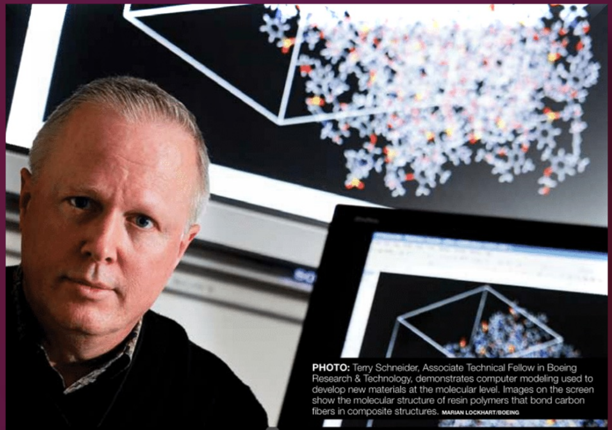


PHOTO: Terry Schneider, Associate Technical Fellow in Boeing Research & Technology, demonstrates computer modeling used to develop new materials at the molecular level. Images on the screen show the molecular structure of resin polymers that bond carbon fibers in composite structures. MARIAN LOCKHART/BOEING

Atoms **to airplanes**

New structures technologies, developed across Boeing, are helping accelerate product development **By Bill Seil**

Terry Schneider, an Associate Technical Fellow in Boeing Research & Technology, works in "atoms to airplanes" modeling, or the complete process of modeling an airplane computationally from a molecular level up to the full-scale, complete airframe.

One important goal of this work is to optimize the chemistry of polymers to increase the load-carrying capability of the carbon fiber in composites, which could significantly reduce the weight of next-generation composite structures.

"This is exciting work because we're able to rapidly assess hundreds of polymer candidates in a matter of weeks—a process that might take years in a lab," Schneider said. "We're also able to quickly determine their performance in large-scale laminated structures and screen for the best-performing candidates. This opens the door to huge cost savings in the future."

Work such as this demonstrates the benefits to Boeing generated by the company's enterprisewide approach to making research investments in key areas such as structures, a term that describes the physical airframe components of airplanes and other aerospace products. Critical aviation design issues—including weight, reliability and safety—all depend on the quality of research and planning that drives structures engineering.

Boeing has long been a leader in structures technology, and research conducted throughout the enterprise has steadily improved the design of structures and the materials used to make them. The challenge today is to increase the company's competitive edge by investing in research that generates maximum benefit for Boeing's range of products, both commercial and military.

That's why, in 2008, the company created its Enterprise Technology Strategy (ETS), which takes a coordinated, "One Company" approach to technology development. The strategy is built around eight technology areas, or domains, that support Boeing's many business programs and can create a sustainable technical competitive advantage that helps the company grow.

3 DECEMBER 2009–JANUARY 2010 / **BOEING FRONTIERS**

From Digital Twin to Industrial Metaverse

Industrial Metaverse

Whole-System DT

Immersive DT

Digital Twin

“

A massively *scaled* and *interoperable* network of real-time rendered 3d virtual worlds that can be experienced *synchronously and persistently* by an effectively *unlimited number of users* with an *individual sense of presence* and with *continuity of data*, such as identity, history, entitlements, objects , communications and payments ”

Matthew Ball, The Metaverse

SIEMENS

A virtual world in which we can **interact in real time with photorealistic, physics-based digital twins** of our real world. We believe **digital twins** are **the building blocks for the Metaverse.**

NVIDIA

Industrial Metaverse enables industrial companies of all sizes to create **closed-loop digital twins with real-time performance data, ideal for running simulations** and AI-accelerated processes for advanced applications such as **autonomous factories that rely on intelligent sensors and connected devices.**

IndustrialMetaverse.org

A real-time, persistent simulation space that is the **sum of all virtual worlds, digital twins, and augmented reality** that **connects digital economic assets and infrastructure** on a global scale in the **industrial and commercial setting.**

Microsoft

Industrial Metaverse enables **humans and AI to work together to design, build, operate, and optimize physical systems** using digital technologies.

sme

A **systematic discipline that combines hardware** [...] **data conversions** through analytics/machine learning, **time histories** through cyber-infrastructure, **cognition** through human-machine interface**, and configuration** through the Metaverse.

COSMOTECH

The Industrial Metaverse enables the creation of **digital twins of places, processes, real-world objects, and the humans** who interact with them.

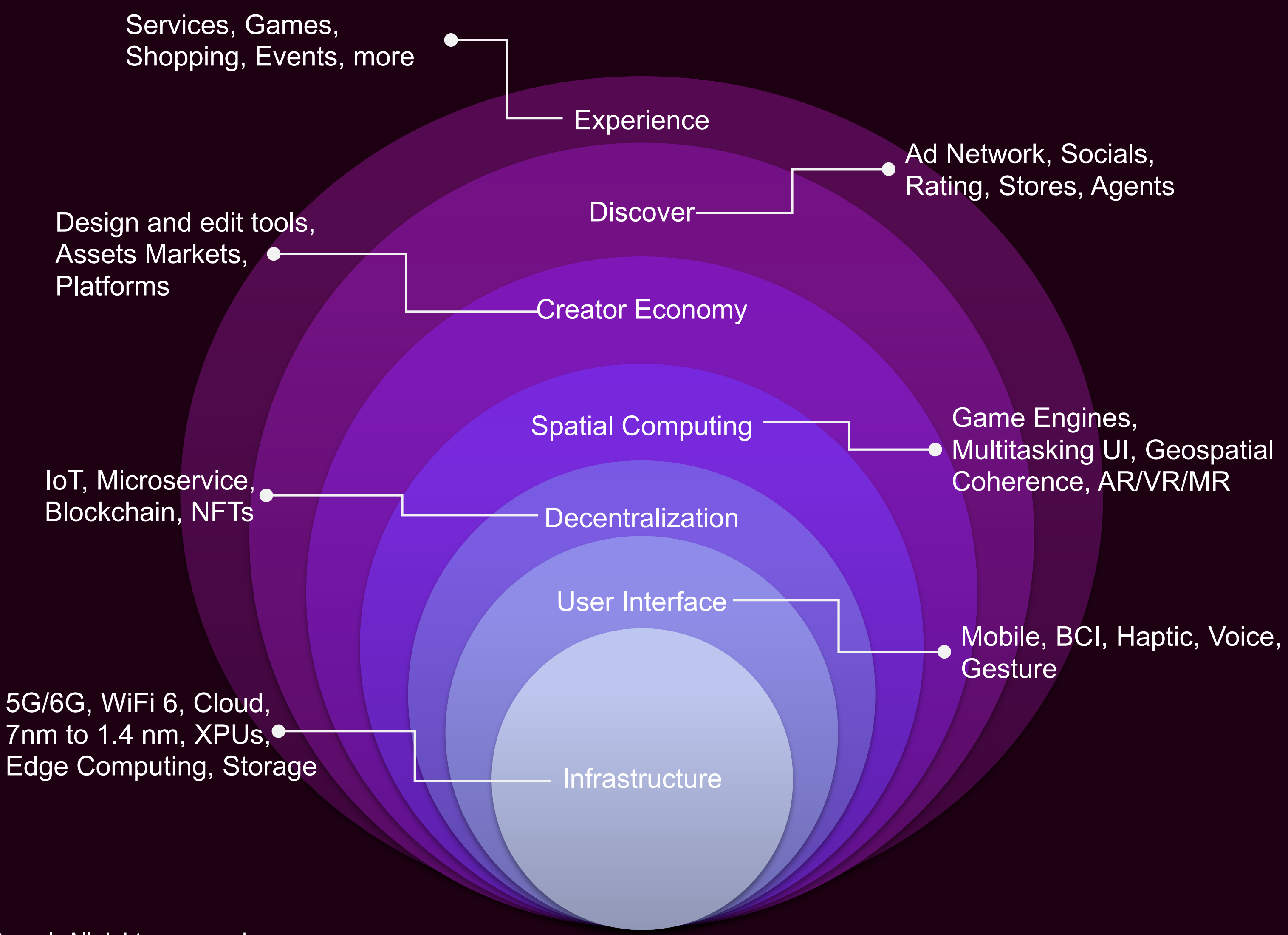
Source: Arthur D. Little

“

Connected whole-system digital twin with functionalities to interact with the real system in its environment, allowing decision makers to better understand the past and forecast the future.”

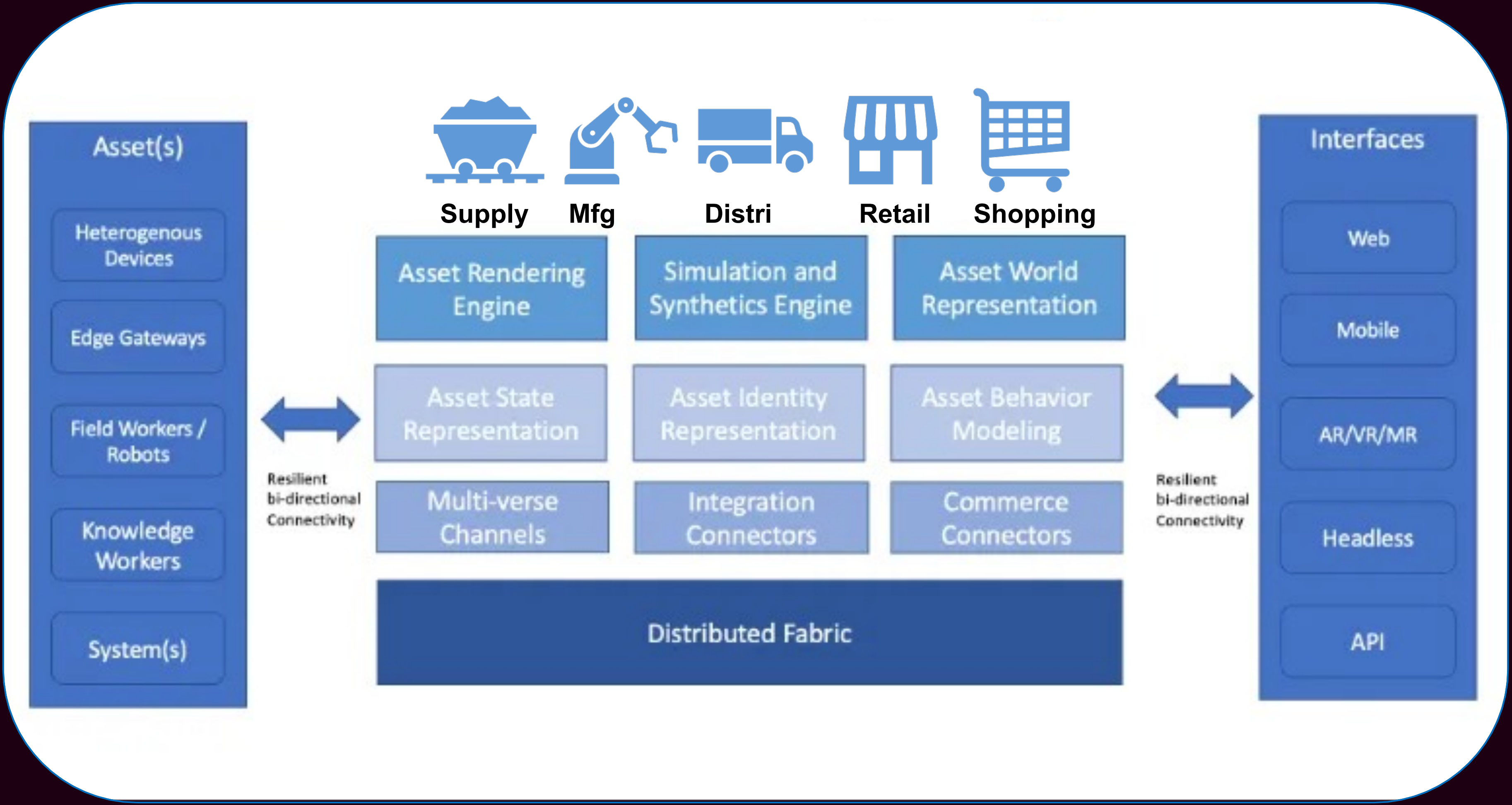
Arthur D. Little

Anatomy of the Metaverse



Thien Huynh-The et al. “Artificial Intelligence for the Metaverse: A Survey” arXiv:2202.10336v1 [cs.CY] 15 Feb 2022

Metaverse System Model



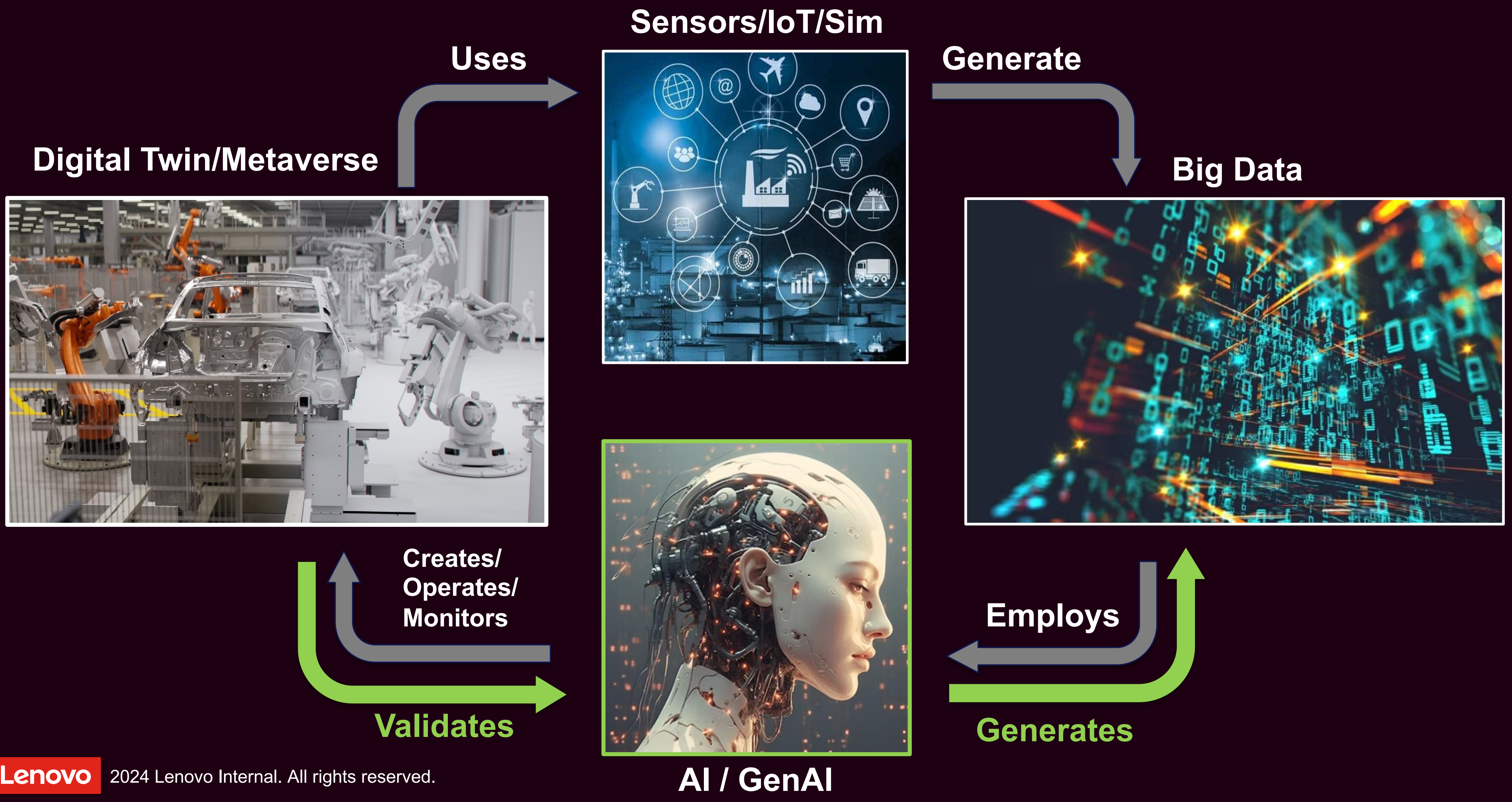
https://medium.com/data-science-at-microsoft/industrial-metaverse-a-software-and-data-perspective-d09950a453f6

"The natural habitat of AI is in the virtual world."

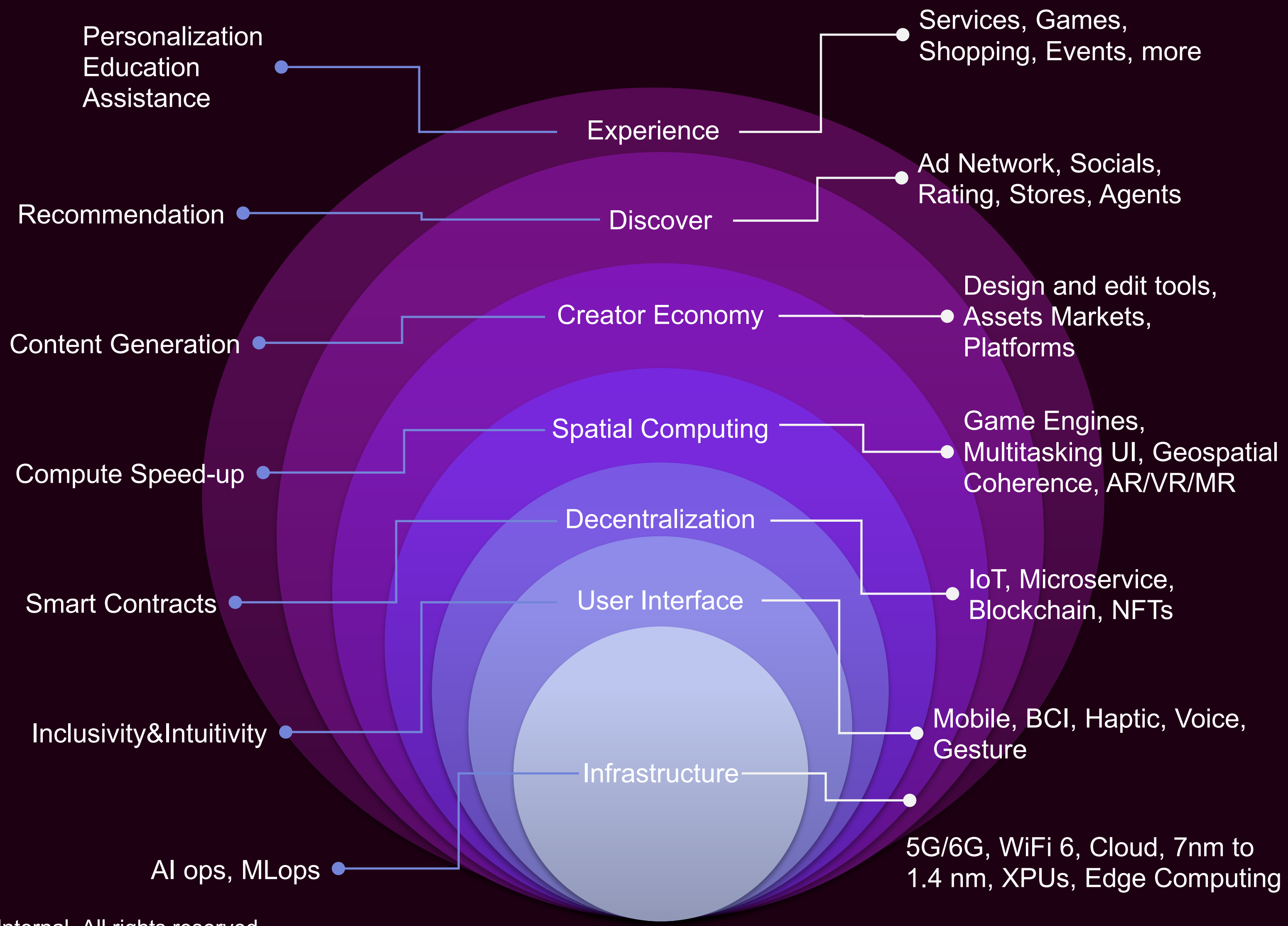
Dr. Michael Grieves

- Intelligent digital twins and the development and management of complex systems -

The Intertwined Nature of Metaverse and AI



AI value for the Metaverse



Thien Huynh-The et al. “Artificial Intelligence for the Metaverse: A Survey” arXiv:2202.10336v1 [cs.CY] 15 Feb 2022

How Today’s AI is Shaping Tomorrow’s Possibilities

3D Modeling & Visualization

Decentralized Computing

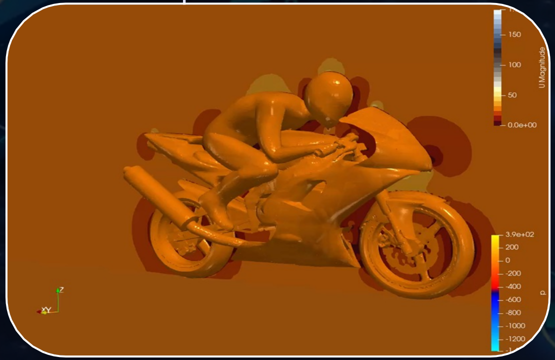
Network Optimization

Confidential AI Solutions

INDUSTRIAL METAVERSE

Spatial Computing

Human-Machine Interactivity



Physically Accurate Simulations

Realistic Interactive Virtual Entities

How Today’s AI is Shaping Tomorrow’s Possibilities

3D Modeling & Visualization

Decentralized Computing

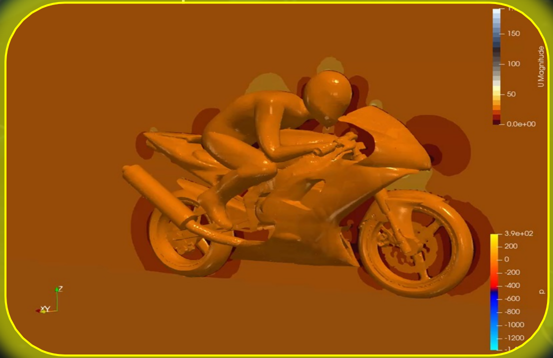
Network Optimization

Confidential AI Solutions

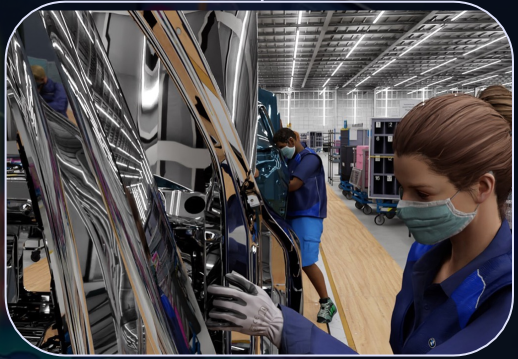
INDUSTRIAL METAVERSE

Spatial Computing

Human-Machine Interactivity



Physically Accurate Simulations



Realistic Interactive Virtual Entities

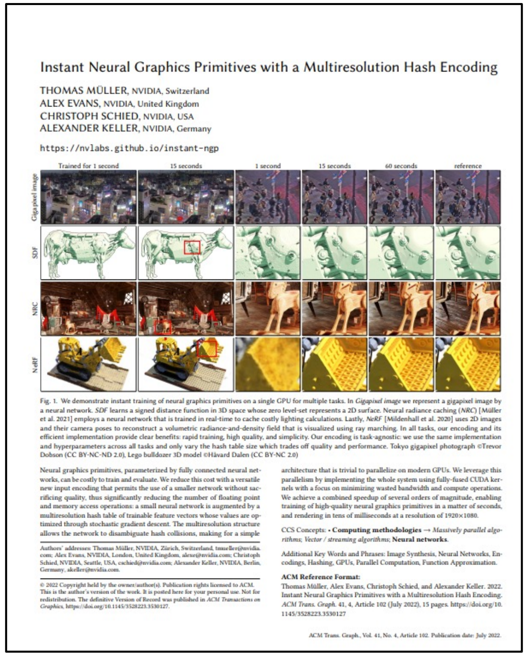
Advances in Neural Rendering and Mesh Generation

PixelNeRF (2021)


pixelNeRF: Neural Radiance Fields from One or Few Images

Alex Yu Vickie Ye Matthew Tancik Angjoo Kanazawa
UC Berkeley

Instant Ngp (2022)


Instant Neural Graphics Primitives with a Multiresolution Hash Encoding

THOMAS MÜLLER, NVIDIA, Switzerland
ALEX EVANS, NVIDIA, United Kingdom
CHRISTOPH SCHIED, NVIDIA, USA
ALEXANDER KELLER, NVIDIA, Germany

https://nvlabs.github.io/instant-ngp

Neuralangelo (2023)


Neuralangelo: High-Fidelity Neural Surface Reconstruction

Zhaoshuo Li[1,2] Thomas Müller[1] Alex Evans[1] Russell H. Taylor[2] Mathias Unberath[2] Ming-Yu Liu[1] Chen-Hsuan Lin[1]
[1]NVIDIA Research [2]Johns Hopkins University
https://research.nvidia.com/labs/dir/neuralangelo

Magic3D (2023)

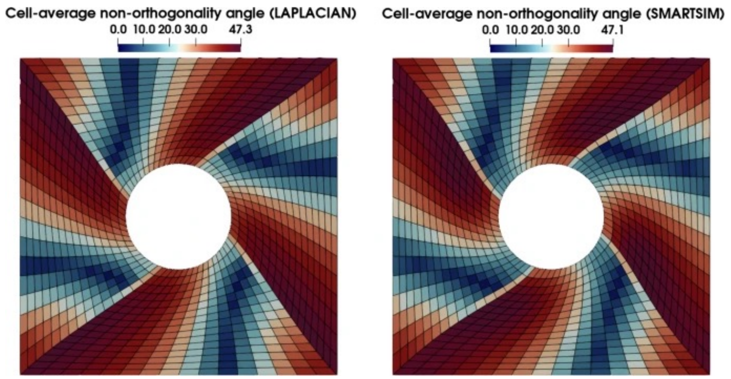

Magic3D: High-Resolution Text-to-3D Content Creation

Chen-Hsuan Lin* Jun Gao* Luming Tang* Towaki Takikawa* Xiaohui Zeng* Xun Huang Karsten Kreis Sanja Fidler† Ming-Yu Liu† Tsung-Yi Lin
NVIDIA Corporation
https://research.nvidia.com/labs/dir/magic3d

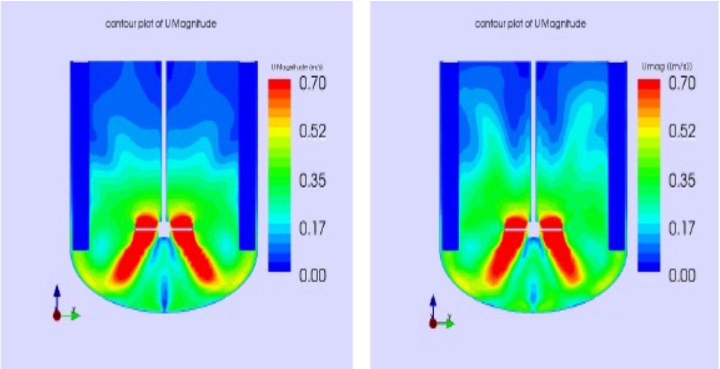
Metric	PixelNeRF	Instant NGP	Neuralangelo
Rendering Time (ms)	10-30 per pixel	<1 per pixel	~100-500 per pixel
Scene Complexity	High	Medium-High	Very High
Photorealism	No/ Limited	Yes	Yes
Real-time Capability	No	Yes	No

“Instant Neural Graphics Primitives with a Multiresolution Hash Encoding” Thomas Müller et al. ACM Transactions on Graphics (SIGGRAPH), July 2022a

Towards Real-Time Physically Accurate Simulations



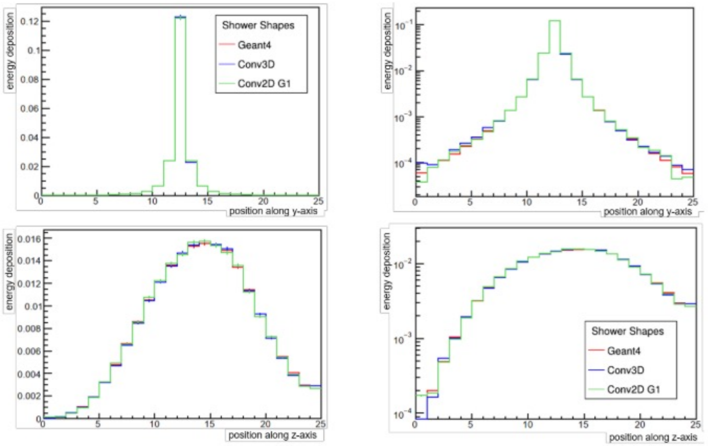
Approximating mesh-motion Laplacian mesh motion solver in OpenFOAM with MLP



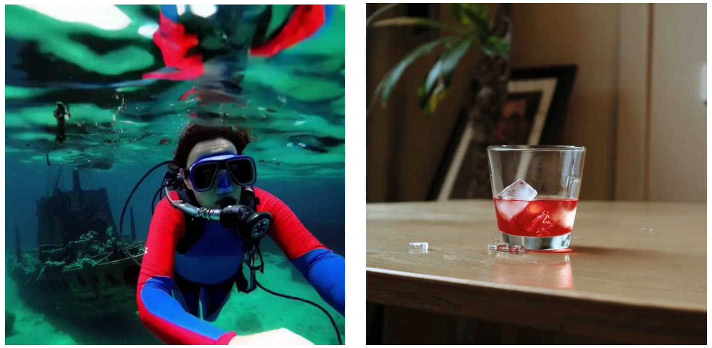
Using CNN to Solve Euler-Lagrange, Momentum Transfer, and Incompressible RANS Equations

AI / ML

Gen AI



Simulating high energy physics calorimeter detector outputs with 2D GAN



Video generation models as general purpose simulators of the physical world?

Tensors Reshape Compute Architectures

HPC

Viz / Render

Edge/Client

AI-optimized Portfolio from Model Development to Inferencing

80+ new and enhanced Infrastructure platforms – Pocket to Cloud, Edge to Core

Data Management

Solutions
High Performance File System (w/WEKA)
Object Storage Solutions (w/Cloudian)
DSS-G / Spectrum Scale
BeeGFS

DM & DE | DG7000
DM7100F | DE6600
DM5100 | DE6600

ML & Data Analytics

4-socket
SR850 V2
SR850 V3 Intel
SR860 V2
SR860 V3 Intel

2-Socket
SR650 V2
SR650 V3 Intel
SR655
SR655 V3 AMD
SR665
SR665 V3 AMD

Deep Learning Training HGX

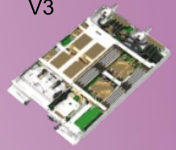
ST650 V3 Intel
SR670/75 V3 4-8x PCIe
SR670/75 V3 4-GPU HGX

NEW SR680a V3 8-GPU HGX
NEW SR685 V3 8-GPU HGX

Liquid Cooling Training

SD650-I V3
SD665-N
V3

NEW SR780a 8-GPU HGX



ThinkSystem

Data Science Workstation

Edge
NEW P3 Ultra
NEW P3 Tiny
Desktop
NEW PX
NEW P7
P620
NEW P5

NEW P3 Tower
Mobile
P16 Gen1
P16 Gen2
P16v Gen1
P1 Gen5
P1 Gen6

Lenovo ThinkStation

ThinkPad with Neural Processing Units

ThinkPad X13s Gen1 – 15 TOPS
ThinkPad Z13 Gen2 – 11 TOPS
ThinkPad Z16 Gen2 – 11 TOPS
ThinkPad T14s AMD Gen4 – 11 TOPS
ThinkPad T14 AMD Gen4 – 11 TOPS
ThinkPad T16 AMD Gen2 – 11 TOPS
ThinkPad X13 AMD Gen4 – 11 TOPS

Lenovo ThinkPad

Edge AI

Server
SE350
NEW SE350 V2
NEW SE360 V2
SE450
NEW SE455

Clients
SE10, SE10-I
M90
SE30
SE50
SE70

AI Appliance
SE70 AWS
Panorama

Lenovo ThinkEdge

Appliances

ThinkAgile MX Systems (Microsoft)
MX3330-F
MX3330-H
MX3331-F
MX3331-H
MX3530-F
MX3530-H
MX3531-F
MX3531-H

ThinkAgile HX Systems (Nutanix)
HX1330
HX1331
HX2330
HX2331
HX3330
HX3331
HX5530
HX5531

ThinkAgile VX Systems (VMware)
VX3331
VX3530-G
VX7531

ThinkAgile

Challenges ahead

Scalability
& Energy Efficiency

Security &
Privacy

Interoperability
& Standards

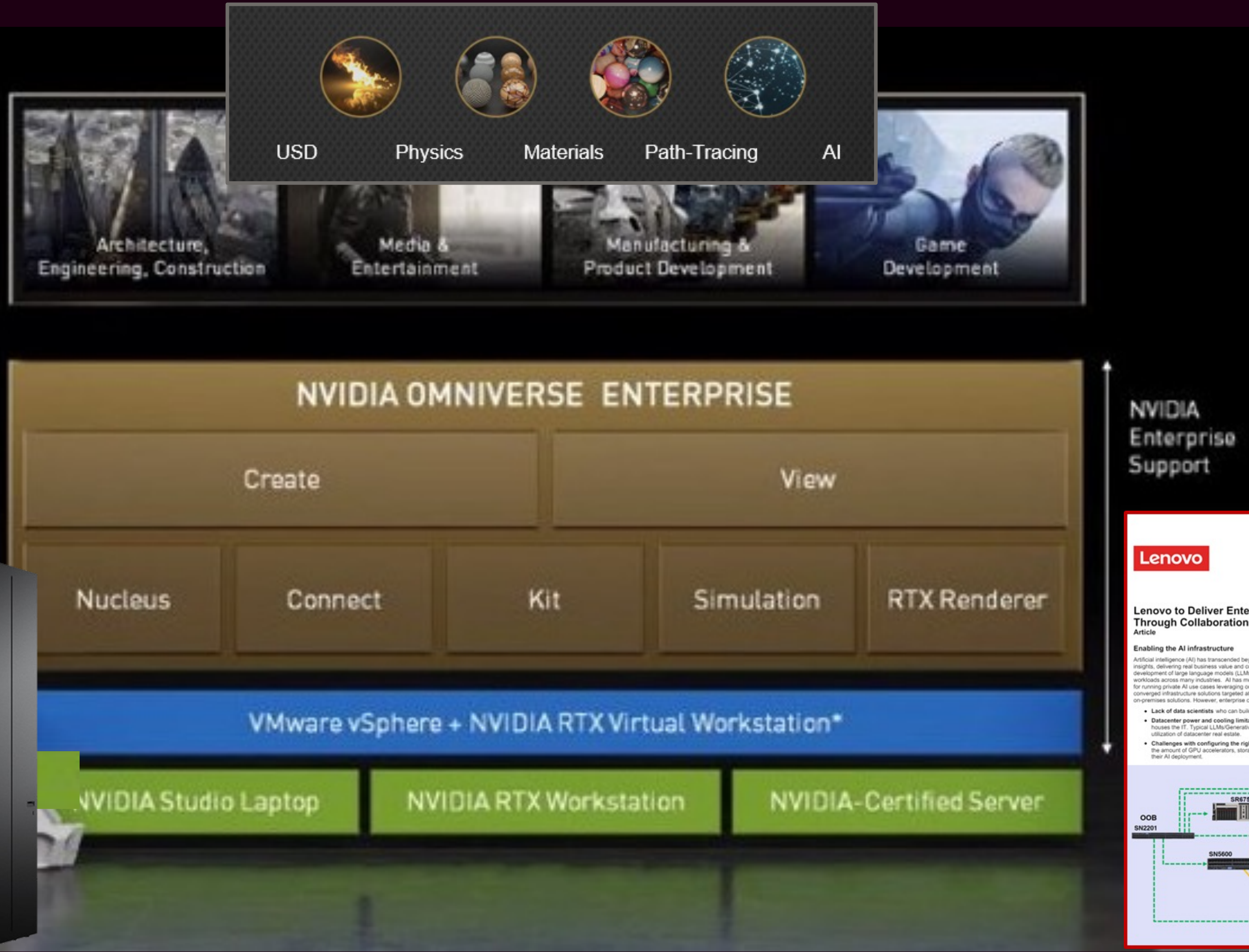
Compute & Storage
Optimization

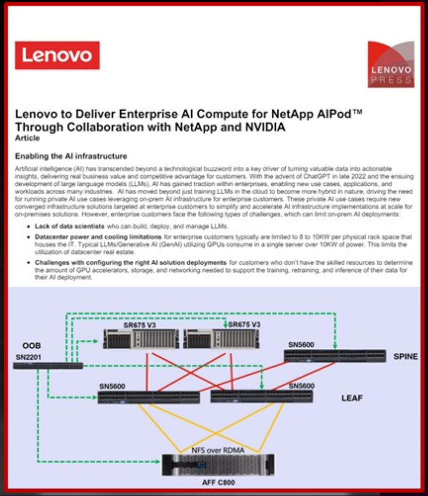
Ethic & Regulations

Lenovo E2E – OVX Infrastructure Solutions

Through Collaboration with NetApp and NVIDIA

LENOVO - E2E – OVX INFRASTRUCTURE




Lenovo

LENOVO PRESS

Lenovo to Deliver Enterprise AI Compute for NetApp AIPod™ Through Collaboration with NetApp and NVIDIA

Article

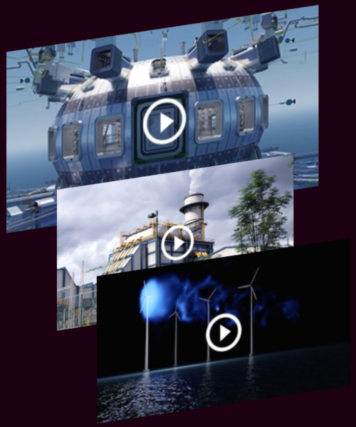
Enabling the AI infrastructure

The benefits of MV tech application embrace all industries.

Automotive

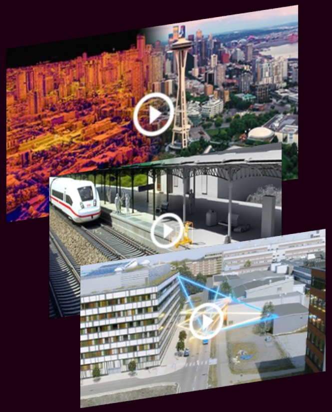
- Fast-Track Industrial Factory Planning
- Developing Custom Applications for Factory Planners

Energy



- Accelerating Fusion Reactor Design and Development
- Reducing Downtime and Unplanned Maintenance
- Optimizing Wind Farm Design and Electricity Generation

Infrastructure

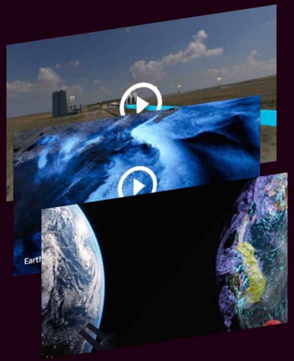


- Transforming Telco Network Planning and Operations
- Simulating and Optimizing Autonomous Railway Networks
- Testing and Optimizing 5G Deployment

Retail

- Autonomous Warehouse Robots
- Retail Layout
- Optimizing Distribution Center Throughput

Science



- Accelerating Carbon Capture and Storage
- Visualizing High-Resolution, Global-Scale Climate Data
- Accelerating Climate Research
- Visualizing Molecular Dynamics
- Brain Digital Twin

Industrial Metaverse

Are we there yet?

Takeaways:

Evolving DT Concept	The extended and enhanced use of digital twins is at the core of the Industrial Metaverse. AI applications can speed up 3D asset creation and prototyping while providing more intelligent capabilities to DT
HPC&AI-Powered Metaverse	Integrating AI into the HPC framework for the Industrial Metaverse unlocks new capabilities, driving innovation and efficiency in high-fidelity rendering and physical simulations.
Metaverse-Ready Infrastructure	The key technologies for achieving extended whole-system digital twins are not yet mature, but advances in AI, edge computing, and cloud infrastructure are rapidly closing the gap.
Challenges	Key issues include security, scalability, latency, costs, skill gaps, and regulatory compliance (including AI and data governance)
Future Trends	Accelerators mem bw will keep increasing, AI eats HPC, Raytracing engine will be integrated into AI superchips (i.e.: NVIDIA DGX) or Viz card will start employing DGX-like architectures

Smarter technology for all

Lenovo

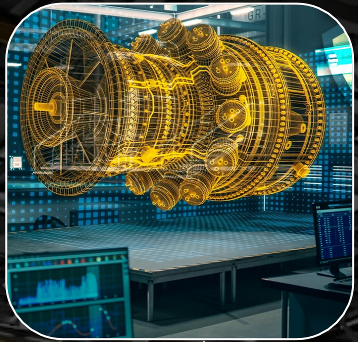
thanks.

Towards an Industrial Metaverse

A glimpse of the transformative power of Metaverse and AI

Synthetic Data

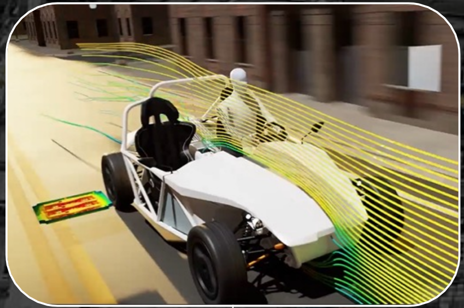
Design



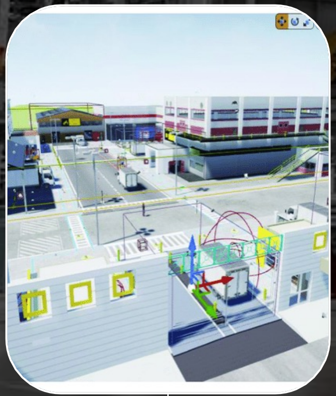
Additive Manufacturing

Process Optimization

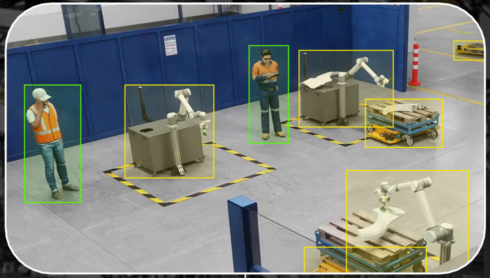
Simulations



Infrastructure



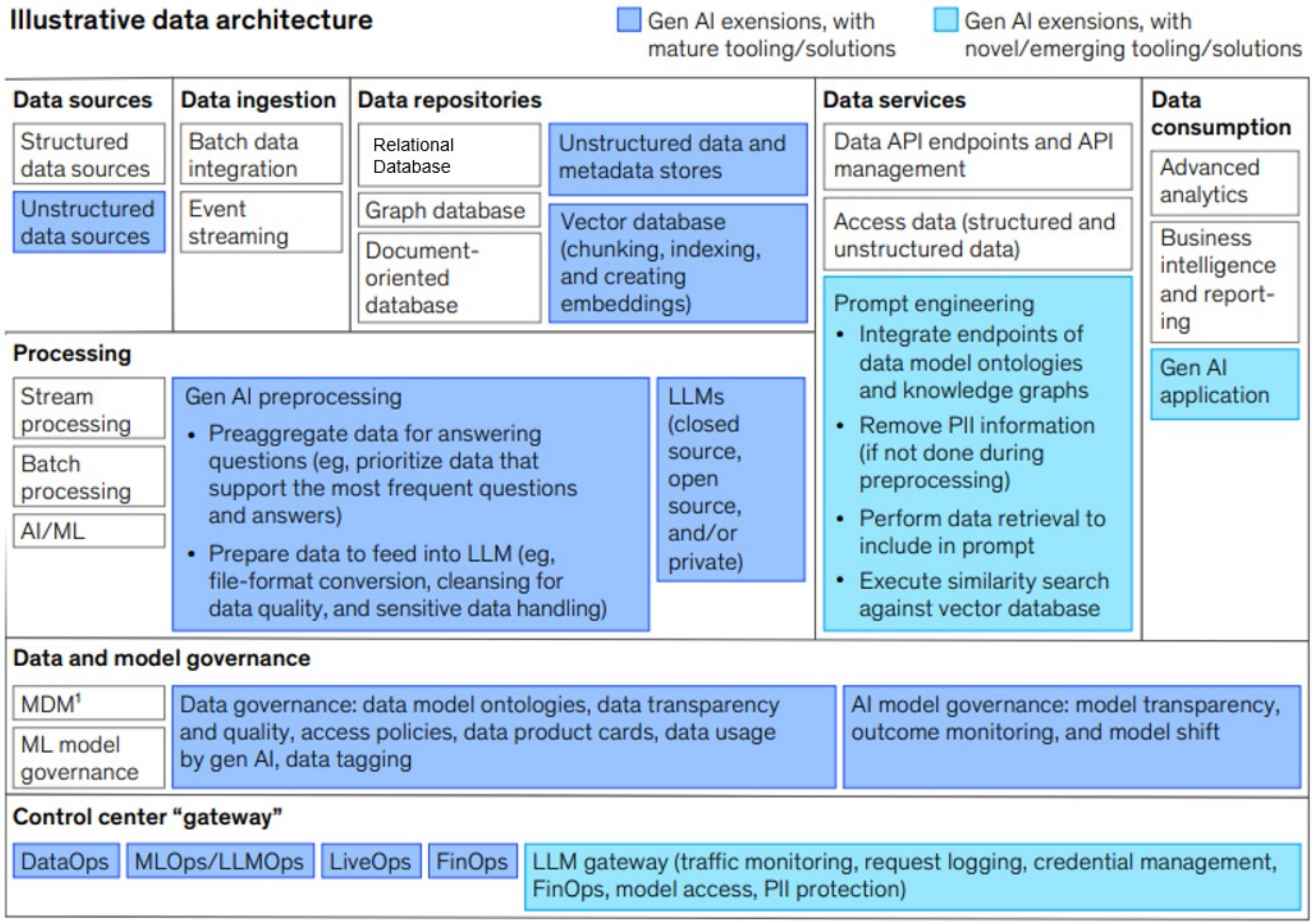
Proactive Safety



Augmented Assistance

Lenovo

Data Architecture Upgrades Needed for AI, GenAI and Industrial Metaverse



Source: Mckinsey
"The data dividend: Fueling generative AI" Sept 2023

Managing the transmission

1. DNN, Deep Transfer Learning and Federated Learning might be used for intelligent radio resource allocation in 5G/6G networks while meeting a very low latency.
2. RL was leveraged to address the resource-slicing problem for enhanced mobile broadband (eMBB) and uRLLC.
3. Efficient radio resource management with a distributed risk-aware ML approach to monitor and manage the transmission of non-scheduled and scheduled uRLLC traffics.
4. Two advanced CNN architectures, namely MCNet and SCGNet, were designed in the physical layer to automatically identify the modulation types of incoming signals
5. Online channel state information (CSI) prediction method was proposed a supervised learning framework by combining CNN and LSTM, in which two-stage training mechanism was deployed to improve the robustness and stableness of CSI estimation in practical 5G wireless systems
6. An end-to end 3D CNN architecture named ST-3DNet was designed for data traffic forecasting.

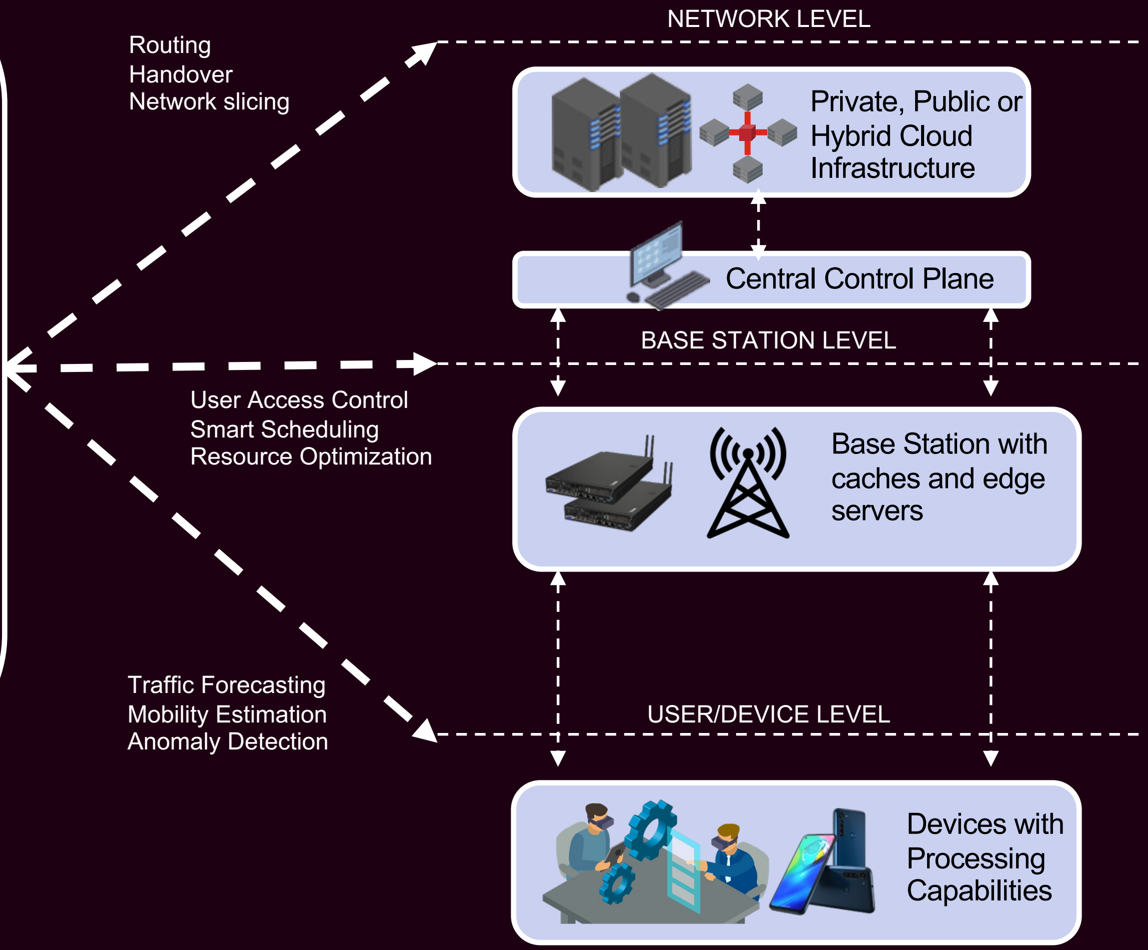
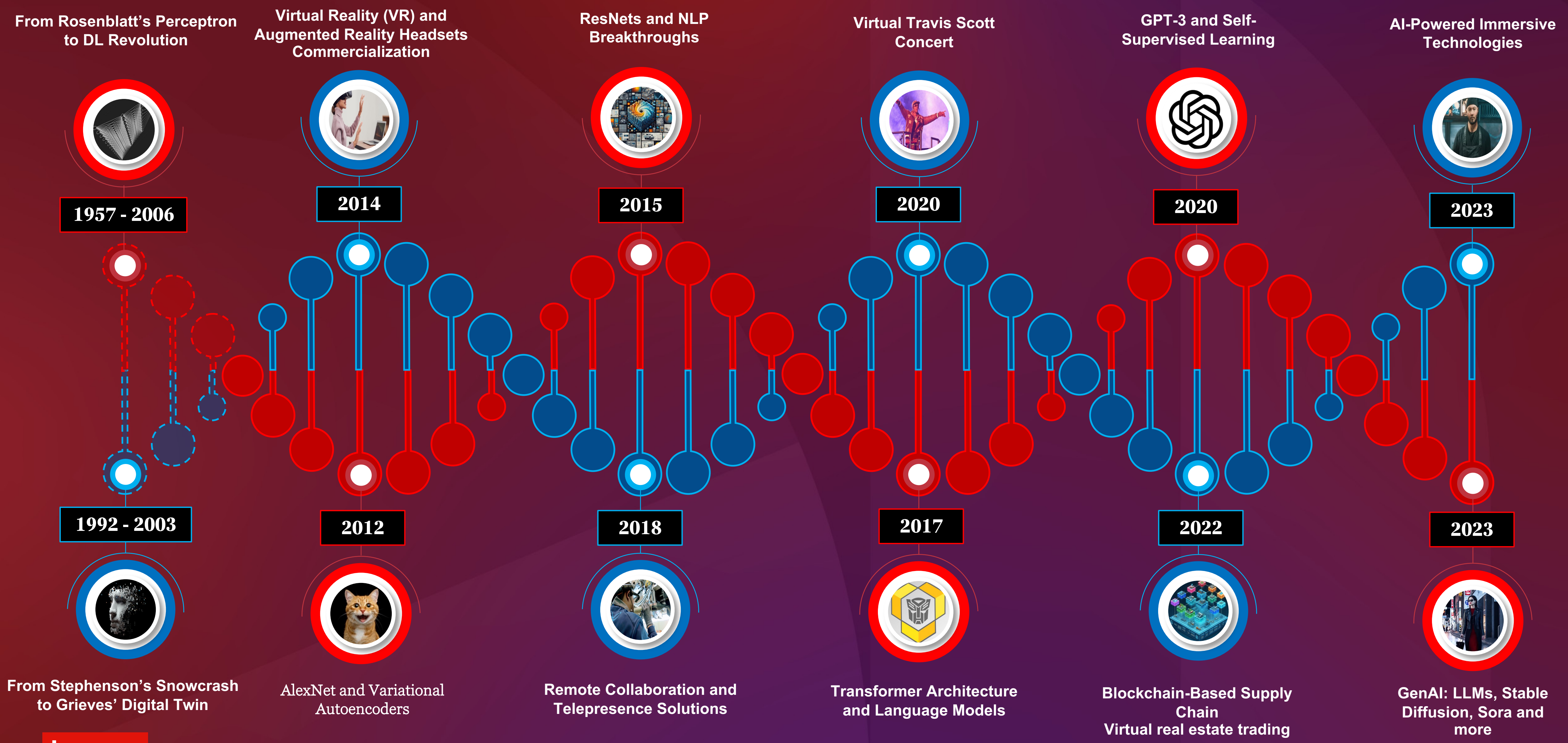


Diagram with courtesy of Thien Huynh-The

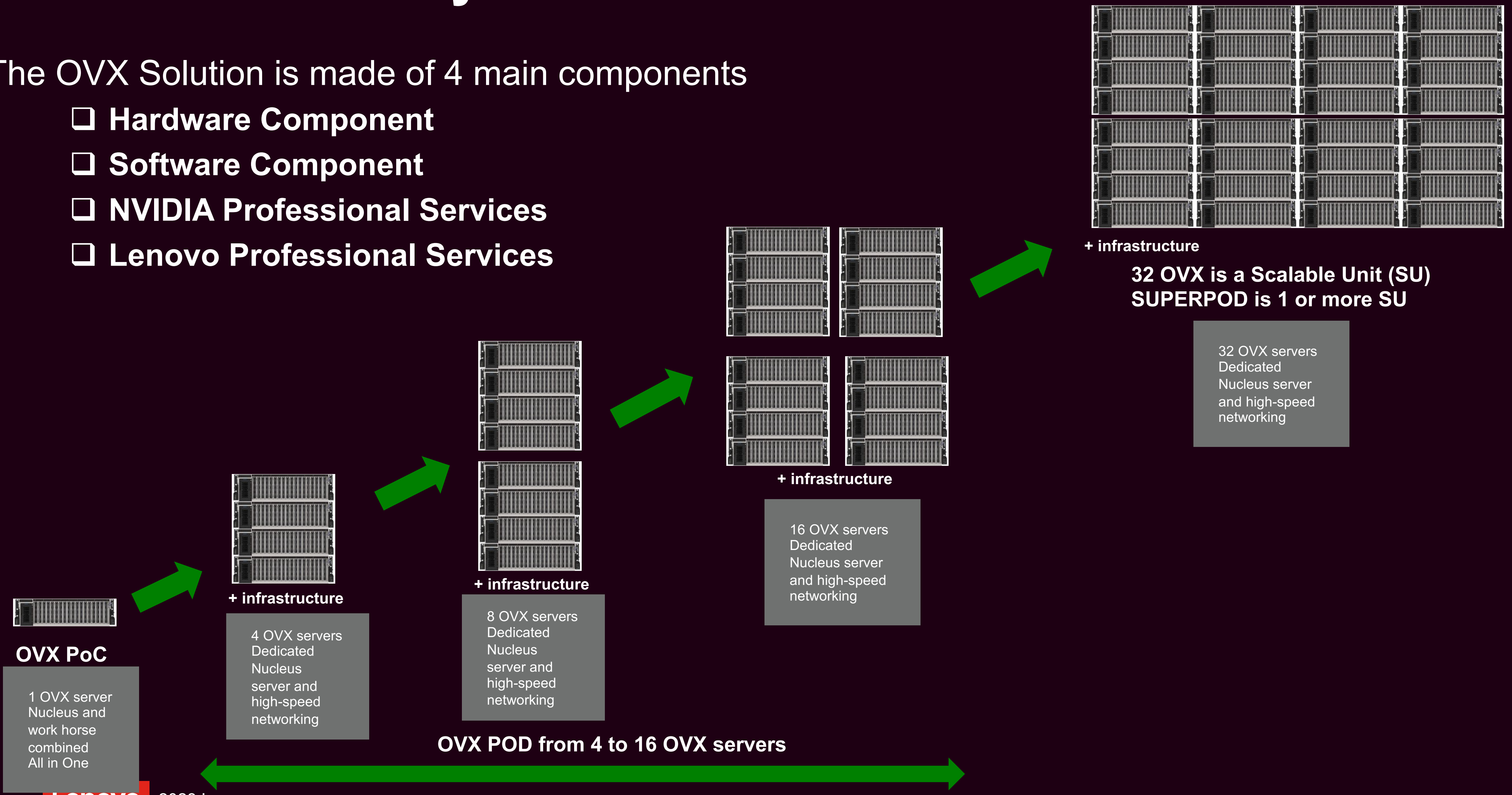
Evolution of AI and Metaverse



Your OVX Journey

The OVX Solution is made of 4 main components

- ❑ **Hardware Component**
- ❑ **Software Component**
- ❑ **NVIDIA Professional Services**
- ❑ **Lenovo Professional Services**

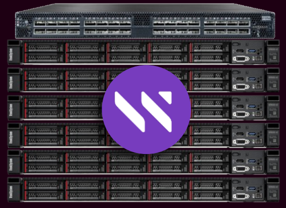


Lenovo HPC Data Management Portfolio

Lenovo DSS-G

IBM Spectrum Scale RAID

NVMe storage



WEKA

Lenovo DE series

IBM Spectrum Scale RAID

BeeGFS

Lenovo DM/DG series

ONTAP

Lenovo Ceph solutions

ceph

Red Hat

Ubuntu

Lenovo ThinkSystem Enterprise Storage Array Portfolio

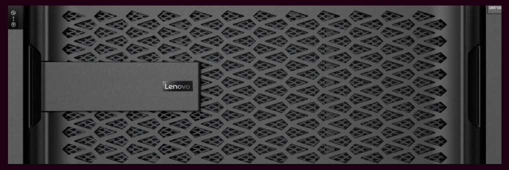
Efficient, secure solutions to maximize performance and value for AI and data intensive workloads

ThinkSystem DE Series

- Simplified data management
- Entry to High performance block
- Flash and Hybrid models
- Easy to configure, manage, and scale.

ThinkSystem DG Series

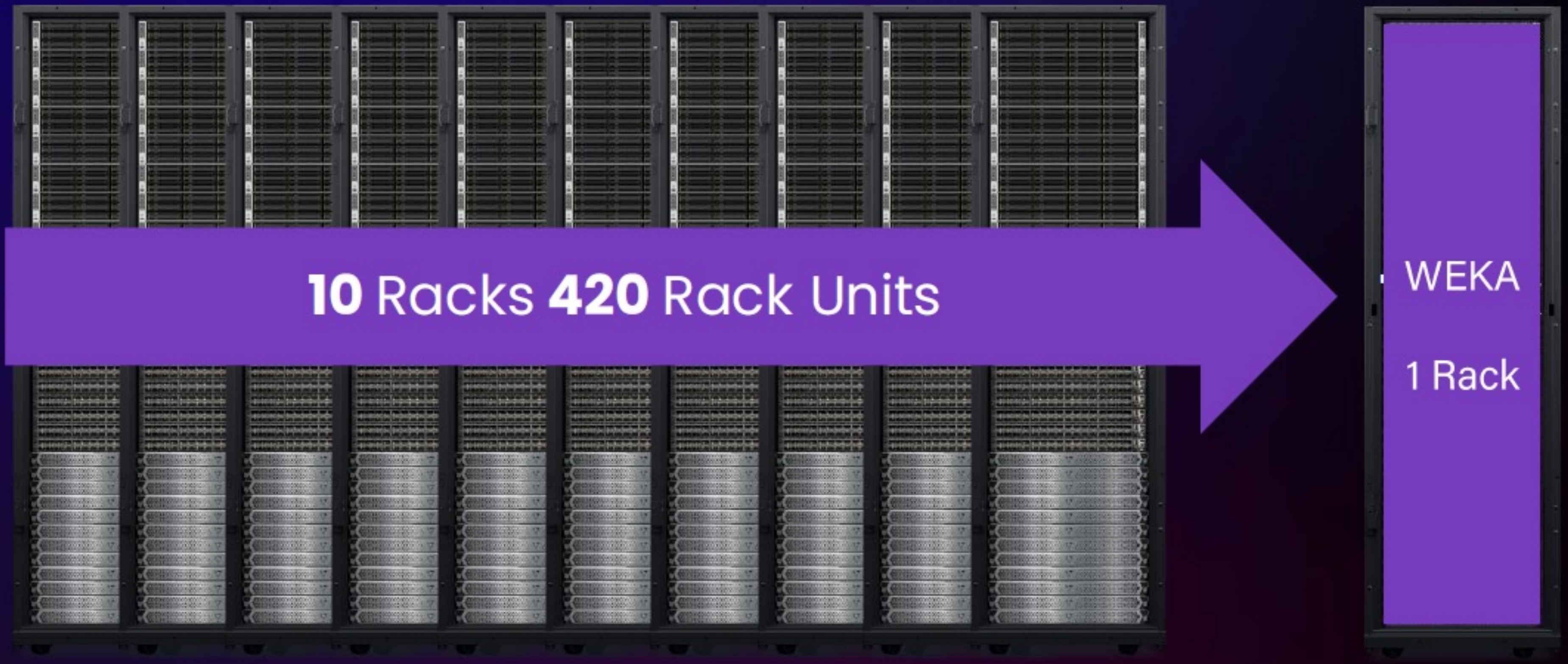
- Efficient all flash data consolidation
- Unified File/Block/Object
- All Flash at HDD economics
- Secure hybrid cloud management
- Integrated ransomware protection



ThinkSystem DM Series

- Leadership flash performance
- Unified File/Block/Object
- Flash and hybrid models to optimize performance and scale
- Secure hybrid cloud management
- Integrated ransomware protection

Sustainability: Write Performance Efficiency



1.0 TB/s
W BW

28.8M
IOPS

327 kW
draw

42 Rack Units
1.0 TB/s W BW
90M IOPS
25.4 kW draw

Same
Bandwidth

More
IOPS

10x Less
Rack
Space

12X Less
Power
Draw