What's Next for Deep Learning › Another AI winter or eternal sunshine?
P.26

Inside DeepMind's Robot Lab › An AI powerhouse takes on "catastrophic forgetting"
P.34

The 7 Biggest Weaknesses of Neural Nets › Surprise! One of them is math
P.42

FOR THE TECHNOLOGY INSIDER

OCTOBER 2021

IEEE Spectrum

Why Is AI So Dumb?

A SPECIAL REPORT

IEEE

Liming Chen, The University of Manchester, UK

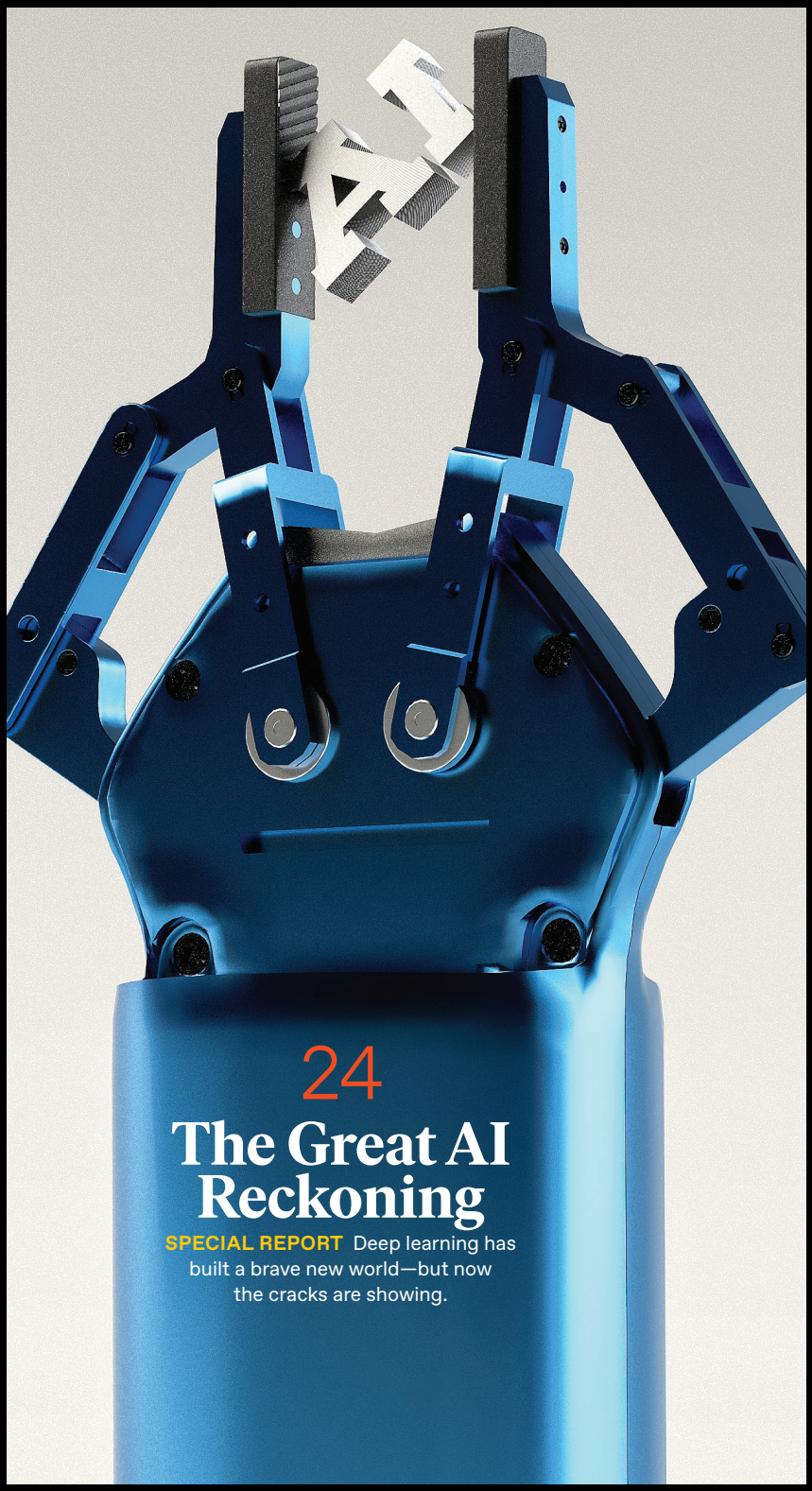
Redefining measurement boundaries

We applaud the electromagnetic sensing and wearable textiles joint lab at the University of Manchester for developing wearable capacitive sensors that will improve the monitoring critical for physical rehabilitation. Thanks to improvements in binding electrodes to the textile, the sensor is robust yet comfortable to wear, and provides realtime high sensitivity information on movement. The sensor is a perfect example of internet of medical things which will help doctors to treat patients remotely.

We look forward to more groundbreaking results in the field of wearable sensors and electronics.

Zurich Instruments

Contents



AI EXPERTS SPEAK The quotes scattered throughout this issue come from *IEEE Spectrum*'s prior AI coverage. Go to spectrum.ieee.org/aiquotes for links to the full articles.

Cover and above: Illustration by Eddie Guy

MATTHEW S. SMITH

BACK STORY

Gadgeteering

A once-famous comic-strip character called L'il Abner would come home tired after a hard day's work...testing mattresses. That's the layman's idea of what it must be like to review cool gadgets for a living. True, but only up to a point, says Matthew S. Smith, who writes Gizmo, our new consumer-electronics column.

Smith, 37, grew up in Indiana and got hooked on consumer electronics in 1995, when his mother bought a computer. After graduating with a degree in English, he started writing product reviews, first for free, then for a nominal fee, and finally for a living. Fourteen years ago he got closer to the West Coast action by moving to Portland, Ore., "where young people go to retire," he quips.

That's him in the photo at the most recent in-person CES, in January 2020, in Las Vegas. He's been going to this influential consumer-electronics convention for a decade.

This month's column, on page 20, reviews some machines that measure indoor air quality and others that perhaps improve it. It's a problem Smith faced firsthand during last year's wildfires. "So far this year it's not been too bad," he says. "Right now the winds are off the ocean, blowing the smoke out to the east, but that might change in a month."

In previous Gizmo articles, he tackled electric bikes, computer monitors, and the new M1 chip that's powering Apple computers.

In each case, Smith strives to get an early look at products, which generally involves agreeing not to publish until the official release date. "But these embargoes can be pretty tight," he says. "You might end up having just three or four days to use the device, maybe run some benchmark tests, and write about it."

So what does the guy who has everything—at least for a short time—do for fun? Smith's tech hobby is gaming, often using classic gaming gear he's found and restored. His nontech hobby is gardening: "Tomatoes—one of my favorite things—onions, peppers, raspberries, fruit trees, and finally grapes," which he says take a few years to yield a crop. "It's nice to have something where you don't have to look at a screen." ■

OPTICA
Advancing Optics and Photonics Worldwide
Formerly OSA

A new light.
A new day.
A new brand.

We are excited to announce the launch of our society's new name, Optica (formerly OSA). Inspired by our work advancing optics and photonics worldwide, our new name and look reflect our community.

Optica is a society for today and for tomorrow. Now, more than ever, we unite a diverse community, empower discovery and drive innovation for the future.

To learn more about our new brand, visit **optica.org/brand**.

RODNEY BROOKS

Brooks was a cofounder of iRobot Corp, Rethink Robotics, and most recently, Robust AI. A Fellow of the IEEE, he was director of the MIT Computer Science & Artificial Intelligence Laboratory (CSAIL). In his article, on page 48, he argues that today we can trust AI only when the stakes aren't high or when there's a human in the loop. "Although we willingly put our lives in the hands of AI systems that choose who we meet on dating sites, I'm not ready to let one drive me around town," he says.

TOM CHIVERS

Chivers is science editor at UnHerd.com and a freelance science writer. He is the author of *How to Read Numbers* and *The Rationalist's Guide to the Galaxy*, and is the Association of British Science Writers' current science journalist of the year. He wrote "a whole book about whether AI is going to kill us all," he notes, but he still went bravely into DeepMind's offices to report "How to Train an All-Purpose Robot," on p. 34.

CHARLES Q. CHOI

Choi, a contributing editor to *IEEE Spectrum*, has written for *Scientific American*, *The New York Times*, *Science*, and *Nature*, among others. In his spare time, he has traveled to all seven continents and published science fiction in *Analog* magazine. On page 42 he writes about AI failures. "It's very interesting how the ways in which AI works poorly shed light on the foibles of human intelligence as well," says Choi.

NEIL C. THOMPSON

Thompson is a research scientist at MIT's Computer Science & Artificial Intelligence Laboratory. On page 50, he and coauthors Kristjan Greenewald of the MIT–IBM Watson AI Lab, Keeheon Lee of Yonsei University, in Seoul, and Gabriel F. Manso of the University of Brasília consider the swiftly growing computational resources dedicated to training deep neural networks. "Continuing on the same path is not going to be sustainable," says Thompson.

IEEE Spectrum

EDITOR IN CHIEF Susan Hassler, s.hassler@ieee.org
EXECUTIVE EDITOR Glenn Zorpette, g.zorpette@ieee.org
EDITORIAL DIRECTOR, DIGITAL
Harry Goldstein, h.goldstein@ieee.org
MANAGING EDITOR Elizabeth A. Bretz, e.bretz@ieee.org
SENIOR ART DIRECTOR
Mark Montgomery, m.montgomery@ieee.org
PRODUCT MANAGER, DIGITAL
Erico Guizzo, e.guizzo@ieee.org
SENIOR EDITORS
Evan Ackerman (Digital), ackerman.e@ieee.org
Stephen Cass (Special Projects), cass.s@ieee.org
Jean Kumagai, j.kumagai@ieee.org
Samuel K. Moore, s.k.moore@ieee.org
Tekla S. Perry, t.perry@ieee.org
Philip E. Ross, p.ross@ieee.org
David Schneider, d.a.schneider@ieee.org
Eliza Strickland, e.strickland@ieee.org
DEPUTY ART DIRECTOR Brandon Palacio, b.palacio@ieee.org
PHOTOGRAPHY DIRECTOR Randi Klett, randi.klett@ieee.org
ONLINE ART DIRECTOR Erik Vrielink, e.vrielink@ieee.org
NEWS MANAGER Mark Anderson, m.k.anderson@ieee.org
ASSOCIATE EDITORS
Willie D. Jones (Digital), w.jones@ieee.org
Michael Koziol, m.koziol@ieee.org
SENIOR COPY EDITOR Joseph N. Levine, j.levine@ieee.org
COPY EDITOR Michele Kogon, m.kogon@ieee.org
EDITORIAL RESEARCHER Alan Gardner, a.gardner@ieee.org
ADMINISTRATIVE ASSISTANT
Ramona L. Foster, r.foster@ieee.org
CONTRIBUTING EDITORS Robert N. Charette, Steven Cherry, Charles Q. Choi, Peter Fairley, Maria Gallucci, W. Wayt Gibbs, Mark Harris, Jeremy Hsu, Allison Marsh, Prachi Patel, Megan Scudellari, Lawrence Ulrich, Emily Waltz

EDITOR IN CHIEF, *THE INSTITUTE*
Kathy Pretz, k.pretz@ieee.org
ASSISTANT EDITOR, *THE INSTITUTE*
Joanna Goodrich, j.goodrich@ieee.org
DIRECTOR, PERIODICALS PRODUCTION SERVICES
Peter Tuohy
MULTIMEDIA PRODUCTION SPECIALIST Michael Spector
ASSOCIATE ART DIRECTOR, PUBLICATIONS
Gail A. Schnitzer

ADVERTISING PRODUCTION +1 732 562 6334
ADVERTISING PRODUCTION MANAGER
Felicia Spagnoli, f.spagnoli@ieee.org
SENIOR ADVERTISING PRODUCTION COORDINATOR
Nicole Evans Gyimah, n.gyimah@ieee.org

MANAGING DIRECTOR, PUBLICATIONS Steven Heffner

EDITORIAL CORRESPONDENCE
IEEE Spectrum, 3 Park Ave., 17th Floor,
New York, NY 10016-5997
TEL: +1 212 419 7555 FAX: +1 212 419 7570
BUREAU Palo Alto, Calif.; Tekla S. Perry **+1 650 752 6661**

DIRECTOR, BUSINESS DEVELOPMENT, MEDIA & ADVERTISING Mark David, m.david@ieee.org

ADVERTISING INQUIRIES Naylor Association Solutions, Erik Henson +1 352 333 3443, ehenson@naylor.com

REPRINT SALES +1 212 221 9595, ext. 319

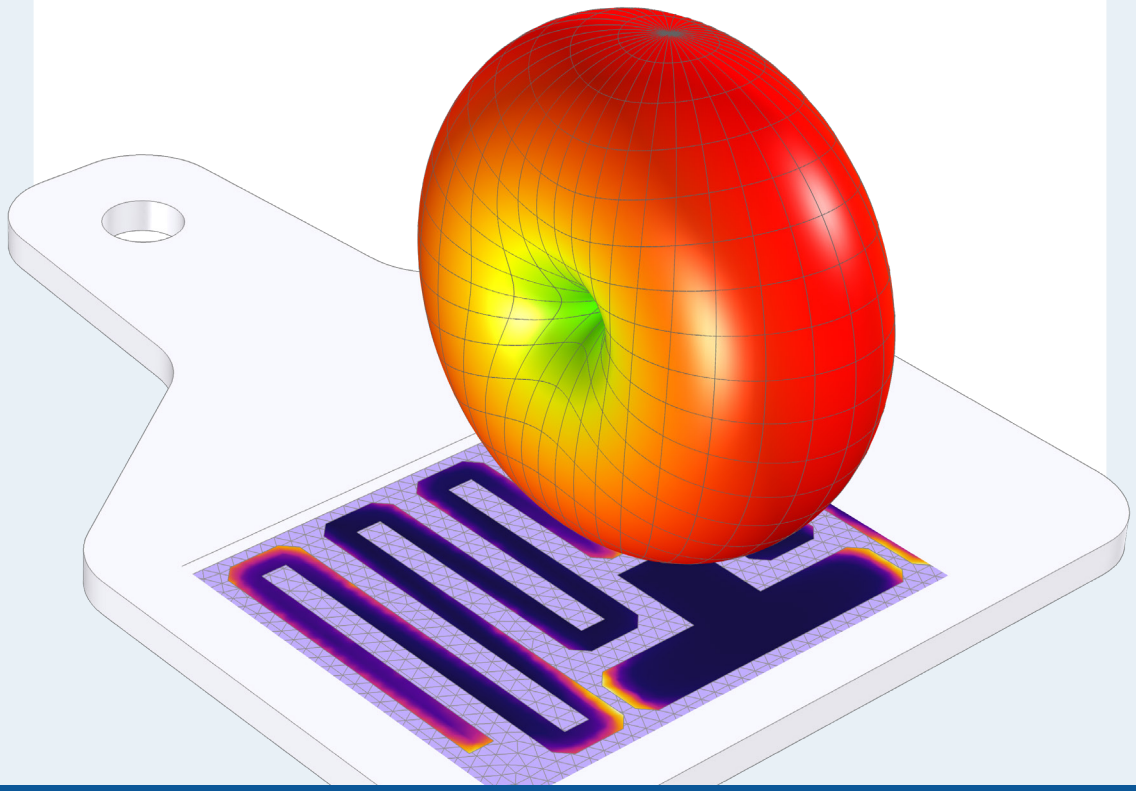
IEEE SPECTRUM (ISSN 0018-9235) is published monthly by The Institute of Electrical and Electronics Engineers, Inc. All rights reserved. © 2021 by The Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, New York, NY 10016-5997, U.S.A. Volume No. 58, Issue No. 10. The editorial content of IEEE Spectrum magazine does not represent official positions of the IEEE or its organizational units. Canadian Post International Publications Mail (Canadian Distribution) Sales Agreement No. 40013087. Return undeliverable Canadian addresses to: Circulation Department, IEEE Spectrum, Box 1051, Fort Erie, ON L2A 6C7. Cable address: ITRIPLEE. Fax: +1 212 419 7570. INTERNET: spectrum@ieee.org. ANNUAL SUBSCRIPTIONS: IEEE Members: $21.40 included in dues. Libraries/institutions: $399. POSTMASTER: Please send address changes to IEEE Spectrum, % Coding Department, IEEE Service Center, 445 Hoes Lane, Box 1331, Piscataway, NJ 08855. Periodicals postage paid at New York, NY, and additional mailing offices. Canadian GST #125634188. Printed at 120 Donnelley Dr., Glasgow, KY 42141-1060, U.S.A. IEEE Spectrum circulation is audited by BPA Worldwide. IEEE Spectrum is a member of the Association of Business Information & Media Companies, the Association of Magazine Media, and Association Media & Publishing. IEEE prohibits discrimination, harassment, and bullying. For more information, visit https://www.ieee.org/web/aboutus/whatis/policies/p9-26.html.

SIMULATION CASE STUDY

Smartphones, smart homes, smart...healthcare?

RFID tags are used across many industries, but when it comes to healthcare, there is a major design challenge: size. If wearable RFID tags are too big and bulky, they could cause patient discomfort. Or, if the tag is for a biomedical implant, it has to be smaller than a grain of rice! Design engineers can optimize the size of an RFID tag for its intended purpose using RF simulation.

LEARN MORE *comsol.blog/biomed-RFID-tags*



COMSOL

The COMSOL Multiphysics® software is used for simulating designs, devices, and processes in all fields of engineering, manufacturing, and scientific research.

News

AEROSPACE

China's Lunar Station Megaproject › Moon base could be a stepping-stone to the solar system

BY ANDREW JONES

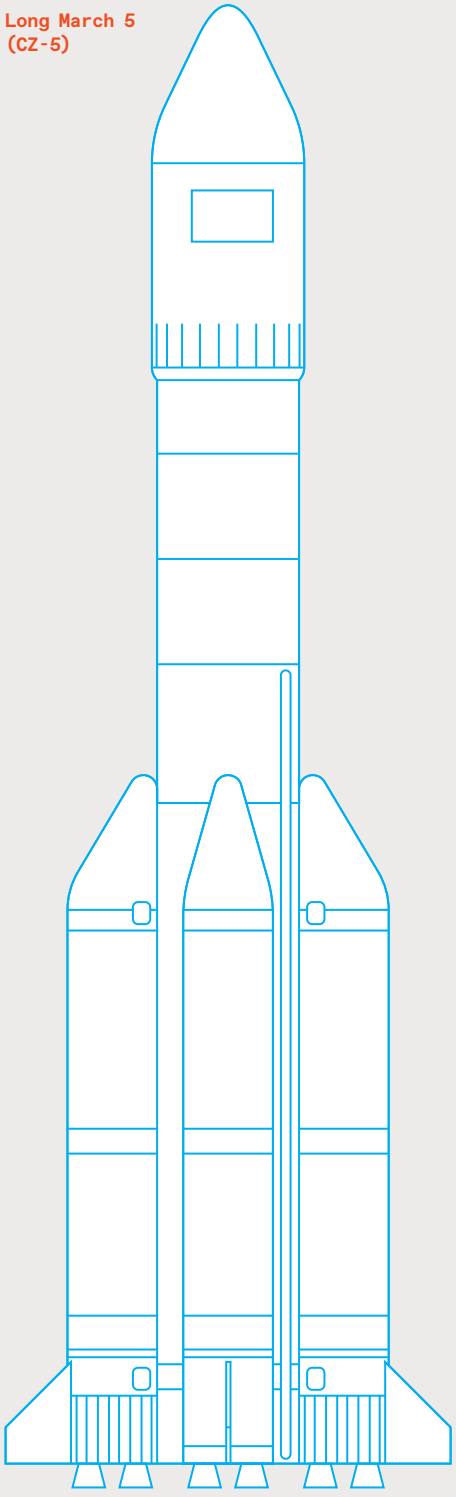
On 3 January 2019, the Chinese spacecraft Chang'e-4 descended toward the moon. Countless craters came into view as the lander approached the surface, the fractal nature of the footage providing no sense of altitude. Su Yan, responsible for data reception for the landing at Miyun ground station, in Beijing, was waiting—nervously and in silence with her team—for vital signals indicating that optical, laser, and microwave sensors had combined effectively with rocket engines for a soft landing. "When the [spectral signals were] clearly visible, everyone cheered enthusiastically. Years of hard work had paid off in the most sweet way," Su recalls.

Chang'e-4 had, with the help of a relay satellite out beyond the moon, made an unprecedented landing on the always-hidden lunar far side. China's space program, long trailing in the footsteps of the U.S. and Soviet (now Russian) programs, had registered an international first. The landing also prefigured grander Chinese lunar ambitions.

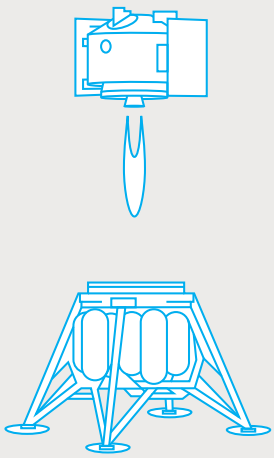
In 2020 Chang'e-5, a complex sample-return mission, returned to Earth with young lunar rocks, completing China's three-step "orbit, land, and return" lunar program conceived in the early 2000s. These successes, together with renewed international scientific and commercial interest in the moon, have emboldened China to embark on a new lunar project that builds on the Chang'e program's newly acquired capabilities.

The International Lunar Research Station (ILRS) is a complex, multiphase megaproject that the China National Space Administration (CNSA) unveiled jointly with Russia in June in St. Petersburg. Starting with robotic landing and orbiting missions in the 2020s, its designers envision a permanently inhabited lunar base by the mid-2030s. Objectives include science, exploration, technology verification, resource and commercial exploitation, astronomical observation, and more.

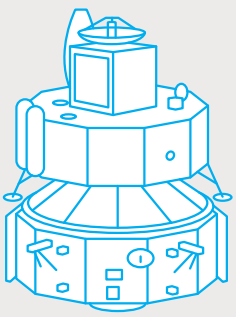
ILRS will begin with a robotic reconnaissance phase running up to 2030, using orbiting and surface spacecraft to survey potential landing areas and resources, conduct technology-verification tests, and assess the prospects for an eventual permanent crewed base on the moon. The phase will consist of Chinese missions Chang'e-4,



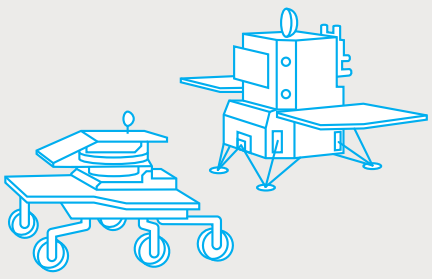
Illustrations by James Provost



2022–24 Chang'e-6



2023–24 Chang'e-7



2025–27 Chang'e-8

The China National Space Administration (CNSA) recently unveiled its plans for a lunar base in the 2030s, the International Lunar Research Station (ILRS). The first phase involves prototyping, exploration, and reconnaissance of possible ILRS locations.

Chang'e-6 sample return, and the more ambitious Chang'e-7, as well as Russian Luna spacecraft, plus potential missions from international partners interested in joining the endeavor. Chang'e-7 will target a lunar south pole landing and consist of an orbiter, relay satellite, lander, and rover. It will also include a small spacecraft capable of "hopping" to explore shadowed craters for evidence of potential water ice, a resource that, if present, could be used in the future for both propulsion and supplies for astronauts.

CNSA will help select the site for a two-stage construction phase that will involve in situ resource utilization (ISRU) tests with Chang'e-8, massive cargo delivery with precision landings, and the start of joint operations between partners. ISRU, in this case using the lunar regolith (the fine dust, soil, and rock that makes up most of the moon's surface) for construction and extraction of resources such as oxygen and water, would represent a big breakthrough. Being able to use resources already on the moon means fewer things need to be delivered, at great expense, from Earth.

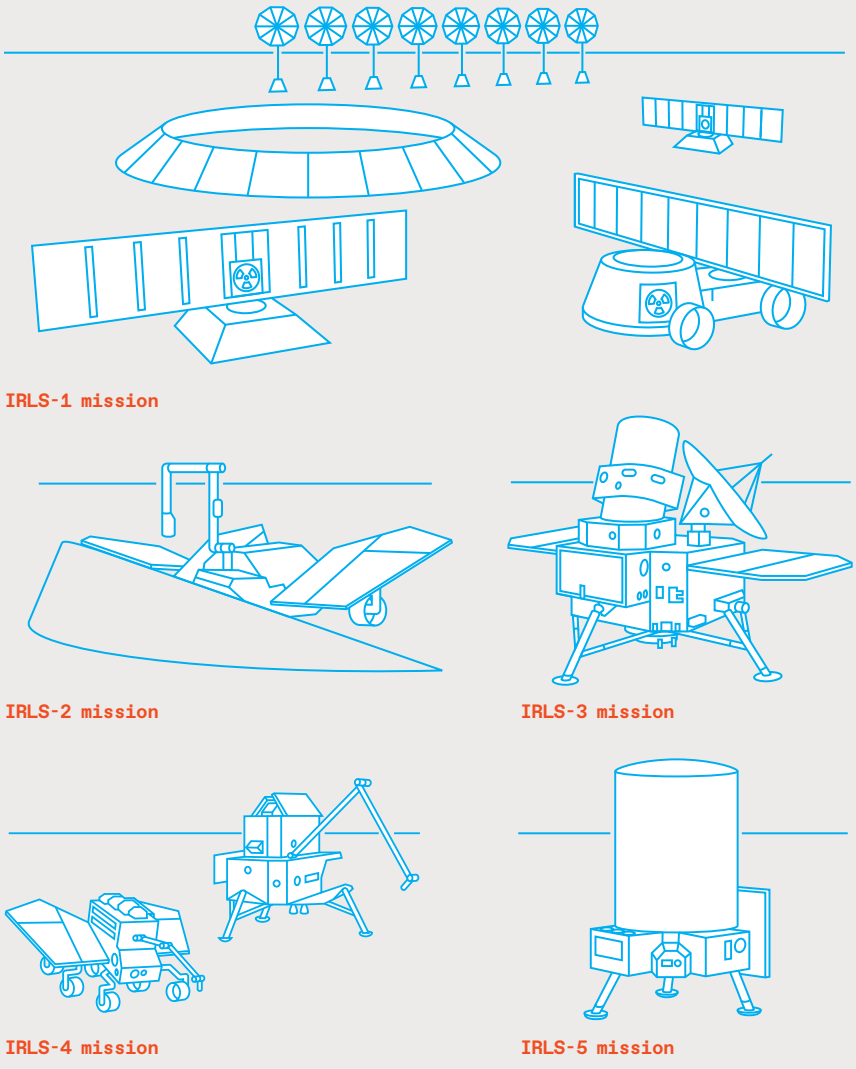
The utilization phase will begin in the early 2030s. It tentatively consists of missions numbered ILRS-1 through 5 and relies on heavy-lift launch vehicles to establish command, energy, and telecommunications infrastructure; experiment, scientific, and IRSU facilities; and Earth- and astronomical-observation capabilities. CNSA artist renderings indicate spacecraft will use the lunar regolith to make structures that would provide shielding from radiation while also exploring lava tubes as potential alternative areas for habitats.

The completed ILRS would then host and support crewed missions to the moon in around 2036. This phase, CNSA says, will feature lunar research and exploration, technology verification, and expanding and maintaining modules as needed.

These initial plans are vague, but senior figures in China's space industry have noted huge, if challenging, possibilities that could greatly contribute to development on Earth. Ouyang Ziyuan, a cosmochemist and early driving force for Chinese lunar exploration, notes in a July talk the potential extraction of helium-3, delivered to the lunar surface by unfiltered solar wind, for nuclear fusion (which would require major breakthroughs on Earth and in space). Another possibility is 3D printing of solar panels at the moon's equator, which would capture solar energy to be transmitted to Earth by lasers or microwaves. China is already conducting early research toward this end. As with NASA's Artemis plan, Ouyang notes that the moon is a stepping-stone to other destinations in the solar system, both through learning and as a launchpad.

The more distant proposals currently appear beyond reach, but in its space endeavors China has demonstrated a willingness to develop capabilities and apply these for new possibilities. Sample-return tech from Chang'e-5 will next be used to collect material from a near-Earth asteroid around 2024. Near the end of the decade, this tech will contribute to the Tianwen-1 Mars mission's capabilities for an unprecedented Mars sample-return attempt. How the ILRS develops will then depend on success and science and resource findings of the early missions.

China is already well placed to implement the early phases of the ILRS blueprint. The Long March 5, a heavy-lift rocket, had its first flight in 2016 and has since enabled the country to begin constructing a space station and to launch spacecraft such as a first independent interplanetary mission and Chang'e-5. To develop the rocket, China had to make break-



CNSA's plans for its international moon base involve a set of missions, dubbed ILRS-1 through ILRS-5, now projected between 2031 and 2035. IRLS-1, as planned, will in 2031 establish a command center and basic infrastructure. Subsequent missions over the ensuing four years would set up research facilities, sample-collection systems, and Earth- and space-observation capabilities.

throughs in using cryogenic propellant and machining a new, wider-diameter rocket body.

This won't be enough for larger missions, however. Huang Jun, a professor at Beihang University, in Beijing, says a super heavy-lift rocket, the high-thrust Long March 9, is a necessity for the future of Chinese aerospace. "Research and breakthroughs in key technologies are progressing smoothly, and the project may at any time enter the engineering-development stage."

The roughly 100-meter-long, Saturn V–like Long March 9 will be capable of launching around 50 tonnes of payload to translunar injection. The project requires precision manufacturing of thin yet strong, 10-meter-diameter rocket stages and huge new engines. In Beijing, propulsion institutes under the China Aerospace Science and Technology Corp., recently produced an engineering prototype of a 220-tonne-thrust staged-combustion liquid hydrogen/liquid oxygen engine. In a ravine near Xi'an, in north China, firing tests of a dual-chamber 500-tonne-thrust kerosene/liquid oxygen engine for the first stage have been carried out. Long March 9 is expected to have its first flight around 2030, which would come just in time to launch the robotic ILRS construction missions.

A human-rated rocket is also under development, building on technologies from the Long March 5. It will feature similar but uprated versions of the YF-100 kerosene/liquid oxygen engine and use three rocket cores, in a similar fashion to SpaceX's Falcon Heavy. Its task will be sending a deep-space-capable crew spacecraft into lunar orbit, where it could dock with a lunar-landing stack launched by a Long March 9.

The spacecraft itself is a new-generation advance on the Shenzhou, which currently ferries astronauts to and from low Earth orbit. A test launch in May 2020 verified that the new vessel can handle the greater heat of a higher-speed atmospheric reentry from higher, more energetic orbits. Work on a crew lander is also assumed to be underway. The Chang'e-5 mission was also seen as a scaled test run for human landings, as it followed a profile similar to NASA's Apollo missions. After lifting off from the moon, the ascent vehicle reunited and docked with a service module, much in the way that an Apollo ascent vehicle rejoined a command module in lunar orbit before the journey home.

China and Russia are inviting all interested countries and partners to cooperate in the project. The initiative will be separate from the United States' Artemis moon program, however. The United States has long opposed cooperating with China in space, and recent geopolitical developments involving both Beijing and Moscow have made things worse still. As a result, China and Russia, its International Space Station partner, have looked to each other as off-world partners. "Ideally, we would have an international coalition of countries working on a lunar base, such as the Moon Village concept proposed by former ESA director-general Jan Wörner. But so far geopolitics have gotten in the way of doing that," says Brian Weeden, director of program planning for the Secure World Foundation.

The final details and partners may change, but China, for its part, seems set on continuing the accumulation of expertise and technologies necessary to get to the moon and back, and stay there in the long term. ■

Illustrations by James Provost

ENERGY

China's Thorium Gambit > A prototype power reactor could go on line by 2030

BY PRACHI PATEL

There is no denying the need for nuclear power in a world that hungers for clean, carbon-free energy. At the same time, there's a need for safer nuclear technologies that bear less proliferation risk. Molten salt reactors (MSRs) fit the bill—and, according to at least one source, China may be well on its way to developing MSR technology.

Government researchers there unveiled a design for a commercial MSR that uses thorium as fuel, the *South China Morning Post* reported recently. Construction of the first commercial MSR, being built in Gansu province, should be complete, they noted, by 2030. According to the Australian Broadcasting Corp., citing a Gansu provincial government statement, first tests on a prototype reactor were expected sometime in September.

If all goes well with this prototype, says a report in Live Science, the Chinese government plans to build several large MSRs. According to the World Nuclear Association, the country is eyeing thorium MSRs as a source of energy especially for the northwestern portion of the country, which has lower population density and an arid climate.

MSRs are attractive for arid regions because instead of the water used by conventional uranium reactors, MSRs use molten fluoride salts to cool their cores. Uranium or thorium fuel can be mixed into the coolant salt. But thorium is more abundant and cheaper.

China's experimental reactor won't be the world's first. Researchers at Oak Ridge National Laboratory (ORNL), in Tennessee, pioneered thorium-based MSRs in the late 1940s for nuclear aircraft propulsion. In the 1960s, a 7.4-megawatt-hour experimental reactor operated at the laboratory over a period of four years—although only a portion of its fuel was derived from uranium-233 bred from thorium in other reactors. This MSR technology was eventually shelved because the U.S. government favored the uranium fast-breeder reactor, says Charles Forsberg, principal research scientist at MIT's department of Nuclear Science and Engineering and a former nuclear researcher at ORNL.

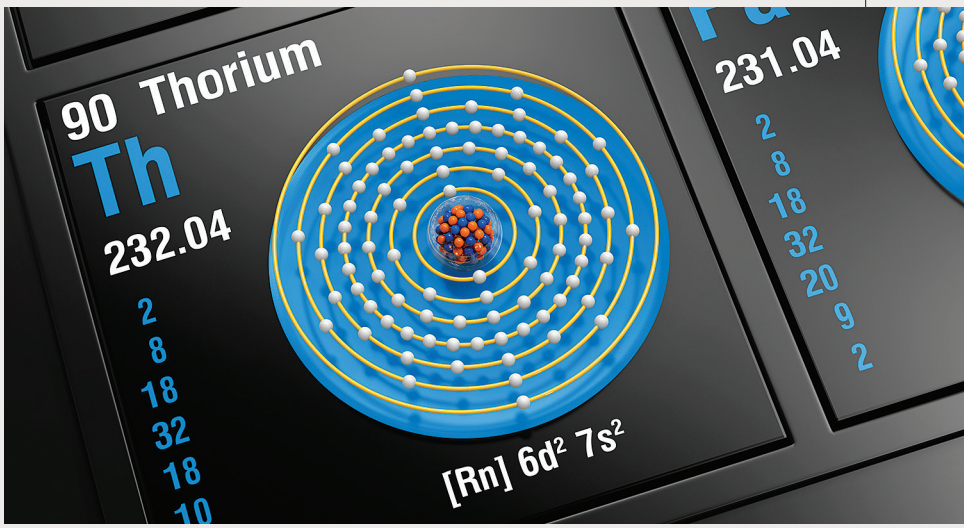
Scientists in China are now building on the same basic MSR technology developed at Oak Ridge. The Chinese government had a small, short-lived knowledge exchange program with ORNL. But most of the thorium-reactor-related intellectual property from the lab is in the public domain, and China appears to have made some use of it. "The real data mine is the thousands of published reports in the 1960s and '70s that are found in the open literature," Forsberg says.

Plus, recent technology advances have made it more feasible to build MSRs, he adds. These include modern instrumentation that can unveil exactly what goes on in the reactor—but also equipment developed in parallel, such as high-temperature salt pumps used in today's concentrated solar power (CSP) plants that store heat via molten salts.

"So now if you want to build a salt pump for an MSR you go talk to your local friendly CSP pump suppliers for a slightly different salt composition," Forsberg says. "That makes a tremendous difference in your development program. You have 50 years' worth of new technology to tap into."

But even though France, India, Japan, Norway, and the United States are all reportedly working on thorium nuclear reactors, none of these countries have outlined plans for commercial reactors yet. A handful of private-sector developers aim to deploy MSRs within the next decade. The closest is probably Alameda, Calif.–based Kairos Power, which plans to have a 50-megawatt demonstration reactor operational in Oak Ridge by 2026.

Yet China leads global MSR research, according to the World Nuclear Association, and it's no surprise that the country is forging ahead faster, Forsberg says. The country's talent pool in nuclear engineering, he says, is quite substantial. "You put a lot of talented people on a project, and it works," he says. "They'll be successful even if it takes them a while." ■



CARLOS CLARIVAN/SCIENCE SOURCE

CORONAVIRUS

Five COVID Breathalyzers > Blow into a tube, get the results in as little as 30 seconds

BY EMILY WALTZ

Concert venues, international airports and even restaurants are increasingly asking patrons for a recent negative COVID-19 test before entering their premises. Some organizations offer to test people on the spot as they enter.

But current COVID-19 testing options aren't convenient enough for the kind of mass daily screening that some businesses would like to implement. Rapid antigen tests take about 15 minutes and are in short supply. Molecular test results—the gold standard—often take days to become available. Both typically require twirling a swab up the nose—not exactly something people like to do as they head for a cocktail.

This has led many scientists to develop super-rapid testing methods using breath samples. Such devices are less socially awkward and can deliver results in under a minute—fast enough to feasibly screen large crowds as they pass through hubs.

Here, *IEEE Spectrum* has selected five different approaches to analyzing breath for SARS-CoV-2, the virus that causes COVID-19. Some of these technologies can sense the virus directly. Others pick up indirect indicators, such as volatile organic compounds (VOCs). These molecules are present in healthy breath, but change in ratio when a person is infected with the virus.

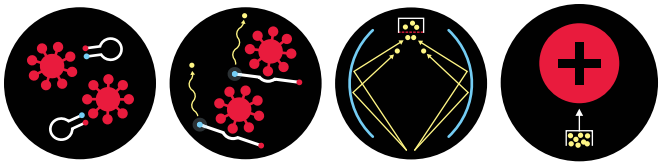
The technologies come from companies working in a wide range of applications that pivoted to COVID-19 when the pandemic hit. Steradian Technologies in the United States, for example, was building a product for human supersight, and managed to turn its optics technology into a diagnostic.

No COVID-19 breathalyzer is widely available yet, but we might soon see them popping up in select settings globally. In May, Singapore provisionally approved a breath-based device from Breathonix and may use it to test travelers at a Singapore-Malaysia checkpoint, according to the company. A device from similarly named Breathomix, in the Netherlands, was recently used by a port company in Rotterdam to check about 3,500 employees daily.

After more clinical validation, COVID-19 breath-based tests might finally give the world a more convenient and comfortable testing option. ■

PHOTONICS BIOSENSOR
Steradian Technologies, Houston
Rumi

TIME TO RESULTS: 30 SECONDS



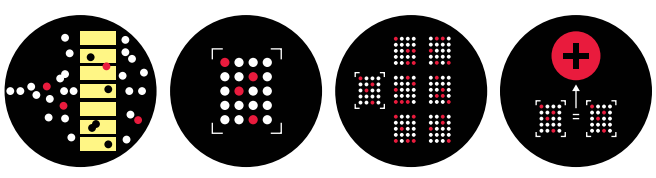
The user blows into a tube, and if the virus is present, its protein receptors will bind with a chemically reactive biosensor. The binding causes the biosensor to emit light in the form of photons. Mirrors inside the handheld device concentrate the light emissions to a single focal point, amplifying the signal and allowing a measurement to be made. Light emissions indicate a positive sample.

About the size of: A glue gun

Good for: Screening people before entering a business, concert, or school

ELECTRONIC NOSE
Breathomix, Leiden, Netherlands
SpiroNose

TIME TO RESULTS: < 1 MINUTE



A cylindrical electronic nose containing seven different biosensors detects exhaled VOCs in breath. The sensors, made of metal oxide semiconductors, react with compounds in the breath. The reactions cause measurable changes in the flow of electrons and indicate the presence of certain compounds. Pattern-recognition algorithms then compare the readings from the sample to those of healthy and infected sample profiles in its database. The device either delivers a negative result, or it recommends further testing. It does not, by itself, definitively provide a positive COVID-19 diagnosis.

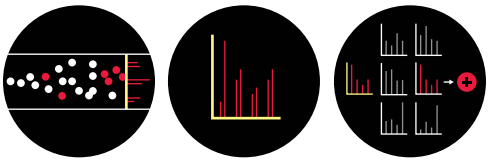
About the size of: A 500-milliliter water bottle

Good for: Screening students in school or employees at large companies

Illustration by StoryTK

MASS SPECTROMETRY
Breathonix, Singapore
BreFence Go COVID-19 Breath Test System

TIME TO RESULTS: < 1 MINUTE



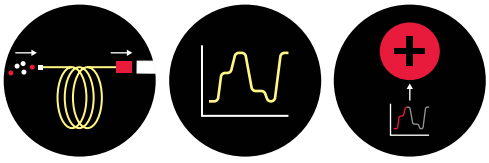
In this "time-of-flight" mass spectrometry approach, VOCs in a breath sample are fragmented, given an electric charge and subjected to magnetic fields. This causes the fragments to take different trajectories depending on their mass-to-charge ratios. A detector records their abundance in a mass spectrum based on these ratios and the time it takes for the molecules to travel a known distance through the machine. The data helps identify the VOCs present in the sample.

About the size of: A dishwasher

Good for: hospitals, clinical laboratories, and point-of-care settings with trained operators

GAS CHROMATOGRAPHY–ION MOBILITY SPECTROMETRY
Imspex Diagnostics, Abercynon, Wales
BreathSpec

TIME TO RESULTS: 8 MINUTES



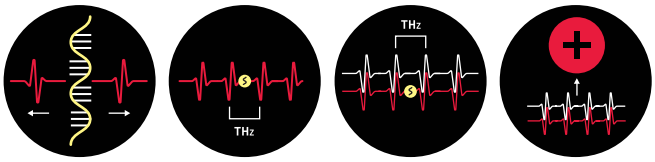
A breath sample moves through a gas-chromatography column, which separates particles based on their size. Molecules then enter an ion-mobility spectrometry chamber where they are ionized, accelerated across the chamber, and hit a Faraday plate. This results in a current that is specific to each molecule and is used to produce a 3D chromatogram that can be analyzed using machine learning algorithms. For COVID screening, the system looks for specific changes in the ratio of VOCs present in breath samples compared with those found in healthy breath.

About the size of: A microwave oven

Good for: Testing travelers at airports, crowds at cultural festivals, staff at large companies

TERAHERTZ SPECTROSCOPY
RAM Group DE, Zweibrücken, Germany
ThEA Terahertz Express Analyzer

TIME TO RESULTS: 2 MINUTES



A metamaterial nanoantenna deposited on glass creates resonances at specific points in the 1-to-2-terahertz range. These waves uniquely interact with SARS-CoV-2 and its protein structure. When the virus is present in a person's breath or throat-swab sample, it is drawn to the minuscule structures on the antenna. The presence of this viral matter generates disturbances in the resonances while interacting with the terahertz wave, creating a change in the spectrum. Detection of this signal indicates a positive result.

About the size of: A large microwave oven

Good for: Screening people coming through transportation hubs and commercial centers

NUCLEAR WEAPONS

One Atmospheric Nuclear Explosion Could Cripple the Entire Grid › New study identifies vulnerabilities to EMP attack

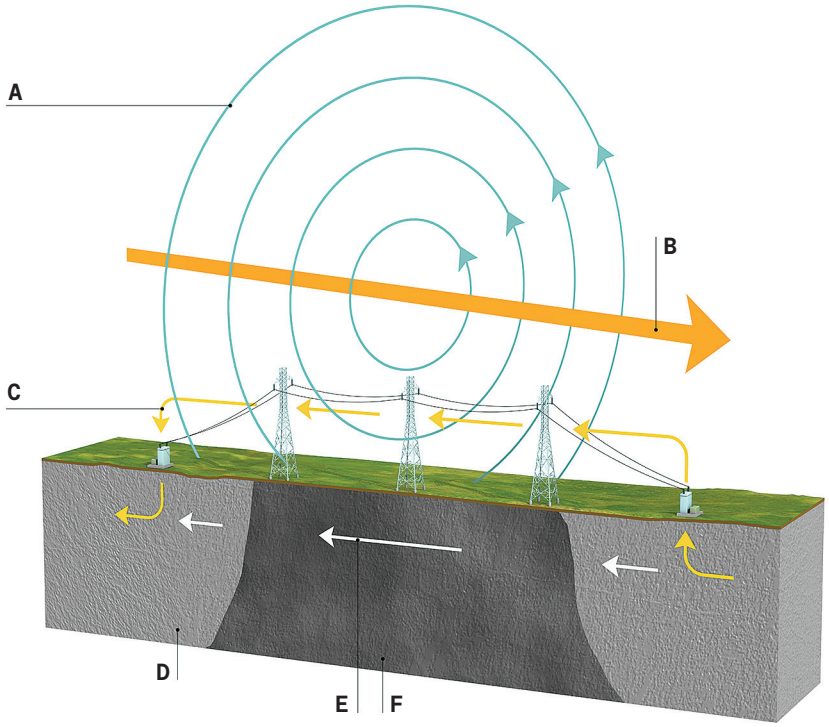
BY NATASHA BAJEMA

In July, Chinese researchers urged their government to increase the country's readiness for defending against a high-altitude electromagnetic pulse (EMP) attack. Just over a year ago, a group of American researchers released a report warning that China possessed the capability to conduct an EMP attack against the United States. Military and nonproliferation experts are worried about the growing temptation by nuclear-armed countries to engage in a first-strike EMP attack using nuclear weapons that, while avoiding direct casualties, could prove devastating to electric grids and electronic devices from smartwatches to supercomputers.

The enormous potential of an electromagnetic pulse released by the high-altitude detonation of a nuclear weapon has been recognized for some time. In 1962, the U.S. conducted an atmospheric test of a 1.45-megaton thermonuclear weapon, code-named Starfish Prime, 400 kilometers above Johnston Island in the Pacific Ocean. Over 1,600 kilometers away, the blast knocked out electricity supply in parts of Hawaii and disrupted telephone service for a period of time. In addition, radiation from the test damaged several satellites in low Earth orbit, taking them out of service. Decades later, the Commission to Assess the Threat to the United States From Electromagnetic Pulse Attack determined as early as 2008 that the United States would face catastrophic consequences from an EMP attack given its growing dependence on electronics of all forms and complete reliance on the electrical grid.

And yet, until now, government and industry risk assessments about EMP attacks and their effects on the power grid have been based on oversimplified models of the solid Earth that assume zero variation in depth or composition. But, as it turns out, the actual effects on the power grid of an electromagnetic pulse in outer space are strongly determined by the three-dimensional distribution of rocks beneath our feet.

Ground-breaking research, the product of collaboration between the U.S. Geological Survey (USGS) and the University of Colorado, illuminates the role of the solid Earth in determining the magnitude of any EMP hazard. In an interview,



A nuclear explosion in the upper atmosphere or in space [not shown] would push a current of ions and electrons through the atmosphere [B], producing a magnetic field [A] that would in turn induce currents and electric fields in the earth beneath it. Regions of high conductivity [D] would more readily carry that current, while regions of low conductivity [F], would not. Instead, the larger electric field in these regions [E] could help coax current out of the ground and through the wires of high-capacity power lines [C]—yielding a potentially grid-crippling power surge induced by the explosion's electromagnetic pulse.

Illustration by John MacNeill

USGS geophysicist Jeffrey J. Love, the lead author on the new report, explains that a high-altitude EMP produces three sequential waveforms with different impacts on electrical systems: E1, E2, and E3.

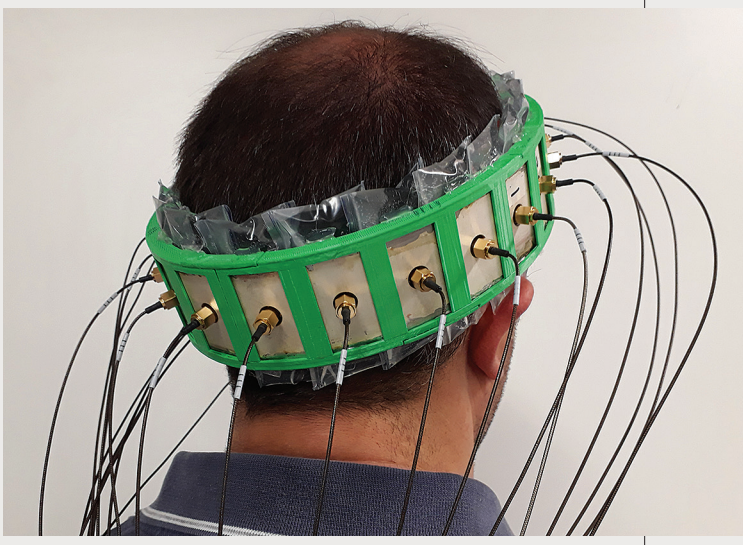
The high-frequency E1 pulse disrupts consumer electronics and tends to get the most attention; E2 behaves more or less like lightning, and, fortunately, our electrical systems are (largely) hardened against its effects. The E3 waveform is the lowest-amplitude part of the EMP signal, but because it is the longest-lasting part, covering periods from about 0.1 seconds to several hundred seconds, it has the potential to cause catastrophic damage to the electrical grid through its interactions with the solid earth. [See illustration.]

The three factors, Love says, that conspire to form a geoelectric hazard for power grids are, "the level of EMP magnetic disturbance, the conductivity of the surrounding earth, and the specific parameters of the grid itself." The team's new study used existing survey data—originally collected for geological exploration—in a small region of the eastern-midcontinental United States, covering portions of Arkansas, Illinois, Kentucky, Mississippi, Missouri, and Tennessee. Researchers from the USGS then secured permission from property owners to place sensors on the ground for measuring the natural variation in the local magnetic field over several weeks. They also used voltmeters to measure the time-varying electric field at the same locations. Together, these two measurements provide estimates of surface impedance—an electromagnetic property that depends on rock conductivity.

Love and his coauthors used these survey data to assess the impact of an E3 EMP waveform generated by a high-altitude detonation of a nuclear weapon with a yield of several hundred kilotons—the benchmark description of an EMP event in the literature. They also included USGS research on natural, magnetic-storm disturbances across the continental United States—in regions with electrically *resistive* metamorphic and igneous rocks, such as northern Minnesota and the area east of the Appalachian Mountains, as well as those with electrically *conductive* sedimentary rocks, such as Illinois and Michigan.

All in all, they found that EMP hazards had not been accurately mapped in complex geological settings. They call for the USGS to analyze surface impedance across regions like the eastern midcontinent. They add that the USGS should pay special attention to the eastern United States, where magnetic-storm hazards are already known to be high. Of course this is also where many of the nation's largest cities are.

"Better coordination across scientific disciplines is necessary,' Love says. "By bringing together weapons engineers, space scientists, and geophysicists, we can achieve a holistic approach to EMP threat assessment and, with that, prioritize improvements." ■



A new smart helmet uses electromagnetic waves to estimate the size and position of a stroke inside a patient's brain.

FOS S.P.A./UNIVERSITY OF GENOA

JOURNAL WATCH

Smart Helmet Provides Early Stroke Diagnosis

When someone experiences a stroke, every passing moment leading up to treatment is critical. Ideally, patients should be diagnosed and treated within the first hour, often referred to as "the golden hour," in order to have the best chance at recovery.

Electromagnetic (EM) measurements are particularly helpful in diagnosing a stroke, because the two types of stroke—ischemic and hemorrhagic—have different dielectric properties, which can be detected with electromagnetism. So a new device developed by Alessandro Fedeli and his colleagues at the University of Genoa, in Italy, can not only confirm the presence of stroke but can also determine what kind occurred. Fideli is an assistant professor in the Department of Electrical, Electronic, Telecommunications Engineering, and Naval Architecture.

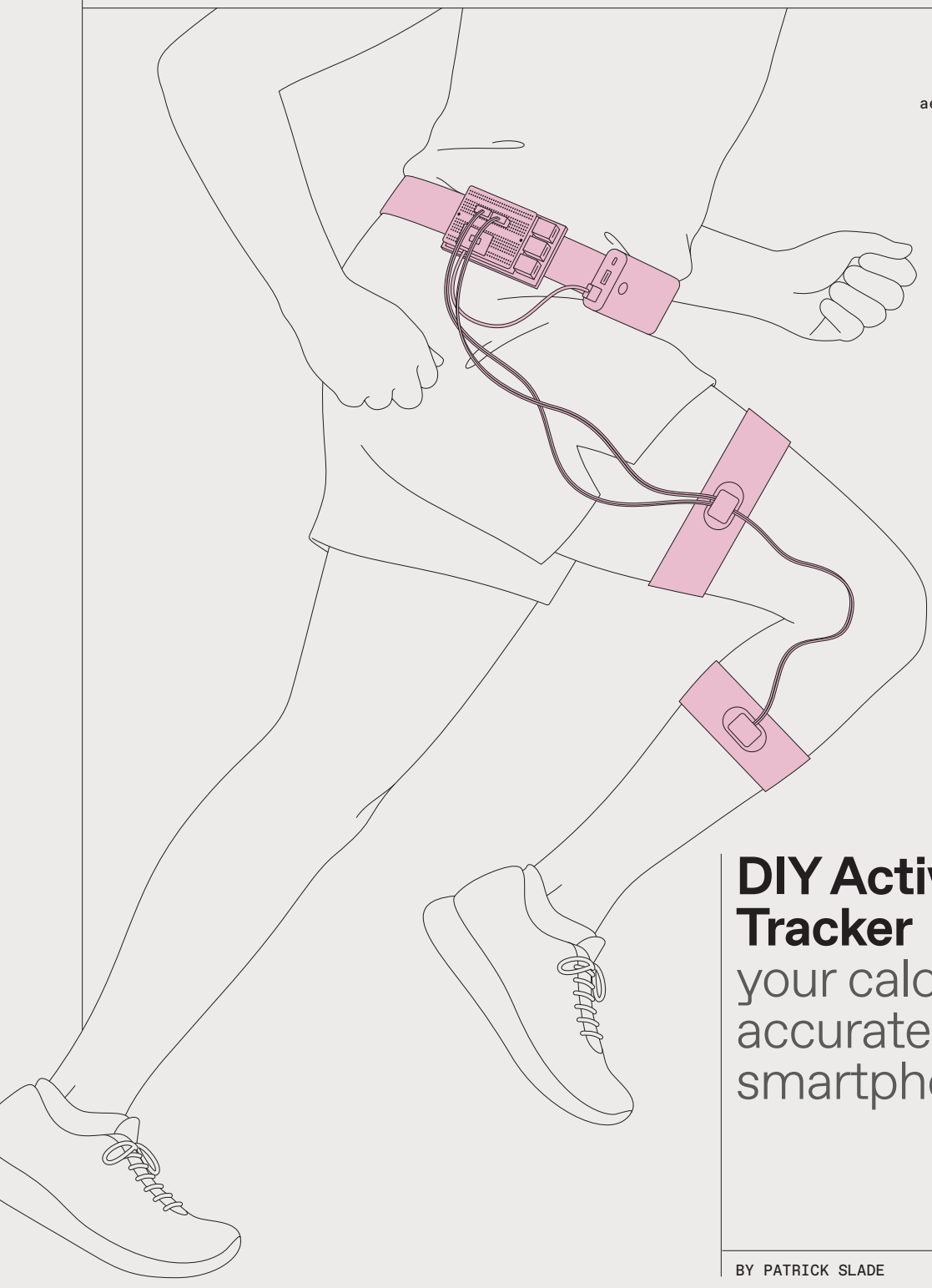
The "smart helmet" contains a series of antennas that are selectively activated to direct EM waves throughout the brain. A simple signal-processing algorithm alerts paramedics and other health professionals whether or not a stroke has occurred. A more complex algorithm, which requires more computational power, can then determine the type, size, and position of the stroke.

Through simulations described in a study published recently in *IEEE Wireless Communications*, the device was found effective at diagnosing various types of stroke.

"We hope to be able to perform clinical trials in the near future, possibly in cooperation with local hospitals," says Fedeli. "We also know that there are several other EM-based systems for stroke detection proposed by other research teams that have been already validated or are in [the] course of validation with clinical trials, with positive and promising results." —Michelle Hampson

Hands On

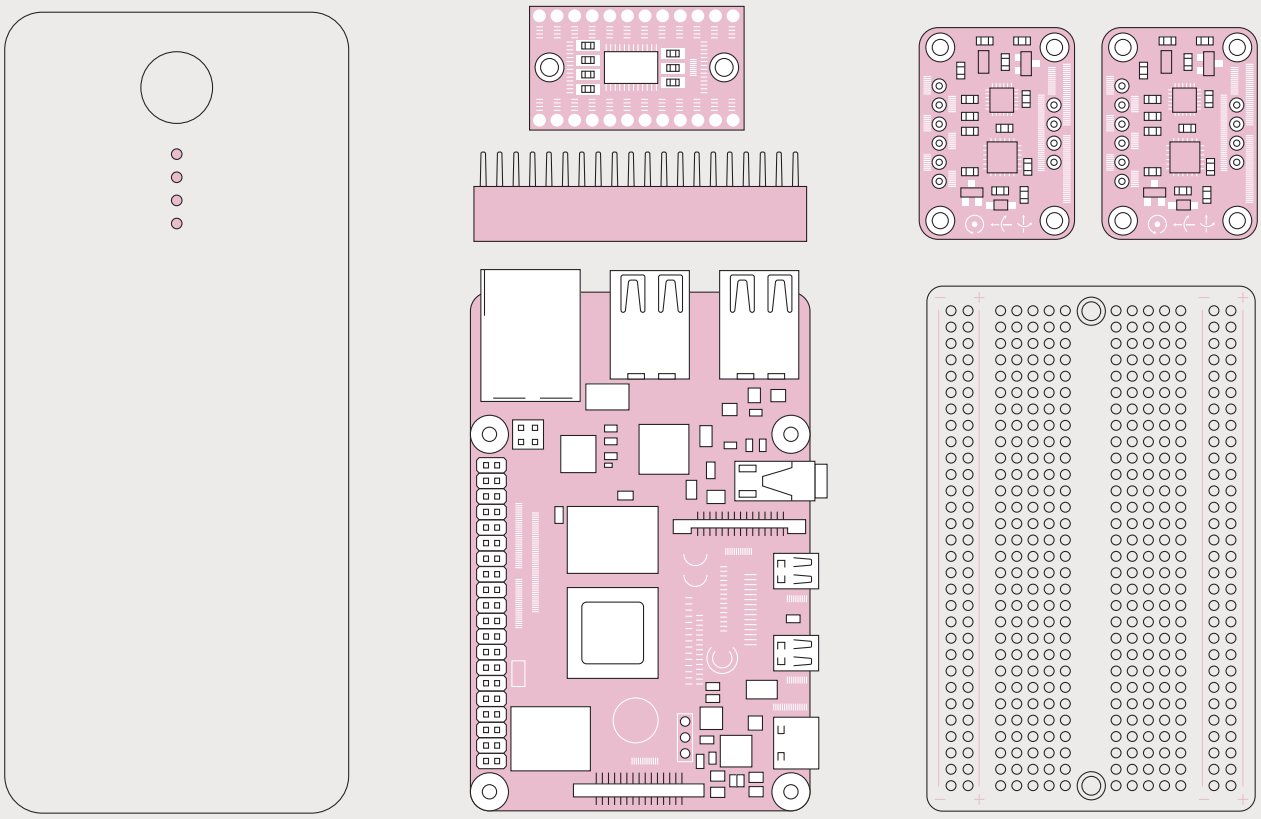
Inexpensive components will let you monitor your aerobic workout much better than a smartphone would.



DIY Activity Tracker › Count your calories more accurately than a smartphone

BY PATRICK SLADE

Illustrations by James Provost



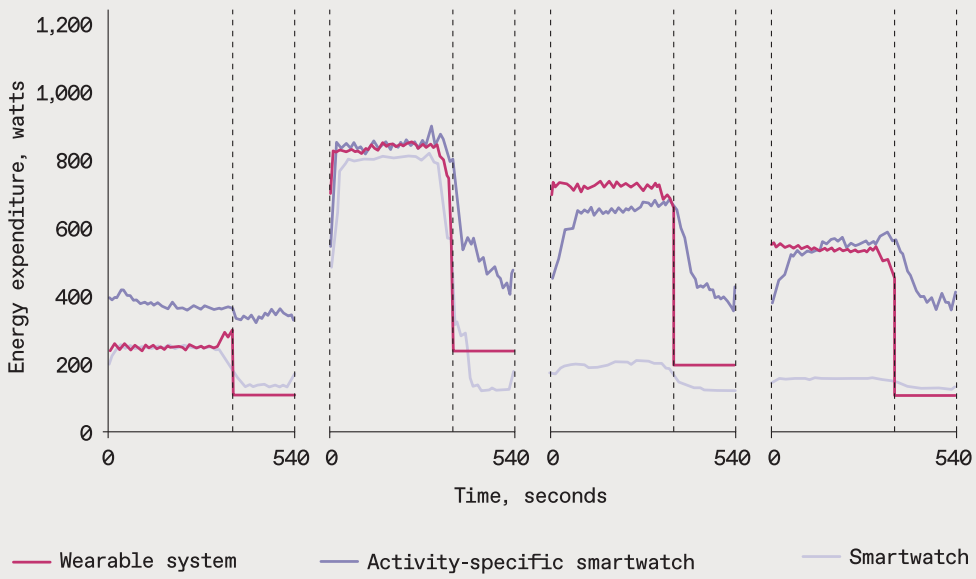
Two small inertial measurement units [top right] are strapped to the user's thigh and shank. Using a I2C expansion board [top middle], the IMUs feed data into a Raspberry Pi [bottom middle] powered by a USB battery pack [left].

Physical activity is essential to both physical and mental health, something brought home to many people following sedentary pandemic lockdowns. Even without the lockdowns, many parts of the world have been facing an obesity epidemic, which has created a need to help people manage their weight. For such people there are a wealth of fitness and diet apps that rely on smartphone and smartwatch sensors to monitor activity levels and track the calories they have burned. The problem is that smartphones and smartwatches do a terrible job at calorie counting.

A 2017 study looked at seven such typical devices, and found their counts were off between 27 and 93 percent, depending on the device. We decided to see if a better calorie counter could be made, at least for some forms of activity. The answer is yes—and it's one you can build yourself with parts any maker can easily obtain.

The road to the calorie counter began in our lab (the Human Performance Laboratory in the Stanford University School of Engineering), where we study things like the metabolic cost of walking. We broke out every possible sensor we had in the lab and attached them to participants. This included sensors that monitor muscle activity, inertial measurement units (IMUs) to monitor movement on different parts of the body, and instrumented insoles in shoes to monitor forces produced by walking and running. We used respirometry, a lab-based method of measuring energy expenditure by

About 20 to 40 percent of the steps we take each day occur in bouts of walking that are 20 seconds or less.



The red line shows how quickly the wearable system responds to the user starting and stopping different aerobic activities (from left to right: walking, running, stair climbing, bicycling). The stop is indicated by the dotted line in the middle of each activity.

monitoring the oxygen intake and carbon dioxide expelled with each breath, so as to get a ground-truth measure of the calories burned as participants moved. With all of this data we looked at building a fundamental relationship between the movement of the body (and thus the activity of the muscles burning calories) and the actual energy expended by the whole body.

We found that by looking at the motion of the thigh and the shank (lower leg) we could estimate caloric expenditure during aerobic activities with an accuracy of about 13 percent. (For the full details of our analysis, see our recent paper in *Nature Communications*.) What's more, we could do it using inexpensive IMUs.

We used Adafruit Precision NXP 9-DOF breakout board IMUs. This board combines two sensor chips, a six-degrees-of-freedom accelerometer/magnetometer, and a three-degrees-of-freedom gyroscope. The counter uses two of them, one attached at midthigh, the other at midshank. In our tests, we sometimes used toupee adhesive to hold various IMUs in place, but a Velcro strap works great too!

The IMUs are connected to a Raspberry Pi using the I2C protocol. Although the Pi is bigger and has a higher power draw than, say, a Teensy board, we chose it because it was easy to stream data wirelessly from the Pi and monitor it during testing and calibration.

The Pi, running a stock version of the standard operating system, also allowed us to use established Python libraries to do onboard data processing. We used the NumPy scientific computing library for storing data in a convenient format, and the Scikit-learn machine-learning library for analyzing motion data from the IMUs.

We trained a linear regression model to perform gait detection to segment each step and make an estimate of the calories burned. (Our Python scripts are available for download from a public repository.) We pull in sensor data at a fixed rate of 100 hertz. We determine when a step occurs by detecting when the leg stops rotating in one direction and starts rotating in the other direction, occurring around the time the heel touches the ground. Then we pass the leg motion from that step into our model that estimates energy expenditure.

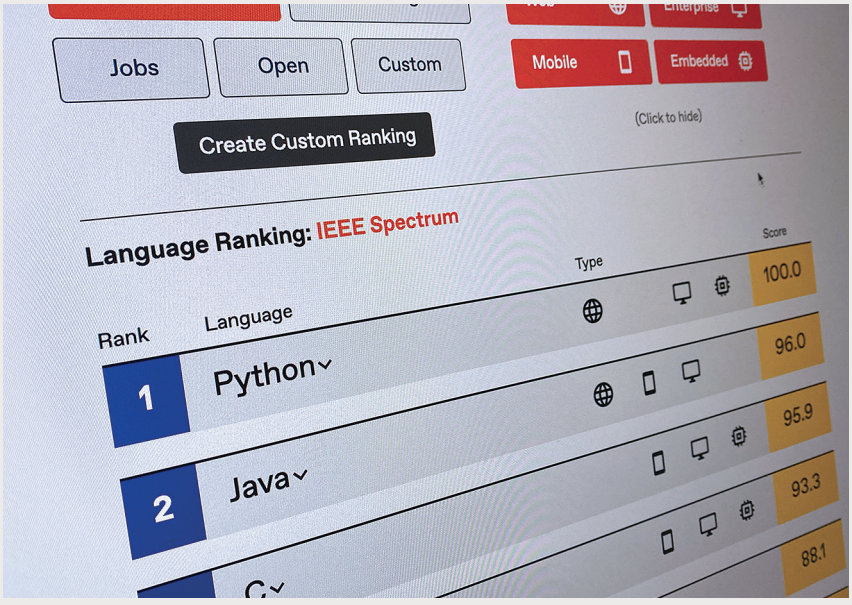
The high accuracy in calorie counting comes from providing a per-step estimate of energy expenditure. Within each step we look at the motion and interpolate the energy expended, regardless of activity: Our system doesn't need wearers to inform it if they're running versus biking versus climbing stairs, for example, and the system is sensitive enough to detect a single step and the moment when a walk turns into a run.

This kind of direct analysis provides a much better way of tracking instantaneous energy expenditure over other measures such as heart rate or respiratory rate. These latter indicators take up to a minute to reflect changes in activity, so if you got up from the couch and walked 10 steps, they wouldn't detect that movement. About 20 to 40 percent of the steps we take each day occur in bouts of walking that are 20 seconds or less, which contributes significantly to the total energy expended every day.

We are currently working to create a more compact version of our calorie counter. We're interested in tracking energy expenditure over longer periods of time to hopefully improve weight management and sports training. What happens if you wear this device for a week or two at a time? We are also thinking about how we can integrate tracking upper-body activity by, say, looking at the motion of a paired smartwatch.

We plan to validate the data using the gold standard technique of so-called doubly labeled water, which lets us track energy expenditures on this kind of timescale very accurately. A subject drinks water containing deuterium and oxygen-18. Tracking how these are eliminated from the body in urine gives us the rate of carbon dioxide production, which is directly related to energy expenditure. In the even longer term, we hope to replace the current IMU boards with technologies being developed for flexible electronics that can be pressed on like a sticker or even possibly for direct printing on the skin. ■

At Work



Top Programming Languages

› Our eighth annual probe into what's hot and not

BY STEPHEN CASS

Learn Python. That's the biggest takeaway we can give you from its continued dominance of *IEEE Spectrum's* annual online interactive rankings of the top programming languages. You don't have to become a dyed-in-the-wool Pythonista, but learning the language well enough to use one of the vast number of libraries written for it is probably worth your time.

Once you've got the basics of Python down, it's all about the ins and outs of particular libraries for things like embedded projects and large-scale AI systems. Frankly, depending on the domain, complexity, and/or quality of documentation, grokking a library can be considerably tougher than learning Python itself.

But Python has its limits, as the continued popularity of languages better suited to solving particular problems, such as R, SQL, and Matlab, shows. C, C++, Java, and JavaScript also continue to dominate at the top of the rankings, both on their own merits and because of the huge existing base of code written in them. (Indeed, significant parts of Python itself and its libraries are written in C for performance reasons.) And while many a high-level language has come and gone, there'll always be a place for those willing to write as close to the metal as possible in some flavor of assembly code.

It's because one size doesn't fit all that our rankings are interactive.

It's precisely because one size doesn't fit all that our rankings are interactive. Want to just see languages that are used for embedded development? Those most in demand by employers? Use one of our filters or presets, or adjust the weights of the individual metrics as you like. Application domains, such as Web or mobile, that you can filter on are based on typical usage, not outliers.

The default ranking is designed to reflect the interests of a typical IEEE member. The metrics are drawn from sources that we think are good proxies for gauging the popularity of languages, since it's impossible to know exactly what everyone is doing at their keyboards. Some were queried through publicly available interfaces, such as Stack Overflow or Google. Other metrics are drawn from private sources, such as the IEEE's Xplore article database, or the data on what languages are in demand by employers, which comes from the IEEE Job Site and courtesy of CareerBuilder.

Some of the metrics reflect the peculiarities of a peculiar time: For example, with our Twitter metric, Cobol dropped from seventh place to 34th place. But this is because Cobol was briefly a hot topic on Twitter in 2020, following the pleas from government officials who needed to update legacy systems in the face of the COVID-19 pandemic. (Dealing with this kind of noise is the reason we combine multiple metrics.)

Other movers in the *Spectrum* default rankings include Microsoft's C#, which has risen from 25th place last year to sixth this year. This most likely reflects that version 9.0 of the language was released toward the end of 2020, the upcoming launch of Windows 11, and continued growing general interest in distributed systems, which C# is designed to enable.

So find the ranking that suits your needs, and let us know if there are any new languages we should include in next year's edition. ■

RANDI KLETT

Crosstalk

What Goes Up...

World population growth has slowed, and some countries are actually declining

In 1960, *Science* published a paper by Heinz von Foerster predicting that on Friday, 13 November 2026, the "human population will approach infinity if it grows as it has grown in the last two millennia." Just a few years after this preposterous doomsday alarm, the annual growth of global population peaked at about 2.1 percent and immediately began to decline. By 2020 the growth rate stood at just a bit more than 1 percent, the result of the steadily declining total fertility rate (TFR), the number of children born to a woman during her reproductive period.

In preindustrial societies this rate stood commonly at 5 or higher; during the United States' baby-boom years (1945–1964) its rate peaked at about 3.2. The replacement rate in developed countries is roughly 2.1 children per woman. Some affluent nations have had below-replacement TFRs for several decades (Germany since 1970, Italy since 1976), but this fertility retreat has now deepened to such an extent that substantial population declines by 2050 are now inescapable in at least a quarter of the world's nations.

As long as the total fertility rate remains just below the replacement rate, its rebound is quite likely. But when the TFR falls very far it means that an increasing share of families are having just one child or none at all, and that makes it much harder to lift fertility through pronatalist policies, such as paying people to have additional children. TFRs below 1.5 lead to demographically uncharted territory. This group of countries now includes many states in Central and Eastern Europe and also such populous countries as Japan, Germany, Italy and South Korea.

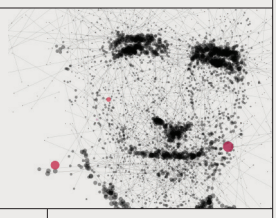
Near-term demographic forecasts are far from perfect, but there is no danger of making very large errors, say, of 50 percent. That's because so many future mothers are already with us, and because TFRs do not quickly double. The latest U.N. population projections for 2050 (released in 2019) show continued global growth, mainly because African TFRs are still mostly above 3. But the medium-growth forecast sees slight declines both in Europe (–5 percent) and in China (–2.5 percent), while the low-growth forecast sees declines of 26 percent in Ukraine, 16 percent in Italy, 15 percent in Russia, 13 percent in Spain, and nearly 9 percent in China.

Shrinking population together with a higher average age erodes the tax base, raises infrastructure costs, and leads to social isolation, as settlements dwindle and die.

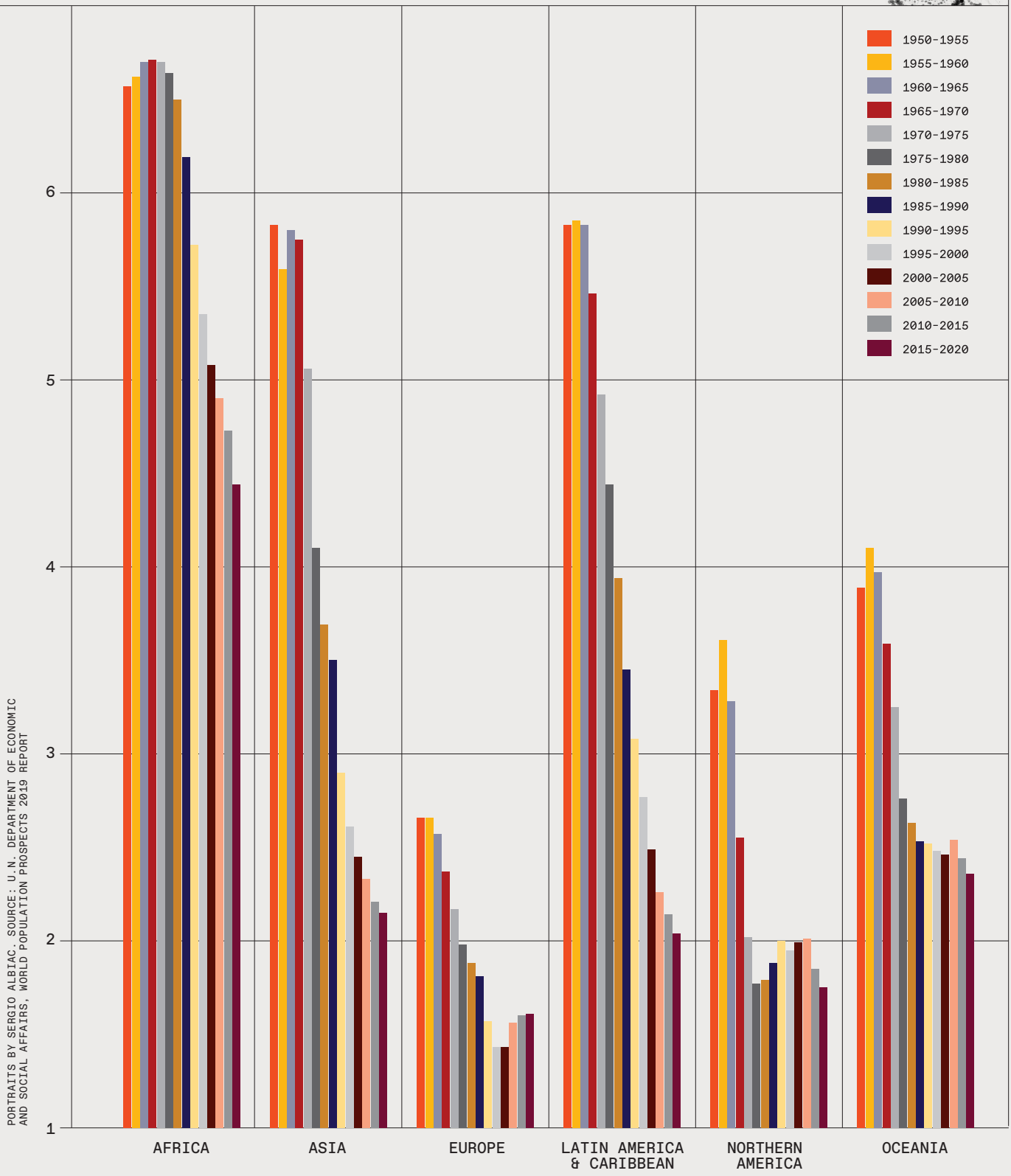
The decline has been underway for some time in villages and small towns, where the sequence is much the same everywhere: First they lose their school, then the post office, gas station, and grocery store. Finally, a settlement is administratively amalgamated with its similarly fated neighbors. You can see what is left behind without leaving your room by taking Google Street View tours of desolate mountain villages in Tohoku, the northern (and the poorest) part of Japan's largest island, where almost every third person is now over 65 years old. Or look at the forlorn places not far from Bucharest, Romania's capital, where all but a few young people have left for Western Europe and the TFR is below 1.4.

This process can be found even in certain parts of countries that are still growing, thanks to immigration. The United States is losing people across much of the Great Plains, Germany throughout most of the former German Democratic Republic, Spain in Castile and Léon and in Galicia. Shrinking population together with a higher average age erodes the tax base, raises infrastructure costs, and leads to social isolation, as settlements dwindle and die. It is all very depressing to contemplate.

Of course, in a truly long-range perspective this is hardly surprising. Ten thousand years ago there were perhaps just 5 million people on Earth—too few, it would have seemed, to become the dominant species. Now we are closing in on 8 billion, and the total may peak at more than 10 billion. We may start losing that global primacy sooner than we think, leaving more room for bacteria, birds, and bears. ■



ESTIMATED TOTAL FERTILITY (LIVE BIRTHS PER WOMAN) BY REGION, 1950-2020



PORTRAITS BY SERGIO ALBIAC. SOURCE: U.N. DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS, WORLD POPULATION PROSPECTS 2019 REPORT

The Indoor Air-Quality Paradox

Easy to Measure, Tough to Fix

The summer of 2020 brought wildfire to Portland, Ore., as it did to so many other cities across the world. All outdoor activity in my neighborhood ceased for weeks, yet staying indoors didn't guarantee relief. The worst days left me woozy as my lone air purifier, whirring like a jet engine, failed to keep up.

The air in my home was bad. But I had no idea of how bad because I had no way to measure it.

That's changing, thanks to indoor air-quality monitors like Airthings' View Plus. Sold for US $299, the View Plus can gauge seven critical metrics: radon, particulates, carbon dioxide, humidity, temperature, volatile organic compounds, and air pressure.

The monitor proved useful. I learned that cooking dinner can spike particulates into unhealthy territory for several hours, a sign that my oven vent is not working properly. The monitor also reported low levels of radon, proof that my home's radon-mitigation system is doing its job.

I had the monitor installed, working, and connected to the Airthings app less than 10 minutes after it arrived at my doorstep, in June. Reading the app was easy: It color-coded the results as good, fair, or poor. I have only one monitor, but the system can support multiple devices, making it possible to sniff out how air quality differs between rooms. You can also just move the device, though it needs time to update its readings.

Airthings' monitor is unusual because it combines a radon sensor with other air-quality metrics, but it's certainly not alone. Alternatives are available from IQAir, Kaiterra, and Temtop, among others, and they range in price from $80 to $300. These monitors don't require permanent installation, so they're suitable for renters as well as owners.

But it's not enough to detect air pollutants; you must also remove them. That problem is harder.

Air purifiers surged in popularity through the second half of 2020 in response to dual airborne threats of COVID-19 and wildfire smoke. Companies responded to this demand at 2021's all-digital Consumer Electronics Show. LG led its presentation with personal air purifiers instead of televisions. Coway, Luft, and Scosche all showed new models, with Coway winning a CES Innovation Award for its new Design Flex purifiers.

Ionization can itself create ozone. The state of California has banned such ozone generators entirely.

Unfortunately, consumers newly educated on indoor air quality will be puzzled about which air purifier, if any, is appropriate. Purifiers vary widely in the pollutants they claim to clean and how they claim to clean them. Most models advertise a HEPA air filter, which promises a specific standard of efficiency based on its rating, but this is often combined with unproven UV light, ionization, and ozone technologies that vaguely claim to catch toxins and kill pathogens, even COVID-19. This is the wild, wild west of air purification.

It's true that an activated carbon filter can remove volatile organic compounds and ozone from the air. There's no common standard for efficiency, however, so shoppers must cross their fingers and hope for the best. Ionization is no better. Studies suggest ionization can destroy viruses and bacteria in the air but, again, there's no common standard.

In fact, ionization can itself create ozone. The state of California has banned such ozone generators entirely, but you'll still find these products on Amazon and other retailers. Studies even suggest the ionization feature in some purifiers may interact with the air in unpredictable ways, adding new pollutants.

It's vital that companies designing air purifiers police their products and work together on standards that make sense to consumers. 2021's harsh fire season will keep demand high, but new, easy-to-use monitors like the Airthings View Plus will leave homeowners better informed about air quality—and ready to kick unproven purifiers to the curb. ■

Illustration by Harry Campbell

IEEE Foundation

IEEE

We have 30 million reasons to be proud.

Thanks to our donors, supporters and volunteers who answered the call of the ***Realize the Full Potential of IEEE*** **Campaign**, helping impact lives around the world through the power of technology and education.

Illuminate

Educate

Engage

Energize

Realize Your Impact

Learn how: ieeefoundation.org/campaign

Bricked by Age

Manufacturers should supply the software to keep things working indefinitely

I recently did some Marie Kondo–inspired housecleaning: Anything that didn't bring me joy got binned. In the process, I unearthed some old gadgets that made me smile. One was my venerable Nokia N95, a proto-smartphone, the first to sport GPS. Another was a craptastic Android tablet—a relic of an era when each year I would purchase the best tablet I could for less than $100 (Australian!), just to see how much you could get for that little. And there was my beloved Sony PlayStation Portable. While I rarely used it, I loved what the PSP represented: a high-powered handheld device, another forerunner of today's smartphone, though one designed for gaming rather than talking.

These nifty antiques shared a common problem: Although each booted up successfully, none of them really work anymore. In 2014, Nokia sold off its smartphone division to Microsoft in a fire sale; then Microsoft spiked the whole effort. These moves make my N95 an orphan product from a defunct division of a massive company. Without new firmware, it's essentially useless. My craptastic tablet and PSP similarly need a software refresh. Yet neither of them can log into or even locate the appropriate update servers.

Device makers are apt to drop support for old gadgets faster than the gadgets themselves wear out.

You might think that a 15-year-old gaming console wouldn't even be operating, but Sony's build quality is such that, with the exception of a very tired lithium-Ion battery, the unit is in perfect condition. It runs but can't connect to modern Wi-Fi without an update, which it can't access without an update to its firmware (a classic catch-22). I've wasted a few hours trying to work out how to get new firmware on it (and on the tablet), without success. Two perfectly good pieces of electronic gear have become useless, simply for want of software updates.

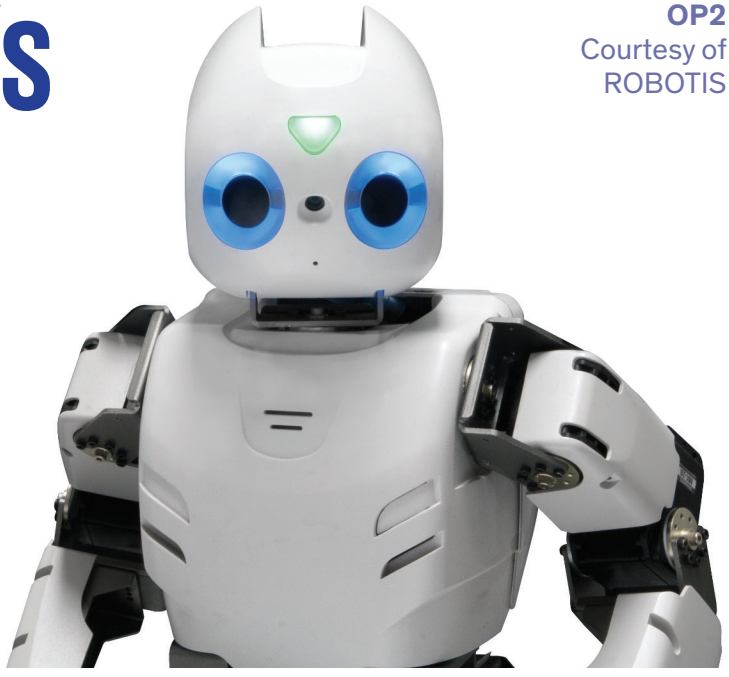
Consumers have relied on the good graces of device makers to keep our gadget firmware and software secure and up-to-date. Doing so costs the manufacturer some of its profits. As a result, many of them are apt to drop support for old gadgets faster than the gadgets themselves wear out. This corporate stinginess consigns far too many of our devices to the trash heap before they have exhausted their usability. That's bad for consumers and bad for the planet. It needs to stop.

We have seen a global right-to-repair movement emerge from maker communities and start to influence public policy around such things as the availability of spare parts. I'd argue that there should be a parallel *right-to-maintain* movement. We should mandate that device manufacturers set aside a portion of the purchase price of a gadget to support ongoing software maintenance, forcing them to budget for a future they'd rather ignore. Or maybe they aren't ignoring the future so much as trying to manage it by speeding up product obsolescence, because it typically sparks another purchase.

Does this mean Sony and others should still be supporting products nearly two decades old, like my PSP? If that keeps them out of the landfill, I'd say yes: The benefits easily outweigh the costs. The devilish details come in decisions about who should bear those costs. But even if they fell wholly on the purchaser, consumers would, I suspect, be willing to pay a few dollars more for a gadget if that meant reliable access to software for it—indefinitely. Yes, we all want shiny new toys—and we'll have plenty of them—but we shouldn't build that future atop the prematurely discarded remains of our electronic past. ■

Photo-illustration by Stuart Bradford

The World's Best ROBOTS GUIDE Is Here!

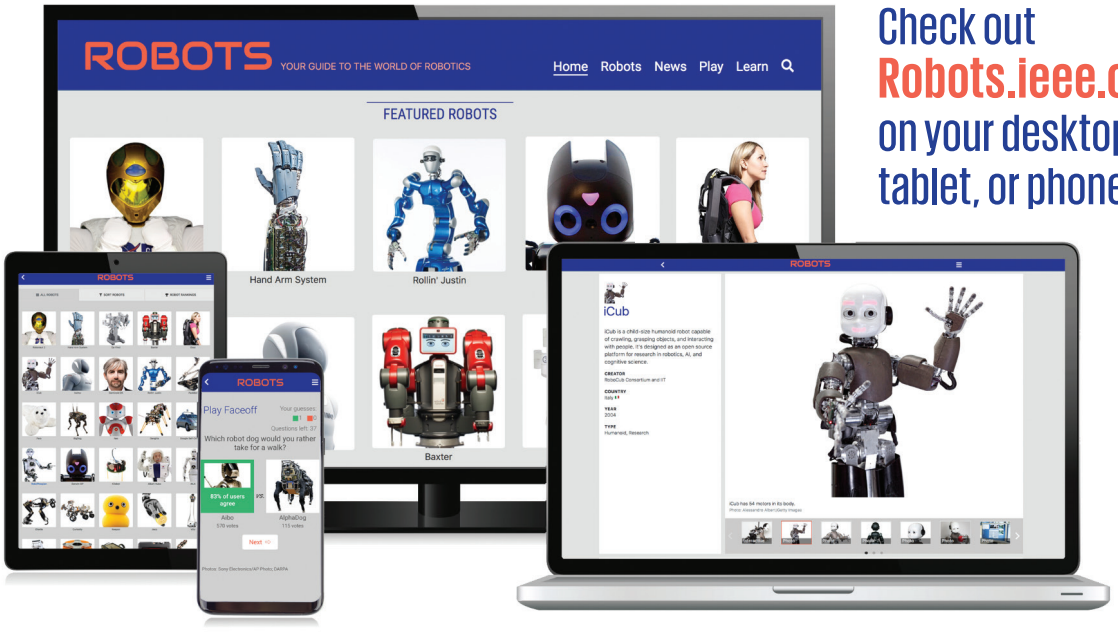


OP2
Courtesy of ROBOTIS

ROBOTS.IEEE.ORG

***IEEE Spectrum's* new ROBOTS site features more than 200 robots from around the world.**

- Spin, swipe and tap to make robots move.
- Read up-to-date robotics news.
- Rate robots and check their ranking.
- View photography, videos and technical specs.
- Play *Faceoff*, an interactive question game.



Check out Robots.ieee.org on your desktop, tablet, or phone now!

SPECIAL REPORT

The Great AI RECKONING

Deep learning may have reached its limits. What comes next?

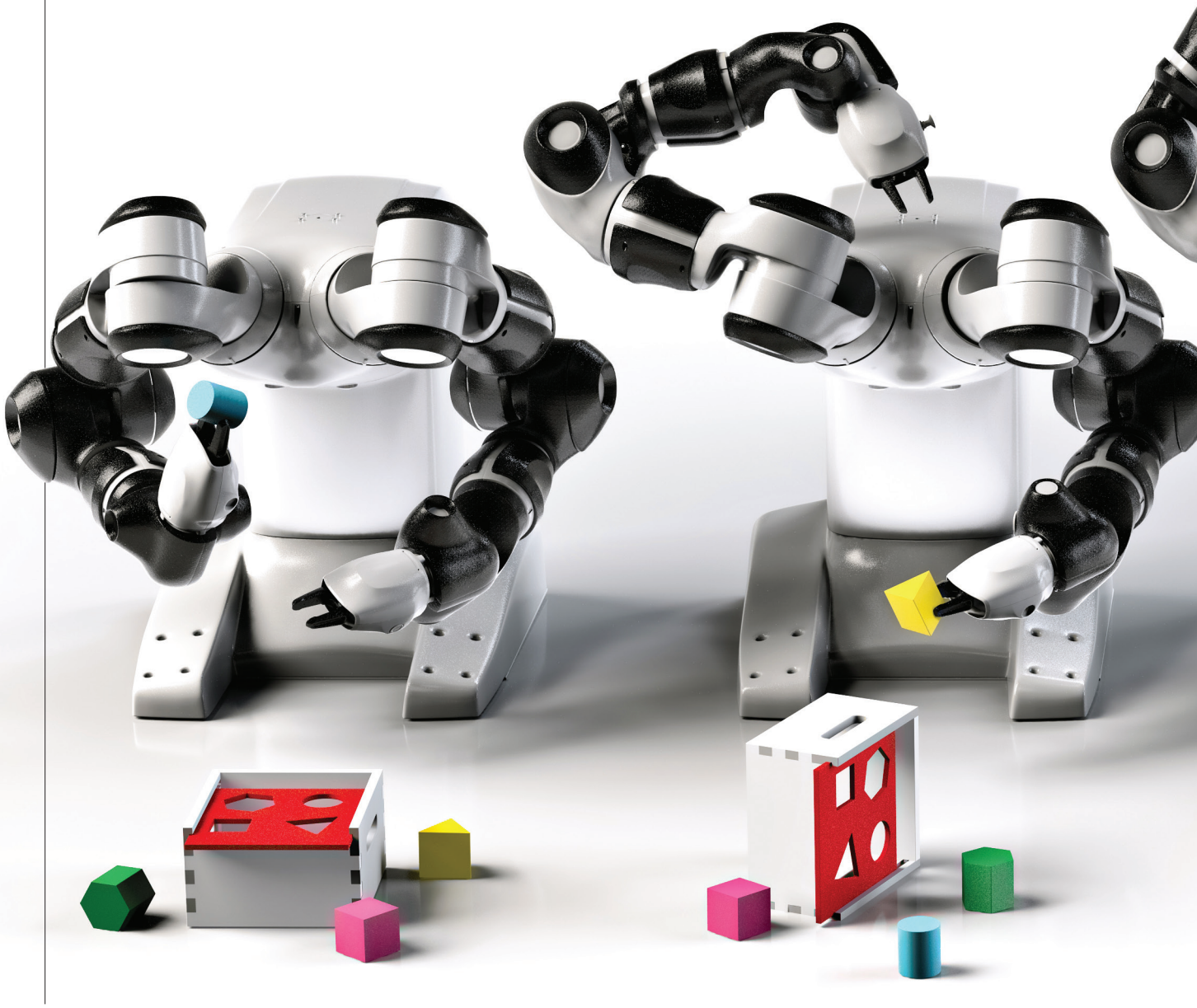
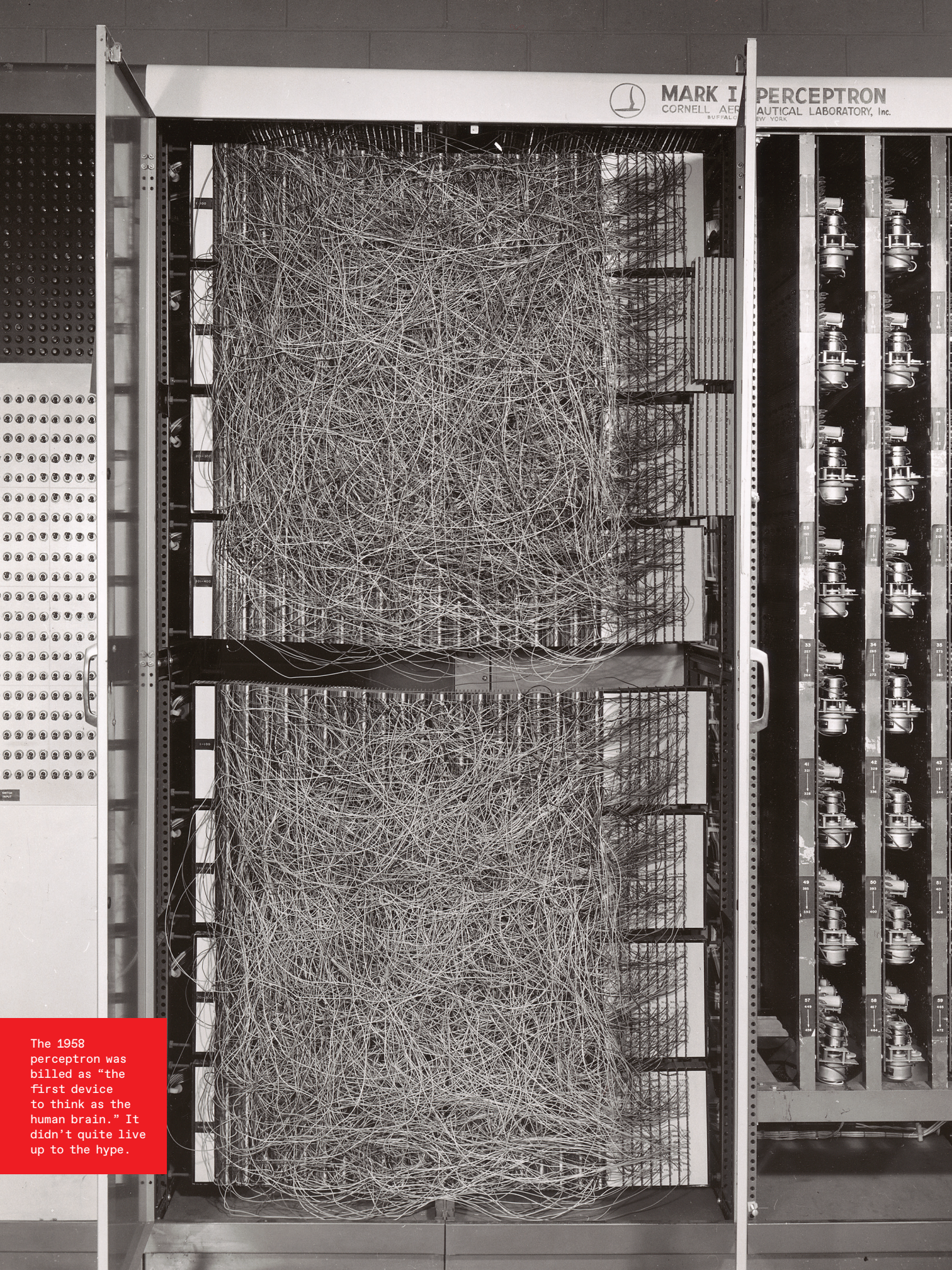


Illustration by Eddie Guy



The 1958 perceptron was billed as “the first device to think as the human brain.” It didn’t quite live up to the hype.

The TURBULENT PAST and UNCERTAIN FUTURE of AI

Is there a way out of AI's boom-and-bust cycle?

BY ELIZA STRICKLAND

IN THE SUMMER OF 1956, a group of mathematicians and computer scientists took over the top floor of the building that housed the math department of Dartmouth College. For about eight weeks, they imagined the possibilities of a new field of research. John McCarthy, then a young professor at Dartmouth, had coined the term "artificial intelligence" when he wrote his proposal for the workshop, which he said would explore the hypothesis that "every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." • The researchers at that legendary meeting sketched out, in broad strokes, AI as we know it today. It gave rise to the first camp of investigators: the "symbolists," whose expert systems reached a zenith in the 1980s. The years after the meeting also saw the emergence of the "connectionists," who toiled for decades on the artificial neural networks that took off only recently. These two approaches were long seen as mutually exclusive, and competition for funding among researchers

Frank Rosenblatt [above] invented the perceptron, the first artificial neural network.

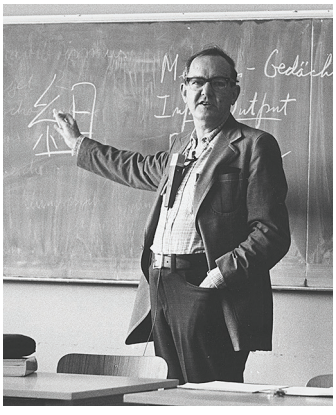
created animosity. Each side thought it was on the path to artificial general intelligence.

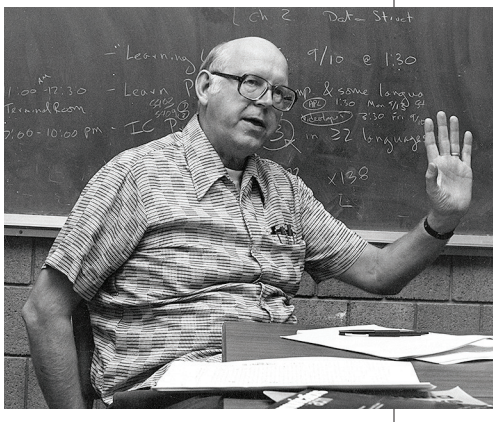
A look back at the decades since that meeting shows how often AI researchers' hopes have been crushed—and how little those setbacks have deterred them. Today, even as AI is revolutionizing industries and threatening to upend the global labor market, many experts are wondering if today's AI is reaching its limits. As Charles Q. Choi delineates in "Seven Revealing Ways AIs Fail" [p. 42], the weaknesses of today's deep-learning systems are becoming more and more apparent. Yet there's little sense of doom among researchers. Yes, it's possible that we're in for yet another AI winter in the not-so-distant future. Or this might just be the time when inspired engineers finally usher us into an eternal summer of the machine mind.

RESEARCHERS DEVELOPING SYMBOLIC AI set out to explicitly teach computers about the world. Their founding tenet held that knowledge can be represented by a set of rules, and computer programs can use logic to manipulate that knowledge. Leading symbolists Allen Newell and Herbert Simon argued that if a symbolic system had enough structured facts and premises, the aggregation would eventually produce broad intelligence.

The connectionists, on the other hand, inspired by biology, worked on "artificial neural networks" that would take in information and make sense of it themselves. The pioneering example was the perceptron, an experimental machine built by the Cornell psychologist Frank Rosenblatt with funding from the U.S. Navy. It had 400 light sensors that together acted as a retina, feeding information to

PREVIOUS PAGES AND THIS PAGE: CORNELL UNIVERSITY DIVISION OF RARE AND MANUSCRIPT COLLECTIONS





about 1,000 "neurons" that did the processing and produced a single output. In 1958, a *New York Times* article quoted Rosenblatt as saying that "the machine would be the first device to think as the human brain."

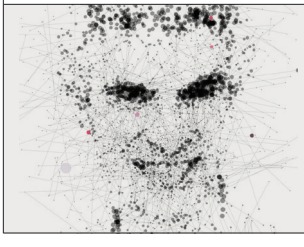
Unbridled optimism encouraged government agencies in the United States and United Kingdom to pour money into speculative research. In 1967, MIT professor Marvin Minsky wrote: "Within a generation...the problem of creating 'artificial intelligence' will be substantially solved." Yet soon thereafter, government funding started drying up, driven by a sense that AI research wasn't living up to its own hype. The 1970s saw the first AI winter.

True believers soldiered on, however. And by the early 1980s renewed enthusiasm brought a heyday for researchers in symbolic AI, who received acclaim and funding for "expert systems" that encoded the knowledge of a particular discipline, such as law or medicine. Investors hoped these systems would quickly find commercial applications. The most famous symbolic AI venture began in 1984, when the researcher Douglas Lenat began work on a project he named Cyc that aimed to encode common sense in a machine. To this day, Lenat and his team continue to add terms (facts and concepts) to Cyc's ontology and to explain the relationships between them via rules. By 2017, the team had 1.5 million terms and 24.5 million rules. Yet Cyc is still nowhere near achieving general intelligence.

The field of AI began at a 1956 workshop [above, left] attended by, from left, Oliver Selfridge, Nathaniel Rochester, Ray Solomonoff, Marvin Minsky, an unidentified person, workshop organizer John McCarthy, and Claude Shannon. Symbolists such as Herbert Simon [above, middle] and Allen Newell [above, right] wanted to teach AI rules about the world.

In the late 1980s, the cold winds of commerce brought on the second AI winter. The market for expert systems crashed because they required specialized hardware and couldn't compete with the cheaper desktop computers that were becoming common. By the 1990s, it was no longer academically fashionable to be working on either symbolic AI or neural networks, because both strategies seemed to have flopped.

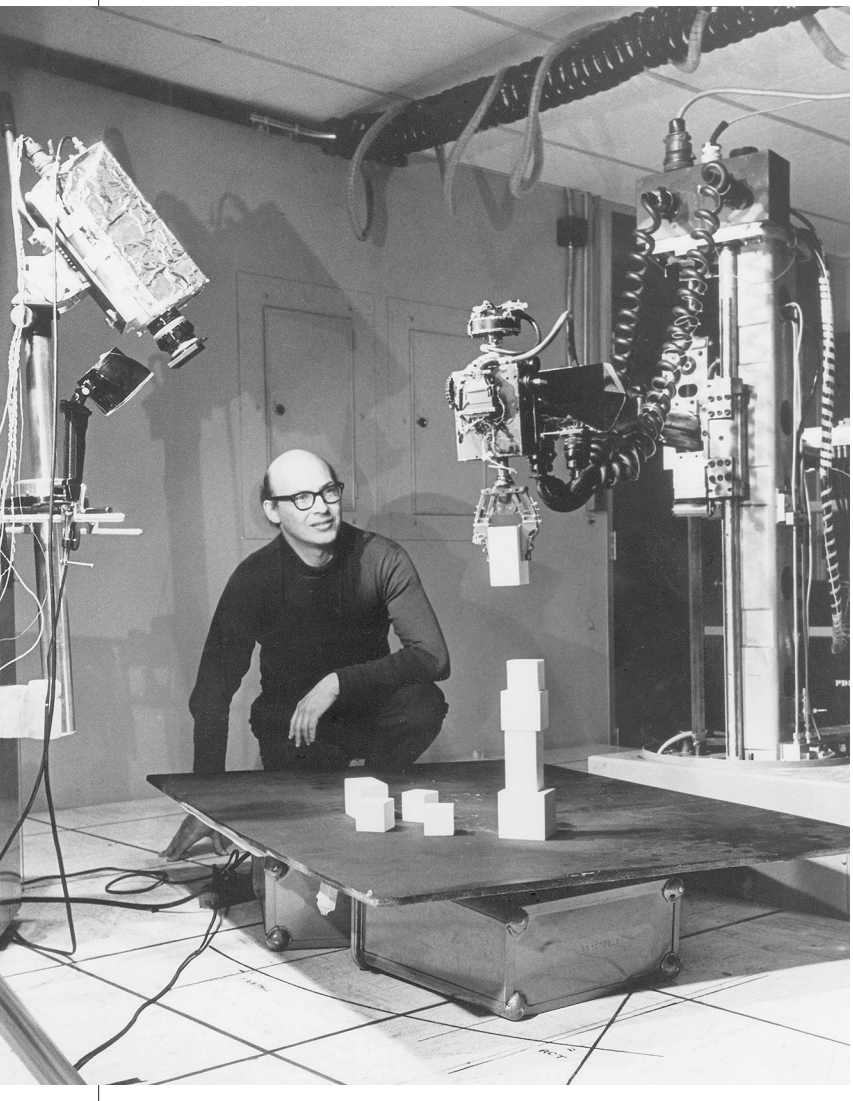
But the cheap computers that supplanted expert systems were a boon for the connectionists, who suddenly had access to enough computer power to run neural networks with many layers of artificial neurons. Such systems became known as deep neural networks, and the approach they enabled was called deep learning. Geoffrey Hinton, at the University of Toronto, applied a principle called back-propagation to make neural nets learn from their mistakes (see "How Deep Learning Works," p. 32). One of Hinton's postdocs, Yann LeCun, went on to AT&T Bell Laboratories in 1988, where he and a postdoc named Yoshua Bengio used neural nets for optical character recognition; U.S. banks soon adopted the technique for processing checks. Hinton, LeCun, and Bengio eventually won the 2019 Turing Award and are sometimes called the godfathers of deep learning.



"In terms of how much progress we've made in this work over the last two decades: I don't think we're anywhere close today to the level of intelligence of a 2-year-old child. But maybe we have algorithms that are equivalent to lower animals for perception."

YOSHUA BENGIO, founder and scientific director of Mila—Quebec AI Institute

FROM LEFT: THE MINSKY FAMILY; CARNEGIE MELLON UNIVERSITY (2). PORTRAIT BY SERGIO ALBIAC



MIT professor Marvin Minsky [above] predicted in 1967 that true artificial intelligence would be created within a generation.

But the neural-net advocates still had one big problem: They had a theoretical framework and growing computer power, but there wasn't enough digital data in the world to train their systems, at least not for most applications. Spring had not yet arrived.

OVER THE LAST TWO DECADES, everything has changed. In particular, the World Wide Web blossomed, and suddenly, there was data everywhere. Digital cameras and then smartphones filled the Internet with images, websites such as Wikipedia and Reddit were full of freely accessible digital text, and YouTube had plenty of videos. Finally, there was enough data to train neural networks for a wide range of applications.

The other big development came courtesy of the gaming industry. Companies such as Nvidia had developed chips called graphics processing units (GPUs) for the heavy processing required to render images in video games. Game developers used GPUs to do sophisticated kinds of shading and geometric transformations. Computer scientists in need of serious compute power realized that they could essentially trick a GPU into doing other tasks—such as training neural networks. Nvidia noticed the trend and created CUDA, a platform that enabled researchers to use GPUs for general-purpose processing. Among these researchers was a Ph.D. student in Hinton's lab named Alex Krizhevsky, who used CUDA to write the code for a neural network that blew everyone away in 2012.

He wrote it for the ImageNet competition, which challenged AI researchers to build computer-vision systems that could sort more than 1 million images into 1,000 categories of objects. While Krizhevsky's AlexNet wasn't the first neural net to be used for image recognition, its performance in the 2012 contest caught the world's attention. AlexNet's error rate was 15 percent, compared with the 26 percent error rate of the second-best entry. The neural net owed its runaway victory to GPU power and a "deep" structure of multiple layers containing 650,000 neurons in all. In the next year's ImageNet competition, almost everyone used neural networks. By 2017, many of the contenders' error rates had fallen to 5 percent, and the organizers ended the contest.

Deep learning took off. With the compute power of GPUs and plenty of digital data to train deep-learning systems, self-driving cars could navigate roads, voice assistants could recognize users' speech, and Web browsers could translate between dozens of languages. AIs also trounced human champions at several games that were previously thought to be unwinnable by machines, including the ancient board game Go and the video game *StarCraft II*. The current boom in AI has touched every industry, offering new ways to recognize patterns and make complex decisions.

But the widening array of triumphs in deep learning have relied on increasing the number of layers in neural nets and increasing the GPU time dedicated to training them. One analysis from the AI research company OpenAI showed that the amount of computational power required to train the biggest AI systems doubled every two years until 2012—and after that it doubled every 3.4 months. As Neil C. Thompson and his colleagues write in

THE MIT MUSEUM

A look back across the decades shows how often AI researchers' hopes have been crushed—and how little those setbacks have deterred them.

"Deep Learning's Diminishing Returns" [p. 50], many researchers worry that AI's computational needs are on an unsustainable trajectory. To avoid busting the planet's energy budget, researchers need to bust out of the established ways of constructing these systems.

WHILE IT MIGHT SEEM as though the neural-net camp has definitively tromped the symbolists, in truth the battle's outcome is not that simple. Take, for example, the robotic hand from OpenAI that made headlines for manipulating and solving a Rubik's cube. The robot used neural nets *and* symbolic AI. It's one of many new neuro-symbolic systems that use neural nets for perception and symbolic AI for reasoning, a hybrid approach that may offer gains in both efficiency and explainability.

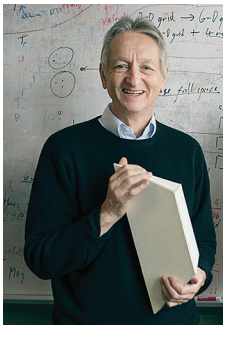
Although deep-learning systems tend to be black boxes that make inferences in opaque and mystifying ways, neuro-symbolic systems enable users to look under the hood and understand how the AI reached its conclusions. The U.S. Army is particularly wary of relying on black-box systems, as Evan Ackerman describes in "Deep Learning Goes to Boot Camp" [p. 56], so Army researchers are investigating a variety of hybrid approaches to drive their robots and autonomous vehicles.

Imagine if you could take one of the Army's road-clearing robots and ask it to make you a cup of coffee. That's a laughable proposition today, because deep-learning systems are built for narrow purposes and can't generalize their abilities from one task to another. What's more, learning a new task usually requires an AI to erase everything it knows about how to solve its prior task, a conundrum called catastrophic forgetting. At DeepMind, Google's London-based AI lab, the renowned roboticist Raia Hadsell is tackling this problem with a variety of sophisticated techniques. In "How to Train an All-Purpose Robot" [p. 34], Tom Chivers explains why this issue is so important for robots acting in the unpredictable real world. Other researchers are investigating new types of meta-learning in hopes of creating AI systems that learn how to learn and then apply that skill to any domain or task.

All these strategies may aid researchers' attempts to meet their loftiest goal: building AI with the kind of fluid intelligence that we watch our children develop. Toddlers don't need a massive amount of data to draw conclusions. They simply observe the world, create a mental model of how it works, take action, and use the results of their action to adjust that mental model. They iterate until they understand. This process is tremendously efficient and effective, and it's well beyond the capabilities of even the most advanced AI today.

Although the current level of enthusiasm has earned AI its own Gartner hype cycle, and although funding for AI has reached an all-time high, there's scant evidence there's a fizzle in our future. Companies around the world are adopting AI systems because they see immediate improvements to their bottom lines, and they'll never go back. It just remains to be seen whether researchers will find ways to adapt deep learning to make it more flexible and robust, or devise new approaches that haven't yet been dreamed of in the 65-year-old quest to make machines more like us. ■

Neither symbolic AI projects such as Cyc from Douglas Lenat [below, left] nor the deep-learning advances pioneered by [from left] Geoffrey Hinton, Yann LeCun, and Yoshua Bengio have yet produced human-level intelligence.



FROM LEFT: BOB E. DAEMMRICH/SYGMA/GETTY IMAGES; CHRISTOPHER WAHL/THE NEW YORK TIMES/REDUX; BRUNO LEVY/REA/REDUX; COLE BURSTON/BLOOMBERG/GETTY IMAGES

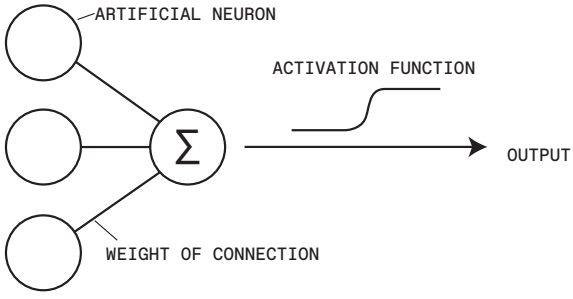
How Deep Learning WORKS

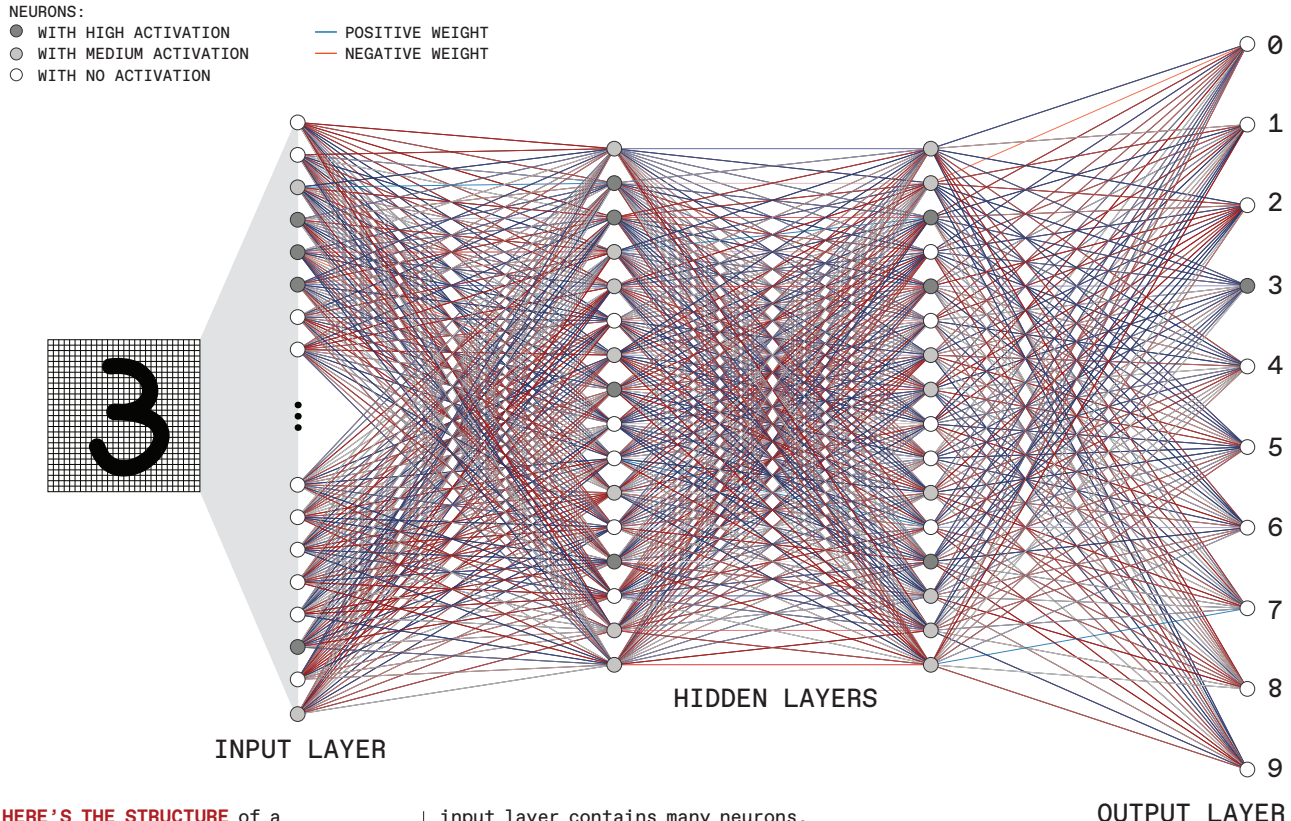
Inside the neural networks that power today's AI

BY SAMUEL K. MOORE, DAVID SCHNEIDER & ELIZA STRICKLAND

ARCHITECTURE

EACH NEURON IN AN ARTIFICIAL neural network sums its inputs and applies an activation function to determine its output. This architecture was inspired by what goes on in the brain, where neurons transmit signals between one another via synapses.





HERE'S THE STRUCTURE of a hypothetical feed-forward deep neural network ("deep" because it contains multiple hidden layers). This example shows a network that interprets images of hand-written digits and classifies them as one of the 10 possible numerals. • The input layer contains many neurons, each of which has an activation set to the gray-scale value of one pixel in the image. These input neurons are connected to neurons in the next layer, passing on their activation levels after they have been multiplied by a certain value, called a weight. Each neuron in the second layer sums its many inputs and applies an activation function to determine its output, which is fed forward in the same manner.

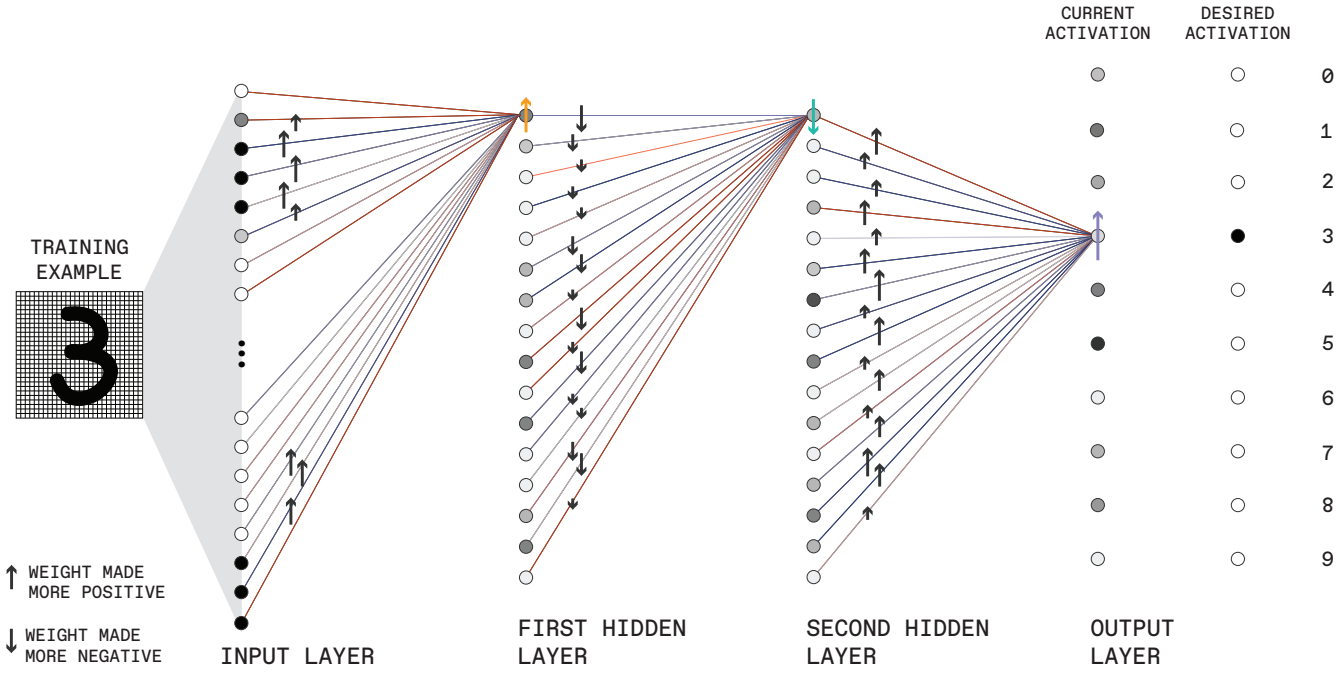
ILLUSTRATIONS: DAVID SCHNEIDER

TRAINING

THIS KIND OF NEURAL NETWORK is trained by calculating the difference between the actual output and the desired output. The mathematical optimization problem here has as many dimensions as there are adjustable parameters in the network—primarily the weights of the connections between neurons, which can be positive [blue lines] or negative [red lines].

Training the network is essentially finding a minimum of this multidimensional "loss" or "cost" function. It's done iteratively over many training runs, incrementally changing the network's state. In practice, that entails making many small adjustments to the network's weights based on the outputs that are computed for a random set of input examples, each time starting with the weights that control the output layer and moving backward through the network. (Only the connections to a single neuron in each layer are shown here, for simplicity.) This backpropagation process is repeated over many random sets of training examples until the loss function is minimized, and the network then provides the best results it can for any new input.

BACKPROPAGATION



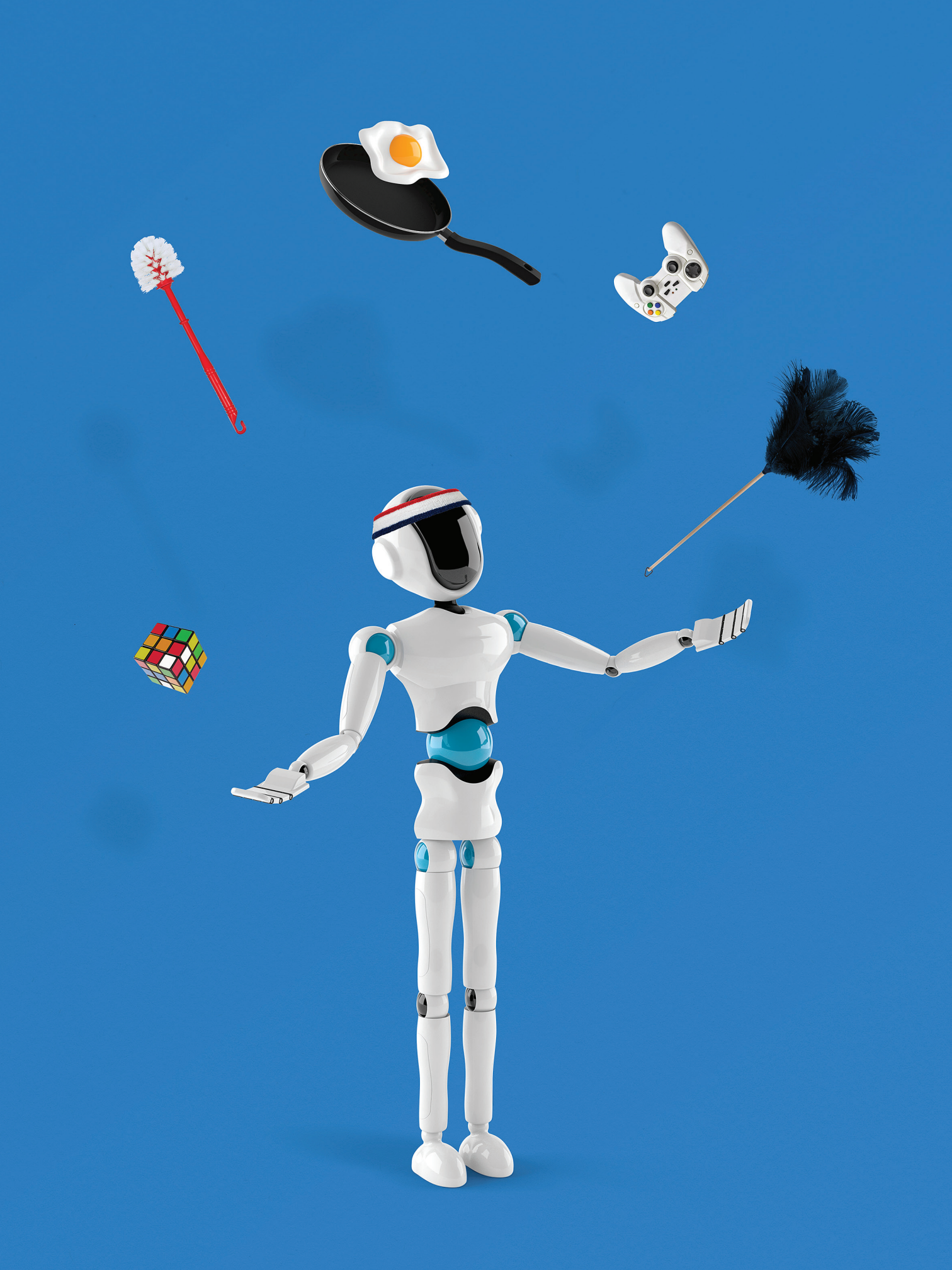
STEP 4 | STEP 3 | STEP 2 | STEP 1

STEP 4
The process is then repeated for the first hidden layer. For example, the first neuron in this layer may need to have its activation increased [orange arrow].

STEP 3
A similar process is then performed for the neurons in the second hidden layer. For example, to make the network more accurate, the top neuron in this layer may need to have its activation reduced [green arrow]. The network can be pushed in that direction by adjusting the weights of its connections with the first hidden layer [black arrows above].

STEP 2
To do that, the weights of the connections from the neurons in the second hidden layer to the output neuron for the digit "3" should be made more positive [black arrows above], with the size of the change being proportional to the activation of the connected hidden neuron.

STEP 1
When presented with a handwritten "3" at the input, the output neurons of an untrained network will have random activations. The desire is for the output neuron associated with 3 to have high activation [dark shading] and other output neurons to have low activations [light shading]. So the activation of the neuron associated with 3, for example, must be increased [purple arrow].



How to Train an ALL-PURPOSE ROBOT

DeepMind is tackling one of the hardest problems for AI

BY TOM CHIVERS

ARTIFICIAL INTELLIGENCE has reached deep into our lives, though you might be hard pressed to point to obvious examples of it. Among countless other behind-the-scenes chores, neural networks power our virtual assistants, make online shopping recommendations, recognize people in our snapshots, scrutinize our banking transactions for evidence of fraud, transcribe our voice messages, and weed out hateful social-media postings. What these applications have in common is that they involve learning and operating in a constrained, predictable environment. • But embedding AI more firmly into our endeavors and enterprises poses a great challenge. To get to the next level, researchers are trying to fuse AI and robotics to create an intelligence that can make decisions and control a physical body in the messy, unpredictable, and unforgiving real world. It's a potentially revolutionary objective that has caught the attention of some of the most powerful tech-research organizations on the planet. "I'd say that robotics as a field is probably 10 years behind where computer vision is," says Raia Hadsell, head of robotics at DeepMind, Google's London-based AI partner. (Both companies are subsidiaries of Alphabet.) • Even for Google, the challenges are daunting. Some are hard but straightforward:

Photo-illustration by Edmon de Haro

For most robotic applications, it's difficult to gather the huge data sets that have driven progress in other areas of AI. But some problems are more profound, and relate to long-standing conundrums in AI. Problems like, how do you learn a new task without forgetting the old one? And how do you create an AI that can apply the skills it learns for a new task to the tasks it has mastered before?

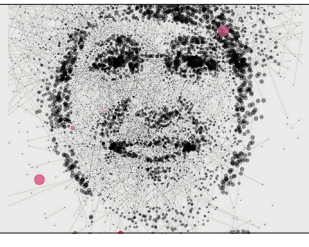
Success would mean opening AI to new categories of application. Many of the things we most fervently want AI to do—drive cars and trucks, work in nursing homes, clean up after disasters, perform basic household chores, build houses, sow, nurture, and harvest crops—could be accomplished only by robots that are much more sophisticated and versatile than the ones we have now.

Beyond opening up potentially enormous markets, the work bears directly on matters of profound importance not just for robotics but for all AI research, and indeed for our understanding of our own intelligence.

Catastrophic forgetting: When an AI learns a new task, it has an unfortunate tendency to forget all the old ones.

LET'S START WITH the prosaic problem first. A neural network is only as good as the quality and quantity of the data used to train it. The availability of enormous data sets has been key to the recent successes in AI: Image-recognition software is trained on millions of labeled images. AlphaGo, which beat a grandmaster at the ancient board game of Go, was trained on a data set of hundreds of thousands of human games, and on the millions of games it played against itself in simulation.

To train a robot, though, such huge data sets are unavailable. "This is a problem," notes Hadsell. You can simulate thousands of games of Go in a few minutes, run in parallel on hundreds of CPUs. But if it takes 3 seconds for a robot to pick up a cup, then you can only do it 20 times per minute per robot. What's more, if your image-recognition system gets the first million images wrong, it might not matter much. But if your bipedal robot falls over the first 1,000 times it tries to walk, then you'll have a badly dented robot, if not worse.



"The beautiful thing about AI and robotics is that you're never done."

MANUELA VELOSO, head of AI research at J.P. Morgan

The problem of real-world data is—at least for now—insurmountable. But that's not stopping DeepMind from gathering all it can, with robots constantly whirring in its labs. And across the field, robotics researchers are trying to get around this paucity of data with a technique called sim-to-real.

The San Francisco–based lab OpenAI recently exploited this strategy in training a robot hand to solve a Rubik's Cube. The researchers built a virtual environment containing a cube and a virtual model of the robot hand, and trained the AI that would run the hand in the simulation. Then they installed the AI in the real robot hand, and gave it a real Rubik's Cube. Their sim-to-real program enabled the physical robot to solve the physical puzzle.

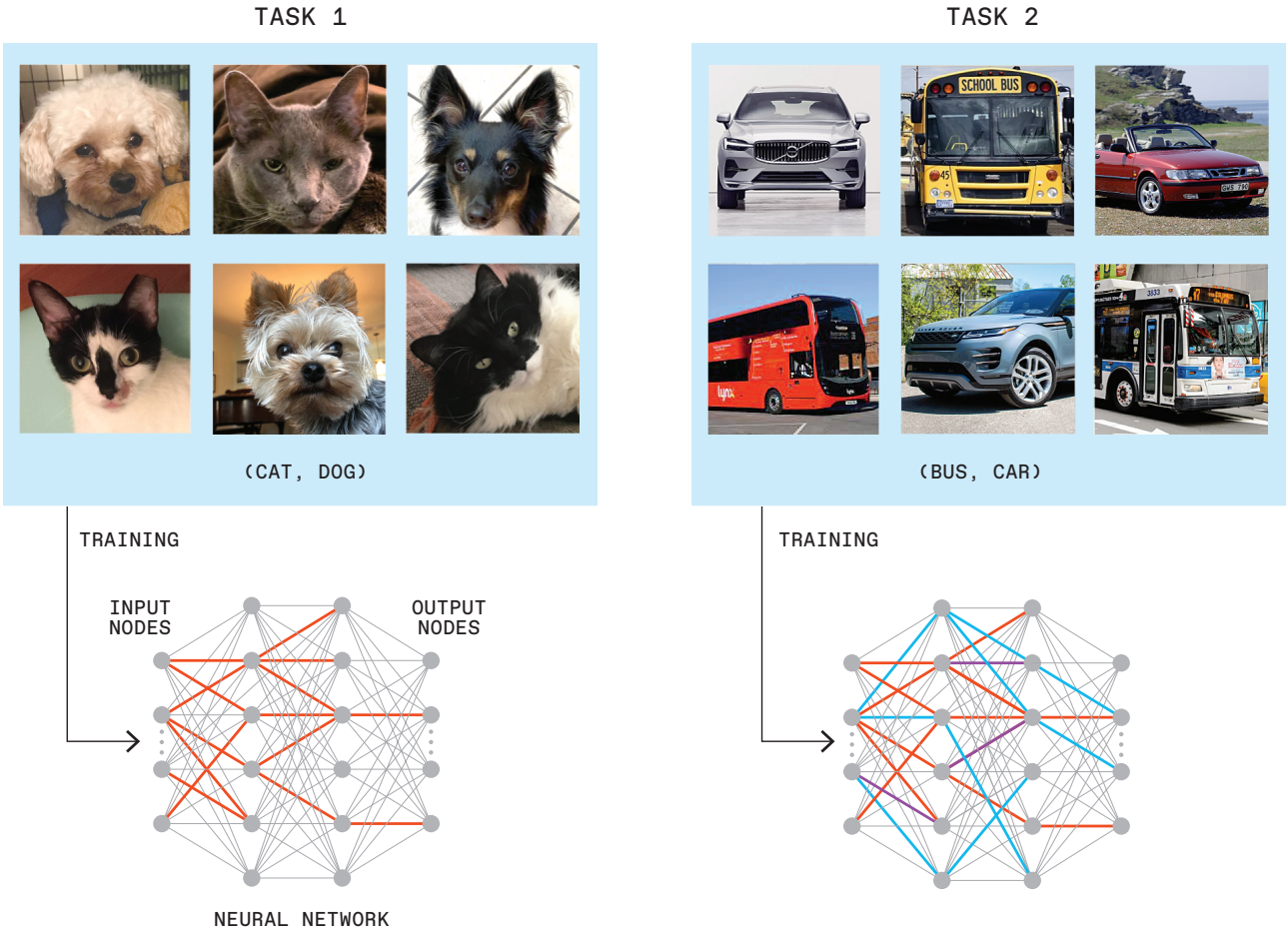
Despite such successes, the technique has major limitations, Hadsell says, noting that AI researcher and roboticist Rodney Brooks "likes to say that simulation is 'doomed to succeed.' " The trouble is that simulations are too perfect, too removed from the complexities of the real world. "Imagine two robot hands in simulation, trying to put a cellphone together," Hadsell says. If you allow them to try millions of times, they might eventually discover that by throwing all the pieces up in the air with *exactly* the right amount of force, with *exactly* the right amount of spin, that they can build the cellphone in a few seconds: The pieces fall down into place precisely where the robot wants them, making a phone. That might work in the perfectly predictable environment of a simulation, but it could never work in complex, messy reality. For now, researchers have to settle for these imperfect simulacrums. "You can add noise and randomness artificially," Hadsell explains, "but no contemporary simulation is good enough to truly recreate even a small slice of reality."

THERE ARE MORE profound problems. The one that Hadsell is most interested in is that of catastrophic forgetting: When an AI learns a new task, it has an unfortunate tendency to forget all the old ones.

The problem isn't lack of data storage. It's something inherent in how most modern AIs learn. Deep learning, the most common category of artificial intelligence today, is based on neural networks that use neuronlike computational nodes, arranged in layers, that are linked together by synapselike connections.

Before it can perform a task, such as classifying an image as that of either a cat or a dog, the neural network must be trained. The first layer of nodes receives an input image of either a cat or a dog. The

PORTRAIT BY SERGIO ALBIAC



Training of a neural network to distinguish whether a photograph is of a cat or a dog uses a portion of the nodes and connections in the network [shown in red, at left]. Using a technique called elastic weight consolidation, the network can then be trained on a different task, distinguishing images of cars from buses. The key connections from the original task are "frozen" and new connections are established [blue, at right]. A small fraction of the frozen connections, which would otherwise be used for the second task, are unavailable [purple, right diagram]. That slightly reduces performance on the second task.

nodes detect various features of the image and either fire or stay quiet, passing these inputs on to a second layer of nodes. Each node in each layer will fire if the input from the layer before is high enough. There can be many such layers, and at the end, the last layer will render a verdict: "cat" or "dog."

Each connection has a different "weight." For example, node A and node B might both feed their output to node C. Depending on their signals, C may then fire, or not. However, the A-C connection may have a weight of 3, and the B-C connection a weight of 5. In this case, B has greater influence over C. To give an implausibly oversimplified example, A might fire if the creature in the image has sharp teeth, while B might fire if the creature has a long snout. Since the length of the snout is more helpful than the sharpness of the teeth in distinguishing dogs from cats, C pays more attention to B than it does to A.

Each node has a threshold over which it will fire, sending a signal to its own downstream connections. Let's say C has a threshold of 7. Then if only A fires, it will stay quiet; if only B fires, it will stay quiet; but if A and B fire together, their signals to C will add up to 8, and C will fire, affecting the next layer.

What does all this have to do with training? Any learning scheme must be able to distinguish between correct and incorrect responses and improve itself accordingly. If a neural network is shown a picture of a dog, and it outputs "dog," then the connections that fired will be strengthened; those that did not will be weakened. If it incorrectly outputs "cat," then the reverse happens: The connections that fired will be weakened; those that did not will be strengthened.

But imagine you take your dog-and-cat-classifying neural network, and now start training it to distinguish a bus from a car. All its previous training will be useless. Its outputs in response to vehicle images will be random at first. But as it is trained, it will reweight its connections and gradually become effective. It will eventually be able to classify buses and cars with great accuracy. At this point, though, if you show it a picture of a dog, all the nodes will have been reweighted, and it will have "forgotten" everything it learned previously.

This is catastrophic forgetting, and it's a large part of the reason that programming neural networks with humanlike flexible intelligence is so difficult. "One of our classic examples was training an agent to play *Pong*," says Hadsell. You could get it playing so that it would win every game against the computer 20 to zero, she says; but if you perturb the weights just a little bit, such as by training it on *Breakout* or *Pac-Man*, "then the performance will—boop!—go off a cliff." Suddenly it will lose 20 to zero every time.

This weakness poses a major stumbling block not only for machines built to succeed at several different tasks, but also for any AI systems that are meant to adapt to changing circumstances in the world around them, learning new strategies as necessary.

Raia Hadsell [top] leads a team of roboticists at DeepMind in London. At OpenAI, researchers used simulations to train a robot hand [above] to solve a Rubik's Cube.

THERE ARE WAYS around the problem. An obvious one is to simply silo off each skill. Train your neural network on one task, save its network's weights to its data storage, then train it on a new task, saving those weights elsewhere. Then the system need only recognize the type of challenge at the outset and apply the proper set of weights.

But that strategy is limited. For one thing, it's not scalable. If you want to build a robot capable of accomplishing many tasks in a broad range of environments, you'd have to train it on every single one of them. And if the environment is unstructured, you won't even know ahead of time what some of

TOP: DEEPMIND; BOTTOM: OPENAI

those tasks will be. Another problem is that this strategy doesn't let the robot transfer the skills that it acquired solving task A over to task B. Such an ability to transfer knowledge is an important hallmark of human learning.

Hadsell's preferred approach is something called "elastic weight consolidation." The gist is that, after learning a task, a neural network will assess which of the synapselike connections between the neuronlike nodes are the most important to that task, and it will partially freeze their weights. "There'll be a relatively small number," she says. "Say, 5 percent." Then you protect these weights, making them harder to change, while the other nodes can learn as usual. Now, when your *Pong*-playing AI learns to play *Pac-Man*, those neurons most relevant to *Pong* will stay mostly in place, and it will continue to do well enough on *Pong*. It might not keep winning by a score of 20 to zero, but possibly by 18 to 2.

There's an obvious side effect, however. Each time your neural network learns a task, more of its neurons will become inelastic. If *Pong* fixes some neurons, and *Breakout* fixes some more, "eventually, as your agent goes on learning Atari games, it's going to get more and more fixed, less and less plastic," Hadsell explains.

This is roughly similar to human learning. When we're young, we're fantastic at learning new things. As we age, we get better at the things we have learned, but find it harder to learn new skills.

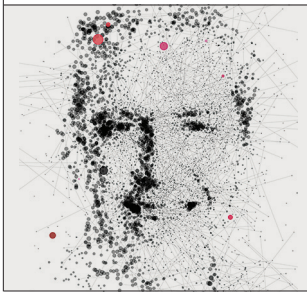
"Babies start out having much denser connections that are much weaker," says Hadsell. "Over time, those connections become sparser but stronger. It allows you to have memories, but it also limits your learning." She speculates that something like this might help explain why very young children have no memories: "Our brain layout simply doesn't support it." In a very young child, "everything is being catastrophically forgotten all the time, because everything is connected and nothing is protected."

If you want to build a robot capable of accomplishing many tasks in a broad range of environments, you'd have to train it on every single one of them.

The loss-of-elasticity problem is, Hadsell thinks, fixable. She has been working with the DeepMind team since 2018 on a technique called "progress and compress." It involves combining three relatively recent ideas in machine learning: progressive neural networks, knowledge distillation, and elastic weight consolidation, described above.

Progressive neural networks are a straightforward way of avoiding catastrophic forgetting. Instead of having a single neural network that trains on one task and then another, you have one neural network that trains on a task—say, *Breakout*. Then, when it has finished training, it freezes its connections in place, moves that neural network into storage, and creates a new neural network to train on a new task—say, *Pac-Man*. Its knowledge of each of the earlier tasks is frozen in place, so cannot be forgotten. And when each new neural network is created, it brings over connections from the previous games it has trained on, so it can transfer skills forward from old tasks to new ones. But, Hadsell says, it has a problem: It can't transfer knowledge the other way, from *new* skills to old. "If I go back and play *Breakout* again, I haven't actually learned anything from this [new] game," she says. "There's no backwards transfer."

That's where knowledge distillation, developed by the British-Canadian computer scientist Geoffrey Hinton, comes in. It involves taking many different neural networks trained on a task and compressing them into a single one, averaging their predictions. So, instead of having lots of neural networks, each trained on an individual game, you have just two: one that learns each new game, called the "active column," and one that contains all the learning from previous games, averaged out, called the "knowledge base." First the active column is trained on a new task—the "progress" phase—and then its connections are added to the knowledge base, and distilled—the "compress" phase. It helps to picture the two networks as, literally, two columns. Hadsell does, and draws them on the whiteboard for me as she talks.



PORTRAIT BY SERGIO ALBIAC

"General intelligence doesn't mean you have to understand humans. There will be a lot of intelligent machines that don't need to know that stuff. [Take the] example of robotic construction workers on Mars: I don't think they need to understand human emotions and human desires to be able to construct things."

JEFF HAWKINS, cofounder and chief scientist of Numenta

The trouble is, by using knowledge distillation to lump the many individual neural networks of the progressive-neural-network system together, you've brought the problem of catastrophic forgetting back in. You'll change all the weights of the connections and render your earlier training useless. To deal with this, Hadsell adds in elastic weight consolidation: Each time the active column transfers its learning about a particular task to the knowledge base, it partially freezes the nodes most important to that particular task.

By having two neural networks, Hadsell's system avoids the main problem with elastic weight consolidation, which is that all its connections will eventually freeze. The knowledge base can be as large as you like, so a few frozen nodes won't matter. But the active column itself can be much smaller, and smaller neural networks can learn faster and more efficiently than larger ones. So the progress-and-compress model, Hadsell says, will allow an AI system to transfer skills from old tasks to new ones, and from new tasks back to old ones, while never either catastrophically forgetting or becoming unable to learn anything new.

Other researchers are using different strategies to attack the catastrophic forgetting problem; there are half a dozen or so avenues of research. Ted Senator, a program manager at the Defense Advanced Research Projects Agency (DARPA), leads a group that is using one of the most promising, a technique called internal replay. "It's modeled after theories of how the brain operates," Senator explains, "particularly the role of sleep in preserving memory."

The theory is that the human brain replays the day's memories, both while awake and asleep: It reactivates its neurons in similar patterns to those that arose while it was having the corresponding experience. This reactivation helps stabilize the patterns, meaning that they are not overwritten so easily. Internal replay does something similar. In between learning tasks, the neural network recreates patterns of connections and weights, loosely mimicking the awake-sleep cycle of human neural activity. The technique has proved quite effective at avoiding catastrophic forgetting.

"You know the cat is never going to learn language, and I'm okay with that."

RAIA HADSELL

THERE ARE MANY other hurdles to overcome in the quest to bring embodied AI safely into our daily lives. "We have made huge progress in symbolic, data-driven AI," says Thrishantha Nanayakkara, who works on robotics at Imperial College London. "But when it comes to contact, we fail miserably. We don't have a robot that we can trust to hold a hamster safely. We cannot trust a robot to be around an elderly person or a child."

Nanayakkara points out that much of the "processing" that enables animals to deal with the world doesn't happen in the brain, but rather elsewhere in the body. For instance, the shape of the human ear canal works to separate out sound waves, essentially performing "the Fourier series in real time." Otherwise that processing would have to happen in the brain, at a cost of precious microseconds. "If, when you hear things, they're no longer there, then you're not embedded in the environment," he says. But most robots currently rely on CPUs to process all the inputs, a limitation that he believes will have to be surmounted before substantial progress can be made.

His colleague Petar Kormushev says another problem is proprioception, the robot's sense of its own physicality. A robot's model of its own size and shape is programmed in directly by humans. The problem is that when it picks up a heavy object, it has no way of updating its self-image. When we pick up a hammer, we adjust our mental model of our body's shape and weight, which lets us use the hammer as an extension of our body. "It sounds ridiculous but they [robots] are not able to update their kinematic models," he says. Newborn babies, he notes, make random movements that give them feedback not only about the world but about their own bodies. He believes that some analogous technique would work for robots.

At the University of Oxford, Ingmar Posner is working on a robot version of "metacognition." Human thought is often modeled as having two main "systems"—system 1, which responds quickly and intuitively, such as when we catch a ball or

"The policy goal is to get the regulators to enforce their laws in the age of algorithms. You can't give up on enforcing antidiscrimination laws because you don't understand the technology." **CATHY O'NEIL**, founder and CEO of O'Neil Risk Consulting and Algorithmic Auditing

PORTRAIT BY SERGIO ALBIAC

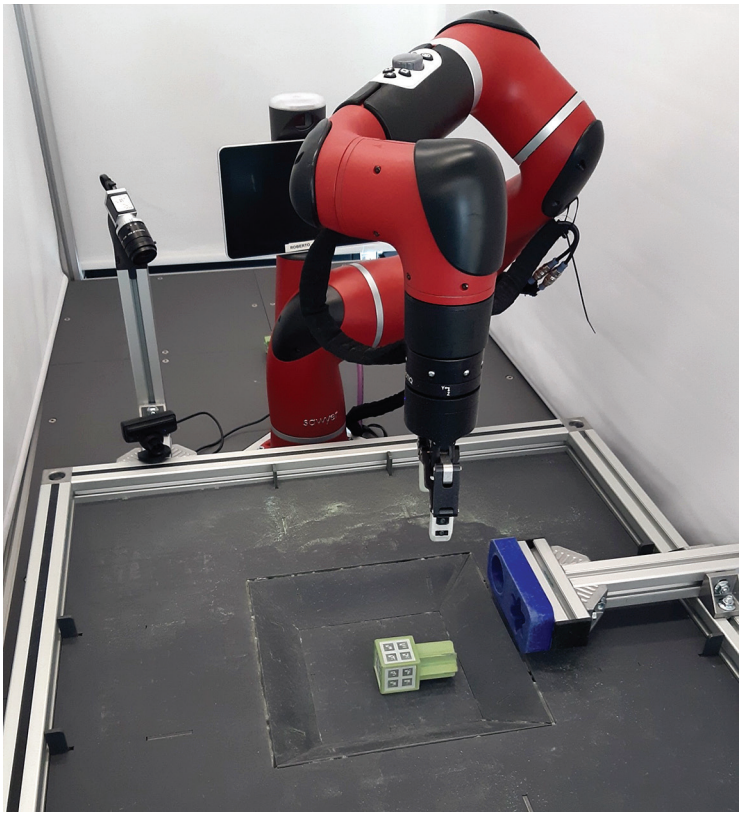
answer questions like "which of these two blocks is blue?," and system 2, which responds more slowly and with more effort. It comes into play when we learn a new task or answer a more difficult mathematical question. Posner has built functionally equivalent systems in AI. Robots, in his view, are consistently either overconfident or underconfident, and need ways of knowing when they don't know something. "There are things in our brain that check our responses about the world. There's a bit which says don't trust your intuitive response," he says.

For most of these researchers, including Hadsell and her colleagues at DeepMind, the long-term goal is "general" intelligence. However, Hadsell's idea of an artificial general intelligence isn't the usual one—of an AI that can perform all the intellectual tasks that a human can, and more. Motivating her own work has "never been this idea of building a superintelligence," she says. "It's more: How do we come up with general methods to develop intelligence for solving particular problems?" Cat intelligence, for instance, is general in that it will never encounter some new problem that makes it freeze up or fail. "I find that level of animal intelligence, which involves incredible agility in the world, fusing different sensory modalities, really appealing. You know the cat is never going to learn language, and I'm okay with that."

Hadsell wants to build algorithms and robots that will be able to learn and cope with a wide array of problems in a specific sphere. A robot intended to clean up after a nuclear mishap, for example, might have some quite high-level goal—"make this area safe"—and be able to divide that into smaller subgoals, such as finding the radioactive materials and safely removing them.

I CAN'T RESIST ASKING about consciousness. Some AI researchers, including Hadsell's DeepMind colleague Murray Shanahan, suspect that it will be impossible to build an embodied AI of real general intelligence without the machine having some sort of consciousness. Hadsell herself, though, despite a background in the philosophy of religion, has a robustly practical approach.

"I have a fairly simplistic view of consciousness," she says. For her, consciousness means an ability to think outside the narrow moment of "now"—to use memory to access the past, and to use imagination to envision the future. We humans do this well. Other creatures, less so: Cats seem to have a smaller time horizon than we do, with less planning for the future. Bugs, less still. She is not keen to be drawn out on the hard problem of consciousness and other philosophical ideas. In fact, most roboticists seem to want to avoid it. Kormushev likens it to asking "Can submarines swim?...It's pointless to debate. As long as they do what I want, we don't have to torture ourselves with the question."



Pushing a star-shaped peg into a star-shaped hole may seem simple, but it was a minor triumph for one of DeepMind's robots.

DEEPMIND

In the DeepMind robotics lab it's easy to see why that sort of question is not front and center. The robots' efforts to pick up blocks suggest we don't have to worry just yet about philosophical issues relating to artificial consciousness.

Nevertheless, while walking around the lab, I find myself cheering one of them on. A red robotic arm is trying, jerkily, to pick up a star-shaped brick and then insert it into a star-shaped aperture, as a toddler might. On the second attempt, it gets the brick aligned and is on the verge of putting it in the slot. I find myself yelling "Come on, lad!," provoking a raised eyebrow from Hadsell. Then it successfully puts the brick in place.

One task completed, at least. Now, it just needs to hang on to that strategy while learning to play *Pong*. ■

7 Revealing Ways AIs FAIL

Neural networks can be disastrously brittle, forgetful, and surprisingly bad at math BY CHARLES Q. CHOI

ARTIFICIAL INTELLIGENCE systems can perform more quickly, accurately, reliably, and impartially than humans on a wide range of problems, from detecting cancer to deciding who receives an interview for a job. But AIs have also suffered numerous, sometimes deadly, failures. And the increasing ubiquity of AI means that failures can affect not just individuals but millions of people.

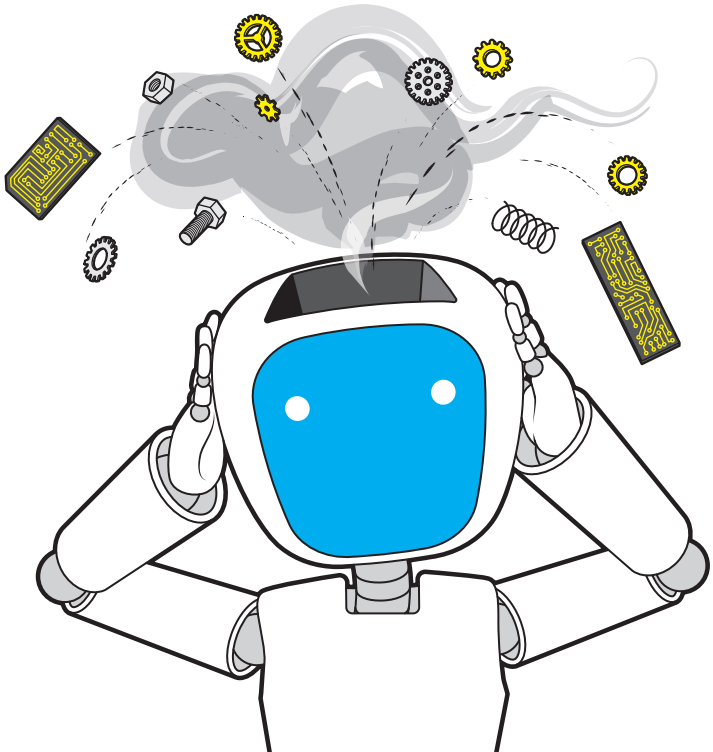
Increasingly, the AI community is cataloging these failures with an eye toward monitoring the risks they may pose. "There tends to be very little information for users to understand how these systems work and what it means to them," says Charlie Pownall, founder of the AI, Algorithmic and Automation Incident & Controversy Repository. "I think this directly impacts trust and confidence in these systems. There are lots of possible reasons why organizations are reluctant to get into the nitty-gritty of what exactly happened in an AI incident or controversy, not the least being potential legal exposure, but if looked at through the lens of trustworthiness, it's in their best interest to do so."

Part of the problem is that the neural network technology that drives many AI systems can break down in ways that remain a mystery to researchers. "It's unpredictable which problems artificial intelligence will be good at, because we don't understand intelligence itself very well," says computer scientist Dan Hendrycks at the University of California, Berkeley.

Here are seven examples of AI failures and what current weaknesses they reveal about artificial intelligence. Scientists are discussing possible ways to deal with some of these problems; others currently defy explanation or may, philosophically speaking, lack any conclusive solution altogether.

1) BRITTLENESS

Take a picture of a school bus. Flip it so it lays on its side, as it might be found in the case of an accident



in the real world. A 2018 study found that state-of-the-art AIs that would normally correctly identify the school bus right-side-up failed to do so on average 97 percent of the time when it was rotated.

"They will say the school bus is a snowplow with very high confidence," says computer scientist Anh Nguyen at Auburn University, in Alabama. The AIs are not capable of a task of mental rotation "that even my 3-year-old son could do," he says.

Such a failure is an example of brittleness. An AI often "can only recognize a pattern it has seen before," Nguyen says. "If you show it a new pattern, it is easily fooled."

There are numerous troubling cases of AI brittleness. Fastening stickers on a stop sign can make an AI misread it. Changing a single pixel on an image can make an AI think a horse is a frog. Neural networks can be 99.99 percent confident that multicolor static is a picture of a lion. Medical images can get modified in a way imperceptible to the human eye so that AI systems misdiagnose cancer 100 percent of the time. And so on.

One possible way to make AIs more robust against such failures is to expose them to as many confounding "adversarial" examples as possible, Hendrycks says. However, they may still fail against rare "black swan" events. "Black-swan problems such as COVID or the recession are hard for even humans to address—they may not be problems just specific to machine learning," he notes.

2) EMBEDDED BIAS

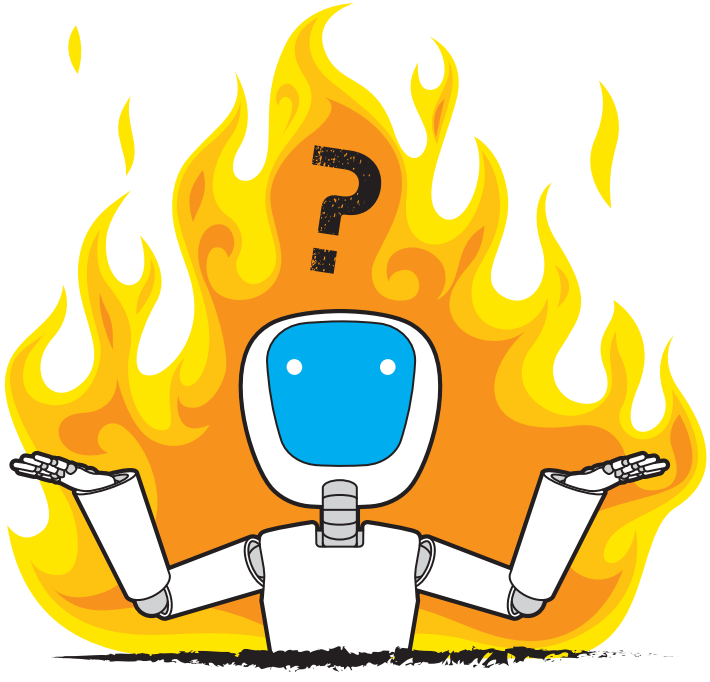
Increasingly, AI is used to help support major decisions, such as who receives a loan, the length of a jail sentence, and who gets health care first. The hope is that AIs can make decisions more impartially than people often have, but much research has found that biases embedded in the data on which these AIs are trained can result in automated discrimination en masse, posing immense risks to society.

For example, in 2019, scientists found a nationally deployed health care algorithm in the United States was racially biased, affecting millions of Americans. The AI was designed to identify which patients would benefit most from intensive-care programs, but it routinely enrolled healthier white patients into such programs ahead of black patients who were sicker.

Physician and researcher Ziad Obermeyer at the University of California, Berkeley, and his colleagues found the algorithm mistakenly assumed that people with high health care costs were also the sickest patients and most in need of care. However, due to systemic racism, "black patients are less likely to get health care when they need it, so are less likely to generate costs," he explains.

After working with the software's developer, Obermeyer and his colleagues helped design a new algorithm that analyzed other variables and displayed 84 percent less bias. "It's a lot more work, but accounting for bias is not at all impossible," he says. They recently drafted a playbook that outlines a few basic steps that governments, businesses, and other groups can implement to detect and prevent bias in existing and future software they use. These include identifying all the algorithms they employ, understanding this software's ideal target and its performance toward that goal, retraining the AI if needed, and creating a high-level oversight body.

Illustrations by Chris Philpot



3) CATASTROPHIC FORGETTING

Deepfakes—highly realistic artificially generated fake images and videos, often of celebrities, politicians, and other public figures—are becoming increasingly common on the Internet and social media, and could wreak plenty of havoc by fraudulently depicting people saying or doing things that never really happened. To develop an AI that could detect deepfakes, computer scientist Shahroz Tariq and his colleagues at Sungkyunkwan University, in South Korea, created a website where people could upload images to check their authenticity.

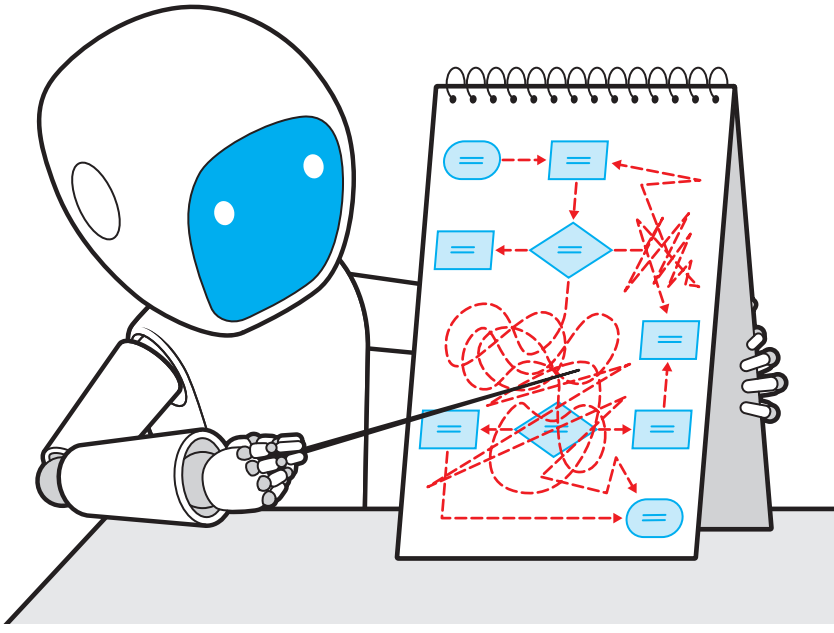
In the beginning, the researchers trained their neural network to spot one kind of deepfake. However, after a few months, many new types of deepfake emerged, and when they trained their AI to identify these new varieties of deepfake, it quickly forgot how to detect the old ones.

This was an example of catastrophic forgetting—the tendency of an AI to entirely and abruptly forget information it previously knew after learning new information, essentially overwriting past knowledge with new knowledge. "Artificial neural networks have a terrible memory," Tariq says.

AI researchers are pursuing a variety of strategies to prevent catastrophic forgetting so that neural networks can, as humans seem to do, continuously learn effortlessly. A simple technique is to create a specialized neural network for each new task one wants performed—say, distinguishing cats from dogs or apples from oranges—"but this is obviously not scalable, as the number of networks increases linearly with the number of tasks," says machine-learning researcher Sam Kessler at the University of Oxford, in England.

One alternative Tariq and his colleagues explored as they trained their AI to spot new kinds of deepfakes was to supply it with a small amount of data on how it identified older types so it would not forget how to detect them. Essentially, this is like reviewing a summary of a textbook chapter before an exam, Tariq says.

However, AIs may not always have access to past knowledge—for instance, when dealing with private information such as medical records. Tariq and his colleagues wanted to make an AI that didn't rely on data from prior tasks. They had it train itself how to spot new deepfake types while also learning from another AI that was previously trained how to recognize older deepfake varieties. They found this "knowledge distillation" strategy was roughly 87 percent accurate at detecting the kind of low-quality deepfakes typically shared on social media.



Illustrations by Chris Philpot

4) EXPLAINABILITY

Why *does* an AI suspect a person might be a criminal or have cancer? The explanation for this and other high-stakes predictions can have many legal, medical, and other consequences. The way in which AIs reach conclusions has long been considered a mysterious black box, leading to many attempts to devise ways to explain AIs' inner workings. "However, my recent work suggests the field of explainability is getting somewhat stuck," says Auburn's Nguyen.

Nguyen and his colleagues investigated seven different techniques that researchers have developed to attribute explanations for AI decisions—for instance, what makes an image of a matchstick a matchstick? Is it the flame or the wooden stick? They discovered that many of these methods "are quite unstable," Nguyen says. "They can give you different explanations every time."

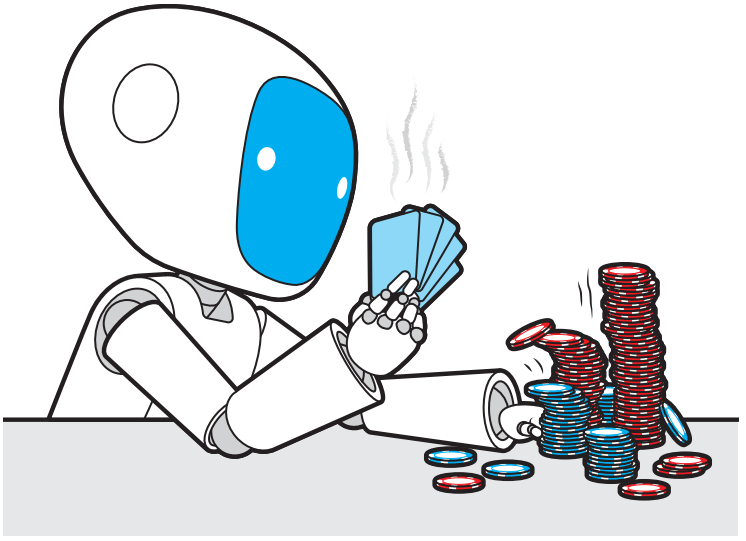
In addition, while one attribution method might work on one set of neural networks, "it might fail completely on another set," Nguyen adds. The future of explainability may involve building databases of correct explanations, Nguyen says. Attribution methods can then go to such knowledge bases "and search for facts that might explain decisions," he says.

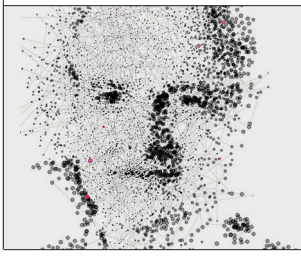
5) QUANTIFYING UNCERTAINTY

In 2016, a Tesla Model S car on autopilot collided with a truck that was turning left in front of it in northern Florida, killing its driver—the automated driving system's first reported fatality. According to Tesla's official blog, neither the autopilot system nor the driver "noticed the white side of the tractor trailer against a brightly lit sky, so the brake was not applied."

One potential way Tesla, Uber, and other companies may avoid such disasters is for their cars to do a better job at calculating and dealing with uncertainty. Currently AIs "can be very certain even though they're very wrong." Oxford's Kessler says that if an algorithm makes a decision, "we should have a robust idea of how confident it is in that decision, especially for a medical diagnosis or a self-driving car, and if it's very uncertain, then a human can intervene and give [their] own verdict or assessment of the situation."

For example, computer scientist Moloud Abdar at Deakin University in Australia and his colleagues applied several different uncertainty quantification techniques as an AI classified skin-cancer images as malignant or benign, or melanoma or not. The researcher found these methods helped prevent the AI from making overconfident diagnoses.





PORTRAIT BY SERGIO ALBIAC

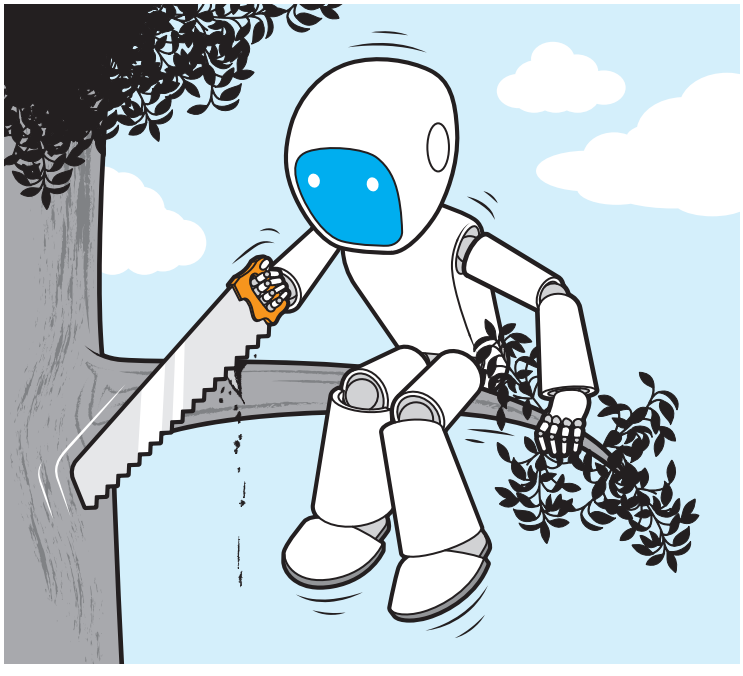
"Those who argue that the risk from AI is negligible have failed to explain why superintelligent AI systems will necessarily remain under human control; and they have not even tried to explain why superintelligent AI systems will never be developed."

STUART RUSSELL, professor at University of California, Berkeley

Autonomous vehicles remain challenging for quantifying uncertainty, because the current uncertainty-quantification techniques are often relatively time consuming, "and cars cannot wait for them," Abdar says. "We need to have much faster approaches."

6) COMMON SENSE

AIs lack common sense—the ability to reach acceptable, logical conclusions based on a vast context of everyday knowledge that people usually take for granted, says computer scientist Xiang Ren at the University of Southern California. "If you don't pay very much attention to what these models are actually learning, they can learn shortcuts that lead them to misbehave," he says.

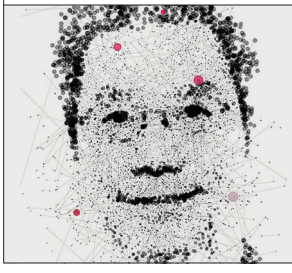


For instance, scientists may train AIs to detect hate speech on data where such speech is unusually high, such as white supremacist forums. However, when this software is exposed to the real world, it can fail to recognize that black and gay people may respectively use the words "black" and "gay" more often than other groups. "Even if a post is quoting a news article mentioning Jewish or black or gay people without any particular sentiment, it might be misclassified as hate speech," Ren says. In contrast, "humans reading through a whole sentence can recognize when an adjective is used in a hateful context."

Previous research suggested that state-of-the-art AIs could draw logical inferences about the world with up to roughly 90 percent accuracy, suggesting they were making progress at achieving common sense. However, when Ren and his colleagues tested these models, they found even the best AI could generate logically coherent sentences with slightly less than 32 percent accuracy. When it comes to developing common sense, "one thing we care a lot [about] these days in the AI community is employing more comprehensive checklists to look at the behavior of models on multiple dimensions," he says.

7) MATH

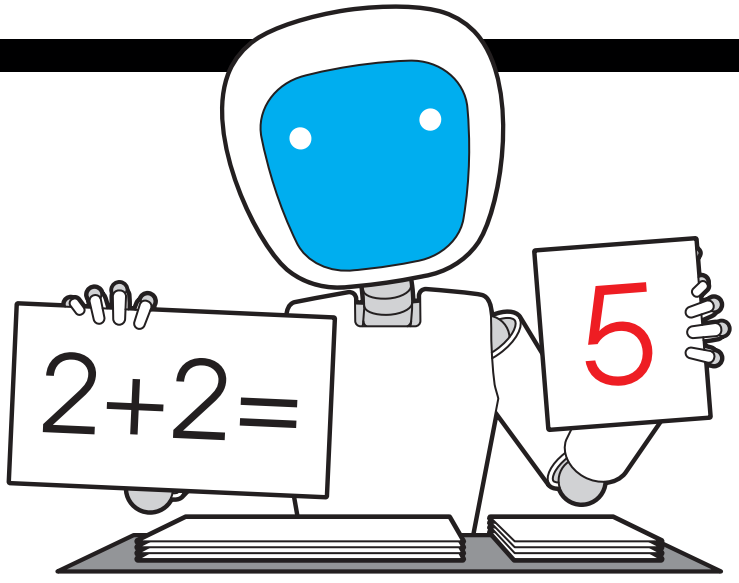
Although conventional computers are good at crunching numbers, AIs "are surprisingly not good at mathematics at all," Berkeley's Hendrycks says. "You might have the latest and greatest models that



"AI will take many single-task, single-domain jobs away. You can argue that humans have abilities that AI does not: We can conceptualize, strategize, create. Whereas today's AI is just a really smart pattern recognizer that can take in data [and] optimize. But how many jobs in the world are simple repetitions of tasks that can be optimized?" **KAI-FU LEE**, chairman and CEO of Sinovation Ventures

PORTRAIT BY SERGIO ALBIAC

Illustrations by Chris Philpot



take hundreds of GPUs to train, and they're still just not as reliable as a pocket calculator."

For example, Hendrycks and his colleagues trained an AI on hundreds of thousands of math problems with step-by-step solutions. However, when tested on 12,500 problems from high school math competitions, "it only got something like 5 percent accuracy," he says. In comparison, a three-time International Mathematical Olympiad gold medalist attained 90 percent success on such problems "without a calculator," he adds.

Neural networks nowadays can learn to solve nearly every kind of problem "if you just give it enough data and enough resources, but not math," Hendrycks says. Many problems in science require a lot of math, so this current weakness of AI can limit its application in scientific research, he notes.

It remains uncertain why AI is currently bad at math. One possibility is that neural networks attack problems in a highly parallel manner like human brains, whereas math problems typically require a long series of steps to solve, so maybe the way AIs process data is not as suitable for such tasks, "in the same way that humans generally can't do huge calculations in their head," Hendrycks says. However, AI's poor performance on math "is still a niche topic: There hasn't been much traction on the problem," he adds. ■

A HUMAN *in the Loop*

AI won't surpass human intelligence anytime soon

BY RODNEY BROOKS

WE ARE WELL INTO THE THIRD WAVE OF MAJOR INVESTMENT in artificial intelligence. So it's a fine time to take a historical perspective on the current success of AI. In the 1960s, the early AI researchers often breathlessly predicted that human-level intelligent machines were only 10 years away. That form of AI was based on logical reasoning with symbols, and was carried out with what today seem like ludicrously slow digital computers. Those same researchers considered and rejected neural networks. • In the 1980s, AI's second age was based on two technologies: rule-based expert systems—a more heuristic form of symbol-based logical reasoning—and a resurgence in neural networks triggered by the emergence of new training algorithms. Again, there were breathless predictions about the end of human dominance in intelligence.

The third and current age of AI arose during the early 2000s with new symbolic-reasoning systems based on algorithms capable of solving a class of problems called 3SAT and with another advance called simultaneous localization and mapping. SLAM is a technique for building maps incrementally as a robot moves around in the world.

In the early 2010s, this wave gathered powerful new momentum with the rise of neural networks learning from massive data sets. It soon turned into a tsunami of promise, hype, and profitable applications.

Regardless of what you might think about AI, the reality is that just about every successful deployment has either one of two expedients: It has a person somewhere in the loop, or the cost of failure, should the system blunder, is very low. In 2002, iRobot, a company that I cofounded, introduced the first mass-market autonomous home-cleaning robot, the Roomba, at a price that severely constricted how much AI we could endow it with. The limited AI wasn't a problem, though. Our worst failure scenarios had the Roomba missing a patch of floor and failing to pick up a dustball.

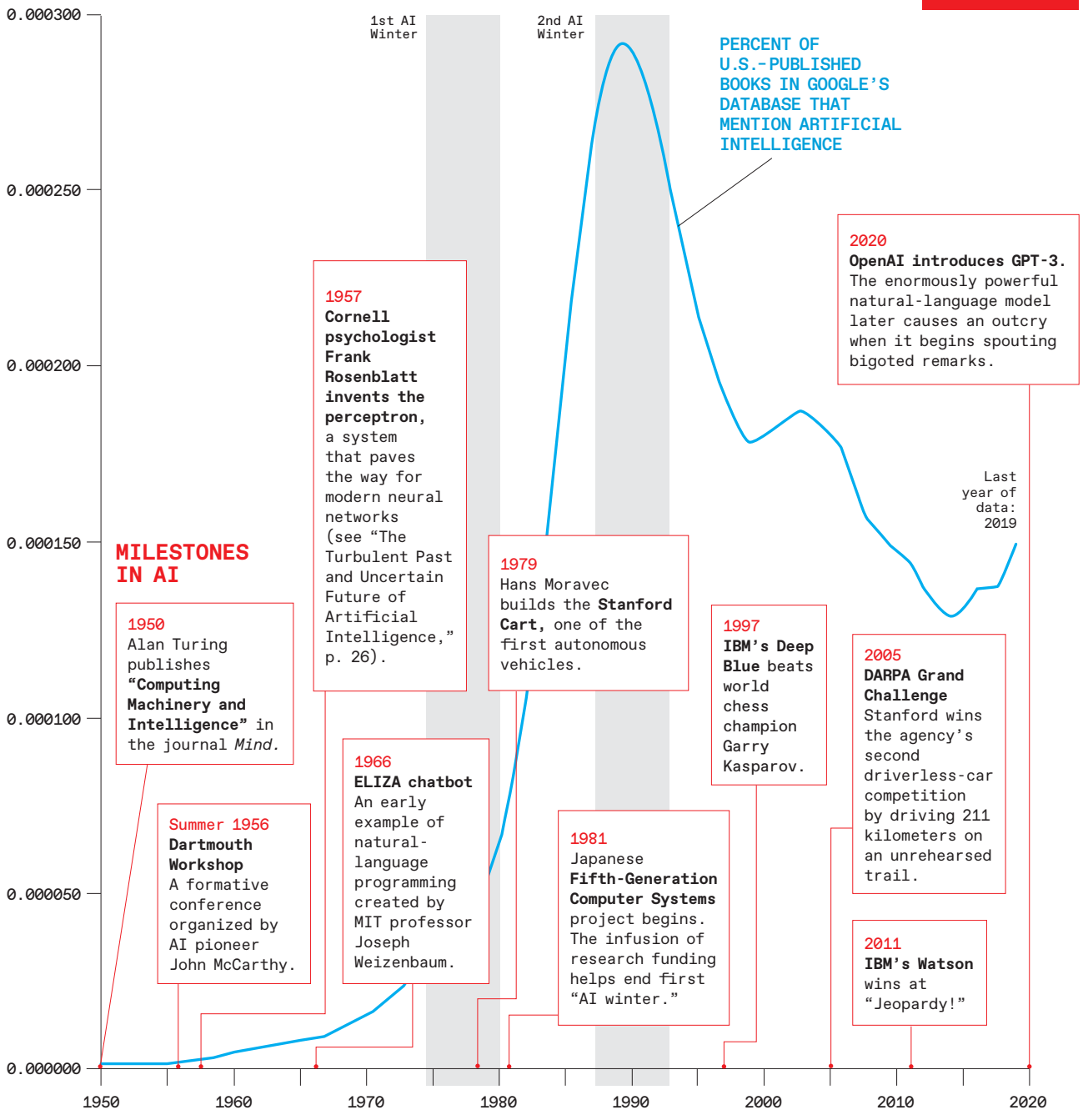
That same year we started deploying the first of thousands of robots in Afghanistan and then Iraq to be used to help troops disable improvised explosive devices. Failures there could kill someone, so there was always a human in the loop giving supervisory commands to the AI systems on the robot.

Just about every successful deployment of AI has either one of two expedients: It has a person somewhere in the loop, or the cost of failure, should the system blunder, is very low.

These days AI systems autonomously decide what advertisements to show us on our Web pages. Stupidly chosen ads are not a big deal; in fact, they are plentiful. Likewise search engines, also powered by AI, show us a list of choices so that we can skip over their mistakes with just a glance. On dating sites, AI systems choose who we see, but fortunately those sites are not arranging our marriages without us having a say in it.

So far the only self-driving systems deployed on production automobiles, no matter what the marketing people may say, are all Level 2. These systems require a human driver to keep their hands on the wheel and to stay attentive at all times so that they can take over immediately if the system is making a mistake. And there have already been fatal consequences when people were not paying attention.

These haven't been the only terrible failures of AI systems when no person was in the loop. For example, people have been wrongly arrested based

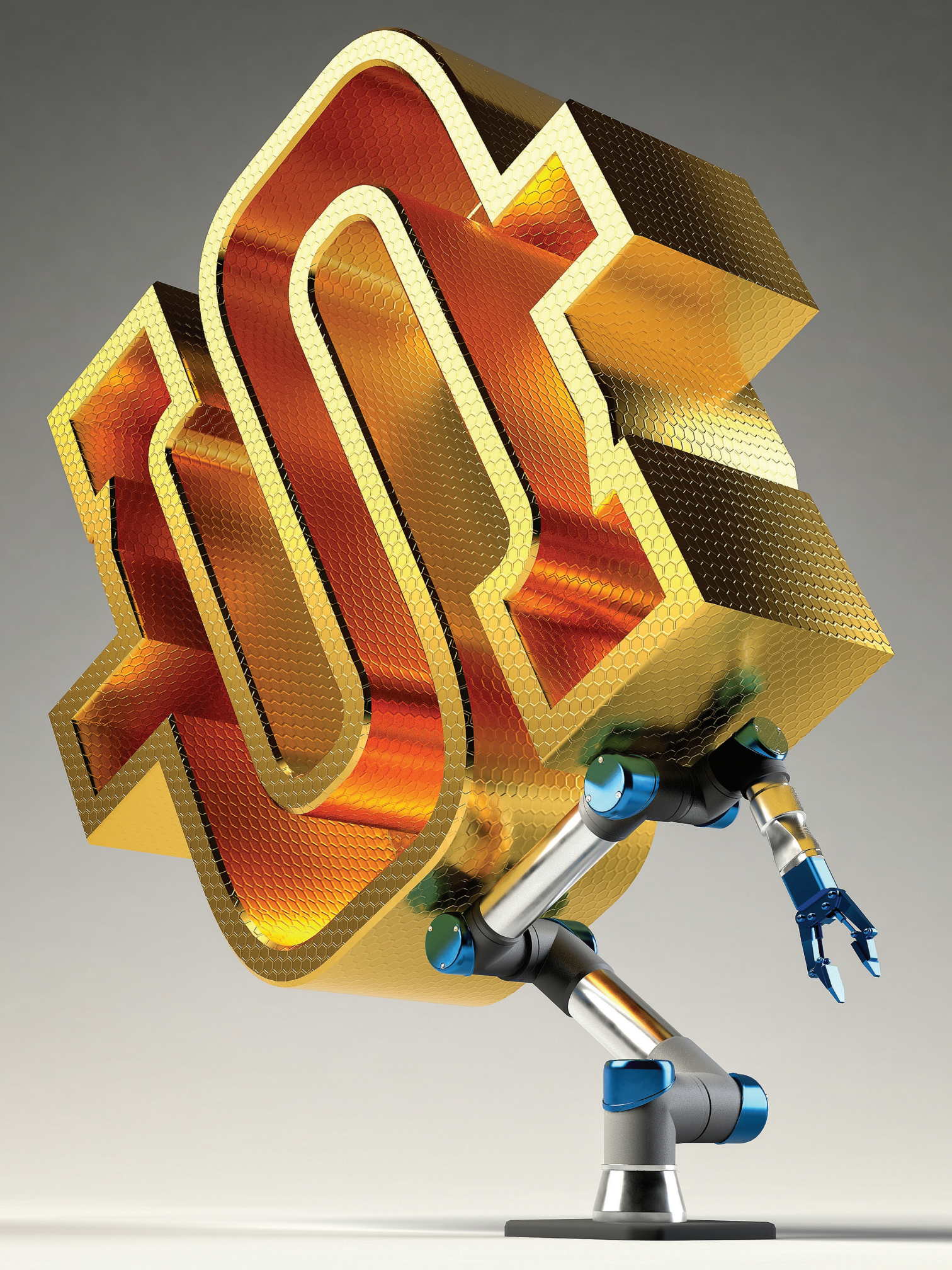


SOURCE: GOOGLE NGRAMS

on face-recognition technology that works poorly on racial minorities, making mistakes that no attentive human would make.

Sometimes we are in the loop even when the consequences of failure aren't dire. AI systems power the speech and language understanding of our smart speakers and the entertainment and navigation systems in our cars. We, the consumers, soon adapt our language to each such AI agent, quickly learning what they can and can't understand, in much the same way as we might with our children and elderly parents. The AI agents are cleverly designed to give us just enough feedback on what they've heard us say without getting too tedious, while letting us know about anything important that may need to be corrected. Here, we, the users, are the people in the loop. The ghost in the machine, if you will.

Ask not what your AI system can do for you, but instead what it has tricked you into doing for it. ■



Deep Learning's DIMINISHING RETURNS

The cost of improvement is becoming unsustainable

BY NEIL C. THOMPSON, KRISTJAN GREENEWALD, KEEHEON LEE & GABRIEL F. MANSO

DEEP LEARNING IS NOW being used to translate between languages, predict how proteins fold, analyze medical scans, and play games as complex as Go, to name just a few applications of a technique that is now becoming pervasive. Success in those and other realms has brought this machine-learning technique from obscurity in the early 2000s to dominance today. • Although deep learning's rise to fame is relatively recent, its origins are not. In 1958, back when mainframe computers filled rooms and ran on vacuum tubes, knowledge of the interconnections between neurons in the brain inspired Frank Rosenblatt at Cornell to design the first artificial neural network, which he presciently described as a "pattern-recognizing device." But Rosenblatt's ambitions outpaced the capabilities of his era—and he knew it. Even his inaugural paper was forced to acknowledge the voracious appetite of neural networks for computational power, bemoaning that "as the number of connections in the network increases...the burden on a conventional digital computer soon becomes excessive." • Fortunately for such artificial neural networks—later rechristened "deep learning" when

Illustration by Eddie Guy

they included extra layers of neurons—decades of Moore's Law and other improvements in computer hardware yielded a roughly 10-million-fold increase in the number of computations that a computer could do in a second. So when researchers returned to deep learning in the late 2000s, they wielded tools equal to the challenge.

These more-powerful computers made it possible to construct networks with vastly more connections and neurons and hence greater ability to model complex phenomena. Researchers used that ability to break record after record as they applied deep learning to new tasks.

While deep learning's rise may have been meteoric, its future may be bumpy. Like Rosenblatt before them, today's deep-learning researchers are nearing the frontier of what their tools can achieve. To understand why this will reshape machine learning, you must first understand why deep learning has been so successful and what it costs to keep it that way.

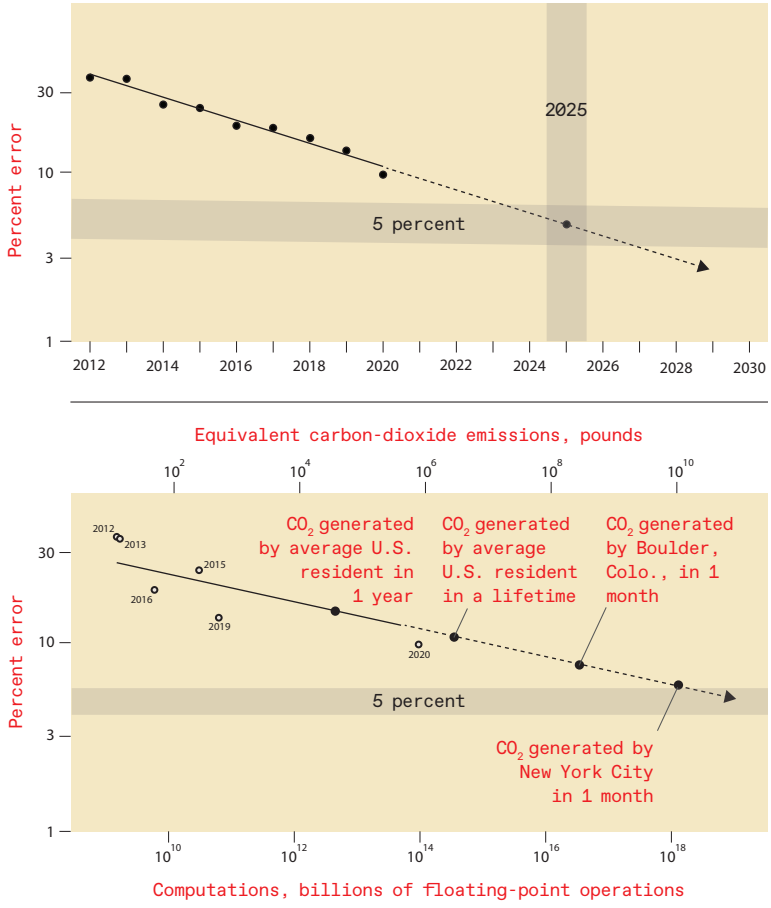
DEEP LEARNING is a modern incarnation of the long-running trend in artificial intelligence that has been moving from streamlined systems based on expert knowledge toward flexible statistical models. Early AI systems were rule based, applying logic and expert knowledge to derive results. Later systems incorporated learning to set their adjustable parameters, but these were usually few in number.

Today's neural networks also learn parameter values, but those parameters are part of such flexible computer models that—if they are big enough—they become universal function approximators, meaning they can fit any type of data. This unlimited flexibility is the reason why deep learning can be applied to so many different domains.

The flexibility of neural networks comes from taking the many inputs to the model and having the network combine them in myriad ways. This means the outputs won't be the result of applying simple formulas but instead immensely complicated ones.

For example, when the cutting-edge image-recognition system Noisy Student converts the pixel values of an image into probabilities for what the object in that image is, it does so using a network with 480 million parameters. The training to ascertain the values of such a large number of parameters is even more remarkable because it was done with only 1.2 million labeled images—which may understandably confuse those of us who remember from high school algebra that we are supposed to have more equations than unknowns. Breaking that rule turns out to be the key.

Deep-learning models are overparameterized, which is to say they have more parameters than there are data points available for training. Classically, this would lead to overfitting, where the model not only learns general trends but also the random vagaries of the data it was trained on. Deep learning avoids this trap by initializing the parameters randomly and then iteratively adjusting sets of them to better fit the



Extrapolating the gains of recent years might suggest that by 2025 the error level in the best deep-learning systems designed for recognizing objects in the ImageNet data set should be reduced to just 5 percent [top]. But the computing resources and energy required to train such a future system would be enormous, leading to the emission of as much carbon dioxide as New York City generates in one month [bottom].

SOURCE: N.C. THOMPSON, K. GREENEWALD, K. LEE, G.F. MANSO

data using a method called stochastic gradient descent. Surprisingly, this procedure has been proven to ensure that the learned model generalizes well.

The success of flexible deep-learning models can be seen in machine translation. For decades, software has been used to translate text from one language to another. Early approaches to this problem used rules designed by grammar experts. But as more textual data became available in specific languages, statistical approaches—ones that go by such esoteric names as maximum entropy, hidden Markov models, and conditional random fields—could be applied.

Initially, the approaches that worked best for each language differed based on data availability and grammatical properties. For example, rule-based approaches to translating languages such as Urdu, Arabic, and Malay outperformed statistical ones—at first. Today, all these approaches have been outpaced by deep learning, which has proven itself superior almost everywhere it's applied.

So the good news is that deep learning provides enormous flexibility. The bad news is that this flexibility comes at an enormous computational cost. This unfortunate reality has two parts.

The first part is true of all statistical models: To improve performance by a factor of k, at least k^2 more data points must be used to train the model. The second part of the computational cost comes explicitly from overparameterization. Once accounted for, this yields a total computational cost for improvement of *at least* k^4. That little 4 in the exponent is very expensive: A 10-fold improvement, for example, would require at least a 10,000-fold increase in computation.

To make the flexibility-computation trade-off more vivid, consider a scenario where you are trying to predict whether a patient's X-ray reveals cancer. Suppose further that the true answer can be found if you measure 100 details in the X-ray (often called variables or features). The challenge is that we don't know ahead of time which variables are important, and there could be a very large pool of candidate variables to consider.

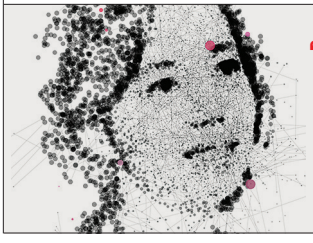
The expert-system approach to this problem would be to have people who are knowledgeable in radiology and oncology specify the variables they think are important, allowing the system to examine only those. The flexible-system approach is to test as many of the variables as possible and let the system figure out on its own which are important, requiring more data and incurring much higher computational costs in the process.

Models for which experts have established the relevant variables are able to learn quickly what values work best for those variables, doing so with limited amounts of computation—which is why they were so popular early on. But their ability to learn stalls if an expert hasn't correctly specified all the variables that should be included in the model. In contrast, flexible models like deep learning are less efficient, taking vastly more computation to match the performance of expert models. But, with enough computation (and data), flexible models can outperform ones for which experts have attempted to specify the relevant variables.

CLEARLY, YOU CAN GET improved performance from deep learning if you use more computing power to build bigger models and train them with more data. But how expensive will this computational burden become? Will costs become sufficiently high that they hinder progress?

To answer these questions in a concrete way, we recently gathered data from more than 1,000 research papers on deep learning, spanning the areas of image classification, object detection, question answering, named-entity recognition, and machine translation. Here, we will only discuss image classification in detail, but the lessons apply broadly.

Over the years, reducing image-classification errors has come with an enormous expansion in computational burden. For example, in 2012 AlexNet, the model that first showed the power of training deep-learning systems on graphics processing units (GPUs), was trained for five to six days using two GPUs. By 2018, another model, NASNet-A,



PORTRAIT BY SERGIO ALBIAC

"AI is fundamentally an applied technology that's going to serve our society. Humanistic AI not only raises the awareness of the importance of the technology, it's a really important way to attract diverse students, technologists, and innovators to participate."

FEI-FEI LI, codirector of the Stanford Institute for Human-Centered AI

had cut the error rate of AlexNet in half, but it used more than 1,000 times as much computing to achieve this.

Our analysis of this phenomenon also allowed us to compare what actually happened with theoretical expectations. Theory tells us that computing needs to scale with at least the fourth power of the improvement in performance. In practice, the actual requirements have scaled with at least the *ninth* power.

This ninth power means that to halve the error rate, you can expect to need more than 500 times the computational resources. That's a devastatingly high price. There may be a silver lining here, however. The gap between what's happened in practice and what theory predicts might mean that there are still undiscovered algorithmic improvements that could greatly improve the efficiency of deep learning.

As we noted, Moore's Law and other hardware advances have provided massive increases in chip performance. Does this mean that the escalation in computing requirements doesn't matter? Unfortunately, no. Of the 1,000-fold difference in the computing used by AlexNet and NASNet-A, only a sixfold improvement came from better hardware; the rest came from using more processors or running them longer, incurring higher costs.

Having estimated the computational cost-performance curve for image recognition, we can use it to estimate how much computation would be needed to reach even more impressive performance benchmarks in the future. For example, achieving a 5 percent error rate would require 10^{19} billion floating-point operations.

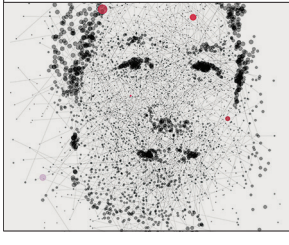
Important work by scholars at the University of Massachusetts Amherst allows us to understand the economic cost and carbon emissions implied by this computational burden. The answers are grim: Training such a model would cost US $100 billion and would produce as much carbon emissions as New York City does in a month. And if we estimate the computational burden of a 1 percent error rate, the results are considerably worse.

Is extrapolating out so many orders of magnitude a reasonable thing to do? Yes and no. Certainly, it is important to understand that the predictions aren't precise, although with such eye-watering results, they don't need to be to convey the overall message of unsustainability. Extrapolating this way *would* be unreasonable if we assumed that researchers would follow this trajectory all the way to such an extreme outcome. We don't. Faced with skyrocketing costs, researchers will either have to come up with more efficient ways to solve these problems, or they will abandon working on these problems and progress will languish.

On the other hand, extrapolating our results is not only reasonable but also important, because it conveys the magnitude of the challenge ahead. The leading edge of this problem is already becoming apparent. When Google subsidiary DeepMind trained its system to play Go, it was estimated to have cost $35 million. When DeepMind's researchers designed a system to play the *StarCraft II* video game, they purposefully didn't try multiple ways of architecting an important component, because the training cost would have been too high.

At OpenAI, an important machine-learning think tank, researchers recently designed and trained a much-lauded deep-learning language system called GPT-3 at the cost of more than $4 million. Even though they made a mistake when they implemented the system, they didn't fix it, explaining simply in a supplement to their scholarly publication that "due to the cost of training, it wasn't feasible to retrain the model."

Even businesses outside the tech industry are now starting to shy away from the computational expense of deep learning. A large European supermarket chain recently abandoned a deep-learning-based system that markedly improved its ability to predict which products would be purchased. The company executives dropped that attempt because they judged that the cost of training and running the system would be too high.



"Those of us in machine learning are really good at doing well on a test set, but unfortunately deploying a system takes more than doing well on a test set. All of AI...has a proof-of-concept-to-production gap." **ANDREW NG**, CEO and cofounder of Landing AI

PORTRAIT BY SERGIO ALBIAC

FACED WITH RISING economic and environmental costs, the deep-learning community will need to find ways to increase performance without causing computing demands to go through the roof. If they don't, progress will stagnate. But don't despair yet: Plenty is being done to address this challenge.

One strategy is to use processors designed specifically to be efficient for deep-learning calculations. This approach was widely used over the last decade, as CPUs gave way to GPUs and, in some cases, field-programmable gate arrays and application-specific ICs (including Google's Tensor Processing Unit). Fundamentally, all of these approaches sacrifice the generality of the computing platform for the efficiency of increased specialization. But such specialization faces diminishing returns. So longer-term gains will require adopting wholly different hardware frameworks—perhaps hardware that is based on analog, neuromorphic, optical, or quantum systems. Thus far, however, these wholly different hardware frameworks have yet to have much impact.

Another approach to reducing the computational burden focuses on generating neural networks that, when implemented, are smaller. This tactic lowers the cost each time you use them, but it often increases the training cost (what we've described so far in this article). Which of these costs matters most depends on the situation. For a widely used model, running costs are the biggest component of the total sum invested. For other models—for example, those that frequently need to be retrained—training costs may dominate. In either case, the total cost must be larger than just the training on its own. So if the training costs are too high, as we've shown, then the total costs will be, too.

And that's the challenge with the various tactics that have been used to make implementation smaller: They don't reduce training costs enough. For example, one allows for training a large network but penalizes complexity during training. Another involves training a large network and then "prunes" away unimportant connections. Yet another finds as efficient an architecture as possible by optimizing across many models—something called neural-architecture search. While each of these techniques can offer significant benefits for implementation, the effects on training are muted—certainly not enough to address the concerns we see in our data. And in many cases they make the training costs higher.

One up-and-coming technique that could reduce training costs goes by the name meta-learning. The idea is that the system learns on a variety of data and then can be applied in many areas. For example, rather than building separate systems to recognize dogs in images, cats in images, and cars in images, a single system could be trained on all of them and used multiple times.

We must either adapt how we do deep learning or face a future of much slower progress.

Unfortunately, recent work by Andrei Barbu of MIT has revealed how hard meta-learning can be. He and his coauthors showed that even small differences between the original data and where you want to use it can severely degrade performance. They demonstrated that current image-recognition systems depend heavily on things like whether the object is photographed at a particular angle or in a particular pose. So even the simple task of recognizing the same objects in different poses causes the accuracy of the system to be nearly halved.

Benjamin Recht of the University of California, Berkeley, and others made this point even more starkly, showing that even with novel data sets purposely constructed to mimic the original training data, performance drops by more than 10 percent. If even small changes in data cause large performance drops, the data needed for a comprehensive meta-learning system might be enormous. So the great promise of meta-learning remains far from being realized.

Another possible strategy to evade the computational limits of deep learning would be to move to other, perhaps as-yet-undiscovered or underappreciated types of machine learning. As we described, machine-learning systems constructed around the insight of experts can be much more computationally efficient, but their performance can't reach the same heights as deep-learning systems if those experts cannot distinguish all the contributing factors. Neuro-symbolic methods and other techniques are being developed to combine the power of expert knowledge and reasoning with the flexibility often found in neural networks.

Like the situation that Rosenblatt faced at the dawn of neural networks, deep learning is today becoming constrained by the available computational tools. Faced with computational scaling that would be economically and environmentally ruinous, we must either adapt how we do deep learning or face a future of much slower progress. Clearly, adaptation is preferable. A clever breakthrough might find a way to make deep learning more efficient or computer hardware more powerful, which would allow us to continue to use these extraordinarily flexible models. If not, the pendulum will likely swing back toward relying more on experts to identify what needs to be learned. ■

Deep Learning GOES TO BOOT CAMP

The U.S. Army wants to team humans and robots on the battlefield

BY EVAN ACKERMAN

"I SHOULD PROBABLY not be standing this close," I think to myself, as the robot slowly approaches a large tree branch on the floor in front of me. It's not the size of the branch that makes me nervous—it's that the robot is operating autonomously, and that while I know what it's *supposed* to do, I'm not entirely sure what it *will* do. If everything works the way the roboticists at the Army Research Laboratory (ARL) in Adelphi, Md., expect, the robot will identify the branch, grasp it, and drag it out of the way. These folks know what they're doing, but I've spent enough time around robots that I take a small step backward anyway. • The robot, named RoMan, for Robotic Manipulator, is about the size of a large lawn mower, with a tracked base that helps it handle most kinds of terrain. At the front, it has a squat torso equipped with cameras and depth sensors, as well as a pair of arms that were harvested from a prototype disaster-response robot originally developed at NASA's Jet Propulsion Laboratory (JPL) for a DARPA robotics competition. RoMan's job today is roadway clearing, a multistep task that ARL wants the robot to complete as autonomously as possible. Instead of instructing the robot to grasp specific objects in specific ways and move them to specific places, the operators tell RoMan

RoMan, the Army Research Laboratory's robotic manipulator, considers the best way to grasp and move a tree branch at the Adelphi Laboratory Center, in Maryland.

to "go clear a path." It's then up to the robot to make all the decisions necessary to achieve that objective.

The ability to make decisions autonomously is not just what makes robots useful, it's what makes robots *robots*. We value robots for their ability to sense what's going on around them, make decisions based on that information, and then take useful actions without our input. In the past, robotic decision making followed highly structured rules—if you sense this, then do that. In structured environments like factories, this works well enough. But in chaotic, unfamiliar, or poorly defined settings, reliance on rules makes robots notoriously bad at dealing with anything that cannot be precisely predicted and planned for in advance.

ROMAN, ALONG WITH many other types of robots including home vacuums, drones, and autonomous cars, handles the challenges of semistructured environments through artificial neural networks—a computing approach that loosely mimics the structure of neurons in biological brains. About a decade ago, artificial neural networks began to be applied to a wide variety of semistructured data that had previously been very difficult for computers running rules-based programming (generally referred to as symbolic reasoning) to interpret. Rather than recognizing specific data structures, an artificial neural network is able to recognize data patterns, identifying novel data that are similar (but not identical) to data that the network has encountered before. Indeed, part of the appeal of artificial neural networks is that they are trained by example, by letting the network ingest annotated data and learn its own system of pattern recognition. For neural networks with multiple layers of abstraction, this technique is called deep learning.

Even though humans are typically involved in the training process, and even though artificial neural networks were inspired by the neural networks in human brains, the kind of pattern recognition a deep-learning system does is fundamentally different from the way humans see the world. It's often nearly impossible to understand the relationship between the data input into the system and the interpretation of the data that the system outputs. And that difference—the "black box" opacity of deep learning—poses a potential problem for robots like RoMan and for ARL.

This opacity means that robots that rely on deep learning have to be used carefully. A deep-learning system is good at recognizing patterns, but lacks the world understanding that a human typically uses to make decisions, which is why such systems do best

In chaotic, unfamiliar, or poorly defined settings, reliance on rules makes robots notoriously bad at dealing with anything that cannot be precisely predicted and planned for in advance.

when their applications are well defined and narrow in scope. "When you have well-structured inputs and outputs, and you can encapsulate your problem in that kind of relationship, I think deep learning does very well," says Tom Howard, who directs the University of Rochester's Robotics and Artificial Intelligence Laboratory and has developed natural-language interaction algorithms for RoMan and other ground robots. "The question when programming an intelligent robot is, at what practical size do those deep-learning building blocks exist?" Howard explains that when you apply deep learning to higher-level problems, the number of possible inputs becomes very large, and solving problems at that scale can be challenging. And the potential consequences of unexpected or unexplainable behavior are much more significant when that behavior is manifested through a 170-kilogram two-armed military robot.

AFTER A COUPLE of minutes, RoMan hasn't moved—it's still sitting there, pondering the tree branch, arms poised like a praying mantis. For the

PREVIOUS PAGES: EVAN ACKERMAN

ARMY RESEARCH LABORATORY

Robots at the Army Research Lab test autonomous navigation techniques in rough terrain [opposite] with the goal of being able to keep up with their human teammates. ARL is also developing robots with manipulation capabilities [above] that can interact with objects so that humans don't have to.

last 10 years, the lab's Robotics Collaborative Technology Alliance (RCTA) has been working with roboticists from Carnegie Mellon University, Florida State University, General Dynamics Land Systems, JPL, MIT, QinetiQ North America, the University of Central Florida, the University of Pennsylvania, and other top research institutions to develop robot autonomy for use in future ground-combat vehicles. RoMan is one part of that process.

The "go clear a path" task that RoMan is slowly thinking through is difficult for a robot because the task is so abstract. RoMan needs to identify objects that might be blocking the path, reason about the physical properties of those objects, figure out how to grasp them and what kind of manipulation technique might be best to apply (like pushing, pulling, or lifting), and then make it happen. That's a lot of steps and a lot of unknowns for a robot with a limited understanding of the world.

This limited understanding is where the ARL robots begin to differ from other robots that rely on deep learning, says Ethan Stump, chief scientist of the AI for Maneuver and Mobility program at ARL. "The Army can be called upon to operate basically anywhere in the world. We do not have a mechanism for collecting data in all the different domains in which we might be operating. We may be deployed to some unknown forest on the other side of the world, but we'll be expected to perform just as well as we would in our own backyard," he says. Most deep-learning systems function reliably only within the domains and environments in which they've been trained. Even if the domain is something like "every drivable road in San Francisco," a robot will do fine, because that's a data set that has already been collected. But, Stump says, that's not an option for the military. If an Army deep-learning system doesn't perform well, they can't simply solve the problem by collecting more data.

ARL's robots also need to have a broad awareness of what they're doing. "In a standard operations order for a mission, you have goals, constraints, a paragraph on the commander's intent—basically a narrative of the purpose of the mission—which pro-

vides contextual info that humans can interpret and gives them the structure for when they need to make decisions and when they need to improvise," Stump explains. In other words, RoMan may need to clear a path quickly, or it may need to clear a path quietly, depending on the mission's broader objectives. That's a big ask for even the most advanced robot. "I can't think of a deep-learning approach that can deal with this kind of information," Stump says.

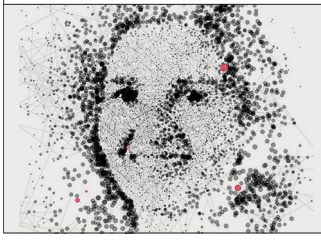
While I watch, RoMan is reset for a second try at branch removal. ARL's approach to autonomy is modular, where deep learning is combined with other techniques, and the robot is helping ARL figure out which tasks are appropriate for which techniques. At the moment, RoMan is testing two different ways of identifying objects from 3D sensor data: UPenn's approach is based on deep learning, while Carnegie Mellon is using a method called perception through search, which relies on a more traditional database of 3D models. Perception through search works only if you know exactly which objects you're looking for in advance, but training is much faster since you need only a single model per object. It can also be more accurate when perception of the object is difficult—if the object is partially hidden or upside-down, for example. ARL is testing these strategies to determine which is the most versatile and effective, letting them run simultaneously and compete against each other.

PERCEPTION IS ONE of the things that deep learning tends to excel at. "The computer vision community has made crazy progress using deep learning for this stuff," says Maggie Wigness, a computer scientist at ARL. "We've had good success with some of these models that were trained in one environment generalizing to a new environment, and we intend to keep using deep learning for these sorts of tasks, because it's the state of the art."

ARL's modular approach might combine several techniques in ways that leverage their particular strengths. For example, a perception system that uses deep-learning-based vision to classify terrain could work alongside an autonomous driving system based on an approach called inverse reinforcement learning, where the model can rapidly be created or refined by observations from human soldiers. Traditional reinforcement learning optimizes a solution based on established reward functions, and is often applied when you're not necessarily sure what optimal behavior looks like. This is less of a concern for the Army, which can generally assume that well-trained humans will be nearby to show a robot the right way to do things. "When we deploy these robots, things can change very quickly," Wigness says. "So we wanted a technique where we could have a soldier intervene, and with just a few examples from a user in the field, we can update the system if we need a new behavior." A deep-learning technique would require "a lot more data and time," she says.

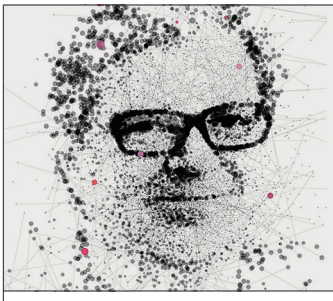
It's not just data-sparse problems and fast adaptation that deep learning struggles with. There are also questions of robustness, explainability, and safety. "These questions aren't unique to the military," says Stump, "but it's especially important when we're talking about systems that may incorporate lethality." To be clear, ARL is not currently working on lethal autonomous weapons systems, but the lab is helping to lay the groundwork for autonomous systems in the U.S. military more broadly, which means considering ways in which such systems may be used in the future.

Safety is an obvious priority, and yet there isn't a clear way of making a deep-learning system verifiably safe, according to Stump. "Doing deep learning with safety constraints is a major research effort. It's hard to add those constraints into the system, because you don't know where the constraints already in the system came from. So when the mission changes, or the context changes, it's hard to deal with that. It's not even a data question; it's an architecture question." In ARL's modular architecture, whether it's a perception module that uses deep learning or an autonomous driving module that uses inverse reinforcement learning or something else, such pieces can form parts of a broader autonomous system that incorporates the kinds of safety and



"Dig into every industry and you'll find AI changing the nature of work." **DANIELA RUS**, director of MIT's Computer Science & Artificial Intelligence Laboratory

PORTRAITS BY SERGIO ALBIAC



"What's missing is a principle that would allow our machine to learn how the world works by observation and by interaction with the world. A learning predictive world model is what we're missing today, and in my opinion is the biggest obstacle to significant progress in AI." YANN LECUN, professor at New York University and chief AI scientist at Facebook

adaptability that the military requires. Other modules in the system can operate at a higher level, using different techniques that are more verifiable or explainable and that can step in to protect the overall system from adverse unpredictable behaviors. "If other information comes in and changes what we need to do, there's a hierarchy there," Stump says. "It all happens in a rational way."

Nicholas Roy, who leads the Robust Robotics Group at MIT and describes himself as "somewhat of a rabble-rouser" due to his skepticism of some of the claims made about the power of deep learning, agrees with the ARL roboticists that deep-learning approaches often can't handle the kinds of challenges that the Army has to be prepared for. "The Army is always entering new environments, and the adversary is always going to be trying to change the environment so that the training process the robots went through simply won't match what they're seeing," Roy says. "So the requirements of a deep network are to a large extent misaligned with the requirements of an Army mission, and that's a problem."

Roy, who has worked on abstract reasoning for ground robots as part of the RCTA, emphasizes that deep learning is a useful technology when applied to problems with clear functional relationships, but when you start looking at abstract concepts, it's not clear whether deep learning is a viable approach. "I'm very interested in finding how neural networks and deep learning could be assembled in a way that supports higher-level reasoning," Roy says. "I think it comes down to the notion of combining multiple low-level neural networks to express higher-level concepts, and I do not believe that we understand how to do that yet." Roy gives the example of using two separate neural networks, one to detect objects that are cars and the other to detect objects that are red. It's harder to combine those two networks into one larger network that detects red cars than it would be if you were using a symbolic reasoning system based on structured rules with logical relationships. "Lots of people are working on this, but I haven't seen a real success that drives abstract reasoning of this kind," he says.

"The requirements of a deep network are to a large extent misaligned with the requirements of an Army mission, and that's a problem." NICHOLAS ROY

FOR THE FORESEEABLE FUTURE, ARL is making sure that its autonomous systems are safe and robust by keeping humans around for both higher-level reasoning and occasional low-level advice. Humans might not be directly in the loop at all times, but the idea is that humans and robots are more effective when working together as a team. When the most recent phase of the RCTA program began in 2009, ARL's Ethan Stump says, "we'd already had many years of being in Iraq and Afghanistan, where robots were often used as tools. We've been trying to figure out what we can do to transition robots from tools to acting more as teammates within the squad."

RoMan gets a little bit of help when a human supervisor points out a region of the branch where grasping might be most effective. The robot doesn't have any fundamental knowledge about what a tree branch actually is, and this lack of world knowledge (what we think of as common sense) is a fundamental problem with autonomous systems of all kinds. Having a human leverage our vast experience into a small amount of guidance can make RoMan's job much easier. And indeed, this time RoMan manages to successfully grasp the branch and noisily haul it across the room.

Turning a robot into a good teammate can be difficult, because it can be tricky to find the right amount of autonomy. Too little and it would take most or all of the focus of one human to manage one robot, which may be appropriate in special situations like explosive-ordnance disposal but is otherwise not efficient. Too much autonomy and you'd start to have issues with trust, safety, and explainability.

"I think the level that we're looking for here is for robots to operate on the level of working dogs," explains Stump. "They understand exactly what we need them to do in limited circumstances, they have a small amount of flexibility and creativity if they

are faced with novel circumstances, but we don't expect them to do creative problem-solving. And if they need help, they fall back on us."

ROMAN IS NOT LIKELY to find itself out in the field on a mission anytime soon, even as part of a team with humans. It's very much a research platform. But the software being developed for RoMan and other robots at ARL, called Adaptive Planner Parameter Learning (APPL), will likely be used first in autonomous driving, and later in more complex robotic systems that could include mobile manipulators like RoMan. APPL combines different machine-learning techniques (including inverse reinforcement learning and deep learning) arranged hierarchically underneath classical autonomous navigation systems. That allows high-level goals and constraints to be applied on top of lower-level programming. Humans can use teleoperated demonstrations, corrective interventions, and evaluative feedback to help robots adjust to new environments, while the robots can use unsupervised reinforcement learning to adjust their behavior parameters on the fly. The result is an autonomy system that enjoys many of the benefits of machine learning, while also providing the kind of safety and explainability that the Army needs. With APPL, a learning-based system like RoMan can operate in predictable ways even under uncertainty, falling back on human tuning or human demonstration if it ends up in an environment that's too different from what it trained on.

It's tempting to look at the rapid progress of commercial and industrial autonomous systems (autonomous cars being just one example) and wonder why the Army seems to be somewhat behind the state of the art. But as Stump finds himself having to explain to Army generals, when it comes to autonomous systems, "there are lots of hard problems, but industry's hard problems are different from the Army's hard problems." The Army doesn't have the luxury of operating its robots in structured environments with lots of data, which is why ARL has put so much effort into APPL, and into maintaining a place for humans. Going forward, humans are likely to remain a key part of the autonomous framework that ARL is developing. "That's what we're trying to build with our robotics systems," Stump says. "That's our bumper sticker: 'From tools to teammates.' " ■

UCF

Stands For Opportunity

University of Central Florida
Department of Electrical & Computer Engineering

The University of Central Florida (UCF) has established several interdisciplinary clusters to strengthen its academic offerings and research mission. Accordingly, we are recruiting one tenured/tenure-track Assistant or Associate Professor for the university's research cluster on Resilient, Intelligent and Sustainable Energy Systems (RISES) with a start date of August 2022. Ideal candidates will have research impact, as reflected in high-quality publications and the ability to build a funded and sustainable research program. They work at the intersection of several areas such as Battery Storage Technologies, Renewable Electrolysis, Grid Integration, and/or Energy Storage Analysis. Prior work experience with energy storage systems is highly preferred. All relevant technical areas will be considered.

This is an interdisciplinary position that is expected to strengthen both the RISES cluster and the chosen tenure home department. A strong advantage of this position is the ability of the candidate to choose multiple units for their appointment in the College of Engineering and Computer Science, the College of Sciences, or both. The position will carry a rank commensurate with the candidate's prior experience and record.

Questions regarding this search can be directed to ECE-FacultySearch@cecs.ucf.edu. Please apply at: **https://jobs.ucf.edu/en-us/job/500854/assistant-professor-or-associate-professor-rises-university-research-center**.

UCF is an equal opportunity/affirmative action employer. All qualified applicants are encouraged to apply, including women, veterans, individuals with disabilities and members of traditionally underrepresented populations. As a Florida public university, UCF's Equal Opportunity Statement can be viewed at: ***http://www.oie.ucf.edu/documents/PresidentsStatement.pdf****. UCF makes all application materials and selection procedures available to the public upon request.*

COLLEGE OF ENGINEERING
ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY OF MICHIGAN

THE ELECTRICAL AND COMPUTER ENGINEERING (ECE)Division of the Electrical Engineering and Computer Science Department at the University of Michigan, Ann Arbor invites applications for junior and senior faculty positions to help enact our new strategic plan (**www.ece.umich.edu**).

Successful candidates will have a relevant doctorate or equivalent experience and an outstanding record of research in academia, industry and/or at national laboratories. They will have a strong commitment to teaching at undergraduate and graduate levels, to providing service to the university and profession, and to broadening the intellectual diversity of the ECE Division; and have expansive world-views on the potential impact of their research.

We invite diverse candidates across all research areas to apply. The highly ranked ECE Division prides itself on the mentoring of junior faculty toward successful careers. Ann Arbor is highly rated as a family friendly best-place-to-live.

Please see application instructions at:
https://ece.engin.umich.edu/people/faculty-positions/

Applications will be reviewed as they are received. The application site will remain open until **15 January 2022.**

The University of Michigan is an Affirmative Action, Equal Opportunity Employer with an Active Dual-Career Assistance Program. The College of Engineering is especially interested in candidates who contribute, through their research, teaching, and/or service, to the diversity and excellence of the academic community.

Tenure Track Assistant Professor

RIT

The Department of Electrical and Microelectronic Engineering at the Rochester Institute of Technology invites applications from candidates in the area of analog & mixed signal integrated circuits and systems. Applicants must have a Ph.D. in Electrical, Microelectronic and/or Computer Engineering or a closely related field. The candidate's research must specialize in analog & mixed signal integrated circuits and systems, with a record of refereed publication in the area. Candidates must have prior teaching experience; a potential for establishing and conducting sponsored research; excellent written and oral communication skills; and ability to contribute in meaningful ways to the university's continuing commitment to cultural diversity, pluralism and individual differences. Faculty responsibilities include teaching at both the undergraduate and graduate levels, conducting sponsored research programs, and providing service to the university community.

Apply online at http://careers.rit.edu/faculty. Keyword Search: 5966BR. Review of applications will begin immediately and continue until a suitable candidate is found. Inquiries may be sent to Dr. Stefan Preble, sfpeen@rit.edu.

RIT does not discriminate. RIT is an equal opportunity employer that promotes and values diversity, pluralism, and inclusion. For more information or inquiries, please visit RIT/TitleIX or the U.S. Department of Education at ED.Gov.

ARIZONA

DEPARTMENT HEAD

Electrical and Computer Engineering
The University of Arizona
Tucson, AZ

Position Summary
The Department of Electrical and Computer Engineering at the University of Arizona is seeking nominations and applications for a department head with excellent leadership skills and enthusiasm for world-class research, innovative teaching, and industry/government collaborations. The future ECE Department Head must be an innovative and visionary academic leader who can spearhead the continuing transformation of ECE for the challenges of the 21st century. The Department Head is responsible for leading a dynamic department which includes a diverse group of faculty, staff, and students, as well as modern teaching and research laboratories and other facilities. Successful candidates are expected to have extensive research and teaching experience at a level sufficient to qualify for appointment as a tenured full professor. This position carries the possibility of an endowed professor title.

Professional Qualifications and Personal Qualities
The successful candidate will have a distinguished record of achievement in scholarship, research and/or professional practice commensurate with an appointment at the rank of full professor with tenure. The candidate should also demonstrate effective managerial leadership; clear communication; and a commitment to shared governance, community engagement, and diversity, equity and inclusion. The successful candidate must demonstrate high ethical standards and is expected to operate in a transparent and collegial way.

Additional information about the position, department, and application package is available at **ece.engineering.arizona.edu/faculty-staff/open-positions**.

The University of Arizona provides equal employment opportunities to applicants and employees without regard to race, color, religion, sex, national origin, age, disability, veteran status, sexual orientation, gender identity, or genetic information.

Past Forward

NATIONAL POSTAL MUSEUM/ SMITHSONIAN INSTITUTION

Anthrax by Mail

Twenty years ago this month, fear gripped Americans as the threat of bioterrorism became real. Five people died and 18 more became seriously ill when anthrax spores were sent in letters to high-profile targets in politics and the media. The technological solution: Irradiate all mail destined for certain regions. Passing letters and packages through a high-energy beam of ionizing radiation did kill the harmful bacteria, but not without consequences. It made envelopes brittle and discolored, warped plastics, exposed film, fogged glass products, weakened the potency of pharmaceuticals, and destroyed biological samples from doctors' offices and scientific labs. Today the U.S. Postal Service continues to sanitize mail bound for federal buildings. To avoid any damage to incoming artifacts intended for its collections, the Smithsonian Institution—a quasi-government entity—uses alternative addresses. ■

FOR MORE ON THE HISTORY OF IRRADIATED MAIL, SEE spectrum.ieee.org/pastforward-oct2021

IEEE Member Group Disability Income Insurance.

Keep the ability to make ends meet,
even if you lose the ability to work.

To learn more*, visit IEEEinsurance.com/Lift

Group Disability Income Insurance is available only for residents of the U.S. (except VT and territories), Puerto Rico and Canada (except Quebec). Underwritten by New York Life Insurance Company, 51 Madison Ave., New York, NY 10010 on Policy Form GMR. It is available to residents of Canada (except Quebec). Mercer (Canada) Limited, represented by its employees Nicole Swift and Pauline Tremblay, acts as a broker with respect to residents of Canada.

*For information on features, costs, eligibility, renewability, limitations and exclusions visit IEEEinsurance.com/Lift.

Program Administered by Mercer Health & Benefits Administration LLC
In CA d/b/a Mercer Health & Benefits Insurance Services LLC
AR Insurance License #100102691 | CA Insurance License #0G39709

IEEE

MathWorks®

MATLAB SPEAKS DEEP LEARNING

With MATLAB,® you can build and deploy deep learning models for signal processing, reinforcement learning, automated driving, and other applications. Preprocess data, train models, generate code for GPUs, and deploy to production systems.

mathworks.com/deeplearning

Semantic segmentation for wildlife conservation.