MACHINE LEARNING

MAKING AI MORE HUMAN

Illustration by Simon Prades

Artificial intelligence has staged a revival by starting to incorporate what we know about how children learn

By Alison Gopnik

Alison Gopnik is a professor of psychology and an affiliate professor of philosophy at the University of California, Berkeley. Her research focuses on how young children learn about the world around them.

IF YOU SPEND MUCH TIME WITH CHILDREN, YOU'RE BOUND TO WONDER HOW young human beings can possibly learn so much so quickly. Philosophers, going all the way back to Plato, have wondered, too, but they've never found a satisfying answer. My five-year-old grandson, Augie, has learned about plants, animals and clocks, not to mention dinosaurs and spaceships. He also can figure out what other people want and how they think and feel. He can use that knowledge to classify what he sees and hears and make new predictions. He recently proclaimed, for example, that the newly discovered species of titanosaur on display at the American Museum of Natural History in New York City is a plant eater, so that means it really isn't that scary.

Yet all that reaches Augie from his environment is a stream of photons hitting his retina and disturbances of air contacting his eardrums. The neural computer that sits behind his blue eyes manages somehow to start with that limited information from his senses and to end up making predictions about plant-eating titanosaurs. One lingering question is whether electronic computers can do the same.

During the past 15 years or so computer scientists and psychologists have been trying to find an answer. Children acquire a great deal of knowledge with little input from teachers or parents. Despite enormous strides in machine intelligence, even the most powerful computers still cannot learn as well as a five-year-old does.

Figuring out how the child brain actually functions—and then creating a digital version that will work as effectively—will challenge computer scientists for decades to come. But in the meantime, they are beginning to develop artificial intelligence that incorporates some of what we know about how humans learn.

THIS WAY UP

AFTER THE FIRST BURST of enthusiasm in the 1950s and 1960s, the quest for AI languished for decades. In the past few years, though, there have been striking advances, especially in the field of machine learning, and AI has become one of the hottest developments in technology. Many utopian or apocalyptic predictions have emerged about what those advances mean. They have, quite literally, been taken to presage either immortality or the end of the world, and a lot has been written about both these possibilities.

I suspect that developments in AI lead to such strong feelings because of our deep-seated fear of the almost human. The idea that creatures might bridge the gap between the human and the artificial has always been deeply disturbing, from the medieval golem to Frankenstein's monster to Ava, the sexy robot fatale in the movie *Ex Machina*.

But do computers really learn as well as humans? How much of the heated rhetoric points to revolutionary change, and how much is just hype? The details of how computers learn to recognize, say, a cat, a spoken word or a Japanese character can be hard to follow. But on closer inspection, the basic ideas behind machine learning are not as baffling as they first seem.

One approach tries to solve the problem by starting with the stream of photons and air vibrations that Augie, and all of us, receives—and that reaches the computer as pixels of a digital image and sound samples of an audio recording. It then tries to extract a series of patterns in the digital data that can detect and identify whole objects in the surrounding world. This so-called bottom-up approach has roots in the ideas of philosophers such as David Hume and John Stuart Mill and psychologists such as Ivan Pavlov and B. F. Skinner, among others.

In the 1980s scientists figured out a compelling and inge-

IN BRIEF

How do young children know what they know? That question has long preoccupied philosophers and psychologists—and now computer scientists.

Specialists in artificial intelligence are studying the mental reasoning powers of preschoolers to develop ways to teach machines about the world.

Two rival machine-learning strategies—both halting attempts to mimic what children do naturally—have begun to transform AI as a discipline.

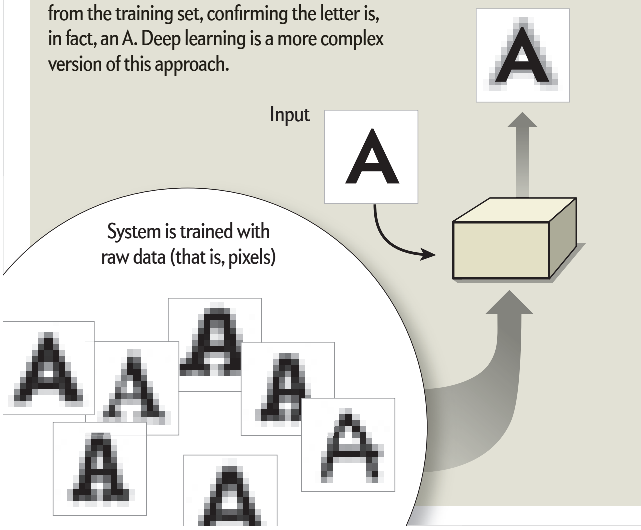
OPPOSING STRATEGIES

Two Paths to AI's Resurgence

Problems the average five-year-old solves readily can stump even the most powerful computers. AI has made a spirited comeback in recent years by teaching computers to learn about the world somewhat like a child does. The machine recognizes the letter "A" either from raw sensory information—a bottom-up approach—or by making a guess from preexisting knowledge—a top-down approach.

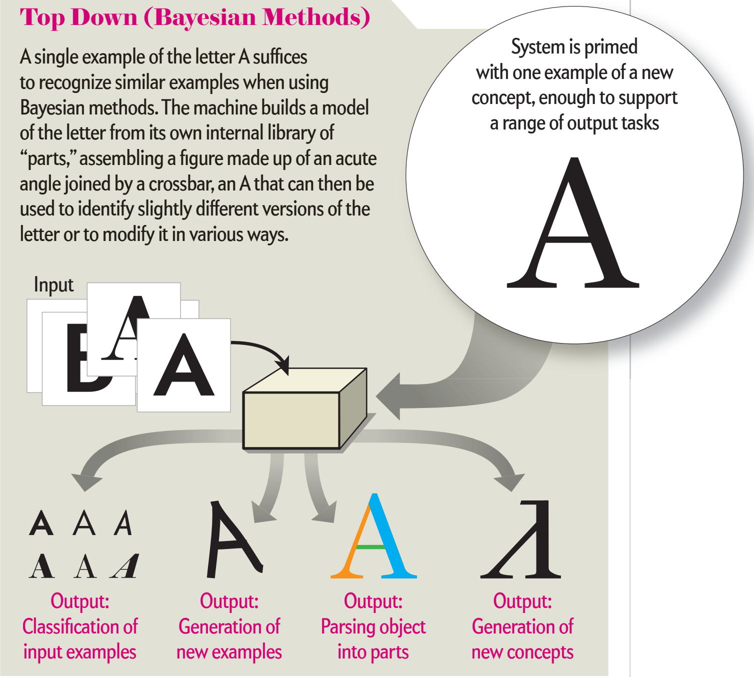
Bottom Up (Deep Learning)

Examples of the letter A teach a computer to distinguish patterns of light and dark pixels for various versions of the letter. Then, when the machine receives a new input, it assesses whether the pixels match the configuration from the training set, confirming the letter is, in fact, an A. Deep learning is a more complex version of this approach.


Output: Pixel by pixel, this character resembles the training raw data set; therefore, it is an A
Input
System is trained with raw data (that is, pixels)

Top Down (Bayesian Methods)

A single example of the letter A suffices to recognize similar examples when using Bayesian methods. The machine builds a model of the letter from its own internal library of "parts," assembling a figure made up of an acute angle joined by a crossbar, an A that can then be used to identify slightly different versions of the letter or to modify it in various ways.


System is primed with one example of a new concept, enough to support a range of output tasks
Input
Output: Classification of input examples
Output: Generation of new examples
Output: Parsing object into parts
Output: Generation of new concepts

SOURCE: "HUMAN-LEVEL CONCEPT LEARNING THROUGH PROBABILISTIC PROGRAM INDUCTION," BY BRENDEN M. LAKE, RUSLAN SALAKHUTDINOV AND JOSHUA B. TENENBAUM, IN *SCIENCE*, VOL. 350; DECEMBER 11, 2015 (*Bayesian example*)

nious way to apply bottom-up methods to let computers hunt for meaningful patterns in data. "Connectionist," or "neural network," systems take inspiration from the way that neurons convert light patterns at your retina into representations of the world around you. A neural network does something similar. It uses interconnected processing elements, akin to biological cells, to transform pixels at one layer of the network into increasingly abstract representations—a nose or an entire face—as data are crunched at progressively higher layers.

Neural-network ideas have gone through a recent revival because of new techniques called deep learning—technology now being commercialized by Google, Facebook and other tech giants. The ever increasing power of computers—the exponential increase in computing capability that is captured by what is known as Moore's law—also has a part in the new success of these systems. So does the development of enormously large data sets. With better processing capabilities and more data to crunch, connectionist systems can learn far more effectively than we might have once thought.

Over the years the AI community has seesawed between favoring these kinds of bottom-up solutions to machine learning and alternative top-down approaches. Top-down approaches leverage what a system already knows to help it learn something new. Plato, as well as so-called rationalist philosophers such as René Descartes, believed in a top-down approach to learning—and it played a big role in early AI. In the 2000s such methods also experienced their own rebirth in the form of probabilistic, or Bayesian, modeling.

Like scientists, top-down systems start out by formulating abstract and wide-ranging hypotheses about the world. The systems then make predictions about what the data should look like if those hypotheses are correct. Also like scientists, the systems then revise their hypotheses, depending on the outcome of those predictions.

NIGERIA, VIAGRA AND SPAM

BOTTOM-UP METHODS are perhaps the most readily understood, so let's consider them first. Imagine that you are trying to get your computer to separate important messages from the spam that arrives in your in-box. You might notice that spam tends to have certain distinguishing characteristics: a long list of recipient addressees, an originating address in Nigeria or Bulgaria, references to $1-million prizes or perhaps mention of Viagra. But perfectly useful messages might look the same. You don't want to miss the announcement that you have earned a promotion or an academic award.

If you compare enough examples of spam against other types of e-mails, you might notice that only the spam tends to have qualities that combine in certain telltale ways—Nigeria, for instance, plus a promise of a $1-million prize together spell

Illustration by Jen Christiansen

trouble. In fact, there might be some quite subtle higher-level patterns that discriminate between the spam messages and the useful ones—misspellings and IP addresses that are not at all obvious, for example. If you could detect them, you could accurately filter out the spam—without fear of missing a notice that your Viagra has shipped.

Bottom-up machine learning can ferret out the relevant clues to solve this kind of task. To do this, a neural network must go through its own learning process. It evaluates millions of examples from huge databases, each labeled as spam or as an authentic e-mail. The computer then extracts a set of identifying features that separate spam from everything else.

In a similar way, the network might inspect Internet images labeled "cat," house," "stegosaurus," and so on. By extracting the common features in each set of images—the pattern that distinguishes all the cats from all the dogs—it can identify new images of a cat, even if it has never seen those particular images before.

One bottom-up method, called unsupervised learning, is still in its relative infancy, but it can detect patterns in data that have no labels at all. It simply looks for clusters of features that identify an object—noses and eyes, for example, always go together to form a face and differ from the trees and mountains in the background. Identifying an object in these advanced deep-learning networks takes place through a division of labor in which recognition tasks are apportioned among different layers of the network.

An article in *Nature* in 2015 demonstrated just how far bottom-up methods have come. Researchers at DeepMind, a company owned by Google, used a combination of two different bottom-up techniques—deep learning and reinforcement learning—in a way that enabled a computer to master Atari 2600 video games. The computer began knowing nothing about how the games worked. At first, it made random guesses about the best moves while receiving constant feedback about its performance. Deep learning helped the system identify the features on the screen, and reinforcement learning rewarded it for a high score. The computer achieved a high proficiency level with several games; in some cases, it performed better than expert human players. That said, it also completely bombed on other games that are just as easy for humans to master.

The ability to apply AI to learn from large data sets—millions of Instagram images, e-mail messages or voice recordings—allows solutions to problems that once seemed daunting, such as image and speech recognition. Even so, it is worth remembering that my grandson has no trouble at all recognizing an animal or responding to a spoken query even with much more limited data and training. Problems that are easy for a human five-year-old are still extremely perplexing to computers and much harder than learning to play chess.

Computers that learn to recognize a whiskered, furry face often need millions of examples to categorize objects that we can classify with just a few. After extensive training, the computer might be able to identify an image of a cat that it has never seen before. But it does so in ways that are quite different from human generalizations. Because the computer software reasons differently, slipups occur. Some cat images will not be labeled as cats. And the computer may incorrectly say an image is a cat, although it is actually just a random blur, one that would never fool a human observer.

ALL THE WAY DOWN

THE OTHER APPROACH to machine learning that has transformed AI in recent years works in the opposite direction, from the top down. It assumes that we can get abstract knowledge from concrete data because we already know a lot and especially because the brain is already capable of understanding basic abstract concepts. Like scientists, we can use those concepts to formulate hypotheses about the world and make predictions about what data (events) should look like if those hypotheses are right—the reverse of trying to extract patterns from the raw data themselves, as in bottom-up AI.

APPLYING AI TO LEARN FROM LARGE DATA SETS—MILLIONS OF INSTAGRAM IMAGES OR E-MAIL MESSAGES—ALLOWS SOLUTIONS TO PROBLEMS THAT ONCE SEEMED DAUNTING.

This idea can best be illustrated by revisiting the spam plague through considering a real case in which I was involved. I received an e-mail from the editor of a journal with a strange name, referring specifically to one of my papers and proposing that I write an article for the publication. No Nigeria, no Viagra, no million dollars—the e-mail had none of the common indications of a spam message. But by using what I already knew and thinking in an abstract way about the process that produces spam, I could figure out that this e-mail was suspicious.

To start, I knew that spammers try to extract money from people by appealing to human greed—and academics can be as greedy to publish as ordinary folks are for $1-million prizes or better sexual performance. I also knew that legitimate "open access" journals have started covering their costs by charging authors instead of subscribers. Also, my work has nothing to do with the journal title. Putting all that together, I produced a plausible hypothesis that the e-mail was trying to sucker academics into paying to "publish" an article in a fake journal. I could draw this conclusion from just one example, and I could go on to test my hypothesis further by checking the editor's bona fides through a search-engine query.

A computer scientist would call my reasoning process a "generative model," one that is able to represent abstract concepts, such as greed and deception. This same model can also describe the process that is used to come up with a hypothesis—the reasoning that led to the conclusion that the message might be an e-mail scam. The model lets me explain how this form of spam works, but it also lets me imagine other kinds of spam or even a type that differs from any I have seen or heard about before. When I receive the e-mail from the journal, the model lets me work backward—tracing step by step why it must be spam.

SCIENTIFIC AMERICAN ONLINE Read more about how researchers use Bayesian methods at ScientificAmerican.com/jun2017/gopnik

Generative models were essential in the first wave of AI and cognitive science in the 1950s and 1960s. But they also had limitations. First, most patterns of evidence might, in principle, be explained by many different hypotheses. In my case, it could be that the e-mail really was legitimate, even though it seemed unlikely. Thus, generative models have to incorporate ideas about probability, one of the most important recent developments for these methods. Second, it is often unclear where the basic concepts that make up generative models come from. Thinkers such as Descartes and Noam Chomsky suggested that you are born with them firmly in place, but do you really come into this world knowing how greed and deception lead to cons?

Bayesian models—a prime example of a recent top-down method—attempt to deal with both issues. Named after 18th-century statistician and philosopher Thomas Bayes, they combine generative models with probability theory using a technique called Bayesian inference. A probabilistic generative model can tell you how likely it is that you will see a specific pattern of data if a particular hypothesis is true. If the e-mail is a scam, it probably appeals to the greed of the reader. But of course, a message could appeal to greed without being spam. A Bayesian model combines the knowledge you already have about potential hypotheses with the data you see to let you calculate, quite precisely, just how likely it is that an e-mail is legitimate or spam.

This top-down method fits better than its bottom-up counterpart with what we know about how children learn. That is why, for the past 15 years, my colleagues and I have used Bayesian models in our work on child development. Our lab and others have used these techniques to understand how children learn about cause-and-effect relationships, predicting how and when youngsters will develop new beliefs about the world and when they will change the beliefs they already have.

Bayesian methods are also an excellent way to teach machines to learn like people. In 2015 Joshua B. Tenenbaum of the Massachusetts Institute of Technology, with whom I sometimes collaborate, Brenden M. Lake of New York University and their colleagues published a study in *Science*. They designed an AI system that could recognize unfamiliar handwritten characters, a job that is simple for people but extremely taxing for computers.

Think of your own recognition skills. Even if you have never seen a character in a Japanese scroll, you can probably tell if it is the same or different from one on another scroll. You can probably draw it and even design a fake Japanese character—and understand as well that it looks quite different from a Korean or Russian character. That is just what Tenenbaum's team members got their software to do.

With a bottom-up method, the computer would be presented with thousands of examples and would use the patterns found in those examples to identify new characters. Instead the Bayesian program gave the machine a general model of how to draw a character: for example, a stroke can go right or left. And after the software finishes one character, it goes on to the next.

When the program saw a given character, it could infer the sequence of strokes that were needed to draw it, and it went on to produce a similar set of strokes on its own. It did so the same way that I inferred the series of steps that led to my dubious spam e-mail from the journal. Instead of weighing whether a marketing scam was likely to lead to that e-mail, Tenenbaum's model guessed whether a particular stroke sequence was likely to produce the desired character. This top-down program worked much better than deep learning applied to exactly the same data, and it closely mirrored the performance of human beings.

A PERFECT MARRIAGE

THESE TWO LEADING APPROACHES to machine learning—bottom up and top down—have complementary strengths and weaknesses. With a bottom-up method, the computer does not need to understand anything about cats to begin with, but it does need a great deal of data.

The Bayesian system can learn from just a few examples, and it can generalize more widely. This top-down approach, though, requires a lot of work up front to articulate the right set of hypotheses. And designers of both types of systems can run into similar hurdles. The two approaches work only on relatively narrow and well-defined problems, such as recognizing written characters or cats or playing Atari games.

Children do not labor under the same constraints. Developmental psychologists have found that young children somehow combine the best qualities of each approach—and then take them much further. Augie can learn from just one or two examples, the way a top-down system does. But he also somehow extracts new concepts from the data themselves, like a bottom-up system. These concepts were not there to begin with.

Augie can actually do much more. He immediately recognizes cats and tells letters apart, but he can also make creative and surprising new inferences that go far beyond his experience or background knowledge. He recently explained that if an adult wants to become a child again he or she should try not eating any healthy vegetables, because they make a child grow into an adult. We have almost no idea how this kind of creative reasoning emerges.

We should recall the still mysterious powers of the human mind when we hear claims that AI is an existential threat. Artificial intelligence and machine learning sound scary. And in some ways, they are. The military is researching ways to use these systems to control weapons. Natural stupidity can wreak far more havoc than artificial intelligence, and we humans will need to be much smarter than we have been in the past to properly regulate the new technologies. Moore's law is an influential force: even if advances in computing result from quantitative increases in data and computer power, rather than conceptual revolutions in our understanding of the mind, they can still have momentous, practical consequences. That said, we shouldn't think that a new technological golem is about to be unleashed on the world. SA

MORE TO EXPLORE

Bayesian Networks, Bayesian Learning and Cognitive Development. Alison Gopnik et al. in *Developmental Science,* Vol. 10, No. 3, pages 281–287; May 2007.

Human-Level Concept Learning through Probabilistic Program Induction. Brenden M. Lake et al. in *Science,* Vol. 350, pages 1332–1338; December 11, 2015.

The Gardener and the Carpenter: What the New Science of Child Development Tells Us about the Relationship between Parents and Children. Alison Gopnik. Farrar, Straus and Giroux, 2016.

FROM OUR ARCHIVES

Machines Who Learn. Yoshua Bengio; June 2016.

scientificamerican.com/magazine/sa