SCIENCE FOR THE CURIOUS

Discover®

APRIL 2017

ARTIFICIAL INTELLIGENCE
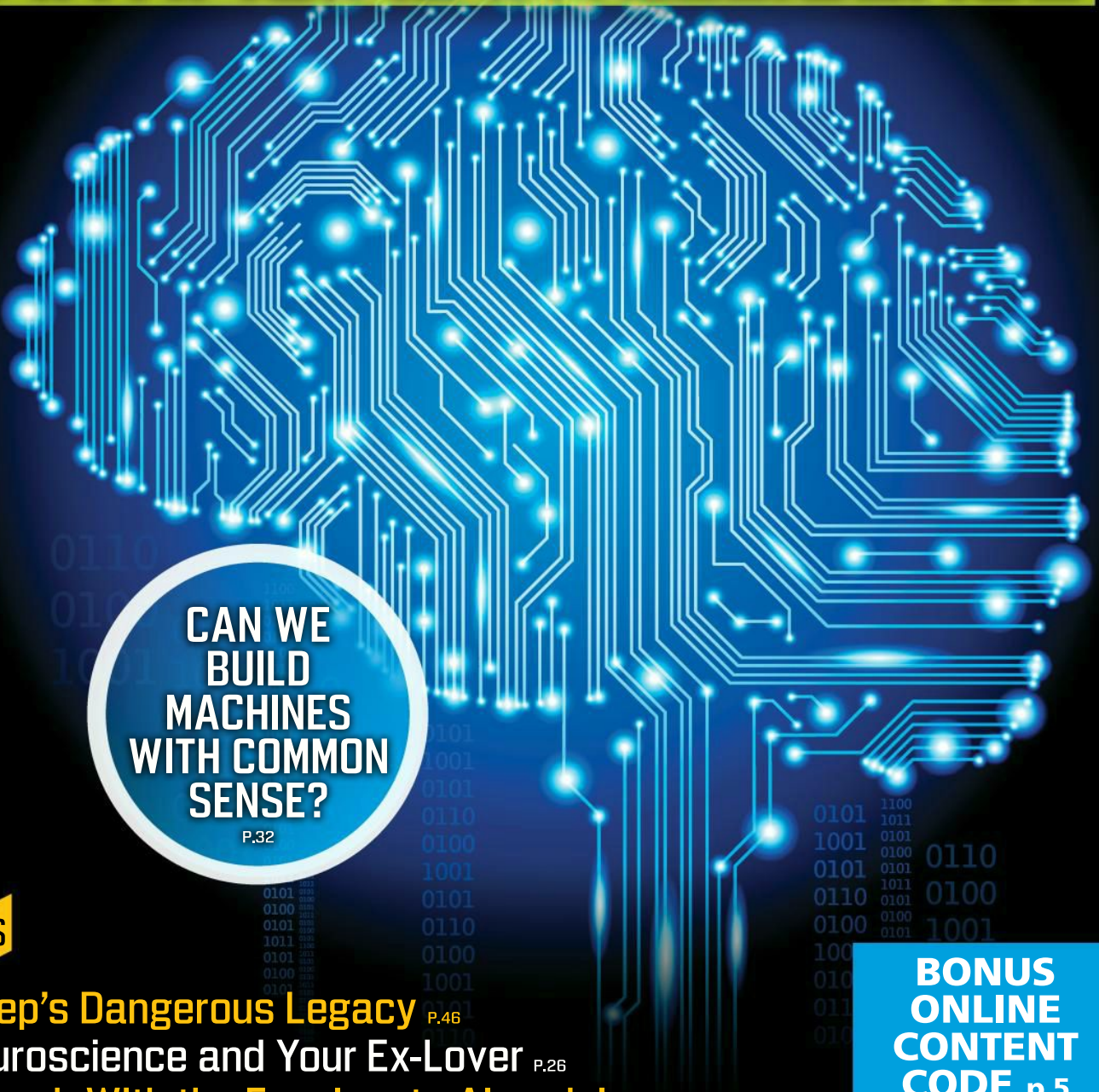
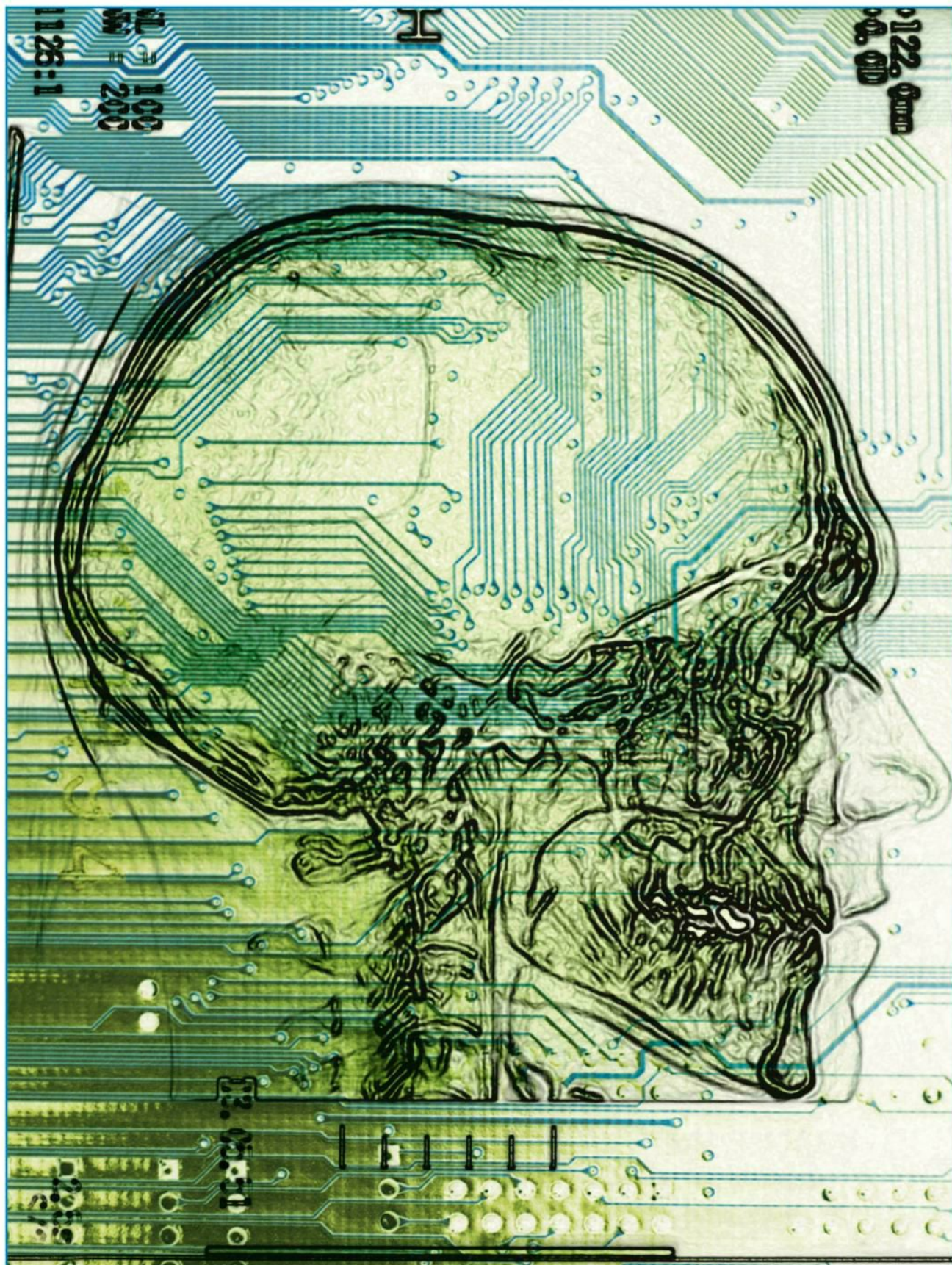## CAN WE BUILD MACHINES WITH COMMON SENSE?
P.32

PLUS

BONUS ONLINE CONTENT CODE p.5

CULTIVATING
COMMON SENSE

A band of Seattle computer scientists
is on a mission to make artificial intelligence
actually INTELLIGENT. BY **CARL ENGELKING**

estled among Seattle's gleaming lights on a gloomy September day, a single nonprofit wants to change the world, one computer at a time. Its researchers hope to transform the way machines perceive the world: to have them not only see it, but understand what they're seeing.

At the Allen Institute for Artificial Intelligence (AI2), researchers are working on just that. AI2, founded in 2014 by Microsoft visionary Paul Allen, is the nation's largest nonprofit AI research institute. Its campus juts into the northern arm of Lake Union, sharing the waterfront with warehouses and crowded marinas. Across the lake, dozens of cranes rise above the Seattle skyline — visual reminders of the city's ongoing tech boom. At AI2, unshackled by profit-obsessed boardrooms, the mandate from its CEO Oren Etzioni is simple: Confront the grandest challenges in artificial intelligence research and serve the common good, profits be damned.

AI2's office atmosphere matches its counterculture ethos. Etzioni's hand-curated wall of quotes is just outside the table tennis room.

attributes, how force is applied and the laws of physics. Computers aren't quite there yet.

If these are the frontiers in AI research, then our much-prophesied computer overlords might be a long time coming: Artificial intelligence overall is still pretty dumb. Even today's "smart" programs are driven by narrow, or weak, AI. Strong AI, also called general AI, doesn't exist.

Narrow AI systems are like savants. They're fantastic at single, well-defined tasks: a Roomba vacuuming the floor, for example, or a digital chess master. But a computer that can recognize images of cats can't play chess. Humans can do both; we possess general intelligence. The AI2 team wants to pull these computer savants away from their lonely tasks and plant seeds of common sense. "We still have a long way to go," Etzioni tells me.

Etzioni's 20-year vision is to build an AI system that would serve as a scientist's apprentice. It would read *and* understand scientific literature, connecting the dots between studies and suggesting hypotheses that could lead to significant breakthroughs. When I ask Etzioni if IBM's Watson is already doing this, I feel I've struck a nerve. "They've made some very strong claims, but I'm waiting to see the data," he says.

AI2's researchers hope to transform the way machines perceive the world: to have them not only see it, but understand what they're seeing.

Equations litter ceiling-to-floor whiteboards and random glass surfaces, like graffiti. Employees are encouraged to launch the company kayak for paddle breaks. Computer scientist Ali Farhadi can savor the Seattle skyline from the windows of his democratically chosen office; researchers vote on the locations of their workspaces. It's where he and I meet to explore the limits of computer vision.

At one point, he sets a dry-erase marker on the edge of his desk and asks, "What will happen if I roll this marker over the edge?"

"It will fall on the floor," I reply, wondering if Farhadi could use one of those kayak breaks.

"Exactly! Clearly it's going to fall. This is so trivial," he says, laughing. "But this is still so difficult for a machine to do." Predicting the effects of forces on objects — something I do instantly — requires first perceiving that object; today's computer vision systems excel here. But estimating an object's *future* location demands understanding scene geometry, an object's

But there's also a darker side to this noble endeavor. If we grow to depend on these emerging technologies, certain skills could become obsolete. I can't help but wonder: If smarter AIs gobble up more human-driven tasks, how can we keep up with them?

IT'S ONLY MATH

Grunge rock grew up in Seattle during the late 1980s and '90s in clubs like the Off Ramp and the Vogue. The dirty guitar licks and angst-filled lyrics were a giant middle finger to mainstream acts of the time — those spandexed, Bedazzled, hair-metal bands selling out arenas. Grunge wasn't a cog in the corporate machine, man.

The so-called "Seattle sound" still resonates in the damp concrete of the Emerald City. I see it in the graffiti coloring the gray city, and I hear it in Etzioni. The 52-year-old Harvard grad smiles more than Kurt Cobain, and he prefers a button-up to a thrift-store flannel plaid. But
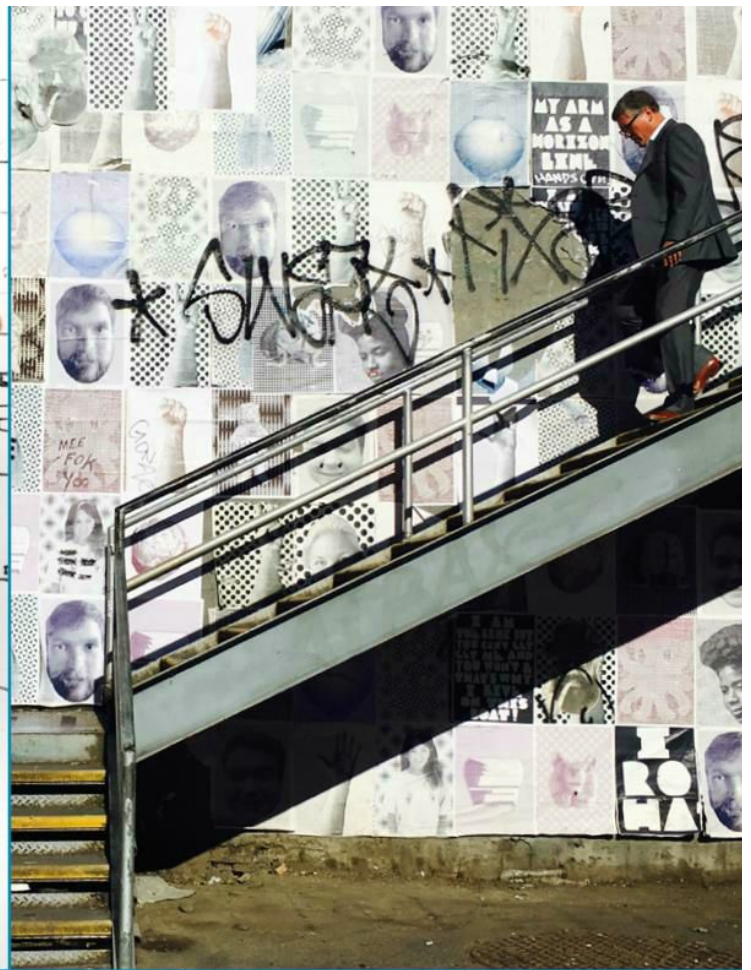
underneath his friendly demeanor, there's an us-versus-the-world edge, a longing to chart his own path. AI2 isn't like Facebook, Google or the other tech behemoths, and Etzioni doesn't want it to be. When we spoke, he used AlphaGo's story as an example.

In March 2016, Google researchers pulled off the year's crowning achievement in the field when their AI, AlphaGo, mastered the ancient Chinese board game Go. Due to the astounding number of board combinations (approximately a 2 followed by 170 zeroes), Go was considered the white whale in computer science. In a highly publicized showdown in South Korea with Lee Sedol, the world's top Go player, AlphaGo came out on top, 4 games to 1.

AlphaGo was soon cited in various click-baity "news" stories as a harbinger of superintelligence and *Terminator*-inspired apocalypse, but Etzioni takes issue with these simplified narratives. "AI isn't magic. It's math," he says with a sigh. AlphaGo isn't a sign of the end times. It's a powerful demonstration of deep learning, a hot subfield of AI research thanks to renewed interest in artificial neural networks, or ANNs.

BRAINY COMPUTERS

ANNs are algorithms — sets of rules — inspired by the way researchers believe the human brain processes information. To understand how they work, it's easiest to start from the beginning, in 1943, when neurophysiologist Warren McCulloch and mathematician Walter Pitts used math to describe the function of neurons in animal brains.

The McCulloch-Pitts neural model is an equation used to convert a series of weighted inputs into a binary output. Lots of data go in, and a 0 or 1 comes out. Add up a mess of numbers and if the solution is greater than or equal to a predetermined total, the output is a 1. If the solution falls below the total, the output is a 0. It's a simplified simulation of how neurons in the brain work: They either fire or don't fire.

Over decades, computer scientists have built upon this foundation, subtly tweaking the mathematical logic of model neurons, connecting multiple neurons and assembling them into hierarchical, layered networks — ANNs. Many ANNs in use today were actually fully described and theoretically executable decades ago, but they weren't as useful then. "AI's overnight success has been 30 years in the making," says Etzioni.
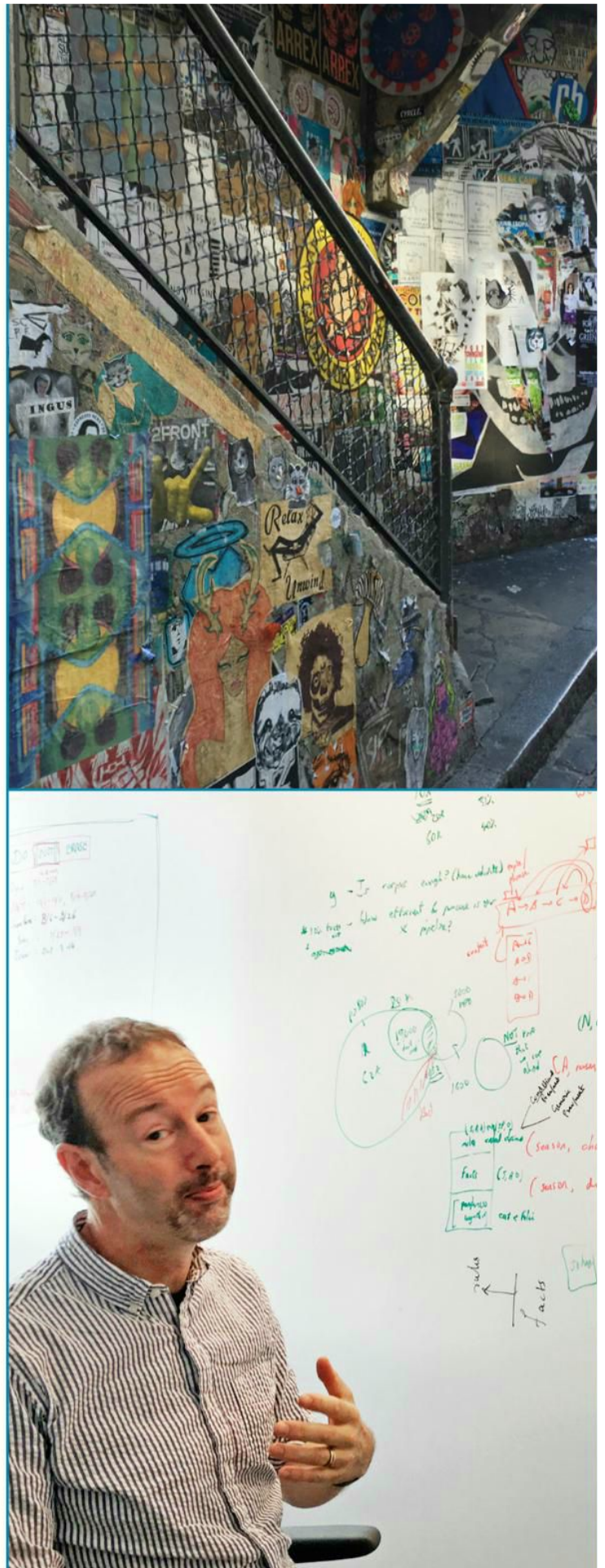
AI researchers configure ANNs for specific tasks, dictating how data flows through them in order to "teach" machines. To have an ANN learn to recognize images of Seattle's iconic Space Needle, for example, scientists might use neurons in the ANN's first layer to compute the brightness of a single pixel. Layers above it in the hierarchy might zero in on the structure's shape. As more Space Needle images are fed through the network, the weighted math that links these digital neurons automatically adjusts, based on the algorithm's parameters, strengthening connections that are unique to the Space Needle while weakening others.

This was the secret to AlphaGo's victory. It extracted winning strategies from thousands of Go games played by humans, pushing them through ANNs. It then played itself millions of times, tuning its networks to optimal Go strategies, always improving. "It was a huge success, but it was a narrow success that took

within 24 hours. As Farhadi explains, AIs are only as effective as the data they are fed.

"Data is the golden key," Farhadi says. "The minute the data are lacking, it's going to cause us trouble." We know a butterfly is smaller than an elephant, but if no one took the time to write that, it's tough for a machine to learn it. If a tree falls in the forest and generates no data, that tree never existed, as far as an AI is concerned.

MAKING THE GRADE

Meanwhile, down the hall from Farhadi, AI2's senior research manager Peter Clark takes a different approach to learning. He forces his subjects to complete the New York Regents Science Test over and over again. It would be cruel and unusual were it not inflicted on machines.

"Passing even a fourth-grade science test isn't a single task. It's a collection of skills that need to come together," he says. In February 2016, AI2 challenged thousands of researchers worldwide to

ANNs are algorithms — sets of rules — inspired by the way researchers believe the human brain processes information.

years of work from a large group of people," Etzioni says. "AlphaGo can't even play chess. It can't talk about the game. My 6-year-old is smarter than AlphaGo."

AlphaGo isn't alone. Virtually every AI we interact with can be startlingly dense. A Roomba teaches itself the layout of your living room, but it will still run over dog poop on the rug and turn the house into a fecal Jackson Pollock painting. Microsoft's chatbot Tay, programmed to generate human-like conversations based on inputs from Twitter, morphed into a foul-mouthed racist

develop an AI that could pass a standard eighth-grade science test. The top prize went to Israel's Chaim Linhart, whose program scored 59 percent.

Science tests serve as a gateway toward commonsense computers. The exams require specific and general knowledge to pass, and Clark can easily check his research's progress by grading the computer's performance. The tests contain diagrams, open-response questions, reading comprehension questions and more.

Teaching machines just one facet of the test — understanding diagrams — exhausted Clark,

Opposite: Human brains can easily perceive the layers of posters and scrawled messages around a Seattle stairwell (top). The artificial intelligence of the computers that Peter Clark (below) works with would struggle to comprehend such a tableau — for now.

How We Worry About Artificial Intelligence

Charting the percentage of AI-related articles in *The New York Times* that mention a specific concern reveals our changing attitudes.

LOSS OF CONTROL

WORK (NEGATIVE)

ETHICAL CONCERNS FOR AI
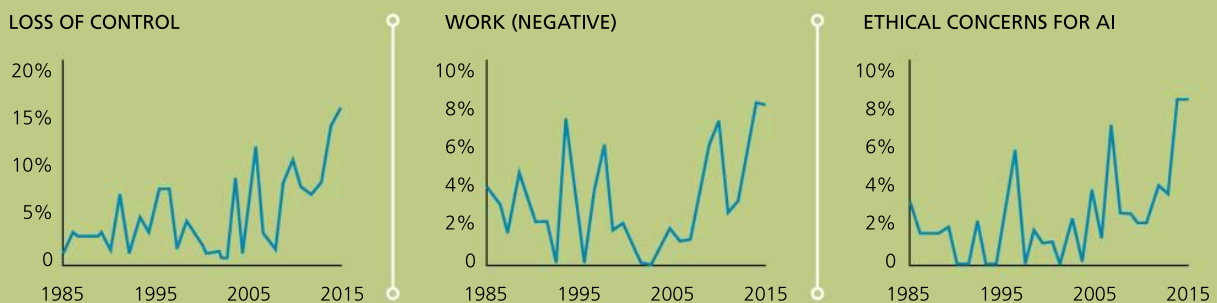
SOURCE: "Long-Term Trends in the Public Perception of Artificial Intelligence," Association for the Advancement of Artificial Intelligence, Dec 2016.

who needed to build a new database of 5,000 annotated diagrams and 15,000 multiple-choice questions. All the data were then annotated, keystroke by keystroke, clarifying relationships and what the diagrams were saying. Only then could Clark's team design and train a system that could answer questions about diagrams.

Every new dataset created at AI2 — and every diagram, video or block of text parsed by a machine — improves upon the other, bringing Etzioni's vision of the scientist's apprentice closer to reality. Eventually, rather than eighth-grade science-test diagrams, Etzioni's team will design algorithms that interpret images, diagrams and text in advanced scientific papers to make new connections and insights, based on its knowledge. Currently, AI2's Semantic Scholar search engine is a glimpse of what's to come; it's the keystone project where all their research will flow.

Semantic Scholar uses numerous ANNs in parallel to identify valuable information from studies. It combines these skills to understand not only the information conveyed in a given study, but also its relevance to the larger body of research. "Medical breakthroughs should not be hindered by the cumbersome process of searching the scientific literature," Etzioni says. AI2 isn't alone in building AI-enhanced search engines, but again, this is just a first step.

It sounds great, and I'm sure Etzioni has the best intentions, but I admit, it's hard not to worry a little. The robot apocalypse presaged in *The Terminator* might not (and almost certainly won't) come to pass, but smarter machines aren't exactly risk-free.

A GRUNGY FUTURE

After my time at AI2's headquarters, I walk past several sagging tents beneath an overpass in downtown Seattle. Two feet stick out of one. A block away, a man without teeth yells incoherently at four police officers imploring him to stand and put on his shoes. He can't. His clothes are in tatters. Is this a glimpse of the future, where more and more people are left behind, replaced by machines that think better and act faster than humanly possible?

"We do need to think hard about the impact on jobs," Etzioni says. A World Economic Forum analysis last year estimated that by 2020,

Software engineer Roie Levin (left) works in the airy, countercultural offices of AI2's Seattle headquarters.
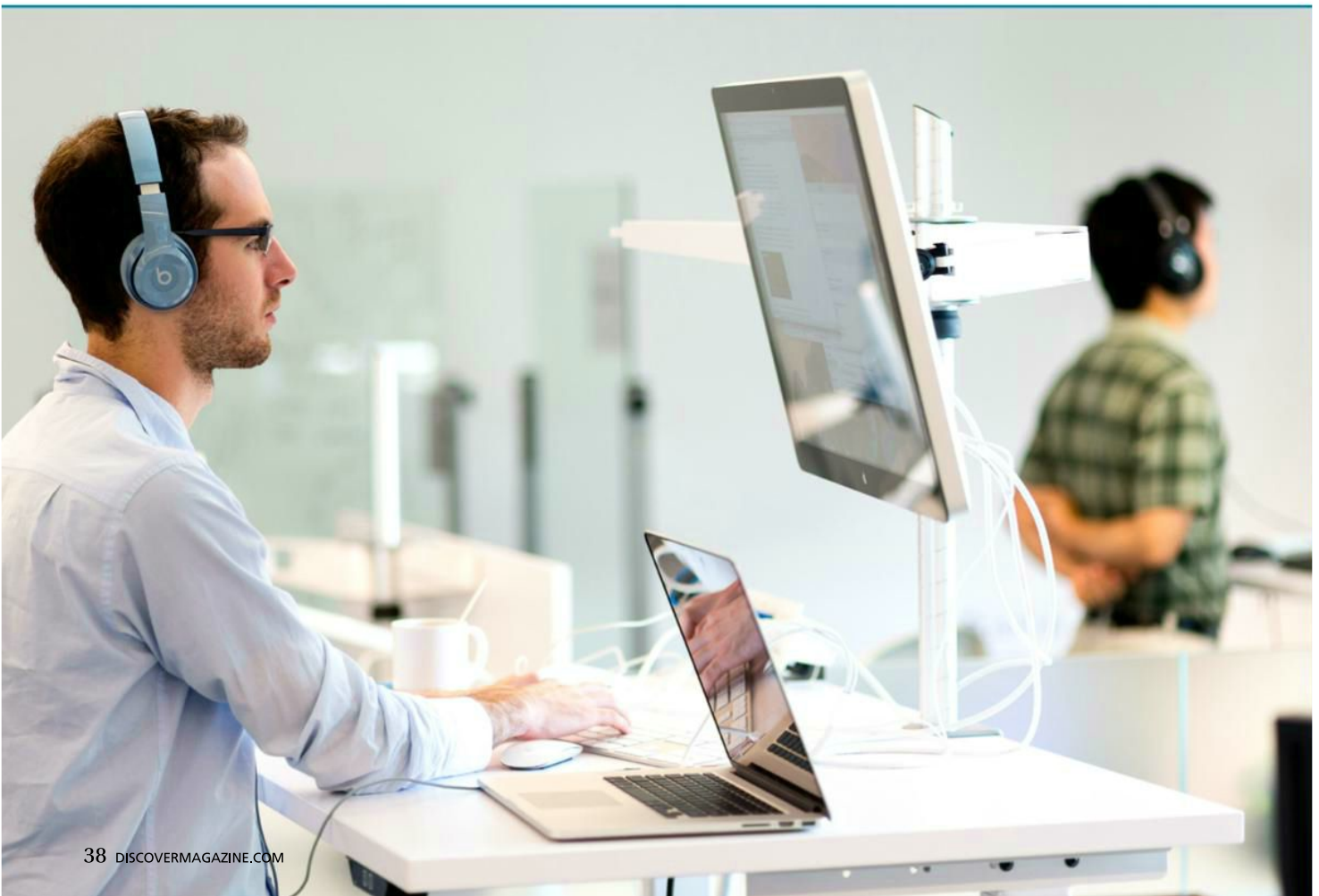
automation and robots will eliminate roughly 5 million jobs in 15 of the world's developed and emerging economies. In a 2016 global survey of 800 CEOs, 44 percent indicated they believe AI will make people "largely irrelevant" in the future of work.

Not all predictions are gloomy. The Obama administration published a 2016 report that outlined a generally optimistic future, with AI serving as a major driver of economic growth and social progress. Sure, AI technologies could displace low-wage, uneducated workers, but the report suggests it's the job of policymakers to ensure these people are "retrained and able to succeed in occupations that are complementary to ... automation." Bernie Meyerson, chief

reoffenders, and it mislabeled as low risk white defendants who went on to commit additional crimes more often. The way the algorithms were designed, and the data that programmers chose to feed them, affected these results.

Etzioni has a theoretical workaround to these ethical quandaries: guardian AIs that would use deep-learning techniques to keep tabs on other AIs working on socially important tasks, like approving loans or assessing criminal behavior. "The guardian AI would have unlimited attention, unflagging patience and can keep up." It could ensure another AI doesn't fall off the rails.

But who will program the guardian AIs? Imperfect humans. AI studies quickly

The robot apocalypse presaged in *The Terminator* might not come to pass, but smarter machines aren't exactly risk-free.

innovation officer at IBM, assured me that AI technologies won't displace us — they'll make us better. These things are resources, he says; they work by amplifying what a person already does best. We'll see if the pessimists or the optimists were closer to the mark.

But there's another difficulty with growing reliance on AI: It's a thoroughly human endeavor. Choosing what's in a dataset or what's not in it, adjusting parameters in algorithms and so on are all subjective decisions. Seattle grunge band Alice in Chains opened one of their most iconic songs, "Man in the Box," with the line, "I'm the man in the box / Buried in my s---." It reminds us of the messiness of existence, of addiction, of being buried in the filth of our imperfections. All of those shortcomings will be reflected in the designs of our machines. "Machine learning is 99 percent human learning," Etzioni says.

Take deep-learning software, widely used in the legal system today. These systems generate risk scores that assess the likelihood that a defendant will commit another violent crime. Independent journalism nonprofit *ProPublica* investigated 7,000 people arrested in Broward County in Florida, finding that only 20 percent of those pegged as high risk by the particular system, called COMPAS, went on to commit another violent crime. Other issues: COMPAS was twice as likely to flag black defendants as

branch into questions of philosophy, ethics and spirituality. Researchers are already hard at work addressing them, but there are no easy answers.

AN END, OR A NEW BEGINNING
Down the street from AI2 is Seattle's iconic Gas Works Park. Its primary feature is the rusted guts of an old plant that fueled the city decades ago. For an outstanding view of the skyline, you can climb the switchbacks of the Great Mound, a pile of rubble from the old plant now covered in dirt and grass. Late in the afternoon, when the sun is low, I stand on top of the mound, casting a 15-foot shadow on the hulking machinery.

Staring at my shadow as it dances across the dormant pipes and barrels, I wonder if I'll share the fate of this industrial artifact within my lifetime. AIs are already writing sports recaps and financial news. Is it just a matter of time before they move on to science journalism? Will imperfectly programmed machines impact my life without my knowledge?

But the evening is perfect. The clouds have lifted, and the sky is clear — a luxury in this city. For now, I enjoy the setting sun. **D**

Carl Engelking is the generally intelligent being serving as web editor at Discover. *Follow him on Twitter, @CarlJamesKing*