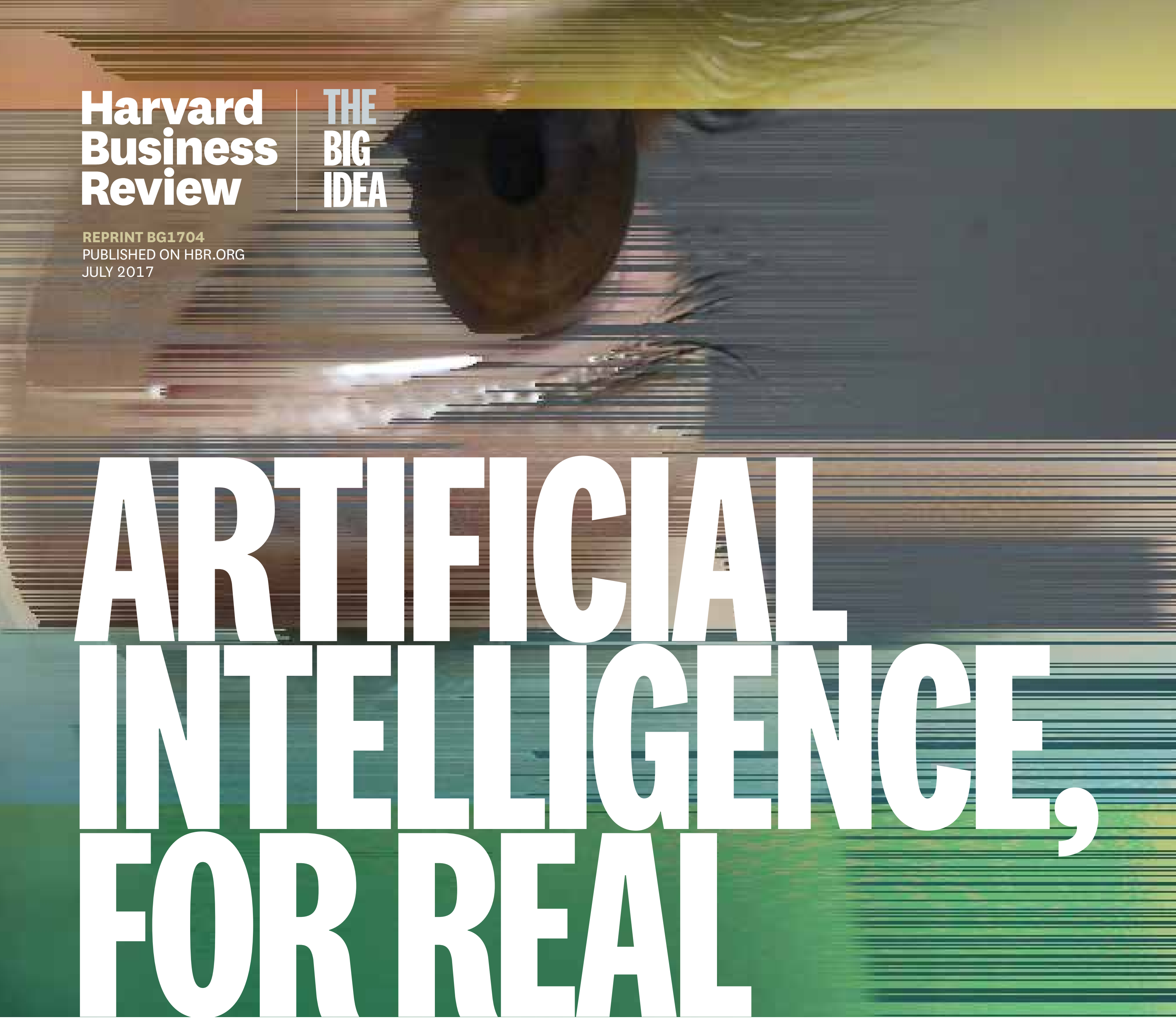


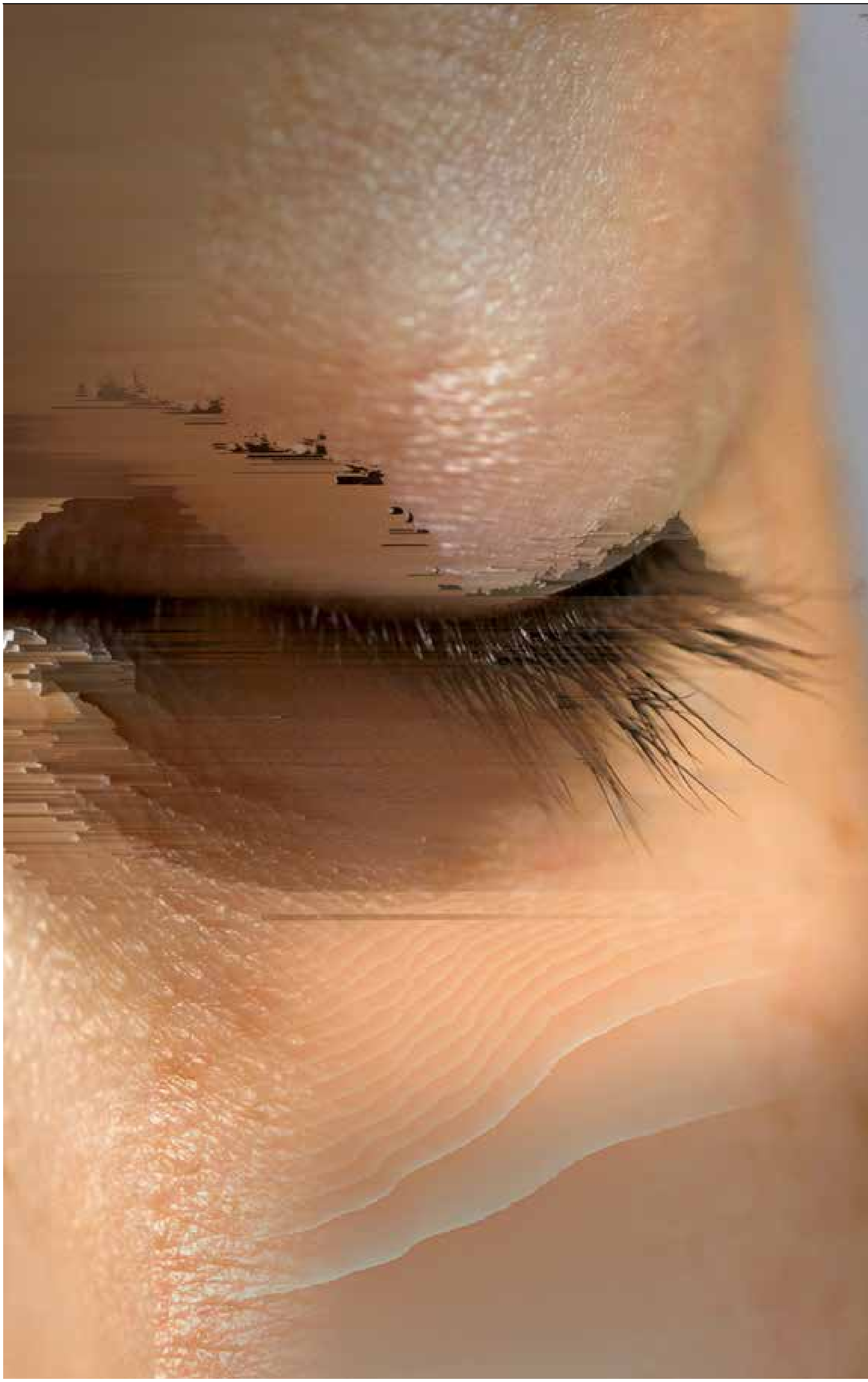
Harvard Business Review
THE BIG IDEA

REPRINT BG1704
PUBLISHED ON HBR.ORG
JULY 2017

# ARTIFICIAL INTELLIGENCE, FOR REAL

## YOU'VE BEEN TOLD IT WILL TRANSFORM EVERYTHING. YOU'VE BEEN TOLD YOU NEED TO INVEST IN IT. BUT YOU HAVEN'T BEEN TOLD HOW. START HERE.

BY ERIK BRYNJOLFSSON AND ANDREW MCAFEE





**COVER:** WHAT DOES IT MEAN TO BE HUMAN? WHAT DO WE RECOGNIZE AS ARTIFICIAL? THE ART FOR THIS SERIES WAS GENERATED FROM A SERIES OF PHOTOGRAPHS OF HUMANS, BUT BECAUSE OF AN APPLICATION OF DISTORTION, YOU MAY NOT RECOGNIZE THE FEATURES THEY PORTRAY. OR YOU MAY. (SOURCE: HBR DESIGN STAFF)

NEXT IN THE BIG IDEA
**SEPTEMBER 2017**
**Dr. Vivek H. Murthy,** the 19th surgeon general of the United States and a tech entrepreneur, makes a case for why addressing social isolation and cultivating emotional well-being at work can make a real difference in fighting loneliness in the U.S. Drawing on his experience as both the nation's doctor and an internist, Murthy shares his insights on how our colleagues and actions at work hold the keys to our health and the impact of our work.

# THE BUSINESS OF ARTIFICIAL INTELLIGENCE

## WHAT IT CAN — AND CANNOT — DO FOR YOUR ORGANIZATION

BY ERIK BRYNJOLFSSON AND ANDREW MCAFEE

For 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.

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.

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.

Brynjolfsson was among the first researchers to measure IT's productivity contributions and the complementary role of organizational capital and other intangibles. His research provided the first quantification of online product variety value, known as the "long tail," and developed pricing and bundling models for information goods. He earned his AB and his SM in applied mathematics and decision sciences at Harvard and his PhD in managerial economics at the Sloan School.

Brynjolfsson is the author of several books, including, with Andrew McAfee, *Machine, Platform, Crowd: Harnessing Our Digital Future* (2017) and the *New York Times* best seller *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies* (2014). You can find his papers here.

**Andrew McAfee** (@amcafee), a principal research scientist at MIT, studies how digital technologies are changing business, the economy, and society. With Erik Brynjolfsson he coauthored *Machine, Platform, Crowd: Harnessing Our Digital Future* (2017) and *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies* (2014), which was a *New York Times* best seller and was shortlisted for the Financial Times/ McKinsey Business Book of the Year Award. McAfee writes academic papers, a blog for the *Financial Times*, and articles for publications including *Harvard Business Review, the Economist, the Wall Street Journal,* and the *New York Times.* He has talked about his work on *Charlie Rose* and *60 Minutes;* at TED, Davos, and the Aspen Ideas Festival; and in front of many other audiences.

McAfee was educated at Harvard and MIT, where he is a cofounder of the institute's Initiative on the Digital Economy.

## 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.

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

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.

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.

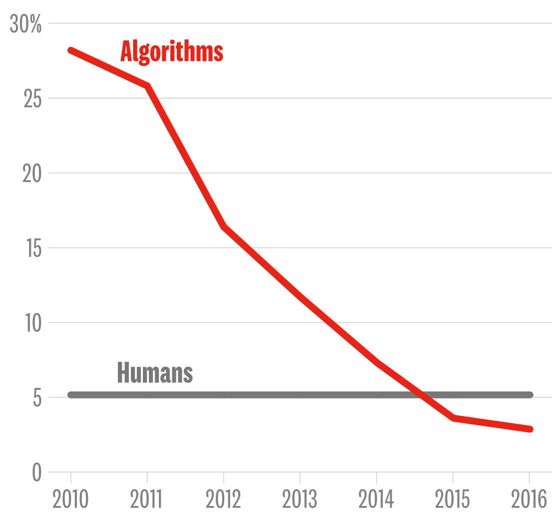
The second type of major improvement has been in cognition and problem solving. Machines have already beaten the finest (human) players of poker and Go — achievements that experts had predicted would take at least another decade. Google's DeepMind team has used ML systems to improve the cooling efficiency 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

## 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 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 start-ups 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.

## 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 |

## 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

ABOVE: THIS IS WHAT IT MEANS TO WORK WITH ARTIFICIAL INTELLIGENCE. THE RESULTS ARE HUMAN AND NOT HUMAN. RECOGNIZABLE BUT ALSO UNEXPECTED. ARE THEY BEAUTIFUL? FRIGHTENING? DELIGHTFUL?

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.

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.

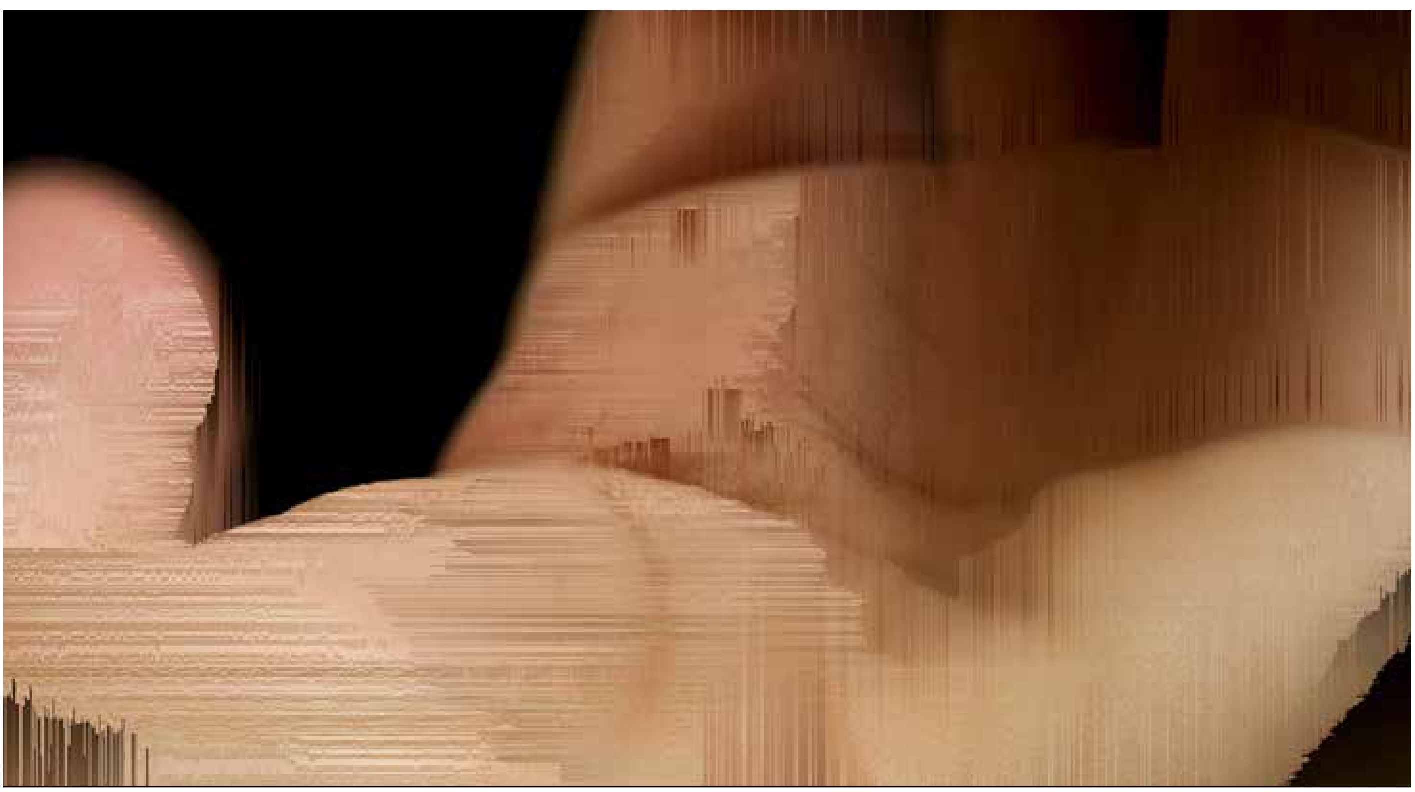
It's comparatively straightforward to label a body of data and use it to train a supervised learner; that's why supervised ML systems are more common than *un*supervised ones, at least for now. Unsupervised learning systems seek to learn on their own. We humans are excellent unsupervised learners: We 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.

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

BELOW: LOOK DEEP, AND YOU'LL SEE THE HUMAN IN THE ALGORITHM. LOOK DEEPER, AND YOU'LL SEE THE ALGORITHM IN THE INTELLIGENCE.

**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.**



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 start-up 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

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

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.

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. ■

ARTICLE

# WHAT'S DRIVING THE MACHINE LEARNING EXPLOSION?

Three factors make this AI's moment.

**by Erik Brynjolfsson and Andrew McAfee**

Machine learning systems have been around since the 1950s, so why are we suddenly seeing breakthroughs in so many diverse areas? Three factors are at play: enormously increased data, significantly improved algorithms, and substantially more-powerful computer hardware. Over the past two decades (depending on the application) data availability has increased as much as 1,000-fold, key algorithms have improved 10-fold to 100-fold, and hardware speed has improved by at least 100-fold. According to MIT's Tomaso Poggio, these can combine to generate improvements of up to a millionfold in applications such as the pedestrian-detection vision systems used in self-driving cars.

**Let's look at each factor in turn.**

**Data.** Music CDs, movie DVDs, and web pages have been adding to the world's stock of digitally encoded information for decades, but over the past few years the rate of creation has exploded. Signals from sensors in smartphones and industrial equipment, digital photos and videos, a nonstop global torrent of social media, and many other sources combine to put us in a totally unprecedented era of data abundance. Ninety percent of the digital data in the world today has been created in the past two years alone. With the burgeoning internet of things (IoT) promising to connect billions of new devices and their data streams, it's a sure bet we'll have far more digital data to work with in the coming decade.

**Algorithms.** The data deluge is important not only because it makes existing algorithms more effective but also because it encourages, supports, and accelerates the development of better algorithms. The algorithms and approaches that now dominate the discipline — such as deep supervised learning and reinforcement learning — share a vital basic property: Their results improve as the amount of training data they're given increases. The performance of an algorithm usually levels off at some point, after which feeding it more data has little or no effect. But that does not yet appear to be the case for many of the algorithms being widely used today. At the same time, new algorithms are transferring the learning from one application to another, making it possible to learn from fewer and fewer examples.

**Computer hardware.** Moore's Law — that integrated circuit capability steadily doubles every 18 to 24 months — celebrated its 50th anniversary in 2015, at which time it was still going strong. Some have commented recently that it's running up against the limits of physics and so will slow down in the years to come; and indeed, clockspeed for standard microprocessors has leveled off. But by a fortuitous coincidence, a related type of computer chip, called a graphic processing unit, or GPU, turns out to be very effective when applied to the types of calculations needed for neural nets. In fact, speedups of 10X are not uncommon when neural nets are moved from traditional central processing units to GPUs. GPUs were initially developed to rapidly display graphics for applications such as computer gaming, which provided scale economies and drove down unit costs, but an increasing number of them are now used for neural nets. As neural net applications become even more common, several companies have developed specialized chips optimized for this application, including Google's tensor processing unit, or TPU. According to Shane Legg, a cofounder of Google DeepMind, a training run that takes one day on a single TPU device would have taken a quarter of a million years on an 80486 from 1990. This can generate about another 10-fold improvement.

These improvements have a synergistic effect on one another. For instance, the better hardware makes it easier for engineers to test and develop better algorithms and, of course, enables machines to crunch much larger data sets in a reasonable amount of time. Some of the applications being solved today — converting sound waves from speech into meaningful text, for example — would take literally centuries to run on 1990s-vintage hardware. Successes motivate more bright researchers to go into the field and more investors and executives to fund further work.

Further amplifying these synergies are two additional technologies: global networking and the cloud. The mobile internet can now deliver digital technologies virtually anywhere on the planet, connecting billions of potential customers to AI breakthroughs. Think about the intelligent assistants you're probably already using on your smartphone, the digital knowledge bases that large companies now share globally, and the crowdsourced systems, like Wikipedia and Kaggle, whose main users and contributors are smart people outside your organization.

Perhaps even more important is the potential of cloud-based AI and robotics to accelerate learning and diffusion. Consider a robot in one location that struggles with a task, such as recognizing an object. Once it has mastered that task, it will be able to upload that knowledge to the cloud and share it with other robots that use a compatible knowledge-representation system (Rethink Robotics is working on such a platform). In this way robots, working independently, can effectively gather data from hundreds, thousands, and eventually millions of eyes and ears. By combining their information in a single system, they can learn vastly more rapidly and share their insights almost instantaneously. ■

ARTICLE

# INSIDE FACEBOOK'S AI WORKSHOP

At the social network behemoth, machine learning has become a platform for the platform.

**by Scott Berinato**

Within Facebook's cavernous Building 20, about halfway between the lobby (panoramic views of the Ravenswood Slough) and the kitchen (hot breakfast, smoothies, gourmet coffee), in a small conference room called Lollapalooza, Joaquin Candela is trying to explain artificial intelligence to a layperson.

Candela — bald, compact, thoughtful — runs Facebook's Applied Machine Learning (AML) group — the engine room of AI at Facebook, which, increasingly, makes it the engine room of Facebook in general. After some verbal searching, he finally says:

> *"Look, a machine learning algorithm really is a lookup table, right? Where the key is the input, like an image, and the value is the label for the input, like 'a horse.' I have a bunch of examples of something. Pictures of horses. I give the algorithm as many as I can. 'This is a horse. This is a horse. This isn't a horse. This is a horse.' And the algorithm keeps those in a table. Then, if a new example comes along — or if I tell it to watch for new examples — well, the algorithm just goes and looks at all those examples we fed it. Which rows in the table look similar? And how similar? It's trying to decide, 'Is this new thing a horse? I think so.' If it's right, the image gets put in the 'This is a horse' group, and if it's wrong, it gets put in the 'This isn't a horse' group. Next time, it has more data to look up.*
>
> *"One challenge is how do we decide how similar a new picture is to the ones stored in the table. One aspect of machine learning is to learn similarity functions. Another challenge is, What happens when your table grows really large? For every new image, you would need to make a zillion comparisons.... So another aspect of machine learning is to approximate a large stored table with a function instead of going through every image. The function knows how to roughly estimate what the corresponding value should be. That's the essence of ML — to approximate a gigantic table with a function. This is what learning is about."*

There's more to it than that, obviously, but it's a good starting point when talking about AI, because it makes it sound real, almost boring. Mechanical. So much of the conversation around AI is awash in mystical descriptions of its power and in reverence for its near-magic capabilities. Candela doesn't like that and tries to use more-prosaic terms. It's powerful, yes, but not magical. It has limitations. During presentations, he's fond of showing a slide with a wizard and a factory, telling audiences that Facebook thinks of AI like the latter, because "wizards don't scale."

And that's what Facebook has done with AI and machine learning: scaled it at a breakneck pace. A few years ago the company's machine learning group numbered just a few and needed days to run an experiment. Now, Candela says, several hundred employees run thousands of experiments a day. AI is woven so intricately into the platform that it would be impossible to separate the products — your feed, your chat, your kid's finsta — from the algorithms. Nearly everything users see and do is informed by AI and machine learning.

Understanding how and why Facebook has so fully embraced AI can help any organization that's ready to invest in an algorithmic future. It would be easy to assume that Facebook, with all its resources, would simply get the best talent and write the best algorithms — game over. But Candela took a different approach. Certainly the talent is strong, and the algorithms are good. Some of them are designed to "see" images or automatically filter them. Some understand conversations and can respond to them. Some translate between languages. Some try to predict what you'll like and buy.

But in some ways the algorithms are not his main focus. Instead, he's been busy creating an AI workshop in which anyone in the company can use AI to achieve a goal. Basically, Candela built an AI platform for the platform. Whether you're a deeply knowledgeable programmer or a complete newbie, you can take advantage of his wares.

Here's how he did it and what you can learn from it.

## SOYUZ

Candela, a veteran of Microsoft Research, arrived at Facebook in 2012 to work in the company's ads business. He and a handful of staffers inherited a ranking algorithm for better targeting users with ads.

Candela describes the machine learning code he inherited as "robust but not the latest." More than once he compares it to Soyuz, the 1960s Soviet spacecraft. Basic but reliable. Gets the job done even if it's not the newest, best thing. "It'll get you up there and down. But it's not the latest covnet [convolutional neural net] of the month."

You might assume, then, that the first thing Candela set out to do was to upgrade the algorithm. Get rid of Soyuz in favor of a space plane. It wasn't. "To get more value, I can do three things," he says. "I can improve the algorithm itself, make it more sophisticated. I can throw more and better data at the algorithm so that the existing code produces better results. And I can change the speed of experimentation to get more results faster.

"We focused on data and speed, not on a better algorithm."

Candela describes this decision as "dramatic" and "hard." Computer scientists, especially academic-minded ones, are rewarded for inventing new algorithms or improving existing ones. A better statistical model is the goal. Getting cited in a journal is validation. Wowing your peers gives you cred.

It requires a shift in thinking to get those engineers to focus on business impact before optimal statistical model. He thinks many companies are making the mistake of structuring their efforts around building the best algorithms, or hiring developers who claim to have the best algorithms, because that's how many AI developers think.

But for a company, a good algorithm that improves the business is more valuable than vanguard statistical models. In truth, Candela says, real algorithmic breakthroughs are few and far between — two or three a year at best. If his team focused its energies there, it would take lots of effort to make marginal gains.

He hammers these points home constantly: Figure out the impact on the business first. Know what you're solving for. Know what business challenge you need to address. "You might look for the shiniest algorithm or the people who are telling you they have the most advanced algorithm. And you really should be looking for people who are most obsessed with getting any algorithm to do a job. That's kind of a profound thing that I think is lost in a lot of the conversation. I had a conversation with our resident machine learning geek at our office, and we were just talking about different people doing AI. He said, 'Nobody really thinks their algorithms are very good or whatever.' It makes me think, maybe that's fine.

"I'm not saying don't work on the algorithm at all. I'm saying that focusing on giving it more data and better data, and then experimenting faster, makes a lot more sense."

So rather than defining success as building the best natural language processing algorithm, he defines it as deploying one that will help users find a restaurant when they ask their friends, "Where can I get a good bite around here?" Instead of being thrilled that some computer vision algorithm is nearing pixel-perfect object recognition, he gets excited if that AI is good enough to notice that you post a lot of pictures of the beach and can help you buy a swimsuit.

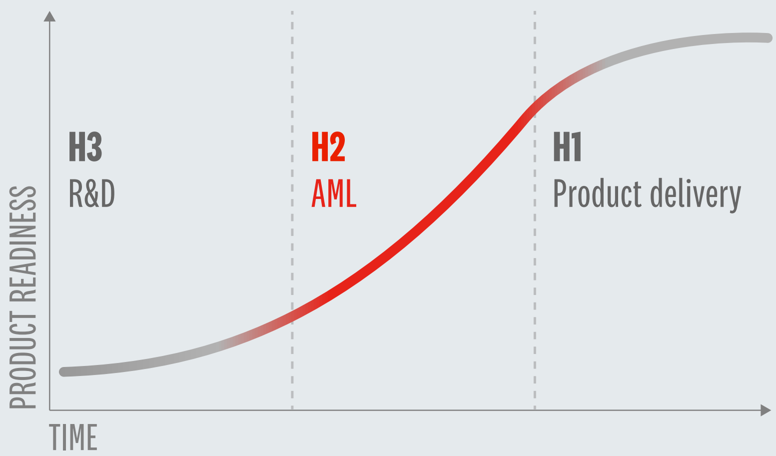
The strategy worked when he started at Facebook. Ad revenues rose. Candela's profile rose. It was suggested that AML become a centralized function for all of Facebook. Candela said no. Twice. "I was concerned about the 'If you build it, they will come' phenomenon." Just creating bits of artificial intelligence in the hope that people would see the value and adopt it wouldn't work.

But he did pick his spots. He collaborated with the feeds team while saying no to many other groups. Then he worked with the Messenger team. His team grew and took on more projects with other teams.

By 2015 Candela could see that his group would need to centralize, so he turned his attention to how he'd build such an operation. He was still worried about the "build it and they will come" phenomenon, so he focused less on how his team would be structured and more on how the group would connect to the rest of Facebook. "You build a factory that makes amazing widgets, and you forget to design the loading docks into your factory?" He laughs. "Well, enjoy your widgets."

Only then, about three years in, did Candela think about upgrading some of his algorithms. (Incidentally, even today, the emergency escape spacecraft attached to the International Space Station is a Soyuz.)

## Where AI Fits In at Facebook




SOURCE FACEBOOK © HBR.ORG

### H2

Candela goes to a whiteboard to describe how he built his AI factory inside Facebook. The key, he says, was figuring out where on the product development path AI fits. He draws something like the graph on this page (see the exhibit "Where AI Fits In at Facebook").

H3 — Horizon 3 or three years out from product — is the realm of R&D and science. Often, data scientists who work on AI think of themselves as here, improving algorithms and looking for new ways to get machines to learn. Candela didn't put his team here for the reasons already mentioned. It's too far from impact on the business. H1, approaching product delivery, is where the product teams live — the feeds team, the Instagram team, the ads team. AI doesn't go here either, because it would be difficult to retrofit products this deeply developed. It would be like building a car and then deciding that it should be self-driving after you started to put it together.

That leaves H2, between the science and the product, as the place AML lives at Facebook. AML is a conduit for transferring the science into the product. It does not do research for research's sake, and it does not build and ship products. As the upward slope in the product's readiness shows, it's a dynamic space. Pointing to H2, Candela says, "This needs to feel uncomfortable all the time. The people you need to hire need to be okay with that, and they need to be incredibly selfless. Because if your work is successful, you spin it out. And you need to fail quite a bit. I'm comfortable with a 50% failure rate."

If the team is failing less, Candela suspects its members are too risk averse, or they're taking on challenges that are sliding them closer to H1's product focus. "Maybe we do something like that and it works, but it's still a failure, because the product teams should be taking that on, not us. If you own a piece of technology that the ads team should operate themselves to generate value, give it to them, and then increase your level of ambition in the machine learning space before something becomes product."

So Candela's team is neither earning the glory of inventing new statistical models nor putting products out into the world. It's a factory of specialists who translate others' science for others' products and fail half the time.

### PUSH/PULL

All that being said, the lines between the three realms — H3, H2, and H1 — still aren't crisp. In some cases Candela's team does look at the science of machine learning, to solve specific problems. And sometimes it does help build the products.

That was especially true as AML got off the ground, because many people in the business hadn't yet been exposed to AI and what it could do for them. In one case AML built a translation algorithm. The team dipped into the research space to look at how existing translation algorithms worked and could be improved, because bad translations, which either don't make sense or create a misleading interpretation, are in some ways worse than no translation.

"Early on it was more push, more tenacity on our part," Candela says. "But it was gentle tenacity. We weren't going to throw something over the fence and tell the product team, 'This is great, use it.'" That meant that his team helped write some product code. Doing a little bit of the science and a little bit of the product in addition to its core function was meant to inspire the product team members to see what AML could do for them.

What the two teams built — a product that allowed community pages to instantly translate into several languages — worked. Other projects were similarly pushed out, and now the international team and other product groups at Facebook are pulling from AML, asking to use code in their products themselves.

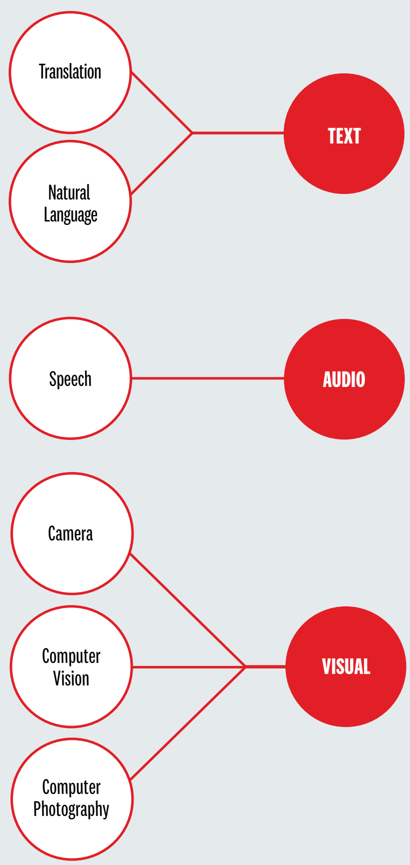
"Look, it's nowhere near where I want it to be," Candela says. "I'd like to have all the product leaders in the company get together quarterly for AI reviews. That will certainly happen. But the conversation in the past two years has completely changed. Now if I walk from one end of this building to the other and I bump into, I don't know, the video team or the Messenger team, they'll stop me and say, 'Hey, we're excited to try this. We think we can build a product on this.' That didn't happen before."

AML's success, though, has created a new challenge for Candela. Now that everyone wants a piece of AML, the factory has to scale.

## LAYER CAKE

Candela couldn't scale just by saying yes to every project and adding bodies to get the work done. So he organized in other ways. First he subdivided his team according to the type of AI its members would focus on:

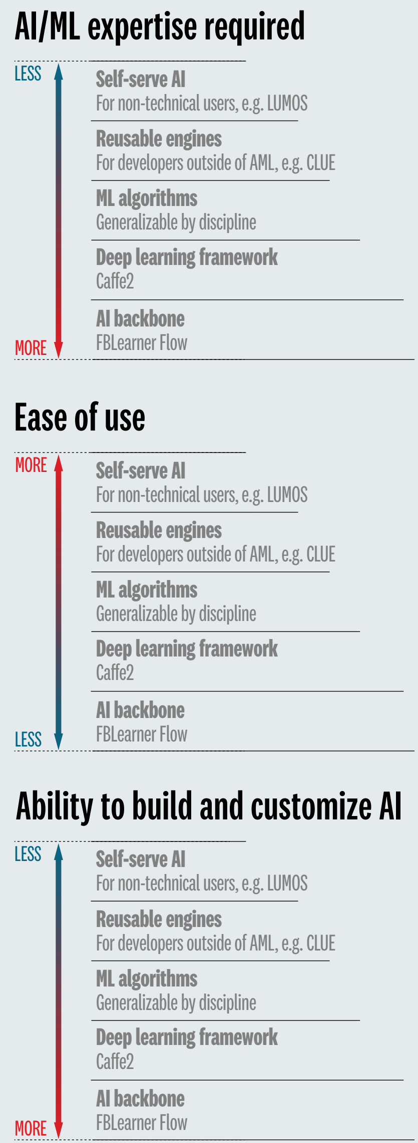
## Applied Machine Learning




SOURCE FACEBOOK © HBR.ORG

This created common denominators so that one team — say, computer vision — could work on any machine learning application involving parsing images and reuse its work whenever possible.

Next came a large-scale engineering effort to build Facebook's own AI backbone, called FBLearner Flow. Here algorithms are deployed once and made reusable for anyone who may need them. The time-consuming parts of setting up and running experiments are automated, and past results are stored and made available and easily searchable. And the system runs through a serious hardware array, so many experiments can be run simultaneously. (The system allows for more than 6 million predictions a second). All of this is to increase the velocity of running experiments on the data and scale.




SOURCE FACEBOOK © HBR.OR

The system was also designed to accommodate many kinds of possible users. Candela believes that for AI to work, and to scale even further, he must help people outside AML do the work themselves. He created what he calls a layer cake of artificial intelligence.

The bottom layers focus on AML's work: refining the core engine (with a strong focus on optimizing performance, especially for mobile) and working with machine learning algorithms. The upper layers focus on tools that make it possible for those outside AML to exploit the algorithms with less AML involvement. "It's all about what you expose to the user," Candela says. In some cases he's built systems that developers outside AML can take advantage of to build and run their own models.

## REX

A good example of Candela's team structure and the push/pull dynamic comes from some AI built to surface content on the basis of what you type. The natural-language machine learning team created an engine to understand conversational typing.

This bit of intelligence first found its way into the Messenger chat client. AML developed the models while the product team developed use cases and "intents" — lingo for the types of tasks you want the engine to learn. For example, training natural language AI to recognize and reliably respond to a phrase like "I'm looking for the best…" is an intent.

The first few such intents were deployed to Messenger through a product called M Suggestions.

If you sent a chat to a friend that said "I'll meet you there in 30 minutes," M Suggestions might prompt you with an offer to hire a car.

As the tools for building intent models developed and the product team became more conversant with them, AML's role diminished. Now the Messenger team has improved M Suggestions by building dozens more intents on its own.

Still, this bit of natural language AI wasn't built just for chat. It's reusable. It was codified as CLUE, for "conversational learning understanding engine." It found its way into more Facebook applications. It's being adapted for status updates and feeds. Social recommendations — or social rex, as everyone calls them — are increasingly driven by AI. If you typed "I'm traveling to Omaha and I really want to find a good steak downtown," AI might respond as if it were one of your friends, with a comment on your post, rex such as a list of steakhouses, and a tagged map of where they are relative to downtown. If your friend replied to you and said, "It also has some great vegetarian restaurants," the algorithm might again reply with pertinent data.

Social rex intents are not yet being developed without AML, but the goal is to have them move out of Candela's group, just as M Suggestions did.

In general, the idea is to make product teams AI-capable themselves. "We'll teach you to fish," Candela says, "and you go fish, and we'll drag up the next thing. We'll build a fishing boat. And once you're using the fishing boat, I'm going to build a cannery, right?"

At the moment, about 70% of the AI work on the backbone is done by people outside Candela's team. That's possible in part because of the interface with AI. In some cases, as with a tool called Lumos, machine learning can be used by nondevelopers.

## HORSEBACK RIDING AND CEREAL BOXES

Lumos is computer vision AI, a tool that can comb through photos on Facebook or Instagram or other platforms and learn what they contain. You can train it to see anything. It has helped automate the discovery and banning of pornographic or violent content, IP appropriation (improper use of brands and logos), and other unwelcome content. It can also help identify things you like and do (to drive personalized advertising and recommendations), on the basis of photos in your feeds.

I watch a demo in which engineers select "horseback riding" as our intent, the thing we'll be looking for. The interface is simple: A few clicks, a couple of forms to fill out — What are you looking for? How much data do you want to look at? — and the algorithm gets to work finding pictures of horseback riding. Thumbnails start to fill the page.

The algorithm has searched for horseback riding before, so it's already quite good at finding it. My guess is that north of 80% of the images that pop up are indeed of horseback riding, and they show remarkable variety. Here's one with someone posing at a standstill. Here's one with the horse rearing. Here's an equestrian jumping. The algorithm finds shapes and boundaries between shapes and builds on previous knowledge of what those interactions mean. It knows things about what combination of pixels is most likely a person, for example, and what's a horse. It knows when it "sees" a person and a horse together with the person situated close above the horse. And it decides that this looks like horseback riding.

We also find pictures that aren't horseback riding — one is a person standing next to a horse; another is a person on a mule — and check those off as not matches. They're framed in red, in case there's any doubt. The algorithm internalizes that information — adds it to the lookup table — for use next time. A simple chart at the top of the page shows the algorithm's accuracy and confidence over time. It's always an S curve, slow to learn at first, then rapidly improving, then tapering off on how much more accurate it can get. It's very good at seeing horseback riding.

Other potentially valuable pictures are harder for AI to parse. "Receipts" is tricky to suss out because it can look to a computer just like type on a page; but there would be some interesting apps for AI that could recognize and "read" receipts. The engineers show how bowling alleys and escalators often confuse the algorithm because they have similar shapes and visual properties.

I ask, "What about something like 'food'?" This brings us to an important point about machine learning: It's only as good as its training.

We call up food as a topic to train. Indeed, we see lots of pictures of fruits and vegetables, a few of plates at restaurants. All food. We also see a cereal box. Is that food?

Well, yes. Or no. It's a box. But there's food in it. When we buy it, we're buying food, not the box. If I asked if there was any food in the cupboard, you wouldn't say, "No, just a cereal box." (Or, more pertinent to Facebook, if I posted a picture of a cereal box, should it think I'm posting about food or about a box?) As a picture, as a piece of data, it's a box.

Should we mark this as a match or a miss? Here's part of the art of machine learning. When training algorithms, one needs to use clearly definable categories. Food is probably too general in some ways, and the algorithm will either improperly hit or miss on images because it's hard to know what we mean when we say, "Show me pictures of food." "Vegetable" is a better idea to train on. And when training, everyone must define terms in the same way. Imagine two people training the algorithm when one always marks cereal boxes as food, and the other marks them as not food. Now imagine that happening at scale, on terabytes of visual data.

The same applies to natural language processing. Humans are very good at interpreting text in context to find sophisticated meaning. For example, I may type, "Gee, I love that movie about the superheroes. It's so, so original! I hope they make a hundred more of them." My friends, who know me and know some of the mechanics of sarcasm, may readily understand that my meaning is the opposite of what I'm typing. Artificial intelligence is still learning how to decide the meaning of something like that. To figure out if I'm being sarcastic, it has to go much further than just learning how to parse grammar and vocabulary. It has to see what else I've said and posted, and try to find other clues that will tell it whether I really loved the movie and I want 100 more or I actually detested it — because getting that wrong is not good for a platform that wants to create affinities with me. If I was being sarcastic and my feed starts filling up with superhero movie ads, I'm probably not enjoying the experience.

## NOT MAGIC

It's details like these — showing where AI is still limited, and how humans have such a core role in training it, and how solving problems and creating value are more important than finding great models — that Candela is thinking about near the end of the day, when he's talking about the mythic status AI has gained. He's railing against what he perceives as laziness in those who find the idea of AI-as-magic-bullet appealing and don't apply critical thinking to it.

"What frustrates me," he says, "is that everybody knows what a statistician is and what a data analyst can do. If I want to know 'Hey, what age segment behaves in what way?' I get the data analyst.

"So when people skip that, and they come to us and say, 'Hey, give me a machine learning algorithm that will do what we do,' I'm like, 'What is it that I look like? What problem are you trying to solve? What's your goal? What are the trade-offs?'" Sometimes they're surprised that there are trade-offs. "If that person doesn't have answers to those questions, I'm thinking, 'What the hell are you thinking AI is?'"

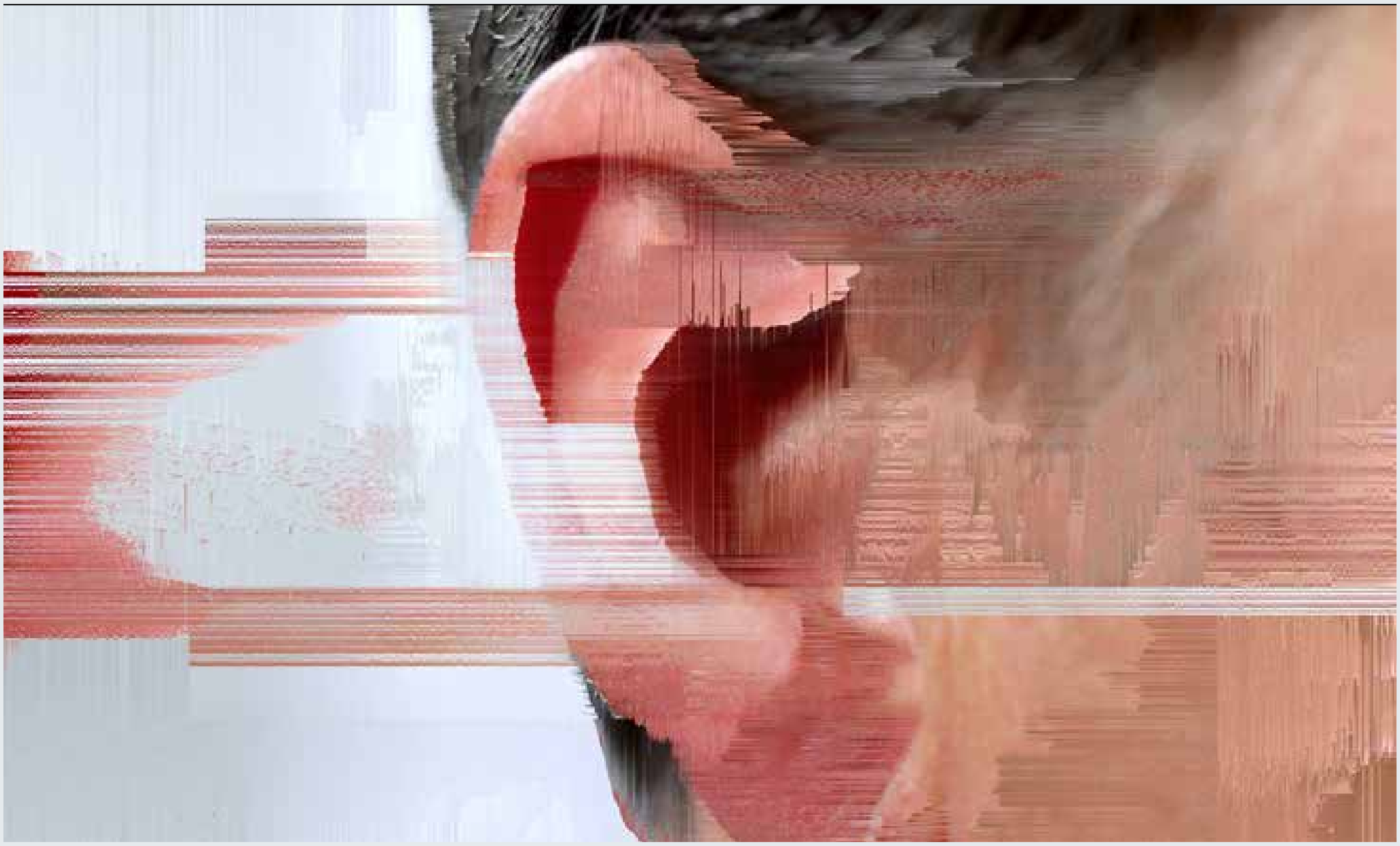
They are thinking it's magic.

"But it's not. That's the part where I tell people, 'You don't need machine learning. You need to build a data science team that helps you think through a problem and apply the human litmus test. Sit with them. Look at your data. If you can't tell what's going on, if you don't have any intuition, if you can't build a very simple, rule-based system — like, *Hey, if a person is younger than 20 and living in this geography, then do this thing* — if you can't do that, then I'm extremely nervous even talking about throwing AI at your problem.'

"I'm delighted when other executives come to me and start not from wanting to understand the technology but from a problem they have that they've thought very, very deeply about. And sometimes — often, in fact — a simple, good old rule-based system, if you have the right data, will get you 80% of the way to solving the problem."

"And guess what? It's going to have the benefit that everybody understands it. Exhaust the human brain first." ■

**About the author: Scott Berinato** is a senior editor at *Harvard Business Review* and the author of *Good Charts: The HBR Guide to Making Smarter, More Persuasive Data Visualizations* (2016).



## ARTICLE
# AI CAN BE A TROUBLESOME TEAMMATE

AI is a focused intelligence, groomed for maximum perfection. That's why, research shows, most people don't trust it.

**by Kurt Gray**

Artificial intelligence promises to make decisions better and faster than humans can — even smart humans. AI's superiority is clear when the choice is "Which road should I take home?" or "How should I organize distribution chains?" But in life-or-death situations, can AI deliver?

I'm a social psychologist who studies technology, but when I was in college, I worked for a geophysical surveying company. We looked for natural gas in the frozen forests of northern Canada. Most sites were remote and very cold. Many could be reached only by helicopter.

One winter afternoon a pilot at one of those sites radioed with bad news: A storm had moved in, making visibility poor and flying dangerous. My crew chief, Ian, had to make a difficult decision: Should he risk our lives by flying in the storm or by staying overnight in the frigid wilderness with no food or shelter? He chose to stay overnight. Although we faced freezing temperatures, I had full faith in Ian's decision. He had worked for years as a wilderness firefighter, and he knew about survival. I literally trusted him with my life.

If my company had been using AI, Ian might not have been making decisions that night. A computer program could have weighed the weather against the costs of losing the crew against the costs of losing the helicopter against many other factors. That intelligent machine might have come to the same conclusion Ian did — that stranding us overnight was the best possible choice — but would I have trusted that decision? Would I have felt safe?

My work since suggests that I would not have trusted AI with my life. And that lack of trust raises serious roadblocks for the full implementation of AI in the workforce, even when no lives are at stake.

My research examines how people understand other minds — human minds, animal minds, and computer minds — and reveals that their contents are more ambiguous than we often think. We can never directly experience the thoughts and feelings of others, and so we're left to make our best guesses about questions such as: Does your baby love you as much as you love him? When your boss smiles, is she actually happy? Does your dog feel embarrassed when you catch it doing something naughty?

Although biological minds can be hard to understand, the nature of computer minds is even more opaque. When Deep Blue beats Garry Kasparov at chess, does it *want* to win or is it just programmed to do so? When Google alerts us to the best route home after work, does it really understand what it means to commute? When Netflix recommends a movie we might like, does it care about our enjoyment?

People perceiving the minds of AI see them as very one-sided — capable of powerful thought but totally incapable of feeling. It's a pretty accurate perception of current technology, because neither Google nor Netflix can fall in love or enjoy the taste of chocolate. But what truly limits AI — or at least its role in the workforce — is that people believe that robots *will never* feel.

In part, it's that inability to feel that makes people regard AI as untrustworthy. This is incredibly important for the deployment of AI. Will employees trust something that views them in purely functional terms — as workers with certain skill sets — rather than as individuals with hopes and concerns?

Trusting team members requires at least three things: Mutual concern, a shared sense of vulnerability, and faith in competence. Mutual concern — knowing that your teammates care about your well-being — is perhaps the most basic element of trust. When a platoon leader risks being shot by going behind enemy lines to rescue one of his soldiers, he is not making the optimal decision from a functional perspective. However, the very fact that — unlike an AI system — he will choose this "irrational" course of action makes everyone in the platoon trust him more, which leads to better overall team performance.

In everyday situations, where careers and promotions are at stake, we still want to know that supervisors and coworkers see us as people rather than as variables in a giant optimization problem. We want to be something more than a row in an inventory spreadsheet. But that's all AI understands us to be.

We mistrust AI not only because it seems to lack emotional intelligence but also because it lacks vulnerability. If humans mess up in a job, they can be fired, lose a bonus, or even die. But in an AI workplace, if an expert decision-making system wrongly recommends one course of action over another, the computer suffers no consequences. AI systems are gambling only with the fates of others, never with their own.

The third impediment to trust is actually AI's strength: its superhuman ability to calculate and predict. We are quick to trust AI's competence after seeing firsthand how it can arrive at huge sums in seconds or forecast the movement of stocks. Unfortunately, this can work against AI, because it performs well only under narrow conditions. When it is pushed to operate outside its limits — when a whole family uses the same Netflix account, or when Google is asked to predict the outcome of a relationship — disappointment is inevitable.

I recently spoke with someone in the Office of Naval Research, part of the U.S. Department of Defense, who outlined how technologically inexperienced sailors operate AI systems. First they approach AI with a sense of awe, expecting it to complete every job perfectly. But if a system makes mistakes that seem — from the point of view of humans — obviously stupid, the sailors stop using it altogether, even in the structured situations in which AI would actually excel. To build trust, AI needs to communicate its confidence or, even better, express its fear of failure.

No one can dispute that AI is leaping ahead in sophistication, but our ability to trust it is lagging behind. This is important because in many industries success requires deep and implicit trust within teams. On oil rigs and in army platoons, trusting your teammates can be a matter of life or death. In less dangerous businesses, trust can make the difference between succeeding and failing to close a deal or finish a project. We trust other people not because they are incredibly smart — like AI — but because they have emotional connections, specifically with *us*.

That doesn't mean AI isn't useful. Quite the contrary. It represents a deconstructed mind, a focused intelligence groomed for maximum performance. In so many ways, it is unlike the well-rounded human mind, which can comprehend language, solve problems, and understand others' feelings all at the same time.

If I were working at that surveying job in northern Canada today, I still might not trust a computer to save my life in the forest, but I would trust AI to screen the weather and decide against our even venturing forth that morning. I'm glad I had a human crew chief, but I wish a computer had prevented our being stranded in the first place.

**About the author: Kurt Gray** is an associate professor of psychology and neuroscience at the University of North Carolina, Chapel Hill. He received his PhD from Harvard University. Gray studies mind perception, moral judgment, social dynamics, and creativity and is an award-winning researcher and teacher. He is a coauthor (with Daniel Wegner) of *The Mind Club: Who Thinks, What Feels, and Why It Matters*.

## Q&A: HILARY MASON
# HOW AI FITS INTO YOUR DATA SCIENCE TEAM

It helps to know the three things data scientists do.

In their HBR Big Idea feature, Erik Brynjolfsson and Andrew McAfee argue that AI and machine learning will soon become "general-purpose technologies" as significant as electricity or the internal combustion engine. They represent a landmark change in our technical capabilities and will power the next wave of economic growth.

But how will we put them into practice? Where in the organization will these new capabilities sit, and how will companies take advantage of them?

To get a practical, on-the-ground view, HBR senior editor Walter Frick spoke with Hilary Mason, the founder of Fast Forward Labs, a machine intelligence research firm. Here are excerpts from their conversation.

**HBR: AI is a hot topic right now. As a data scientist and a researcher, how do you think about the recent progress in your field?**

**MASON:** If we were having this conversation eight or 10 years ago, it would have been about big data — about whether we could even build the infrastructure to get all the data in one place and query it. Once you can do that, you can do analytics — which is essentially counting things to answer questions that have business value or product value. People could always count things in data, but the change we saw about eight years ago was that new software made doing it affordable and accessible for a wide variety of people who never could do it before.

And that led to the rise of data science, which is about counting things cleverly, predicting things, and building models on data. Because that modeling was now so much cheaper, it was applied not just to very high value problems, like actuarial science, but to things that may seem fairly trivial, like recommendations, search results, and that kind of stuff.

Then we had machine learning, which is a set of tools inside data science that let you count things cleverly and incorporate feedback loops. We began using the models to get more data from the world and fed it back into those models so that they improved over time.

Now today we talk about AI. The term itself is a little bit loose and has both a technical meaning and a marketing meaning, but it's essentially about using machine learning — and specifically deep learning — to enable applications that are built on top of this stack. That means that you can't do AI without machine learning. You also can't do machine learning without analytics, and you can't do analytics without data infrastructure. And so that's how I see them all being related.

**How do machine learning and AI fit into companies' existing data capabilities?**

Data science is used in multiple ways inside an organization, and a really common mistake I see people make in managing it is assuming that because it runs on one tech stack, it's just one thing. But I'd break it down into three capabilities, all of which rely on the same technology. The first capability is understanding the business. That's analytics, or business intelligence — being able to ask questions and analyze information to make better decisions. It's usually run out of the CFO or COO's office. It's not necessarily a technical domain.

The second capability is product data science: building algorithms and systems — which may use machine learning and AI — that actually improve the product. This is where things like spam filters, recommendation systems, search algorithms, and data visualization come in. This capability usually sits under a line of business and is run out of product development or engineering.

The last data capability is one that tends to get neglected or lumped in with product data science. It's an R&D capability — using data to open up new product, new business, and new revenue opportunities.

**And are all three capabilities changed by machine learning and AI?**

Let's take a moment and look more closely at what deep learning offers, since it's central to a lot of what people now call AI and a big part of the progress in machine learning in recent years. First, deep learning makes data that was previously inaccessible to any kind of analysis accessible — you can actually find value in video and audio data, for example. The number of companies that have a large amount of that kind of data is still fairly small, but I do think it's likely to increase over time. Even analytics is impacted by the ability to use image data rather than just text or structured data. Second, deep learning enables new approaches to solving very difficult data science problems — text summarization, for example. Deep learning allows you to create predictive models at a level of quality and sophistication that was previously out of reach. And so deep learning also enhances the product function of data science because it can generate new product opportunities. For example, several companies are using deep learning very successfully in e-commerce recommendation systems. Then of course deep learning affects the R&D function by pushing the frontier of what is technically possible.

**So data science is about analytics, product development, and R&D. Is this a walk-before-you-run situation? Or should companies attempt all three at once?**

It's a little bit of both. You'll leave opportunities on the table if you pursue only one of these use cases. However, it really helps to get your infrastructure and analytics piece to be fairly solid before jumping into R&D. And in practice we see that people are much more comfortable investing in cost-saving initiatives before they invest in new revenue opportunities. It's just more culturally acceptable.

**What other mistakes do you see companies making in their data science efforts?**

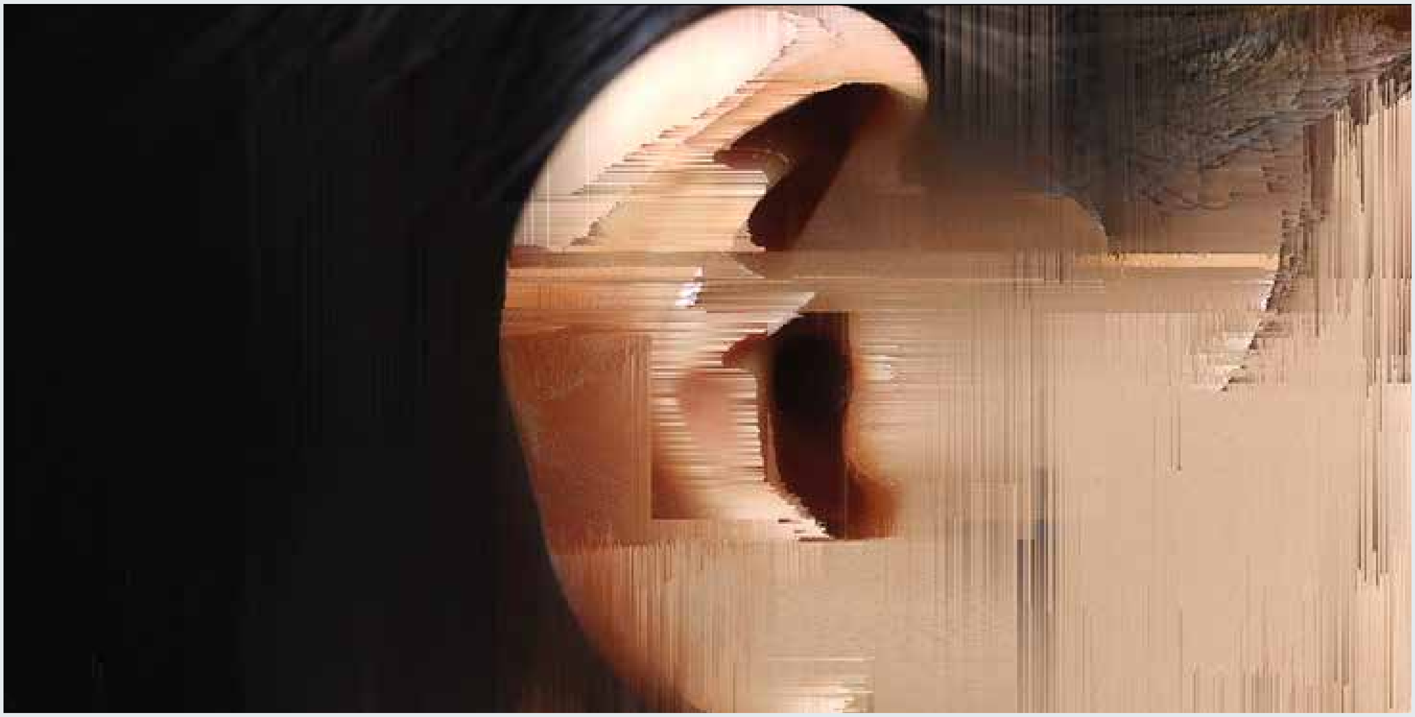
A big one involves process. We've noticed that people shoehorn this kind of stuff into the software-engineering process, and that doesn't work. Developing data science systems is fundamentally different in several ways. At the outset of a data science project, you don't know if it's going to work. At the outset of a software-engineering project, you know it's going to work.

This means that software-engineering processes fail when they encounter uncertainty. By contrast, data science requires an experimental process that allows for uncertainty.

Also, every company has its own cultural hurdle to get over. A lot of companies aren't places where you can work on something that doesn't succeed, so the poor data scientists who do the risky research projects end up getting penalized in their annual reviews because they worked on something for two months that didn't pay off, even though they did great work. Data science requires having that cultural space to experiment and work on things that might fail. Companies need to understand that they're investing in a portfolio of initiatives, some of which will eventually pay off, generating dramatically more value than incremental product improvements do.

**How do you navigate all the buzz around this topic, and how do you recommend executives do so?**

I remain a relentless optimist about the potential of what we're now calling AI, but I'm also a pragmatist in the sense that I need to deliver systems that work to our clients, and that is quite a constraint. There are some folks running around making claims that are clearly exaggerated and ridiculous. In other cases things that a few years ago we would have called a regression analysis are now being called AI, just to enhance their value from a marketing perspective. So my advice is to keep in mind that there is no magic. At a conceptual level nothing here is out of reach of any executive's understanding. And if someone is pitching you on an idea and says, "I don't want to explain how it works, but it's AI," it's really important to keep asking: How does it work? What data goes in? What patterns might be in the data that the system could be learning? And what comes out? Because what comes out of a deep learning system is generally just a previously unlabeled data point that now has a label, along with some confidence in that label, and that's it. It's not intelligent in the sense that you and I are — and we're still a long, long way away from anything that looks like the kind of intelligence a human has. ■



ARTICLE

# WHY AI CAN'T WRITE THIS ARTICLE (YET)

Machines are learning faster and better than ever, but they still have limitations. **by Walter Frick**

Roger Schank, a researcher and former professor, once proposed a novel goal for artificial intelligence: A computer should be able to watch *West Side Story* and recognize the plot of *Romeo and Juliet*. Schank and his students believed that stories are central to intelligence, reasoning, and meaning. By Schank's measure, today's AI isn't intelligent at all.

The feature article on AI that HBR.org published earlier this week is, ironically, a good example of the kind of work that computers can't yet do. It was written by two experts who drew on decades of experience to formulate a thesis, assemble evidence, and construct a narrative. And three editors helped to shape the nearly 5,000 words that made it into the final piece.

The fact that software can't yet write an article like that isn't a knock on AI, or evidence that it won't be transformative. But that fact offers a window into how, exactly, machine learning technologies work, what they are and aren't good at right now, and how they'll develop as writing tools — or even writers — in the future.

## NOT READY FOR LONG-FORM

Today's AI works by formulating tasks as prediction problems and then using statistical techniques and lots of data to make predictions. One simple example of a text-based prediction problem is auto-complete. When I type "How'd" into a text message, my phone uses data and statistical modeling to predict what's coming next. It offers "it," "you," or "the." "It" is what I had in mind, and once I select that, my phone moves on to predicting the next word. This time it's so confident that I'm going to select "go" (which is right) that it doesn't even offer other options but instead moves on to the next word, suggesting "go with" or "go today." In machine learning, prediction problems like this are called *supervised learning*. Given a data set containing the right answer — in this case lots of completed text messages — an algorithm learns to recognize patterns, such as that "go" often follows "How'd it." (Another kind of machine learning, *unsupervised learning*, works differently, but supervised learning has driven most of the recent progress in the field.)

The process of writing a magazine feature can't easily be distilled into a prediction problem, however — at least not yet. As Sam Bowman, a professor at New York University, told a recent conference on AI and journalism, "The notion of really generating long-form coherent text without a very clear, journalist-specified template is quite far away." Researchers have shown that machine learning can generate coherent text in specific settings, Bowman notes, but "really building systems that are able to go all the way from an abstract idea or a set of facts to a long-form coherent text is still something that's quite difficult."

To illustrate that difficulty, Bowman pointed to a screenplay, titled *Sunspring*, written last year using machine learning. The script was generated by feeding dozens of science fiction screenplays into a neural network — a type of machine learning algorithm — at the character level, meaning that the unit of data the algorithm was learning from was a single character of text. Given the characters that had come before, the


algorithm was learning to predict which character would come next.

The result appears above. You don't want to skip it.

That the actors in *Sunspring* are speaking actual English words is itself impressive (even if they don't make much sense): Before the neural network "read" those scripts, it not only didn't know how to write a screenplay but also had no knowledge of the English language. It learned some of the features of a screenplay — for instance, that lines of text should be assigned to characters and that stage directions should be included. Again, it learned all this just by reading a few dozen scripts.

What it didn't pick up from all those screenplays was the art of narrative. *Sunspring* has no story. Its characters exist only in the sense that they've been given sentences to speak. The script shows how far machine learning has to go before it masters storytelling, or becomes "intelligent." Yet the algorithm's ability to construct sentences and to recognize basic features of a screenplay suggests that AI could play a role in the future of writing. But that future, at least in the near term, is limited.

## AI-GENERATED SUMMARIES

One area of writing in which machine learning is already making useful progress is summaries. Finding the most important parts of a text and producing a summary is an extremely common writing task: Press teams compile "clips" of the day's news, reporters summarize previous developments while writing a story, think tanks summarize a new study, book editors summarize a chapter. Some of that work can now be done by machines, and start-ups and tech companies alike are racing to build tools and products to make it more accessible.

Auto-summarization techniques usually fit one of two categories: *extractive* or *abstractive*. Extractive methods try to identify the most important sentences in a document and then create a summary by stitching them together. Modern versions of this technique are quite complicated, but the original idea, which Hans Peter Luhn introduced at IBM in 1958, gives a sense of the approach. Luhn proposed that the words used most frequently in a document

## THREE SUMMARIES: HUMAN, EXTRACTIVE, AND ABSTRACTIVE

Here are summaries of the Big Idea article "The Business of Artificial Intelligence," by Erik Brynjolfsson and Andrew McAfee.

### Human

"General-purpose technologies," such as the internal combustion engine, have been the fundamental drivers of economic growth for 250 years. Artificial intelligence — particularly machine learning (ML) — is the most important such technology of our era. In the coming decade, practically every industry will transform its keys processes and business models to take advantage of ML. But not all expectations surrounding AI are realistic. In this article the authors describe its real potential, its practical implications, and the barriers to its adoption.

They note three pieces of good news for organizations looking to put ML to use today: AI skills are spreading quickly, through online educational resources as well as universities; the necessary algorithms and hardware for modern AI can be bought or rented as needed; and companies may not need all that much data to start making productive use of ML.

They also note three risks: Machines may have hidden biases, derived from the data used to train them; neural networks deal with statistical rather than literal truths; and diagnosing and correcting system errors is often difficult, because a solution's underlying structure can be unimaginably complex.

### Extractive

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.
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.
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.
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.
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.

—Fast Forward Labs

### Abstractive

companies as diverse as walmart, ups, and uber found ways to manipulate the technology to create profitable new business models. in the sphere of business, ai is poised to have a transformational impact, on the scale of earlier general purpose technologies. in the past few years machine learning has become far more effective and widely available.

—Alexander Rush

(excluding very common words such as "the" and "and") offer clues to the document's subject. Sentences that contain those common words are therefore most representative of the document; by extracting those sentences and combining them into a paragraph, something approximating a summary can be created. (Even in describing this original approach, I'm oversimplifying. For more, see an excellent history of the subfield of Automatic Summarization, by Kathy McKeown, of Columbia University, and Ani Nenkova, of the University of Pennsylvania.)

Abstractive summaries, by contrast, attempt to articulate the information contained in one or more documents in original language written by the algorithm. This approach is more ambitious, and until recently it hasn't worked very well. As the *Sunspring* script illustrates, generating new language is difficult. But progress in deep learning, a subfield of machine learning, has led to renewed interest in abstractive summarization and produced some promising results.

To illustrate what machine learning can and can't do, let's compare an editor-written summary of our AI feature with two automated summaries, one extractive and one abstractive. (See the sidebar "Three Summaries: Human, Extractive, and Abstractive.")

The first summary was written by an HBR editor. It's grammatically correct, it contains the article's main point, and it speaks in the third person ("the authors describe").

The second summary is extractive and was produced using a prototype built by Fast Forward Labs, a research firm. Using actual articles and summaries from a website of reading recommendations, the Fast Forward team trained a neural network to score sentences according to the likelihood that they would be included in a summary. The highest-scoring sentences, combined in the order in which they appear in the original article, become the summary. In the case of our article, the model's highest-scoring sentence is the one beginning "The most important general-purpose technology of our era is artificial intelligence," which is also arguably the article's thesis. In that sense, the extractive summarizer did well. But when the top seven sentences are arranged in their original order, the first sentence includes the pronoun "these" with no mention of what it refers to. (Teaching these systems to recognize the noun to which a pronoun refers is difficult, and the Fast Forward prototype did not attempt it.)

The third summary, courtesy of Alexander Rush, a professor of engineering at Harvard, is abstractive. Rush trained his system to write three-sentence summaries of CNN articles, and although he emphasizes that it isn't state-of-the-art, he offered to try it on the first 450 words of our AI feature. "The system is, in theory, abstractive," he says, "so it can generate anything it wants. In practice, it looks like it is generating mostly sentences it sees in the original article itself." In other words, it avoids the nonsensical results of *Sunspring*, but at the cost of originality. And like the extractive summary, this one captures the article's key themes but includes a reference to "the technology" without providing necessary context.

Are these summaries good enough to replace human-written ones? Perhaps not quite. But that's not the right question. A better one is whether AI-written first drafts of summaries might speed up our process. And here the answer is almost certainly yes.

## AI AS RESEARCH ASSISTANT

Summarization may seem too narrow a task to make much of a difference in the writing process, but combined with related technologies, it creates the opportunity to assist writers in a crucial part of their process: research. And research is "the hardest thing we do as writers," according to David Hill, the editor in chief of SingularityHUB, a niche technology and science publication.

Google, whose search algorithms lean on AI, has already transformed the research process and made writers significantly more productive. But Google isn't a perfect research assistant. Hill describes Google searching as "shallow" and "frenetic." "It's horrifically laborious, all the searching you would do," says Susannah Locke, an editor at Vox.com. She found herself thinking, "Isn't there something that can do this for me?" Tim Lee, of Ars Technica, describes his process of "unstructured" reading: finding 10 to 15 papers on a subject, reading them, and taking notes. He dreams of a tool that can take 1,000 pages on a topic and identify the 10 pages to start with.

The immediate opportunity isn't to fully automate the research process but to make it more structured and efficient. "I don't understand why news sites don't let you just click a name and assemble a backgrounder," says Brian Ulicny, a data scientist at Thomson Reuters Labs. (Disclosure: Ulicny's wife and I are colleagues.) In 2006, while at Lycos, Ulicny authored a paper in which he describes an "information fusion engine." Type in a name or a topic, as you might in Google, and instead of returning a list of links, the system arranges paragraphs from content found across the web into a "coherent summary report or background briefing," what Ulicny calls "something like the level of the first draft of a Wikipedia article."

Ulicny isn't the only one to suggest that topic or news overviews can be automatically generated by software. Computer scientists have been building systems and publishing papers along these lines for more than 15 years. These projects are technically complex, and they vary in important ways. But they face the same challenges and follow a similar process.

Hilary Mason, a data scientist and the founder of Fast Forward Labs, outlines the main tasks these systems must perform: First, they have to identify source data, meaning some number of text documents such as news articles. Then they need to identify the most important information within those documents and extract it. Finally, that information has to be presented to the end user. Somewhere along the way, many of these systems take a fourth step: They try to identify some structure for the story. Is it a chronology of independent events? A biography of a person? Part of a larger story? Structure can both help the system decide what information is important and provide an outline of how to present it to the end user.

This process resembles the way humans approach at least simple research and writing tasks. John O'Neil edits Bloomberg's explanatory QuickTake, but before that he worked on topic pages at the *New York Times*. He describes the process he and his team used to write the text for those topic pages (which have since changed format): First, find four or five key articles the *Times* had published about the subject. Second, identify the background paragraphs in each story (as opposed to the news). Third, write a summary that combined the information from those background paragraphs. At least when writing topic pages, the key steps for both humans and software are the same.

## THE FUTURE OF AI AND WRITING

If these tools have been around for years, even in an imperfect form, why haven't they had more impact on writing? One reason, as with so much revolutionary technology, is culture. On the one hand, many writers don't perceive a need for these tools; on the other, computer scientists haven't always been concerned with how people will actually use their work. In auto-summarization, according to Ani Nenkova, the focus has mostly been on improving accuracy, rather than on thinking about how the technology could be embedded in a tool that people would actually use.

Money is another factor — lots of writers and newsrooms don't have much of it. "Most of the progress [in natural language processing] happened when security analysts and the government were interested in being able to monitor foreign news," says Nenkova, whose PhD was funded by DARPA. Finance, too, is an area in which machine learning and natural language processing have had an impact, in large part because the money has been there to make it happen.

The final reason these tools haven't made more of a dent in writing is simply that the results haven't been good enough to consistently serve readers well on their own. In his paper, Ulicny describes an auto-generated backgrounder for the retired hockey player Mario Lemieux. The system recognized key subtopics that should be part of the explainer, such as "games," "seasons," and "Pittsburgh Penguins." It also added "ice" — a topic that's clearly related in some sense but that no writer would include in a profile of a hockey player.

All of this is changing. The technology is getting both better and easier to use, and more and more writers and media companies are recognizing that smart software can help them do their work. It's clear to me that machine learning does have a near-term role in many types of writing, but for the most part it won't involve producing full-fledged articles. Rather, it will help journalists produce those articles more effectively.

Lots of people are working on tools to make that happen. David Hill has a grant to create an open-source research assistant. Frase, an early-stage start-up in Boston, is working on something similar, though its founders plan to target content marketers as initial customers. Google Docs already has such a tool, but its utility is limited.

Vox built a Slack bot to show writers older articles that they might want to cite in new stories. IBM Watson built a prototype called Watson Angles that summarized news stories, created timelines, and highlighted significant quotations. The prototype, which was removed from the web last fall, also included some key pieces of metadata, such as sentiment analysis of how Reddit users had responded to the news story in question, ranging from positive to negative.

These projects are just the beginning. Imagine a news story on the recent fire in London that mentions the fact that your friend who lives there posted an hour ago that she was safe. Or text that automatically adjusts to the reader's level of background knowledge. Or fact-checking built into a word processor. Or topic pages covering a long tail of niche subjects that smaller audiences are passionate about but that few publishers today can afford to produce. Or a research assistant that recalls a relevant story written a century ago just as readily as one written last week.

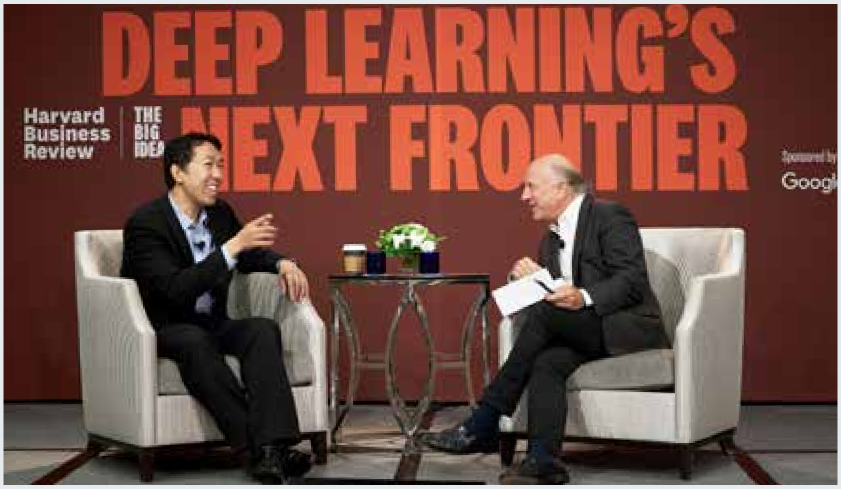
Algorithms still can't craft a narrative the way a person can — they can't write a decent screenplay, or pass Schank's *Romeo and Juliet* test. For the most part, they can't reason about cause and effect. They can't write stirring prose, and they can't persuade a public official to go on the record about an important policy. Still, there's plenty they can do. AI may not be able to tell a great story, but it can help us better tell our own. ■

**About the author: Walter Frick** is a senior associate editor at *Harvard Business Review*. He was a 2016 Knight Visiting Nieman fellow at Harvard University, during which time he researched how machine learning will change the field of explanatory journalism.

# LIVE WEBINAR
# DEEP LEARNING'S NEXT FRONTIER

Watch the recorded event here.

**HBR'S EDITOR IN CHIEF, ADI IGNATIUS, AND ANDREW NG,** former chief scientist at Baidu and a cofounder of Coursera, discuss artificial intelligence and machine learning. They demystify AI and talk about its real-world impact today — and about the most pressing challenges and opportunities it presents for businesses in the future.




PHOTOGRAPHY: ELIE HONEIN

VIDEO

# ARTIFICIAL INTELLIGENCE, REAL FOOD

We asked IBM's AI to create recipes and then had celebrity chef Ming Tsai cook them. Watch what happened. **by *Harvard Business Review* Staff**


## THE GOAL FOR CHEF WATSON, IBM SAYS, IS TO "SURPRISE AND DELIGHT HUMAN CHEFS."

Chef Watson can't chop, dice, or julienne. "He" has no taste buds or appetite. But ask the chef for a recommendation on cooking with green olives, and his knowledge is vast, incorporating data points from a library of recipes and an encyclopedia of flavor profiles.

One of the early applications of IBM's Watson technology, Chef Watson's intelligence is in food. Specifically, how ingredients can come together to form new, never-before-tried recipes. The goal for Chef Watson, IBM says, is to "surprise and delight human chefs."

HBR enlisted two cooks to partner with Watson in the kitchen: Ming Tsai, a renowned professional chef, and Gretchen Gavett, an HBR editor and kitchen novice. We asked each of them to cook with Watson as an experiment in how humans and machines work together. Was it surprising and delightful? Or a recipe for disaster? Watch and find out.