

**The Problem With Digital Voiceprints** › Nabbing criminals using their voices  
P. 28

**Can AI-Generated Code Be Trusted?** › Spotting its flaws takes a keen eye  
P. 34

**RoboCup at 25** › Advancing robotics, training roboticists  
P. 40

FOR THE  
TECHNOLOGY  
INSIDER  
JULY 2023

# IEEE Spectrum

The  
Incredible  
Shrinking  
Lens

Metalenses  
transform  
cameras  
and projectors



# High Impedance or High Throughput?

Impedance analyzers and LCR meters for  
your component manufacturing workflow



MFIA Impedance Analyzer



R&S® LCX LCR-Meter

Materials  
science

Research &  
development

Characterization

Quality  
assurance

Production &  
test

Watch now to find out more

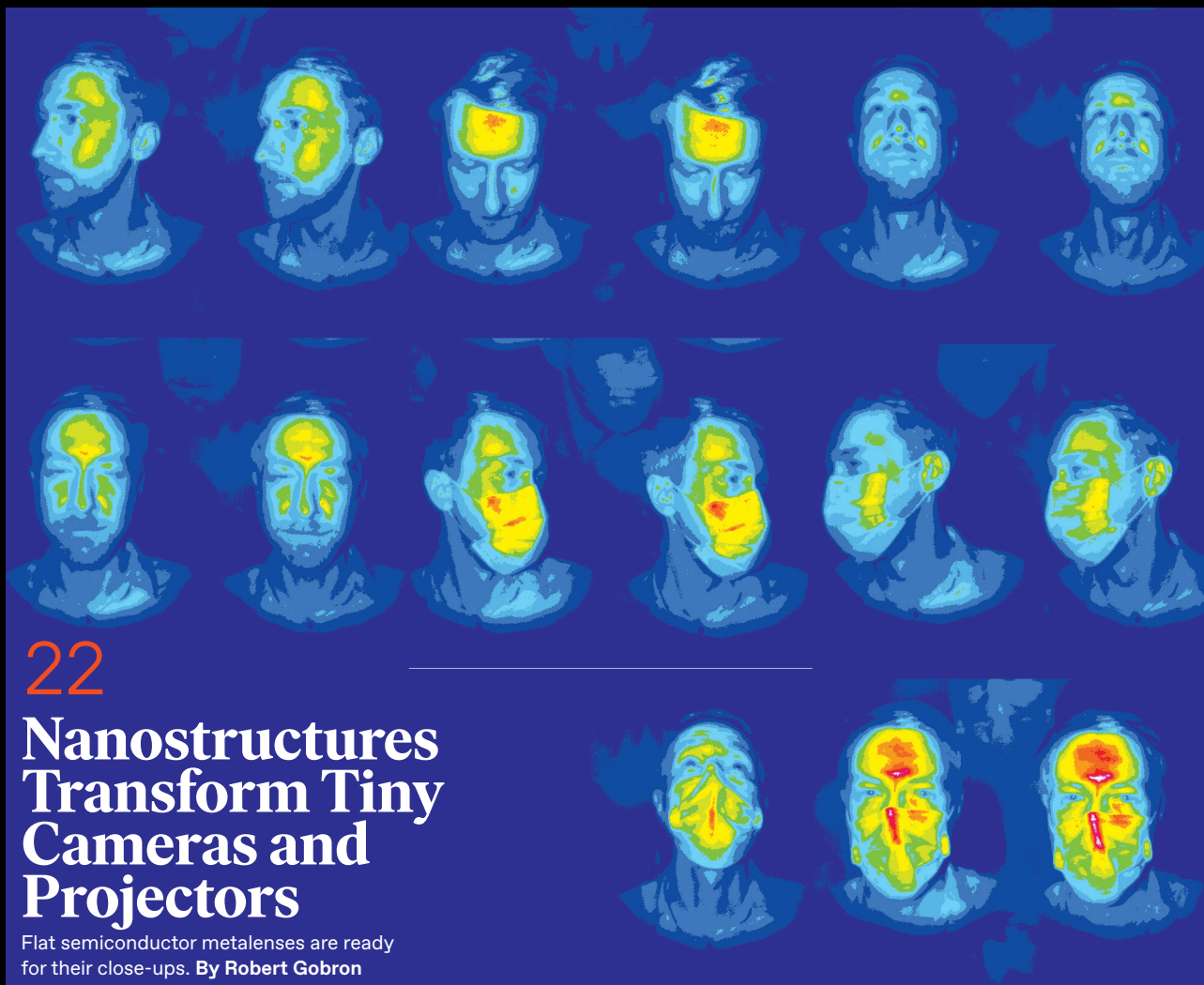


**ROHDE & SCHWARZ**  
Make ideas real





# Contents



22

## Nanostructures Transform Tiny Cameras and Projectors

Flat semiconductor metalenses are ready for their close-ups. **By Robert Gobron**

### Guided by Voices

Digital voiceprinting is back, but will the courts accept it?

**By Peter Andrey Smith**

28

### A Critical Look at AI-Generated Software

ChatGPT and other AI coding assistants hold both promise and peril.

**By Jaideep Vaidya  
& Hafiz Asif**

34

#### ON THE COVER:

Illustration by Chad Hagen

### The Future of RoboCup

Robots will challenge world-champion humans in 2050.

**By Peter Stone**

40

#### EDITOR'S NOTE

For the love of bots.

2

#### NEWS

Ukraine's War Drones  
Kilowatt Solar-to-Hydrogen  
Time-Reversing Mirror  
Large Language Models Shrink

6

#### HANDS ON

Record some mesmerizing views of the sky.

16

#### CAREERS

Sushama Chakraverty designs the fun in video games.

19

#### 5 QUESTIONS

Sam Naffziger on Advanced Micro Devices' chiplet push

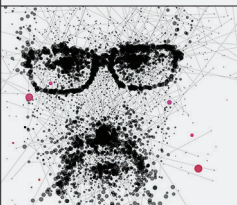
21

#### PAST FORWARD

Beyond the Mechanical Turk

52





# Kicking It With Robots

RoboCup and the new Robots Guide are gateways to the wonderful world of robots

**I**n July of 2010, I traveled to Singapore to take care of my then 6-year-old son Henry while his mother attended an academic conference. But I was really there for the robots.

IEEE Spectrum's digital product manager, Erico Guizzo, was our robotics editor at the time. We had just combined forces with robot blogger *par excellence* and now *Spectrum* senior editor Evan "BotJunkie" Ackerman to supercharge our first and most successful blog, Automaton. When I told Guizzo I was going to be in Singapore, he told me that RoboCup, an international robot soccer competition, was going on at the same time. So of course we wrangled a press pass for me and my plus one.

I brought Henry and a video camera to capture the bustling bots and their handlers. Guizzo told me that videos of robots flailing at balls would do boffo Web traffic, so I was as excited as my first grader (okay, more excited) to be in a convention center filled with robots and teams of engineers toiling away on the sidelines to make adjustments and repairs and talk with each other and us about their creations.

Even better than the large humanoid robots lurching around like zombies and the smaller, wheeled bots scurrying to and fro were the humans who tended to them. They exuded the kind of joy that comes with working together to build cool stuff, and it was infectious. On page 40 of this issue, Peter Stone—past president of the RoboCup Federation, professor in the computer science department of the University of Texas at Austin, and executive director of Sony AI America—captures some of that unbridled enthusiasm and gives us the history of the event. To go along with his story, we include action shots taken at various RoboCups throughout the 25 years of the event. You can check out this year's RoboCup com-

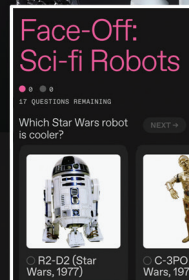
## ROBOTS

Search Robots Rankings Learn Play News

### GITAI G1

GITAI G1 is a general-purpose humanoid robot for space applications. It combines AI and teleoperation to perform tasks like manipulation of tools, control boards, and scientific equipment. It's designed to operate inside and outside space stations and future lunar bases.

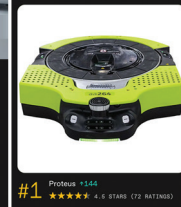
CREATOR  
GITAI



### Robot Rankings

UPDATED EVERY FRIDAY

TOP RATED ROBOTS



petitions going on 6–9 July at the University of Bordeaux, in Nouvelle-Aquitaine, France.

Earlier in 2010, the same year as my first RoboCup, Apple introduced what was in part pitched as the future of magazines: the iPad. Guizzo and photography director Randi Klett instantly grokked the possibilities of the format and the new sort of tactile interactivity (ah, the swipe!) to showcase the coolest robots they could find. Channeling the same spirit I experienced in Singapore, Guizzo, Klett, and the app-maker Tendigi launched the Robots app in 2012. It was an instant hit, with more than 1.3 million downloads.

To reach new audiences on other devices beyond the iOS platform, we ported Robots from appworld to the Web. With the help of founding sponsors—including the IEEE Robotics and Automation Society and Walt Disney Imagineering—and the support of the IEEE Foundation, the Robots site launched in 2018 and quickly found a following among STEM educators, students, roboticists, and the general public.

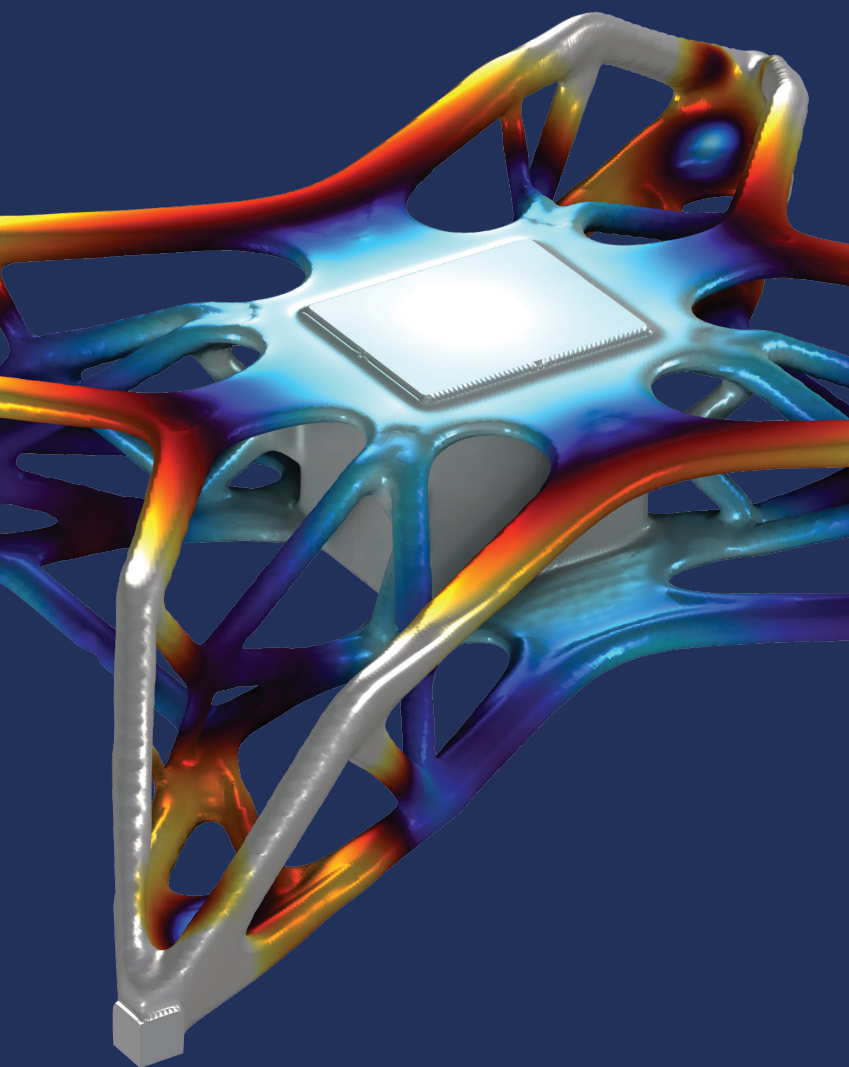
By 2022 it was clear that the site, whose basic design had not changed in years, needed a reboot. We gave it a new name and URL to make it easy for more people to find: RobotsGuide.com. And with the help of Pentagram, the design consultancy that reimaged *Spectrum's* print magazine and website in 2021, in collaboration with Standard, a design and technology studio, we built the site as a modern, fully responsive Web app.

Featuring almost 250 of the world's most advanced and influential robots, hundreds of photos and videos, detailed specs, 360-degree interactives, games, user ratings, educational content, and robot news from around the world, the Robots Guide helps everyone learn more about robotics. ■



# Simulate real-world designs, devices, and processes with COMSOL Multiphysics®

[comsol.com/feature/multiphysics-innovation](https://comsol.com/feature/multiphysics-innovation)



## Innovate faster.

Test more design iterations before prototyping.

---

## Innovate smarter.

Analyze virtual prototypes and develop a physical prototype only from the best design.

---

## Innovate with multiphysics simulation.

Base your design decisions on accurate results with software that lets you study unlimited multiple physical effects on one model.



## ● BOB GOBRON

Gobron is an applications engineer with Metalenz, whose metasurface optical technology he describes in this issue [p. 22]. His passion for optics began when his high school physics teacher brought a telescope out to the football field one night after practice to show his students the stars. Gobron has since worked in optical engineering and design for LED lighting, avionics, laser engraving, and chemical detection based on quantum-cascade lasers.

## ● CHAD HAGEN

Hagen, the designer and illustrator who created our cover art, has a long list of well-known publishing and corporate clients. But he's probably best known for his series of colorful "nonsensical infographics," which playfully invert an infographic's purpose by their total lack of data. Hagen acknowledges that he sometimes has to work on "projects that are meaningful and provide good information," but he likes to "flip the 'form follows function' mantra" whenever he gets the chance.

## ● PETER ANDREY SMITH

Smith, based in Maine, has written for *Undark*, *The New York Times Magazine*, *Wired*, and WNYC's "Radiolab." As a reporter, he keeps tabs on the use of questionable scientific evidence in the courts. Upon learning that federal investigators had resurrected a once-discredited technique known as voiceprinting to solve a case involving hoax mayday calls, he knew he had to investigate. For more on the technique, which is still in dispute, see Smith's article on page 28.

## ● JAIDEEP VAIDYA & HAFIZ ASIF

Vaidya, a professor at Rutgers, the State University of New Jersey, works at the business school, where he teaches courses on data privacy and cybersecurity. His coauthor, Hafiz Asif, currently a postdoctoral researcher at Rutgers, will be joining the faculty of Hofstra University. In discussing the dangers of AI-based systems for writing code [p. 34], they note the relevance of an adage about computing: "garbage in, garbage out." Here, though, "the garbage looks very realistic," says Vaidya.

# IEEE Spectrum

**EDITOR IN CHIEF** Harry Goldstein, h.goldstein@ieee.org

**EXECUTIVE EDITOR** Jean Kumagai, j.kumagai@ieee.org

**MANAGING EDITOR** Elizabeth A. Bretz, e.bretz@ieee.org

**CREATIVE DIRECTOR**  
Mark Montgomery, m.montgomery@ieee.org

**PRODUCT MANAGER, DIGITAL** Erico Guizzo, e.guizzo@ieee.org

**EDITORIAL DIRECTOR, CONTENT DEVELOPMENT**  
Glenn Zorpette, g.zorpette@ieee.org

**SENIOR EDITORS**  
Evan Ackerman (Digital), ackerman.e@ieee.org  
Stephen Cass (Special Projects), cass.s@ieee.org  
Samuel K. Moore, s.k.moore@ieee.org  
Tekla S. Perry, tperry@ieee.org  
Philip E. Ross, p.ross@ieee.org  
David Schneider, d.a.schneider@ieee.org  
Eliza Strickland, e.strickland@ieee.org

**ART & PRODUCTION**  
**DEPUTY ART DIRECTOR** Brandon Palacio, b.palacio@ieee.org  
**PHOTOGRAPHY DIRECTOR** Randi Klett, randi.klett@ieee.org  
**ONLINE ART DIRECTOR** Erik Vrieling, e.vrieling@ieee.org  
**PRINT PRODUCTION SPECIALIST**  
Sylvana Meneses, s.meneses@ieee.org  
**MULTIMEDIA PRODUCTION SPECIALIST**  
Michael Spector, m.spector@ieee.org

**NEWS MANAGER** Margo 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  
**CONTRACT SPECIALIST** Ramona L. Foster, r.foster@ieee.org

**CONTRIBUTING EDITORS** Robert N. Charette, Steven Cherry, Charles Q. Choi, Peter Fairley, Edd Gent, W. Wayt Gibbs, Mark Harris, Allison Marsh, Prachi Patel, Julianne Pepitone, Lawrence Ulrich, Emily Waltz

**THE INSTITUTE**  
**EDITOR IN CHIEF** Kathy Pretz, k.pretz@ieee.org  
**ASSOCIATE EDITOR** Joanna Goodrich, j.goodrich@ieee.org  
**DIRECTOR, PERIODICALS PRODUCTION SERVICES** Peter Tuohy

**ADVERTISING PRODUCTION MANAGER**  
Felicja Spagnoli, f.spagnoli@ieee.org

**ADVERTISING PRODUCTION** +1 732 562 6334

**EDITORIAL ADVISORY BOARD, IEEE SPECTRUM**  
Harry Goldstein, *Chair*; Robert Schober, Ella M. Atkins, Sangyeun Cho, Francis J. "Frank" Doyle III, Hugh Durrant-Whyte, Matthew Eisler, Shahin Farshchi, Alissa Fitzgerald, Benjamin Gross, Lawrence O. Hall, Daniel Hissel, Jason K. Hui, Michel M. Maharbiz, Somdeb Majumdar, Lisa May, Carmen S. Menoni, Ramune Nagisetty, Paul Nielsen, Sofia Olhede, Christopher Stiller, Mini S. Thomas, Wen Tong, Haifeng Wang, Boon-Lock Yeo

**EDITORIAL ADVISORY BOARD, THE INSTITUTE**  
Kathy Pretz, *Chair*; Qusi Alqarqaz, Stamatis Dragoumanos, Jonathan Garibaldi, Madeleine Glick, Lawrence O. Hall, Harry Goldstein, Francesca Iacopi, Cecilia Metra, Shashi Raj Pandey, John Purvis, Chenyang Xu

**MANAGING DIRECTOR, PUBLICATIONS** Steven Heffner

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

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

**ADVERTISING INQUIRIES** Naylor Association Solutions, Erik Albin **+1 352 333 3371**, ealbin@naylor.com

**REPRINT SALES** +1 212 221 9595, ext. 319

**REPRINT PERMISSION / LIBRARIES** Articles may be photocopied for private use of patrons. A per-copy fee must be paid to the Copyright Clearance Center, 29 Congress St., Salem, MA 01970. For other copying or republication, contact Managing Editor, *IEEE Spectrum*.

**COPYRIGHTS AND TRADEMARKS**  
*IEEE Spectrum* is a registered trademark owned by The Institute of Electrical and Electronics Engineers Inc. Responsibility for the substance of articles rests upon the authors, not IEEE, its organizational units, or its members. Articles do not represent official positions of IEEE. Readers may post comments online; comments may be excerpted for publication. IEEE reserves the right to reject any advertising.



## IEEE BOARD OF DIRECTORS

**PRESIDENT & CEO** Saifur Rahman, president@ieee.org  
**+1 732 562 3928 Fax: +1 732 981 9515**  
**PRESIDENT - ELECT** Thomas M. Coughlin  
**TREASURER** Mary Ellen Randall  
**SECRETARY** Forrest "Don" Wright  
**PAST PRESIDENT** K.J. Ray Liu

## VICE PRESIDENTS

Rabab Ward, *Educational Activities*; Sergio Benedetto, *Publication Services & Products*; Jill I. Gostin, *Member & Geographic Activities*; John P. Verboncoeur, *Technical Activities*; Yu Yuan, *President, Standards Association*; Eduardo F. Palacio, *President, IEEE-USA*

## DIVISION DIRECTORS

Franco Maloberti (I); Kevin L. Peterson (II); Khaled Ben Letaief (III); Alistair P. Duffy (IV); Cecilia Metra (V); Kamal Al-Haddad (VI); Claudio Cañizares (VII); Leila De Floriani (VIII); Ali H. Sayed (IX); Stephanie M. White (X)

## REGION DIRECTORS

Greg T. Gdowski (I); Andrew D. Lowery (2); Theresa A. Brunasso (3); Vickie A. Ozburn (4); Bob G. Beenal (5); Kathy Hayashi (6); Robert L. Anderson (7); Vincenzo Piuri (8); Enrique A. Tejera (9); ChunChe (Lance) Fung (10)

**DIRECTOR EMERITUS** Theodore W. Hissey

## IEEE STAFF

**EXECUTIVE DIRECTOR & COO** Sophia A. Muirhead  
**+1 732 562 5400**, s.muirhead@ieee.org  
**CHIEF INFORMATION OFFICER** Cherif Amirat  
**+1 732 562 6017**, c.amirat@ieee.org  
**CHIEF HUMAN RESOURCES OFFICER** Liesel Bell  
**+1 732 562 6347**, l.bell@ieee.org  
**GENERAL COUNSEL & CHIEF COMPLIANCE OFFICER**  
Anta Cissé-Green **+1 212 705 8927**, a.cisse-green@ieee.org  
**CHIEF MARKETING OFFICER** Karen L. Hawkins  
**+1 732 562 3964**, k.hawkins@ieee.org  
**PUBLICATIONS** Steven Heffner  
**+1 212 705 8958**, s.heffner@ieee.org  
**CORPORATE ACTIVITIES** Donna Hourican  
**+1 732 562 6330**, d.hourican@ieee.org  
**MEMBER & GEOGRAPHIC ACTIVITIES** Cecelia Jankowski  
**+1 732 562 5504**, c.jankowski@ieee.org  
**STANDARDS ACTIVITIES** Konstantinos Karachalios  
**+1 732 562 3820**, constantin@ieee.org  
**EDUCATIONAL ACTIVITIES** Jamie Moesch  
**+1 732 562 5514**, j.moesch@ieee.org  
**CHIEF FINANCIAL OFFICER** Thomas R. Siegert  
**+1 732 562 6843**, t.siegert@ieee.org  
**TECHNICAL ACTIVITIES** Mary Ward-Callan  
**+1 732 562 3850**, m.ward-callan@ieee.org  
**MANAGING DIRECTOR, IEEE-USA** Russell T. Harrison  
**+1 202 530 8326**, r.t.harrison@ieee.org

## IEEE PUBLICATION SERVICES & PRODUCTS BOARD

Sergio Benedetto, *Chair*; Stefano Galli, Maria Sabrina Greco, Lawrence O. Hall, James Irvine, Charles M. Jackson, Clem Karl, Yong Lian, Fabrizio Lombardi, Aleksandar Mastilovic, Anna Scaglione, Gaurav Sharma, Isabel Trancoso, Peter Winzer, Weihua Zhuang

## IEEE OPERATIONS CENTER

445 Hoes Lane, Box 1331, Piscataway, NJ 08854-1331 U.S.A.  
**Tel: +1 732 981 0060 Fax: +1 732 981 1721**

**IEEE SPECTRUM** (ISSN 0018-9235) is published monthly by The Institute of Electrical and Electronics Engineers, Inc. All rights reserved. © 2023 by The Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, New York, NY 10016-5997, U.S.A. Volume No. 60, Issue No. 7. 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 #125634198. Printed at 3401 N. Heartland Dr, Liberty, MO 64068, U.S.A. Printed at Spenta Multimedia Pvt. Ltd., Plot 15, 16 & 21/1, Village Chikhli, Morivali, MIDC, Ambernath (West), Dist. Thane, 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/about/corporate/governance/p9-26.html>.





**YOU**  
**JOINED  
THE  
GROUP**



**YOU**  
**HAVE  
EXCLUSIVE  
ADVANTAGES**

Visit [ieee.org/discounts](http://ieee.org/discounts) to access your exclusive member savings.

# News

## AEROSPACE

### Budget Drones in Ukraine Are Redefining Warfare > Instead of multimillion-dollar aircraft, near-disposable fliers now patrol Ukrainian skies

BY PHILIP E. ROSS

**T**he war between Russia and Ukraine is making a lot of high-tech military systems look like so many gold-plated irrelevancies. That's why both sides are relying increasingly on low-tech alternatives—dumb artillery shells instead of pricey missiles, and drones instead of fighter aircraft.

“This war is a war of drones, they are the super weapon here,” Anton Gerashchenko, an adviser to Ukraine's minister of internal affairs, told *Newsweek* earlier this year.

In early May, Russia attributed explosions at the Kremlin to drones sent by Ukraine for the purpose of assassinating Vladimir Putin, the Russian leader. Ukraine denied the allegation. True, the mission to Moscow was ineffectual, but it is amazing that it could be managed at all.

Drones fly slower than an F-35, carry a smaller payload, beckon ground fire, and last mere days before being shot out of the skies. But for the most part, the price is right: China's DJI Mavic 3, used by both Russia and Ukraine for surveillance and for delivering bombs, goes for around US \$2,000. You can get 55,000 of them for the price of a single F-35. Also, they're much easier to maintain: When they



A Ukrainian soldier equips a drone with grenades [top left] in the area of the heaviest battles with Russian forces (the Bakhmut, Donetsk region), March 2023; in February 2023, a Ukrainian soldier attaches grenades to a DJI Mavic 3 drone [top right]; this tablet screen [bottom left] gives the remote operator a drone's-eye view of a rural scene with smoke rising from a struck target; a Ukrainian serviceman operates a drone [bottom right] to spot Russian positions near the city of Bakhmut, in the Donbas region, March 2023.





CLOCKWISE FROM TOP LEFT: ROMAN CHOP/AP; JOHN MOORE/GETTY IMAGES; ARTS MESSINIS/AFP/GETTY IMAGES; SERGEY SHESTAK/AFP/GETTY IMAGES

break, you throw them out, and there's no pilot to be paraded through the streets of the enemy capital.

You can do a lot with 55,000 drones. Shovel them at the foe and one in five may make it through. Yoke them together and send them flocking like a murmuration of starlings, and they will overwhelm anti-aircraft defenses. Even individually they can be formidable. One effective tactic is to have a drone "loiter" near a point where targets are expected to emerge, then dash in and drop a small bomb. Videos posted on social media purport to show Ukrainian remote operators dropping grenades on Russian troops or through the hatches of Russian armored vehicles.

A drone gives a lot of bang for the buck, as utterly new weapons often do. Over time, as a weapons system provokes countermeasures, their designers respond with improvements, and the gold plate accumulates.

In 1938, a single British Spitfire cost £9,500 to produce, equivalent to about \$1 million today. In the early 1950s the United States F-86 Sabre averaged about \$250,000 apiece, about \$3 million now. The F-35, today's top-of-the-line U.S. fighter, starts at \$110 million. Behold the modern-day fighter plane: the hypertrophied product of the longest arms race since the days of the dreadnought.

"In the year 2054, the entire defense budget will purchase just one aircraft," wrote Norman Augustine, formerly Under Secretary of the Army, back in 1984. "This aircraft will have to be shared by the Air Force and Navy 3-1/2 days each per week except for leap year, when it will be made available to the Marines for the extra day."

Like fighter planes, military drones started cheap, then got expensive. Unlike the fighters, though, they got cheap again.

Back in 1981, Israel sent modest contraptions sporting surveillance cameras in its war against Syria, to some effect. The U.S. military took hold of the concept, and in its hands, those simple drones morphed into Predators and Reapers, bomber-size machines that

flew missions in Iraq and Afghanistan. Each cost millions of dollars (if not tens of millions). But a technologically powerful country needn't count the cost; the United States certainly didn't.

"We are a country of technologists, we love technological solutions," says Kelly A. Grieco, a strategic analyst at the Stimson Center, a think tank in Washington, D.C. "It starts with the Cold War: Looking at the Soviet Union, their advantages were in numbers and in their close approach to Germany, the famous Fulda Gap. So we wanted technology to offset the Soviet numerical advantage."

A lot of the cost of an F-35 can be traced to the stealth technology that lets it elude even very sophisticated radar. The dreadnoughts of old needed guns of ever-greater range—enough finally to shoot beyond the horizon—so that the other side couldn't hold them at arm's length and pepper them with shells the size of compact cars.

Arms races tend to shift when a long peacetime buildup finally ends, as it has in Ukraine.

"The character of war has moved back toward quantity mattering," Grieco says. "Sophisticated tech is more readily available, and with AI advances and the potential for swarms, there's even more emphasis on quantity."

A recent research paper she wrote with U.S. Air Force Col. Maximilian K. Bremer notes that China has showcased such capabilities, "including a swarm test of 48 loitering munitions loaded with high-explosive warheads and launched from a truck and helicopter."

What makes these things readily available—as the nuclear and stealth technologies were not—is the Fourth Industrial Revolution: 3D printing, easy wireless connections, AI, and the big data that AI consumes. These things are all out there, on the open market.

"You can't gain the same advantage from simply possessing the technology," Grieco says. "What will become

## US \$3,200

Ukrainian grenade-equipped drone  
(including the grenade)

## US \$4 million

Russian T-90 tank destroyed  
by this Ukrainian drone

more important will be how you use it."

One example of how experience has changed use comes from the early days of the war in Ukraine. That country scored early successes with the Baykar Bayraktar TB2, a Turkish drone priced at an estimated at \$5 million each, about one-sixth as much as the United States' Reaper, which it broadly resembles. That's not cheap, except by U.S. standards.

"The Bayraktar was extremely effective at first, but after Russia got its act together with air defense, they were not as effective by so large a margin," says Zach Kallenborn, a military consultant associated with the Center for Strategic and International Studies, a think tank in Washington, D.C. That, he says, led both sides to move to masses of cheaper drones that get shot down so often they have a working life of maybe three to four days. So what? It's a good cost-benefit ratio for drones as cheap as Ukraine's DJIs and for Russia's new equivalent, the Shahed-136, supplied by Iran.

Ukraine has also resorted to home-made drones as an alternative to long-range jet fighters and missiles, which Western donors had been slow to provide. It recently launched such drones from its own territory to targets hundreds of kilometers inside Russia; Ukrainian officials said that they were working on a model that would fly about 1,000 kilometers.

Every military power is now staring at these numbers, not least the United States and China. If those two powers ever clash, it would likely be over Taiwan, which China says it will one day absorb and the United States says it will defend. Such a far-flung maritime arena would be very different from the close-in land war going on now in Eastern Europe. The current war may therefore not be a good guide to future ones.

"I don't buy that drones will transform all of warfare. But even if they do, you'd need to get them all the way to Taiwan. And to do that you'd need [aircraft] carriers," says Kallenborn. "And you'd need a way to communicate with drones. Relays are possible, but now satellites are key, so China's first move might be to knock out satellites. There's reason to doubt they would, though, because they need satellites, too."

In every arms race there is always another step to take. Right now the militaries of the world are working on ways to shoot down small drones with directed-energy weapons based on lasers or microwaves. The marginal cost of a shot would be low—once you've amortized the expense of developing, making, and deploying such weapons systems.

Should such antidrone measures succeed, then succeeding generations of drones will be hardened against them. With gold plating. ■

**Like fighter planes, military drones started cheap, then got expensive. Unlike the fighters, though, they got cheap again.**





This parabolic dish focuses sunlight on a small spot where the solar energy is 1,000 times its normal intensity. Over an 8-hour period, that energy is used to create about 500 grams of hydrogen, enough to propel a fuel-cell vehicle for roughly 100 kilometers.

## ENERGY

# Solar-to-Hydrogen Pilot Plant Reaches Kilowatt Scale

## › Sun-made hydrogen will supplant cracking of natural gas

BY PRACHI PATEL

**R**esearchers have built a kilowatt-scale pilot plant that can produce both green hydrogen and heat using solar energy. The solar-to-hydrogen plant is the largest ever constructed, and it produces about half a kilogram of hydrogen in 8 hours, which amounts to a little over 2 kilowatts of equivalent output power.

“We have cracked the 1-kW ceiling for the production of solar hydrogen,” says

Sophia Haussener, a professor of renewable energy science and engineering at the Swiss Federal Institute of Technology (EPFL), in Lausanne. “With half a kilogram of hydrogen you can drive a car for about 100 kilometers. Or you can use it in a fuel cell to produce electricity and satisfy about half the electricity needs of a four-person household a day.”

About 95 percent of the hydrogen used in the world today—mainly for

producing fertilizers and other chemicals or for oil refining—is made by cracking natural gas. Hydrogen also holds tremendous promise as a fuel for airplanes and ships, heating homes, and producing electricity. But that promise is blunted by the fact that cracking natural gas produces carbon dioxide.

To be a sustainable fuel, hydrogen must be made using renewable energy or nuclear power with minimal emissions. Green hydrogen is now gaining a greater foothold around the world. It is the centerpiece of Australia’s plans to decarbonize its economy, for instance.

One of the most sustainable ways to make hydrogen is to use solar energy to split water into hydrogen and oxygen. This can be done using photoelectrochemical (PEC) systems, which absorb sunlight and generate the electricity that drives the electrolytic splitting of water. “You don’t have to design and pay for two separate systems,” Haussener says. “It’s one single, integrated system, so ultimately there’s a cost advantage.”

PEC systems have shown tremendous promise at laboratory scale, in setups with less than 100 watts of output power. Scaling up to larger systems is not easy, Haussener says, because it involves balancing efficiency, stability, and operating costs while maximizing production rates. To achieve that balance, she and her colleagues focus the sun’s radiation onto a small spot, at which they place a solar-cell module using mirrored reflectors.

For the kilowatt-scale system the team reported on in *Nature Energy*, they built a 7-meter-wide parabolic solar dish whose mirrors concentrate solar radiation to about 1,000 times that of the sun’s normal output. The dish tracks the sun over a 6- to 8-hour period. Not all of the concentrated solar energy that falls on the solar cell is converted to electricity. Some of it is converted into waste heat, which the team extracts using a heat exchanger. Haussener says the heat could be used for space heating or hot water in buildings, or for fueling industrial processes.

Through a startup called SoHHytec, the EPFL team is now scaling up its hydrogen and heat-production technology to commercialize it. The company is building a system with a larger, 9-meter-wide solar dish, Haussener says. Multiple dishes can be tied together to make a system as big or as small as customers need. ■

## TELECOMMUNICATIONS

# Time-Reversing Mirror Makes Light Go Backward

## > Potential wireless, radar, and optical-computing applications on tap

BY CHARLES Q. CHOI

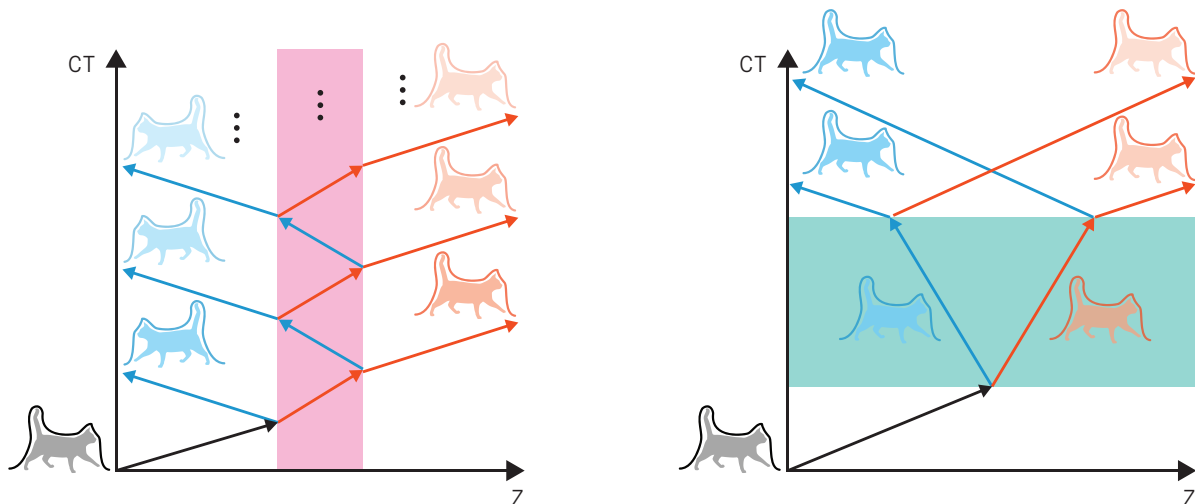
**S**cientists have long theorized about “time reflections,” in which a light or sound signal passing through a time “interface” would act as if it were traveling backward in time. Now, a new study demonstrates time reflections with light waves. This discovery could lead to new, unusual ways to control light, such as photonic time crystals, and for potential applications in wireless communications, radar technologies, and photonic computing.

A standard reflection occurs when a signal bounces off a boundary in space. By contrast, time reflections can materialize when the entire medium in which a light or sound wave is traveling suddenly and drastically changes its optical or sonic properties.

Previously, researchers tried to create a photonic time interface that could generate time reflections for electromagnetic waves. But they couldn’t figure out how to change a material’s optical properties in a way that was fast, robust, and uniform enough to obtain the desired result. Now, after six decades of research, scientists have created the first such time interface for light.

When a light wave enters this new time interface and the device changes its optical properties, the signal keeps moving forward in space—but reversed: If the signal were a spoken word, it would sound as if it was being played backward. With a conventional reflection, a light or sound wave travels back to its source but mostly looks or sounds the same as it did before the reflection.

The scientists were able to accomplish photonic time reflections using a metamaterial—a kind of structure engi-



These conceptual illustrations show how the properties of a metamaterial can create a time reflection. In the illustration at left, a gray cat stands next to a small block of glass [pink bar]. Light reflects off the near edge of the glass and produces a mirror image of the cat. Some of the light transmits through the glass and appears on the opposite side of the glass. Other light rays follow internal reflections off the glass block’s near and far edges and produce their own blurred, distorted, and discolored images of the cat. The illustration at right presents a scenario in which the optical properties of an entire medium—in this case, a metamaterial [green bar]—can be made to change and then very rapidly change back again. Here, the rapidly flickering properties of the medium “tune” the material so that the photons’ orientation in time itself is effectively reversed.



neered with features not generally found in nature, such as the ability to bend light in unexpected ways. Such work has led to “invisibility cloaks,” which can hide objects from light, sound, heat, and other types of waves.

Optical metamaterials, which are designed to manipulate light, have structures with repeating patterns at scales that are smaller than the wavelengths of light they influence. In the study, the researchers transmitted radio signals into a meandering strip of metal about 6 meters long. Then they connected an array of 30 electronic switches, each linked to a capacitor, to this metal strip.

When these switches are triggered at the same time, the impedance of the strip doubles in about 3 nanoseconds. This change occurs more rapidly than the variations in time of the signals involved, causing a time reflection in the light, says study senior author Andrea Alù, a professor of physics and electrical engineering at the City University of New York’s Graduate Center. The new interfaces can also stretch or compress the light signals in time. This can, in turn, abruptly change the color of these signals, Alù notes.

This novel device is a temporal version of the commonplace Fabry-Pérot imaging spectrometers widely used in telecommunications, lasers, and other applications, such as detecting terrestrial methane leaks from orbiting satellites.

The researchers suggest that the time interface’s initial application will likely be reversing the order of signals to help process them. Currently, time reversal is done by first digitizing the signal and putting it in reverse. But this results in time, energy, and memory demands that researchers have hoped to avoid.

Time interfaces sidestep this time and energy crunch, Alù says. Moreover, the new interfaces may help scientists develop exotic new ways to control light, such as photonic time crystals, in which optical properties vary regularly over time. “As a result, one can selectively amplify incoming signals and control their propagation in highly unusual ways,” he says.

The scientists detailed their findings online 13 March in the journal *Nature Physics*. ■

#### JOURNAL WATCH

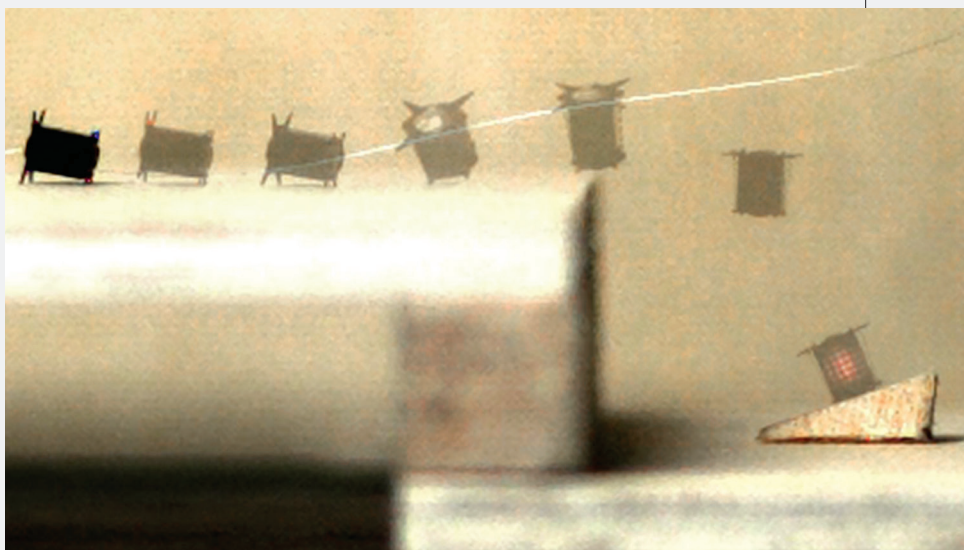
## A Giant Leap for Robotkind

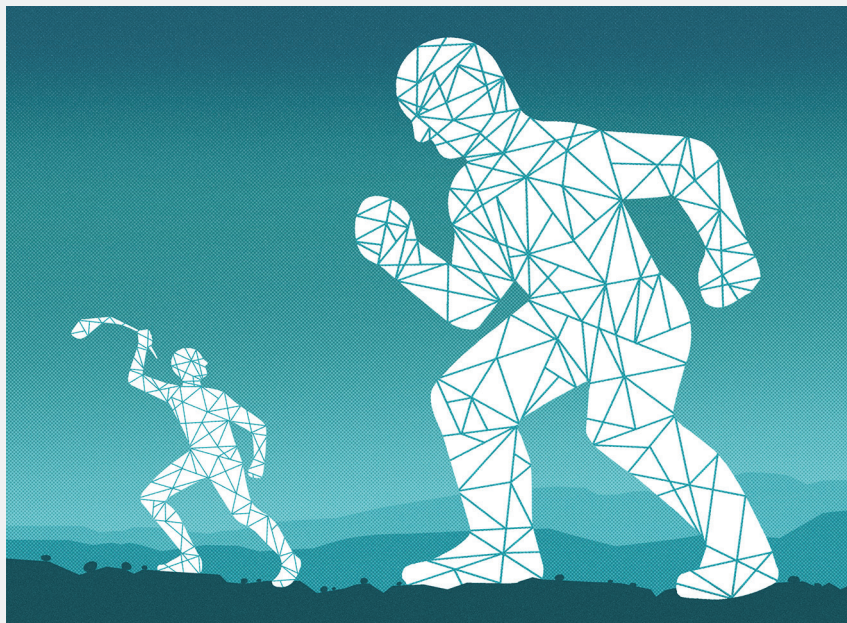
Despite their small size, fleas can unleash tremendous potential and jump close to 100 times their body length. Inspired by these skills, researchers at Beihang University, in Beijing, have created a robot with similar capabilities. The results were published 24 March in *IEEE Robotics and Automation Letters*.

To jump like a flea, a robot has to be able to unleash a lot of energy at once. So Ruide Yun, a third-year Ph.D. student at Beihang, and his colleagues created one that works somewhat like a miniature piston engine, powered by a high-voltage discharge in the piston chamber. In a series of experiments, the robot was able to jump as far as 296 millimeters and as high as 156 mm—equivalent to 87 times its own body length and 45 times its height.

“Our initial experiment was only intended to verify the novel driving method, but we never expected the robot to jump so high like a natural flea, while also exhibiting good crawling performance,” says Yun.

“Next, we will conduct research to help the robots break free from the constraints of wires and achieve free jumping and crawling,” Yun says. The team may also explore the use of actuators to control the leaping robot’s motion, he adds. —Michelle Hampson





## ARTIFICIAL INTELLIGENCE

## When Large Language Models Shrink > Smaller models trained on curated data challenge the preeminence of Big AI

BY EDD GENT

**B**uilding ever larger language models has led to groundbreaking jumps in performance. But it's also pushing state-of-the-art AI beyond the reach of all but the most well-resourced AI labs. For instance, while OpenAI has

remained tight-lipped on the matter, there is speculation that its latest GPT-4 large language model (LLM) has as many as a trillion parameters, far more than most companies or research groups have the computing resources to train or run.

**“The max size [of large language models] is going to continue to grow, and the quality at small sizes is going to continue to grow.”**

-DYLAN PATEL, SEMIANALYSIS

That's a problem, says Dylan Patel, chief analyst at the consultancy SemiAnalysis, because it makes it nearly impossible to reproduce these models.

Efforts to push back against the logic of scaling are underway, however. Last year, researchers at Alphabet's London-based subsidiary DeepMind showed that training smaller models on far more data could significantly boost performance. DeepMind's 70-billion-parameter Chinchilla model outperformed the 175-billion-parameter GPT-3 by training on nearly five times as much data. And in February, Meta's LLaMa model boasted a 13-billion-parameter version that outperformed GPT-3 on most benchmarks.

In March, LLaMa was leaked online. But then, researchers at Stanford fine-tuned the 7-billion-parameter version of LLaMa on 52,000 query responses from GPT-3.5, the model that originally powered ChatGPT and (as of press time) still powers OpenAI's free version. The resulting model, called Alpaca, was able to replicate much of the behavior of the OpenAI model, according to the researchers, who released their data and training recipe so others could replicate it.

Since then, hackers and hobbyists have run with the idea, using the LLaMa weights and the Alpaca training scheme to run their own LLMs on PCs, cellphones, and even a Raspberry Pi microcontroller.

Another promising approach for boosting efficiency is the so-called mixture of experts, which involves training multiple smaller submodels specialized for specific tasks rather than using a single large model to solve all of them.

Nvidia recently used the MoE tactic to build a vision-language model called Prism— and showed that the model could match the performance of models trained on 10 to 20 times as much data. “If we can take advantage of the knowledge already frozen in [existing pretrained] models,” says Jim Fan, AI research scientist at Nvidia, “we should.”

According to Patel, another trick involves exploiting a property known as sparsity. A surprisingly large number of weights in LLMs are set to zero. Finding ways to remove these zeros could help shrink the size of models and reduce computational costs, he says.

Sparsity is one of the most promising future directions for compressing models, says Sara Hooker, who leads the research

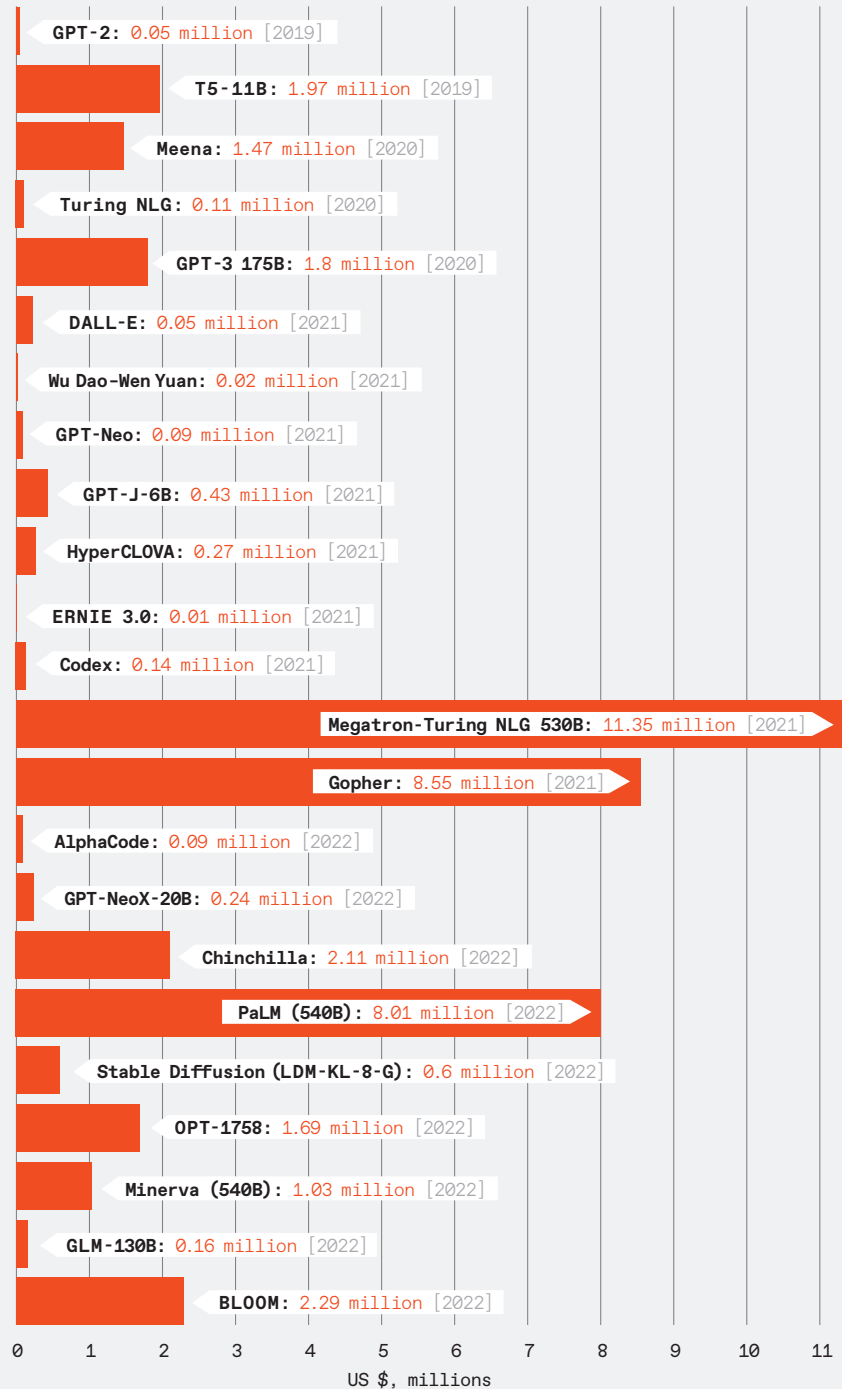


## ARTIFICIAL INTELLIGENCE

# Large Language and Multimodal Models Don't Come Cheap

BY MARGO ANDERSON & TEKLA S. PERRY

**HOVERING IN THE BACKGROUND** behind today's debates over large language models like ChatGPT is the sometimes exorbitant cost it takes to train them. In April, Stanford's Human-Centered AI Institute released a comprehensive report on the state of AI. The report includes this graph, which reveals the many millions of dollars spent in training the biggest AI models. Such trends set the stage for the rise of clever workarounds that don't require nearly as much resources to train and operate, as Edd Gent describes in the adjacent article.



lab Cohere For AI. But, she adds, today's hardware is not well-suited to exploit it. "The interesting challenge is, how do you represent the absence of something without actually representing it?" she says.

Yet another approach is known as quantization, which reduces data

requirements by representing weights using fewer bits—for instance, using 8-bit floating-point numbers rather than 32-bit. Knowledge distillation, in which a large teacher model is used to train a smaller one, also holds promise.

All these approaches face the same

problem, though. As Patel says, it's often easier to throw computing power at a well-understood model architecture than to fine-tune new techniques.

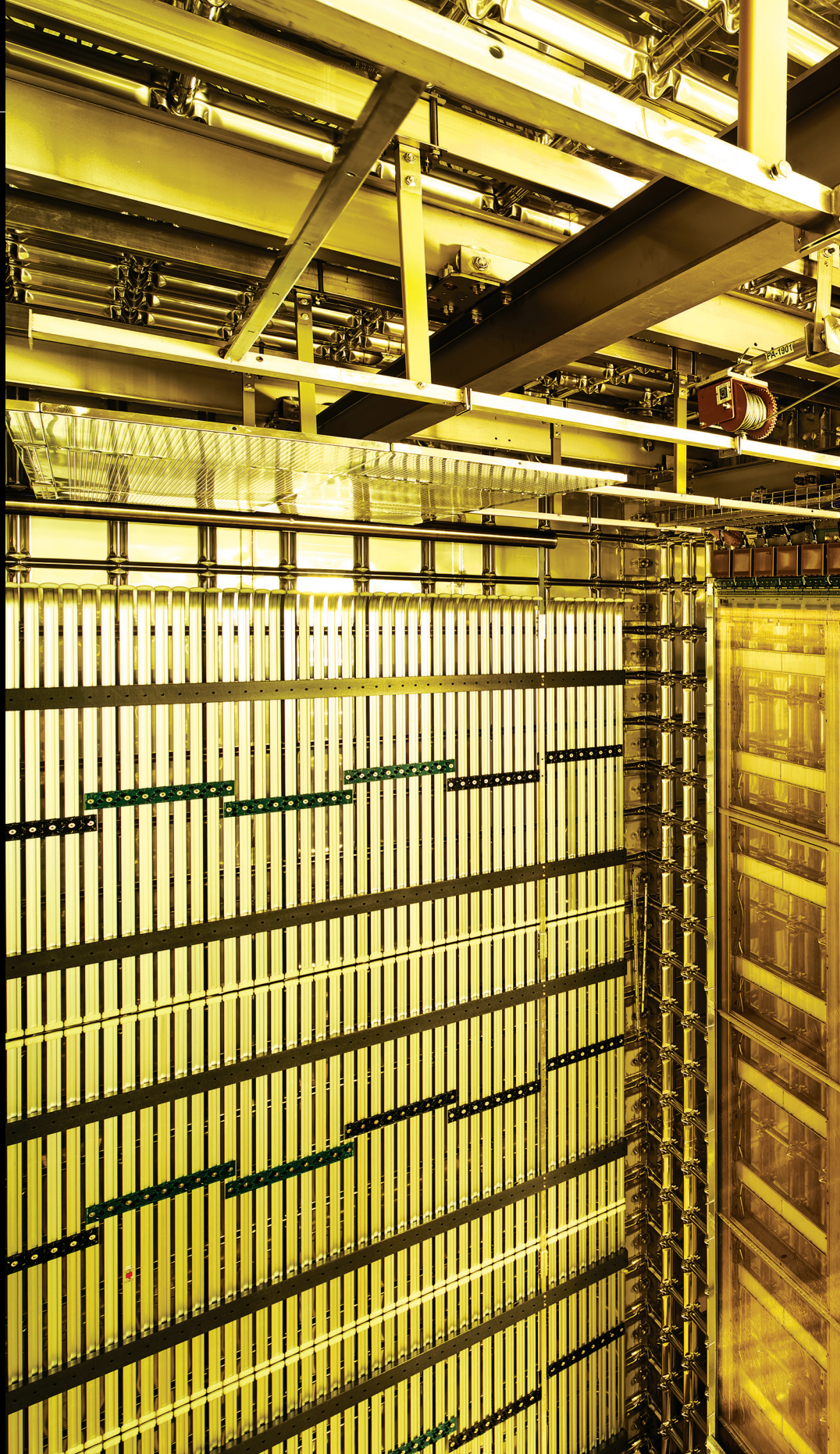
"The max size is going to continue to grow, and the quality at small sizes is going to continue to grow," he says. ■

## Neutrino Detector to Solve Universe's Mysteries

By Willie D. Jones

This is a cryostat for the ProtoDUNE II neutrino detector test bed located at the CERN nuclear research center on the France-Switzerland border. It is the precursor to the DUNE (Deep Underground Neutrino Experiment) detector. This prototype and the DUNE detector now under construction in the United States benefit from the experience gained from the design, construction, and operation of facilities such as the Large Hadron Collider (also at CERN). The DUNE neutrino generator located at Fermilab, in Illinois, will emit a beam made up of trillions of neutrinos. The so-called far detector, located 1,300 kilometers away at the Sanford Underground Research Facility, in South Dakota, will observe how the particles react.

PHOTOGRAPH BY  
ENRICO SACCHETTI

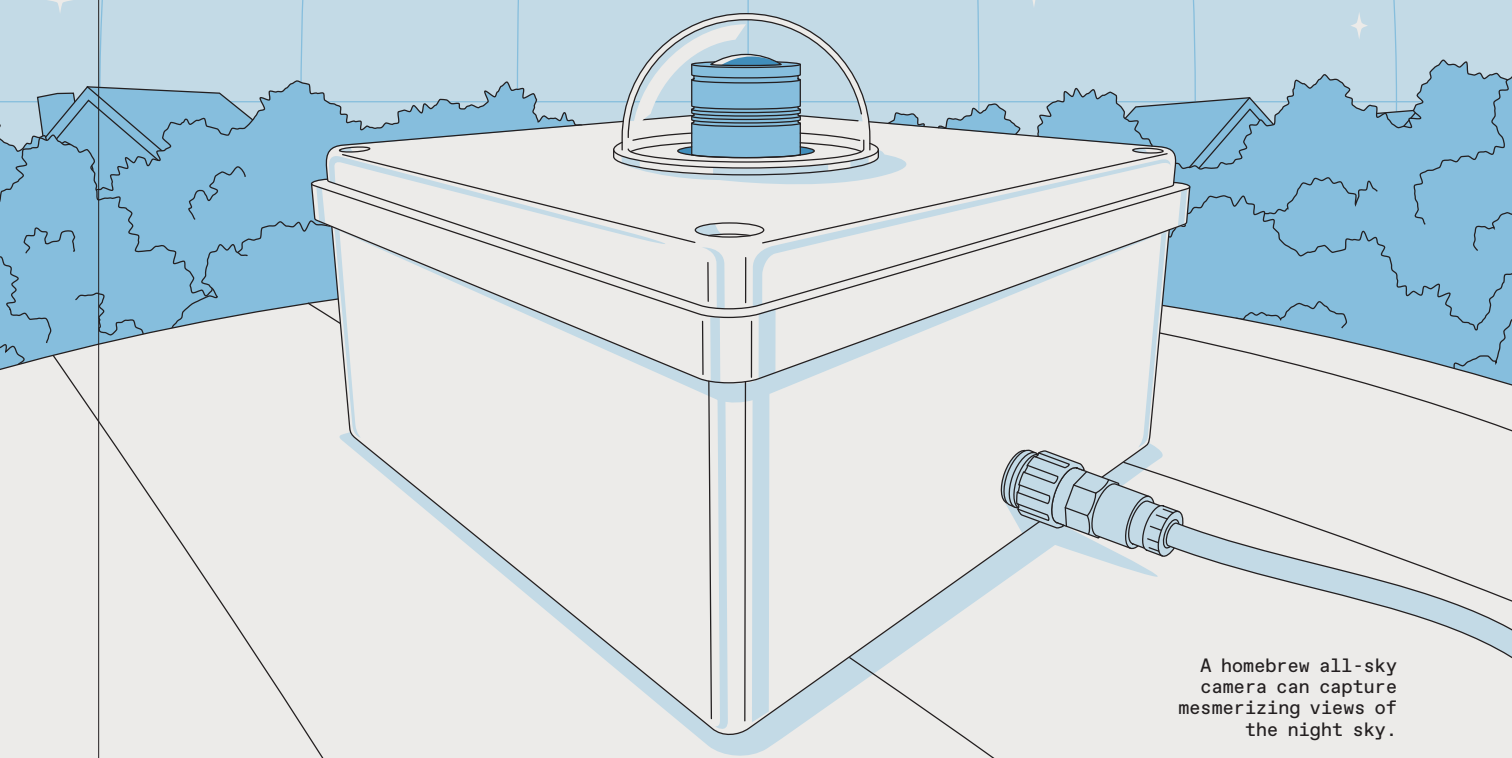








# Hands On



A homebrew all-sky camera can capture mesmerizing views of the night sky.

## Roll Your Own All-Sky Camera > Use Raspberry Pi hardware to capture mesmerizing time-lapse images of the heavens

BY DAVID SCHNEIDER

**W**hile driving home one night recently, I saw a spectacularly bright meteor flash across the sky in front of my car. A good-sized chunk of interplanetary detritus must have been on its way to a crash landing not too far away, I said to myself. My next thought was that if I had a bearing on that luminous streak, and if at least one other person in my region also had such information, we might be able to triangulate on it and narrow down where any landing zone might be. I'm, of course, not the only one to



ponder this possibility—and, I soon learned, people have indeed successfully found meteorites this way.

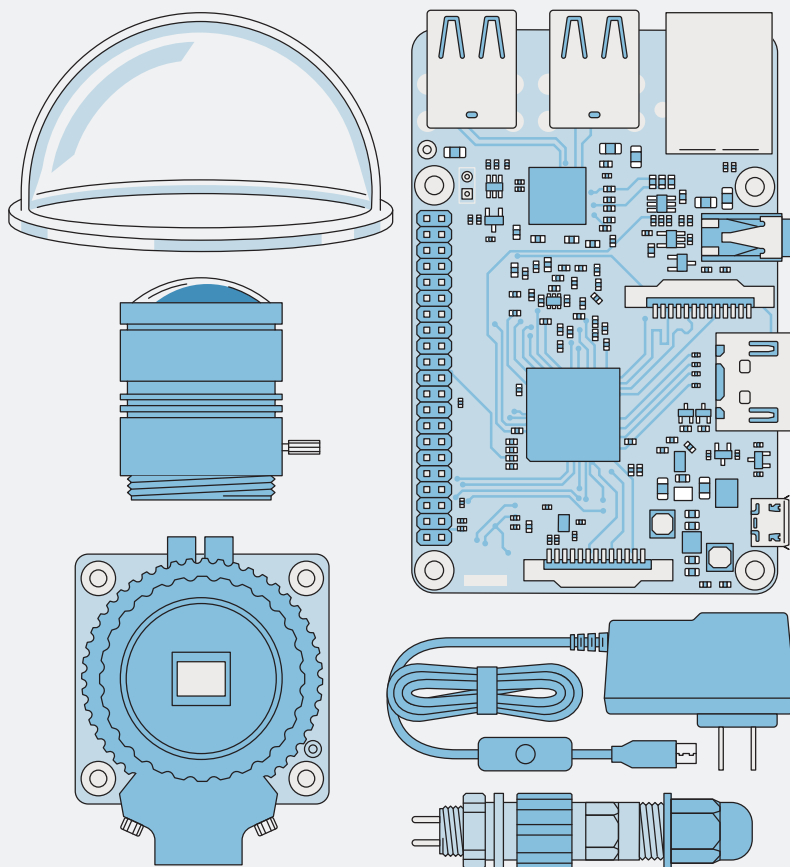
One example occurred in 2012, when a fireball lit up the sky over Northern California. Images of the meteor were recorded by a project called CAMS (Cameras for Allsky Meteor Surveillance)—a project of NASA and the SETI Institute. These observations allowed the object's trajectory and landing zone to be estimated, and coverage of the event in *The San Francisco Chronicle* soon led to the discovery of what became known as the Novato Meteorite.

CAMS is not the only such project looking for meteors. Another is the Global Meteor Network, whose mission is to observe the night sky with “a global science-grade instrument.” Organizers of this network even provide guidance for how anyone can build a suitable camera based on the Raspberry Pi and how to contribute observations that can help determine the orbits of the parent asteroids that spawned particular meteors.

I was tempted to join the cause, but after reading more on the subject I discovered alternative strategies for building a camera to survey the night sky. Ultimately, I decided that I wanted to capture attractive color images more than I wanted to contribute data to the Global Meteor Network, which uses black-and-white cameras because of their greater sensitivity.

So I opted to build a different kind of all-sky camera, one that is also based on a Raspberry Pi but that uses the color Raspberry Pi High Quality Camera, following the lead of a project called, reasonably enough, Allsky Camera.

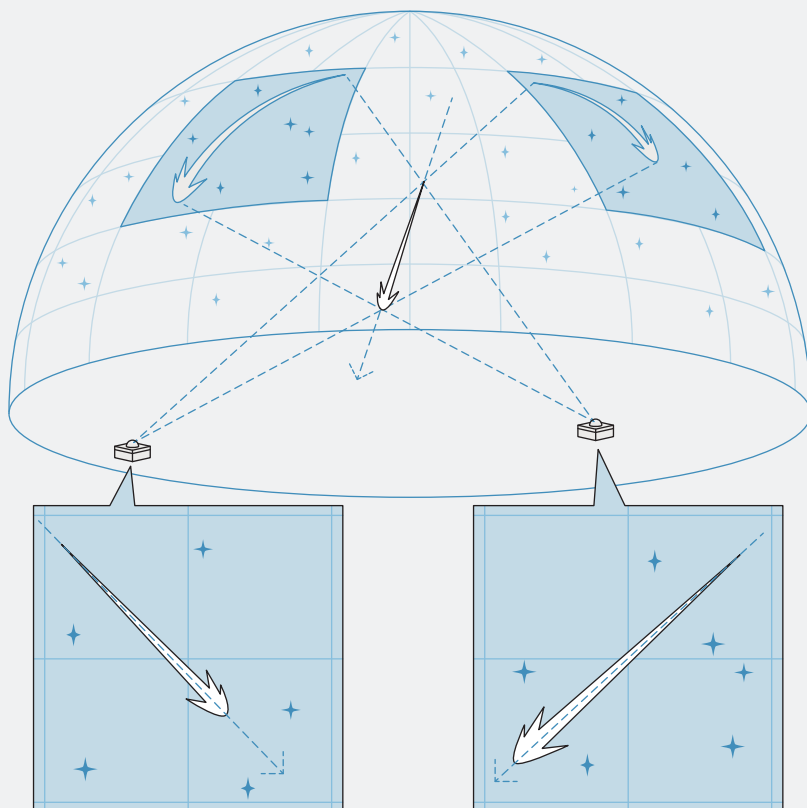
The hardware for this project consists of a Raspberry Pi and either the Raspberry Pi HQ camera or one of the purpose-built planetary cameras made by ZWO. To be truly “all sky,” the camera should be equipped with a fish-eye lens having a 180-degree field of view. Recognizing that my home is surrounded by trees, I opted for a lens with a narrower



The required components include a Raspberry Pi microcomputer [case not shown], a Raspberry Pi High Quality Camera, a lens, a dome-shaped transparent lens cover, a 5-volt power supply, and a waterproof bulkhead connector, allowing AC-mains power to pass through the wall of the waterproof enclosure [not shown] holding the camera.

(120-degree) field of view. A modern Raspberry Pi 4 is recommended, but I used a several-year-old Raspberry Pi 3 Model B simply because I had it on hand. I decided to use a US \$60 Raspberry Pi HQ camera over a ZWO camera because it offered higher resolution.

To protect this hardware from the elements, I housed the Pi, camera, and a suitable wall wart for powering the Pi inside a \$25 waterproof plastic enclosure. The opening I cut for the camera lens is covered with a \$16 clear acrylic dome. The first dome I purchased distorted things,



A meteor's trajectory through the atmosphere can hold clues to the location of any part of it that survives and also reveals the orbit of the parent body around the sun. With images of the meteor trail captured by two cameras, that trajectory can be ascertained: The position of a glowing trail relative to the background stars in an image defines a plane, and the intersection of these two planes defines the trajectory.

but I ordered another one that worked out much better. I also purchased an \$11 case for the Raspberry Pi (one that included a fan) and a long extension cord, which I cut and connected to a waterproof bulkhead connector. This means I can leave the unit outside even when it rains.

Following the guidance provided in a very nice tutorial video, I found it straightforward to set up the Allsky Camera software on my Pi, running it in a “headless” configuration—meaning without a monitor or keyboard. I access it wirelessly from a laptop through my local area network using SSH.

I fired everything up—but the camera didn't work at all. So I turned to the appropriate troubleshooting section in the project's ample documentation and

tried what was advised there—to enable “Glamor” graphic acceleration on the Pi. Still no images, though. Eventually, I discovered some tweaks to a configuration file that are needed when using the HQ camera on a Pi 3B, which allowed me to obtain a hopelessly blurry image of the ceiling of my office.

Through trial and error, I was able to get the manual focus of the camera dialed in properly. And slowly I learned how to adjust the multitude of settings available in the Allsky Camera software, which is done either by editing a configuration file or, more conveniently, through a Web interface this software provides.

For example, I learned that I should reduce the resolution of the images used to make time-lapse videos, lest images saved at the impressive native resolution of the HQ camera (4,056 by 3,040 pixels) overwhelm the processing and storage available on my Pi. While that required tweaking a configuration file, other settings can be adjusted using the Web interface, which also makes it very easy to view live images, browse images collected earlier, and view and save time-lapse videos.

One thing that puzzled me early on was how well such a camera would work to catch meteors flashing by, given that the camera takes still images, not many-frames-per-second videos. But my concerns diminished after capturing images of the night sky over my home, some of which caught the light of passing aircraft. The long trails of light in those images made it apparent that the exposure time must be at least some tens of seconds long. I knew these were aircraft, not meteor trails, because the streaks included parallel tracks (from wingtip lights) and obvious pulsations from strobes.

I hope yet to capture meteors some day with this gizmo. For that, I may go camping in the mountains in mid-August, when the Perseids are hitting their peak. My family and I had taken such a trip years ago, but I didn't have an all-sky camera at the time. So I returned home with only some now-fading memories of the wondrous show nature had put on display above our heads. Next time, I'll have something I can view over and over! ■



# Careers:

## Sushama Chakraverty

How this UI designer finds the “fun factor” that makes video games so compelling

**L**anding a job in the video-game industry is a dream for many young engineers, but for Sushama Chakraverty, it was a happy accident.

Back in 2001, she spotted a poster on her professor’s door advertising for an internship, and she applied without knowing the company she’d be working for. It turned out to be Atari, the legendary video-game company. That internship kicked off a fulfilling career as a video-game programmer.

“They hired me full-time six months later, and I never looked back,” she says. “I was having too much fun.”

In addition to Atari, she’s worked at smaller gaming companies, as well as the leading development studio Ubisoft in Montreal and Toronto, where she contributed to major titles like *Assassin’s Creed 2*, *Watchdogs Legion*, and *Far Cry 4*.

Today, she’s a senior user interface engineering manager for Sledgehammer Games, in Toronto, creator of the iconic *Call of Duty* franchise.

The industry has changed considerably in just a few decades, Chakraverty says, going from a few small firms to a multibillion-dollar economic powerhouse. And she still thanks her lucky stars that she stumbled into one of the most fun jobs an engineer can have.

“There are so many challenges to solve and there are different ones that come up all the time,” she says. “The programming is very technical, but it’s also very creative.”

**Growing up in the Indian city of Pune**, Chakraverty followed a well-worn path for an ambitious soft-

**“We’re here because we want to have fun making entertainment for millions of people across the globe. Just the thought of that is very exciting.”**



While working at Ubisoft as a video-game programmer, Sushama Chakraverty contributed to major titles like *Watchdogs Legion*.

ware engineer. After completing a bachelor’s degree in computer engineering at D.Y. Patil College of Engineering, also in Pune, she got a job at a multinational company in Mumbai. She left India to pursue a master’s degree in computer science at the University of Maryland, Baltimore County. But as her degree program came to a close in 2001, Chakraverty was undecided about her next move.

“I knew I didn’t want to do traditional IT like building databases or working on networking programs, but I was unsure what else was out there,” she says. That’s when she saw the help-wanted poster for “an intern with some programming knowledge, some math, and some computer science experience.”

Chakraverty had played a few early MS-DOS-based games like *Doom* and *Aladdin* at college, but was far from a diehard gamer. The first project she worked on at Atari was the fantasy-themed role-playing game *Dungeons and Dragons: Heroes* for the Xbox. She admits she found the content slightly baffling at first. But she was immediately

drawn in by how varied and stimulating the work was. She wrote code for a variety of game systems, fixed software bugs, and helped to build tools for the game's artists and designers.

"Every day was different," she says.

What sets the video-game industry apart from other programming jobs, Chakraverty says, is the highly iterative nature of the work, requiring multiple revisions. If you're writing a piece of software for a bank, she says, the requirements are gathered at the start of the project and are generally well defined. The architecture the engineers work on generally stays the same from start to finish.

"In the video-game industry, things can pivot six months before the product ships, even though we've been working on it for the last five years," she says.

That's because game designers often make major changes late in the development process, from rewriting the storyline to ripping out whole levels or adding an entirely new multiplayer mode. The team can cycle through hundreds of versions of the game until they hit on a combination that works. There's a good reason for that, Chakraverty says. More than anything, video-game designers are searching for that ephemeral "fun factor" that can make playing the game so rewarding. And that requires a lot of prototyping and a lot of reinvention on the fly.

**Typically a video-game** project starts with the team throwing together a few rough-and-ready first drafts to make sure the concept works. Once that crucial fun factor has been found, the real development work begins. Much of the early code gets tossed out when the programmers begin work on the real game, Chakraverty says.

As different elements of a game are pieced together, from specific levels to the way fighting between combatants works, designers assess whether they are improving the game or just introducing more complexity.

The work itself is technically challenging. It's "hardcore C++ programming," she says, something most developers learn at university but don't always get the chance to put into practice. And everything has to be optimized for the tight performance constraints of a gaming console.

**After landing a job** at the Montreal office of Ubisoft in 2008, Chakraverty began to specialize in user interface design. The interfaces are the host of menus and head-up display elements that players use to navigate through the game. They provide a

**Employer:**

Sledgehammer Games, Toronto

**Title:**

Senior User Interface Engineering Manager

**Education:**

Bachelor's degree in computer engineering, D.Y. Patil College of Engineering, Pune, India; Master's degree in computer science, The University of Maryland, Baltimore County

crucial link between the underlying game systems and the information displayed to the player on the screen. This has to be done as intuitively as possible, Chakraverty says.

At Ubisoft, she also got her first taste of working in large teams, contributing to blockbusters like *Assassin's Creed 2* and *Splinter Cell: Conviction*.

"On smaller teams, you get more exposure to stuff because you have more responsibility," she says. Projects at the smaller companies tend to be smaller in scale.

At larger studios, projects can stretch for years, and the systems being built are more complex and interesting. The roles and responsibilities are more defined, Chakraverty says, which means "you don't get to wear so many different hats. But you're also working alongside the best in the industry, and there's more opportunity for learning and network building. That's a critical part of getting ahead."

"There's never a time where you feel like you know enough," she adds.

**Being a woman** in a male-dominated industry has not always been smooth sailing. "Some companies are still boys' clubs," she says.

But things are changing fast, she adds, and the industry is very different compared to when she started out more than 20 years ago.

While there's still a long way to go to diversify technical teams, she notes that areas like production and art, as well as more business-focused departments like marketing and finance are far more inclusive.

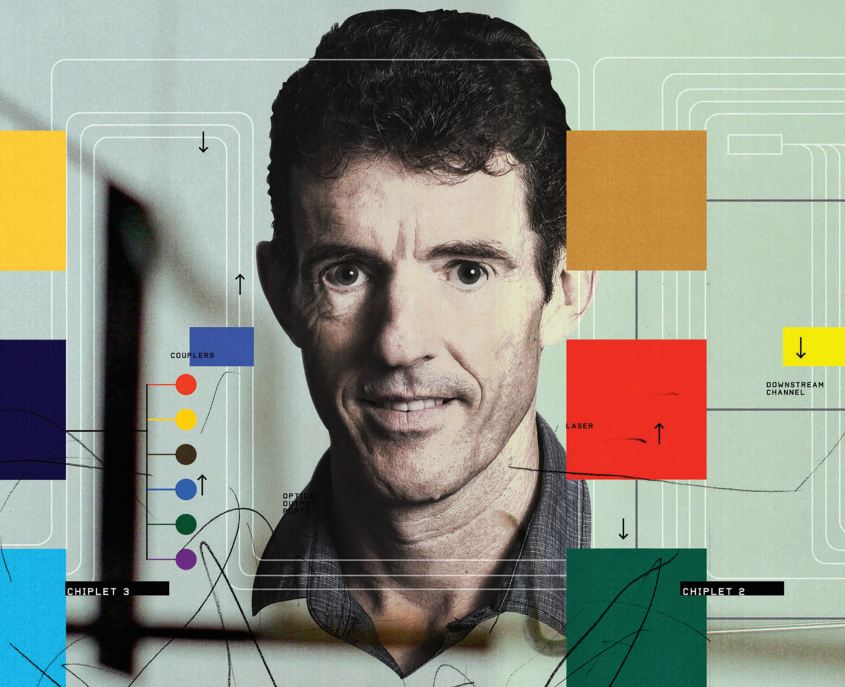
Chakraverty says her experiences have made her passionate about mentoring and helping younger generations of engineers recognize their skills and advocate for themselves, something she thinks women in particular are often not taught to do.

For those interested in a career in video games, it's important to really think about where you want to contribute to the development process, Chakraverty says. Depending on what you're working on, the requirements can be very different. At the most well-known studios, C++ skills are a must, she says, but for mobile games, JavaScript is more in demand. If you want to work on 3D engines, strong math knowledge is important.

Wherever you end up, Chakraverty says, a career in video games can be uniquely rewarding.

"It's a fun job," she says. "We're here because we want to have fun making entertainment for millions of people across the globe. Just the thought of that is very exciting." ■





# 5 Questions for Sam Naffziger

How he drove AMD's pioneering move to chiplets

**O**ver the past five years, the processor has gone from being a single piece of silicon to a collection of smaller chiplets. This approach means that the CPU's functional pieces can be built using the best technology for each piece. Sam Naffziger, a product-technology architect at AMD, was an early proponent of this approach. Naffziger recently answered five chiplet-size questions from *IEEE Spectrum* on the topic.

## What are the main challenges you've seen for chiplets-based processors?

**Sam Naffziger:** We started out five or six years ago with the EPYC and Ryzen CPU lines. And at the time, we cast a pretty broad net to find what package technologies would be best for connecting the die [small block of silicon]. It's a complex equation of cost, capability, bandwidth densities, power consumption, and also manufacturing capacity. It's relatively easy to come up with great package technologies, but it's a completely different thing to actually manufacture them in high volume, cost effectively.

## How might chiplets change the semiconductor manufacturing process?

**Naffziger:** That's definitely something that the industry is working through. There's where we're at

*Sam Naffziger is senior vice president, corporate fellow, and product-technology architect at AMD and an IEEE Fellow. He is the recipient of the IEEE Solid-State Circuits Society's 2023 Industry Impact Award.*

today, and then there's where we might go in 5 to 10 years. I think today, pretty much, the technologies are general purpose. They can be aligned to monolithic die just fine, or they can function for chiplets. With chiplets, we have much more specialized intellectual property. So, in the future one could envision specializing the process technology and getting performance benefits, cost reductions, and other things. But that's not where the industry is at today.

## How will chiplets affect software?

**Naffziger:** One of the goals of our architecture is to have it be completely transparent to software, because software is hard to change. For example, our second-generation EPYC CPU is made up of a centralized I/O [input/output] chiplet surrounded by compute dies. When we went to a centralized I/O die, it reduced memory latency, eliminating a software challenge from the first generation.

Now, with the [AMD Instinct] MI300—AMD's upcoming high-performance computing accelerator—we're integrating both CPU and GPU compute dies. The software implication of that sort of integration is that they can share one memory address space. Because the software doesn't have to worry about managing memory, it's easier to program.

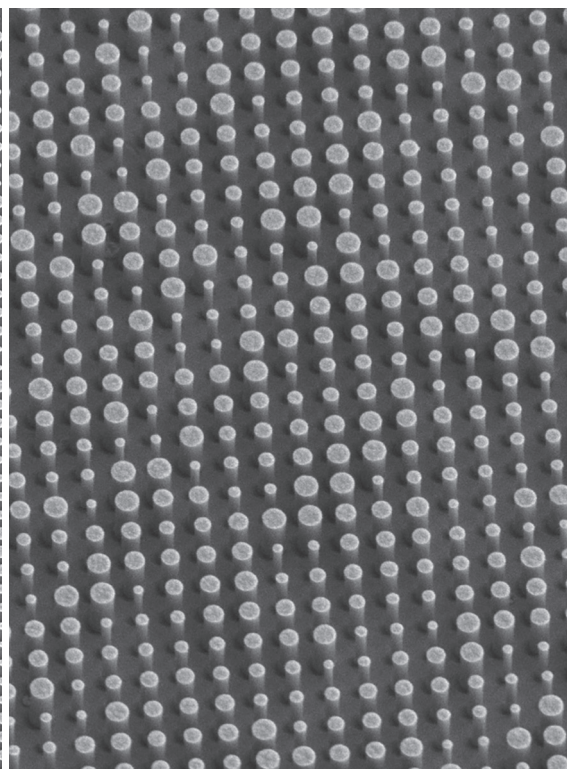
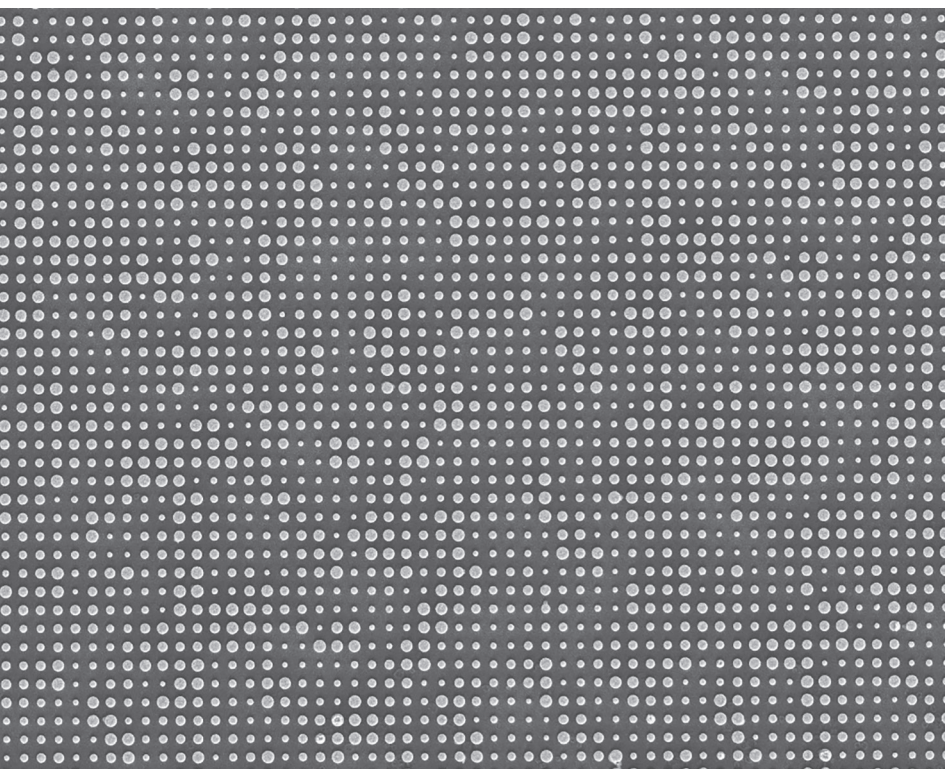
## How much of the architecture can be separated out onto chiplets?

**Naffziger:** We're finding ways to scale logic, but SRAM is more of a challenge, and analog stuff is definitely not scaling. We've already taken the step of splitting off the analog with the central I/O chiplet. With 3D V-Cache—a high-density cache chiplet 3D-integrated with the compute die—we have split off the SRAM. And I would expect in the future there will be lots more of that kind of specialization. The physics will dictate how fine grained we can go, but I'm bullish about it.

## What has to happen for mixing and matching different companies' chiplets into the same package to become a reality?

**Naffziger:** First of all, we need an industry standard on the interface. UCIe, a chiplet interconnect standard introduced in 2022, is an important first step. I think we'll see a gradual move towards this model because it really is going to be essential to deliver the next level of performance per watt and performance per dollar. Then, you will be able to put together a system-on-chip that is market or customer specific. ■

# NANOSTRUCTURES TRANSFORM



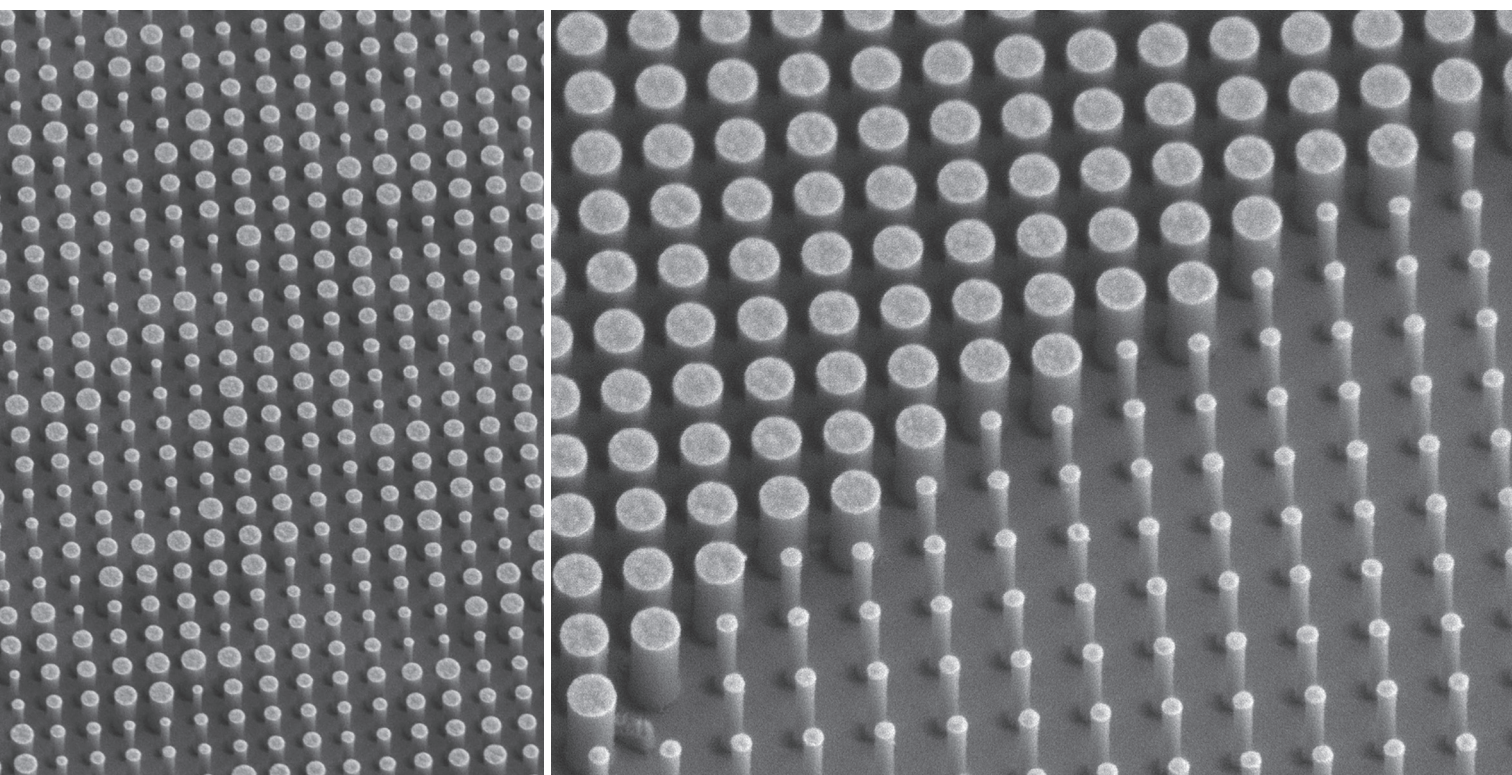
These images of a metalens, taken with a scanning electron microscope, show the variations in size and layout of the nanopyramids used to manipulate light.

Thin semiconductor



# TINY CAMERAS

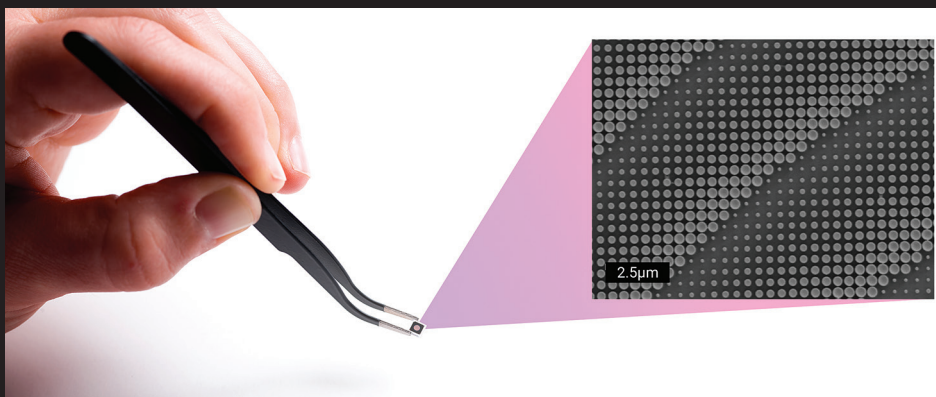
By Robert Gobron



metalenses are finally moving into consumers' hands

In this single metalens [right, held with tweezers], the pillars are less than 500 nanometers in diameter. The black box at the bottom left of the enlargement represents 2.5 micrometers.

A 12-inch wafer [opposite page] can hold up to 10,000 metalenses, which are made using a single semiconductor layer.



**I**nside today's computers, phones, and other mobile devices, more and more sensors, processors, and other electronics are fighting for space. Taking up a big part of this valuable real estate are the cameras—just about every gadget needs a camera, or two, three, or more. And the most space-consuming part of the camera is the lens.

The lenses in our mobile devices typically collect and direct incoming light by refraction, using a curve in a transparent material, usually plastic, to bend the rays. So these lenses can't shrink much more than they already have: To make a small camera, you need a lens with a short focal length; but the shorter the focal length, the greater the curvature and therefore the thickness at the center. These highly curved lenses also suffer from all sorts of aberrations, so camera-module manufacturers use multiple lenses to compensate, adding to the camera's bulk.

With today's lenses, the size of the camera and image quality are pulling in different directions. The only way to make lenses smaller and better is to replace refractive lenses with a different technology.

That technology exists. It's the metalens, a device developed at Harvard and commercialized at Metalenz, where I am an applications engineer. We create these devices using traditional semiconductor-processing techniques to build nanostructures onto a flat surface. These nanostructures use a phenomenon called metasurface optics to direct and focus light. These lenses can be extremely thin—a few hundred micrometers thick, about twice the thickness of a human hair. And we can combine the functionality of multiple curved lenses into just one of our devices, further addressing the space crunch and opening up the possibility of new uses for cameras in mobile devices.

**Before I tell you how** the metalens evolved and how it works, consider a few previous efforts to replace the traditional curved lens.

Conceptually, any device that manipulates light does so by altering its three fundamental properties: phase, polarization, and intensity. The idea that any

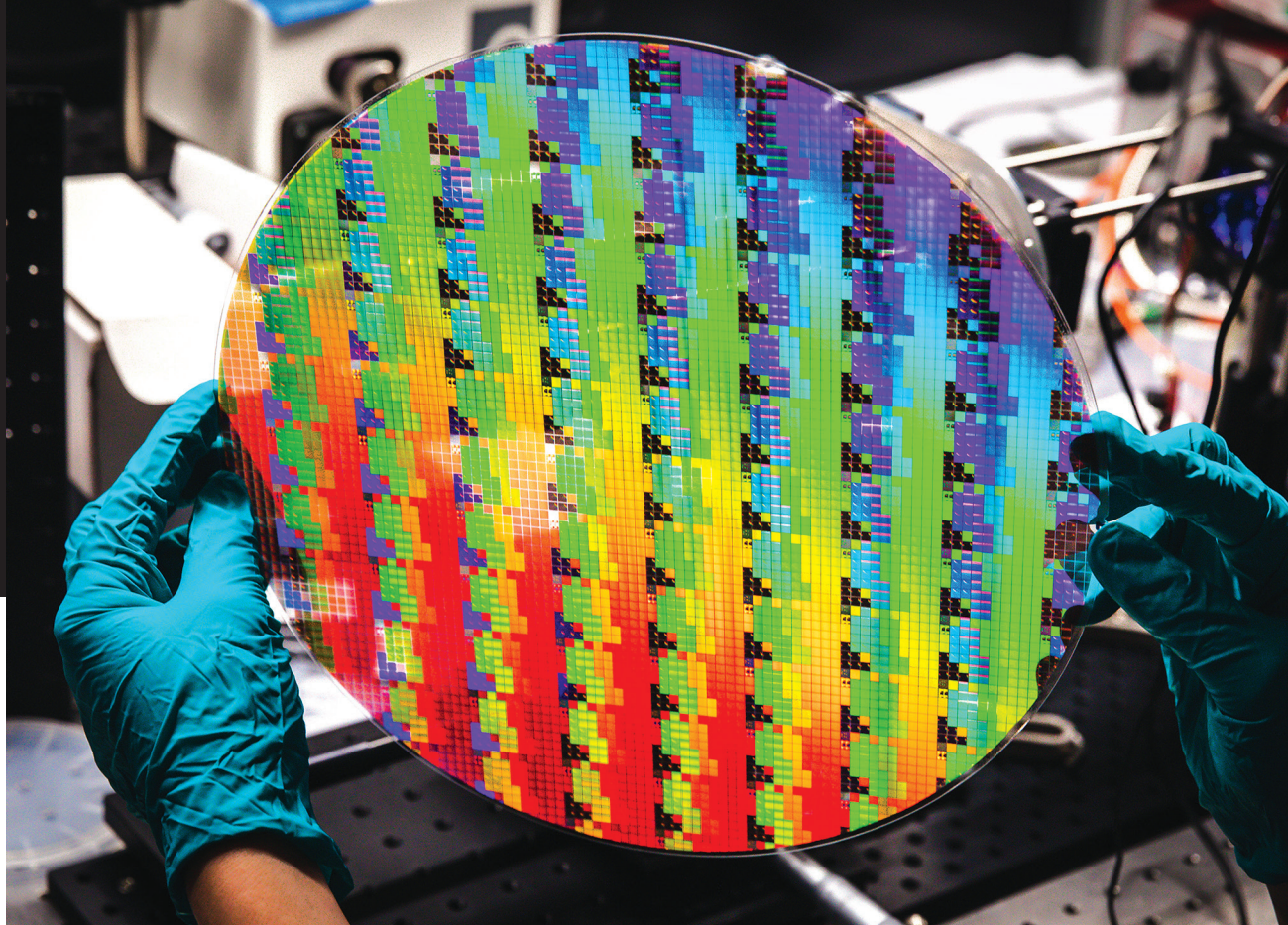
wave or wave field can be deconstructed down to these properties was proposed by Christiaan Huygens in 1678 and is a guiding principle in all of optics.

In the early 18th century, the world's most powerful economies placed great importance on the construction of lighthouses with larger and more powerful projection lenses to help protect their shipping interests. However, as these projection lenses grew larger, so did their weight. As a result, the physical size of a lens that could be raised to the top of a lighthouse and structurally supported placed limitations on the power of the beam that could be produced by the lighthouse.

French physicist Augustin-Jean Fresnel realized that if he cut a lens into facets, he could remove much of the central thickness of the lens but retain the same optical power. The Fresnel lens represented a major improvement in optical technology and is now used in a host of applications, including automotive headlights and brake lights, overhead projectors, and—still—for lighthouse projection lenses. However, the Fresnel lens has limitations. For one, the flat edges of facets become sources of stray light. For another, faceted surfaces are more difficult to manufacture and to polish precisely than continuously curved ones are. It's a no-go for camera lenses, due to the surface accuracy requirements needed to produce good images.

Another approach, now widely used in 3D sensing and machine vision, traces its roots to one of the most famous experiments in modern physics: Thomas Young's 1802 demonstration of diffraction. This experiment showed that light behaves like a wave, and that when the waves meet, they can amplify or cancel one another depending on how far the waves have traveled. The so-called diffractive optical element (DOE) based on this phenomenon uses the wavelike properties of light to create an interference pattern—that is, alternating regions of dark and light, in the form of an array of dots, a grid, or any number of shapes. Today, many mobile devices use DOEs to convert a laser beam into “structured light.” This light pattern is projected, captured by an image sensor, then used by algorithms to create a 3D map of the scene.





These tiny DOEs fit nicely into small gadgets, yet they can't be used to create detailed images. So, again, applications are limited.

**Enter the metalens.** Developed at Harvard by Federico Capasso's team, which included then-graduate student Rob Devlin, research associates Reza Khorasaninejad and Wei Ting Chen, and others, metalenses work in a way that's fundamentally different from any of these other approaches.

A metalens is a flat glass surface with a semiconductor layer on top. Etched in the semiconductor is an array of pillars several hundred nanometers high. These nanopillars can manipulate light waves with a degree of control not possible with traditional refractive lenses.

Imagine a shallow marsh filled with sea grass standing in water. An incoming wave causes the sea grass to sway back and forth, sending pollen flying off into the air. If you think of that incoming wave as light energy, and the nanopillars as the stalks of sea grass, you can picture how the properties of a nanopillar, including its height, thickness, and position next to other nanopillars, might change the distribution of light emerging from the lens.

We can use the ability of a metalens to redirect and change light in a number of ways. We can scatter and project light as a field of infrared dots. Invisible to the eye, these dots are used in many smart devices to measure distance, mapping a room or a face. We can sort light by its polarization (more on that in a moment). But probably the best way to explain how we are using

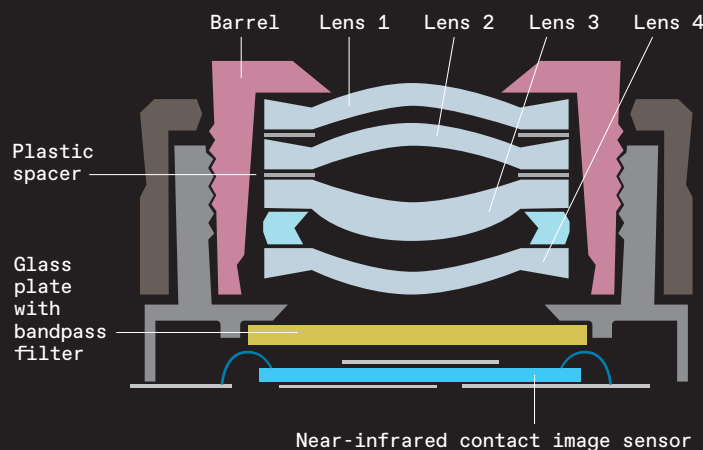
these metasurfaces as a lens is by looking at the most familiar lens application—capturing an image.

The process starts by illuminating a scene with a monochromatic light source—a laser. (While using a metalens to capture a full-color image is conceptually possible, that is still a lab experiment and far from commercialization.) The objects in the scene bounce the light all over the place. Some of this light comes back toward the metalens, which is pointed, pillars out, toward the scene. These returning photons hit the tops of the pillars and transfer their energy into vibrations. The vibrations—called plasmons—travel down the pillars. When that energy reaches the bottom of a pillar, it exits as photons, which can be then captured by an image sensor. Those photons don't need to have the same properties as those that entered the pillars; we can change these properties by the way we design and distribute the pillars.

**R**esearchers around the world have been exploring the concept of metalenses for decades.

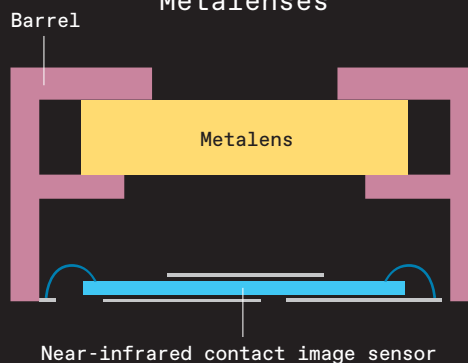
In a paper published in 1968 in *Soviet Physics Uspekhi*, Russian physicist Victor Veselago put the idea of metamaterials on the map, hypothesizing that nothing precluded the existence of a material that exhibits a negative index of refraction. Such a material would interact with light very differently than a normal material would. Where light ordinarily bounces off a material in the form of reflection, it would pass around this type of metamaterial like water going around a boulder in a stream. It wasn't until 2000 that the theory of meta-

## Conventional lenses



A single metalens [below] can replace a stack of traditional lenses [left], simplifying manufacturing and dramatically reducing the size of a lens package.

## Metalenses



materials was implemented in the lab. That year, Richard A. Shelby and colleagues at the University of California, San Diego, demonstrated a negative refractive index metamaterial in the microwave region. They published the discovery in 2001 in *Science*, causing a stir as people imagined invisibility cloaks. (While intriguing to ponder, creating such a device would require precisely manufacturing and assembling thousands of metasurfaces.)

The first metalens to create high-quality images with visible light came out of Federico Capasso's lab at Harvard. Demonstrated in 2016, with a description of the research published in *Science*, the technology immediately drew interest from smartphone manufacturers. Harvard then licensed the foundational intellectual property exclusively to Metalenz, where it has now been commercialized.

Since then, researchers at Columbia University, Caltech, and the University of Washington, working with Tsinghua University, in Beijing, have also demonstrated the technology.

Much of the development work Metalenz does involves fine-tuning the way the devices are designed. In order to translate image features like resolution into nanoscale patterns, we developed tools to help calculate the way light waves interact with materials. We then convert those calculations into design files that can be used with standard semiconductor processing equipment.

The first wave of optical metasurfaces to make their way into mobile imaging systems have on the order of 10 million silicon pillars on a single flat surface only a few millimeters square, with each pillar precisely tuned to accept the correct phase of light, a painstaking process even with the help of advanced software. Future generations of the metalens won't necessarily have more pillars, but they'll likely have more sophisticated geometries, like sloped edges or asymmetric shapes.

**Metalenz came out** of stealth mode in 2021, announcing that it was getting ready to scale up production of devices. Manufacturing was not as big a challenge as design because metasurfaces are manufactured using the same materials, lithography, and etching processes as those used to make integrated circuits.

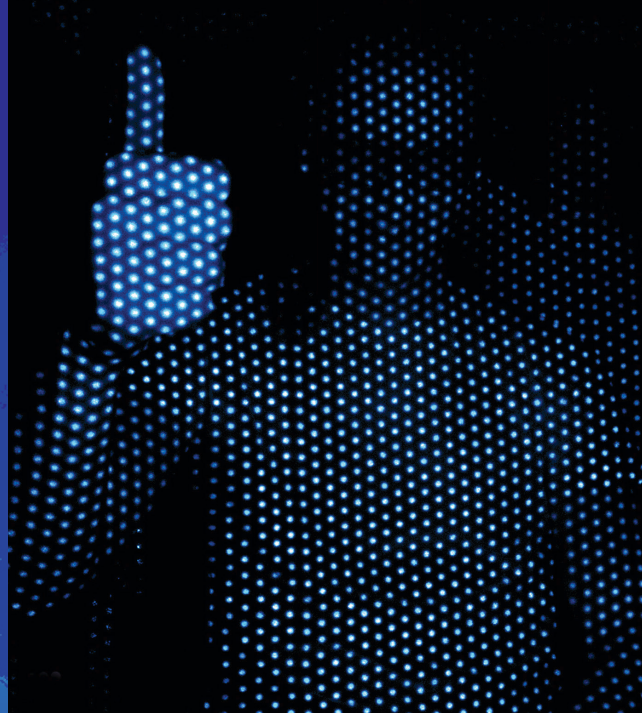
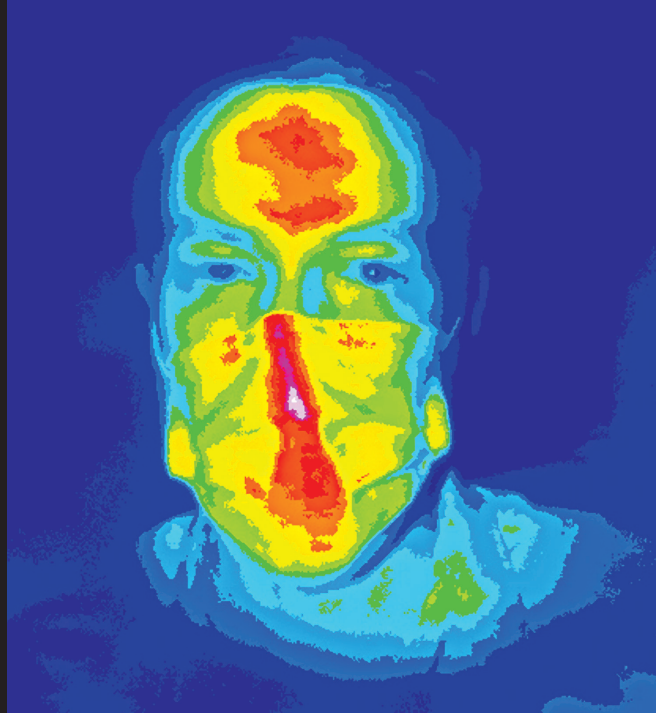
In fact, metalenses are less demanding to manufacture than even a very simple microchip because they require only a single lithography mask as opposed to the dozens required by a microprocessor. That makes them less prone to defects and less expensive. Moreover, the size of the features on an optical metasurface are measured in hundreds of nanometers, whereas foundries are accustomed to making chips with features that are smaller than 10 nanometers.

And, unlike plastic lenses, metalenses can be made in the same foundries that produce the other chips destined for smartphones. This means they could be directly integrated with the CMOS camera chips on site rather than having to be shipped to another location, which reduces their cost still further.

In 2022, STMicroelectronics announced the integration of Metalenz's metasurface technology into its FlightSense modules. Previous generations of FlightSense have been used in more than 150 models of smartphones, drones, robots, and vehicles to detect distance. Such products with Metalenz technology inside are already in consumer hands, though STMicroelectronics isn't releasing specifics.

Indeed, distance sensing is a sweet spot for the current generation of metalens technology, which operates at near-infrared wavelengths. For this application, many consumer electronics companies use a time-of-flight system, which has two optical components: one that transmits light and one that receives it. The transmitting optics are more complicated. These involve multiple lenses that collect light from a laser and transform it to parallel light





Metasurface optics [above] are capable of capturing polarization information from light, revealing a material's characteristics and providing depth information. A single meta-optic [above right], in combination with an array of laser emitters, can be used to create the type of high-contrast, near-infrared dot or line pattern used in 3D sensing.

waves—or, as optical engineers call it, a collimated beam. These also require a diffraction grating that turns the collimated beam into a field of dots. A single metalens can replace all of those transmitting and receiving optics, saving real estate within the device as well as reducing cost.

And a metalens does the field-of-dots job better in difficult lighting conditions because it can illuminate a broader area using less power than a traditional lens, directing more of the light to where you want it.

**C**onventional imaging systems, at best, gather information only about the spatial position of objects and their color and brightness.

But the light carries another type of information: the orientation of the light waves as they travel through space—that is, the polarization. Future metalens applications will take advantage of the technology's ability to detect polarized light.

The polarization of light reflecting off an object conveys all sorts of information about that object, including surface texture, type of surface material, and how deeply light penetrates the material before bouncing back to the sensor. Prior to the development of the metalens, a machine vision system would require complex optomechanical subsystems to gather polarization information. These typically rotate a polarizer—structured like a fence to allow only waves oriented at a certain angle to pass through—in front of a sensor.

They then monitor how the angle of rotation impacts the amount of light hitting the sensor.

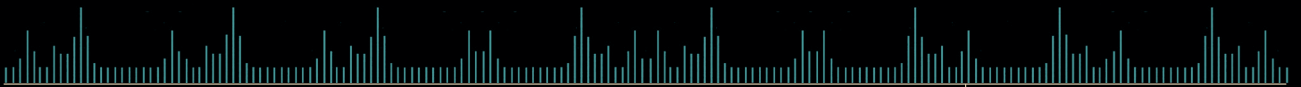
A metalens, by contrast, doesn't need a fence; all the incoming light comes through. Then it can be redirected to specific regions of the image sensor based on its polarization, using a single optical element. If, for example, light is polarized along the x axis, the nanostructures of the metasurface will direct the light to one section of the image sensor. However, if it is polarized at 45 degrees to the x axis, the light will be directed to a different section. Then software can reconstruct the image with information about all its polarization states.

Using this technology, we can replace previously large and expensive laboratory equipment with tiny polarization-analysis devices incorporated into smartphones, cars, and even augmented-reality glasses. A smartphone-based polarimeter could let you determine whether a stone in a ring is diamond or glass, whether concrete is cured or needs more time, or whether an expensive hockey stick is worth buying or contains micro cracks. Miniaturized polarimeters could be used to determine whether a bridge's support beam is at risk of failure, whether a patch on the road is black ice or just wet, or if a patch of green is really a bush or a painted surface being used to hide a tank. These devices could also help enable spoof-proof facial identification, since light reflects off a 2D photo of a person at different angles than a 3D face and from a silicone mask differently than it does from skin. Handheld polarizers could improve remote medical diagnostics—polarization is used in oncology to examine tissue changes, for example.

But like the smartphone itself, it's hard to predict where metalenses will take us. When Apple introduced the iPhone in 2008, no one could have predicted that it would spawn companies like Uber. In the same way, perhaps the most exciting applications of metalenses are ones we can't even imagine yet. ■

By PETER ANDREY SMITH

# Guided by Voices



**DIGITAL VOICEPRINTING** may not be ready for the courts









At 6:36 a.m., on 3 December 2020, the U.S Coast Guard received a call over a radio channel reserved for emergency use: “Mayday, Mayday, Mayday. We lost our rudder...and we’re taking on water fast.” The voice hiccuped, almost as if the man were struggling. He radioed again, this time to say that the pumps had begun to fail. He said he’d try to get his boat, a 42-footer with three people on board, back to Atwood’s, a lobster company on Spruce Head Island, Maine. The Coast Guard asked for his GPS coordinates and received no reply.

That morning, a Maine Marine Patrol officer, Nathan Stillwell, set off in search of the missing vessel. Stillwell rode down to Atwood Lobster Co., which is located at the end of a peninsula, and boarded a lobster boat, motoring out into water so shockingly cold it can induce lethal hypothermia in as little as 30 minutes.

When he returned to shore, Stillwell continued canvassing the area for people who had heard the radio plea for help. Someone told him the voice in the mayday call sounded “messed up,” according to a report obtained through a state-records request. Others said it sounded like Nate Libby, a dockside worker. So Stillwell went inside Atwood’s and used his phone to record his conversation with Libby and another man, Duane Maki. Stillwell asked if they had heard the call.

“I was putting my gloves and everything on the rack,” Libby told him. “I heard it. I didn’t know that word, honestly,” (presumably referring to the word *mayday*). “And I just heard it freaking coming on that he lost his rudder, that he needed pumps.” Both men denied making the call.

Stillwell seemed unsure. In his report, he said he’d received other tips suggesting the VHF call had been made by a man whose first name was Hunter. But then, the next day, a lobsterman, who owned a boat like the one reported to be in distress, called Stillwell. He was convinced that the mayday caller was his former sternman, the crew member who works in the back of the lobster boat: Nate Libby.

The alarm was more than just a prank call. Broadcasting a false distress signal over maritime radio is a violation of international code and, in the United States, a federal Class D felony. The Coast Guard recorded the calls, which spanned about 4 minutes, and investigators isolated four WAV files, capturing 20 seconds of the suspect’s voice. To verify the caller’s identity and solve the apparent crime, the Coast Guard’s investigative service emailed the files to Rita Singh, a computer scientist at Carnegie Mellon University and author of the textbook *Profiling Humans From Their Voice* (Springer, 2019).

In an email obtained through a federal Freedom of Information Act request, the lead investigator wrote Singh, “We are currently working a possible Search and Rescue Hoax in Maine and were wondering if you could compare the voice in the MP3 file with the voice making the radio calls in the WAV files?” She agreed to analyze the recordings.

Historically, such analysis—or, rather, an earlier iteration of the technique—had a bad reputation in the courts. Now, thanks to advances in computation, the technique is coming back. Indeed, forensic scientists hope one day to glean as much information from a voice recording as from DNA.

## WE HEAR WHO YOU ARE

**THE METHODS OF AUTOMATED** speech recognition, which converts speech into text, can be adapted to perform the



more sophisticated task of speaker recognition, which some practitioners refer to as voiceprinting.

Our voices have a lot of special characteristics. “As an identifier,” Singh wrote recently, “voice is potentially as unique as DNA and fingerprints. As a descriptor, voice is more revealing than DNA or fingerprints.” As such, there are many reasons to be concerned about its use in the criminal legal system.

A 2020 U.S. Government Accountability Office report says that the U.S. Secret Service claims to be able to identify an unknown person in a voice-only lineup, comparing a recording of an unknown voice with a recording of a known speaker, as a reference. According to a 2022 paper, there have been more than 740 judgments in Chinese courts involving voiceprints. Border-control agencies in at least eight countries have used language analysis for determination of origin, or LADO, to analyze accents to determine a person’s country of origin and assess the legitimacy of their asylum claims.

Voice-based recognition systems differ from old-school wiretapping and surveillance by going beyond the substance of a conversation to infer information about the speaker from the voice itself. Even something as simple as putting in an order at a McDonald’s drive-through in Illinois has raised legal questions about collecting biometric data without con-

sent. In October, the Texas attorney general accused Google of violating the state’s biometric privacy law, saying the Nest home-automation device “records—without consent—friends, children, grandparents, and guests who stop by, and then stores their voiceprints indefinitely.” Another lawsuit asserts that JPMorgan Chase used a Nuance system called Gatekeeper, which allegedly “collects and considers the unique voiceprint of the person behind the call” to authenticate its banking customers and detect potential fraud.

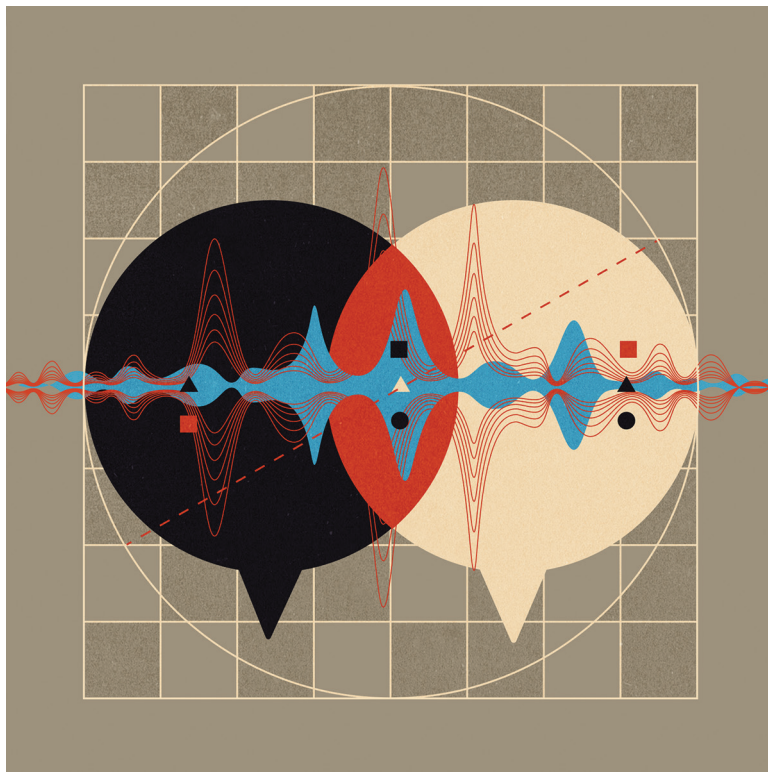
Other state and national authorities allow citizens to use their voices to verify their identity and thus gain access to their tax data records and pension information. “There’s a massive shadow risk, which is that any speaker-verification technology can be turned into speaker identification,” says Wiebke Toussaint Hutiri, a researcher at Delft University of Technology (TU Delft), in the Netherlands, who has studied bias.

### BETTER TO REMAIN SILENT AND BE THOUGHT A FOOL THAN TO OPEN YOUR MOUTH AND REMOVE ALL DOUBT

**SINGH SUGGESTS THAT** speech analysis alone can be used to generate a shockingly detailed profile of an unknown speaker. “If you merge the powerful machine-learning, deep-learning technology that we have today with all of the information that is out there and do it right, you can engineer very powerful systems that can look really deeply into the human voice and derive all kinds of information,” she says.

In 2014, Singh fielded her first query about hoax callers from the Coast Guard. She analyzed the recordings they provided and sent the service several conclusions. “I was able to tell them how old the person was, how tall he was, where he was from, probably where he was at the time of calling, approximately what kind of area, and a bunch of things about the guy.” She did not learn until later that the information apparently helped solve the crime. From then on, Singh says, she and the agency have had an “unspoken pact.”

On 16 December 2020, about two weeks after receiving the relevant audio files, Singh emailed investigators a report that explained how she had used computational algorithms to compare the recordings. “Each recording is studied in its entirety, and all conclusions are based on quantitative measures obtained from complete signals,” she said. Singh wrote that she had performed the automated portion of the analysis after manually labeling two voices Stillwell recorded in his in-person



dockside interview as US410 and US411: Person1 and Person2. Then, she used algorithms to compare the unknown voice—the four short bursts broadcast on the emergency channel—with the two known speakers.

Singh reached the conclusion many others in Maine had: The unknown voice in the four mayday recordings came from the same speaker as Person1, who identified himself as Nate Libby in US410. A little after 5 p.m. on the day Singh returned her report, Stillwell received the news. As he wrote in an incident report obtained through records requests: “The recordings of the distress call and the interview with Mr. Libby were a match.” By comparing the voice of an unknown speaker with two possible suspects, the investigators had apparently verified the mayday caller’s identity as Person1—Nate Libby.

The term “vocal fingerprint” dates to at least as early as 1911, according to Mara Mills and Xiaochang Li, the coauthors of an authoritative history on the topic. Mills says the technique has always been inextricably linked to criminal identification. “Vocal fingerprinting was about identifying people for the purpose of prosecuting them.” Indeed, the Coast Guard’s recent investigation of audio from the hoax distress call and the more general revival of the term “voice-printing” are especially surprising given its checkered history in U.S. courts.

Perhaps the best-known case began in 1965, when a TV reporter for CBS went to Watts, a Los Angeles neighborhood that had been besieged by rioting, and interviewed a man whose face was not depicted. On camera, the man claimed he’d taken part in the violence and had firebombed

a drugstore. Police later arrested a man named Edward Lee King on unrelated drug charges. They found a business card for a CBS staffer in his wallet. Police suspected King was the anonymous source—the looter who confessed to torching a store. Police secretly recorded him and then invited Lawrence Kersta, an engineer who worked at Bell Labs, to compare the two tapes. Kersta popularized the examinations of sound spectrograms, which are visual depictions of audio data.

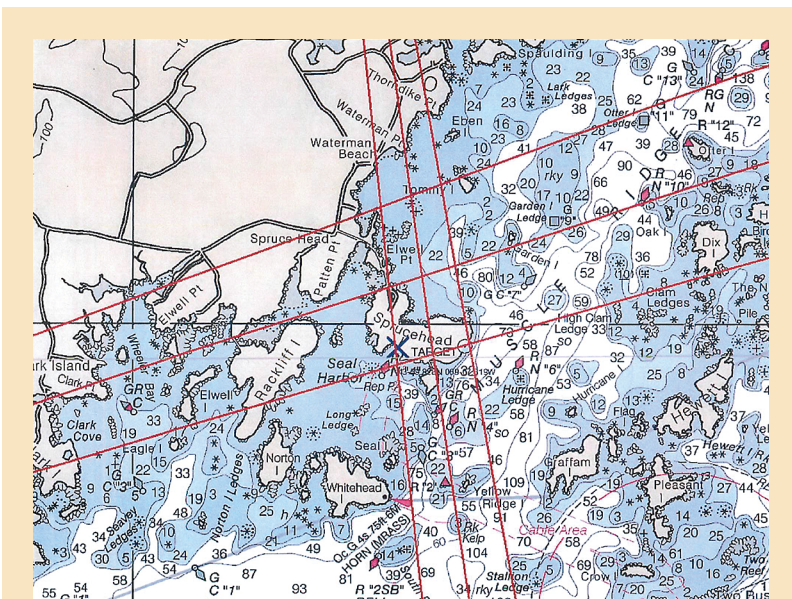
Kersta’s testimony sparked considerable controversy, forcing speech scientists and acoustical engineers to take a public stand on voiceprinting. Experts ultimately convinced a judge to reverse King’s guilty verdict.

## PRETENDING TO PREDICT WHAT YOU ALREADY KNOW

**VOICEPRINTING’S DEBUT** triggered a flurry of research that soon discredited it. As a 2016 paper in the *Journal of Law and the Biosciences* put it: “The eulogy for voiceprints was given by the National Academy of Sciences in 1979, following which the FBI ceased offering such experts...and the discipline slid into decline.” In a 1994 ruling, U.S. District Judge Milton Shadur of the Northern District of Illinois criticized the technique, likening one-on-one comparisons to a kind of card trick, where “a magician forces on the person chosen from the audience the card that the magician intends the person to select, and then the magician purports to ‘divine’ the card that the person has chosen.”

It’s surprising that the old term has come back into vogue, says James L. Wayman, a voice-recognition expert who works on a subcommittee of the U.S. National Institute of Standards and Technology. Despite the recent advances in machine learning, he says, government prosecutors still face significant challenges in getting testimony admitted and convincing judges to allow experts to testify about the technique before a jury. “The FBI has frequently testified against the admissibility of voice evidence in cases, which is a really interesting wrinkle.” Wayman suggested that defense attorneys would have a field day asking why investigators had relied on an academic lab—and not the FBI’s examiners.

The Coast Guard appeared to be aware of these potential hurdles. In January 2021, the lead investigator wrote Singh: “We are working on our criminal complaint and the attorneys are wondering if we could get your CV and if you have ever testified as an expert witness in court.” Singh replied that all the cases she had worked on had been settled out of court.



This map shows maritime details around Spruce Head Island, Maine.



Six months later, on 3 June 2021, Libby pleaded guilty, averting any courtroom confrontation over Singh's voice-based analysis. (The judge said the hoax appeared to be an attempt to get back at an employer who had fired Libby because of his drug use.) Libby was sentenced to time served, three years of supervised release, and the payment of US \$17,500 in restitution. But because of the opacity of the plea-bargaining system, it's hard to say what weight the voice-based analyses played in Libby's decision: His public defender declined to comment, and Libby himself could not be reached.

The outcome nonetheless reflects practice: The use of forensic speaker comparison is primarily investigative. "People do try to use it as evidence in courts, but it's not the kind of thing that would send someone to jail for life," Mills says. "Even with machine learning, that kind of certitude isn't possible with voiceprinting."

Moreover, any technical limitations are compounded by the lack of standards. Wayman contends that there are too many uncontrolled variables, and analysts must contend with so-called channel effects when comparing audio made in different environments and compressed into different formats. In the case of the Maine mayday hoax, investigators had no recording of Libby as he would sound when broadcast over the emergency radio channel and recorded in WAV format.

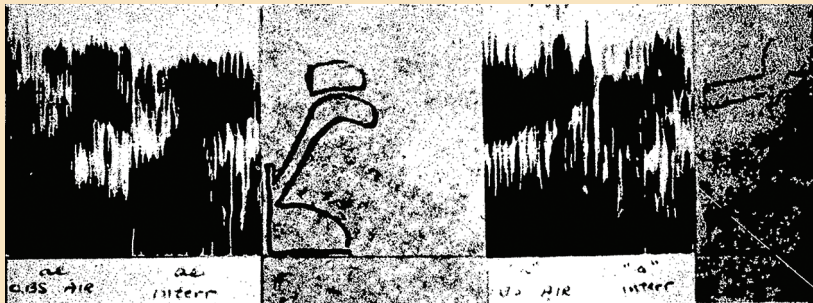
## SHOOT FIRST, DRAW THE TARGET AFTERWARD

**TU DELFT'S HUTIRI** suggests that any bias might not be inherent in the technology; rather, the technology may reinforce systemic biases in the criminal-justice system.

One such bias may be introduced by whoever manually labels the identity of the speakers in template recordings, prior to analysis. That simply reflects the fact that the examiner is applying received information about the suspect. Such unmasking may contribute to what forensic experts call the sharpshooter fallacy: Someone fires a bullet in the side of a barn and then draws a circle around the bullet hole afterward to show they've hit their mark.

Singh did not build a profile from an unidentified voice. She used computational algorithms to draw another circle around the chief suspect, confirming what law enforcement and several Mainers already suspected: that the hoax caller's voice belonged to Libby.

True, Libby's plea suggests that he was indeed guilty. His



At a 1966 trial in Los Angeles, Lawrence Kersta, an engineer from Bell Labs, testified that a set of annotated spectrograms, including the above, could identify a criminal suspect's voiceprint. The suspect was convicted, but the conviction was later overturned, and critics widely denounced voiceprinting.

confession, in turn, suggests that Singh correctly verified the speaker's voice in the distress call. But the case was not published, peer reviewed, or replicated. There is no estimate of the error rate associated with the identification—the probability that the conclusion is inaccurate. This is quite a weakness.

These gaps may hint at larger problems as deep neural networks play an ever-bigger role. Federal evidentiary standards require experts to explain their methods, something the older modeling techniques could do but deep-learning models can't. "We know how to train them, right? But we don't know what it is exactly that they're doing," Wayman says. "These are some major forensic issues."

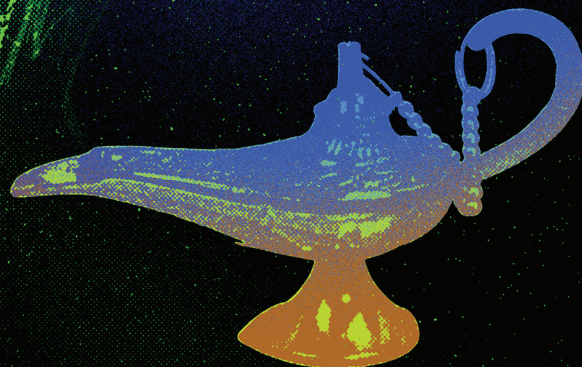
Other, more fundamental questions remain unanswered. How distinctive is an individual human's voice? "Voices change over time," Mills says. "You could lose a couple of your fingerprints, but you'd still have the others; any damage to your voice, you suddenly have a quite different voice." Also, people can train their voices. In the era of deepfakes and voice cloning text-to-speech technologies, such as Overdub and VALL-E, can computers identify who is impersonating whom?

On top of all that, defendants have the right to confront their accusers, but machine testimony, as it's called, may be based on as little as 20 seconds of audiotape. Is that enough to prove guilt beyond a reasonable doubt? The courts have yet to decide.

Singh sometimes boasts that her group was the first to demonstrate a live voice-profiling system and the first to re-create a voice from a mere portrait (that of the 17th-century Dutch painter Rembrandt). That claim, of course, cannot be falsified. And, despite the prevailing skepticism, Singh still contends that it is possible to profile a person from a few sentences, even a single phrase. "Sometimes," she says, "one word is enough."

The courts may not agree. ■







JULY 2023

SPECTRUM.IEEE.ORG 35

# A CRITICAL LOOK at AI- Generated Software

*Coding with  
the new AI tools  
is both irresistible  
and dangerous*

By  
**JAIDEEP VAIDYA  
& HAFIZ ASIF**



*Illustrations by Daniel Zender*

IN MANY WAYS, we live in the world of *The Matrix*. If Neo were to help us peel back the layers, we would find code all around us. Indeed, modern society runs on code: Whether you buy something online or in a store, check out a book at the library, fill a prescription, file your taxes, or drive your car, you are most probably interacting with a system that is powered by software.

And the ubiquity, scale, and complexity of all that code just keeps increasing, with billions of lines of code being written every year. The programmers who hammer out that code tend to be overburdened, and their first attempt at constructing the needed software is almost always fragile or buggy—and so is their second and sometimes even the final version. It may fail unexpectedly, have unanticipated consequences, or be vulnerable to attack, sometimes resulting in immense damage.

Consider just a few of the more well-known software failures of the past two decades. In 2005, faulty software for the US \$176 million baggage-handling system at Denver International Airport forced the whole thing to be scrapped. A software bug in the trading system of the Nasdaq stock exchange caused it to halt trading for several hours in 2013, at an economic cost that is impossible to calculate. And in 2019, a software flaw was discovered in an insulin pump that could allow hackers to remotely control it and deliver incorrect insulin doses to patients. Thankfully, nobody actually suffered such a fate.

These incidents made headlines, but they aren't just rare exceptions. Software failures are all too common, as are security vulnerabilities. Veracode's most recent survey on software security, covering the last 12 months, found that about three-quarters of the applications examined contained at least one security flaw, and nearly one-fifth had at least one flaw regarded as being of high severity.

What can be done to avoid such pitfalls and more generally to prevent software from failing? An influential 2005 article in *IEEE Spectrum* identified several factors that are still quite relevant.

Testing and debugging remain the bread and butter of software reliability and maintenance. Tools such as functional programming, code review, and formal methods can also help to eliminate bugs at the source. Alas, none of these methods has proven absolutely effective, and in any case they are not used consistently. So problems continue to mount.

Meanwhile, the ongoing AI revolution promises to revamp software development, making it far easier for people to program, debug, and maintain code. GitHub Copilot, built on top of OpenAI Codex, a system that translates natural language to code, can make code recommendations in different programming languages based on the appropriate prompts. And this is not the only such system: Amazon CodeWhisperer, CodeGeeX, GPT-Code-Clippy, Replit Ghostwriter, and Tabnine, among others, also provide AI-powered coding and code completion [see "Robo-Helpers," p. 39].

Most recently, OpenAI launched ChatGPT, a large-language-model chatbot that is capable of writing code with a little prompting in a conversational manner. This makes it accessible to people who have no prior exposure to programming.

ChatGPT, by itself, is just a natural-language interface for the underlying GPT-3 (and now GPT-4) language model. But what's key is that it is a descendant of GPT-3, as is Codex, OpenAI's AI model that translates natural language to code. This same model powers GitHub Copilot, which is used even by professional programmers. This means that ChatGPT, a "conversational AI programmer," can write both simple and impressively complex code in a variety of different programming languages.

This development sparks several important questions. Is AI going to replace human programmers? (Short answer: No, or at least, not immediately.) Is AI-written or AI-assisted code better than the code people write without such aids? (Sometimes yes; sometimes no.) On a more conceptual level, are there any concerns with AI-written code and, in particular, with the use of natural-language systems such as ChatGPT for this purpose? (Yes, there are many, some obvious and some more metaphysical in nature, such as whether the AI involved really understands the code that it produces.)



The goal of this article is to look carefully at that last question, to place AI-powered programming in context, and to discuss the potential problems and limitations that go along with it.

While we consider ourselves computer scientists, we do research in a business school, so our perspective here very much reflects on what we see as an industry-shaping trend. Not only do we provide a cautionary message regarding overreliance on AI-based programming tools, but we also discuss a way forward.

#### WHAT IS AI-POWERED PROGRAMMING?

**F**irst, it is important to understand, at least broadly, how these systems work.

Large language models are complex neural networks trained on humongous amounts of data—selected from essentially all written text accessible over the Internet. They are typically characterized by a very large number of parameters—many billions or even trillions—whose values are learned by crunching on this enormous set of training data. Through a process called unsupervised learning, large language models automatically learn meaningful representations (known as “embeddings”) as well as semantic relationships among short segments of text. Then, given a prompt from a person, they use a probabilistic approach to generate new text.

In its most elemental sense, what the neural network does is use a sequence of words to choose the next word to follow in the sequence, based on the likelihood of finding that particular word next in its training corpus. The neural network doesn’t always just choose the most likely word, though. It can also select lower-ranked words, which gives it a degree of randomness—and therefore “interestingness”—as opposed to generating the same thing every time.

After adding the next word in the sequence, it just needs to rinse and repeat to build longer sequences. In this way, large language models can create very human-looking output, of various forms: stories, poems, tweets, whatever, all of which can appear indistinguishable from the works people produce.

In creating AI tools for generating code, computer programs can themselves be treated as text sequences, with

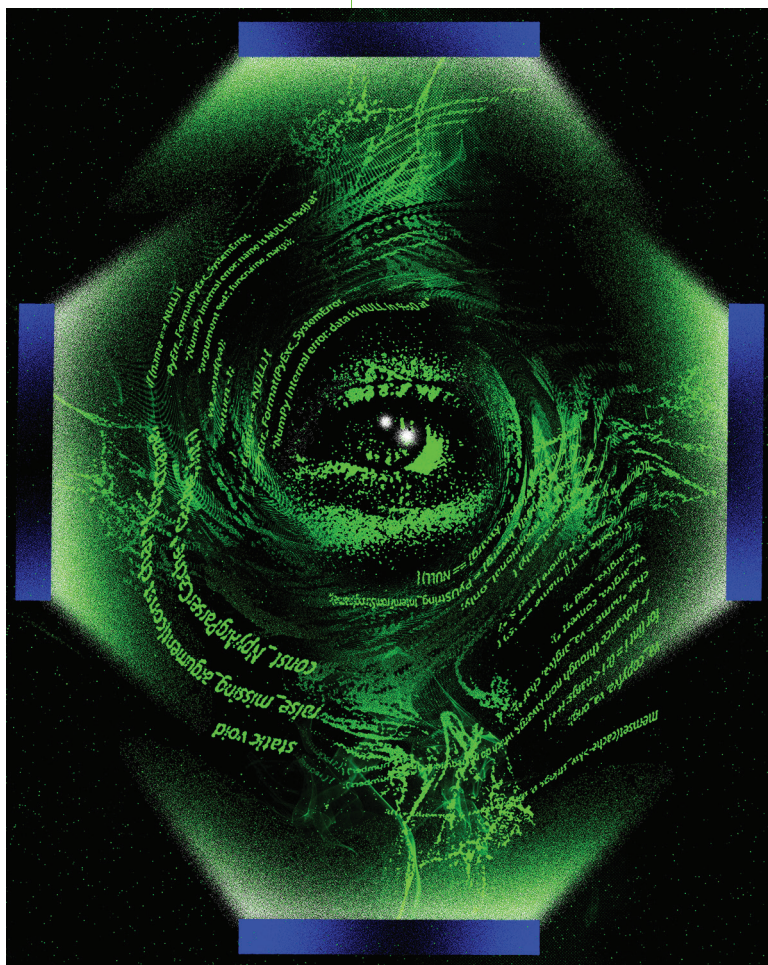
a large language model being trained on code and then used to perform tasks such as code completion, code translation, and even entire programming projects. For example, Codex was trained on a massive dataset of public code repositories, which included billions of lines of code. These models are also fine-tuned to work for specific programming languages or applications, by training the model on a dataset that is specific to the target programming language or type of task at hand.

Even so, the neural network does not have any real understanding of programming, beyond a prescription for how to generate it. So the code that is output can fail on tasks or propagate subtle bugs. One technique these systems use to minimize such issues is to generate a large number of complete programs and then evaluate them against a set of automated tests (the kind many software developers

use), providing as output the program that passes the most tests. In any case, these large language models produce code based on what someone has already written—they cannot come up with genuinely new programming solutions on their own.

#### AYE, ROBOT

**D**espite the many benefits of AI-powered programming, the use of AI here raises significant concerns, many of which have been pointed out recently by researchers and even by the providers of these AI-based tools themselves. Fundamentally, the problem is this: AI programmers are necessarily limited by the data they were trained on, which includes plenty of bad code along with the good. So the code these systems produce may well have problems, too.



# BAD PROGRAMMING ADVICE FROM CHATGPT

The AI seems to know about security issues but sometimes misses the most obvious ones

**CHATGPT LEARNED ABOUT** programming from the code that was available in its massive training set—some of which is good and some of which is bad. So the seemingly authoritative advice it gives can sometimes include code with security vulnerabilities, and you'll need to check it carefully lest you be misled. Here's one illustrative example:

Imagine that you're programming in the Ruby language and want to use a library (which Ruby programmers refer to as a "gem") to perform AES encryption. The Ruby aes gem has encrypt and decrypt functionality, but maybe you want some expert advice on how to use these methods. So you might enter the following prompt into ChatGPT:

## PROMPT:

```
aes ruby gem AES.encrypt
```

## CHATGPT'S RESPONSE:

Sure, here's an example of how to use the 'AES.encrypt' method from the 'aes' Ruby gem to encrypt data with AES-256 encryption:

```
require 'aes'

key = 'mysecretkey12345'
iv = AES.iv(:AES_256_CBC)
plaintext = 'my secret data'

ciphertext = AES.
encrypt(plaintext, key, iv:
iv)
```

In this example, we generate a random initialization vector (IV) using the 'AES.iv' method with the 'AES\_256\_CBC' option to specify the CBC mode of operation. We then call the 'AES.encrypt' method with the plaintext data, the key, and the IV to encrypt the data.

To decrypt the encrypted data, you can use the following code:

```
require 'aes'

key = 'mysecretkey12345'
iv = ciphertext[0..15] #
extract the first 16 bytes as
the IV
encrypted_data = cipher-
text[16..-1] # extract the
```

rest as the encrypted data

```
plaintext = AES.decrypt(en-
crypted_data, key, iv: iv)
```

In this code, we extract the IV and encrypted data from the ciphertext, and then call the 'AES.decrypt' method with the encrypted data, the key, and the IV to decrypt the data.

Note that the 'AES.encrypt' method uses CBC mode by default, which can be vulnerable to certain types of attacks. If you need to use a more secure mode of operation, such as GCM, you can pass additional options to the 'encrypt' and 'decrypt' methods. For example:

```
require 'aes'

key = 'mysecretkey12345'
iv = AES.iv(:AES_256_GCM)
plaintext = 'my secret data'

ciphertext, tag = AES.
encrypt(plaintext, key, iv:
iv, tag: true)
decrypted_data = AES.
decrypt(ciphertext, key, iv:
iv, tag: tag)
```

In this example, we use the 'AES\_256\_GCM' option when generating the random IV to specify the GCM mode of operation. We also pass the 'tag: true' option to the 'AES.encrypt' method to request the authentication tag, and then pass the tag to the 'AES.decrypt' method to verify the integrity of the decrypted data.

ChatGPT appears authoritative. It even warns about the vulnerability of the default CBC encryption mode to certain kinds of attacks. But ChatGPT missed a greater danger: This gem applies the Ruby hex function on the key. So a programmer using it must take care that the key employed is a string that can be interpreted as a series of hexadecimal numbers. Ruby's hex function applied to the key given in ChatGPT's example code would turn it into zero. Using such a key would result in the encryption being completely insecure.

Not only did ChatGPT fail to include a warning about this drastic vulnerability, its example code could also lead a programmer to fall prey to it. And using additional prompts about key security does little to forestall that danger.

First and foremost are issues with security and reliability. Like the code that people write, AI-produced code can contain all manner of security vulnerabilities. Indeed, a recent research study looked at the result of developing 89 different scenarios for Copilot to complete. Of the 1,689 programs that were produced, approximately 40 percent were found to contain vulnerabilities.

To get a better sense of what we mean by a vulnerability, consider something called a buffer-overflow attack, which takes advantage of the way memory is allocated. In such an attack, a hacker tries to input more data into a buffer (a portion of system memory set aside for storing some particular kind of data) than the buffer can accommodate. What happens next depends on the underlying machine architecture as well as the specific code used. It's possible that the extra data will overflow into adjacent memory and thus corrupt it, which could potentially result in unexpected and perhaps even malicious behavior. With carefully crafted inputs, hackers can use buffer overflows to overwrite system files, inject code, or even gain administrative privileges.

Buffer overflows can be prevented through careful programming practices, such as validating user input and limiting the amount of data that can be placed in a buffer, as well as through architectural safeguards. But there are many other kinds of security vulnerabilities: SQL-injection attacks, improper error handling, insecure cryptographic storage and library use, cross-site scripting, insecure direct object references, and broken authentication or session management, to name just a few common attack strategies. Until there is a way to check for all the different kinds of vulnerabilities and automatically remove them, code generated by an AI system is likely to contain these weaknesses.

A more fundamental problem is that there aren't yet ways to formally specify requirements and to verify that these requirements are met. So it's currently impossible to know that the behavior of an AI-generated program matches what it's supposed to do. A related issue is that the code these AI tools produce is not necessarily optimized for any particular attribute, such as scalability. While it may be possible to achieve that with the right prompts, this brings up the question of how to compose such prompts.



Of course, many of these problems exist with the code people write as well. So why should AI-generated code be held to a higher standard?

There are three reasons. First, because the training process utilizes the body of all publicly accessible code, and because there are no straightforward criteria for judging quality, you just don't know how good the code you get from an AI programmer is. The second reason involves psychology. People are apt to believe that computer-generated code will be free of problems, so they may scrutinize it less. And third, because the people using these tools did not create the code themselves, they may not have the skills to debug or optimize it.

There are other thorny issues to consider, too. One is bias, which is insidious: Why did the AI programmer adopt a particular solution when there were multiple possibilities? And what if the approach it adopted is not the best for your application?

Even more problematic are concerns about intellectual property and liability. The data that these models are trained on is often copyrighted. Several legal scholars have argued that the training itself constitutes fair use, but the output of these models may nevertheless infringe on copyrights or violate license terms in the training set. This is particularly relevant because large models can, in many cases, memorize significant parts of the data they are trained on. While there is some very recent work on provable copyright protection for generative models, this area requires significantly more consideration, especially when the notion of a software bill of materials is in the air.

#### PANDORA'S BLACK BOX

**C** Clearly, using any type of automated programming has its dangers. But when these tools are combined with a conversational interface like ChatGPT, the problems are that much more acute. Unlike the AI tools that are primarily used by professional programmers, who should be aware of their limitations, ChatGPT is accessible to everyone. Even novice programmers can use it as a starting point and accomplish quite a lot.

To get a better sense of what is possible, we, along with many others, have

## ROBO-HELPERS

This compilation shows the proliferation of AI-based coding assistants

AI Dev Codes	Code GPT	GPT95	Raycast
AI Query	Code Snippets	Hacker AI	Refact
AirOps AI Data	AI	IntelliCode	Refraction AI
Sidekick	CodeAssist	IntelliSense	Replit
AIWriter	Codefy.ai	Jedi	Safurai
aiXcoder	CodeGeeX	JetBrains	Second
AlphaCode	CodeGen	Datalore	SinCode AI
Amazon	Codeium	K.Explorer	Snappify
CodeWhisperer	CodeSquire	Kite	SourceAI
AskCodi	CodeWP	Kodezi	Sourcegraph
Autocode	CometCore	Krater.ai	Cody
Bard	Denigma	Lightly	SpellBox
Bito	DevBox	Marve Chat	StarCoder
Blackbox AI	DevKit	MutableAI	Stenography
Bloop	ExplainDev	Noya	Tabnine
BotCity	Fig	OpenAI Codex	Vivid
Buildt	Figstack	Phind	Warp AI
ChatGPT	Ghostwriter	Programming	What The Diff
ChatGPT Plus	GitFluence	Helper	Wing Python IDE
Cmd J	GitGab	ProMindGPT	YouChat
Codacy	GitHub Copilot	Q	Zentask

asked ChatGPT to answer some common coding questions posed at hiring interviews. Those carrying out such an exercise have come to a range of conclusions, but in general the results show ChatGPT to be quite an impressive job candidate.

And even if ChatGPT is unable to solve a problem the way you want the first time, you can use additional prompts to get to the desired solution eventually. That's because ChatGPT is conversational and remembers the chat history. This is an immensely attractive feature, which suggests that ChatGPT and its successors will sooner or later become part of the software supply chain. To some extent, these tools are already becoming part of the teaching of computer science, apparently with some benefits to students learning to program.

We nevertheless worry that increased reliance on such technologies will prevent programmers from learning important details about how their code actually functions. That seems inevitable. After all, most programmers, even seasoned professionals, aren't thinking in terms of bit manipulation or what's going on in the registers of a CPU or GPU. They reason at much higher levels of abstraction. While that's generally a good thing, there's a danger that the programs they write with AI assistance will become black boxes to them.

And as we mentioned, the code that ChatGPT and other AI-based programming aids produce often contains security vulnerabilities. Interestingly, ChatGPT itself is sometimes aware of this, and it is able to remove such vulnerabilities if requested to do so. But you have to ask. Otherwise it may give you the simplest possible code, which could be problematic if it is used without further thought.

So where do we go from here? Large language models create a conundrum for the future of computer programming. While it's easy enough to create a fragment of code to tackle a straightforward task, the development of robust software for complex applications is a tricky art, one that requires significant training and experience. Even as the application of large language models for programming deservedly continues to grow, we can't forget the dangers of its ill-considered use.

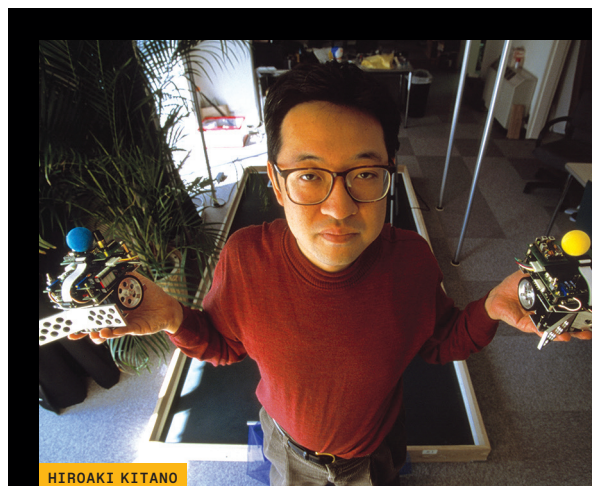
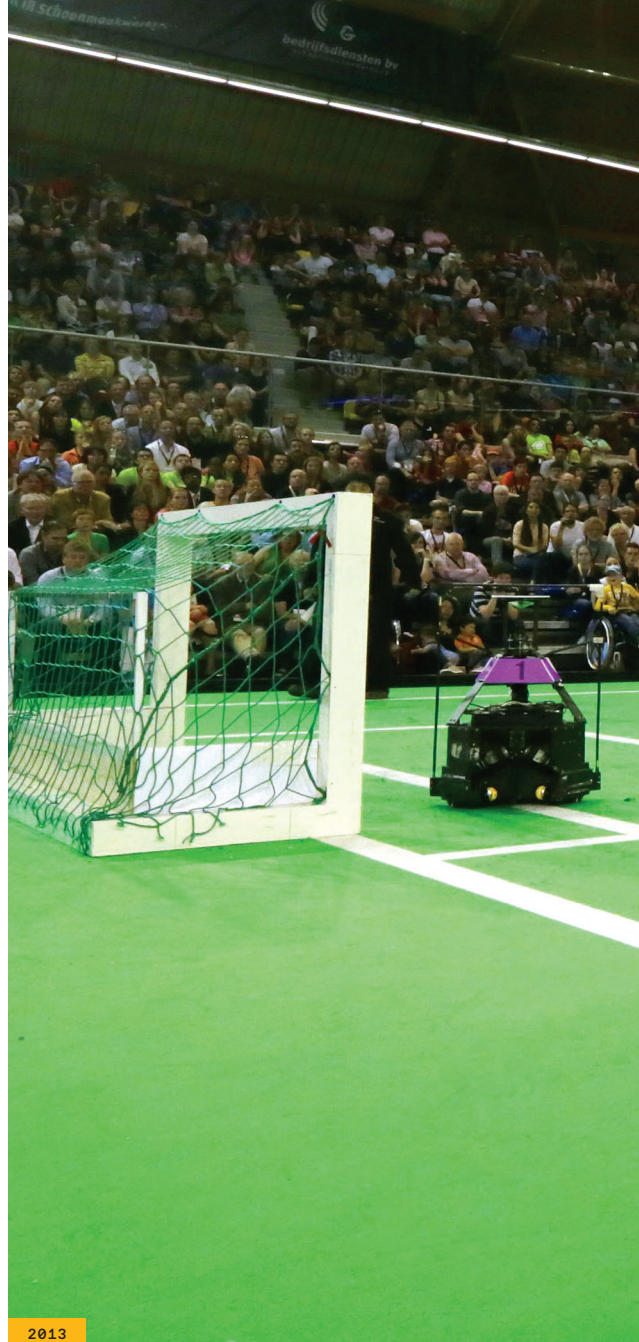
In one way, these models remind us of an aphorism often used to describe working with computers: garbage in, garbage out. And there's plenty of garbage in the training sets these models were built from. Yet they are also immensely capable. ChatGPT, Codex, and other large language models are like the proverbial genie of the lamp, who has the power to give you almost anything you might want. Just be careful what you wish for. ■



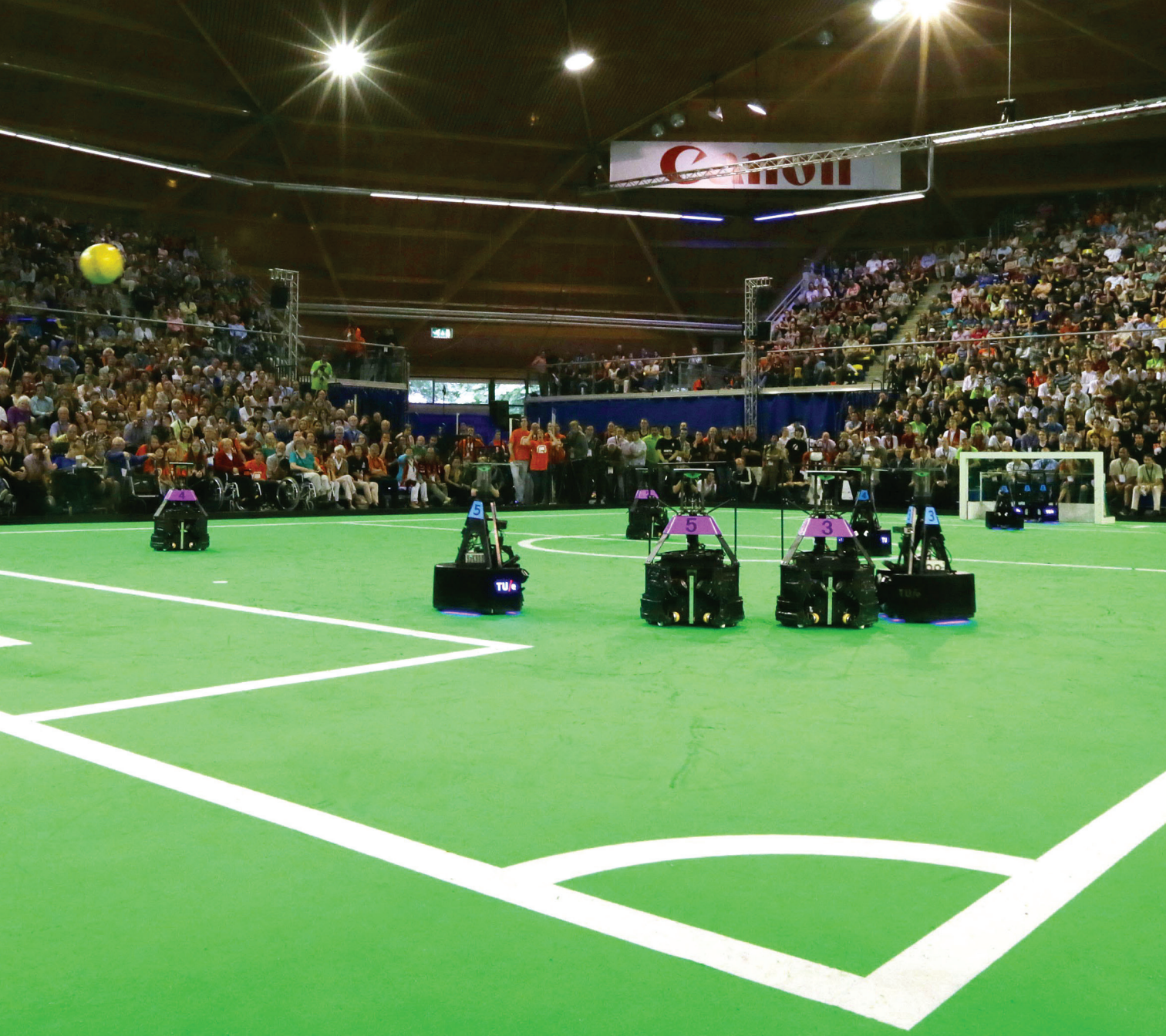
## Designing robots to best World Cup winners has inspired generations of roboticists **By Peter Stone**

**F**ROM 2019 TO 2022, I had the privilege of serving as president of the RoboCup Federation. RoboCup is an annual international competitive event that merges visionary thinking about how AI and robotics will change the world with practical robot design. Participants spend months solving diverse technical problems to enable their robots to autonomously play soccer, do household chores, or search for disaster victims. And their efforts are in turn enabling fundamental advances in a range of fields, including machine learning, multiagent systems, and human-robot interaction. ● RoboCup's original goal, as defined by founding president Hiroaki Kitano [right], was to enable a team of fully autonomous humanoid robots to beat the best human soccer team in the world on a real, outdoor field by the year 2050. Since the first RoboCup competition in 1997 which featured three leagues—small-size wheeled robots, middle-size wheeled robots, and simulation—the event has expanded to include humanoid robot soccer leagues, as well as other leagues devoted to robots with more immediate practicality. The next RoboCup event takes place in July in Bordeaux, France, where 2,500 humans (and 2,000 robots) from 45 countries are expected to compete.

TOP: JOYCE VAN BELKOM/AFP/GETTY IMAGES







1997

**THE BEGINNING** • The first RoboCup, which I attended as a student, was held in 1997 in a small exhibit room at the International Joint Conference on AI (IJCAI) in Nagoya, Japan. The level of competition was, by today's standards, not very high. However, it's important to remember that many "roboticists" back then didn't work with real robots. RoboCup was unusual during its early years in that it forced people to build complete, integrated working systems that could sense, decide, and act.

LEFT: PETER MENZEL/SCIENCE SOURCE;  
RIGHT: PHILIPPE PLAILLY/SCIENCE SOURCE





## Small-Size League

Over the years, RoboCup has seen huge improvements in the level of play, often following a pattern of one team making a discovery and dominating the competition for a year or two and then being supplanted by another. For example, in the Small-Size League, in which the robots use a golf ball and external perception and computing, Team FU-Fighters from Freie Universität Berlin introduced some innovations in the early 2000s. They began controlling the ball using a device that spins it backward toward the robot to “dribble.” A second device propelled the ball quickly forward to shoot it. As the first team to come up with this strategy, the FU-Fighters had a big advantage, but other teams soon followed suit.



1997

PHILIPPE PLAILLY/SCIENCE SOURCE



MICHEL PORRO/GETTY IMAGES





2017



2013

## Standard Platform League

While many RoboCup leagues include a hardware design component, some teams prefer to focus more on software. In the Standard Platform League, each team is provided with identical robots, and thus the best combination of algorithms and software engineering wins. The first standard platform for RoboCup was the Aibo, a small robot dog [at right, top and middle] made by Sony. Ultimately, though, the goal of RoboCup is to achieve human-level performance on a bipedal robot, and so the Standard Platform League now uses a small humanoid robot called Nao, made by SoftBank. Rugged and capable, Nao [bottom right] is able to fall over and quickly get up again, a critical skill for soccer-playing humans and robots alike.

VOLKER STEGER/SCIENCE SOURCE



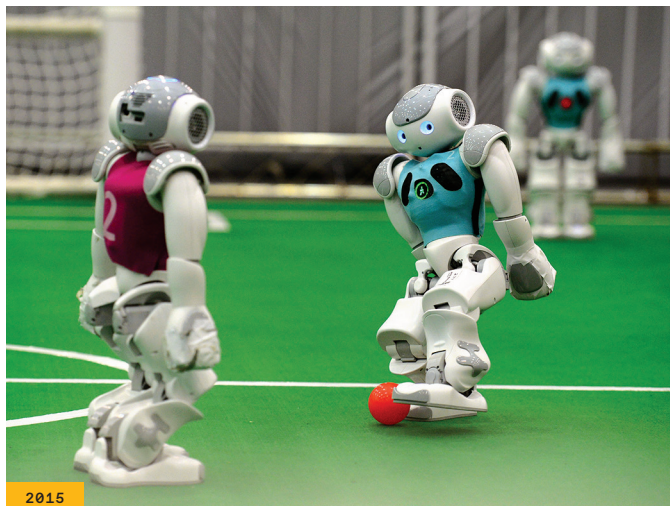
1998

MALTE CHRISTIANS/BONGARTS/GETTY IMAGES

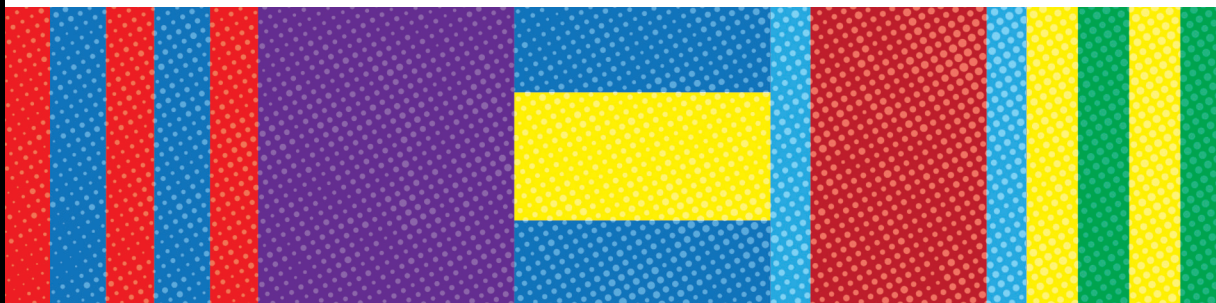


2006

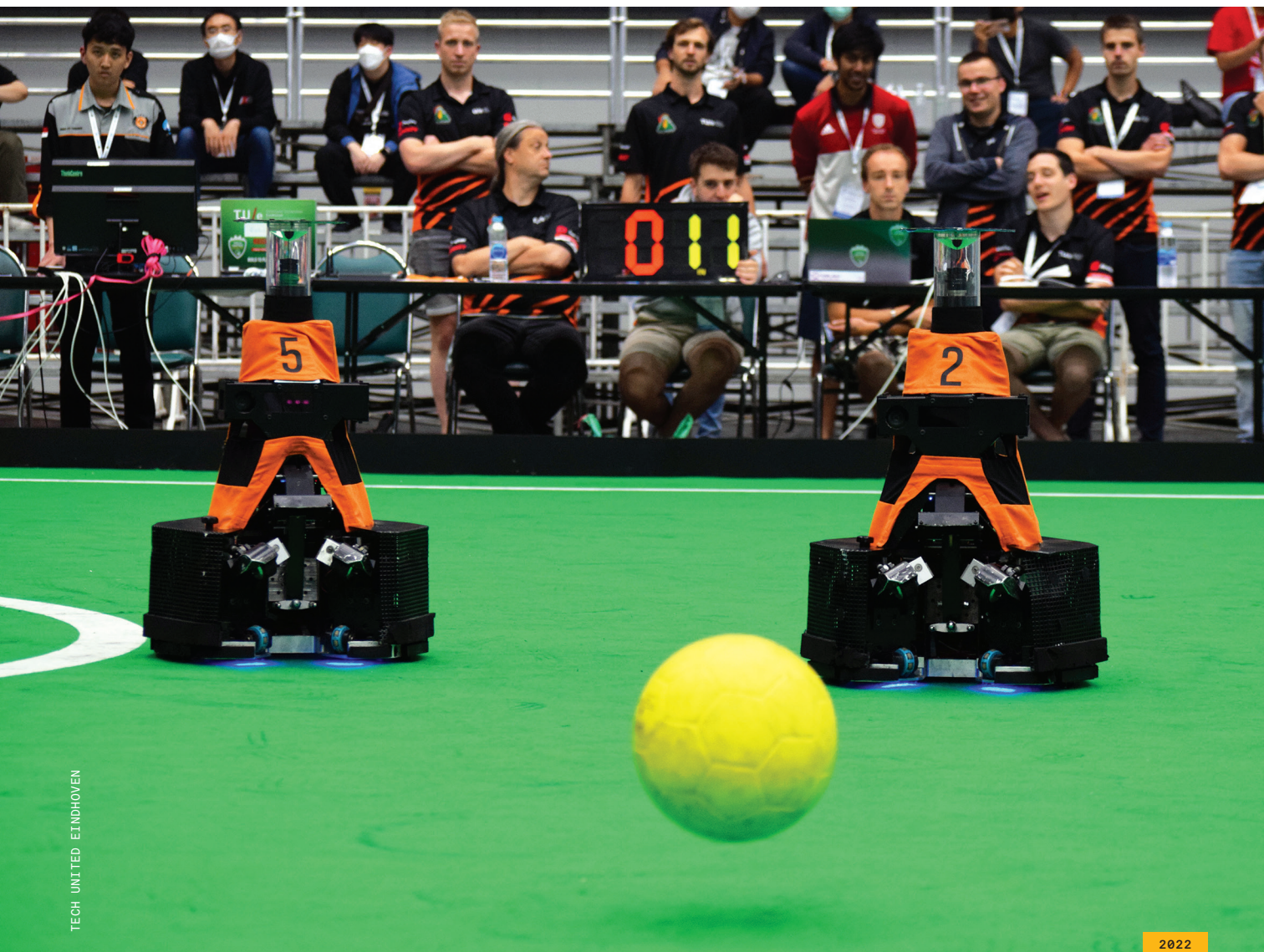
VISUAL CHINA GROUP/GETTY IMAGES



2015



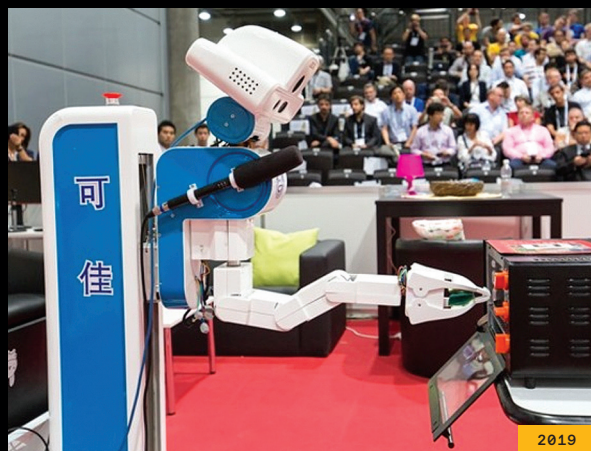




2022

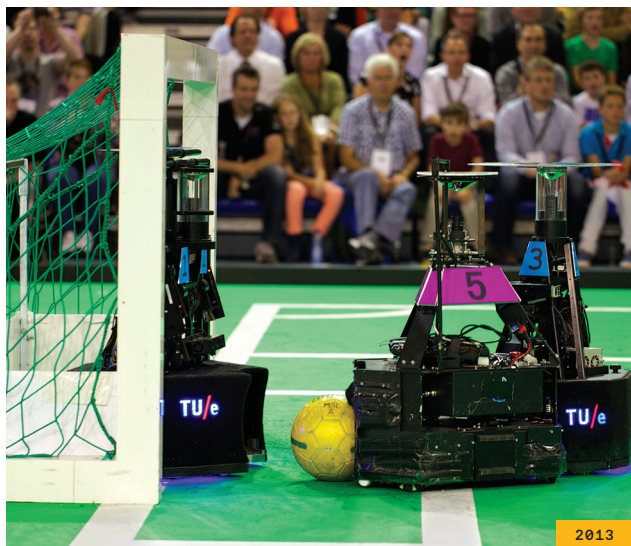
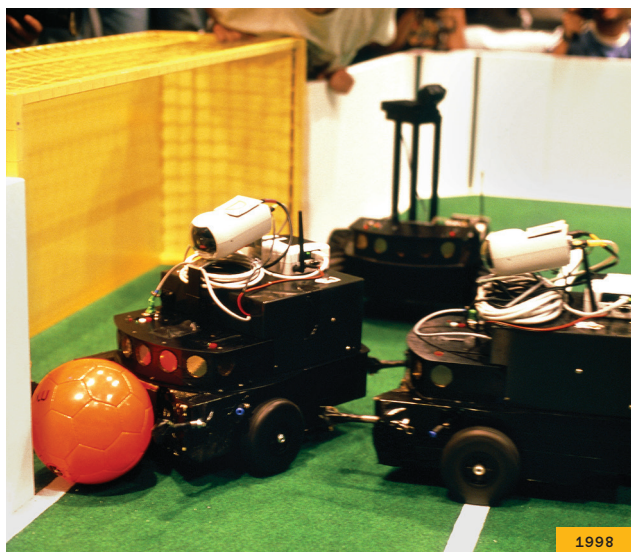
**OTHER LEAGUES** • While soccer is the ultimate goal of RoboCup, and it motivates research on fundamental topics such as robot vision and mobility, it can be hard to see the practicality in a game. Other RoboCup leagues thus focus on more immediate applications. RoboCup@Home [left] features robots for domestic environments, RoboCup Rescue [middle] is for search-and-rescue robots for disaster response, and RoboCup@Work [right] develops robots for industrial and logistics tasks.

ROBOCUP FEDERATION (3)



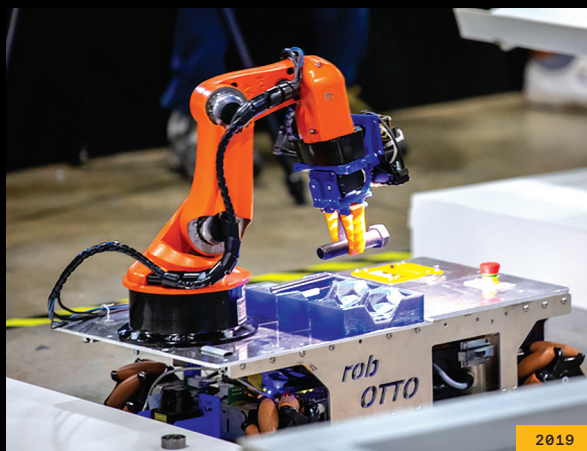
2019



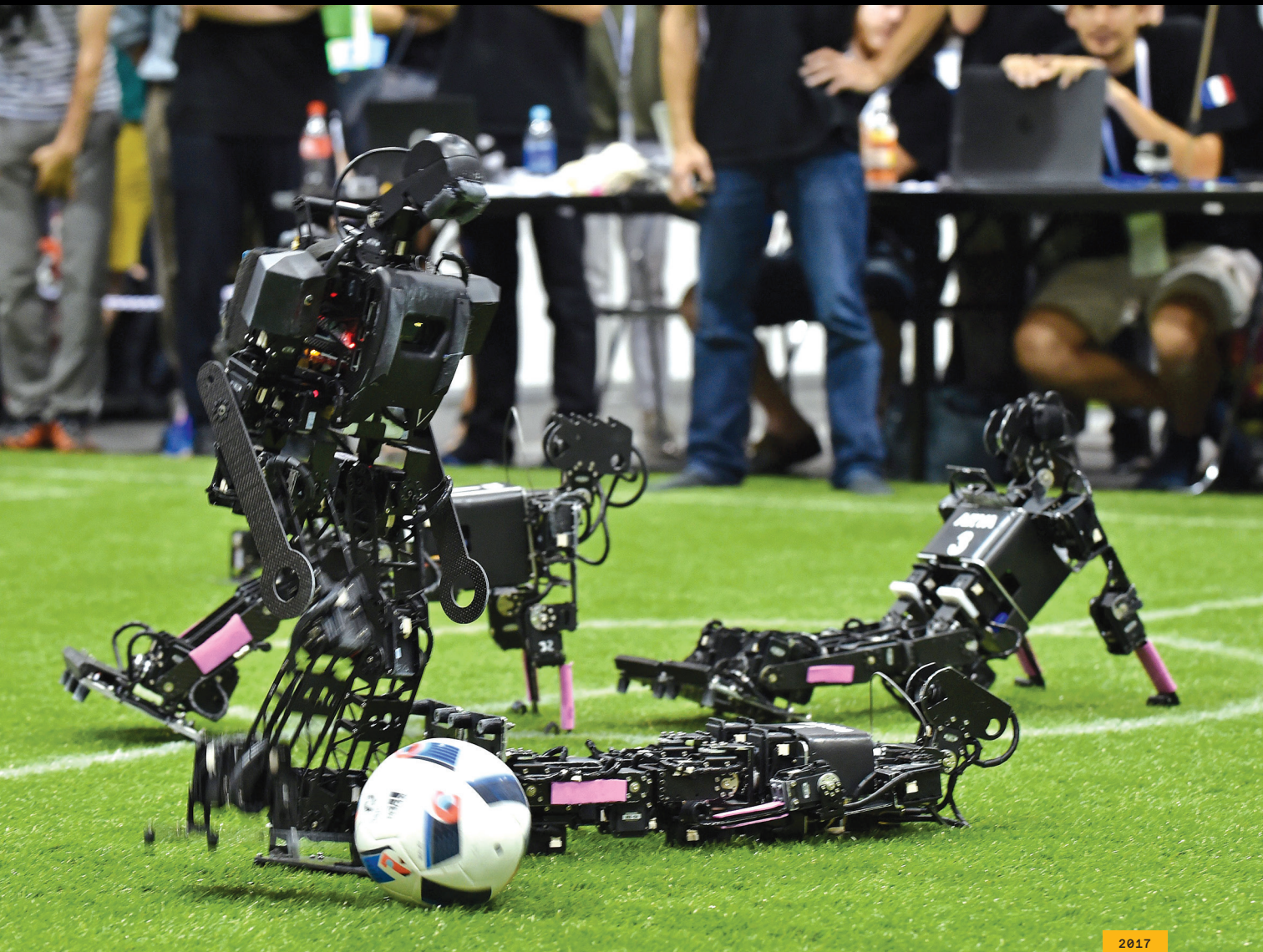


## Middle-Size League

The Middle-Size League uses a full-size soccer ball and onboard perception on waist-high wheeled robots. Over the years, the league has showcased an enormous amount of progress toward human-speed, human-scale soccer. In recent competitions, the robots have moved briskly around a large field, autonomously developing offensive and defensive strategies and coordinating passes and shots on goal. The typical middle-size robot has the skill of a competent primary-school human, although in this league the robots don't have to worry about legs. And in some ways, the middle-size robots have advantages over human players—the robots have omnidirectional sensing, wireless communication, and the ability to consistently place very accurate shots thanks to a mechanical ball-launching system.







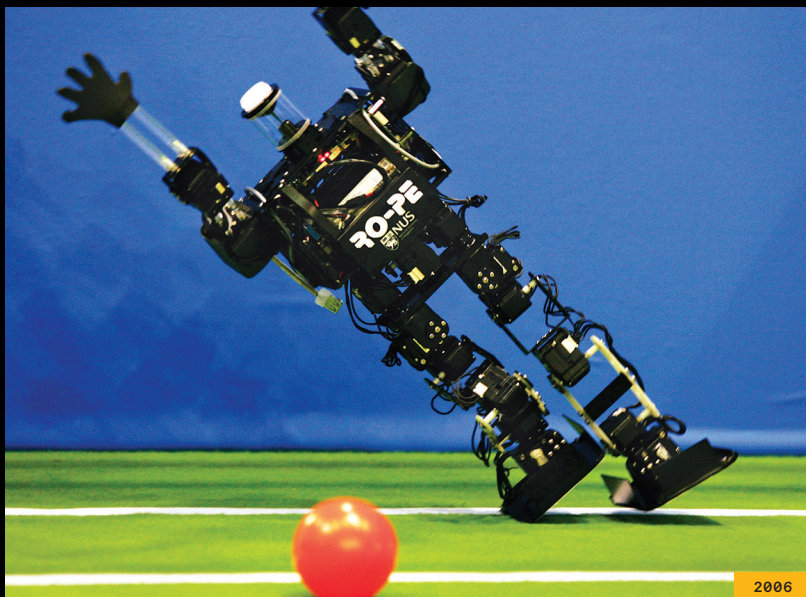
## Humanoid League

The RoboCup Humanoid League, launched in 2002, is critical to meeting our objective of fielding a team of highly skilled humanoid robots by 2050. Bipedal robots are an ongoing challenge, espe-

cially when those robots have to interact with full-size soccer balls, balancing on one leg to kick with the other. The humanoids must have humanlike proportions and sensor configurations akin to human perception—which means, among other things, no omnidirectional sensing.

ABOVE: KAZUHIRO NOGI/AFP/GETTY IMAGES





2006

MALTE CHRISTIANS/BONGARTS/GETTY IMAGES



2017

KAZUHIRO NOSI/AFP/GETTY IMAGES



2022

JACK TAYLOR/AFP/GETTY IMAGES

## RoboCup's Legacy

Compared to 25 years ago, there are now many more robotics competitions to choose from, and applications of AI and robotics are much more widespread. RoboCup inspired many of the other competitions, and it remains the largest such event. Our community is determined to keep pushing the state of the art. The event draws teams from research labs specializing in mechanical engineering, medical robotics, human-robot interaction, learning from demonstration, and many other fields, and there's no better way to train new students than to encourage them to immerse themselves in RoboCup.

The importance of RoboCup can also be measured beyond the competition itself. One of the most notable successes, stemming from the early years of the competition, was the technology spun off from the Small-Size League to form the basis of Kiva Systems. The hardware of Kiva's robot was designed by Cornell's RoboCup team, led by Raffaello D'Andrea. His team racked up Small-Size League victories in 1999, 2000, 2002, and 2003, and D'Andrea went on to cofound Kiva Systems. The company, which developed warehouse robots that moved shelves of goods, was acquired by Amazon in 2012 for US \$775 million.



## Robots vs. Humans

At the conclusion of a RoboCup event, there has been a tradition since 2011 of the trustees of the RoboCup Federation playing a friendly game against the winning team of the Middle-Size League. In recent years, the middle-size robots have become surprisingly competitive, able to keep possession of the ball, dribble around the opposing team, and string together passes across the field. The robots may not be ready to take on the world champions quite yet, but the progress has been impressive—in 2022, Tech United Eindhoven played a friendly match against a Portuguese professional women's team, Vitória SC [right], and the robots managed to score several goals (after the women took it easy on them).

RIGHT: TECH UNITED EINDHOVEN (3)



**SIMULATION** • Even in the Standard Platform League, hardware can be frustrating, so the RoboCup Simulation League allows teams to work entirely in software. This enables more rapid progress using cutting-edge techniques. My own team, UT Austin Villa, started using hierarchical machine learning to develop skills such as walking and kicking in the Simulation League in 2014, which allowed us to dominate the competition. But in 2022, FC Portugal and Magma Offenburg were able to surpass us with deep reinforcement learning methods.





EUROPEAN ROBOCUP 2022



2022



2022

## Future of RoboCup

At this point, you may be wondering what the prospects are for achieving RoboCup's founding goal—enabling a team of autonomous humanoid robots to beat the world's best human team at a game of soccer on a real, outdoor field by the year 2050. Will soccer go the way of chess, checkers, poker, *Gran Turismo*, “Jeopardy!”, and other human endeavors and be conquered by AI? Or will the requirements for real-world perception and humanlike speed and agility keep soccer out of reach for robots? This question remains a source of uncertainty and debate within the RoboCup community. Although 27 years is a very long time in technological terms, physical automation tends to be significantly harder and take much longer than purely software-oriented tasks do.

Ultimately, if the community is going to achieve its goals, we will need to address two challenges: building hardware that can move as quickly and easily as people do, and creating software that can outsmart the best human players in real time. Some experts point to existing state-of-the-art humanoid robots as evidence that sufficiently capable hardware is already available. As impressive as they are, however, I don't think these robots can match the capabilities of the most skilled human athletes just yet. I haven't seen any evidence that even the best humanoid robots today can dribble a soccer ball and deftly change directions at high speed in the way that a professional soccer player can—especially when factoring in the requirement that for professional players to agree to get on the field with robots, the robots will need to be not too heavy or powerful: They will need to be both skilled and eminently safe.

Regardless of how it turns out, there is no question in my mind that RoboCup is an enduring grand challenge for AI and robotics, as well as a great training ground for the next generation of roboticists. The RoboCup community is thriving, generating new ideas and new engineers and scientists. I've been proud to have led the RoboCup organization, and look forward to seeing where it will go from here. ■





The Global Institute of Future Technology (GIFT hereafter) at Shanghai Jiao Tong University sincerely invites applications for tenure-track or tenured professoriate and research scientist positions, mainly at Associate and Full levels, related to Sustainable Energy.

GIFT manages endowments from a variety of sources, including a recent donation of more than 200 million US dollars. GIFT's new building will be completed within two years, adding 50,000 square meters to the current space. With strong government and university support, we will rapidly expand our faculty, including high-level hiring (e.g. NAE/NAS members and Fellows of multiple professional societies). At present, five research centers have been established: Large-scale Energy Storage and Reliability Center, Solid-state Battery Center, Advanced Sustainable Power Systems Center, Green Energy and Modern Agriculture Center, and Perovskite Photovoltaics Center.

Applicants must hold a doctorate degree in a relevant field. Candidates are expected to establish a vigorous research program and contribute to undergraduate and graduate education. For full consideration, please send a CV, teaching plan and research plan, copies of three representative publications and contact information of five referees as a single PDF file to the GIFT Search Committee (email: [gift-facultysearch@sjtu.edu.cn](mailto:gift-facultysearch@sjtu.edu.cn)). For more information, please visit <http://gift.sjtu.edu.cn>.



## One of the most influential reference resources for engineers around the world.

For over 100 years, *Proceedings of the IEEE* has been the leading journal for engineers looking for in-depth tutorial, survey, and review coverage of the technical developments that shape our world.



To learn more and start your subscription today, visit

[ieee.org/proceedings-subscribe](http://ieee.org/proceedings-subscribe)



# IEEE Access®

Multidisciplinary | Rapid Review | Open Access Journal

Harness the publishing power of IEEE Access.™



- Receive high-quality, rigorous peer review in only 4 to 6 weeks
- Reach millions of global users through the IEEE Xplore® digital library by publishing open access

Learn more at [ieeaccess.ieee.org](http://ieeaccess.ieee.org)



# TAP. CONNECT. NETWORK. SHARE.



## Connect to IEEE—no matter where you are—with the IEEE App.



Stay up-to-date with the latest news



Schedule, manage, or join meetups virtually



Get geo and interest-based recommendations



Read and download your IEEE magazines



Create a personalized experience



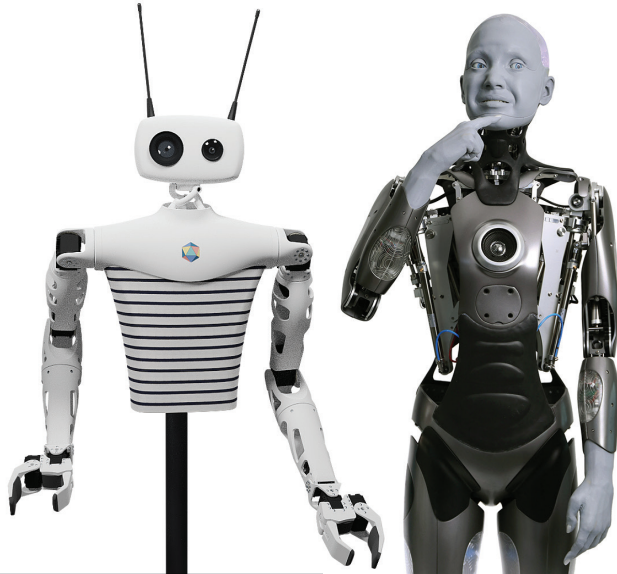
Locate IEEE members by location, interests, and affiliations

**Download Today!**



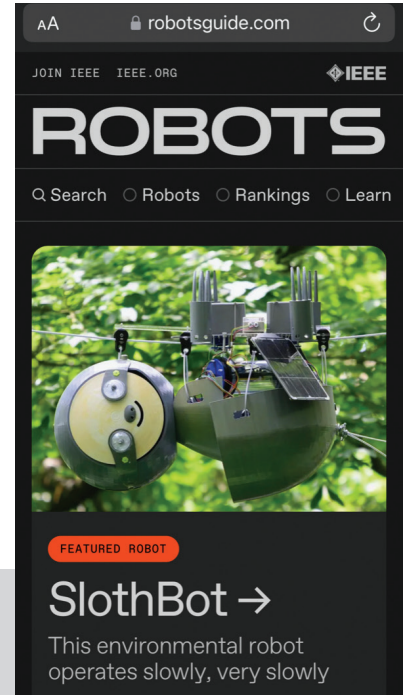


# The World's Best Robots Guide Just Got Better!



LEFT: REACHY BY POLLEN ROBOTICS;  
RIGHT: AMECA BY ENGINEERED ARTS

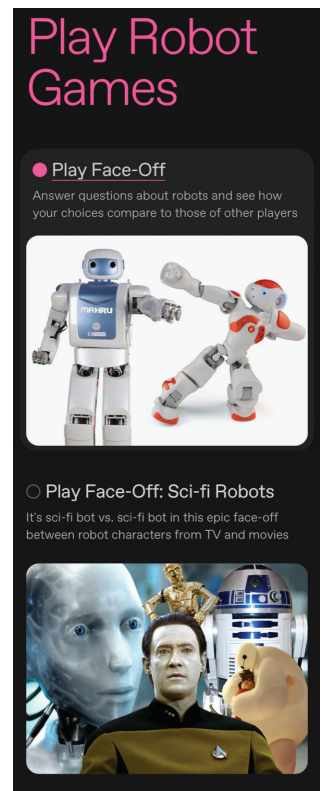
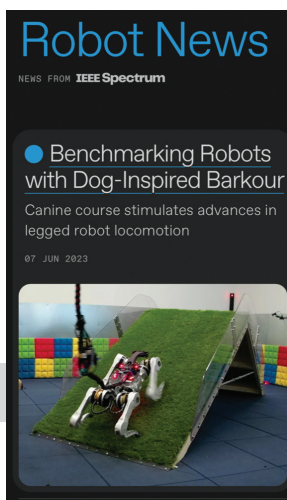
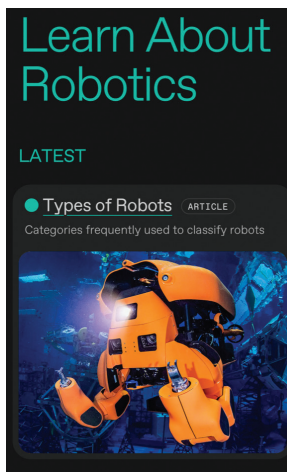
The new redesigned and improved  
**ROBOTSGUIDE.COM**



**IEEE Spectrum's Robots Guide** features almost **250 robots** from around the world. Designed for robot enthusiasts of all ages, researchers, STEM educators, and students.

- **NEW:** Explore robots with new advanced filtering by type and features.
- Make robots move and spin 360 degrees.
- View thousands of photos and videos of robots in action.
- **NEW:** Play the Sci-Fi Robots Face-Off game.
- Keep up to date with daily robot news.
- Read detailed tech specs for each robot.

Check out **Robotsguide.com** on your desktop, tablet, or phone now!



# Past Forward

Spanish engineer Leonardo Torres Quevedo built this chess-playing automaton to demonstrate his ideas about machines that “think.”



## Computer Chess, Circa 1920

In 1912, Leonardo Torres Quevedo invented what might be considered the first computer game—a simplified chess-playing automaton. The machine played a king and a rook against a human's lone king. Eight years later, Torres Quevedo upgraded the appearance and mechanics of his automaton [shown here]. This version moved its pieces using electromagnets concealed below the chessboard. A gramophone recording announced *jaque al rey* (Spanish for “check”) or *mate* (checkmate). If the human attempted an illegal move, a lightbulb gave a warning signal; after three illegal attempts, the game would shut down. Torres Quevedo explained that machines needed to advance in their intellectual capacity if they were to relieve humans of burdensome mental tasks, but the fix was in: His algorithm for the chess-playing automaton guaranteed a win for the machine every time. ■

FOR MORE ON LEONARDO TORRES QUEVEDO, see [spectrum.ieee.org/pastforward-jul2023](https://spectrum.ieee.org/pastforward-jul2023)



A photograph of a man in a wheelchair, seen from behind, holding up a young girl in a white t-shirt and jeans. They are in a grassy field with trees in the background. The man is wearing a blue button-down shirt. The girl is smiling and has her arms outstretched. The image is split by a diagonal line, with the left side being a solid dark blue color.

# BILLS DON'T STOP ON DISABILITY.

## WHY SHOULD YOUR INCOME?

Help protect your  
livelihood and lifestyle.

**1-800-493-IEEE (4333)**

To learn more\*, visit **[IEEEinsurance.com/Lift](https://IEEEinsurance.com/Lift)**



**IEEE**

**DISABILITY INSURANCE**

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). \*For information on features, costs, eligibility, renewability, limitations and exclusions visit [IEEEinsurance.com/Lift](https://IEEEinsurance.com/Lift).

Program Administered by AMBA Administrators, Inc.

In CA d/b/a Association Member Benefits & Insurance Agency

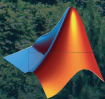
CA Insurance License #0196562 | AR Insurance License #100114462

99695 (7/23) Copyright 2023 AMBA. All rights reserved.

# MATLAB FOR AI

Boost system design and simulation with explainable and scalable AI. With MATLAB and Simulink, you can easily train and deploy AI models.

*[mathworks.com/ai](https://mathworks.com/ai)*



**MathWorks®**

*Accelerating the pace of engineering and science*