



# **Accelerating the speed and accessibility of artificial intelligence technologies**

A path to deeper insight

# Contents

The evolution of artificial intelligence .....	3
Deeper insight, innovation, and intelligence .....	4
Providing the foundation for AI success.....	4
An ecosystem of strategic partnerships.....	5
Conclusion .....	5

The ability of computers to autonomously learn, predict, and adapt using massive datasets is driving innovation and competitive advantage across many industries and applications. These artificial intelligence (AI) technologies are growing faster and more accessible, prompting businesses to hop aboard this next big wave of computing to uncover deeper insight, quickly resolve their most difficult problems, and differentiate their products and services. Whether the goal is to build a smarter city, power an intelligent car, or deliver personalized medicine, we've only just begun to understand the true potential of artificial intelligence.

One of the most exciting prospects of AI is the potential to solve tremendously complex problems that until recently have extended beyond the capabilities of compute. Addressing large social challenges in sectors such as agriculture, healthcare, transportation, and education has historically been limited by the abilities of human beings. However, the growth of [Big Data](#) and advancements in high-performance computing (HPC) have changed that. Advances in the field of AI have served to increase the intelligence of machines rapidly, improving their ability to recognize patterns, iteratively learn from data, and improve accuracy through experience. As computer-led implementations continue to surpass human-led ones, we will have the unprecedented opportunity to leverage compute to solve the world's largest problems and drive higher levels of efficiency across all facets of life.

Use cases of [machine learning](#), a key tool in the AI toolbox, are already beginning to proliferate across a variety of industries and applications. For example, in algorithmic trading supercomputers execute hundreds of trades each millisecond based on complex financial algorithms—machines are in charge and notify humans of the outcome later. This trend is giving birth to a new class of [Machine Intelligence-Led Services](#), where processes normally dominated by humans are being gradually augmented, and in some cases, automated by machines. As computers become increasingly capable to learn freely, reason, and determine the best course of action in real time, businesses stand to reap the rewards.

## The evolution of artificial intelligence

Solving 99 percent of today's most computationally intensive problems requires an **outside-in** approach, meaning that a period of observation and data collection must take place before understanding can begin. As a lack of insight renders writing fundamental analytical equations virtually impossible, we first treat the problem as a **black box** and just observe it. This outside-in approach relies on large amounts of data and sophisticated computing techniques to succeed; however, this is no longer a problem due to Big Data and HPC technologies supplying us with sufficient quantities of both. Enabling technologies for the industry vision of artificial intelligence go hand-in-hand with this approach, as they rely on machines sifting through massive data volumes to understand and extract new features, deriving a data-driven accurate model of the decision making task at hand—one that gets better with time as more data gets crunched.

[Deep learning](#), a subset of machine learning, has recently demonstrated significantly superior performance to some traditional machine learning approaches. Deep learning utilizes a combination of multi-layer artificial neural networks, data- and compute-intensive training, inspired by our latest understanding of human brain behavior. This approach has become so effective it's even begun to surpass human abilities in many areas, such as image and speech recognition and natural language processing.

Deep learning and AI become even more accessible as our reliance on experts continues to fade. For example, in order for computer models to estimate the yield of a corn crop accurately, an agricultural expert would typically be needed to train the algorithm and define features that might be useful for prediction, such as soil quality or chlorophyll levels. Until recently, this feature engineering required time-intensive and deep domain expertise, but all of that is changing with deep learning. With only primitive labeling required now, machines can learn these features independently, and in many cases more accurately, as more data is introduced to the model.

With a long record of accomplishment of pushing the limits of technology innovation and leading major advancements in the computing industry, Intel® solutions provide the foundation for success in artificial intelligence. Intel processors already power over 97 percent of the servers supporting machine learning workloads today. The [Intel® Xeon® processor E5 family](#) is the most widely deployed processor for machine and deep learning inference, and the recently launched [Intel Xeon Phi processor family](#) delivers the scalable performance needed for machine and deep learning training. These processor solutions are specifically designed to help our customers leverage AI technologies to become leaders in their industries, build and test new applications for artificial intelligence, and discover a path to deeper business insight.

## Deeper insight, innovation, and intelligence

Today's data centers need to be efficient, agile, flexible, and capable of handling many types of workloads at the same time. Already the most widely used processor for inference, the Intel Xeon processor E5 family offers excellent performance, enhanced orchestration capabilities, and improved security. The Intel Xeon processor E5 family even allows businesses to quickly scale their artificial intelligence applications and reduce time-to-insight while leveraging their existing hardware infrastructure. This family of processors is the ideal choice for next-generation and cloud architectures, providing more options and greater flexibility as businesses grow and scale their infrastructure. The Intel Xeon processor E5 family delivers the high levels of versatility needed to handle diverse applications and workloads, allowing businesses to design next-generation data centers to support their analytics demands in the future. This flexible processor also delivers industry-leading inference throughput on an existing E5 infrastructure along with other data center workloads, resulting in lower TCO.

The Intel Xeon Phi Processor is the latest addition to the Intel Xeon family, gearing specifically for HPC and machine learning workloads. This new processor family is designed to enable data scientists to train complex machine algorithms faster and run a wider variety of workloads. Intel's first bootable host processor, the Intel Xeon Phi processor family is designed from the ground up to eliminate node bottlenecks and deliver the highly parallel performance to pave the path to deeper insight, innovation, and intelligence. The Intel Xeon Phi processor [features 72 processor cores](#), each with two Intel Advanced Vector Extensions 512 (Intel AVX-512) SIMD processing units for improved per-core floating-point performance. In fact, the new Intel Xeon Phi processor-based system can train or learn an AlexNet image classification system up to 2.3 times faster than a similarly configured system.<sup>1</sup>

The Intel Xeon Phi processor delivers training breakthroughs, without foregoing the benefits of general purpose computing. The ability to scale is crucial when processing large machine learning workloads. The Intel Xeon Phi processor is designed to scale rapidly and efficiently up to thousands of nodes as the computational needs of the business grow more complex and time-sensitive. This proven scalability helps businesses train deep neural networks up to fifty times faster by scaling to 128 nodes.<sup>2</sup> Memory capacity is another key differentiator, as Intel CPUs provide direct access to hundreds of gigabytes of memory with no PCIe performance lag, opening up a much larger set of AI workloads. Finally, the ability to run a wide variety of workloads, amortize development efforts across environments, and maintain code flexibility without vendor lock-in, is a wise investment decision.

## Providing the foundation for AI success

For decades, Intel has been driving innovation and insight through its industry-leading technologies and contributions to parallel computing and HPC. Intel's strategy is geared to continuing to bring to market a strong portfolio of products and reference solutions that deliver the best platform performance for both single-node and multi-node benchmarks or workloads. Intel also continues to refine its roadmap to ensure [leadership in the AI space](#), with teams across the company continually working to perfect hardware, software, and algorithms built for AI applications. This broad initiative will help improve the use of AI technologies, and make machine learning practical on today's most widely deployed and scalable CPU-based data center architecture.

While Intel is fundamentally well positioned to support the AI revolution, to truly lead the industry requires more. Intel is committed to working with the open source community to ensure existing and future frameworks (such as Caffe and Theano) and libraries (Intel Math Kernel Library—Deep Learning Neural Network, Intel Deep Learning SDK) are optimized for CPUs. Intel will continue to develop tools and resources to help our customers accelerate the training and development of artificial intelligence networks and reduce deployment complexity. In short, Intel is continuously striving to bring artificial intelligence technologies to a broader base of users and make machine learning broadly accessible for the first time.

<sup>1</sup> Up to 2.3x faster training per system claim based on AlexNet topology workload (batch size = 256) using a large image database running 4-nodes Intel Xeon Phi processor 7250 (16 GB, 1.4 GHz, 68 Cores) in Intel Server System LADMP2312KXX41, 96GB DDR4-2400 MHz, quad cluster mode, MCDRAM flat memory mode, Red Hat® Enterprise Linux® 6.7 (Santiago), 1.0 TB SATA drive WD1003FZEX-00MK2A0 System Disk, running Intel Optimized DNN Framework, Intel Optimized Caffe (source: [github.com/intelcaffe](https://github.com/intelcaffe)) training 1.33 billion images/day in 10.5 hours compared to 1-node host with four NVIDIA® "Maxwell" GPUs training 1.33 billion images/day in 25 hours (source: [slideshare.net/NVIDIA/gtc-2016-opening-keynote](https://slideshare.net/NVIDIA/gtc-2016-opening-keynote)).

<sup>2</sup> [software.intel.com/en-us/blogs/2016/06/20/how-xeon-phi-processors-benefit-machine-and-deep-learning-apps-frameworks](https://software.intel.com/en-us/blogs/2016/06/20/how-xeon-phi-processors-benefit-machine-and-deep-learning-apps-frameworks)

## An ecosystem of strategic partnerships

End-to-end computing platforms for artificial intelligence are highly complex, and can't be delivered piecemeal or by any one compute vendor. Strategic partnerships between leaders in the hardware and software markets are required to deliver a full suite of offerings optimized for artificial intelligence applications. When co-designing at the system level, everything matters—from the amount of latency, to the memory, to the fabric. Now more than ever, it's crucial for compute leaders to collaborate to deliver solutions that can enable customers to act on Big Data in real time.

In the hardware realm, server solutions from Hewlett Packard Enterprise are built to support the latest Intel processor families and help customers realize real-time performance for their AI workloads. The newly launched [HPE Apollo 6000 system](#) is a deep learning platform designed to deliver unprecedented performance and scale in a relatively small footprint. HPE has led the mainstream server market for years, and a commitment to supporting the latest processor technologies from Intel means that customers will benefit from deep learning intelligence at faster speeds than ever before.

## Conclusion

Artificial intelligence is a [constellation of technologies](#) that is rapidly transforming the way businesses operate, solve problems, and gain competitive advantage. As AI technologies become even faster and more accessible, the computing community will be positioned to help organizations achieve the desired levels of efficiency that are critically needed in order to resolve the world's most complex problems, and increase safety, productivity, and prosperity.

## Learn more at

[intel.com/machinelearning](http://intel.com/machinelearning)



Sign up for updates

---

© Copyright 2016 Hewlett Packard Enterprise Development LP. The information contained herein is subject to change without notice. The only warranties for Hewlett Packard Enterprise products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. Hewlett Packard Enterprise shall not be liable for technical or editorial errors or omissions contained herein.

Intel, Intel Xeon, and the Intel logo are trademarks of Intel Corporation in the U.S. and other countries. Red Hat is a registered trademark of Red Hat, Inc. in the United States and other countries. Linux is the registered trademark of Linus Torvalds in the U.S. and other countries. NVIDIA is a trademark and/or registered trademark of NVIDIA Corporation in the U.S. and other countries. All other third-party trademark(s) is/are property of their respective owner(s).

4AA6-8682ENW, November 2016

