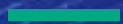




Hewlett Packard
Enterprise

HPE GREENLAKE FOR LARGE LANGUAGE MODELS



Large Language Models as a Service from HPE

Bill Mannel
September, 2023

LARGE LANGUAGE MODELS AS THE FOUNDATION FOR GENERATIVE AI HAVE THE POTENTIAL TO:

Disrupt nearly every industry

promising both competitive advantage and creative destruction¹

Change the anatomy of work

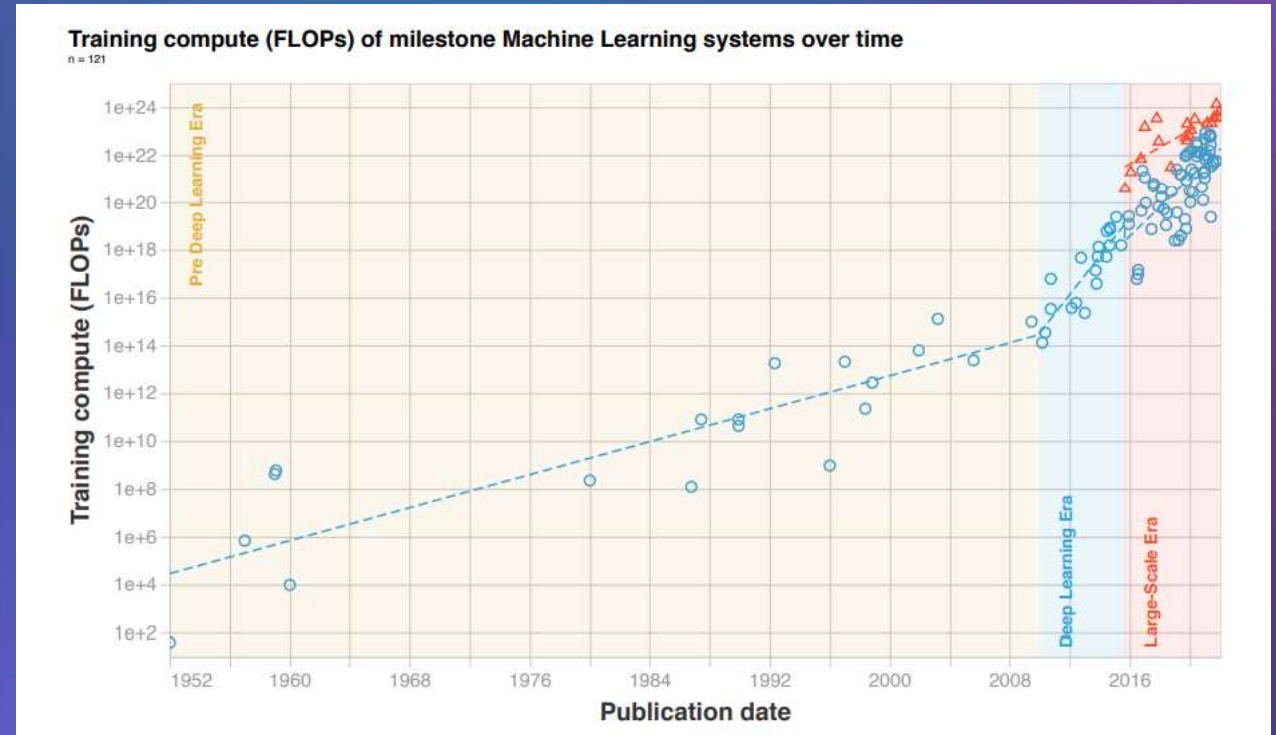
augmenting the capabilities of knowledge workers by automating 60% to 70% of their individual activities²

¹ Boston Consulting Group, [The CEO's Guide to the Generative AI Revolution](#), March 2023

² McKinsey & Company, [The Economic Potential of Generative AI – The next productivity frontier](#), June 2023

Large language models are a supercomputing problem
Development, training, tuning and deployment are very compute-intensive

“The computing requirements for large-scale AI models doubled every 10.7 months from 2016 to 2022.”



Source: [Compute trends across three eras of machine learning](#), University of Aberdeen, Centre for the Governance of AI, University of St. Andrews, MIT, University of Tübingen, Complutense University of Madrid, March 2022



Challenges with scaling generative AI POCs into full scale production

Going from successful POC into reliable and sustainable production is challenging for enterprises



Chief Executive Officer

Productivity

Goals

“I want successful POCs scaled into production asap to drive competitive advantage through productivity”

Challenges

- Huge upfront capital outlay for on-premises AI supercomputer with 100s or 1000s of processing units
- Time lag of several quarters from order to production



Chief AI/Data Officer

Innovation

“I want successful POCs scaled into production asap to drive competitive advantage through innovation”

- Lack of skills to operate an AI supercomputer
- Concerns about ability of public cloud providers to scale to the required level of 100s or 1000s of processing units working in concert to get the job done



Chief Sustainability Officer

Sustainability

“I am concerned about the negative environmental impact of large AI models like Large Language Models”

- The energy consumption of large language models can be prohibitive for organizations with aggressive net zero pledges
- The potentially negative impact of large AI models has been getting a lot of negative press lately.

The industry's first LLM application run on the most advanced
supercomputing software and architectures accessible to you on-demand

HPE GREENLAKE for
**LARGE LANGUAGE
MODELS**

Run on the most advanced supercomputing software and architectures for unmatched scalability and reliability

Built in partnership with Aleph Alpha and their pre-trained GreenLake LLM to train, tune, and deploy models with your data in five different languages

Hosted in AI-native datacenters powered by 100% renewable energy and employing waste heat capture/re-use

Accessible to you on-demand and backed by HPE Services to ensure your success



Large Language Models as a Service

HPE GreenLake for Large Language Models

AI experts that build or fine-tune models

Customers that use a pre-trained model

HPE MLDE

HPE MLDM

LLM as a service with Aleph Alpha LLM

AI Producers

AI Consumers

HPE Project Breckenridge
supercomputing as a service

LLM Inference Service with Aleph Alpha LLM model

What is Aleph Alpha LLM model?

Pre-trained ready-to-use large language model from Aleph Alpha that is designed with a focus on accuracy and explainability

What is the Service?

HPE offers enterprise customers access to a large language model to safely use with their own data to gain business value. The service is addressed to customers in all industries and provides the following benefits:

Upload data
in multiple
formats

Word, ppt, excel, html,
SharePoint

Data privately
and securely
stored

Your data stays private to
your company

Question
answering
in natural
language

Languages supported:
English, Italian, Spanish,
German and Portuguese

Quick access to
source data

Interface shows data source
for the generated answer
and a reliability score

Accelerate strategic generative AI initiatives – in a sustainable way!

HPE GreenLake for Large Language Models addresses all challenges



Chief Executive Officer

Productivity

Goals

“I want successful POCs scaled into production asap to drive competitive advantage through productivity”

HPE GreenLake for Large Language Models

- No large upfront capital outlay
- Fast access to cloud-native AI supercomputing capability to scale successful POCs into full scale production



Chief AI/Data Officer

Innovation

“I want successful POCs scaled into production asap to drive competitive advantage through innovation”

- No need to create in-house capability to maintain and operate a large-scale AI supercomputer
- Piece of mind that the solution will reliably scale to the 100s or 1000s of processing units working in concert to get the job done.



Chief Sustainability Officer

Sustainability

“I am concerned about the negative environmental impact of large AI models like Large Language Models”

- 100% powered by renewable energy – mainly hydropower
- Waste heat capture and re-use contributes to an even more sustainable ecological footprint

Next steps

- Schedule a deep dive briefing of HPE GreenLake for Large Language Models
- Define the optimal solution to meet your business needs with your experts and our experts working together.
- Get access to HPE GreenLake for LLM to embed AI into every aspect of your strategy and deploy it at scale across every facet of your enterprise – backed up HPE know how and enterprise-grade **SLA's** to ensure your success.

Visit hpe.com/hpe-greenlake-large-language-models for more information

THANK YOU

Bill Mannel
bill.mannel@hpe.com

