Managing the Machines
AI is making prediction cheap, posing new challenges for managers

by Ajay Agrawal, Joshua Gans, and Avi Goldfarb

7 October 2016

Introduction

We are living through a renaissance in artificial intelligence (AI). The major tech companies (e.g. Google, Facebook, Amazon, Microsoft, Tesla, and Apple) are prominently including AI in their product launches. They and others are acquiring AI-based startups by the dozen. Depending on how you look at it, this conjures up images that range from augmentation of human labor to replacement with the consequent follow-on logic of excitement to fear. In each case, proponents point to an advance — an AI that plays Go or one that can drive a truck — and define a future path by extrapolating from the human tasks enhanced or replaced. But economic history has taught us that this is a flawed approach. Instead, when looking to assess the impact of radical technological change, one approach stands out: ask yourself, what is this reducing the cost of? It is only then that you can figure out what really might change.

To understand how important this framing can be, let’s step back one technological revolution ago and ask that same question. Moore’s Law — the 18-month doubling of transistor intensity on microprocessors — has dominated information technology for the past four decades. What did those advances reduce the cost of? The answer: arithmetic.

This answer may seem surprising as computers appear to do much more. They allow us to communicate, to play games and music, to design and to create art. This is all true but at their heart computers are direct descendants of electronic calculators. That they appear to do more is testament to the power of arithmetic. In their earliest
days, this relationship was more obvious. Computers focused on arithmetic operations related to the census, artillery tables, and other largely military applications. Prior to the invention of the digital computer, “computers” were humans who spent their days solving arithmetic problems related to these applications. What digital computers did was make arithmetic so inexpensive that thousands of new applications for arithmetic were implemented and discovered; data storage, word processing, and digital photography are all novel applications of arithmetic.

AI presents a similar opportunity: to make something that was once expensive very cheap, and with it, to make resources that were once scarce now abundant. For AI, that particular task is prediction. That is, the ability to take information you have and generate information you do not have.

To reduce the economics of AI down to a single thing may seem like a tall order. To be sure, for many who have looked at the emerging applications of AI, our take may seem far from obvious. The purpose of this article is, however, to convince you otherwise. What we will argue here is that AI — in its modern incarnation — is all about advances in prediction. Based on this thesis, we will then explore the impact of advances in AI on the nature and direction of applications coming from it: how AI will be used as an input for traditionally non-prediction-oriented problems, how the value of some human skills will rise while others fall as AI advances, and the implications for managers. These speculations will be informed by earlier radical technological changes that similarly involved the near elimination of the costs of particular tasks. As we will demonstrate, this allows us to understand how AI is likely to change what workers and managers do.

### Machine Learning and Prediction

The recent advances in AI come under the rubric of a field now called “machine learning.” This involves programming computers to learn from example data or past experience. It is most useful in cases where we cannot directly write a computer program to solve a given problem. In other words, for some tasks, we do not have an algorithm and programming such an algorithm is not feasible. For example, while humans are good at recognizing the faces of their friends, we cannot explain our expertise and therefore cannot write a computer program directly. Machine learning solves this problem by analyzing example face images and determining the pattern that is specific to a face.
This approach to learning is not optimal for all knowledge. But it does turn out to be a highly efficient process for understanding what outcomes will result from a large number of quantifiable observations. Consider the job of identifying objects in a basket of groceries. If we humans could describe what, say, an apple looks like and compare that to an orange, then we could program a computer to recognize both objects; say based on some color and shape classifications. But a basket of groceries includes other objects some of which are apple-like in colour and shape. It may be possible to continue encoding our knowledge of apples in finer detail but as we move towards environments in the real world, potential complexity increases exponentially. In this respect, one starts to appreciate just how hard these jobs can be and how amazing it is that humans can do them so easily.

It is in environments with this type of complexity that machine learning is most useful. In one type of training, the machine starts with a set of objects that it is shown pictures with names attached to them. It is then shown millions of pictures that may or may not include those objects but each with objects named. As a result, it notices correlations. For instance, that objects named “apples” tend to be red. Of course, if the apple is sometimes eaten or half-eaten that may be trickier. But other objects may be red so the machine looks for other correlates within its range of classifiers — shapes, texture, and then most importantly context. The final one requires some sophisticated machine learners, but a red round object is more likely to be an apple in a basket of fruit than in a ball pit at a play centre.

What is happening under the hood is that the machine uses information from past images of apples to predict whether the current image contains an apple. Why use the word “predict”? Prediction uses information you have to generate information you do not have. Machine learning uses data collected from sensors, images, videos, typed notes, or anything else that can be represented in bits. This is information you have. It uses this information to fill in missing information, to recognize objects, and to predict what will happen next. This is information you do not have. In other words, machine learning is a prediction technology.

There are a large variety of possible predictions. We most often think about prediction as determining what will happen in the future. For example, machine learning can be used to predict whether a bank customer will default on a loan. There is also a great deal of useful data we do not have that is not about the future. For example, mass retailer Target predicted which of its customers were pregnant based on their purchasing behavior. This was not predicting the future: The customers were already pregnant. Instead, it was filling in missing data in a way that proved useful to the company. Similarly, medical diagnosis is a prediction problem: using data on symptoms
to diagnose disease. Classifying objects, including the apple described above, is prediction: The missing data is which items are similar to each other.

The use of data for prediction is not new. The mathematical ideas behind machine learning are decades old. Many of the algorithms are even older. So what changed?

Recent advances in computational speed, data storage, data retrieval, sensors, and algorithms have combined to dramatically reduce the cost of machine-learning-based predictions. An approach called “deep learning” has been particularly important to the changes of the past five years. Figure 1 shows improvements in image classification, pedestrian detection, and object detection. The IMAGENET competition has gone from 72% successful image classification in 2010 to 96% success in 2015. Thus over five years, the fraction of images mistakenly classified declined from 28% to 4%. The biggest improvement was between 2011 and 2012, when deep-learning algorithms were applied for the first time.

Similarly, the second graph shows that between 2013 and 2016 there were substantial improvements in pedestrian detection in video taken from moving cars. The third graph shows that the rate of improvement continues to accelerate. Between July and December 2015, object detection in a set of difficult-to-recognize images (the KITTI vision benchmark) improved from 39% to 87% success.

These improvements mean that prediction through machine learning is becoming a cost-effective means of conducting a variety of tasks. For example, until recently, classifying images required a human to perform the classification, a time-consuming (and not easily scalable) process. Advances in machine learning mean that millions of

Figure 1: Recent performance on three vision benchmarks
(Source: NVIDIA slides, page 19)
images can now be recognized with much less time and money required. Image classification is thus an automated task. Similarly, fraud detection in banking is moving from a costly human-based process to a much less expensive and more easily scaled machine-based process. In this sense, recent advances in machine learning have led to a dramatic decrease in the cost of prediction.

**Box 1: Human Intelligence and Prediction**

In his book *On Intelligence*, Jeff Hawkins argues that prediction is the basis for human intelligence and the primary function of the cortex. Hawkins was among the first to put such a strong emphasis on prediction as the main function of the brain and as the primary feature of intelligence. The essence of his theory is that human intelligence, which is at the core of creativity and productivity gains, is due to the way our brains use memories to make predictions. “We are making continuous low-level predictions in parallel across all our senses. But that's not all. I am arguing a much stronger proposition. Prediction is not just one of the things your brain does. It is the primary function of the neocortex, and the foundation of intelligence. The cortex is an organ of prediction.” (p.89) Hawkins’ argues that our brains are constantly making predictions regarding what we are about to experience - what we will see, feel, and hear. As we develop and mature as humans, our brains’ predictions are increasingly accurate; the predictions often come true. However, when predictions do not accurately predict the future, we notice the anomaly, and this information is fed back into our brain, which updates its algorithm, thus learning and further enhancing the model.

Hawkins’ work is controversial. His ideas are debated in the psychology literature and many computer scientists flatly reject his emphasis on the cortex as a model for prediction machines. Irrespective of whether the underlying model is appropriate, his emphasis on prediction as the basis for intelligence is useful for understanding the impact of recent changes in AI. In particular, the falling cost of prediction due to advances in machine learning has meant that many tasks previously considered unique to humans can now be done by machines, including driving, language translation, playing Go, and image classification.

To the extent that our brains really are “organs of prediction,” then advances in AI may indeed lead to increased substitution of humans for machines. However, to the extent that Jeff Hawkins’ model of humans is incorrect - that instinct, or some other process besides prediction, drives human behavior - then humans have an advantage in efficiently undertaking some tasks because they seem to have better algorithms for converting sensory input (data) into actions. It requires much more data for machines to overcome their disadvantage in algorithms. Machine learning remains far less efficient at converting data into predicted actions than the human brain. A teenager learns to drive with tens or hundreds of miles of practice. Automated driving systems are given millions of miles of data and still periodically require human intervention.
Prediction is a key component in any task

While it is straightforward to claim that the latest AI developments involve a dramatic improvement in the ability of machines to predict, this only matters in a context where prediction is important. Given that our context is work, what role does prediction play in the performance of tasks? By answering this question, we can anticipate the avenues by which a reduction in the cost of prediction will change human and machine roles in other areas, outside of prediction.

The following diagram presents the anatomy of a task. A task may be anything from driving a car between A and B to setting prices for a multi-product retailer. The locus of a task is a component called an action that, when taken, generates an outcome. However, actions are not taken in a vacuum. Importantly, the way an action is translated into an outcome is shaped by underlying conditions and the resolution of uncertainty. For example, to drive a car between two points involves a myriad of decisions that impact how quickly, say, that task is achieved. Most notably, it involves adjusting for traffic conditions. Thus, a driver who is performing the action, observes the immediate environment, for example, the behaviour of cars ahead of it on the road.

Those observations are the data the driver uses to forecast (predict) where those cars might be as the car moves forward. On the basis of the forecast, the driver then takes different actions to minimize the risk of accidents or to avoid bottlenecks. In so

![Figure 2: The Anatomy of a Task](image-url)
doing they apply judgment in combination with the prediction. While an experienced driver may not learn much from their own behavior, an inexperienced one will see how their action led to an outcome and then, through a process of feedback, use that information to inform their future predictions.

Seen in this light, it is useful to distinguish between the value and the cost of prediction. As we have mentioned, AI advances have lowered the cost of prediction. Given data and something to predict, it is now much less costly to derive an accurate prediction. But of equal importance is what has happened to the value of prediction. Put simply, prediction has more value in a task if data is more widely available and accessible. The many decades-long improvements in the ability of computers to copy, transmit, and store information have meant that the data available for such predictions has also improved, leading to a higher value of prediction in a wider variety of tasks.

Consider autonomous driving. Engineers have been able to make cars that could drive themselves in specific environments for decades. The modern era of autonomous driving began in the 1980s. The US and Germany were the two nations at the forefront in this line of research. In the US, the research was largely funded by DARPA (Defense Advanced Research Projects Agency). In Germany, large automotive companies such as Mercedes-Benz funded research. The leading projects utilized computer vision-based systems, lidar, and autonomous robotic control. The decision-making systems were essentially driven by optimizing if-then-else algorithms (e.g., optimize speed subject to not exceeding the speed limit and not hitting anything; “if is raining, then slow down…”, “if a pedestrian approaches within 5 feet on the left, then swerve right”). In other words, the systems were algorithmic, that is, codifying an algorithm describing the connection between road conditions and decisions related to speed and steering.

However, to be able to drive on roads in unstructured and unpredictable environments, including ones where other drivers (human or autonomous) also exist, requires predicting the outcomes of a large number of possible actions. Codifying the set of possible outcomes proved too challenging. In the early 2000s, however, several groups began using primitive versions of modern machine-learning techniques. The key difference between the new method and the old method is that while the old method centered around optimizing a long list of if-then-else statements, the new method instead predicts what a human driver would do given the set of inputs (e.g., camera images, lidar information, mapping data, etc). This facilitated significant improvements in autonomous driving performance.
Who Judges?

The ultimate outcome of any task comes from the action that is taken. As already emphasized, one input to what action gets taken is prediction. However, as depicted in Figure 2, what we call “judgment” has a distinct role in determining an action. Judgment is the ability to make considered decisions. In other words, to determine the impact different actions have on outcomes given predictions. As it turns out, this hinges directly on how clear outcomes themselves are. For some tasks, the precise outcome that is desired can be easily described. For instance, in terms of labeling images, you want the label to be accurate, which it either is or is not. In these situations, the separate need for judgment, as might be applied by a human, is limited and the task can be largely automated.

In other situations, describing the precise outcome is difficult. It resides in the mind of humans and cannot be translated into a form a machine can understand. This is why it is often argued that AIs may be less adept at handling emotional tasks. However, machine prediction can significantly impact more pedestrian tasks. For instance, an AI that maps out the optimal route to take to minimize travel time (such as might occur in apps like Waze) cannot easily take into account the preferences of drivers who have to do the work of driving. For instance, the AI may find a route with many turns that reduces travel by a few seconds but a human driver may prefer a slower route with fewer turns or one that takes them closer the dry cleaning store where they remembered they had clothing to pick up. Of course, a clever engineer will resolve these particular issues. But the point here is that the outcome that is desired often has indescribable elements that require the exercise of judgment to mitigate and absorb. Judgment takes predictions and uses them as information that is useful to determine actions. Hence, alongside prediction, judgment is a critical input into many actions. Whether a machine can undertake the action depends on whether the outcome can be described in such a way that the machine can exercise suitable judgment.

This is not to say that our understanding of human judgment cannot evolve and be automated. One of the features of new modes of machine learning is that they can examine the relationship between actions and outcomes and use this feedback to further refine predictions. In other words, prediction machines can learn. This dynamic aspect drives improvements in how a task is performed. For instance, machines may learn to predict better by observing how humans perform in tasks. This is what DeepMind’s AI AlphaGo did when learning the game of Go. It analyzed thousands of human-to-human games and then played itself millions more games, each time
receiving feedback on action/outcomes that allowed it to predict more accurately in order to inform on strategies in the game.

In other situations, the feedback can include data on human judgment and actions. For instance, the startup X.ai launched a service whereby it provides a virtual assistant to interact with people you know in order to schedule appointments. Its goal is to replace the human assistant. But such interactions can be difficult. The AI needs to understand your preferences and also be able to communicate with others and not seem overbearing. Thus, what the X.ai team have been doing is handling the tasks themselves and having their own AIs observe the interactions in order to learn from them. In effect, the AI is trained to predict the human responses and, indeed, what choices a human makes in judgment. While this does not lead to a formal description of outcomes, the idea is to allow the AI to mimic that judgment. In this way, over time, feedback can transform some aspects of judgment into prediction problems: predicting human judgment.

By breaking a task down into its constituent components, we can see different ways in which AI will affect the workplace. While much discussion frames the issue in terms of machines versus humans, this is not the conclusion we draw. Instead, we emphasize the nature of the judgment required to undertake an action. When judgment is easily codified, then computers will indeed begin to replace humans in the workforce. These are situations where the bottleneck to automation has been prediction, for example, object classification and autonomous driving. However, there are many tasks for which judgment is not easily codified. As the cost of prediction falls, the number of such human judgment tasks is likely to grow.

**Employing prediction machines**

Major advances in prediction may facilitate the automation of entire tasks. In other words, while existing technology may allow machines to take action, the full automation of a task requires the ability for the machines to predict and rely on those predictions to determine what to do. Thus, being able to move predictive tasks to machines may facilitate machine led-tasks, that is, automation. For example, for many business-related language translation tasks, as prediction-driven translation improves, the role for human judgment becomes limited, though judgment might still serve a role in critical negotiations (see Box 2 for how prediction may lead to the increased automation of tasks in fulfillment centers).
Box 2: Prediction in fulfillment centers

To see how prediction machines may lead to automation of tasks we do not normally associate with prediction, consider fulfillment. Fulfillment is a central step in retail generally and in electronic commerce in particular. This is the process of taking an order and executing it by making it ready for delivery to its intended customer. In electronic commerce, fulfillment includes a number of steps such as locating items associated with an order in a large warehouse-type facility, picking the items off shelves, scanning them for inventory management, placing them in a tote, packing them in a box, labeling the box, and shipping the box for delivery. The fulfillment industry has grown rapidly over the past two decades due to the rapid growth in online shopping.

Many early applications of machine learning to fulfillment related to inventory management: predicting which products would sell out and which did not need to be reordered because demand was low. These were well-established prediction tasks that have been a key part of offline retail and warehouse management for decades. Machine-learning technologies made these predictions better.

Over the past two decades, much of the rest of the fulfillment process has been automated. For example, research determined that fulfillment centre workers were spending over half their time walking around the warehouse to find items that had been ordered in order to pick them off the shelf and put them in their tote. As a result, several companies developed an automated process for bringing shelves to workers in order to reduce the time spent walking. Amazon acquired the leading company in this market, Kiva, in 2012 for $775m and eventually stopped servicing other Kiva customers. Other providers subsequently emerged to fill the demand for the growing market of in-house fulfillment centers and “3PLs” (third-party logistics firms).

Despite significant automation, fulfillment centres still employ many humans. Perhaps surprisingly, the reason is because of the difficulty of grasping. Although grasping objects is easy for humans – infants develop the skill during the latter half of their first year (6-12 months) – this task has so far eluded automation. The core challenge is not in creating dextrous fingers, but in identifying the right angle to use in grasping a particular object. As a result, Amazon alone employs 40,000 human pickers full time and tens of thousands more part time during the busy holiday season. Human pickers handle approximately 120 picks per hour. It is not that companies that do high-volume fulfillment would not like to automate picking. In fact, for the past two years, Amazon incentivized the best robotics teams in the world to work on this long-studied problem of grasping by hosting the Amazon Picking Challenge, focused on automated picking in unstructured warehouse environments. Even though top teams from institutions such as MIT worked on this problem, many using advanced industrial-grade equipment like Baxters, Yaskawa Motomans, Universal Arms, ABBs, PR2s, and Barrett Arms, as of this writing the problem has not yet been solved in a manner that is satisfactory for industrial use.

It may seem hard to believe that robots are perfectly capable of assembling a car or flying a plane but are not able to pick items off a shelf and place them in a box. However, perhaps most surprising is to learn that prediction is at the root of the physical task of grasping. Robots can assemble an automobile because the components are highly standardized and the process highly routinized. However, there is an almost infinite variety of shapes, sizes, weights, and firmness of items in an Amazon warehouse. Therefore, it is impossible to program robots to pick and place each and every type of item in whatever orientation they happen to be positioned in on the shelf. Instead, they must be able to “see” the object (analyze the image) and predict what approach to grasping the object will work (arm approach, finger positioning, grip pressure, etc.) so as to hold it and not drop or crush it.
In other contexts, however, more abundant prediction could lead to increased value of human-led tasks, that is, where the machine provides predictions but the human exercises judgment and undertakes the action component of a task. For example, Google Inbox’s autoreply processes your incoming email and predicts several short responses. It then lets the human judge which is the most appropriate. Selecting between one of a handful of actions is much quicker than composing a reply, enabling the human to respond to more emails in less time.

Autonomous driving is another example. While autonomous driving is an area where new developments in AI are seen as taking humans out of the picture, many of the most imminent advances still keep humans in the driver’s seat. For instance, many car companies are using AI to predict when a collision might occur. Most of the time, all is clear. However, when an issue arises, the machine alerts the human who then exercises judgment as to the course of action they will take to avoid the collision. While many have argued that machine reactions may be superior in such situations, it is precisely because the full dimensions of the outcome cannot be easily described that humans remain in control. In particular, human control arises when judgment requires decisions not of whether to avoid harm altogether but what type of harm to avoid and when situations are sufficiently unusual that judgment is required to correct for unreliable prediction. Combined with human judgment, prediction has made human-led driving safer and so increased the value of human-led driving overall.

Another example is medical diagnosis. Artificial intelligence is improving diagnosis. While this may reduce the human role in diagnosis, treatment and patient care require judgment. Each patient may have different needs that are difficult to define for the machine. More effective diagnosis is likely to lead to increased use of effective treatment and better patient care.

Overall, human judgment is particularly important when it is difficult to describe outcomes in a manner that can be understood by machines. Part of this inability comes from a lack of fundamental and precise knowledge of what those outcomes involve. But in other situations, it may never be possible to describe the relevant trade-offs to machines in a way that is acceptable to humans.

The Managerial Challenge

As artificial intelligence technology improves, prediction done by machines will increasingly dominate prediction done by humans. In such a world, what is the role of humans in the organization? What are the most salient features of a role that
emphasizes human judgment and de-emphasizes human prediction? Managing for such a future requires understanding three interrelated insights:

1. **Prediction is not the same as automation.** Prediction is an input into automation, but successful automation requires a variety of other tasks. The anatomy of a task involves data, prediction, judgment, and action in order to attain an outcome. Machine learning is only one component of this: prediction. However, automation requires that data collection, judgment, and action also be done by machines. For example, autonomous driving involves a set of tasks such as vision (data), prediction (given sensory inputs, what action would a human take?), assessment of consequences (judgment), and acceleration, braking, and steering (action). Similarly, medical care tasks include imaging (data), diagnostics (prediction), treatment choice (judgment), bedside manner (judgement/action) and physical intervention (action), among others. Prediction is the stage of automation in which the technology is currently improving particularly rapidly, although advances in sensor technology (data) and robotics (action) are also advancing quickly.

2. **The most valuable workforce skills involve judgment.** For many activities, prediction was the bottleneck to automation. This meant a role for human workers in aiding a variety of prediction tasks, including pick-and-place and driving. Going forward, this role for human workers will diminish. Instead, employers will look for workers who augment the value of prediction. In the language of economics, just as the returns to numeracy and literacy increased with the diffusion of computing and the demand for golf balls rises with lower prices of golf clubs, the most valuable skills in the future will be those that are complementary to prediction: those related to judgment. We can only speculate on the aspects of judgment that will be most valuable: ethical judgement, emotional intelligence, artistic ability, task definition, etc. This diverse group of skills suggests that the set of activities in which the human role may increase in value is potentially wide-ranging. Nursing skills related to physical intervention and emotional comfort may be more valuable if prediction leads to better diagnosis of disease. Retail store greeters may become increasingly effective if social interactions help differentiate stores in the presence of reliable predictions on purchase behavior. Private security guard skills related to ethical judgement and use of force may be of increased value as a complement to better prediction of crime. For many activities, human judgment becomes more valuable when paired with effective prediction. It is these skills that will be most valuable as artificial intelligence technology
improves and diffuses. Still, as noted above, the parts of tasks that need human judgment can change over time. In particular, when you have enough observations of judgment, it becomes a prediction problem. The AI can use the data to predict the human judgment. Thus the judgement aspect of a task is a moving target, requiring humans to adapt to new situations in which judgment is required.

3. **Managing may require a new set of talents and expertise.** Managerial tasks may also change. First, many managerial tasks are predictive. For example, many managers emphasize their skills in identifying which applicants to hire and which workers to promote. These are prediction skills: predicting which workers will succeed. As machines improve their performance in prediction, human prediction skills will become less valuable compared to judgement skills, such as mentoring, emotional support, and ethics. Second, managing a workforce whose skills are complementary to prediction may be different from managing a workforce for which a core skill is prediction. For example, promotion cannot be based on the (often well-measured) success of past predictions. Instead, alternative metrics must be developed. Third, and perhaps most importantly, managing a firm with artificial intelligence capabilities will involve managing the artificial intelligence itself. What are the opportunities for prediction? What should be predicted? How should the artificial intelligence agent learn to improve those predictions over time? Managing in this context requires judgment in identifying and applying the most useful predictions and weighing the relative costs of different types of errors.

The usefulness of the technology is limited by the usefulness of the object to be predicted. Sometimes such predictions are clear in the sense that there is a well-acknowledged objective, such as object recognition to identify people from their faces. Other times, however, the objective is less clear and requires judgment to determine. For example, should the team managing Facebook’s Newsfeed, maximize clicks, likes, content sharing, time on Facebook, or profitability to advertisers? In such cases, managers’ judgment becomes a particularly valuable complement to prediction technology.

At the dawn of the twenty-first century, the set of recognized prediction problems were classic statistical questions: inventory management, demand forecasting, etc. However, over the last ten years, researchers learned that image recognition, driving, and translation may also be framed as prediction problems. As the range of tasks that are recast as prediction problems continues to grow, we believe the scope of new applications will be extraordinary. The key managerial challenges will be: 1) shifting the
training of workers from prediction-related to judgement-related skills, 2) assessing the rate and direction of the adoption of AI technologies in order to properly time the shifting of workforce training (not too early, not too late), and 3) developing management processes that build the most effective teams of judgment-focused humans and prediction-focused artificial intelligence agents.