NVIDIA.

Artificial intelligence, made simple

Artificial intelligence, made simple

Written by: **Dale Benton**
Produced by: **Danielle Harris**

THE ARTIFICIAL INTELLIGENCE MARKET IS SET TO EXPLODE AND NVIDIA, ALONG WITH THE TECHNOLOGY ECOSYSTEM INCLUDING PURE STORAGE, HAS STRATEGICALLY POSITIONED ITSELF TO TAP INTO THIS HUGE MARKET GROWTH

After decades in the doldrums, AI has recently exploded in a boom that has unleashed applications used by hundreds of millions of people every day. The impact of this technology has been likened to that of electricity 100 years ago: AI won't be an industry, it will be part of every industry.

The rapid rise of AI has left many businesses scrambling to understand how they can benefit from a technology they don't yet understand. An MIT Sloan Management Review found that 85% of executives surveyed believe AI will transform their company, yet only 39% report having a strategy for AI.

One company has taken the lead in helping business decision makers evaluate, implement and monetise AI. NVIDIA cut its teeth on 3D graphics for gaming and professional design, but the graphics processing unit (GPU) NVIDIA invented back in the 1990s has revealed itself as a processing powerhouse, capable of tackling computing's grandest challenges.

Richard Jackson, Vice President for the EMEA Partner Organisation at NVIDIA, reveals the story behind AI's rapid rise from sci-fi gimmick to reality, and explores how businesses can take their first steps in this brave new world.

DEVELOPING DEEP LEARNING

Founded in 1993, NVIDIA's GPU sparked the growth of the PC gaming market. Now this same tiny piece of silicon is credited with unleashing the Intelligent Industrial Revolution.

Several years ago, researchers discovered that the same parallel architecture designed to handle the vast amount of data required for 3D

THE NEW POSSIBLE

WELCOME TO THE GOLDEN AGE OF AI, POWERED BY THE MODERN DATA PLATFORM

The fourth industrial revolution is upon us, powered by the rapid rise of artificial intelligence (AI) and machine learning. Every industry benefits from greater intelligence: the ability to transform data into intelligence is the new competitive currency. Some advances in industries include:
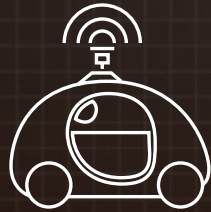
Driving autonomous vehicles

Aiding doctors for more accurate insights

Helping agents to serve customers better

The big bang of AI has been fueled by a perfect storm of three key technologies: deep learning (DL), Graphics Processing Unit (GPU) processors, and big data. NVIDIA is the world leader of deep learning technologies, providing the most advanced GPUs along with the NVIDIA GPU Cloud Deep Learning Stack: optimized versions of today's most popular Deep Learning frameworks.

While deep learning algorithms and GPUs are massively parallel, delivering performance leaps every year, legacy storage systems were largely built on decades-old building blocks, designed in the serial era. And the performance gap between compute and storage continues to grow.

FlashBlade from Pure Storage is the industry's first data platform purpose-built for AI and machine learning. Fast, big and simple, FlashBlade is massively parallel at its core, delivering unprecedented performance and simplicity for data scientists. Like a GPU-accelerated server, FlashBlade is architected to accelerate parallel workloads, delivering the performance of 10 racks of disk in a 4U size.

NVIDIA DGX-1 SERVER	PURE FLASHBLADE
Delivers Performance of 10 Racks of CPU Servers	Delivers Performance of 10 Racks of Disk

Unlock the full potential of your data with Pure Storage and NVIDIA; providing tera ops of performance and terabytes of storage in dense form factors that are significantly more efficient than alternative legacy solutions. Enabling you to put your data to work, with speed, with agility and with intelligence.

Today is The New Possible.

For more information, please visit **purestorage.co.uk/FlashBlade**

graphics was also a perfect fit for the complex parallel computing required by deep learning. This form of artificial intelligence enables computers to learn from data and write software that is too complex for people to code.

NVIDIA recognised the opportunity presented by this affinity between deep learning and the GPU. Since then, it has been investing in a new computing model, GPU-accelerated deep learning, which is helping to create computers, robots and self-driving cars that can perceive and understand the world.

"Although AI has been around for a long time, before now the processing power needed for it to succeed just wasn't available," explains Jackson. "Now, NVIDIA is working to democratise AI for all."

Pure Storage also recognised data is the fuel to drive deep learning, and parallel architecture is the future. It built a new data platform from the ground-up to keep pace with the innovation curve of GPUs.

"Deep learning is unique among all learning algorithms in that it keeps getting better with more data," says James Petter, VP EMEA
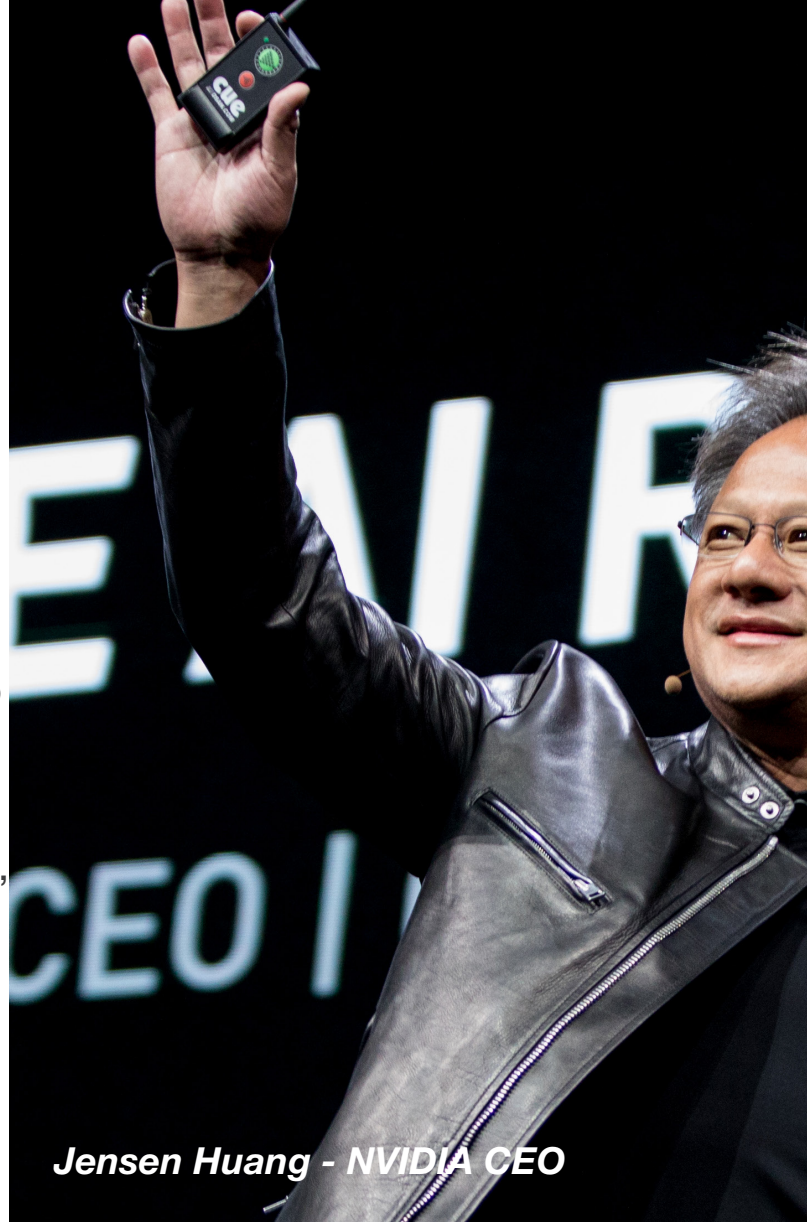
Jensen Huang - NVIDIA CEO

at Pure Storage. "At Pure, we believe data should never be the bottleneck for data scientists."

THE VOCABULARY OF AI

The first step in realising the business benefits of AI is to understand a few fundamentals. When we talk about AI, three terms tend to be used interchangeably: artificial intelligence, machine learning and deep learning. Their relationship

is a bit like Russian dolls. AI is the overarching idea, within which machine learning and deep learning fit.

For a large part, the entertainment industry has molded what we think of when we think of intelligent machines and AI. In reality, what's possible today is known as 'Narrow AI', as opposed to the 'General AI' displayed by C3PO and The Terminator. Narrow AI encompasses technologies that can perform specific tasks, such as image classification or speech recognition, as well as or better than humans.

This human-like intelligence brings us to deep learning. It's a fundamentally new software model where billions of software-neurons and trillions of connections are trained in parallel. The graphics processing unit emerged as the ideal processor to accelerate deep learning.

"GPUs, like artificial neural networks and the human brains on

1999

The year that NVIDIA
invented the Graphics
Processing Unit (GPU)

which they're modelled, process information in parallel, handling multiple tasks simultaneously," says Jackson. "That's why GPUs can now be found accelerating deep learning-based applications from movie recommendations to cancer detection and fraud detection to self-driving cars."

DEMOCRATISING AI

The democratisation of AI brought about by GPU-accelerated deep learning is already finding its way into deployment across industries. As this form of AI expands from research institutions and startups to implementation by large enterprise, new use cases for deep learning are emerging daily.

From intelligent assistants to smart homes to self-driving cars, it's clear that this new computing model will infuse consumer technology as much as it will reinvent enterprise computing.

Jackson comments: "Those businesses looking to grab the competitive advantage offered by AI have a narrow window of opportunity. For those who move quickly, rewards

will include efficiencies in existing processes, and insights based on data and predictive analysis that enable new classes of products and services.

"We see companies like SAP seizing an early-mover advantage by implementing GPU deep learning in their data centres to solve their customers' most challenging problems."

PLUG AND PLAY AI

Thanks to the rapid development of the AI industry, this technology is available in many flavours and at varying scales. For those looking to dip a toe in the water, cloud service providers like Microsoft and AWS offer GPU deep learning cycles on demand.

NVIDIA has also developed an offering aimed at companies seeking a combination of unprecedented computing power with security and support.

Last year, NVIDIA launched the DGX-1. It's essentially an AI supercomputer in a box, purpose-built for deep learning. Instead of building its AI data centre from the ground up, the DGX-1 integrates everything data

Richard Jackson

"THE DGX-1 REPRESENTS A REAL BREAKTHROUGH IN TECHNOLOGY BUILT FOR ARTIFICIAL INTELLIGENCE"

scientists need to get started building, training and running powerful and sophisticated deep neural networks.

But all of this data needs to be stored somewhere. Traditional storage systems were largely built on decades-old, building blocks burdened with serial bottlenecks, and have been proven to lag behind the performance curve needed to keep GPUs busy with data.

However, Pure Storage supports NVIDIA's efforts to democratise AI by enabling companies to store and process vast quantities of data at significant speeds. Pure Storage's FlashBlade is an ideal match for NVIDIA's DGX and complements

NVIDIA gives back to communities through Project Inspire

its deep learning performance.

AI pushes beyond the limits of what's possible with traditional storage technologies. The velocity in which DGX-1 consumes data is unprecedented. The level of parallelism required by deep neural networks and GPUs continues to grow rapidly. A new class of data system was needed. FlashBlade is industry's first data platform purpose-built for AI and deep learning, engineered with a massively parallel architecture from end-to-end.
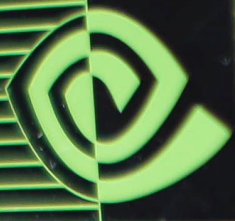
"By using Pure Storages' FlashBlade with NVIDIA's DGX-1, data scientists can enjoy the performance they need when working on AI," explains Petter. "We designed FlashBlade specifically for AI and machine learning applications – and it shows."

Combining the FlashBlade system with DGX-1 means that the GPUs can be continuously and efficiently fed with the large amount of data they need in order to build smarter AI solutions.

"The DGX-1 represents a real breakthrough in technology built for artificial intelligence," states Jackson. "As we continue to push the boundaries of what is possible, we'll continue creating tools which bring AI to the enterprise in ways that enable innovation and drive growth." ■