Nearly There

Bringing an AI experience to life

The making of Antonio “Nearly”

A use case for trusted enterprise AI at scale featuring an avatar of our President and CEO, Antonio Neri.

Explore

How we gave our data a voice

Using proven HPE hardware, software, services, and connectivity to build a cutting edge, personalized generative AI system.

Explore

Ready to have a conversation with *your* data?

What does conversational data look like in your environment?

Explore

Hewlett Packard Enterprise | HPE aruba networking | NVIDIA | Titan ML | DATA MONSTERS | PROTO

The making of Antonio "Nearly"

The inside story

Explore

Ethical review

Explore

Creating the avatar

Explore

The inside story

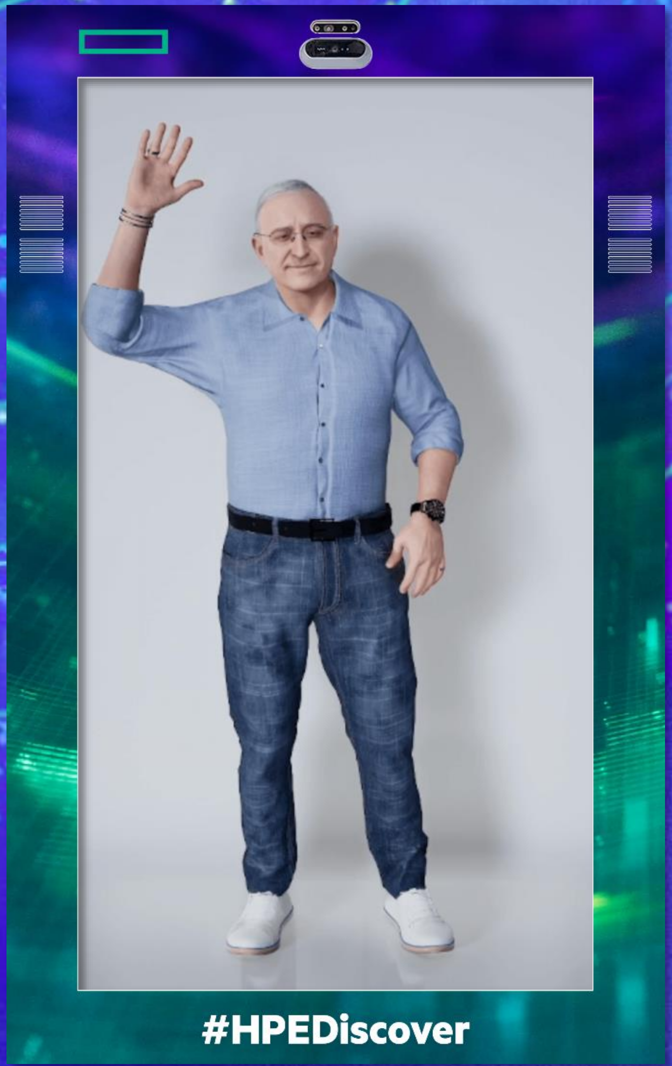
How did we get from customer visits to a customized interactive avatar?

The need

We needed a way to connect our Customers and Partners to our best story tellers and subject matter experts to discuss their business challenges and learn how HPE's portfolio of services and solutions can enable their digital transformation journey. The challenge is that our best people can't be everywhere at once. Especially our CEO, Antonio Neri.

The solution

Turn our data into an interactive Generative AI hologram, armed with the knowledge of the entire HPE portfolio of services and solutions. Built on our trusted Enterprise AI stack, backed by HPE's AI-at-Scale platform, Compute, Storage, Networking, Software and Services.



Ethical AI

Our mission is to provide solutions that advance the way people live and work. HPE seeks to use and develop responsible AI with beneficial outcomes guided by these ethical principles:

Privacy & security

Human focused

Inclusive

Responsible

Robust

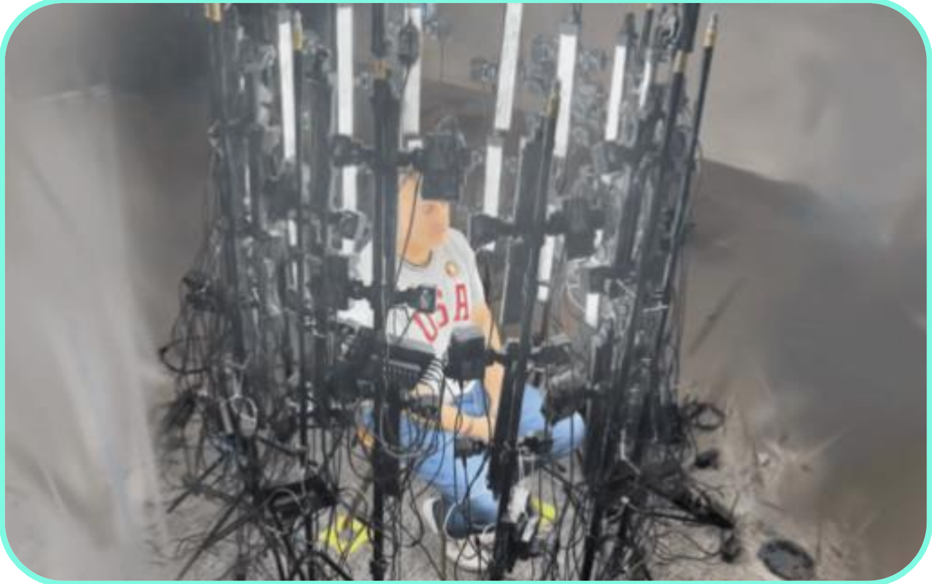
DATA MONSTERS | NVIDIA Elite Partner

Data Monsters is an AI consulting company. Their team created the photorealistic avatar of Antonio Neri and cloned his voice to bring Antonio "Nearly" to life.

The avatar was created from:

- 360° photogrammetry
- 79 cameras
- 3,592 pictures
- 147 GB of data

The data set was used to replicate Antonio Neri's facial expressions, gestures, style, and distinctive characteristics.



DATA MONSTERS | NVIDIA Elite Partner

Avatar & Voice

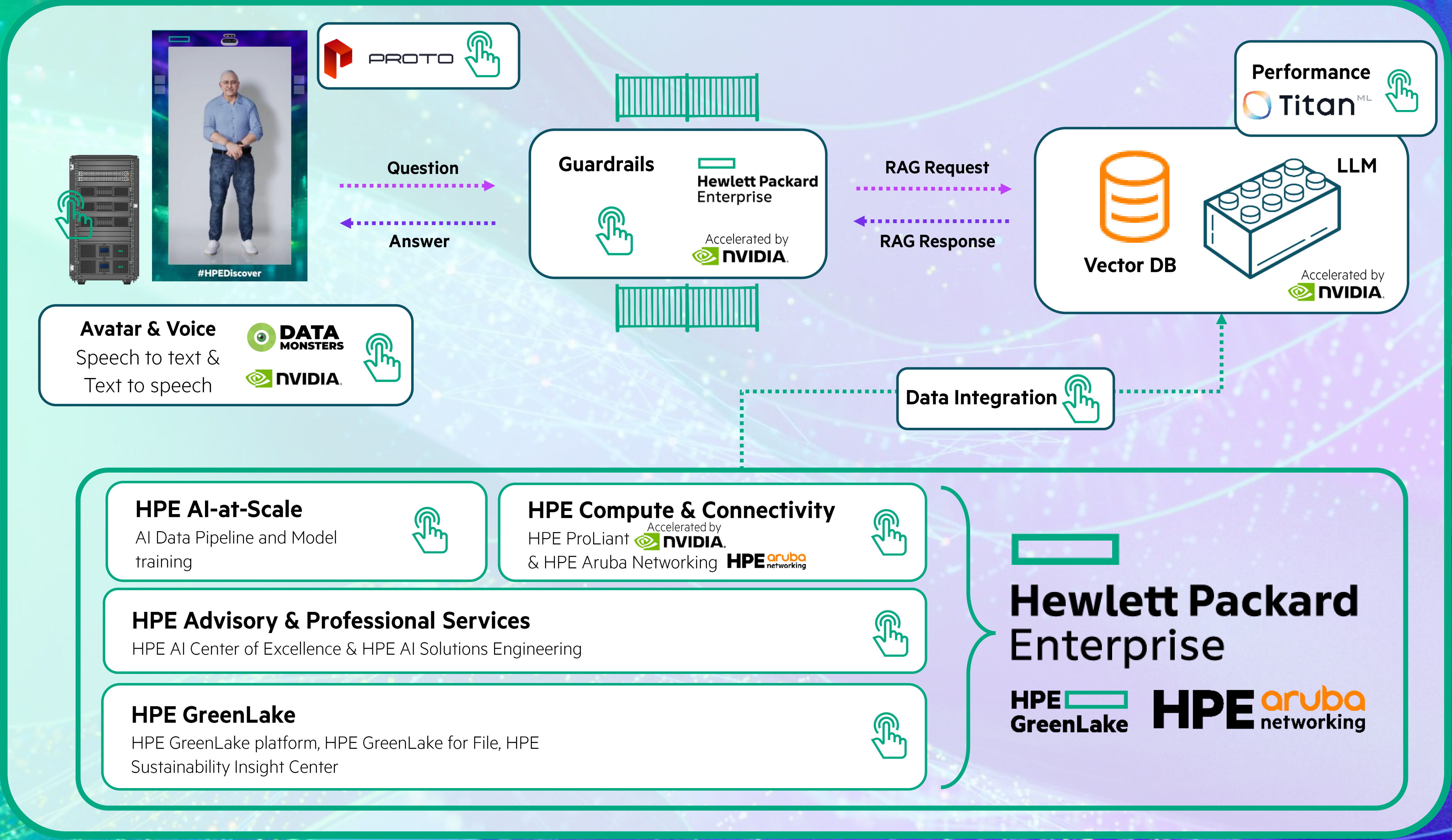
Giving Antonio "Nearly" a voice

Antonio's voice was created using four hours of data recognition training and over an hour of continuous speech. The data was labeled and used to train the Text-to-Speech model. Antonio Nearly's voice has gone through 30 different versions to get the one you hear today!

How are your questions processed?

Visitor speech is quickly converted to text using **NVIDIA RIVA's speech-to-text technology**.

These text requests then undergo a series of accurately-planned steps, including natural language understanding and response generation, ensuring the avatar's responses are relevant to the questions.


PROTO
#HPEDiscover
Question
Answer
Guardrails
Hewlett Packard Enterprise
Accelerated by NVIDIA
RAG Request
RAG Response
Performance
Titan ML
Vector DB
LLM
Accelerated by NVIDIA
Avatar & Voice
Speech to text &
Text to speech
DATA MONSTERS
NVIDIA
Data Integration
HPE AI-at-Scale
AI Data Pipeline and Model training
HPE Compute & Connectivity
HPE ProLiant Accelerated by NVIDIA
& HPE Aruba Networking HPE aruba networking
HPE Advisory & Professional Services
HPE AI Center of Excellence & HPE AI Solutions Engineering
HPE GreenLake
HPE GreenLake platform, HPE GreenLake for File, HPE Sustainability Insight Center
Hewlett Packard Enterprise
HPE GreenLake
HPE aruba networking

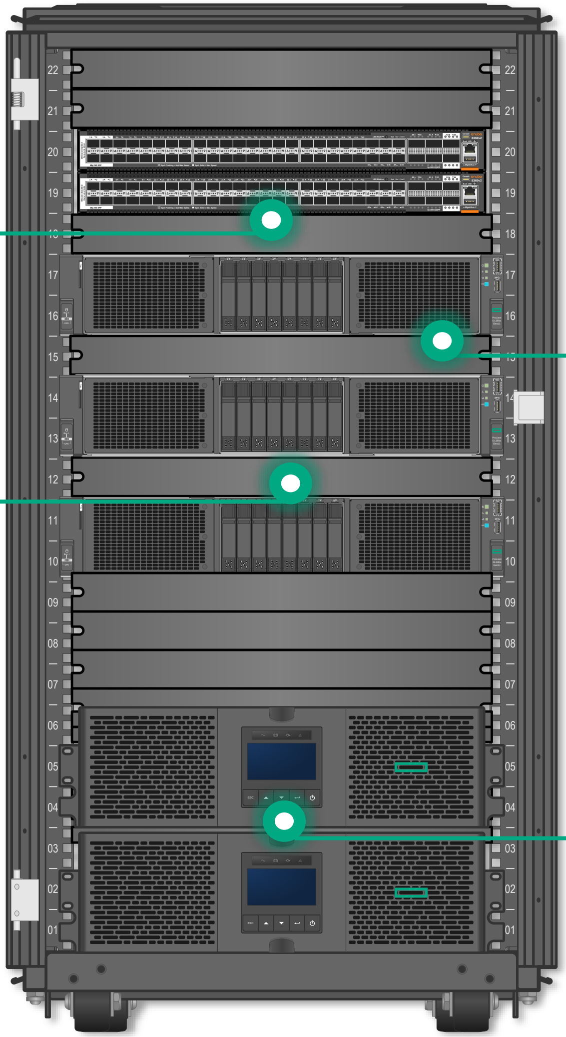
Hewlett Packard Enterprise

What's under the hood? HPE trusted enterprise AI at scale.

2X HPE Aruba CX9300
400GbE switch

2X HPE ProLiant DL385 gen11
4XNVIDIA L40S
Avatar and speech engine

Accelerated by



HPE ProLiant DL380a gen11
4X NVIDIA H100 NVL
Guardrails, vector DB, LLM, MLDM, MLDE

Accelerated by

HPE Uninterruptible Power System (UPS)

HPE 22U portable rack to power the AI stack and 3D digital hologram

PROTO

Proto is the world's first and only holographic communications platform with its own hardware, software, and app ecosystem.

Proto is known today for its Epic life-sized hologram device that is so realistic it seems as if there is actually a person (or object) inside their 7ft tall machine.

A live person can be beamed in real-time from any camera or mobile phone to any Proto device anywhere in the world, but HPE had a different idea...

HPE partnered with Proto to visually bring Antonio "Nearly" to life, making digital real. This partnership integrates Proto's spatial computing displays with HPE's secure AI infrastructure to create the avatar you can meet today.

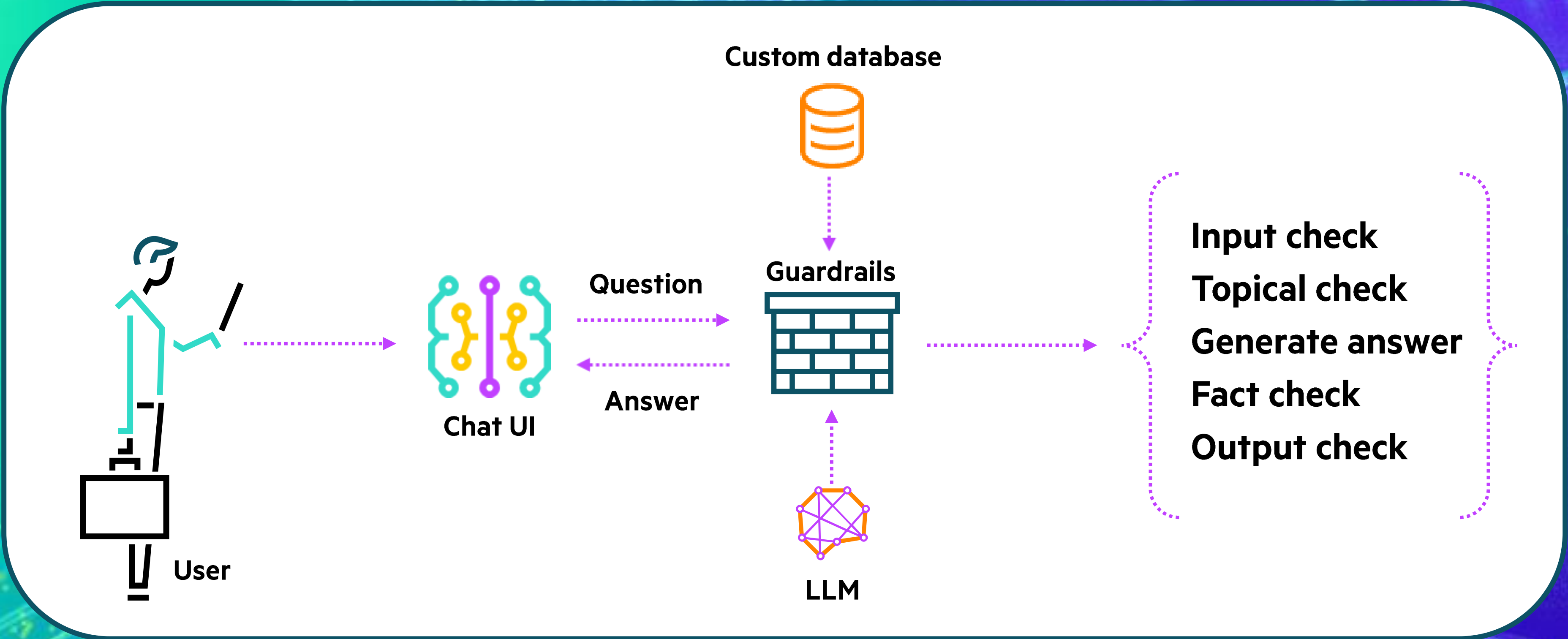
Trusted AI

Guardrails are a set of predefined rules, limitations, and operational protocols that serve to govern the behavior and outputs of advanced AI systems.

Policy enforcement

Contextual understanding

Continuous adaptability

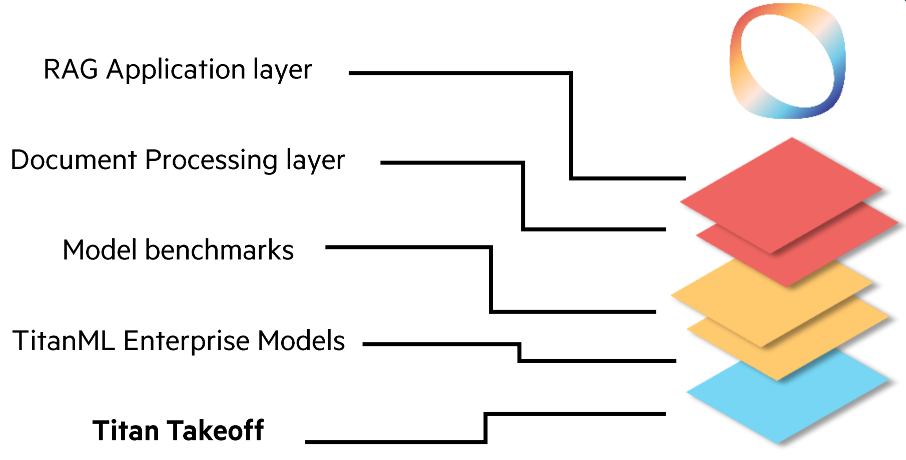


Built and delivered by the HPE AI Center of Excellence

TitanML

TitanML makes Enterprise AI possible offering best-in-class performance tailored to individual use cases. With TitanML we were able to easily use two separate models of today's latest open-source LLM on HPE equipment. This allows us multi-GPU support and interchangeable LLMs with optimized efficiency.

The TitanML Enterprise Inference Stack Unlocks Enterprise-ready AI
State of the Art AI, Streamlined for individual businesses



Antonio Nearly runs two separate LLM models on **Titan Takeoff** — Accelerated by NVIDIA

NVIDIA/Llama3-ChatQA-1.5-70B

- **Model Type:** Big Model
- **Model Size:** Large
- **Accelerators:** **NVIDIA** H100
- **Use Cases:** Chat

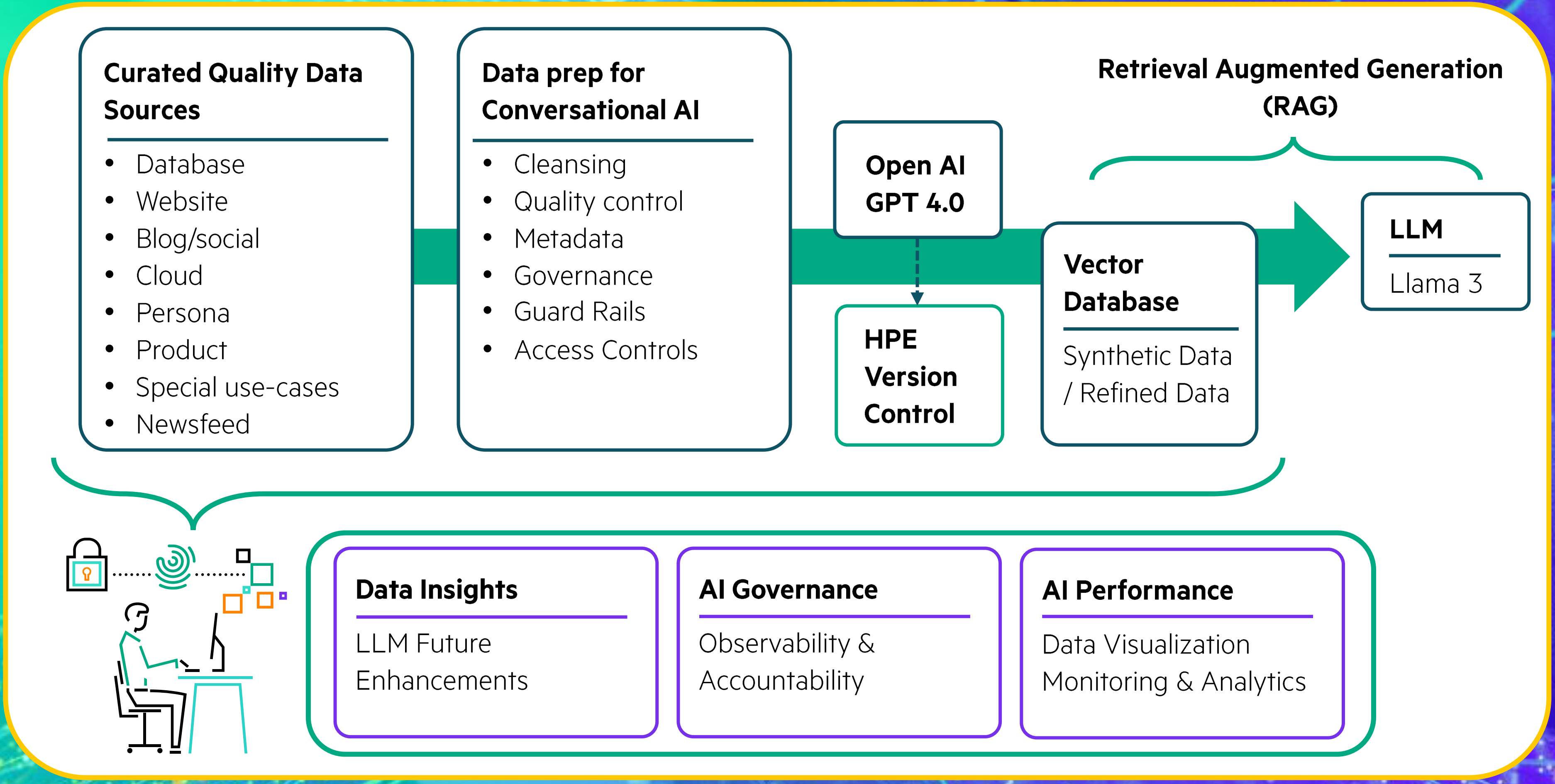
\+

NousResearch/Meta-Llama-3-8B-Instruct

- **Model Type:** Small Model
- **Model Size:** Medium
- **Accelerators:** **NVIDIA** L40S
- **Use Cases:** Text Generation

Data Integration

Data integrity is the foundation to make AI functional. The humans curating, visualizing, and monitoring the data must have the right tools and services to create the right AI framework for each individual use case.



Built and delivered by the HPE Digital Experience Team

HPE AI at Scale

Machine Learning Data Management	Machine Learning Development Environment	Machine Learning Inference Software
Data Processing & Pipelining	**Model Development & Optimization**	**Model Deployment & Monitoring**
• Any data source, format or type • Incremental processing optimizes ingestions of massive amounts of data • Flexible, automated data pipelines • Immutable data versioning and end-to-end data lineage tracking	• Distributed training • Experiment tracking and collaboration • Advanced GPU resource management and monitoring • GenAI and prompt engineering workflows with GenAI studio	• Model serving both both classical ML/DL models and LLMs • Automatic model packaging and deployment • Model observability, load balancing, and auditability

Workflow Management

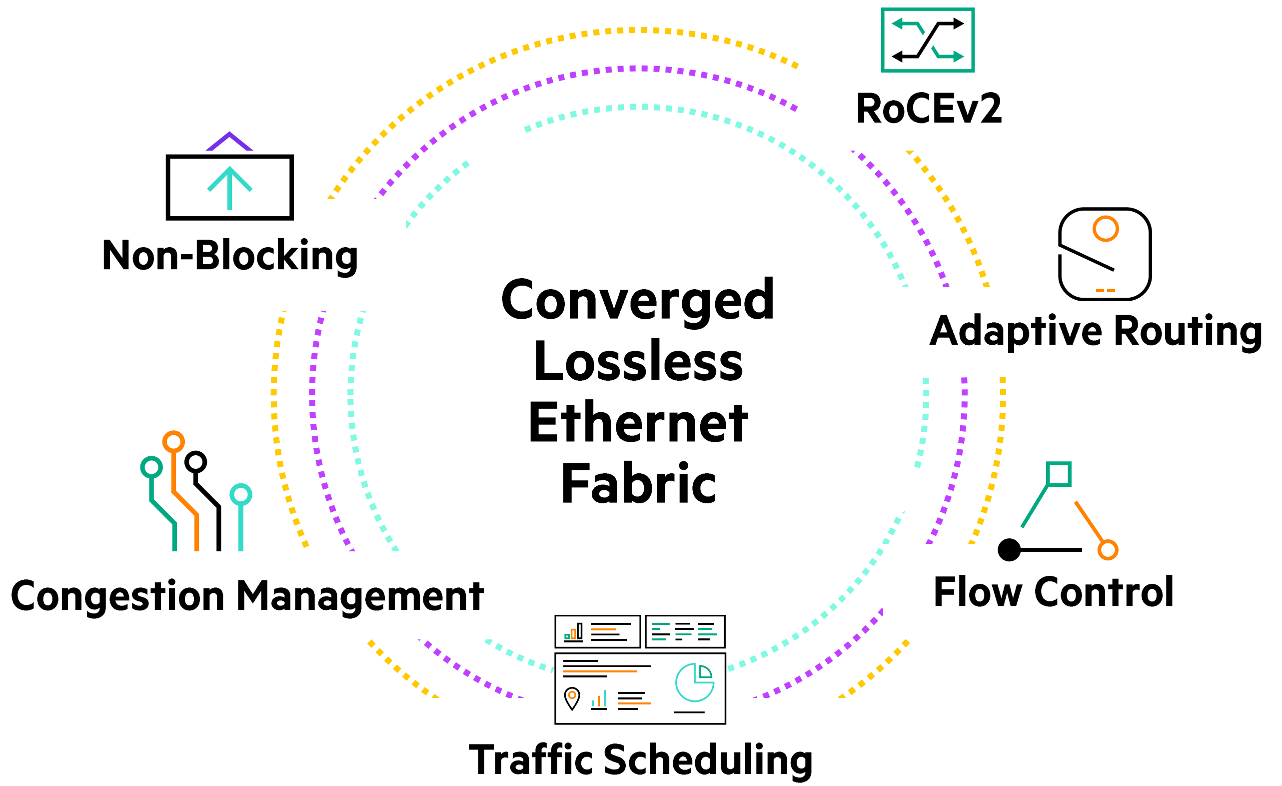
Distributed File & Object Store: HPE Ezmeral Data Fabric | Cloud | On-Premises

GPU Accelerated Infrastructure: Delivered by HPE GreenLake | Cloud | On-Premise

Built and delivered by the HPE AI Solutions Engineering Team

HPE aruba networking

The benefits of running Generative AI on a Converged Lossless Ethernet Network: Combining HPE Storage, HPE Compute and HPE Aruba Networking on a single Fabric.



HPE Advisory & Professional Services

Driving AI innovation and customer outcomes through use case development, project execution, and operation enablement.

Whether it is your first chat bot, or you are harnessing AI to help find a cure for cancer, we have the team of experts to collaborate on your AI solution.

Expand your team with HPE's:

- **Researchers**
- **Solution Architects**
- **AI Engineers**
- **ML & DL Ops**
- **Data Scientists**
- **AI Ethics Specialists**

Domain experts in:

- Machine learning
- Deep learning
- Natural Language Processing
- Computer vision
- Generative AI
- Predictive Analytics

HPE GreenLake

HPE GreenLake for File Storage offers everything you need to help maximize the vast potential of AI: enterprise performance at AI scale that supercharges workloads across every stage of the data pipeline; radical simplicity at AI scale with an intuitive cloud experience; and enhanced efficiency at AI scale for higher ROI.

2x System throughput

2x Performance density

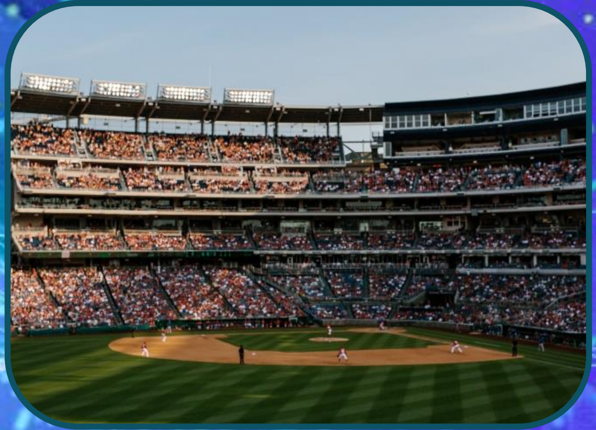
4x Capacity density

1/2 Power consumption

We created an interactive hologram of our CEO with our data. What will you do with your data?

Will you give doctors a deeper understanding of patient data?

Will you improve fraud protection and cyber security?



Will you enhance the fan experience?

Will you create digital concierge services?

HPE products, services, and solutions hold the potential to unlock your ambition and bring your ideas to life.

Ready to have a conversation with your data?

Generative AI

Explore

Transformative impact

Explore

Sustainability

Explore

HPE GreenLake

Explore

AI Engagement Model

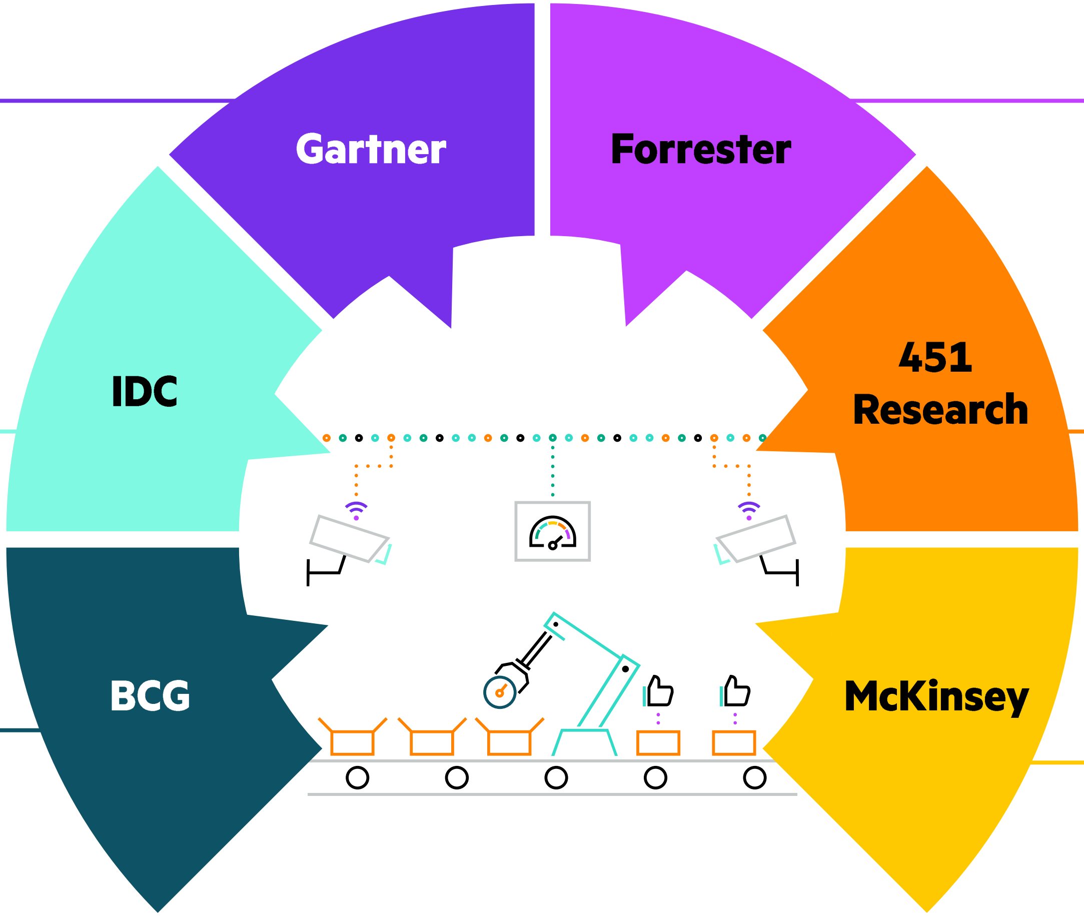
Explore

Generative AI

‘**Generative AI expands the output of AI systems** to include high-value artifacts such as video, narrative, software code, and synthetic data through to designs and schematics’.

‘**Generative AI is a branch of computer science** that involves unsupervised and semi-supervised algorithms that enable computers to create new content using previously created content, such as text, audio, video, images and code’.

‘**Generative AI is a set of algorithms**, capable of generating seemingly new, realistic content—such as text, images, or audio—from the training data. The most powerful generative AI algorithms are built on foundation models trained on a vast quantity of unlabeled data in a self-supervised way to identify underlying patterns for a wide range of tasks’.



‘**Generative AI is defined as a set of technologies and techniques** that **leverage very large corpuses of data, including large language models** like GPT-3, to generate new content. Inputs for generative AI may be natural language prompts or other non-code and non-traditional inputs.’

‘**Generative AI refers to AI models that are taught on large training sets** aimed to predict the probability of a sequence of words and to generate text. Some of the earliest examples were large language models that emerged from Google’s Bert in 2018.’

‘**Generative AI describes algorithms (such as ChatGPT) that can be used to create new content,** including audio, code, images, text, simulations, and videos.’

Source: BCG, IDC, Gartner, Forrester, 451 Research, McKinsey

Transformative Impact

Technology

~88%

Of software developers reported higher productivity when using a generative AI code assistant

Consumer

Automated on-model fashion image generation resulted in

1.5X

Increase in retailer conversion rate

Biopharma

Generative AI Identified a novel drug candidate for the treatment of Idiopathic Pulmonary Fibrosis in

21 days

(vs. years with traditional methods)

Financial Institutions

Synthetic GAN-enhance training set for fraud detection achieved a

~98%

accuracy rate

(vs. 97% with unprocessed original data)

Entertainment

Generate novel animated motions from a single training motion sequence with

~97.2%

quality score on natural movements

(vs. 84.6% with traditional methods)

Insurance

InsureTech platforms leveraging generative AI to reduce up to

~30%

of customer service costs

Source: News Articles, Deloitte, MIT Sloan, BCG, McKinsey

Sustainability

AI solutions have an insatiable appetite for resources that cannot be ignored. A data-driven sustainability approach with measurable goals is essential.

HPE's industry-unique approach to sustainable IT transformation

Approach → The 5 efficiencies to consider: Data efficiency, Software efficiency, Equipment efficiency, Energy efficiency, and Resource efficiency

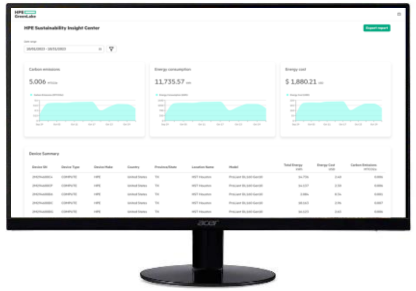
Expertise → We can help you get there with advisory services, design expertise, and a unique workload-in approach to speed your journey

Circular Economy → Leverage our efficient product design, Certified Pre-Owned, and asset upcycling to extend asset life and help self-fund your transformation

HPE GreenLake Sustainability Insight Center

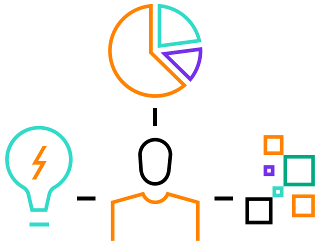
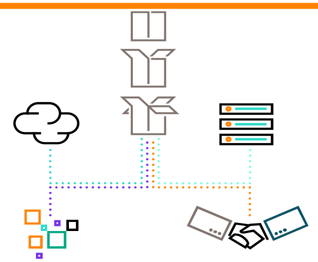
Delivers key insights on your IT estate

- Energy consumption
- Estimated Carbon emissions
- Estimated electricity costs
- By device, location, and date range



HPE GreenLake

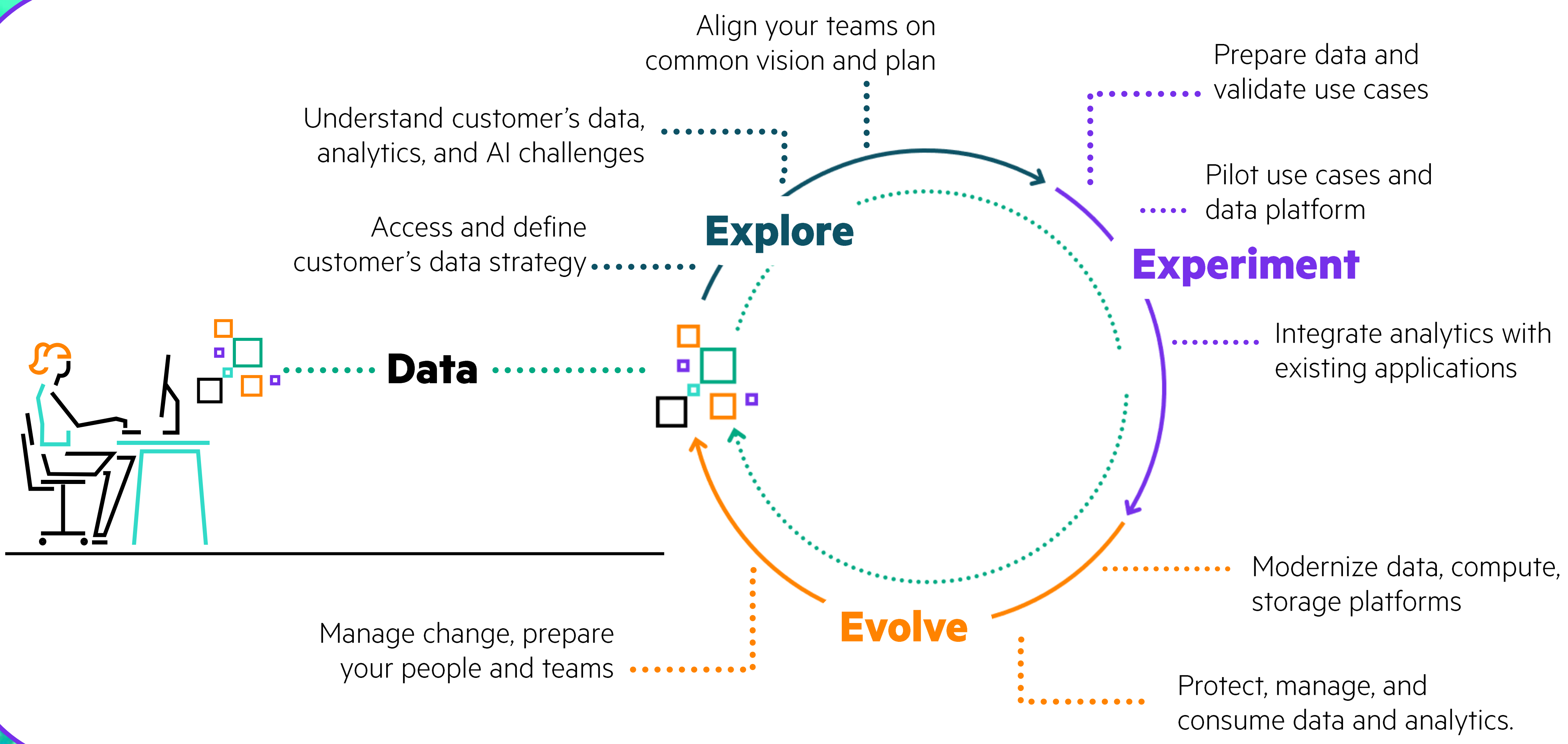
- HPE GreenLake for Large Language Models
- HPE GreenLake for High Performance Computing
- HPE GreenLake for Private Cloud Enterprise
- HPE GreenLake for File Storage
- HPE GreenLake Flex Solutions

We Build	We Optimize	We Integrate
 HPE develops and deploys the AI solution end-to-end	 Bring your own AI model and HPE will optimize it for you	 HPE will work to integrate partner ecosystem into your solution

Advise & Align | Assess & Analyze | Plan & Design

Build & Implement | Enhance & Optimize | Operate & Manage

AI Services Engagement Model



Join us for an AI transformation workshop.

AI transformation workshops, for every AI journey:

AI governance | Best Practices | Defining risk and ROI | Building internal use cases | Executing use cases

Thank you!

Daniel Sherrill
Founder

Dan Kortenhoven
Lead Architect

Nyla Kongsinh
Project Manager

Dr. Eng Lim Goh
Principle Advisor

Morad Qutwut
Advisor

Cyril Hug
Advisor

David Head
Executive Sponsor

Core Team

Akhila Aiyer
Andrew Mendez
Austin Hirsch
Jing Li
Jose Gonzalez
Kevin Musgrave
Sara Diaz
Sekhar Pogula
Sundaram Shandilya

Storytelling

Holly Hubbard
Michael Cooper

Contributors

Aaron Lamond
Alex Vannier
Alison Paisley
Andrew Elisavetski
Andrew Taylor
Beth Bartel
Bijan Varjavand
Chris Bradley
Derek Vanderloo
Dimitrios Poulopoulos
Gary Craze
George Thornton
Heather Otto
Jeff Fougere
Jerome Justet
Ka Wi Leung
Katie Champagne
Khoa Nguyen
Lindsey Hillesheim
Nora Dailey
Piyush Shukla
Radhakrishna Naidu Balineni
Sandy Correa
Tami Wessley
Victoria Lee Hanrahan
Yash Nagaraju

Thank you!

DATA MONSTERS

Alexander Nesterok
Alexey Kuznetsov
Andrew Dvorko
Artem Kozmin
Dan Lesovodski
Dmitry Malkov
Margarita Zomareva
Maxim Pavlov
Sergey Berezin
Vladimir Nechaev

PROTO

Edward Ginis
Eric Chasko
Noah Rothstein
Nolan Cassidy
Raffi Kryszek
Sean Storin

TitanML

Meryem Arik
Peter Bhabra
Rod Rivera

Have a chat with Antonio “Nearly”

Please have your questions ready

If there is a line, please keep to 3 questions

Hold down the mic button to ask your question

Try not to pause too long during your question

Ideas to ask:

What’s so great about HPE GreenLake?

Why should I choose hybrid cloud?

What is Private Cloud AI?

Tell me about the partnership between NVIDIA and HPE.

What's the difference between Private Cloud Business Edition and Private Cloud Enterprise?

Where can I get the best pizza in Las Vegas?

Who is your favorite soccer team?

Who is your favorite Formula 1 team?