No Longer Science Fiction, AI Is Transforming the Enterprise and the Data Center

By: Todd Coleman

Depending on your generation, you may associate Artificial Intelligence with HAL 9000, Arnold and Skynet, or the largely apocryphal story about Facebook shutting down an AI project after developers discovered that the chatbots they'd programmed had created their own unique language, which humans couldn't understand. The reality, though, is that AI is big business—and the trendline is only headed up.

According to MarketsandMarkets, the Artificial Intelligence (AI) business is expected to reach approximately $191 billion globally by 2025, up from roughly $16 billion in 2017 and increasing at a compound annual growth rate (CAGR) of nearly 37 percent. These figures clearly align with other surveys suggesting a bullish attitude among enterprise leaders regarding strategic investments in AI.

Deloitte recently interviewed more than a thousand IT and line-of-business executives from U.S.-based companies on the current state of their AI adoption initiatives. Among the findings was that more than eight out of ten enterprises are already seeing positive returns from their production-level AI projects, with telecommunications, technology, media and entertainment companies earning an average 20 percent or greater return on investment (ROI). Meanwhile, across the border to the north, more than 30 percent of Canadian businesses will introduce AI-based technologies into their operations by the end of this year, according to the International Data Corporation (IDC).

AI is a term used to describe technologies that enable computers to perceive, learn, reason and assist in decision-making to solve problems in ways similar to how human beings do. Common use cases range from natural language processing that enables globally dispersed employees to communicate across languages and borders, to far more complex applications such as autonomous vehicles. Notably, more than half of enterprise executives surveyed say their AI initiatives are needed to either edge slightly ahead or widen their lead against their competitors. The bottom 36 percent state that AI is enabling them to remain parallel to the competition or play catch up.

Hey Siri, Get Me an Uber

AI research began in earnest in the 1950s, based on work by British mathematician and computer scientist Alan Turing during World War II. Over the past decade, however, we have witnessed more rapid advances in AI due to the confluence of cloud computing, a tremendous surge in data volumes, and significant breakthroughs in machine learning algorithms.

Machine Learning (ML) is a method of data analysis that automates analytical model-building. Using algorithms that iteratively learn from data, ML allows computers to find hidden insights without being explicitly programmed on where to look for them. While many ML algorithms, too, have been around for a long time, the ability to automatically apply complex mathematical calculations to Big Data — rapidly, iteratively and repeatedly — is a recent development. Meanwhile, the unmanageable volume and complexity of Big Data that researchers and businesses now have access to has increased the potential of ML—and, increasingly, society's dependence on it.

Search recommendations, speech recognition, and email filtering are all examples of AI that leverage ML. When you search products while shopping online or browse Netflix or Hulu to plan your next binge-watch, the suggested results that follow you, even

across various platforms, are the work of ML algorithms. The voice recognition systems of Siri and Cortana are based on ML. And when the ride-hailing app Uber provides an ETA for when you should pull up to your destination, that's ML at work too.

AI That Emulates the Brain

Deep learning, a subset of ML, emulates the functions of the inner layers of the human brain, thereby creating knowledge from multiple layers of information processing. The present and near-future potential of deep learning applications is so immense that many experts believe it will soon become the dominant technology of the AI market.

In practice, deep learning involves acquiring learning or experience from hierarchical layers of discovery. The computer learns from each layer and then uses that learning in the next layer to learn more, until the learning reaches its full stage through cumulative learning in multiple layers. In this way, the system reaches a highly detailed understanding of the data that amounts to a form of intelligent reasoning. So, in perceiving a picture of an object, the machine will first detect a shape from a matrix of pixels. Then, it might identify the edges of that shape, then its contours, then the object itself, and so on, until it identifies the image.

In the healthcare sector, hospitals and clinics are now using deep learning technology that enables computers to read large numbers of X-rays, MRIs, and CT scans more rapidly than a radiologist. Researchers at Google have built a deep learning model that detects lung cancer as well as or better than human radiologists. Meanwhile, across the financial services sector, deep learning is being used to enhance fraud detection. By analyzing past customer spending patterns, the technology can identify and alert banks to anomalous activity in real time.

Advertising and marketing, agriculture, automotive and transportation, cybersecurity, drug testing, oil and gas exploration and production, and retail are among the industries that will benefit from advances in deep learning.

AI and the Data Center

At eStruxture, we are seeing substantial demand for capacity to support AI-powered applications that are being implemented by enterprises across a variety of verticals. AI labs and other associations are also flocking to multi-tenant data centers to deploy AI platforms with robust power and connectivity requirements that are comparatively richer than what we have previously witnessed.

AI developers and organizations already using AI-based applications do not fit easily within traditional Tier III data center models. When we talk about the applications that drive a lot of AI development and deep learning needs, we find that their requirements are different from a power, cooling and redundancy perspective. Where a standard customer might need three or five kilowatts in a rack, an AI provider could require 10, 15, 20, or even 30 kilowatts in a rack. Additionally, due to the size of the compute workloads and power density requirements of both ML and deep learning environments, servers can run approximately 30 percent hotter, presenting challenges in managing the power and thermal demands of the equipment.

With respect to issues surrounding power and cooling, and for a host of other advantages, AI providers seeking the optimal environment for hosting their applications would do well to consider colocating in Montreal. With its long winters and temperate summers, maintaining proper server temperature is easier and more cost-effective. In data centers, approximately 40 percent of the total energy consumed goes toward cooling IT equipment. Hence, cooling costs are one of the major contributors to the total electricity bill of large data centers. By taking advantage of the naturally cold climate to cool their infrastructure, thereby reducing operating expenses, data centers and colocation facilities in Montreal use less power than their counterparts to the South.

Enterprises and cloud service providers are attracted to colocating in Montreal because of the city's favorable power rates. Compared to all of the major North

American markets, Montreal offers the lowest-cost power for data center operations on the continent, and nearly 100 percent of the energy generated by the local utility, Hydro-Quebec, comes from hydroelectricity. A clean form of energy, hydropower produces greenhouse gas emissions that are 50 times lower than natural gas, five times lower than solar power and about equal to wind power. Montreal is already home to a thriving tech sector that includes AI, digital media, and managed IT services providers. These types of companies depend on high availability, high bandwidth capacity, and ultra-secure facilities to get their offerings to market quickly and efficiently.

While there's much that data center operators can do to meet the needs of customers involved in AI, the emerging technology also has the potential to transform the management of these mission-critical facilities themselves. For example, one program that utilized an AI unit to manage power usage in a cloud provider's data centers achieved a 40 percent reduction in the amount of electricity needed for cooling across its facilities. In another application of deep learning, an enterprise facility planned the most efficient methods of cooling by analyzing data from sensors strategically located among the server racks, including information on inlet temperatures and cooling pump speeds.

AI and deep learning have also been found to be useful in server optimization and load balancing. In fact, Gartner predicts that by 2020, 30 percent of data centers that fail to apply AI and ML effectively in support of enterprise business will cease to be operationally and economically viable.

All this is to say: AI is here to stay. Regardless of the dire, fictionalized prognostications or misperceptions you might associate with the technology, the reality is that AI is now affecting the way we do business, interact, and provide care for one another. If managed judiciously today and tomorrow, AI will remain a force for good that far outweighs any possibilities of harm.