AI and big data: The changing investment landscape

January 2019

What this means to investors:

| AI and big data have achieved mainstream viability for a growing list of business applications in nearly every industry. | This paper provides background, definitions, and examples of how these technologies may produce durable competitive advantages. | Investors should build awareness of these technologies and consider their strategic implications for portfolio candidates. |

Interest in artificial intelligence (AI) and big data has gathered steam in recent years, fanning speculation about how these technologies might deliver efficiencies, unlock hidden value, or even create entirely new business models. Is it hype, or has the future finally arrived for their mass adoption? Digital transformation strategies have gone mainstream in corporate America, backed by a projected $1.1 trillion in global spending in 2018, and accelerating to $1.8 trillion by 2021.1 AI and big data applications are often the centerpiece of these efforts, bringing them into view of investors in nearly every industry.

This paper offers a brief explanation of why these technologies have only recently become commercially viable after an extended incubation period. It provides definitions of common terms, a few examples of practical applications, and some context on how various categories of AI and big data are related. We think these technologies will be highly disruptive and that investors can benefit by proactively exploring the threats and opportunities they will inevitably present for each of their portfolio holdings going forward.

Background

In 1965, Gordon Moore, founder of Intel Corporation, observed that the number of transistors that could be fitted to an integrated circuit doubled every year since its invention and predicted that future growth would follow a similar trajectory.2 Some 50+ years on, the actual rate of doubling has slowed to an 18- to 24-month window. But, even the most conservative application of Moore’s law would approximate that

---

2. The growth rate in processing power is now commonly known as Moore’s law.
computing power doubled at least 26 times since 1965, making the efficiency of a modern computer chip tens of millions of times greater today. Assuming that unit costs of semiconductor production have remained roughly constant,\(^3\) we can use Moore’s law to plot the rapid depreciation in the price of $1 million of computing power in 1998 to the present day.\(^4\) The remarkable plunge of supercomputing costs has made its application viable for an ever-expanding list of commercial uses. What can we do with all that computing power? The answer emerging for many businesses is data analysis.

**Figure 1: Cheap, plentiful, and accessible computing power**

![Figure 1: Cheap, plentiful, and accessible computing power](image)

Data is the feedstock of computer algorithms and its generation has mushroomed in the age of digital networking. In 1998, internet business models based on B2B and B2C\(^5\) interactions became household jargon, yet the volume of business conducted this way and the data it produced were still in their infancy, and social media had not yet been conceived in its modern form. An International Data Corporation study estimated that all of the world’s hard drives in 2006 constituted 0.16 zettabytes of data.\(^6\) By 2025, the global datasphere is expected to reach 163 zettabytes,\(^7\) becoming more than 1,000 times larger in a span of just 20 years, roughly tracking the historical growth rate of processing power.

The rapid buildout of a network ecosystem of human to human, machine to human, and machine to machine interactions is producing more data than businesses know what to do with. Keeping up with and deriving value from this explosion is the realm of computer science aptly named big data.

---

3. The costs of R&D and of semiconductor fabrication plants have increased but so has production efficiency, allowing higher fixed costs to be spread across more units.
4. The calculation is $1 million\(\times 2^{-N/2}\), where N is the number of years in the measurement period. This ignores the impact of inflation.
5. B2B and B2C are acronyms for business-to-business and business-to-consumer, respectively.
Big data

Big data seeks to extrapolate insights from large sets of raw data to improve decision-making in ways that supplement traditional data analysis techniques. There are a number of challenges in dealing with the raw data itself, even before any insights are drawn. Practitioners have settled on three generic issues, called the three Vs:

**THREE Vs**

- **Volume:** Much of the data that can be recorded will end up having little value. To make that determination, a lot of data has to be collected, stored, and retrieved for analysis. Data warehouses capture and store mostly structured data, which can be arranged into traditional database structures. Data lakes have entered the conversation more recently. These act as floodplains for pools of data that are so large that they typically bypass any preprocessing and are held in their native states.

- **Velocity:** The rate at which data are generated, or the flow, is increasing rapidly. Sensors are being connected to nearly everything, from consumer products to business machines (all the nodes that make up the internet of things), producing a deluge of telemetry data. Streaming data, the so-called firehose, must be tagged, metered, and filtered for harmful content to be safely and reliably captured.

- **Variety:** Digital data is composed of 1s and 0s at the foundational level but can be constructed in a variety of different formats. Most data captured or accessed by businesses in the past have been recorded, structured, and often sold by subscription by various vendors. Now, unstructured data make up roughly 80% of captured data and will likely represent a growing proportion. Examples include all forms of multimedia, satellite imagery, email, and earnings conference calls, to name a few. This data must be somehow read and synthesized to produce useful information.

Some investment opportunities are clear. The most obvious involve companies that are directly addressing challenges surrounding the three Vs, often in the technology, media, and telecommunications sectors. Cloud service providers have emerged to help companies outsource the various tasks of building an analytics platform to capture, store, retrieve, process, and generate insights from data. These companies offer a stack of services that range from infrastructure as a service (which may include access to computing or storage among other related services), platform as a service (for software development), and software as a service (for deployment). Investors may have a view on these providers or may ask if companies that are developing analytics programs have made the best make-or-buy decisions for their specific circumstances.

Proprietary data is an increasingly important competitive advantage as the tools to capture, store, and process data are becoming widespread. Many companies are sitting on a trove of legacy data from their operations that can be gleaned and monetized. Others are actively seeking to build out datasets by strategic expansion into new areas. Investors can assess the value of proprietary datasets or consider the option value hidden within the development of experimental datasets. *The cost of seeking breakthrough insights from large datasets is so cheap now that it is viable for low success rate projects.*

---

Real-time information about oil inventories is often guarded by producers and consumers alike. Satellite imagery can be used to estimate oil tank capacity, but this alone does not provide an accurate estimate of current volume. Tanks have floating tops to prevent dangerous buildup of gases in the oil compartment. Satellites are now exploiting this fact by capturing the relative lengths of the sun’s shadows cast inside and outside the tank, allowing for an estimate of current volume (for example, the sun will cast almost no internal shadow when the tank is filled near capacity). Data from tank farms around the world can be aggregated to estimate changes in supply and demand.

This information has already become commercialized through subscription, but it is easy to imagine there are countless other ways to use high-resolution imagery from outer space, or from any type of sensor data, to find unique insights and sustain an informational advantage.

Using vast computing resources to spot useful information that is hidden within data is the natural intersection of big data and AI—advancements in both are related and mutually reinforcing. Because AI has applications outside of big data, we focus the remainder of our discussion on the branches of AI that are most closely related to big data. First, we provide a basic definition of AI.
**Artificial intelligence**

Generally, AI seeks to mimic humanlike abilities to perform tasks, including the ability to learn and reason based on a set of observations and experience. It is a form of automation but differs from more familiar types of automation based on static programming (like robots welding on an assembly line) because AI involves a feedback loop to improve performance over time. Like a human, it “learns” from experience and adjusts its output as a result. People define tasks, interpret results, and ask new questions to guide model development.

![Feedback loop](image)

Modern AI has its roots in neural networks that were initially conceived in the 1950s. These seek to mimic the behavior of brain neurons in which a stimulus is captured and travels directionally through a series of nodes to arrive at some kind of recognition, or in the case of an algorithm, a decision. One aspect of learning occurs through backpropagation: The choice the model selected in an iteration is compared with the correct choice (for example, did the model choose cat or dog when digitally processing a picture of a dog?), and then related pathways that led to the decision are weakened or strengthened to improve future performance. More data is supplied to “train” the model and refine the decision nodes. This iterative process is the basic foundation of specialized subsets of AI that we hear about today, including machine learning and deep learning.

![A simple neural network](image)
Machine learning

Machine learning uses statistical techniques to train a model how to program itself autonomously from raw data or examples rather than direct human programming. This can involve massive amounts of data. Continuing with our neural network example, multiple pictures of cats and dogs are fed to the model and processed so that algorithms autonomously pick out distinguishing features that produce a correct choice. The promise behind this autonomy is that new algorithms may reveal differences that are hidden to humans. They also may highlight human biases, which have numerous investment applications.

Deep learning

Deep learning is a subset of machine learning that uses networks with multiple layers, each responsible for a separate processing function that adapts to improve the model’s accuracy over time (in contrast, our simple example has only one layer). Clearly, the computing power available today is required for complex models. In the appendix, we discuss how the development of self-driving cars may be the quintessential, modern application of deep learning, combining elements of computer vision and natural language processing interfaces with vehicle occupants.

The truly game-changing nature of these data-intensive AI technologies is clear in self-driving cars. Elsewhere the implications are less obvious, but perhaps no less disruptive. For this reason, we think investors should proactively consider their ultimate impact in their assessments of potential portfolio candidates.

Closing thoughts

AI and, more recently, big data have been gestating over a number of decades and have become truly viable for a broad set of commercial applications in just the past few years. Our discussion above serves only to evoke consideration of the impact these technologies will ultimately have on the ways business will be conducted in the future. A number of companies are asking the right questions and deploying resources to these technologies to pursue competitive advantages. We consider this a developing theme that will weigh more heavily into assessments of the strategic vision of management and the direction (or fate) of companies that they lead.

Adept asset managers are also training their sights on their own businesses. In a companion paper, we survey the asset management landscape from a high level and discuss the basic determinants of value creation among various generic investment styles. Because AI and big data are generally aimed at improving operational efficiencies, scale, and insight, we conclude that the benefits of these tools will ultimately accrue to active investment approaches. We go on to discuss some tools we are developing to seek a leading position among our peers.
Appendix:
Self-driving technology is a high-stakes machine learning arms race

Autonomous driving technology may be the quintessential, modern application of deep learning. Tesla, Waymo, Uber, General Motors Cruise, and NVIDIA are some of the most recognized names that are pursuing different paths to deploy commercial technologies in this space. Differences in their approaches can be described along a number of dimensions:

• Types of telemetry data recorded by multiple redundant sensor systems on the vehicle (for example, laser-based radar [LiDAR] versus high-res camera versus radar, etc).

• Number and types of test miles driven to train algorithms (fully autonomous, semiautonomous, or simulation)

• Emphasis on the types and combinations of algorithms deployed for object identification, classification, and prediction (regression, pattern recognition, clustering, decision matrix)

• Level of driving autonomy, measured by SAE Federal Highway Research Institute as levels 0 through 5 (driver assistance, partial autonomy, full autonomy)

Breakthrough advances toward safety, reliability, and affordability may provide a first-mover advantage in an industry expected to achieve $2.8 trillion in revenue globally by 2030 according to UBS. It is likely that multiple supporting actors will stake a profitable claim within the emerging transportation-as-a-service industry.

Figure 6: Industry standard autonomous driving levels

SOCIETY OF AUTOMOTIVE ENGINEERS (SAE) AUTOMOTIVE LEVELS

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No automation; the driver performs all driving tasks.</td>
</tr>
<tr>
<td>1</td>
<td>Driver assistance</td>
</tr>
<tr>
<td>2</td>
<td>Partial automation</td>
</tr>
<tr>
<td>3</td>
<td>Conditional automation</td>
</tr>
<tr>
<td>4</td>
<td>High automation</td>
</tr>
<tr>
<td>5</td>
<td>Full automation</td>
</tr>
</tbody>
</table>


We want to help clients build for successful outcomes, defend portfolios against uncertainty, and create long-term financial well-being. To learn more, investment professionals can contact us:

- To reach our U.S.-based investment professionals, contact your existing client relations director, or contact us at WFAMInstitutional@wellsfargo.com.

- To reach our U.S.-based intermediary sales professionals, contact your dedicated regional director, or call us at 1-888-877-9275.

- To reach our U.S.-based retirement professionals, contact Nathaniel Miles, head of Defined Contribution at Wells Fargo Asset Management, at nathaniel.s.miles@wellsfargo.com.

- To reach our international investment professionals, contact your regional client relations or sales director, or contact Ben Foley at either ben.foley@wellsfargo.com or +44 20 7518 2947.

- To discuss environmental, social and governance (ESG) investing solutions, contact Hannah Skeates, global head of ESG at Wells Fargo Asset Management, at hannah.skeates@wellsfargo.com.