



INSIGHTS | WHITE PAPER

Mastering AI For The Enterprise With HPC: A Comprehensive Guide

Is there any area of today's technology that is more captivating than artificial intelligence (AI)? Perhaps not, as it is something that's gained mainstream attention well outside the industry bubble of technology. But the AI phenomenon didn't happen overnight. In fact, its phrase was coined by an assistant professor at Dartmouth college back in the 1950s and there have been numerous significant milestones since then.²

Who can forget when IBM's Deep Blue made headlines by becoming the first computer system to defeat a reigning world chess champion in 1997? Another landmark event occurred in 2012 when Google's AI system learned to recognize cats from photos. More recently, the introduction of Open AI ChatGPT has democratized AI, enabling people worldwide (and outside the tech world) to casually interact with sophisticated AI systems. Encaptured by the promise of this new AI marvel, nearly 3 in 4 CEOs now believe that generative AI is a top spending priority.³

AI Literacy Is Essential

Despite the considerable hype surrounding AI, there remains a great deal we have yet to understand and discover. One thing we do know is that AI is going to have a profound impact on your industry, your business, and the customers you serve for the generations to follow. It is going to serve as a competitive advantage for some organization in your industry, if not yours. It is also going to have an extraordinary impact on the daily roles of individuals within organizations. While there is a growing fear that AI might lead to widespread job losses across professions like engineering, writing, programming, and legal work, the true shift will likely be subtler, yet profound. It won't be a matter of AI replacing people wholesale, but rather, those proficient in leveraging AI will find themselves at a distinct advantage over those who do not adapt. This transition emphasizes the importance of integrating AI literacy across all levels of a workforce.

All of this means one thing: Whether you're a CEO, a business owner, a manager, an IT administrator, or a language translator, it's crucial to understand AI and how to leverage it in your role. Every organization needs to harness AI effectively, integrating it into its operations to enhance workforce capabilities. By fostering an understanding of AI as a tool for empowerment rather than as a threat to employment, businesses can transform AI into a significant advantage, driving innovation and efficiency across all levels of the organization.

40%

of core IT budgets will be allocated toward AI-related initiatives for Global 2000 companies **by 2025.**¹



**Hewlett Packard
Enterprise**

This WEI white paper aims to deepen the understanding of AI and highlight the critical role of High-Performance Computing (HPC) in managing extensive datasets and advancing sophisticated machine learning models.

More Than One Type of AI

The first thing to realize is that AI is often misinterpreted and lumped together in conversations. An applicable example are terms such as *internet*, *web*, and *cloud*, as they are mentioned interchangeably. This confusion extends to how AI, machine learning (ML), deep learning (DL), and generative pre-trained transformer (GPT) are mentioned synonymously, and incorrectly so. Artificial Intelligence itself is a vast field of computer science dedicated to creating systems capable of executing tasks that usually necessitate human intelligence, such as decision-making, problem-solving, and learning. It encompasses various distinct domains, each with specialized focuses and applications, highlighting the importance of understanding the nuances between them.

Machine Learning

Machine Learning is a subset of AI that involves training algorithms on data to make predictions or decisions without being explicitly programmed. ML is the simplest form of AI where machines are trained to make decisions. An example of simpler ML in action is a smart thermostat, which adjusts home temperatures based on learned user preferences. More sophisticated implementations are seen in services like Netflix or Amazon, which utilize ML to analyze user behavior and preferences to personalize recommendations, significantly enhancing the user experience.

Deep Learning

Deep learning is an advanced subset of machine learning that uses multi-layered neural networks to mimic human decision-making processes. It is with DL that we begin training models. In tasks like image recognition, deep learning systems are trained on large datasets to identify and categorize images with high accuracy. For instance, when recognizing animals, these systems analyze basic features (shapes, colors) and finer details (ear shape,

leg count) to identify images as cats. This hierarchical decision-making improves with more data, enabling the system to refine its predictions and increase accuracy autonomously. Deep learning's self-improving capability allows it to process data more effectively, enhancing its performance over time without human intervention.

Generative AI

All forms of AI operate by processing input to generate output. Back to the smart thermostat example, it adjusts your air conditioner based on the current temperature it senses. Similarly, an autonomous vehicle uses real-time sensor data to control steering, acceleration, or braking. However, generative AI stands out because it produces unique content based on the input it receives, much like two subject matter experts may provide slightly different responses to the same question. This capability allows tools like ChatGPT to give varied answers to identical queries, showcasing a significant evolution from the more predictable input-output models used in traditional machine learning and deep learning.

Large Language Model

A Large Language Model (LLM) refers to advanced generative AI systems that are designed to understand and generate human-like text based on extensive training data. These models, such as generative pre-trained transformer (GPT), use deep learning techniques to process and produce language in a way that mimics human writing and conversation. LLMs can perform a variety of language-based tasks, including answering questions, composing text, summarizing content, and more, making them powerful tools in both academic and industry settings for enhancing natural language processing applications.

Another Acronym: HPC

From an enterprise perspective, there is another acronym that AI-enabled organizations must understand – HPC. High Performance Computing (HPC) is the critical infrastructure that AI-driven organizations need to leverage effectively. HPC refers to a robust computing setup designed to process large volumes of data and

execute complex calculations at high speeds. This system often includes clusters of servers that work in parallel, sophisticated networking, and high-speed storage capabilities.

HPC systems are essential for efficiently managing the hefty datasets typical in AI operations, offering scalable solutions that can grow with the increasing demands of AI computations. These environments not only provide the necessary computational power for rapid processing and real-time decision-making, but they also enable organizations to maximize their AI capabilities by supporting the intense workloads required to train and run AI models effectively. Only HPC can handle the complex intense processing tasks that AI demands.

The Growth Challenge

Here is the challenge. According to Nvidia, AI models will be one million times more powerful than ChatGPT within 10 years.⁴ You have undoubtedly heard of Moore's Law. Well, even in its best days, it would have only delivered 100X in a decade. While the good news is that new GPU processors have contributed to a 1000x increase in our compute capacity, the growth rate of AI models has increased by 8000x according to Bill Mannel, Chief Technologist at HPE.

According to Google DeepMind founder, Mustafa Suleyman, AI training models will be 1,000 times larger in only three years.⁵ A Wall Street Journal article from 2024 highlights the soaring demand for textual data, raising concerns among researchers and executives that the available supply may not suffice to meet AI's escalating needs within the next two years.⁶ Despite a 70-fold increase in data capacity over the past 13 years, the rapid expansion of data-intensive AI technologies could lead to a shortfall, affecting data-driven companies significantly.

Growth Of AI Expenditure

There is something else besides data and computing power that's necessary to fuel the AI boom, and that is money. By 2026, global spending on AI is supposed to exceed \$301 billion.⁷ The generative AI market alone is expected to grow to \$1.3 trillion over the next 10 years. That is from a market size of just \$40 billion in 2022.⁸ According to a Goldman Sachs report, AI-related

investment could peak as high as 2.5 to 4% of GDP in the U.S.⁹ Businesses are investing in AI to gain a competitive edge, gain insights into customer behavior, streamline operations, and innovate product offerings.

The investment boom in AI extends far beyond the tech sector, encompassing industries such as finance, hospitality, manufacturing, and healthcare. This widespread infusion of capital highlights a global transformation, as businesses across all sectors strive to innovate and secure a competitive advantage.

HPE and High-Performance Computing

If the AI revolution is dependent upon high performance computing, it should come as no surprise then that IDC positions HPE in the Major Players category for worldwide AI in 2023. By offering the most comprehensive end-to-end portfolio in the HPC and AI segments that can meet the needs of any business, HPE is helping organizations architect the AI advantage that they are seeking today to set them apart from their competitors.

HPE delivers an open, full stack solution for running an AI-powered business. This solution suite includes technologies such as:

- **Exascale Computing:** Capable of performing a quintillion calculations per second, supporting massive data-intensive applications.
- **Cray Supercomputing:** Advanced supercomputers designed for high-scale performance to handle complex simulations and computations.
- **Slingshot:** HPE's high-speed, purpose-built networking architecture designed to connect distributed supercomputing nodes with unprecedented speed and efficiency.
- **Data Storage Solutions:** High-capacity, scalable storage systems optimized for high-performance and data-intensive tasks.

What makes their HPC platform even more powerful is the strength of their vast, open ecosystem that offers access to diverse tools, effectively avoiding vendor lock-in and providing users with the freedom to choose their preferred technologies.

It Isn't Just About Hardware

However, HPE goes beyond just providing powerful hardware and technology integration. Their AI software tools prepare your data to be AI-ready, automating and scaling complex data pipelines to expedite the development and deployment of AI models. With HPE's Machine Learning Inferencing Software (MLIS) you can quickly deploy machine learning models into production environments. The software ensures that ML models run efficiently, with tools that monitor performance, manage resource allocation, and scale operations as needed, catering to enterprise needs for robust AI-driven decision-making systems.

Alongside these technological solutions are HPE's team of AI experts that include data scientists and engineers specializing in machine learning and ML Operations to provide crucial support when needed. They evaluate business goals, align them with technological needs and performance metrics, and ensure all operations are conducted within a secure framework.

The HPE Machine Learning Development Environment

The HPE Machine Learning Development Environment is part of HPE's broader efforts to streamline AI integration and deployment for businesses. This environment provides a comprehensive platform that combines hardware, software, and tools optimized for machine learning and deep learning applications. It helps businesses by simplifying the configuration and management of necessary resources, reducing the time from development to deployment. By offering a ready-to-use, scalable solution, it enables enterprises to focus on developing innovative AI-driven applications without the need to invest heavily in underlying infrastructure setup and maintenance. This environment supports a range of industries in adapting AI to enhance operational efficiencies, improve decision-making processes, and create new product offerings.

HPC-as-a-Service With HPE GreenLake

While there are certainly many companies applying significant capital investments into HPC, that isn't the only way to harness the power you need to drive your AI initiatives. HPE GreenLake is known for its expansive as-a-Service offerings that provides cloud services and infrastructure management for various IT resources. This portfolio is something the WEI team is very familiar with, and that includes AI solution. HPE GreenLake offers a scalable, flexible way to deploy and manage AI workloads on a pay-as-you-go basis.

This platform helps organizations use AI without large upfront investments in physical infrastructure by hosting and managing the heavy data and compute demands of AI systems. It provides AI-focused solutions for analytics, data protection, and business insights, all within a centrally managed hybrid cloud environment.

Talk To WEI Today

As time advances, leveraging cutting-edge technology to stay competitive is crucial. HPE provides an expansive suite of high-performance computing and AI solutions tailored to meet diverse business needs.

With deep expertise and scalable HPC solutions, WEI and HPE equips your business to tackle complex challenges and transform operations in a way that only AI can. Harness the transformative power of HPE's AI technology to propel your business forward and maintain a competitive edge in your industry. Contact WEI today to get started.

Sources:

1. IDC FutureScape: Artificial Intelligence Will Reshape the IT Industry and the Way Businesses Operate, October 2023
2. Dartmouth workshop - Wikipedia
3. CEOs are prioritizing generative AI investments: KPMG | HR Dive
4. Nvidia predicts AI models one million times more powerful than ChatGPT within 10 years | PC Gamer
5. AI training models will be 1,000x larger in three years (futuretimeline.net)
6. The Internet May Be Too Small for the AI Boom, Researchers Say - Tech News Briefing - WSJ Podcasts
7. IDC Worldwide Artificial Intelligence Spending Guide, August 2022
8. Generative AI to Become a \$1.3 Trillion Market by 2032, Research Finds | Press | Bloomberg LP
9. AI investment forecast to approach \$200 billion globally by 2025 (goldmansachs.com)

About WEI

WEI is an innovative, full service, customer centric IT solutions provider.

Why WEI? Because we care. We go further.

WEI is an expert in business technology improvement, helping clients optimize their technology environments and work efficiently. WEI works with clients to understand goals, integrate strategy with technology solutions, and leverage their current IT environment into one company-wide model to increase utilization and efficiencies around their unique business processes.