

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

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

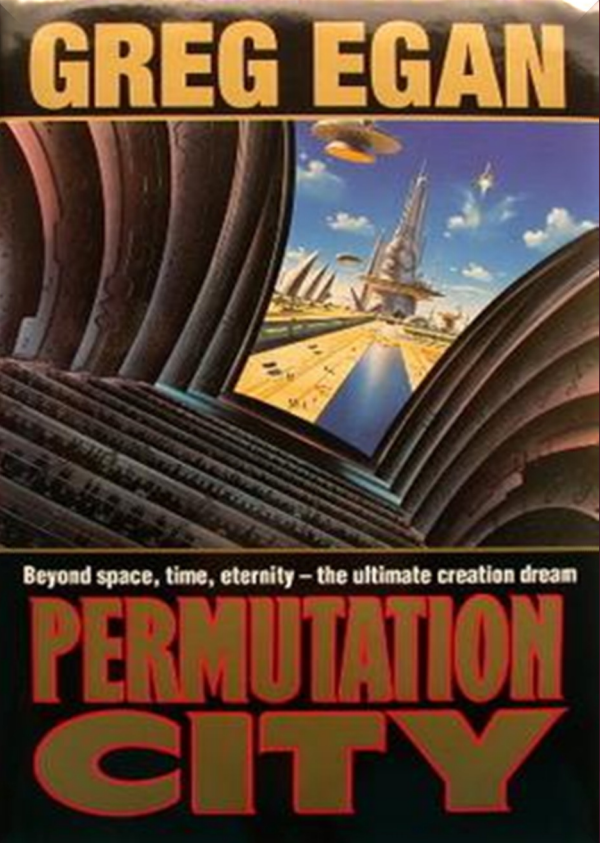
- PhD in Neuroscience and Neurophysiology
- Researcher, Lecturer, Reviewer and Associate Editor in neuroscience and neurophysiology
- Extensive professional experience in Immersive technology applied to pre-clinical research and M&E



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 from 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. MARAH LOCKMART/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.

DECEMBER 2009—JANUARY 2010 / BOEING FRONTIERS

DECEMBER 2009—JANUARY 2010 / BOEING FRONTIERS

From Digital Twin to Industrial Metaverse

Industrial Metaverse

Digital Twin



Immersive DT



Whole-System DT





“

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.**



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.**



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



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.



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

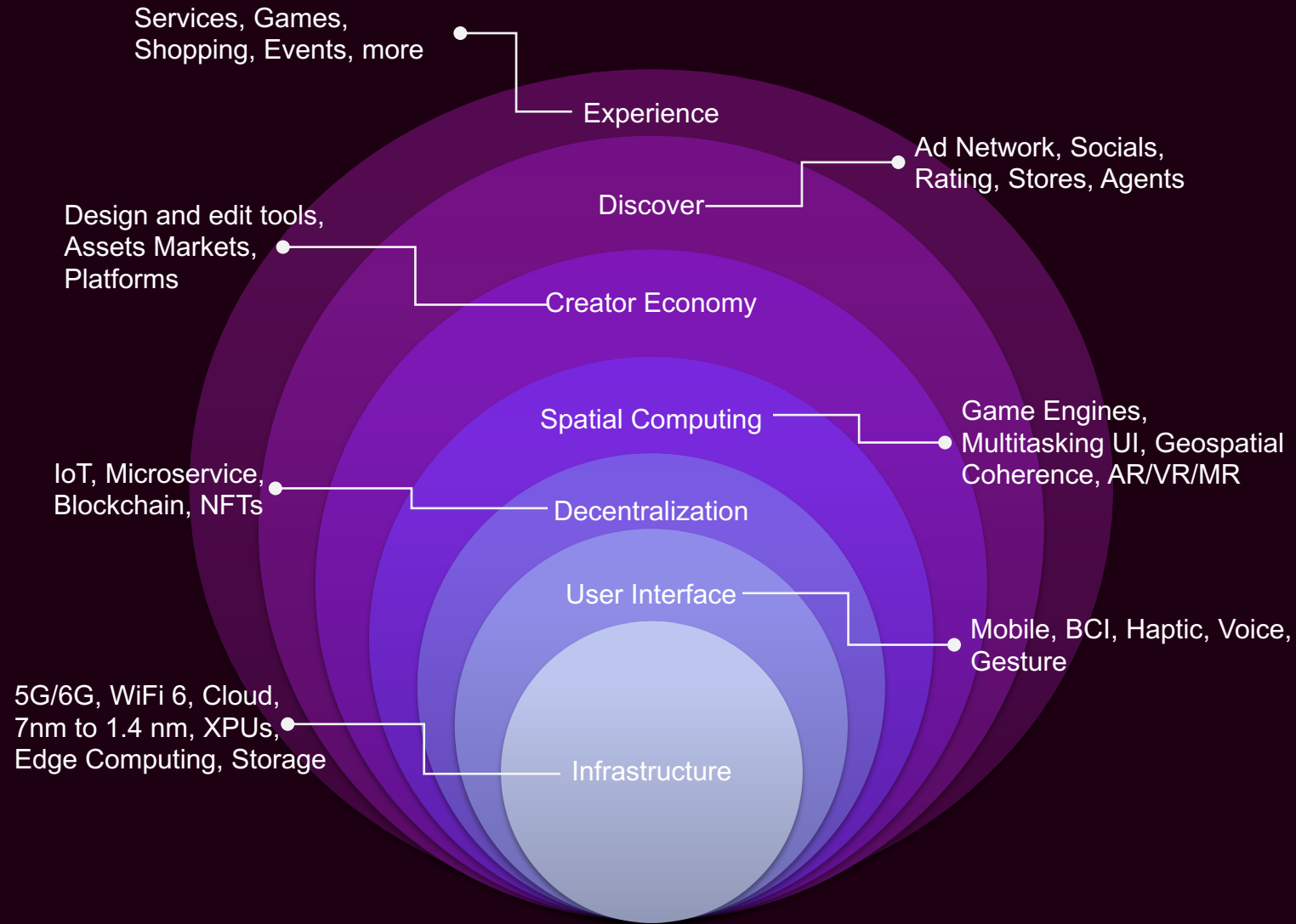


“

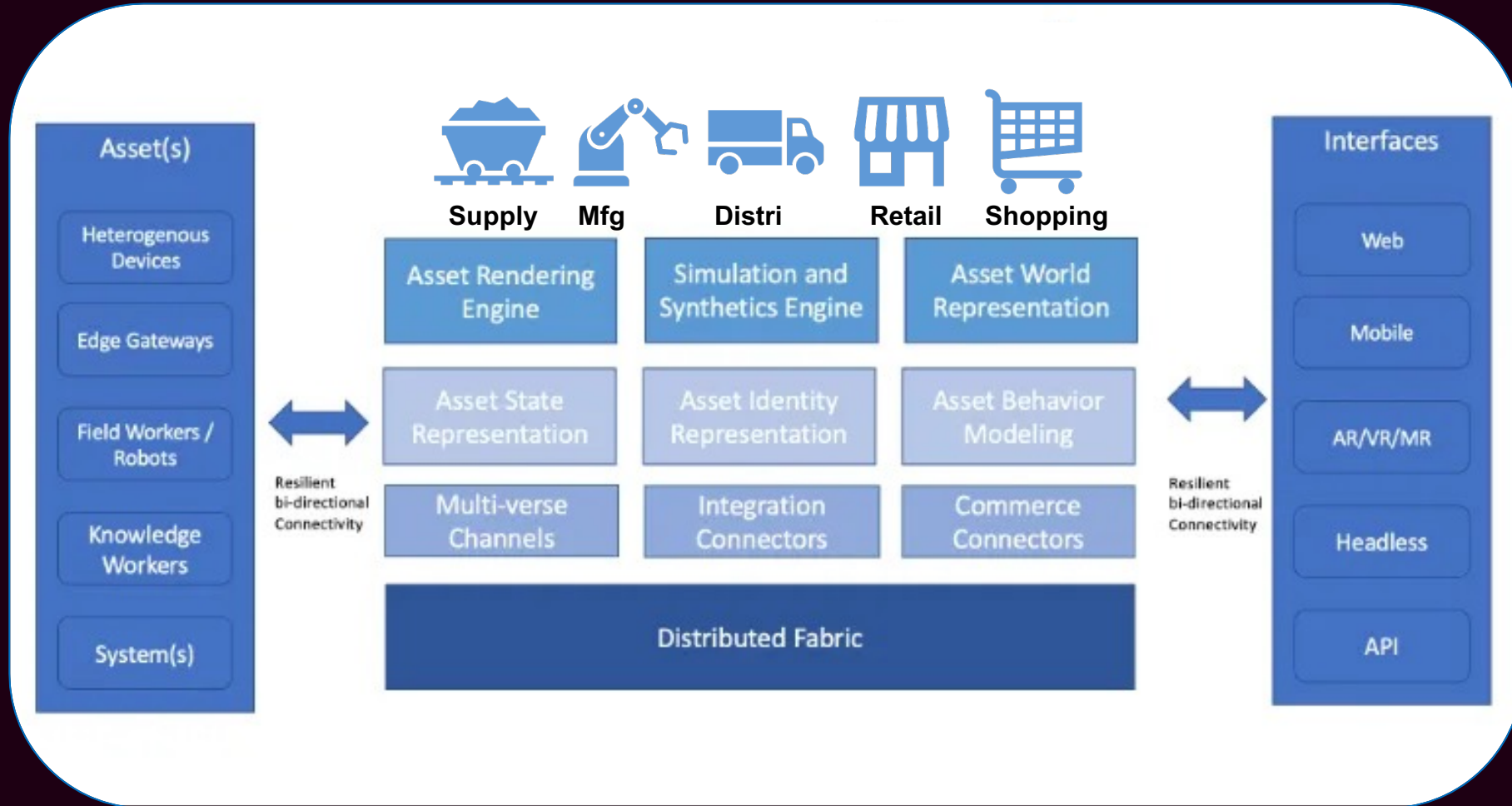
Connected whole-system digital twin with functionalities to interact with the real system in its environment, allowing decision makers to better understand the past and forecast the future.”

Arthur D. Little

Anatomy of the Metaverse



Metaverse System Model



“

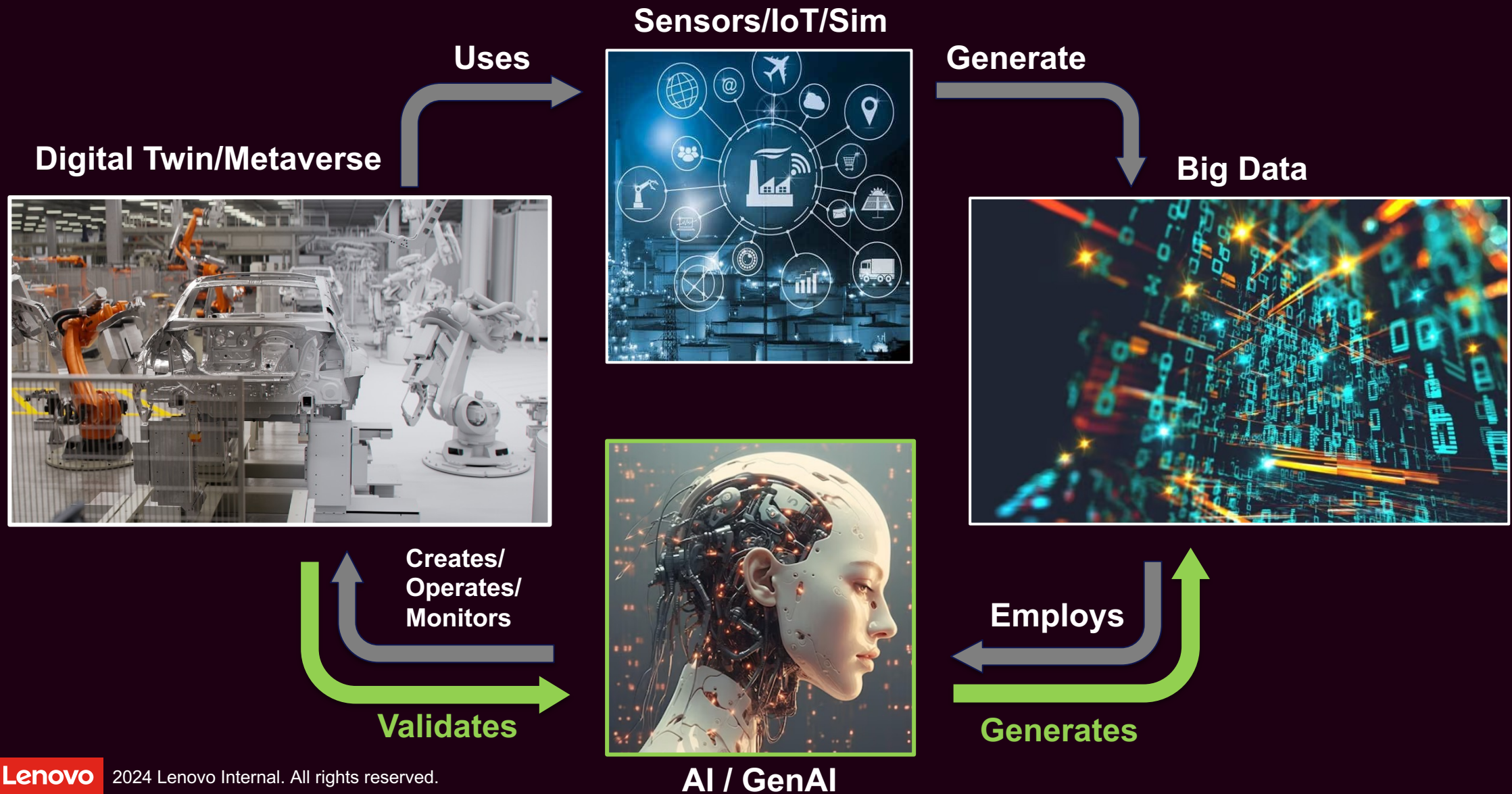
The natural habitat of AI is in the virtual world.”

Dr. Michael Grieves

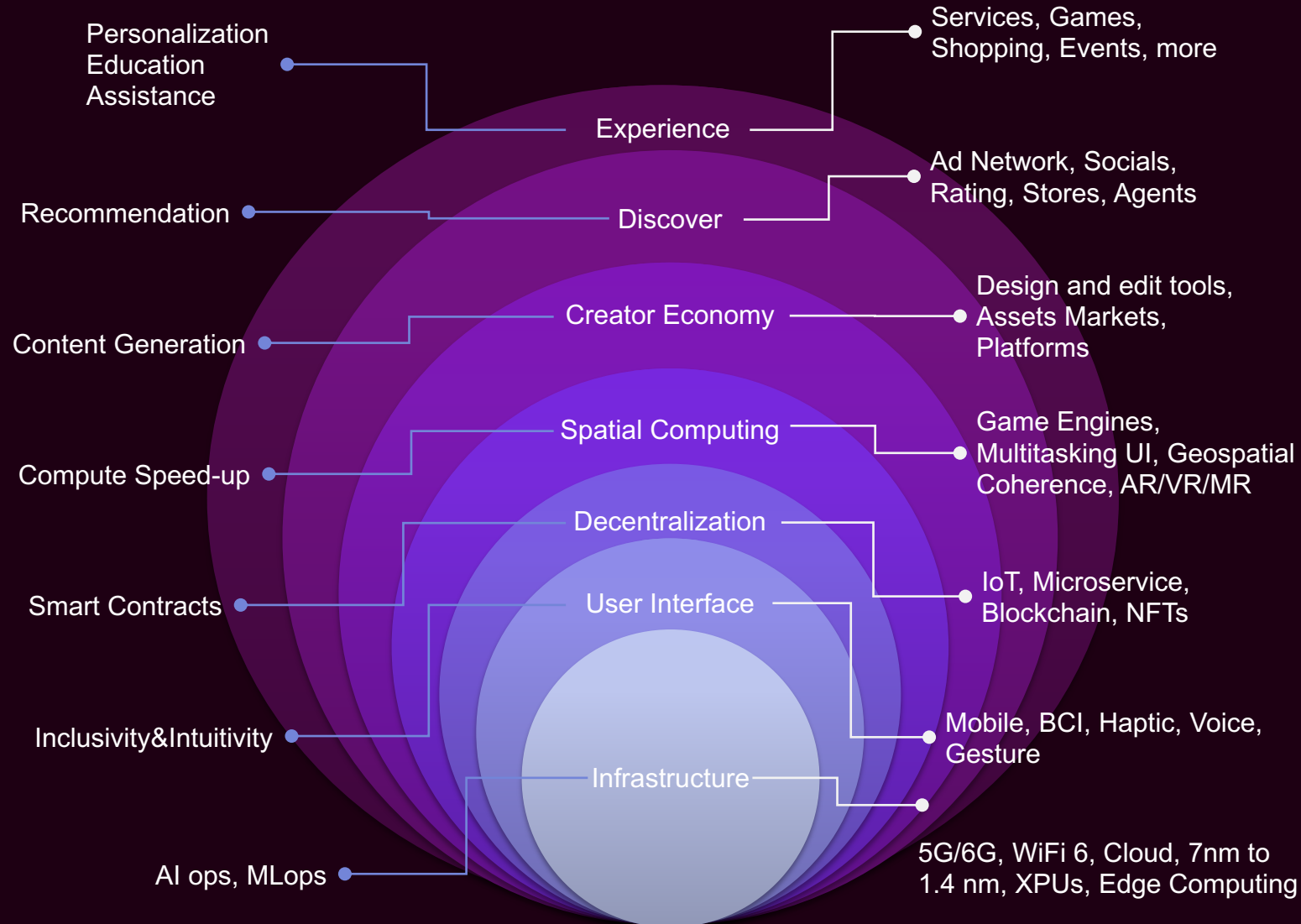
- Intelligent digital twins and the development and management of complex systems -



The Intertwined Nature of Metaverse and AI



AI value for the Metaverse



How Today's AI is Shaping Tomorrow's Possibilities

3D Modeling & Visualization



Decentralized Computing



Network Optimization



Confidential AI Solutions



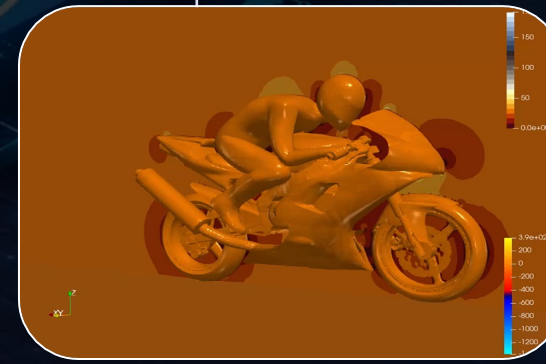
INDUSTRIAL METAVERSE



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



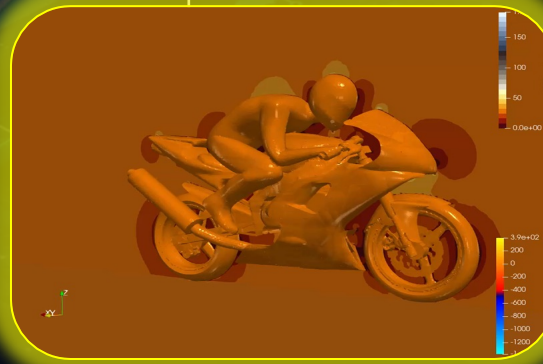
INDUSTRIAL METAVERSE



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 Tatoni, Angjoo Kanazawa
UC Berkeley

Figure 1. NeRF from one or few images. We present pixelNeRF, a learning framework that predicts a Neural Radiance Field (NeRF) representation from a single (top) or few posed images (bottom). PixelNeRF can be trained on a set of multi-view images, allowing it to generate plausible novel view synthesis from very few input images without test-time optimization (bottom left). In contrast, NeRF has no generalization capabilities and performs poorly when only three input views are available (bottom right).

Abstract

We propose pixelNeRF, a learning framework that predicts a continuous neural scene representation conditioned on one or few input images. The existing approach for constructing neural radiance fields [27] involves optimizing the representation to every scene independently, requiring many calibrated views and significant compute time. We take a step towards resolving these shortcomings by introducing an architecture that conditions a NeRF on image inputs in a fully convolutional manner. This allows the network to be trained across multiple scenes to learn a scene prior, enabling it to perform novel view synthesis in a feed-forward manner from a sparse set of views (as few as one). Leveraging the volume rendering approach of NeRF, our model can be trained directly from images with no explicit 3D supervision. We conduct extensive experiments on ShapeNet benchmarks for single image novel view synthesis tasks with held-out objects as well as entire unseen categories. We further demonstrate the feasibility of pixelNeRF by demonstrating it on multi-object ShapeNet scenes and real scenes from the DTU dataset. In all cases, pixelNeRF outperforms current state-of-the-art baselines for novel view synthesis and single image 3D reconstruction. For the video and code, please visit the project website: <https://alexeyuy.net/pixelnerf/>.

arXiv:2012.02190v3 [cs.CV] 30 May 2021

Instant Ngg (2022)

Instant Neural Graphics Primitives with a Multiresolution Hash Encoding

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

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

Figure 1. We demonstrate instant training of neural graphics primitives on a single GPU for multiple tasks. In clustered image we represent a gigapixel image by a neural network. NeRF learns a signed distance function in 3D space whose zero level set represents a 2D surface. Neural radiance caching (NRC) [Müller et al. 2022] employs a neural network that is trained in real-time to cache costly lighting calculations. Lastly, NeRF [Mildenhall et al. 2020] uses 2D images and their camera poses to reconstruct a volumetric radiance-and-density field that is visualized using ray marching. In all tasks, our encoding and its efficient implementation provide clear benefits: rapid training, high quality, and simplicity. Our encoding is task-agnostic: we use the same implementation and hyperparameters across all tasks and only vary the hash table size which trades off quality and performance. Tokyo gigapixel photograph ©Toson Dobson (CC BY-NC-ND 2.0), Lego bulldozer 3D model ©Hazard Dales (CC BY-NC 2.0)

Abstract

Neural graphics primitives, parametrized by fully connected neural networks, can be easily trained and evaluated. We reduce this cost with a versatile new input encoding that permits the use of a smaller network without sacrificing quality. This significantly reduces the number of floating point and memory access operations: a small neural network is augmented by a multiresolution hash table of trainable feature vectors whose values are optimized through stochastic gradient descent. The multiresolution structure allows the network to disentangle hash collisions, enabling for a simple author address: Thomas Müller, NVIDIA, Zürich, Switzerland, cmu@nvidia.com, Alex Evans, NVIDIA, London, United Kingdom, alex@nvidia.com, Christoph Sched, NVIDIA, Zürich, Switzerland, sched@nvidia.com, Alexander Keller, NVIDIA, Berlin, Germany, akeller@nvidia.com

CCS Concepts: • Computing methodologies → Massively parallel architectures; • Vector • streaming algorithms; • Neural networks

Additional Key Words and Phrases: Image Synthesis, Neural Networks, Encodings, Hashing, GPUs, Parallel Computation, Function Approximation.

ACM Reference Format: Thomas Müller, Alex Evans, Christoph Sched, and Alexander Keller. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. ACM Trans. Graph., 41, 4, Article 102 (July 2022), 15 pages. <https://doi.org/10.1145/3528223.3530127>

© 2022 Copyright held by the owner/authors. Publication rights licensed to ACM. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in ACM Transactions on Graphics, August 2022, 1140-1154.

ACM Trans. Graph., Vol. 41, No. 4, Article 102. Publication date: July 2022.

Neuralangelo (2023)

Neuralangelo: High-Fidelity Neural Surface Reconstruction

Zhaoshuo Li^{1,2}, Thomas Müller¹, Alex Evans¹, Russell H. Taylor², Mathias Unberath², Ming-Yu Liu¹, Chen-Hsuan Lin¹

¹NVIDIA Research ²Johns Hopkins University
<https://research.nvidia.com/labs/diz/neuralangelo>

Figure 1. We present Neuralangelo, a framework for high-fidelity 3D surface reconstruction from RGB images using neural volume rendering, even without auxiliary data such as segmentation or depth. Shown in the figure is an extracted 3D mesh of a courthouse.

Abstract

Neural surface reconstruction has been shown to be powerful for recovering dense 3D surfaces via image-based neural rendering. However, current methods struggle to recover detailed structures of real-world scenes. To address this, we present Neuralangelo, which combines the representation power of multi-resolution 3D hash grids with neural surface rendering. Two key ingredients enable our approach: (1) numerical gradients for computing higher-order derivatives as a smoothing operation and (2) coarse-to-fine optimization on the hash grids controlling different levels of details. Even without auxiliary inputs such as depth, Neuralangelo can effectively recover dense 3D surface structures from multi-view images with fidelity significantly surpassing previous methods, enabling detailed large-scale scene reconstruction from RGB video captures.

1. Introduction

3D surface reconstruction aims to recover dense geometric scene structures from multiple images observed at different viewpoints [9]. The recovered surfaces provide structural information useful for many downstream applications, such as 3D asset generation for augmented/virtual/mixed reality or environment mapping for autonomous navigation of robotics. Photogrammetric surface reconstruction using a monocular RGB camera is of particular interest, as it equips users with the capability of casually creating digital twins of the real world using ubiquitous mobile devices.

Classically, multi-view stereo algorithms [6, 16, 33, 39] had been the method of choice for sparse 3D reconstruction. An inherent drawback of these algorithms, however, is their inability to handle ambiguous observations, e.g. regions with large areas of homogeneous colors, repetitive texture

arXiv:2306.03092v2 [cs.CV] 12 Jun 2023

Magic3D (2023)

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

Chen-Hsuan Lin¹, Jun Gao², Luming Tang², Towaki Takikawa³, Xiaohui Zeng¹, Xun Huang, Karsten Kreis, Sanja Fidler¹, Ming-Yu Liu¹, Tsung-Yi Lin¹

NVIDIA Corporation
<https://research.nvidia.com/labs/diz/magic3d>

Abstract

DreamFusion [13] has recently demonstrated the utility of a pre-trained text-to-image diffusion model to optimize Neural Radiance Fields (NeRF) [25], achieving remarkable text-to-3D synthesis results. However, the method has two inherent limitations: (a) extremely slow optimization of NeRF and (b) low-resolution image space supervision on NeRF, leading to low-quality 3D models with a long processing time. In this paper, we address these limitations by utilizing a two-stage optimization framework. First, we obtain a coarse model using a low-resolution diffusion prior and an encoder with a sparse 3D hash grid structure. Using the coarse representation as the initialization, we further optimize a textured 3D mesh model with an efficient differentiable renderer interacting with a high-resolution latent diffusion model. Our method, dubbed Magic3D, can create high quality 3D mesh models in 40 minutes, which is 2x faster than DreamFusion (reportedly taking 1.5 hours on average), while also achieving higher resolution. User studies show 61.7% users to prefer our approach over DreamFusion. Together with the image-conditioned generation capabilities, we provide users with new ways to control 3D synthesis, opening up new avenues to various creative applications.

1. Introduction

3D digital content has been in high demand for a variety of applications, including gaming, entertainment, architecture, and robotics simulation. It is slowly finding its way into virtually every possible domain: retail, online conferencing, virtual social presence, education, etc. However, creating professional 3D content is not for anyone — it requires immense artistic and aesthetic training with 3D modeling expertise. Developing these skill sets takes a significant amount of time and effort. Augmenting 3D content creation with natural language could considerably help democratize 3D content creation for novices and turbocharge expert artists.

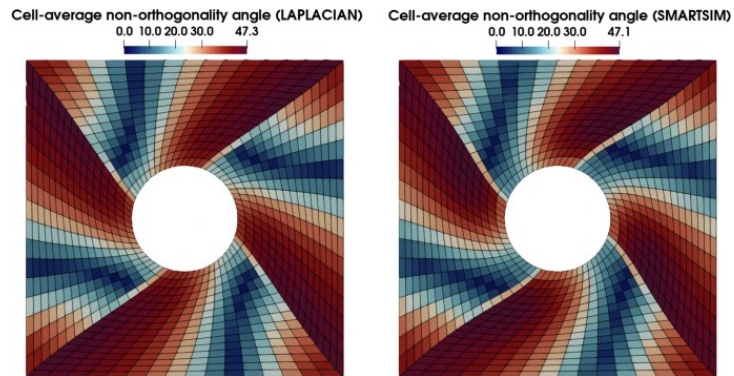
Recently, DreamFusion [13] demonstrated its remarkable ability for text-conditioned 3D content generation by utilizing a pre-trained text-to-image diffusion model [10] that generates images as a strong image prior. The diffusion model acts as a critic to optimize the underlying 3D representation. The optimization process ensures that rendered images from a 3D model, represented by Neural Radiance Fields (NeRF) [25], match the distribution of photorealistic images across different viewpoints, given the input text prompt. Since the supervision signal in DreamFusion operates on diverse low-resolution images (64 × 64), DreamFusion cannot synthesize high-frequency 3D geometric and texture details. Due to the use of inefficient MLP architectures for the NeRF representation, practical high-resolution synthesis may not even be possible as the required memory footprint and the computation budget grows quickly with the resolution. Even at a resolution of 64 × 64, optimization times are in hours (1.5 hours per prompt on average using TPUv4).

In this paper, we present a method that can synthesize highly detailed 3D models from text prompts within a reduced computation time. Specifically, we propose a coarse-

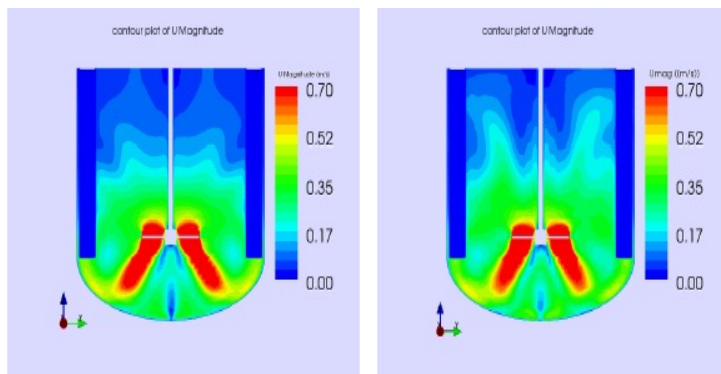
[†] equal contribution.

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

Towards Real-Time Physically Accurate Simulations



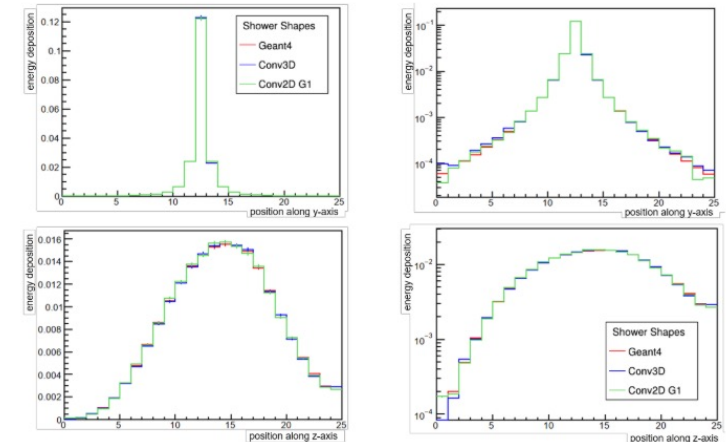
Approximating mesh-motion Laplacian mesh motion solver in OpenFOAM with MLP



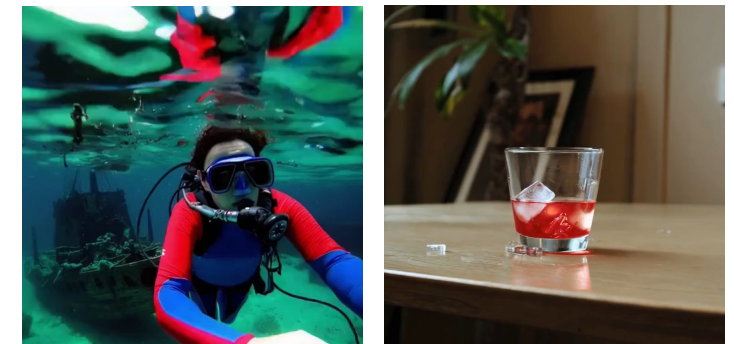
Using CNN to Solve Euler-Lagrange, Momentum Transfer, and Incompressible RANS Equations

Gen AI

AI / ML



Simulating high energy physics calorimeter detector outputs with 2D GAN



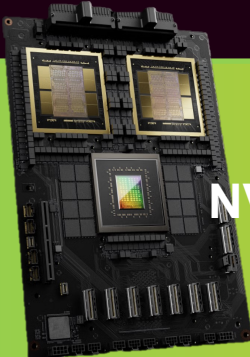
Video generation models as general purpose simulators of the physical world?

Tensors Reshape Compute Architectures

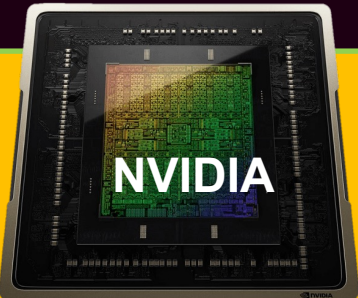
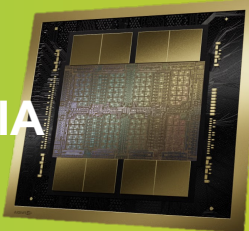
HPC

Viz / Render

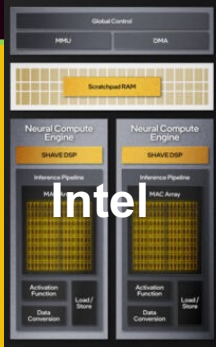
Edge/Client



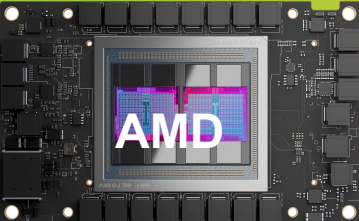
NVIDIA



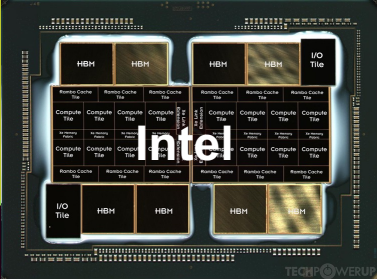
NVIDIA



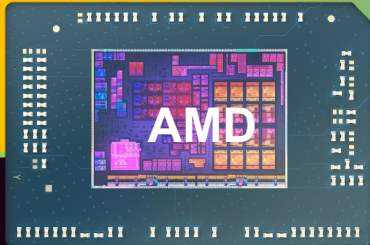
Intel



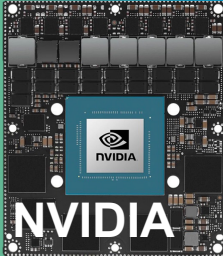
AMD



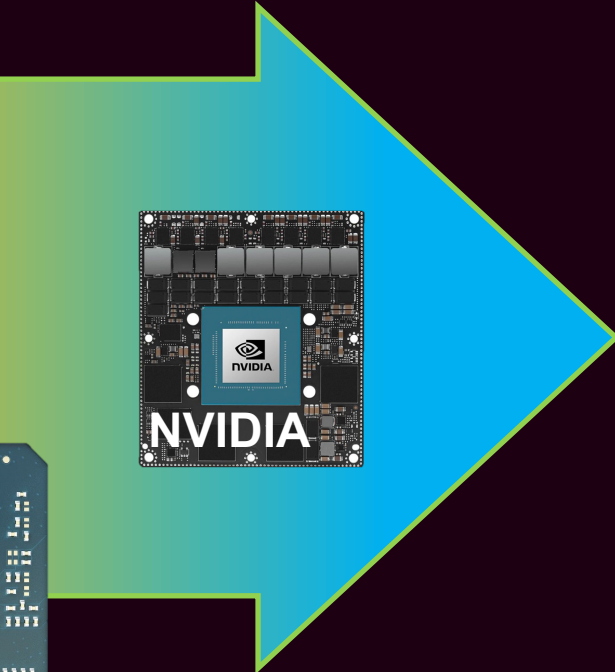
Intel



AMD



NVIDIA



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



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



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



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



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



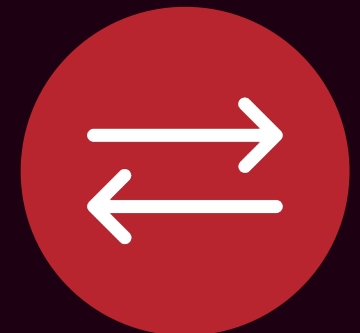
Challenges ahead



Scalability
& Energy Efficiency



Security &
Privacy



Interoperability
& Standards



Compute & Storage
Optimization

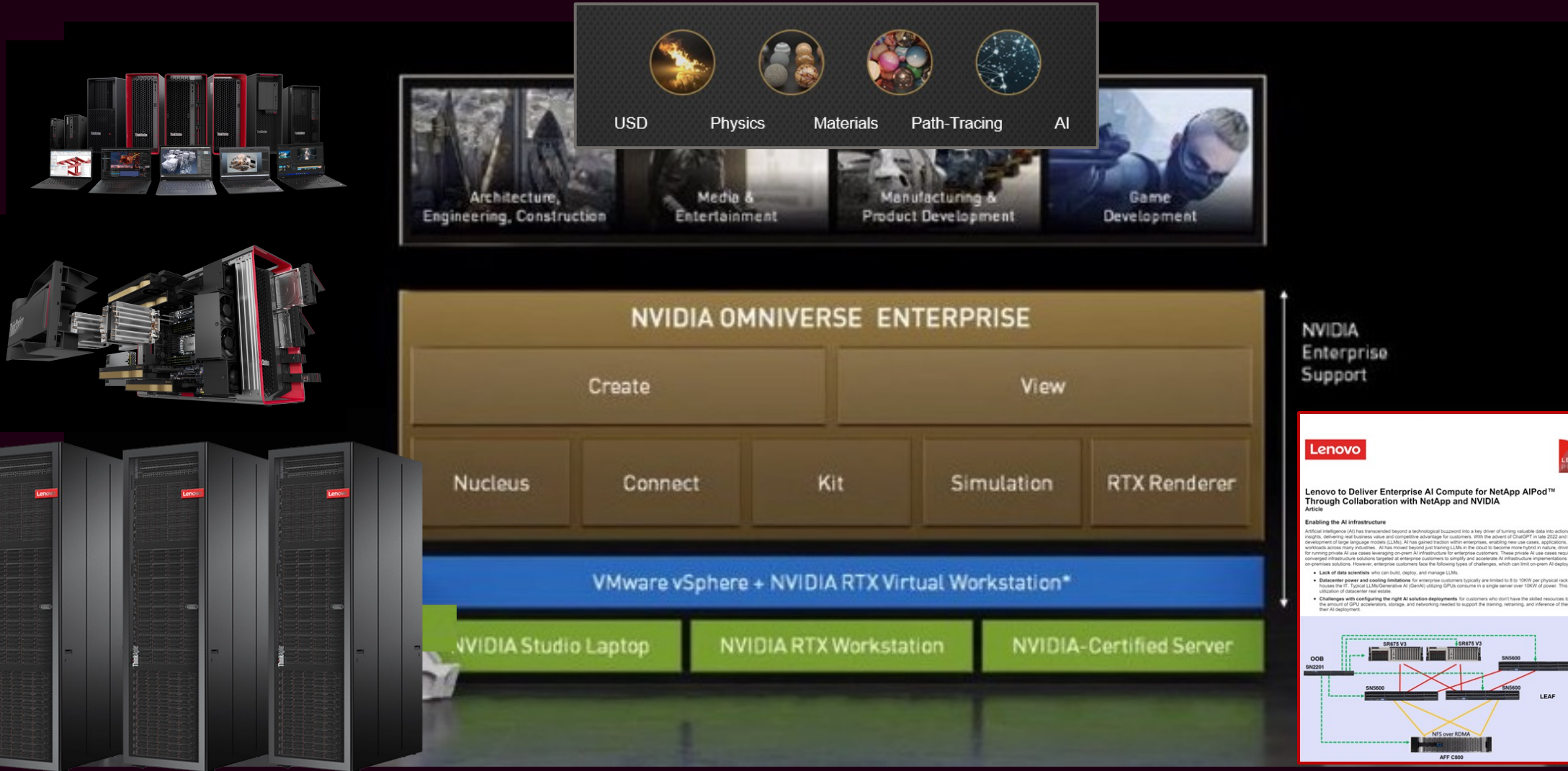


Ethic & Regulations

Lenovo E2E – OVX Infrastructure Solutions

Through Collaboration with NetApp and NVIDIA

LENOVO - E2E – OVX INFRASTRUCTURE



NVIDIA
Enterprise
Support

Lenovo

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

Article

Enabling the AI Infrastructure

Artificial intelligence (AI) has transcended beyond a technological buzzword into a key driver of turning valuable data into actionable insights, delivering real business value and competitive advantage for customers. With the advent of ChatGPT in late 2022 and the ensuing development of large language models (LLMs), AI has gained traction within enterprises, enabling new use cases, applications, and workflows across many industries. AI has moved beyond just training LLMs on the cloud to become more tightly integrated, driving the need for running private AI use cases leveraging on-prem AI infrastructure for enterprise customers. These private AI use cases require new converged infrastructure solutions targeted at enterprise customers to simplify and accelerate AI infrastructure implementations at scale for on-premises solutions. However, enterprise customers face the following types of challenges, which can limit on-prem AI deployments:

- Lack of data scientists who can build, deploy, and manage LLMs.
- Datacenter power and cooling limitations: for enterprise customers typically are limited to 8 to 10KW per physical rack space that houses the IT. Typical LLMs/Generative AI (GenAI) utilizing GPUs consume in a single server over 10KW of power. This limits the utilization of datacenter real estate.
- Challenges with configuring the right AI solution deployments for customers who don't have the skilled resources to determine the impact of GPU accelerators, storage, and networking needed to support the training, reasoning, and inference of their data for their AI deployment.

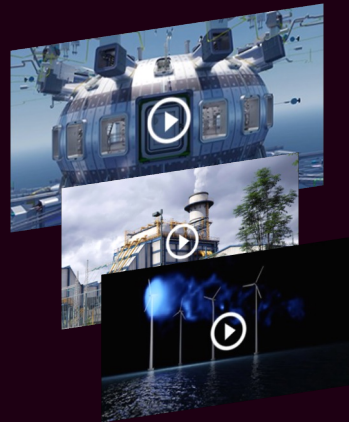
The benefits of MV tech application embrace all industries.

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



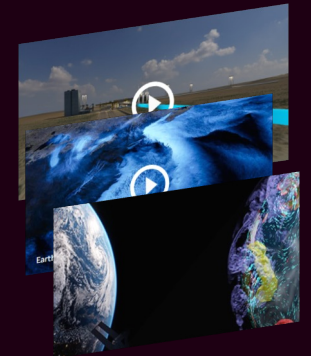
- Transforming Telco Network Planning and Operations
- Simulating and Optimizing Autonomous Railway Networks
- Testing and Optimizing 5G Deployment

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.

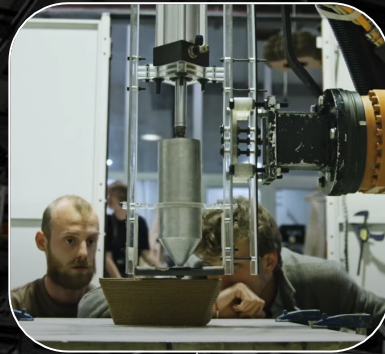
Towards an Industrial Metaverse

A glimpse of the transformative power of Metaverse and AI

Process Optimization



Additive Manufacturing



Design



Synthetic Data



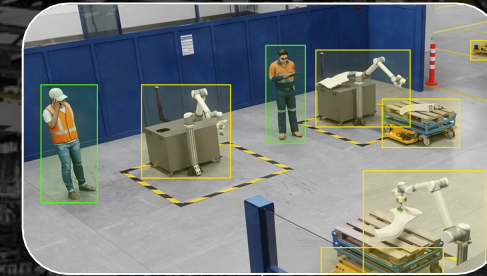
Simulations



Infrastructure



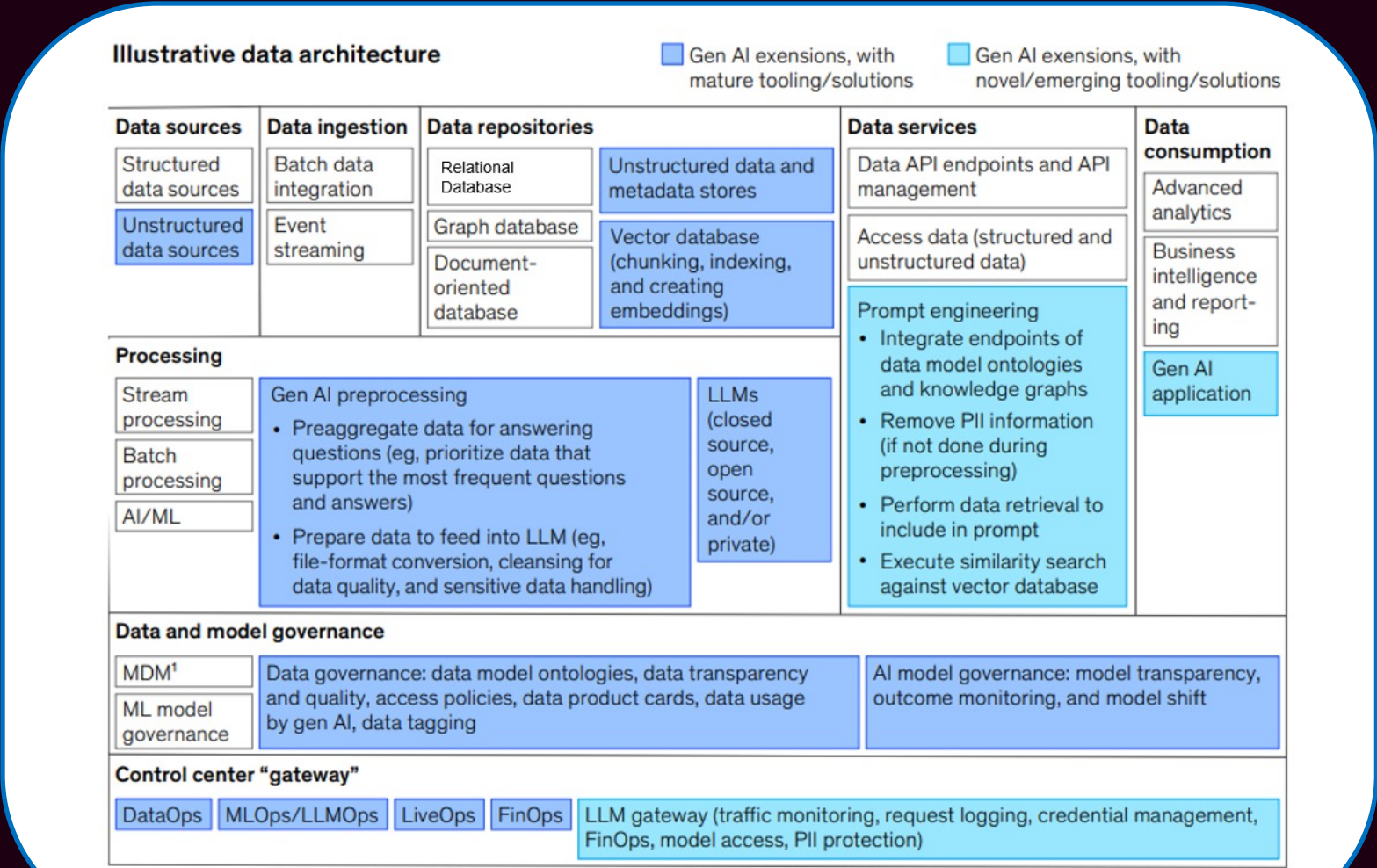
Proactive Safety



Augmented Assistance



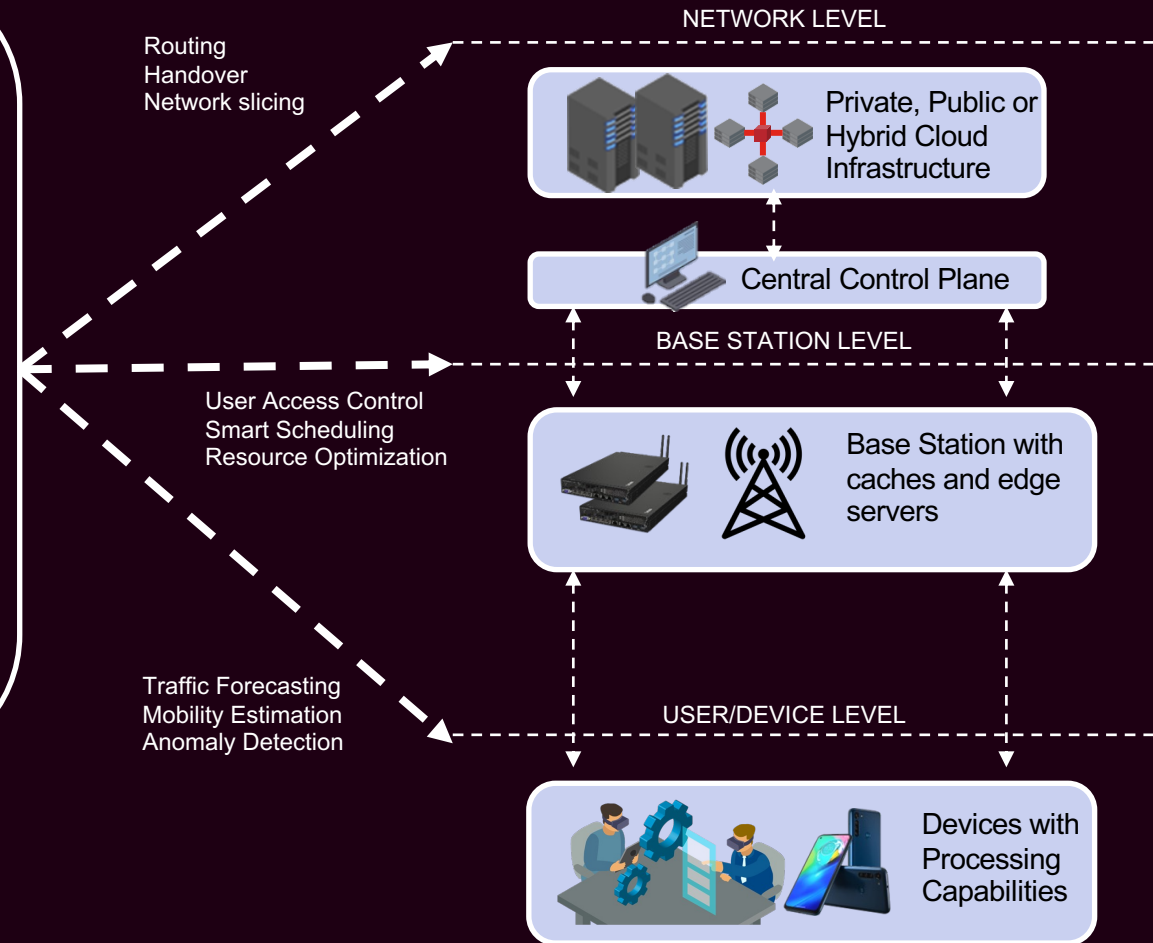
Data Architecture Upgrades Needed for AI, GenAI and Industrial Metaverse



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.



Evolution of AI and Metaverse

From Rosenblatt's Perceptron to DL Revolution



1957 - 2006

Virtual Reality (VR) and Augmented Reality Headsets Commercialization



2014

ResNets and NLP Breakthroughs



2015

Virtual Travis Scott Concert



2020

GPT-3 and Self-Supervised Learning



2020

AI-Powered Immersive Technologies



2023

1992 - 2003



From Stephenson's Snowcrash to Grieves' Digital Twin

2012



AlexNet and Variational Autoencoders

2018



Remote Collaboration and Telepresence Solutions

2017



Transformer Architecture and Language Models

2022



Blockchain-Based Supply Chain
Virtual real estate trading

2023



GenAI: LLMs, Stable Diffusion, Sora and more

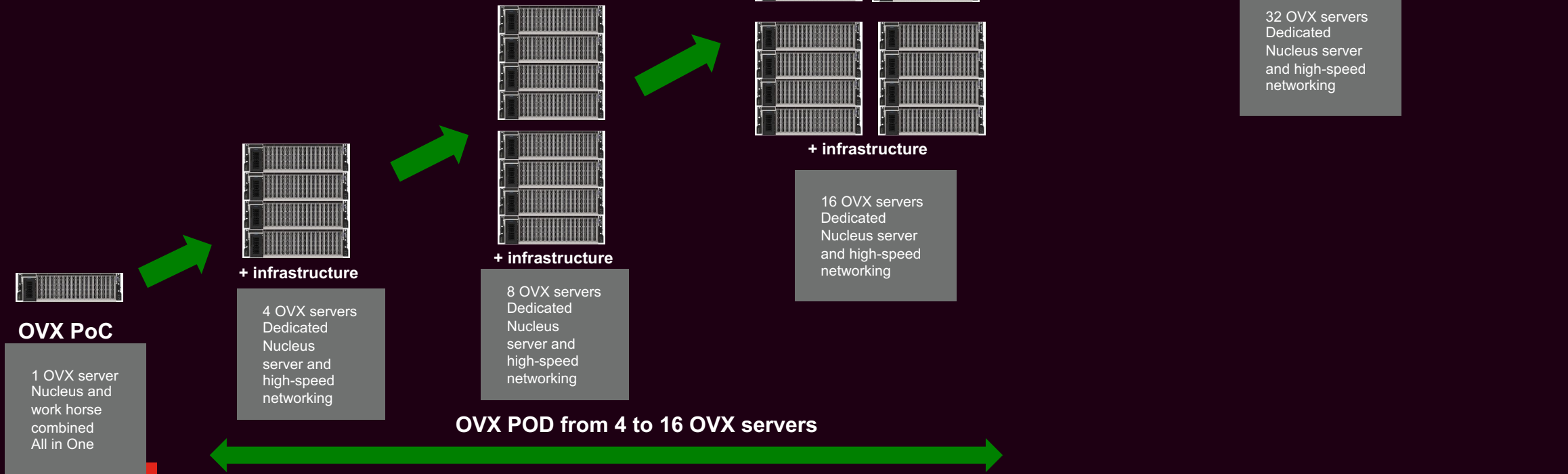


2024 Lenovo Internal. All rights reserved.

Your OVX Journey

The OVX Solution is made of 4 main components

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



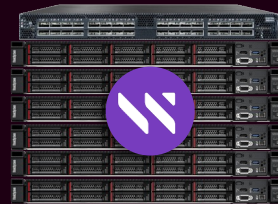
Lenovo HPC Data Management Portfolio

Lenovo DSS-G



IBM
Spectrum
Scale
RAID

NVMe storage



Lenovo DE series



IBM
Spectrum
Scale
RAID



Lenovo DM/DG series



Lenovo Ceph solutions



Lenovo ThinkSystem Enterprise Storage Array Portfolio

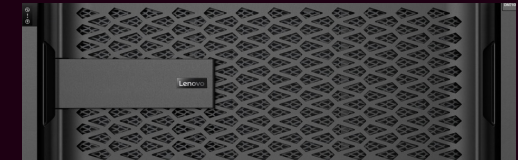
Efficient, secure solutions to maximize performance and value for AI and data intensive workloads



ThinkSystem DE Series



ThinkSystem DG Series



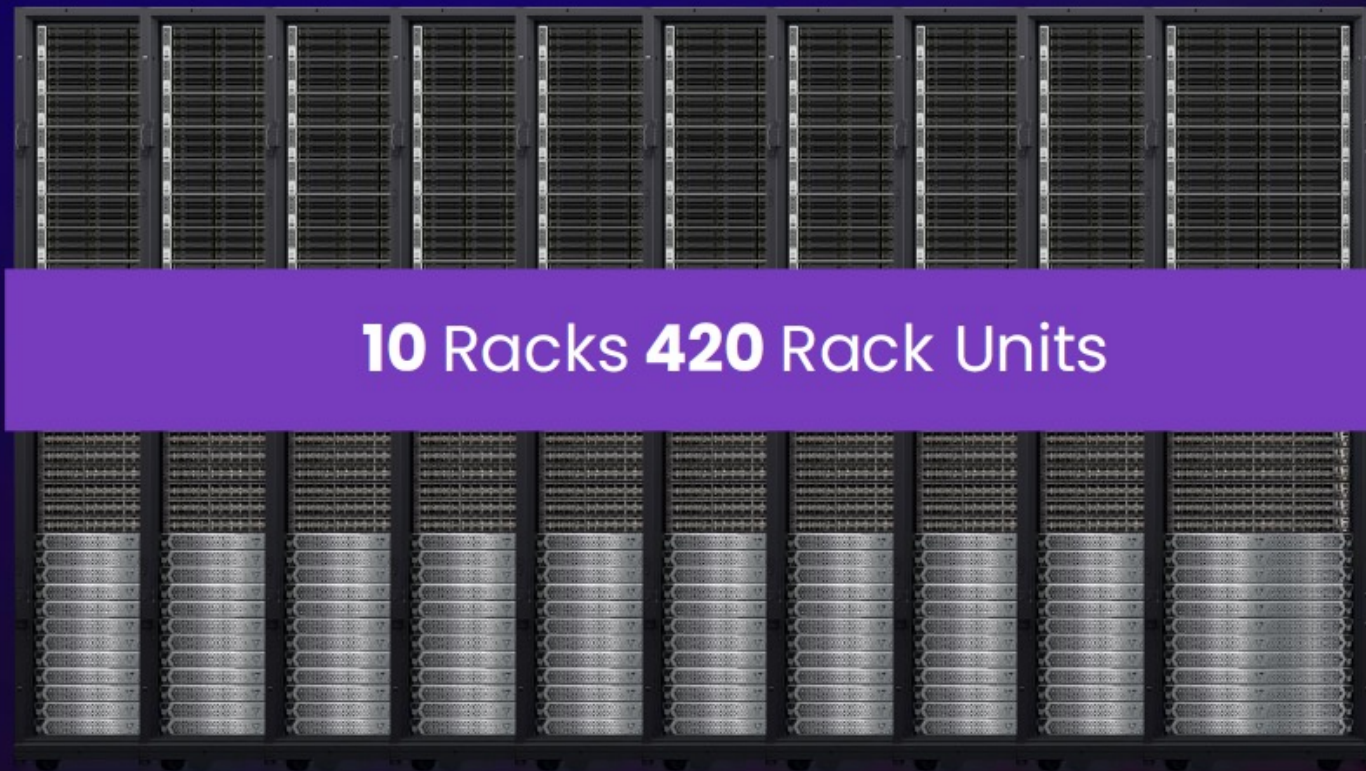
ThinkSystem DM Series

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

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

- 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

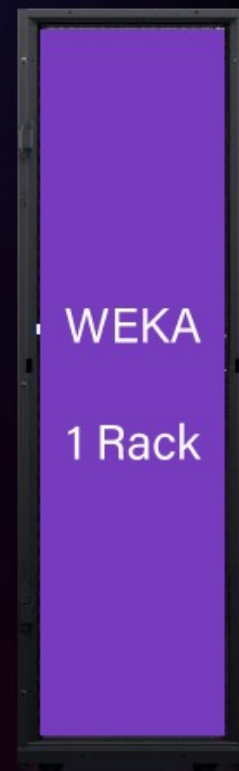


10 Racks 420 Rack Units

1.0 TB/s
W BW

28.8M
IOPS

327 kW
draw



WEKA
1 Rack

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