

**BY ERIK BRYNJOLFSSON AND ANDREW MCAFEE**

**LATEST JUL 21**

How AI Fits into Your  
Science Team

DIGITAL ARTICLE

**READ** 

What it can — and cannot — do for your organization

 **THE LINEUP**

**F**or more than 250 years the fundamental drivers of economic growth have been technological innovations. The most important of these are what economists call general-purpose technologies – a category that includes the steam engine, electricity, and the internal combustion engine. Each one catalyzed waves of complementary innovations and opportunities. The internal combustion engine, for example, gave rise to cars, trucks, airplanes, chain saws, and lawnmowers, along with big-box retailers, shopping centers, cross-docking warehouses, new supply chains, and, when you think about it, suburbs.

Companies as diverse as Walmart, UPS, and Uber found ways to leverage the technology to create profitable new business models.

The most important general-purpose technology of our era is artificial intelligence, particularly machine learning (ML) – that is, the machine’s ability to keep improving its performance without humans having to explain exactly how to accomplish all the tasks it’s given. Within just the past few years machine learning has become far more effective and widely available. We can now build systems that learn how to perform tasks on their own.

Why is this such a big deal? Two reasons. First, we humans know more than we can tell: We can’t explain exactly how we’re able to do a lot of things – from recognizing a face to making a smart move in the ancient Asian strategy game of Go. Prior to ML, this inability to articulate our own knowledge meant that we couldn’t automate many tasks. Now we can.

Second, ML systems are often excellent learners. They can achieve superhuman performance in a wide range of activities, including detecting fraud and diagnosing disease. Excellent digital learners are being deployed across the economy, and their impact will be profound.

---

THE AUTHORS

# ERIK BRYNJOLFSSON AND ANDREW MCAFEE



Erik Brynjolfsson (@erikbryn) is the director of MIT's Initiative on the Digital Economy, the Schussel Family Professor of Management Science at the MIT Sloan School of Management, and a research associate at NBER. His research examines the effects of information technologies on business strategy, productivity and performance, digital commerce, and intangible assets. At MIT he teaches courses on the economics of information and the Analytics Lab.



---

In the sphere of business, AI is poised have a transformational impact, on the scale of earlier general-purpose technologies. Although it is already in use in thousands of companies around the world, most big opportunities have not yet been tapped. The effects of AI will be magnified in the coming decade, as manufacturing, retailing, transportation, finance, health care, law, advertising, insurance, entertainment, education, and virtually every other industry transform their core processes and business models to take advantage of machine learning. The bottleneck now is in management, implementation, and business imagination.

Like so many other new technologies, however, AI has generated lots of unrealistic expectations. We see business plans liberally sprinkled with references to machine learning, neural nets, and other forms of the technology, with little connection to its real capabilities. Simply calling a dating site “AI-powered,” for example, doesn’t make it any more effective, but it might help with fundraising. This article will cut through the noise to describe the real potential of AI, its practical implications, and the barriers to its adoption.

## WHAT CAN AI DO TODAY?

The term *artificial intelligence* was coined in 1955 by John McCarthy, a math professor at Dartmouth who organized the seminal conference on the topic the following year. Ever since, perhaps in part because of its evocative name, the field has given rise to more than its share of fantastic claims and promises. In 1957 the economist Herbert Simon predicted that computers would beat humans at chess within 10 years. (It took 40.) In 1967 the cognitive scientist Marvin Minsky said, “Within a generation the problem of creating ‘artificial intelligence’ will be substantially solved.” Simon and Minsky were both intellectual giants, but they erred badly. Thus it’s understandable that dramatic claims about future breakthroughs meet with a certain amount of skepticism.

Let’s start by exploring what AI is already doing and how quickly it is improving. The biggest advances have been in two broad areas: perception and cognition. In the former category some of the most practical advances have been made in relation to speech. Voice recognition is still far from perfect, but millions of people are now using it – think Siri, Alexa, and Google Assistant. The text you are now reading was originally dictated to a computer and transcribed with sufficient accuracy to make it faster than typing. A study by the Stanford computer scientist James Landay and colleagues found that speech recognition is now about three times as fast, on average, as typing on a cell phone. The error rate, once 8.5%, has dropped to 4.9%. What’s striking is that this substantial improvement has come not over the past 10 years but just since the summer of 2016.

# ALTHOUGH AI IS ALREADY IN USE IN THOUSANDS OF COMPANIES AROUND THE WORLD, MOST BIG OPPORTUNITIES HAVE NOT YET BEEN TAPPED.

Image recognition, too, has improved dramatically. You may have noticed that Facebook and other apps now recognize many of your friends' faces in posted photos and prompt you to tag them with their names. An app running on your smartphone will recognize virtually any bird in the wild. Image recognition is even replacing ID cards at corporate headquarters. Vision systems, such as those used in self-driving cars, formerly made a mistake when identifying a pedestrian as often as once in 30 frames (the cameras in these systems record about 30 frames a second); now they err less often than once in 30 million frames. The error rate for recognizing images from a large database called ImageNet, with several million photographs of common, obscure, or downright weird images, fell from higher than 30% in 2010 to about 4% in 2016 for the best systems. (See the exhibit "Puppy or Muffin?")

The speed of improvement has accelerated rapidly in recent years as a new approach, based on very large or "deep" neural nets, was adopted. The ML approach for vision systems is still far from flawless – but even people have trouble quickly recognizing puppies' faces or, more embarrassingly, see their cute faces where none exist.

---

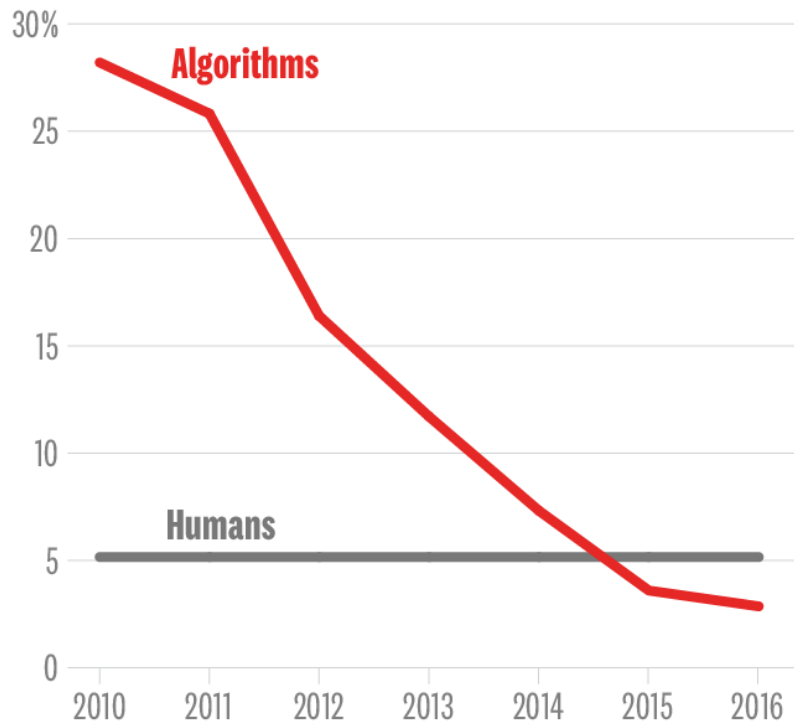
## **Puppy or Muffin? Progress in Image Recognition**

Machines have made real strides in distinguishing among similar-looking categories of images.



Karen Zack/@teenybiscuit

## VISION ERROR RATE



SOURCE ELECTRONIC FRONTIER FOUNDATION © HBR.ORG

The second type of major improvement has been in cognition and problem solving. Machines have already beaten the finest (human) players of poker and Go – achievements that experts had predicted would take at least another decade. Google’s DeepMind team has used ML systems to improve the cooling efficiency at data centers by more than 15%, even after they were optimized by human experts. Intelligent agents are being used by the cybersecurity company Deep Instinct to detect malware, and by PayPal to prevent money laundering. A system using IBM technology automates the claims process at an insurance company in Singapore, and a system from Lumidatum, a data science platform firm, offers timely advice to improve customer support. Dozens of companies are using ML to decide which trades to execute on Wall Street, and more and more credit decisions are made with its help. Amazon employs ML to optimize inventory and improve product recommendations to customers. Infinite Analytics developed one ML system to predict whether a user would click on a particular ad, improving online ad placement for a global consumer packaged goods company,

and another to improve customers' search and discovery process at a Brazilian online retailer. The first system increased advertising ROI threefold, and the second resulted in a \$125 million increase in annual revenue.

Machine learning systems are not only replacing older algorithms in many applications, but are now superior at many tasks that were once done best by humans. Although the systems are far from perfect, their error rate – about 5% – on the ImageNet database is at or better than human-level performance. Voice recognition, too, even in noisy environments, is now nearly equal to human performance. Reaching this threshold opens up vast new possibilities for transforming the workplace and the economy. Once AI-based systems surpass human performance at a given task, they are much likelier to spread quickly. For instance, Aptonomy and Sanbot, makers respectively of drones and robots, are using improved vision systems to automate much of the work of security guards. The software company Affectiva, among others, is using them to recognize emotions such as joy, surprise, and anger in focus groups. And Enlitic is one of several deep-learning startups that use them to scan medical images to help diagnose cancer.

These are impressive achievements, but the applicability of AI-based systems is still quite narrow. For instance, their remarkable performance on the ImageNet database, even with its millions of images, doesn't always translate into similar success "in the wild," where lighting conditions, angles, image resolution, and context may be very different. More fundamentally, we can marvel at a system that understands Chinese speech and translates it into English, but we don't expect such a system to know what a particular Chinese character means – let alone where to eat in Beijing. If someone performs a task well, it's natural to assume that the person has some competence in related tasks. But ML systems are trained to do specific tasks, and typically their knowledge does not generalize. The fallacy that a computer's narrow understanding implies broader understanding is perhaps the biggest source of confusion, and exaggerated claims, about AI's progress. We are far from machines that exhibit general intelligence across diverse domains.

## **UNDERSTANDING MACHINE LEARNING**



The most important thing to understand about ML is that it represents a fundamentally different approach to creating software: The machine learns from examples, rather than being explicitly programmed for a particular outcome. This is an important break from previous practice. For most of the past 50 years, advances in information technology and its applications have focused on codifying existing knowledge and procedures and embedding them in machines. Indeed, the term “coding” denotes the painstaking process of transferring knowledge from developers’ heads into a form that machines can understand and execute. This approach has a fundamental weakness: Much of the knowledge we all have is tacit, meaning that we can’t fully explain it. It’s nearly impossible for us to write down instructions that would enable another person to learn how to ride a bike or to recognize a friend’s face.



In other words, we all know more than we can tell. This fact is so important that it has a name: Polanyi’s Paradox, for the philosopher and polymath Michael Polanyi, who described it in 1964. Polanyi’s Paradox not only limits what we can tell one another but has historically placed

a fundamental restriction on our ability to endow machines with intelligence. For a long time that limited the activities that machines could productively perform in the economy.

Machine learning is overcoming those limits. In this second wave of the second machine age, machines built by humans are learning from examples and using structured feedback to solve on their own problems such as Polanyi's classic one of recognizing a face.

## DIFFERENT FLAVORS OF MACHINE LEARNING

Artificial intelligence and machine learning come in many flavors, but most of the successes in recent years have been in one category: supervised learning systems, in which the machine is given lots of examples of the correct answer to a particular problem. This process almost always involves mapping from a set of inputs, X, to a set of outputs, Y. For instance, the inputs might be pictures of various animals, and the correct outputs might be labels for those animals: dog, cat, horse. The inputs could also be waveforms from a sound recording and the outputs could be words: "yes," "no," "hello," "good-bye." (See the exhibit "Supervised Learning Systems.")

Successful systems often use a training set of data with thousands or even millions of examples, each of which has been labeled with the correct answer. The system can then be let loose to look at new examples. If the training has gone well, the system will predict answers with a high rate of accuracy.

---

### Supervised Learning Systems

As two pioneers in the field, Tom Mitchell and Michael I. Jordan, have noted, most of the recent progress in machine learning involves mapping from a set of inputs to a set of outputs. Some examples:

INPUT X	OUTPUT Y	APPLICATION
Voice recording	Transcript	Speech recognition
Historical market data	Future market data	Trading bots

Photograph	Caption	Image tagging
Drug chemical properties	Treatment efficacy	Pharma R&D
Store transaction details	Is the transaction fraudulent?	Fraud detection
Recipe ingredients	Customer reviews	Food recommendations
Purchase histories	Future purchase behavior	Customer retention
Car locations and speed	Traffic flow	Traffic lights
Faces	Names	Face recognition

The algorithms that have driven much of this success depend on an approach called *deep learning*, which uses neural networks. Deep learning algorithms have a significant advantage over earlier generations of ML algorithms: They can make better use of much larger data sets. The old systems would improve as the number of examples in the training data grew, but only up to a point, after which additional data didn't lead to better predictions. According to Andrew Ng, one of the giants of the field, deep neural nets don't seem to level off in this way: More data leads to better and better predictions. Some very large systems are trained by using 36 million examples or more. Of course, working with extremely large data sets requires more and more processing power, which is one reason the very big systems are often run on supercomputers or specialized computer architectures.

Any situation in which you have a lot of data on behavior and are trying to predict an outcome is a potential application for supervised learning systems. Jeff Wilke, who leads Amazon's consumer business, says that supervised learning systems have largely replaced the memory-based filtering algorithms that were used to make personalized recommendations to customers. In other cases, classic algorithms for setting inventory levels and optimizing supply chains have been replaced by more efficient and robust systems based on machine learning.

JPMorgan Chase introduced a system for reviewing commercial loan contracts; work that used to take loan officers 360,000 hours can now be done in a few seconds. And supervised learning systems are now being used to diagnose skin cancer. These are just a few examples.

It's comparatively straightforward to label a body of data and use it to train a supervised learner; that's why supervised ML systems are more common than *unsupervised* ones, at least for now. Unsupervised learning systems seek to learn on their own. We humans are excellent unsupervised learners: We pick up most of our knowledge of the world (such as how to recognize a tree) with little or no labeled data. But it is exceedingly difficult to develop a successful machine learning system that works this way.

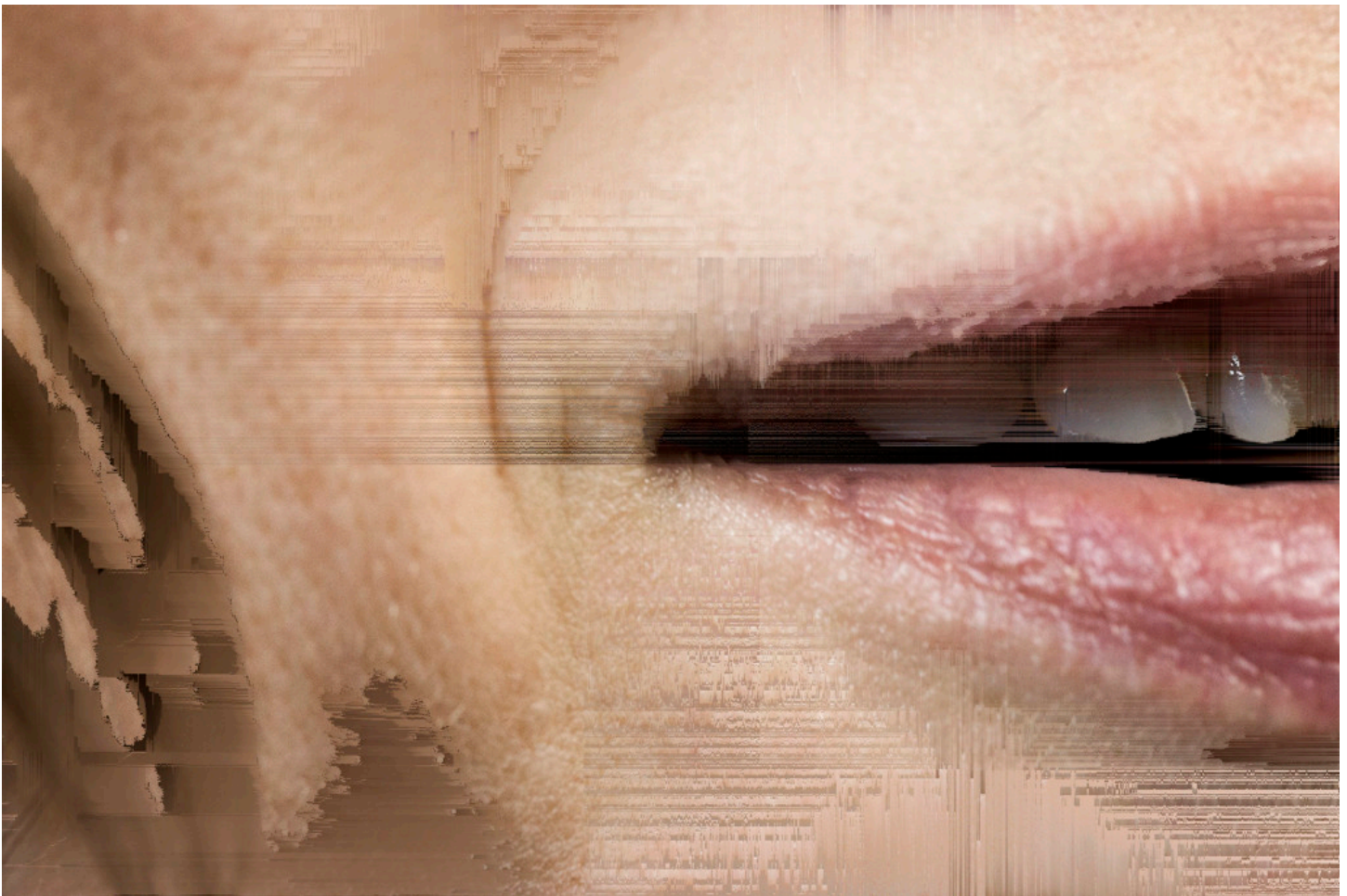
If and when we learn to build robust unsupervised learners, exciting possibilities will open up. These machines could look at complex problems in fresh ways to help us discover patterns – in the spread of diseases, in price moves across securities in a market, in customers' purchase behaviors, and so on – that we are currently unaware of. Such possibilities lead Yann LeCun, the head of AI research at Facebook and a professor at NYU, to compare supervised learning systems to the frosting on the cake and unsupervised learning to the cake itself.

ONCE AI-BASED SYSTEMS SURPASS HUMAN  
PERFORMANCE AT A GIVEN TASK, THEY ARE MUCH  
LIKELIER TO SPREAD QUICKLY.

Another small but growing area within the field is *reinforcement learning*. This approach is embedded in systems that have mastered Atari video games and board games like Go. It is also helping to optimize data center power usage and to develop trading strategies for the stock market. Robots created by Kindred use machine learning to identify and sort objects they've never encountered before, speeding up the "pick and place" process in distribution centers for



consumer goods. In reinforcement learning systems the programmer specifies the current state of the system and the goal, lists allowable actions, and describes the elements of the environment that constrain the outcomes for each of those actions. Using the allowable actions, the system has to figure out how to get as close to the goal as possible. These systems work well when humans can specify the goal but not necessarily how to get there. For instance, Microsoft used reinforcement learning to select headlines for MSN.com news stories by “rewarding” the system with a higher score when more visitors clicked on the link. The system tried to maximize its score on the basis of the rules its designers gave it. Of course, this means that a reinforcement learning system will optimize for the goal you explicitly reward, not necessarily the goal you really care about (such as lifetime customer value), so specifying the goal correctly and clearly is critical.



## **PUTTING MACHINE LEARNING TO WORK**

There are three pieces of good news for organizations looking to put ML to use today. First, AI skills are spreading quickly. The world still has not nearly enough data scientists and machine learning experts, but the demand for them is being met by online educational resources as well as by universities. The best of these, including Udacity, Coursera, and fast.ai, do much more than teach introductory concepts; they can actually get smart, motivated students to the point of being able to create industrial-grade ML deployments. In addition to training their own people, interested companies can use online talent platforms such as Upwork, Topcoder, and Kaggle to find ML experts with verifiable expertise.

The second welcome development is that the necessary algorithms and hardware for modern AI can be bought or rented as needed. Google, Amazon, Microsoft, Salesforce, and other companies are making powerful ML infrastructure available via the cloud. The cutthroat competition among these rivals means that companies that want to experiment with or deploy ML will see more and more capabilities available at ever-lower prices over time.

The final piece of good news, and probably the most underappreciated, is that you may not need all that much data to start making productive use of ML. The performance of most machine learning systems improves as they're given more data to work with, so it seems logical to conclude that the company with the most data will win. That might be the case if "win" means "dominate the global market for a single application such as ad targeting or speech recognition." But if success is defined instead as significantly improving performance, then sufficient data is often surprisingly easy to obtain.

For example, Udacity cofounder Sebastian Thrun noticed that some of his salespeople were much more effective than others when replying to inbound queries in a chat room. Thrun and his graduate student Zayd Enam realized that their chat room logs were essentially a set of labeled training data – exactly what a supervised learning system needs. Interactions that led to a sale were labeled successes, and all others were labeled failures. Zayd used the data to predict what answers successful salespeople were likely to give in response to certain very

common inquiries and then shared those predictions with the other salespeople to nudge them toward better performance. After 1,000 training cycles, the salespeople had increased their effectiveness by 54% and were able to serve twice as many customers at a time.

The AI startup WorkFusion takes a similar approach. It works with companies to bring higher levels of automation to back-office processes such as paying international invoices and settling large trades between financial institutions. The reason these processes haven't been automated yet is that they're complicated; relevant information isn't always presented the same way every time ("How do we know what currency they're talking about?"), and some interpretation and judgment are necessary. WorkFusion's software watches in the background as people do their work and uses their actions as training data for the cognitive task of classification ("This invoice is in dollars. This one is in yen. This one is in euros..."). Once the system is confident enough in its classifications, it takes over the process.

Machine learning is driving changes at three levels: tasks and occupations, business processes, and business models. An example of task-and-occupation redesign is the use of machine vision systems to identify potential cancer cells – freeing up radiologists to focus on truly critical cases, to communicate with patients, and to coordinate with other physicians. An example of process redesign is the reinvention of the workflow and layout of Amazon fulfillment centers after the introduction of robots and optimization algorithms based on machine learning. Similarly, business models need to be rethought to take advantage of ML systems that can intelligently recommend music or movies in a personalized way. Instead of selling songs à la carte on the basis of consumer choices, a better model might offer a subscription to a personalized station that predicted and played music a particular customer would like, even if the person had never heard it before.

Note that machine learning systems hardly ever replace the entire job, process, or business model. Most often they complement human activities, which can make their work ever more valuable. The most effective rule for the new division of labor is rarely, if ever, "give all tasks to the machine." Instead, if the successful completion of a process requires 10 steps, one or two of them may become automated while the rest become more valuable for humans to do. For

instance, the chat room sales support system at Udacity didn't try to build a bot that could take over all the conversations; rather, it advised human salespeople about how to improve their performance. The humans remained in charge but became vastly more effective and efficient. This approach is usually much more feasible than trying to design machines that can do everything humans can do. It often leads to better, more satisfying work for the people involved and ultimately to a better outcome for customers.

Designing and implementing new combinations of technologies, human skills, and capital assets to meet customers' needs requires large-scale creativity and planning. It is a task that machines are not very good at. That makes being an entrepreneur or a business manager one of society's most rewarding jobs in the age of ML.

## **RISKS AND LIMITS**

The second wave of the second machine age brings with it new risks. In particular, machine learning systems often have low "interpretability," meaning that humans have difficulty figuring out how the systems reached their decisions. Deep neural networks may have hundreds of millions of connections, each of which contributes a small amount to the ultimate decision. As a result, these systems' predictions tend to resist simple, clear explanation. Unlike humans, machines are not (yet!) good storytellers. They can't always give a rationale for why a particular applicant was accepted or rejected for a job, or a particular medicine was recommended. Ironically, even as we have begun to overcome Polanyi's Paradox, we're facing a kind of reverse version: Machines know more than they can tell us.

This creates three risks. First, the machines may have hidden biases, derived not from any intent of the designer but from the data provided to train the system. For instance, if a system learns which job applicants to accept for an interview by using a data set of decisions made by human recruiters in the past, it may inadvertently learn to perpetuate their racial, gender, ethnic, or other biases. Moreover, these biases may not appear as an explicit rule but, rather, be embedded in subtle interactions among the thousands of factors considered.



A second risk is that, unlike traditional systems built on explicit logic rules, neural network systems deal with statistical truths rather than literal truths. That can make it difficult, if not impossible, to prove with complete certainty that the system will work in all cases – especially in situations that weren't represented in the training data. Lack of verifiability can be a concern in mission-critical applications, such as controlling a nuclear power plant, or when life-or-death decisions are involved.

Third, when the ML system does make errors, as it almost inevitably will, diagnosing and correcting exactly what's going wrong can be difficult. The underlying structure that led to the solution can be unimaginably complex, and the solution may be far from optimal if the conditions under which the system was trained change.

While all these risks are very real, the appropriate benchmark is not perfection but the best available alternative. After all, we humans, too, have biases, make mistakes, and have trouble explaining truthfully how we arrived at a particular decision. The advantage of machine-based systems is that they can be improved over time and will give consistent answers when presented with the same data.

Does that mean there is no limit to what artificial intelligence and machine learning can do? Perception and cognition cover a great deal of territory – from driving a car to forecasting sales to deciding whom to hire or promote. We believe the chances are excellent that AI will soon reach superhuman levels of performance in most or all of these areas. So what *won't* AI and ML be able to do?

We sometimes hear “Artificial intelligence will never be good at assessing emotional, crafty, sly, inconsistent human beings – it's too rigid and impersonal for that.” We don't agree. ML systems like those at Affectiva are already at or beyond human-level performance in discerning a person's emotional state on the basis of tone of voice or facial expression. Other systems can infer when even the world's best poker players are bluffing well enough to beat them at the

amazingly complex game Heads-up No-Limit Texas Hold'em. Reading people accurately is subtle work, but it's not magic. It requires perception and cognition – exactly the areas in which ML is currently strong and getting stronger all the time.

A great place to start a discussion of the limits of AI is with Pablo Picasso's observation about computers: "But they are useless. They can only give you answers." They're actually far from useless, as ML's recent triumphs show, but Picasso's observation still provides insight. Computers are devices for answering questions, not for posing them. That means entrepreneurs, innovators, scientists, creators, and other kinds of people who figure out what problem or opportunity to tackle next, or what new territory to explore, will continue to be essential.

WHILE ALL THE RISKS OF AI ARE VERY REAL, THE  
APPROPRIATE BENCHMARK IS NOT PERFECTION  
BUT THE BEST AVAILABLE ALTERNATIVE.

Similarly, there's a huge difference between passively assessing someone's mental state or morale and actively working to change it. ML systems are getting quite good at the former but remain well behind us at the latter. We humans are a deeply social species; other humans, not machines, are best at tapping into social drives such as compassion, pride, solidarity, and shame in order to persuade, motivate, and inspire. In 2014 the TED Conference and the XPrize Foundation announced an award for "the first artificial intelligence to come to this stage and give a TED Talk compelling enough to win a standing ovation from the audience." We doubt the award will be claimed anytime soon.

We think the biggest and most important opportunities for human smarts in this new age of superpowerful ML lie at the intersection of two areas: figuring out what problems to work on next, and persuading a lot of people to tackle them and go along with the solutions. This is a decent definition of leadership, which is becoming much more important in the second machine age.

The status quo of dividing up work between minds and machines is falling apart very quickly. Companies that stick with it are going to find themselves at an ever-greater competitive disadvantage compared with rivals who are willing and able to put ML to use in all the places where it is appropriate and who can figure out how to effectively integrate its capabilities with humanity's.

A time of tectonic change in the business world has begun, brought on by technological progress. As was the case with steam power and electricity, it's not access to the new technologies themselves, or even to the best technologists, that separates winners from losers. Instead, it's innovators who are open-minded enough to see past the status quo and envision very different approaches, and savvy enough to put them into place. One of machine learning's greatest legacies may well be the creation of a new generation of business leaders.

In our view, artificial intelligence, especially machine learning, is the most important general-purpose technology of our era. The impact of these innovations on business and the economy will be reflected not only in their direct contributions but also in their ability to enable and inspire complementary innovations. New products and processes are being made possible by better vision systems, speech recognition, intelligent problem solving, and many other capabilities that machine learning delivers.

Some experts have gone even further. Gil Pratt, who now heads the Toyota Research Institute, has compared the current wave of AI technology to the Cambrian explosion 500 million years ago that birthed a tremendous variety of new life forms. Then as now, one of the key new capabilities was vision. When animals first gained this capability, it allowed them to explore the environment far more effectively; that catalyzed an enormous increase in the number of

species, both predators and prey, and in the range of ecological niches that were filled. Today as well we expect to see a variety of new products, services, processes, and organizational forms and also numerous extinctions. There will certainly be some weird failures along with unexpected successes.

Although it is hard to predict exactly which companies will dominate in the new environment, a general principle is clear: The most nimble and adaptable companies and executives will thrive. Organizations that can rapidly sense and respond to opportunities will seize the advantage in the AI-enabled landscape. So the successful strategy is to be willing to experiment and learn quickly. If managers aren't ramping up experiments in the area of machine learning, they aren't doing their job. Over the next decade, AI won't replace managers, but managers who use AI will replace those who don't. | **THE BIG IDEA**

