

# Harvard Business Review

PUBLISHED ON HBR.ORG  
NOVEMBER - DECEMBER 2016

## INSIGHT CENTER COLLECTION

# The Next Analytics Age: Machine Learning

SPONSORED BY



# The Next Analytics Age: Machine Learning

What management and leadership challenges will the next wave of analytic technology bring? This Insight Center on HBR.org went beyond the buzz of what machine learning can do, to talk about how it will change companies and the way we manage them.

- 1 How to Make Your Company Machine Learning Ready**  
James Hodson
- 4 Machine Learning Is No Longer Just for Experts**  
Josh Schwartz
- 7 Teaching an Algorithm to Understand Right and Wrong**  
Greg Satell
- 11 The Simple Economics of Machine Intelligence**  
Ajay Agrawal, Joshua Gans, and Avi Goldfarb
- 15 Robots and Automation May Not Take Your Desk Job After All**  
Dan Finnigan
- 18 How to Tell If Machine Learning Can Solve Your Business Problem**  
Anastassia Fedyk

CONTINUES ON NEXT PAGE



country's history. Comparatively, there has been very little discussion of what artificial intelligence *is*, and where we should expect it to actually affect business.

When people talk about AI, machine learning, automation, big data, cognitive computing, or deep learning, they're talking about the ability of machines to learn to fulfill objectives based on data and reasoning. This is tremendously important, and is already changing business in practically every industry. In spite of all the bold claims, there remain several core problems at the heart of Artificial Intelligence where little progress has been made (including learning by analogy, and natural language understanding). Machine learning isn't magic, and the truth is we have neither the data nor the understanding necessary to build machines that make routine decisions as well as human beings.

That may come as a disappointment to some, and potentially disrupt some very expensive marketing campaigns. But the likelihood of self-directed, super-intelligent computational agents emerging in the foreseeable future is extremely low — so keep it out of the yearly business plan for now. Having said that, an enormous amount can already be achieved with the machinery we have today. And that's where forward-thinking managers should be focusing.

Over the next five to 10 years, the biggest business gains will likely stem from getting the right information to the right people at the right time. Building upon the business intelligence revolution of the past years, machine learning will turbocharge finding patterns and automate value extraction in many areas. Data will increasingly drive a real-time economy, where resources are marshaled more efficiently, and the production of goods and services becomes on-demand, with lower failure rates and much better predictability. This will mean different things for different industries.

In services, we will not only get better at forecasting demand, but will learn to provide the right product on a hyper-individualized basis (the Netflix approach).

In retail we will see more sophisticated supply chains, a deeper understanding of consumer preferences, and the ability to customize products and purchase experiences both on- and off-line. Retailers will focus on trend creation and preference formation/brand building.

In manufacturing there will be an evolution towards real-time complete system monitoring, an area known as "anomaly detection." The components will become increasingly connected, allowing for streams of real-time data that machine learning algorithms can use to reveal problems before they happen, optimize the lifetime of components, and reduce the need for human interventions.

In agriculture, data will be used to decide which crops to grow, in what quantities, in what locations, and will render the growing process more efficient year after year. This will create more efficient supply chains, better food, and more sustainable growth with fewer resources.

In short, AI may be a ways off, but machine learning already offers huge potential. So how can managers incorporate it into daily decision-making and longer-term planning? How can a company become *ML-ready*?

**First, catalogue your business processes.** Look for procedures and decisions that are made frequently and consistently, like approving or denying a loan application. Make sure you're collecting as much data as is feasible about how the decision was made, along with any data used to make it. And make sure to collect the decision itself. In the hypothetical loan example, you want to record whether the loan was approved; the data used to make that decision; and any other information about the circumstances behind the decision. (Who made it? At what time of day? How confident did they feel in the decision?) This is the kind of data that can be used to fuel machine learning in the future.

**Second, focus on simple problems.** Automation and machine learning will work well where the problem is well defined and well understood, and where the available data exemplifies the information necessary to make a decision. A good problem for machine learning is identifying a fraudulent transaction. The question "What makes customers feel happy?" is vaguer, more challenging, and not the place to start.

**Third, don't use machine learning where standard business logic will suffice.** Machine learning is useful when the set of rules is unclear, or follows complex, non-linear patterns. If you want transparency and reliability, go for the simplest possible approach that meets your performance criteria.

**Fourth, if a process is complicated, use machine learning to create decision support systems.** If the objective is too unclear to define with respect to the data, try to create intermediate results to help your teams be more effective. You can think of machine learning as part of the hierarchical decision-making path, and it will engender a better understanding of the problem in future.

The point is that there is a lot that can be done without needing to dig very deep. The majority of your workforce will continue to have a job, and you can help them to be more productive, and work on more interesting and demanding (read: *more valuable*) tasks by digitizing more of the mechanical parts of your business. For now, artificial intelligence cannot turn a business's performance from bad to good, but it *can* make some aspects of a *good* business *great*.

If you run out of low hanging fruit — though I'll wager you won't — it may be time to consider building a team to attack more complex problems with machine learning. Be patient, as this investment will not pay off immediately. If you do decide to create such a team, be open, engage with the research community, and you will be contributing to building tomorrow's economy.

---

**James Hodson** is CEO of the [AI for Good Foundation](#), Chairman of the [Financial Data Science Association](#), and an early-stage researcher in Artificial Intelligence and Finance at the [Jozef Stefan Institute AI Lab](#).

ANALYTICS

# Machine Learning Is No Longer Just for Experts

by Josh Schwartz  
OCTOBER 26, 2016



If you're not using deep learning already, you should be. That was the message from legendary Google engineer Jeff Dean at the end of his keynote earlier this year at a conference on web search and data mining. Dean was referring to the rapid increase in machine learning algorithms' accuracy, driven by recent progress in deep learning, and the still untapped potential of these improved algorithms to change the world we live in and the products we build.

But breakthroughs in deep learning aren't the only reason this is a big moment for machine learning. Just as important is that over the last five years, machine learning has become far more accessible to nonexperts, opening up access to a vast group of people.

For most software developers, there have historically been many barriers to entry in machine learning, most notably software libraries designed more for academic researchers than for software engineers as well as a lack of sufficient data. With massive increases in the data being generated and stored by many applications, though, the set of companies with data sets on which machine learning algorithms could be applied has significantly expanded.

In tandem, the last few years have seen a proliferation of cutting-edge, commercially usable machine learning frameworks, including the highly successful [scikit-learn](#) Python library and well-publicized releases of libraries like [Tensorflow](#) by Google and [CNTK](#) by Microsoft Research. The last two years have also seen the major cloud providers Amazon Web Services and Google Cloud Services release machine learning-specific services — both Machine Learning as a Service platforms and graphics processor unit machines optimized for machine learning work.

The net effect of these new technologies is that a person interested in using machine learning need not understand the science of deep learning algorithms in order to experiment with cutting-edge techniques. Tutorials and public code exist for applications as diverse as [AI-driven art generation](#), [language translation](#), and [automated image captioning](#).

The accessibility of this code creates a virtuous cycle. Use by nonexperts creates even more demand for easier-to-use systems and uncovers new applications of machine learning, which inspires further research and development by experts.

And these new technologies affect who works in machine learning as well. When hiring into applied machine learning positions, exceptional quantitative skills are critical, but direct education in machine learning itself has become less important.

In many ways, this change in accessibility mimics the progression we've seen in software development as a whole. Over the last 50 years, software development has gradually migrated from “low-level” languages — highly technical languages that closely relate to a computer's underlying architecture — to high-level languages with significantly lower barriers to entry. Similarly, software deployment has migrated from hosted machines and data centers to cloud-based services, with massive decreases in the time and capital required to deploy a new system.

These changes have not simply made software developers more efficient; they have allowed a much broader set of people to develop software and start software companies. Software bootcamps now train working engineers in a matter of months, and startups can turn ideas into products in a few development cycles.

All of that is not to say, of course, that there's no place for experts — as in software engineering, untold amounts of scientific progress have yet to be made in machine learning. But for the first time in history it's possible, for example, for a person with knowledge of programming but no machine learning experience to create in one afternoon a neural network that can read handwritten digits.

Try it for yourself.

---

**Josh Schwartz** is Chief of Engineering and Data Science at Chartbeat.

---

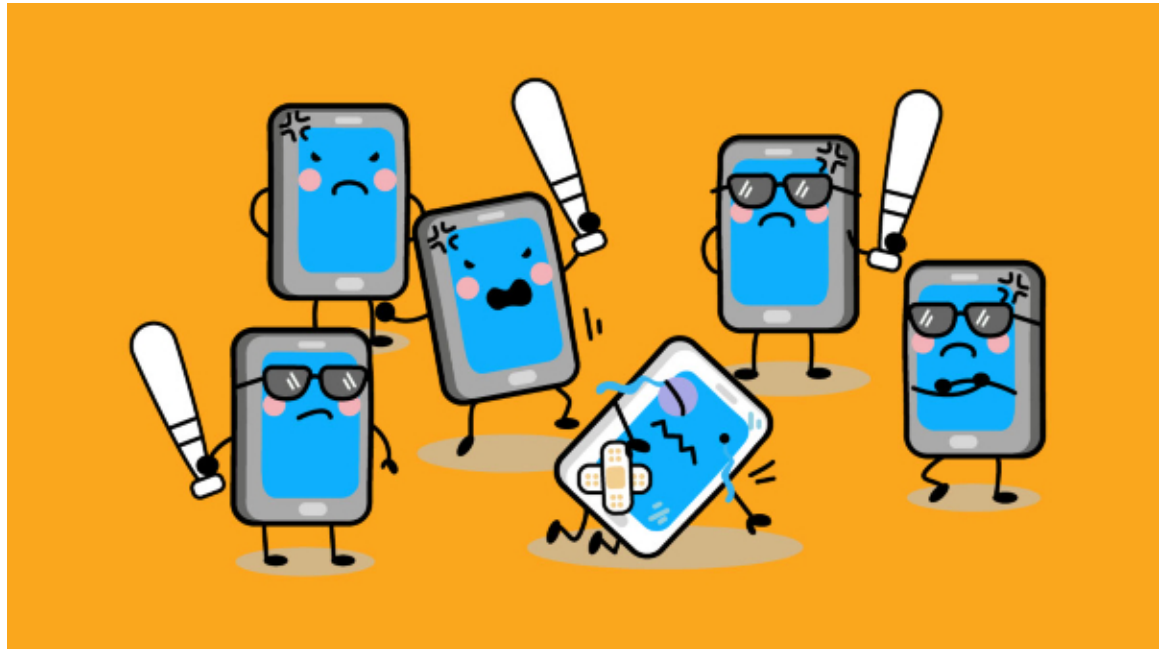


ETHICS

# Teaching an Algorithm to Understand Right and Wrong

by Greg Satell

NOVEMBER 15, 2016



In his *Nicomachean Ethics*, Aristotle states that it is a fact that “all knowledge and every pursuit aims at some good,” but then continues, “What then do we mean by the good?” That, in essence, encapsulates the ethical dilemma. We all agree that we should be good and just, but it’s much harder to decide what that entails.

Since Aristotle's time, the questions he raised have been continually discussed and debated. From the works of great philosophers like [Kant](#), [Bentham](#), and [Rawls](#) to modern-day cocktail parties and late-night dorm room bull sessions, the issues are endlessly mulled over and argued about but never come to a satisfying conclusion.

Today, as we enter a “cognitive era” of thinking machines, the problem of what should guide our actions is gaining newfound importance. If we find it so difficult to denote the principles by which a person should act justly and wisely, then how are we to encode them within the artificial intelligences we are creating? It is a question that we need to come up with answers for soon.

### **Designing a Learning Environment**

Every parent worries about what influences their children are exposed to. What TV shows are they watching? What video games are they playing? Are they hanging out with the wrong crowd at school? We try not to overly shelter our kids because we want them to learn about the world, but we don't want to expose them to too much before they have the maturity to process it.

In artificial intelligence, these influences are called a “machine learning corpus.” For example, if you want to teach an algorithm to recognize cats, you expose it to thousands of pictures of cats and things that are not cats. Eventually, it figures out how to tell the difference between, say, a cat and a dog. Much as with human beings, it is through learning from these experiences that algorithms become useful.

However, the process can go horribly awry, as in the case of [Microsoft's Tay](#), a Twitter bot that the company unleashed on the microblogging platform. In under a day, Tay went from being friendly and casual (“Humans are super cool”) to downright scary (“Hitler was right and I hate Jews”). It was profoundly disturbing.

[Francesca Rossi](#), an AI researcher at IBM, points out that we often encode principles regarding influences into societal norms, such as what age a child needs to be to watch an R-rated movie or whether they should learn evolution in school. “We need to decide to what extent the legal principles that we use to regulate humans can be used for machines,” she told me.

However, in some cases [algorithms can alert us to bias](#) in our society that we might not have been aware of, such as when we [Google “grandma”](#) and see only white faces. “There is a great potential for machines to alert us to bias,” Rossi notes. “We need to not only train our algorithms but also be open to the possibility that they can teach us about ourselves.”

### **Unraveling Ethical Dilemmas**

One thought experiment that has puzzled ethicists for decades is the [trolley problem](#). Imagine you see a trolley barreling down the tracks and it's about to run over five people. The only way to save them is to pull a lever to switch the trolley to a different set of tracks, but if you do that, one person standing on the other tracks will be killed. What should you do?

Ethical systems based on moral principles, such as Kant's [Categorical Imperative](#) (act only according to that maxim whereby you can, at the same time, will that it should become a universal law) or [Asimov's first law](#) (a robot may not injure a human being or, through inaction, allow a human being to come to harm) are thoroughly unhelpful here.

Another alternative would be to adopt the [utilitarian principle](#) and simply do what results in the most good or the least harm. Then it would be clear that you should kill the one person to save the five. However, the idea of killing somebody intentionally is troublesome, to say the least. While we do apply the principle in some limited cases, such as in the case of a Secret Service officer's duty to protect the president, those are rare exceptions.

The rise of artificial intelligence is forcing us to take abstract ethical dilemmas much more seriously because we need to code in moral principles concretely. Should a self-driving car risk killing its passenger to save a pedestrian? To what extent should a drone take into account the risk of collateral damage when killing a terrorist? Should robots make life-or-death decisions about humans at all? We will have to make concrete decisions about what we will leave up to humans and what we will encode into software.

These are tough questions, but IBM's Rossi points out that machines may be able to help us with them. Aristotle's teachings, often referred to as [virtue ethics](#), emphasize that we need to learn the meaning of ethical virtues, such as wisdom, justice, and prudence. So it is possible that a powerful machine learning system could provide us with new insights.

### **Cultural Norms vs. Moral Values**

Another issue that we will have to contend with is that we will have to decide not only *what* ethical principles to encode in artificial intelligences but also *how* they are coded. As noted above, for the most part, "Thou shalt not kill" is a strict principle. Other than a few rare cases, such as the Secret Service or a soldier, it's more like a preference that is greatly affected by context.

There is often much confusion about what is truly a moral principle and what is merely a cultural norm. In many cases, as with LGBT rights, societal judgments with respect to morality change over time. In others, such as teaching creationism in schools or allowing the sale of alcohol, we find it reasonable to let different communities make their own choices.

What makes one thing a moral value and another a cultural norm? Well, that's a tough question for even the most-lauded human ethicists, but we will need to code those decisions into our algorithms. In some cases, there will be strict principles; in others, merely preferences based on context. For some tasks, algorithms will need to be coded differently according to what jurisdiction they operate in.

The issue becomes especially thorny when algorithms have to make decisions according to conflicting professional norms, such as in medical care. How much should cost be taken into account

when regarding medical decisions? Should insurance companies have a say in how the algorithms are coded?

This is not, of course, a completely new problem. For example, firms operating in the U.S. need to abide by GAAP accounting standards, which rely on strict rules, while those operating in Europe follow IFRS accounting standards, which are driven by broad principles. We will likely end up with a similar situation with regard to many ethical principles in artificial intelligences.

### **Setting a Higher Standard**

Most AI experts I've spoken to think that we will need to set higher moral standards for artificial intelligences than we do for humans. We do not, as a matter of course, expect people to supply a list of influences and an accounting for their logic for every decision they make, unless something goes horribly wrong. But we will require such transparency from machines.

“With another human, we often assume that they have similar common-sense reasoning capabilities and ethical standards. That’s not true of machines, so we need to hold them to a higher standard,” Rossi says. Josh Sutton, global head, data and artificial intelligence, at [Publicis.Sapient](#), agrees and argues that both the logical trail and the learning corpus that lead to machine decisions need to be made available for examination.

However, Sutton sees how we might also opt for less transparency in some situations. For example, we may feel more comfortable with algorithms that make use of our behavioral and geolocation data but don't let humans access that data. Humans, after all, can always be tempted. Machines are better at following strict parameters.

Clearly, these issues need further thought and discussion. Major industry players, such as Google, IBM, Amazon, and Facebook, [recently set up a partnership](#) to create an open platform between leading AI companies and stakeholders in academia, government, and industry to advance understanding and promote best practices. Yet that is merely a starting point.

As pervasive as artificial intelligence is set to become in the near future, the responsibility rests with society as a whole. Put simply, we need to take the standards by which artificial intelligences will operate just as seriously as those that govern how our political systems operate and how are children are educated.

It is a responsibility that we cannot shirk.

---

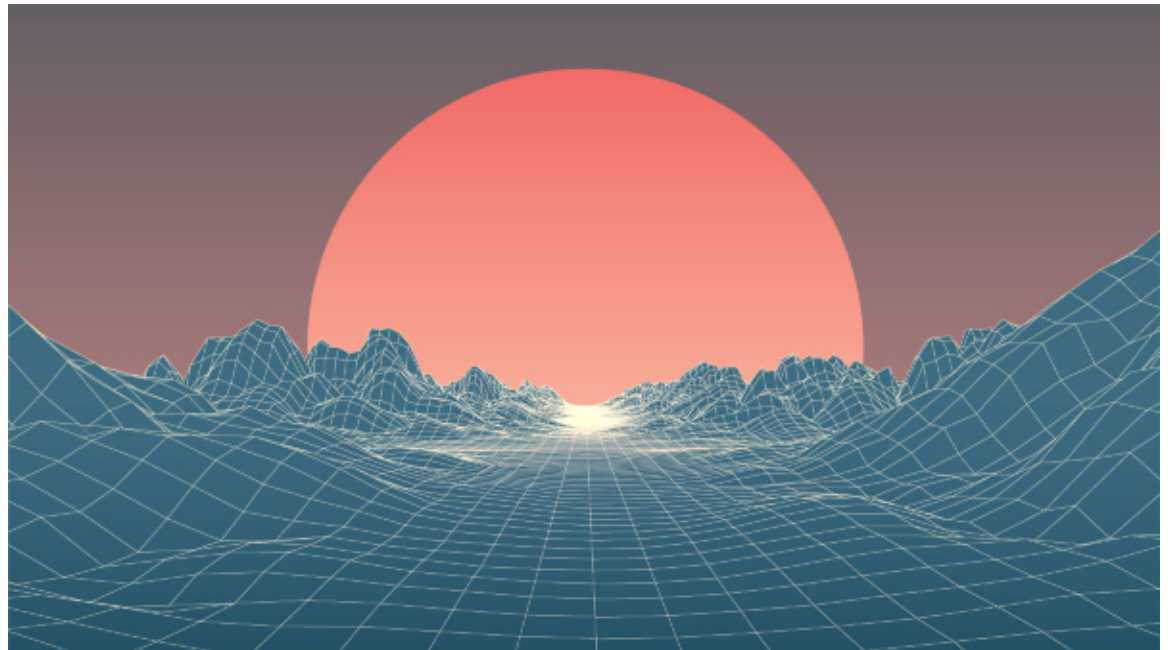
**Greg** Satell is a popular speaker and consultant. His first book, *Mapping Innovation: A Playbook for Navigating a Disruptive Age*, is coming out in 2017. Follow his blog at [Digital Tonto](#) or on twitter [@DigitalTonto](#).

---

ECONOMICS

# The Simple Economics of Machine Intelligence

by Ajay Agrawal, Joshua Gans and Avi Goldfarb  
NOVEMBER 17, 2016



The year 1995 was heralded as the beginning of the “New Economy.” Digital communication was set to upend markets and *change everything*. But economists by and large didn’t buy into the hype. It wasn’t that we didn’t recognize that something changed. It was that we recognized that the old economics lens remained useful for looking at the changes taking place. The economics of the “New Economy” could be described at a high level: Digital technology would cause a reduction in the cost of search and communication. This would lead to more search, more communication, and more activities that go together with search and communication. That’s essentially what happened.

Today we are seeing similar hype about machine intelligence. But once again, as economists, we believe some simple rules apply. Technological revolutions tend to involve some important activity becoming cheap, like the cost of communication or finding information. Machine intelligence is, in its essence, a prediction technology, so the economic shift will center around a drop in the cost of prediction.

The first effect of machine intelligence will be to lower the cost of goods and services that rely on prediction. This matters because prediction is an input to a host of activities including transportation, agriculture, healthcare, energy manufacturing, and retail.

When the cost of any input falls so precipitously, there are two other well-established economic implications. First, we will start using prediction to perform tasks where we previously didn't. Second, the value of other things that complement prediction will rise.

### **Lots of tasks will be reframed as prediction problems**

As machine intelligence lowers the cost of prediction, we will begin to use it as an input for things for which we never previously did. As a historical example, consider semiconductors, an area of technological advance that caused a significant drop in the cost of a different input: arithmetic. With semiconductors we could calculate cheaply, so activities for which arithmetic was a key input, such as data analysis and accounting, became much cheaper. However, we also started using the newly cheap arithmetic to solve problems that were not historically arithmetic problems. An example is photography. We shifted from a film-oriented, chemistry-based approach to a digital-oriented, arithmetic-based approach. Other new applications for cheap arithmetic include communications, music, and drug discovery.

The same goes for machine intelligence and prediction. As the cost of prediction falls, not only will activities that were historically prediction-oriented become cheaper — like inventory management and demand forecasting — but we will also use prediction to tackle other problems for which prediction was not historically an input.

Consider navigation. Until recently, autonomous driving was limited to highly controlled environments such as warehouses and factories where programmers could anticipate the range of scenarios a vehicle may encounter, and could program if-then-else-type decision algorithms accordingly (e.g., “If an object approaches the vehicle, then slowdown”). It was inconceivable to put an autonomous vehicle on a city street because the number of possible scenarios in such an uncontrolled environment would require programming an almost infinite number of if-then-else statements.

Inconceivable, that is, until recently. Once prediction became cheap, innovators reframed driving as a prediction problem. Rather than programming endless if-then-else statements, they instead simply asked the AI to predict: “What would a human driver do?” They outfitted vehicles with a variety of sensors – cameras, lidar, radar, etc. – and then collected millions of miles of human driving data. By

linking the incoming environmental data from sensors on the outside of the car to the driving decisions made by the human inside the car (steering, braking, accelerating), the AI learned to predict how humans would react to each second of incoming data about their environment. Thus, prediction is now a major component of the solution to a problem that was previously not considered a prediction problem.

### **Judgment will become more valuable**

When the cost of a foundational input plummets, it often affects the value of other inputs. The value goes up for complements and down for substitutes. In the case of photography, the value of the hardware and software components associated with digital cameras went up as the cost of arithmetic dropped because demand increased – we wanted more of them. These components were complements to arithmetic; they were used together. In contrast, the value of film-related chemicals fell – we wanted less of them.

All human activities can be described by five high-level components: data, prediction, judgment, action, and outcomes. For example, a visit to the doctor in response to pain leads to: 1) x-rays, blood tests, monitoring (data), 2) diagnosis of the problem, such as “if we administer treatment A, then we predict outcome X, but if we administer treatment B, then we predict outcome Y” (prediction), 3) weighing options: “given your age, lifestyle, and family status, I think you might be best with treatment A; let’s discuss how you feel about the risks and side effects” (judgment); 4) administering treatment A (action), and 5) full recovery with minor side effects (outcome).

As machine intelligence improves, the value of human prediction skills will decrease because machine prediction will provide a cheaper and better substitute for human prediction, just as machines did for arithmetic. However, this does not spell doom for human jobs, as many experts suggest. That’s because the value of human judgment skills will increase. Using the language of economics, judgment is a complement to prediction and therefore when the cost of prediction falls demand for judgment rises. We’ll want more human judgment.

For example, when prediction is cheap, diagnosis will be more frequent and convenient, and thus we’ll detect many more early-stage, treatable conditions. This will mean more decisions will be made about medical treatment, which means greater demand for the application of ethics, and for emotional support, which are provided by humans. The line between judgment and prediction isn’t clear cut – some judgment tasks will even be reframed as a series of predictions. Yet, overall the value of prediction-related human skills will fall, and the value of judgment-related skills will rise.

Interpreting the rise of machine intelligence as a drop in the cost of prediction doesn’t offer an answer to every specific question of how the technology will play out. But it yields two key implications: 1) an expanded role of prediction as an input to more goods and services, and 2) a change in the value of other inputs, driven by the extent to which they are complements to or substitutes for prediction. These changes are coming. The speed and extent to which managers should invest in judgment-related capabilities will depend on the how fast the changes arrive.

---

**Ajay Agrawal** is the Peter Munk Professor of Entrepreneurship at the Rotman School of Management, University of Toronto. He is also a Research Associate at the National Bureau of Economic Research in Cambridge, MA, co-founder of The Next 36, founder of Creative Destruction Lab, and Academic Director of the Centre for Innovation and Entrepreneurship.

---

**Joshua Gans** is professor of strategic management at the Rotman School of Management. His latest book, [\*The Disruption Dilemma\*](#), is published by MIT Press.

---

**Avi Goldfarb** is the Ellison Professor of Marketing at the Rotman School of Management, University of Toronto. He is also a Research Associate at the National Bureau of Economic Research, Chief Data Scientist at the Creative Destruction Lab, and Senior Editor at Marketing Science.

---

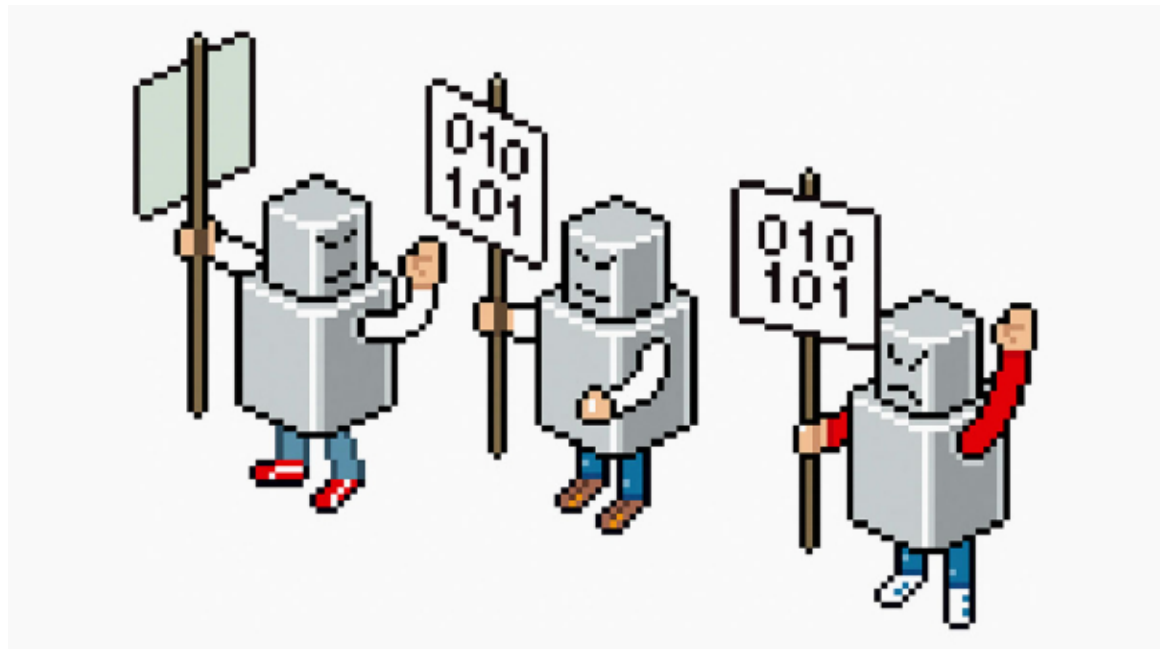


HIRING

# Robots and Automation May Not Take Your Desk Job After All

by Dan Finnigan

NOVEMBER 22, 2016



Despite talk of job automation in practically every industry, the question no longer seems to be *whether* jobs will be automated, but rather *which* jobs are at greatest risk and how concerned we should be about the changes. But is this anxiety actually based in reality?

Let's be clear: People *are* afraid. According to the [Job Seeker Nation Report](#) released by Jobvite earlier this year, 55% of job seekers are at least a little worried about job automation. But it turns out that

those actually in charge of hiring aren't there yet. Jobvite's [Recruiter Nation Report](#) also says that just 10% of recruiters anticipate automating some jobs in the next 2–3 years (compared to 2015's 25%).

We are entering the second wave of the jobs automation process, Automation 2.0. We've seen predictable, repeatable, physical tasks (think forklifts on the factory floor or assembly-line work) be automated. But now, just as the recruiters predicted, automating the next wave of cognitive work (think diagnosing diseases or reviewing legal documents) will prove much more difficult, slowing the process until we figure out how to do it.

We don't need to fear this new wave, though — because humans occupy a space that machines probably never will. They manage each other, they take responsibility for actions, and they understand the dynamics of working relationships in a way that an algorithm simply can't (at least for the foreseeable future). Automation 2.0 will undoubtedly change the labor market, but in a way we can adapt to.

### **The Realities of Automation 2.0**

We're never going to automate every job out of existence. Human beings are relentless in their capacity to invent new ways to serve others, so we will always have new work. A recent [McKinsey study](#) shows that there really are tasks that only humans can do for the foreseeable future, particularly in areas such as education and health care. Among them are managing others, applying expertise — such as knowledge of the stock market or a knack for creative messaging — and interacting with key stakeholders. There are still aspects of many jobs that cannot be automated away, so let's all take a deep breath for a second.

It's important to recognize that we will inevitably automate *parts* of almost every job. We've seen it happen already across manufacturing, food service, and retail operations. As we move into the cognitive work arena, the idea will be the same: We'll see automation of many predictable, scalable, and repeatable tasks.

But while computers can do the rote analysis, even they have limits. Instead of physical work, these tasks will be related to things such as data collection or the processing of algorithms. My own doctor doesn't talk to me unless he's sitting in front of a terminal. Much of what he used to do as a rote part of seeing patients is now automated. In education, many of the daily tasks that have beleaguered teachers can now be given to computers. In fact, entire university educations can be delivered online. Does that mean we don't need doctors or teachers? No.

As this next wave of automation transpires, it will unveil the real value of human beings at work. Humans have an understanding of the dynamic nuances inherent in working relationships and the ability to respond and adapt to those nuances; machines probably will not get there. This is where strong managers and mentors excel. Humans can provide judgment and hold each other accountable

for their actions. This is why [a single fatal accident with a self-driving car](#) reminds us that machines, for all their excellent attributes, cannot be blamed or accept responsibility.

While fears about Automation 2.0 aren't as founded as we thought, the movement will impact the labor market significantly. As technology enables experts to offload rote tasks and analysis, they gain time to tend to the other work that *cannot* be automated – the work that makes them uniquely necessary. In that sense, people performing expert cognitive labor will become significantly more productive.

It's already happening. We're seeing doctors who can effectively oversee the treatment of hundreds of patients because they have computers. We're seeing lawyers who can effectively handle hundreds of clients because they have technology to assist in research and analysis.

But as people in these professions leverage technology to apply their expertise, we likely won't need as many of them. And those who succeed will *have* to be people who are extremely well prepared to function in a technological and automated work environment. In fact, they might not even need to go as far in their expert education to perform these jobs. With the right technology skills, and with automation in play, we're learning that physician assistants and paralegals can do the same work with less investment.

These changes aren't negative, but people concerned about how the automation of cognitive work might impact their careers can't stick their heads in the sand. If we can continue to innovate and accept technology in the labor market and better incorporate technological skills into our education system, then we will continue to create the kinds of lucrative opportunities that only humans can effectively seize.

---

**Dan Finnigan** is the CEO of [Jobvite](#).

---

ANALYTICS

# How to Tell If Machine Learning Can Solve Your Business Problem

by Anastassia Fedyk  
NOVEMBER 25, 2016



“AI,” “big data,” and “machine learning” are all trending buzzwords, and you might be curious about how they apply to your domain. You might even have startups beating down your door, pitching you their new “AI-powered” product. So how can you know which problems in your business are amenable to machine learning? To decide, you need to think about the problem to be solved and the available data, and ask questions about feasibility, intuition, and expectations.

Start by distinguishing between automation problems and learning problems. Machine learning can help automate your processes, but not all automation problems require learning.

Automation without learning is appropriate when the problem is relatively straightforward. These are the kinds of tasks where you have a clear, predefined sequence of steps that is currently being executed by a human, but that could conceivably be transitioned to a machine. This sort of automation has been happening in businesses for decades. Screening incoming data from an outside data provider for well-defined potential errors is an example of a problem ready for automation. (For example, hedge funds automatically filtering out bad data in the form of a negative value for trading volume, which can't be negative.) On the other hand, encoding human language into a structured dataset is something that is just a tad too ambitious for a straightforward set of rules.

For the second type of problems, standard automation is not enough – they require learning from data. And we now venture into the arena of machine learning. Machine learning, at its core, is a set of statistical methods meant to find patterns of predictability in datasets. These methods are great at determining how certain features of the data are related to the outcomes you are interested in. What these methods cannot do is access any knowledge outside of the data you provide. For example, researchers at the University of Pittsburgh in the late 1990s [evaluated machine learning algorithms](#) for predicting mortality rates from pneumonia. The algorithms recommended that hospitals send home pneumonia patients who were also asthma sufferers, estimating their risk of death from pneumonia to be lower. It turned out that the dataset fed into the algorithms did not account for the fact that asthma sufferers had been immediately sent to intensive care, [and had fared better only due to the additional attention](#).

So what are good business problems for machine learning methods? Essentially, any problems that: (1) require prediction rather than causal inference; and (2) are sufficiently self-contained, or relatively insulated from outside influences. The first means that you are interested in understanding how, on average, certain aspects of the data relate to each other, and not in the causal channels of their relationship. Keep in mind that the statistical methods do not bring to the table the intuition, theory, or domain knowledge of human analysts. The second means that you are relatively certain that the data you feed to your learning algorithm includes more or less all there is to the problem. If, in the future, the thing you're trying to predict changes unexpectedly – and no longer matches prior patterns in the data – the algorithm will not know what to make of it.

Examples of good machine learning problems include predicting the likelihood that a certain type of user will click on a certain kind of ad, or evaluating the extent to which a piece of text is similar to previous texts you have seen.

Bad examples include predicting profits from the introduction of a completely new and revolutionary product line, or extrapolating next year's sales from past data, when an important new competitor just entered the market.

Once you verify that your problem is suitable for machine learning, the next step is to evaluate whether you have the right data to solve it. The data might come from you, or from an external provider. In the latter case, make sure to ask enough questions to get a good feel for the data's scope and whether it is likely to be a good fit for your problem.

Say you have determined that your problem is the classic machine learning problem and you have the data to fit that problem. The last step of the process is your intuition check. Yes, intuition: machine learning methods, however proprietary and seemingly magical, are statistics. And statistics *can* be explained in intuitive terms. Instead of trusting that the brilliant proposed method will seamlessly work, ask lots of questions.

Get yourself comfortable with how the method works. Does the intuition of the method roughly make sense? Does it fit, conceptually, in your framework of the particular setting or problem you are dealing with? What makes this method especially well suited to your problem? If you are encoding a set of steps, perhaps sequential models or decision trees are a good choice. If you need to separate two classes of outcome, perhaps a binary support vector machine would be best aligned with your needs.

With understanding come more realistic expectations. Once you ask enough questions and receive enough answers to have an intuitive understanding of how the methodology works, you will see that it is far from magical. Every human makes mistakes, and every algorithm is error prone, too. For all but the simplest of problems, there *will* be times when things go wrong. The machine learning prediction engine will get things right *on average* but will reliably make mistakes. Mistakes will happen, and they will happen most often in ways that you cannot anticipate.

So the last step is to evaluate the extent to which you can allow for exceptions or statistical errors in your process. Is your problem the kind of problem where getting things right 80% of the time is enough? Can you deal with a 10% error rate? 5%? 1%? Are there certain kinds of errors that should never be allowed? Be clear and upfront about your needs and expectations, both with yourself and with your solution-provider. And once both of you are comfortably on the same page, go ahead. Armed with knowledge, understanding, and reasonable expectations, you are set to reap the benefits of machine learning. Just please be patient.

---

**Anastassia Fedyk** is a Ph.D. Candidate in business economics at Harvard Business School. Her research focuses on finance and behavioral economics.

---

# Leadership and Big Data Innovation

**PRESENTER:**

Michael Schrage, Fellow, MIT Sloan School's Initiative on the Digital Economy

**MODERATOR:**

Angelia Herrin, Editor, Special Projects and Research, *Harvard Business Review*

## Overview

As devices, sensors and new technology platforms generate ever more big data, organizations have new opportunities to gain important insights that can spur innovation, bring new products and services to the market and create a competitive edge. But unless organizations begin to treat data as a key asset that is just as important as their financial assets, those opportunities will be missed.

To survive and thrive in the big data era, organizations must implement new governance models, create dashboards of information shared across the company, encourage an environment of “experiment and scale,” and cultivate a culture that incorporates data into the everyday activities of the business. And the message needs to come from the top: the kind of data-driven transformation that most companies need will demand commitment from the C-suite.

## Context

Michael Schrage discussed how leaders can promote innovation through big data and analytics.

## Key Takeaways

**To create new value, organizations must treat data as an asset.**

In many cases, organizations use analytics to optimize existing sources of value, rather than leveraging data and analytics to innovate and create new value. To create new value, organizations must treat data as an asset, just like capital and human resources.

In just the next 12 to 18 months, organizations can expect to gather 10 to 1,000 times more data. As data assets grow, opportunities exist to create new value through more efficient processes, new value propositions, enhanced customer experiences, and more. Organizations must ask themselves how they would sell or market to customers and prospects differently, if they had 100 times more data about them.

“Organizations should expect 10 to 1,000 times more data in the next 12 to 18 months. To make data more valuable, organizations must consider how to define, measure, and assess value creation inside the enterprise and outside.” —Michael Schrage

**Purposeful analytics like KPIs and dashboards support innovation and value creation.**

Purposeful data and analytics can be applied to a variety of business activities, such as customer retention, upselling and bundling, and identifying new products and services. Both KPIs and dashboards support purposeful analytics.

To harness data to outcomes that matter, enterprises must align application programming interfaces (APIs) with KPIs. It is also important to think about dashboards related to innovation and new value creation. Key questions include what these dashboards would look like and which KPIs they would include.

From an organizational perspective, siloed organizations lead to siloed KPIs. Schrage suggests viewing data as a “solvent” that breaks down siloes and creates virtuous cycles within organizations. Insights gathered from analytics related to customer retention, for example, may inform analytics related to innovation.

**Big data requires “little experiments.”**

When an insight is identified in data, the next step is to determine which tests will validate whether the insight is true or not. More and better experimentation is needed. This means socializing experiments across the enterprise. New data should facilitate collaboration around experiments focused on new value creation.

“Organizations need a fundamental paradigm shift. Innovation will no longer be based on an R&D pipeline, but on an ‘experiment and scale’ model. Experiments are the future of business innovation.”  
—Michael Schrage

**Data leadership demands data governance.**

Data oversight is the next leadership challenge for organizations. Explicit data governance is needed to determine who has access to data, who can experiment with it, and more. Observations related to data governance include:

- *Data stewards are a new category of colleague.* The data steward role involves planning, implementing, and managing the sourcing, use, and maintenance of data assets in an organization.



Data stewards enable organizations to take control and govern all the types and forms of data and their associated libraries or repositories.

- *Data governance plays a major role in company growth.* Serious growth and value companies have C-suites that are committed to data as an asset and to data management.
- *Data governance working groups must be unified around the idea that data resources can be transformed into new value.* Data governance working groups create the vision for the organization. It is important to have members who are committed to collaboration, data sharing, customer data protection, and developing data as an asset. Schrage recommends against having CFOs run data governance working groups, as they may overweight capital as an asset.



FIGURE 1: ELEMENTS OF DATA GOVERNANCE

### **As data volumes increase, the importance of an enterprise data culture grows.**

To promote an enterprise “value of data” culture, organizations often need to make behavioral, cultural, and operational improvements. Examples of these shifts include:

- *API development.* To generate value, people expect to share data sets across the organization. APIs are essential to accomplish this.
- *Testing “gut feel” ideas with data-based experiments.* Many senior managers make decisions based on “gut feel,” rather than using experiments to challenge their perceptions. Ideas based on intuition aren’t bad, but they must be accompanied by testable business hypotheses. As Schrage noted, “Many good ideas evaporate on contact with the real world.” Data helps extract signals from the noise.
- *Deploying incentives to promote information sharing.* If data is equated with power, some people are inclined to hoard it. To promote information sharing at a large telecom, Schrage created an internal competition called “we were robbed” and “thief of the week.” These programs rewarded people for “stealing” data from other parts of the organization to create new value.
- *Revising performance review criteria.* Employees may be evaluated on how well they share data, as well as to what degree their decision making is driven by richer sources of data over time.

### **Other Important Points**

**Self-quantification.** Schrage believes that individuals who engage in self-improvement every day will have an edge in the job market. People need to be more introspective about ways they add value in the workplace and must use data as a mirror to see themselves differently. For example, at the end of the day employees could analyze the content of their emails and texts using a Wordle. Being data driven and self-aware are two sides of the same coin.

**Data and risk appetite.** Organizations may use data to rethink their risk appetite. Insights derived from data are one way to create “guardrails” that can increase comfort with risk. Major “bet the company” decisions should always be data driven.

**Data scientists.** Some organizations place data scientists in a center of excellence, while others implant them in different groups. Regardless of where data scientists “live” in the organization, they should be expected to collaborate with the business and serve as a resource, like a good financial expert. Data scientists help teams achieve their goals by drawing on data as an asset.

## FROM OUR SPONSOR

SAS continues to foster innovation through predictive analytics across industries. Leading organizations are bringing about the next analytics age by working with SAS for real-time streaming analytics, in-memory analytics, self-service analytics for business users, and machine learning applications that solve today's business problems and propel their organizations forward. Frequently sought out for quality, scalability, and compatibility of its solutions with open-source platforms, turn to SAS to make your new possible happen with analytics.

### Start the journey to your new possible with these suggested next steps:

---

#### **THE EVOLUTION OF ANALYTICS: OPPORTUNITIES AND CHALLENGES FOR MACHINE LEARNING IN BUSINESS →**

This report by O'Reilly Media addresses modern applications for machine learning, including recommendation systems, streaming analytics, deep learning and cognitive computing. It includes details about the experiences of two companies that have successfully navigated both organizational and technological challenges to adopt machine learning and embark on their own analytics evolution.

---

#### **TDWI BEST PRACTICES REPORT: OPERATIONALIZING AND EMBEDDING ANALYTICS FOR ACTION →**

This report, explores how forward-looking organizations are beginning to not only analyze their data but also take action on it to derive business value by taking into account organizational and technical dimensions, such as infrastructure, culture, data and processes. It addresses challenges and how organizations are overcoming them, and provides recommendations for successfully operationalizing analytics in the organization.

SPONSORED BY 