FOR THE TECHNOLOGY INSIDER

JUNE 2023

IEEE Spectrum

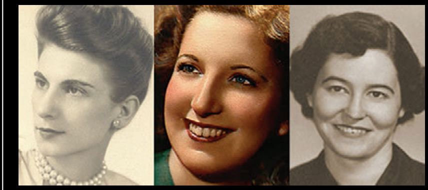
The Great AI Treasure Hunt

We're running out of critical metals for EV batteries. Can machine learning help discover more?

IEEE

VOLUME 60 / ISSUE 6 JUNE 2023

Contents



FROM LEFT: THE SPENCE FAMILY; MELTZER FAMILY; TEITELBAUM FAMILY

ON THE COVER:
Illustration by Max-o-matic

Jayne Griffith is an environmental engineering student at Ohio State University, in Columbus.

Photo by Peter Adams

AI Everywhere, All at Once

It's time to get serious about regulating artificial intelligence

BY HARRY GOLDSTEIN

It's been a frenetic six months since OpenAI introduced its large language model ChatGPT to the world at the end of last year. Every day since then, I've had at least one conversation about the consequences of the global AI experiment we find ourselves conducting. We aren't ready for this, and by we, I mean everyone—individuals, institutions, governments, and even the corporations deploying the technology today.

The sentiment that we're moving too fast for our own good is reflected in an open letter calling for a pause in AI research, which was posted by the Future of Life Institute and signed by many AI luminaries, including some prominent IEEE members. As News Manager Margo Anderson reports online in *The Institute*, signatories include senior member and IEEE's AI Ethics Maestro Eleanor "Nell" Watson and IEEE Fellow and chief scientist of software engineering at IBM, Grady Booch. He told Anderson, "These models are being unleashed into the wild by corporations who offer no transparency as to their corpus, their architecture, their guardrails, or the policies for handling data from users. My experience and my professional ethics tell me I must take a stand...."

But research and deployment haven't paused, and AI is becoming essential across a range of domains. For instance, Google has applied deep-reinforcement learning to optimize placement of logic and memory on chips, as Senior Editor Samuel K. Moore reports in "An Ugly Chapter in Chip Design," on page 5. The cofounders of KoBold Metals explain in "AI Hunts for Hidden Minerals," page 22, how they use machine-learning models to search for minerals needed for electric-vehicle batteries.

Somewhere between the proposed pause and headlong adoption of AI lie the social, economic, and political challenges of crafting the regulations that tech CEOs like OpenAI's Sam Altman and Google's Sundar Pichai have asked governments to create.

> **"These models are being unleashed into the wild by corporations who offer no transparency as to their corpus, their architecture, their guardrails, or the policies for handling data from users."**
> —Grady Booch, IBM

To help make sense of the current AI moment, I talked with *IEEE Spectrum* senior editor Eliza Strickland, who recently won a Jesse H. Neal Award for best range of work by an author, for her biomedical, geoengineering, and AI coverage. Trustworthiness, we agreed, is probably the most pressing near-term concern. Addressing the provenance of information and its traceability is key. Otherwise people may be swamped by so much bad information that the fragile consensus among humans about what is and isn't real totally breaks down.

The European Union is ahead of the rest of the world with its proposed Artificial Intelligence Act. It assigns AI applications to three risk categories: Those that create unacceptable risk would be banned, high-risk applications would be tightly regulated, and applications deemed to pose few if any risks would be left unregulated.

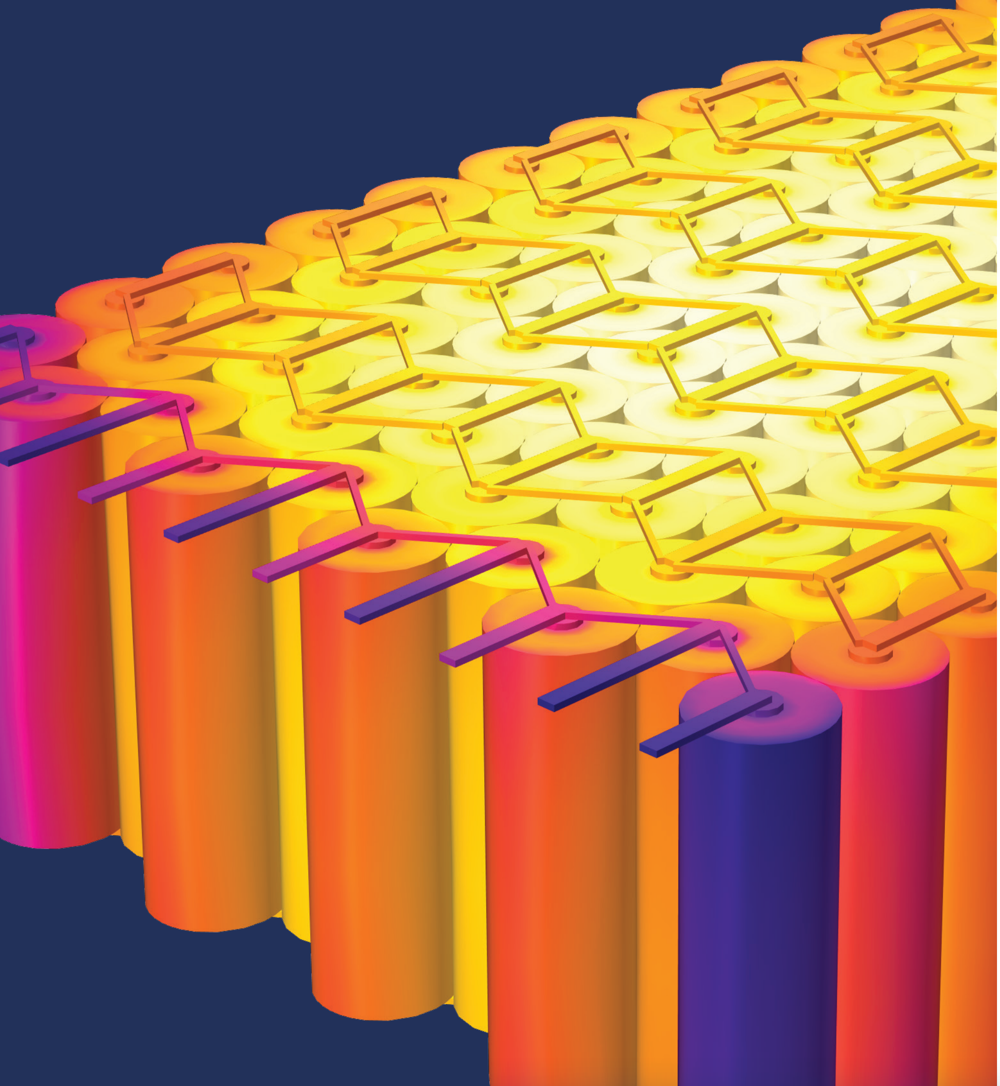
The EU's draft AI Act touches on traceability and deepfakes, but it doesn't specifically address generative AI—deep-learning models that can produce high-quality text, images, or other content based on its training data. However, a recent article in *The New Yorker* by the computer scientist Jaron Lanier directly takes on provenance and traceability in generative AI systems.

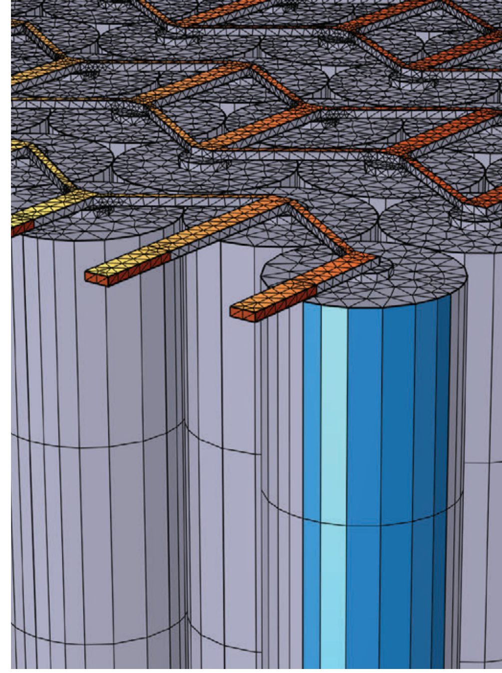
Lanier views generative AI as a social collaboration that mashes up work done by humans. He has helped develop a concept dubbed "data dignity," which loosely translates to labeling these systems' products as machine generated based on data sources that can be traced back to humans, who should be credited with their contributions. "In some versions of the idea," Lanier writes, "people could get paid for what they create, even when it is filtered and recombined through big models, and tech hubs would earn fees for facilitating things that people want to do."

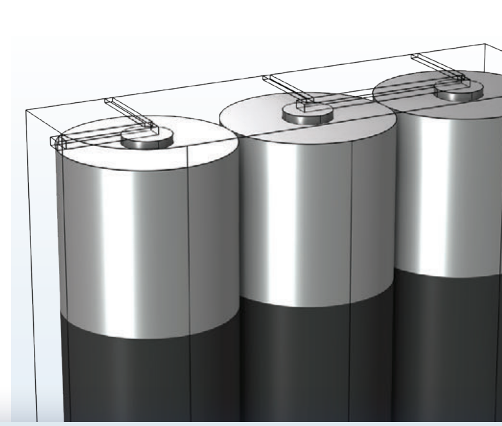
That's an idea worth exploring right now. Unfortunately, we can't prompt ChatGPT to spit out a global regulatory regime to guide how we should integrate AI into our lives. Regulations ultimately apply to the humans currently in charge, and only we can ensure a safe and prosperous future for people and our machines. ■

PORTRAIT BY SERGIO ALBIAC

Illustration by Daniel Zender







PRACHI PATEL

Patel, a longtime contributing editor, covers energy, battery technology, and other topics for *IEEE Spectrum*. For "Building a U.S. Semiconductor Workforce" [p. 28] she explores how universities are adapting to the renewed focus on semiconductor manufacturing brought on by the CHIPS Act. Several new fab projects are underway in states in the U.S. Midwest; Patel says she was struck by how these states try "to leverage their traditional manufacturing efforts for the chip industry."

SCOTT J. SHAPIRO

Shapiro is a professor of law and philosophy at Yale, where he teaches a class on cybersecurity to law students. They typically "don't even know what Linux is" at the start, but "by the end of 12 classes, they are kind of dangerous," Shapiro says. Teaching that class prompted him to write a book about hacking, part of which is the basis for his article about a New Jersey teenager's malware and how it threatened the entire Internet [p. 42].

MATTHEW S. SMITH

Smith is a freelance consumer-tech journalist, an avid gamer, and a longtime aviation enthusiast. He took to the virtual skies with a version of *Flight Simulator 95* that was bundled with his family's first home computer. His grandfather Chester had a pilot's license, and he helped Matthew learn how to keep the blue side up. In this issue, Smith relates the long journey of one of the most consequential computer games ever [p. 36].

MÁXIMO TUJA

Known as Max-o-matic, Tuja is the Barcelona-based designer and illustrator who created our cover art. He describes what he does as "creating imaginary worlds from torn pieces of the real one." Using bold colors and imagery and a mixture of different styles and media, Max-o-matic produces handmade and digital collages that tell complex stories in a compelling way. His past clients include Apple, Spotify, Nike, *The New York Times*, *The Guardian*, *Billboard*, *Mother Jones*, *The Wall Street Journal*, and *Rolling Stone*.

IEEE Spectrum

ACTING 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, t.perry@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 Vrielink, e.vrielink@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
Felicia Spagnoli, f.spagnoli@ieee.org
SENIOR ADVERTISING PRODUCTION COORDINATOR
Nicole Evans Gyimah, n.gyimah@ieee.org
ADVERTISING PRODUCTION +1 732 562 6334

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**

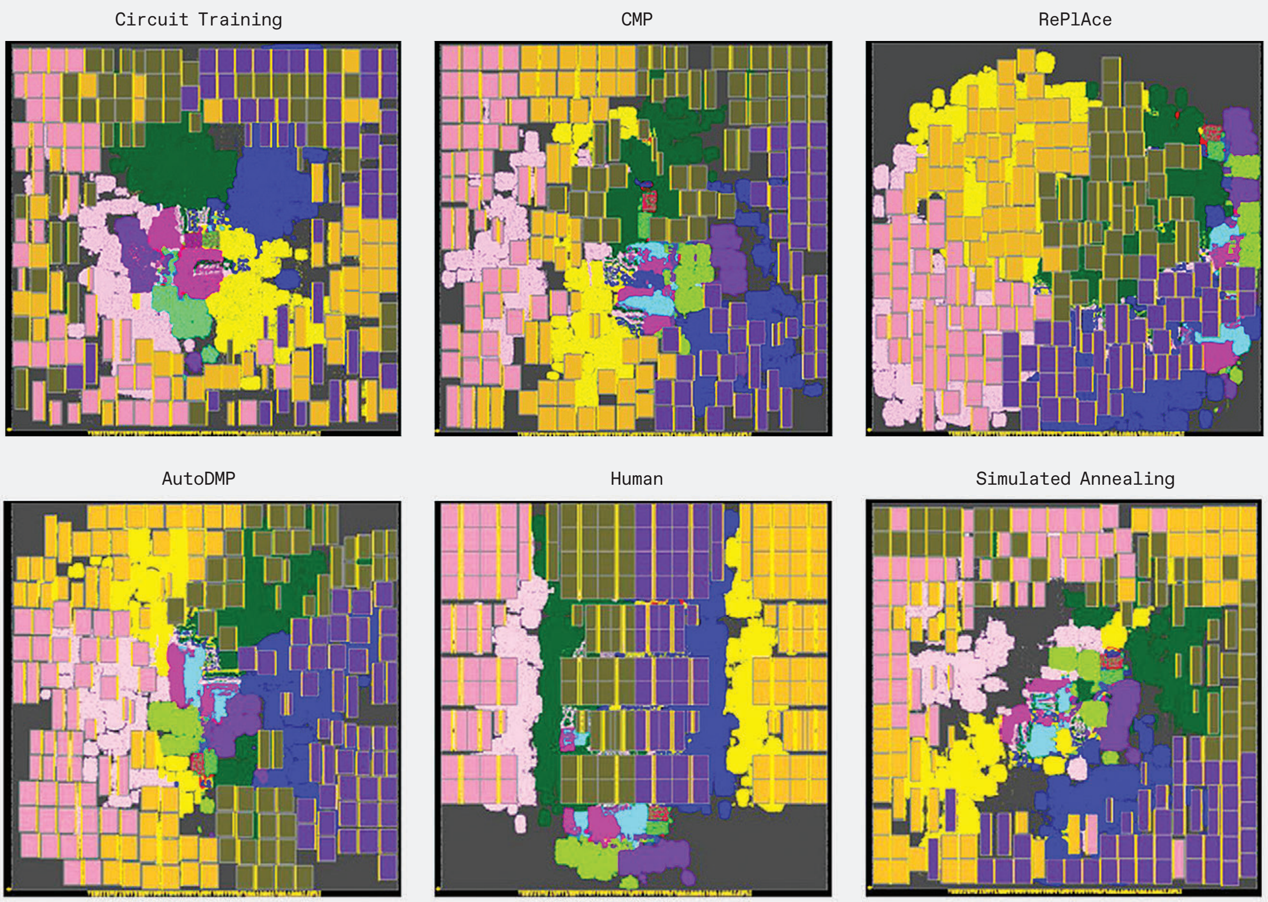
ADVERTISING INQUIRIES Naylor Association Solutions, Erik Albin +1 352 333 3371, ealbin@naylor.com

REPRINT SALES +1 212 221 9595, ext. 319

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. 6. The editorial content of IEEE Spectrum magazine does not represent official positions of the IEEE or its organizational units. Canadian Post International Publications Mail (Canadian Distribution) Sales Agreement No. 40013087. Return undeliverable Canadian addresses to: Circulation Department, IEEE Spectrum, Box 1051, Fort Erie, ON L2A 6C7. Cable address: ITRIPLEE. Fax: +1 212 419 7570. INTERNET: spectrum@ieee.org. ANNUAL SUBSCRIPTIONS: IEEE Members: $21.40 included in dues. Libraries/institutions: $399. POSTMASTER: Please send address changes to IEEE Spectrum, ℅ Coding Department, IEEE Service Center, 445 Hoes Lane, Box 1331, Piscataway, NJ 08855. Periodicals postage paid at New York, NY, and additional mailing offices. Canadian GST #125634188. Printed at 3401 N. Heartland Dr, Liberty, MO 64068, U.S.A. Printed at Spenta Multimedia Pvt. Ltd., Plot 15, 16 & 21/1, Village Chikhloli, 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.

News



Circuit Training [top left], a reinforcement-learning method developed by Google, competed with various other AI methods and a human to see which could best place large blocks of circuitry on a chip.

CHUNG-KUAN CHENG ET AL.

SEMICONDUCTORS

An Ugly Chapter in Chip Design › Study tries to settle a bitter disagreement over Google's AI tool

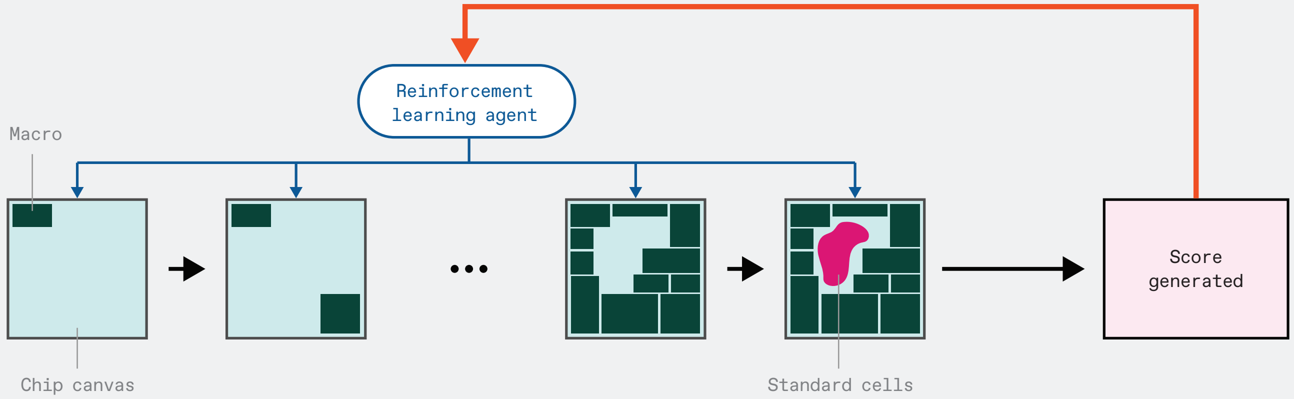
SAMUEL K. MOORE

Discussions at chip design conferences rarely get heated. But a year ago at the International Symposium on Physical Design (ISPD), things got out of hand. It was described by observers as a "trainwreck" and an "ambush." The crux of the clash was whether Google's AI-fueled solution to one of chip design's thornier problems was really better than those of humans and state-of-the-art algorithms. It pitted established male electronic design automation (EDA) experts against two young female Google computer scientists, and the underlying argument had already led to the firing of one Google researcher.

This year at that same conference, a leader in the field, IEEE Fellow Andrew Kahng, hoped to put an end to the acrimony once and for all. He and colleagues at the University of California, San Diego, delivered what he called "an open and transparent implementation and assessment" of Google's reinforcement learning approach. Using Google's open-source version of its process, called Circuit Training, and reverse engineering some parts that were not clear enough for Kahng's team, they set reinforcement learning against a human designer, commercial software, and state-of-the-art academic algorithms.

In most cases, Circuit Training was not the winner, but it was competitive. And for various reasons, the study did not test Circuit Training's signature ability—to improve its performance by learning from other chip designs.

"Our goal has been clarity of understanding that will allow the community to move on," he told engineers this year at ISPD 2023. Only time will tell whether it worked.



Google's reinforcement learning system treats the placement of large circuit blocks, called macros, as a game. The agent places one block at a time on the chip canvas. Then a separate algorithm fills in smaller parts called standard cells. The placement is scored according to several metrics, and that score is used as feedback to improve the agent.

The Hows and the Whens

The problem in question is called placement—determining where chunks of logic or memory should be placed on a chip. Competing considerations must be weighed against one another, with an eye toward maximizing the chip's operating frequency while minimizing its power consumption and the area it takes up. Finding an optimal solution to this puzzle is among the most difficult problems around, with more possible permutations than the game Go.

But Go was ultimately defeated by a type of AI called deep reinforcement learning, and that's what former Google Brain researchers Azalia Mirhoseini and Anna Goldie applied to the placement problem. The scheme, then called Morpheus, treats placing large pieces of circuitry, called macros, as a game, learning how to find an optimal solution as it plays. (The locations of macros have an outsize impact on the chip's characteristics. In Circuit Training and Morpheus, a separate algorithm fills in the gaps with the smaller parts, called standard cells. Other methods use the same process for both macros and standard cells.)

Briefly, this is how it works: The chip's design file starts as what's called a netlist—specifying which macros and cells are connected to which others according to what constraints. The standard cells are then collected into clusters to help speed up the training process. Circuit Training then starts placing the macros on the chip "canvas" one at a time. When the last one is down, a separate algorithm fills in the gaps with the standard cells, and the system spits out a quick evaluation of the attempt, encompassing the wiring's length (longer is worse), how densely packed it is (more dense is worse), and its degree of congestion (you guessed it, worse). Called proxy cost, this acts like the score in a video game. The score is used as feedback to adjust the neural network, and it tries again. Wash, rinse, repeat. When the system has finally learned its task, commercial software does a full evaluation of the complete placement, generating the kind of metrics that chip designers care about, such as area, power consumption, and constraints on frequency.

Mirhoseini and Goldie published the results and method of Morpheus in *Nature* in June 2021, following a seven-month review process. (Kahng was reviewer No. 3.) And the technique was used to design more than one generation of Google's TPU AI accelerator chips. In January 2022, Google released an open-source version of Circuit Training on GitHub. But Kahng and others claim that even this version was not complete enough to reproduce the research.

In response to the *Nature* publication, a separate group of engineers, mostly within Google, began research aimed at what they believed to be a better way of comparing reinforcement learning to established algorithms. But this was no friendly rivalry. According to press reports, the group's leader, Google researcher Satrajit Chatterjee, repeatedly undermined Mirhoseini and Goldie personally and was fired for it in 2022.

While Chatterjee was still at Google, his team produced a paper titled "Stronger Baselines," critical of the research published in *Nature*. He sought to have it presented at a conference, but after review by an independent resolution committee, Google refused. After his termination, an early version of the paper was leaked via an anonymous Twitter account just ahead of ISPD in 2022, leading to the public confrontation.

Benchmarks, Baselines & Reproducibility

When *IEEE Spectrum* spoke with EDA experts following ISPD 2022, detractors had three interrelated concerns—benchmarks, baselines, and reproducibility.

Benchmarks are openly available circuit designs on which researchers test their new algorithms. Commonly used benchmarks when Google began its work were from designs that were already about two decades old. Although there is some debate about their relevance to modern chips, they continue to enjoy broad support.

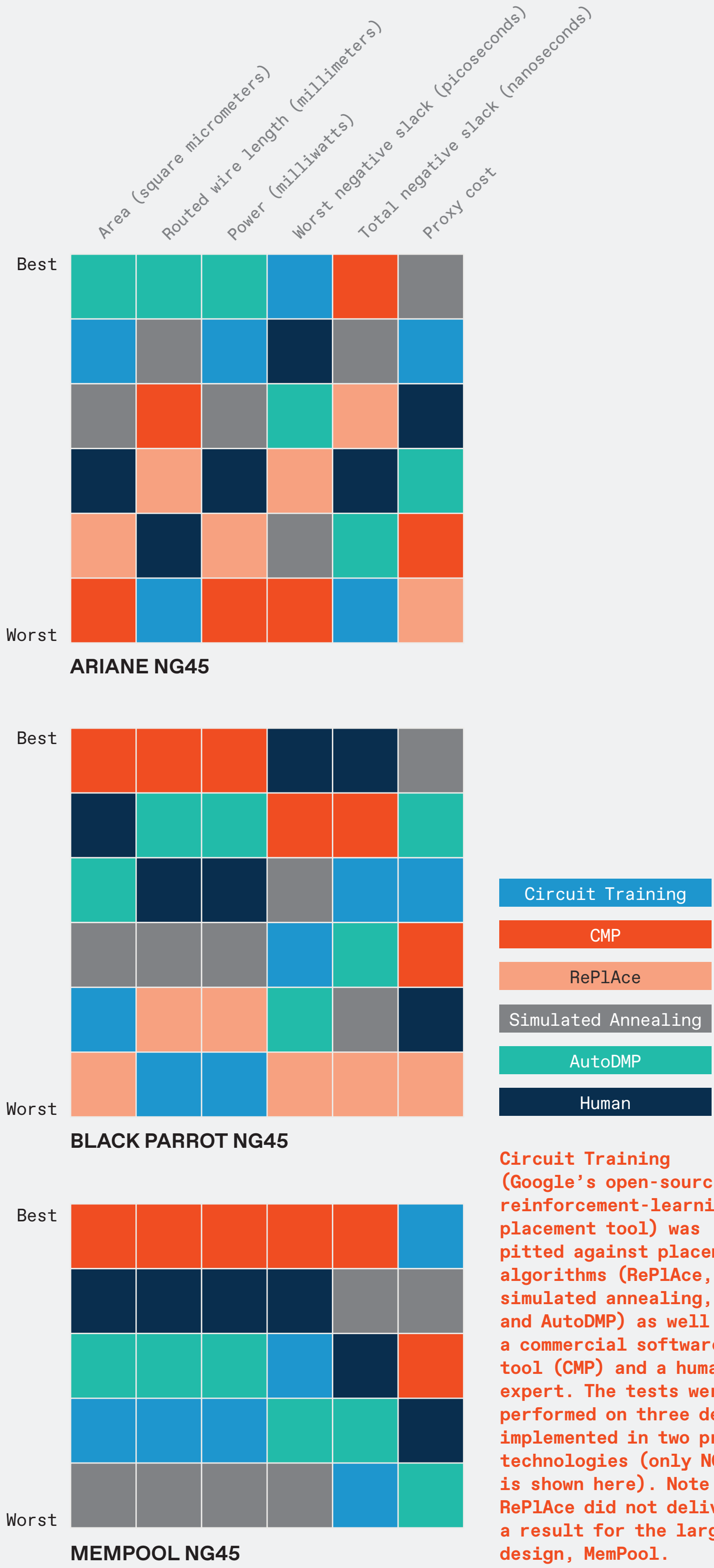
Instead of those benchmarks, the *Nature* paper focused mostly on doing the placement for Google's TPU, a complex and cutting-edge chip whose design is not available to researchers outside of the company.

Baselines are the state-of-the-art algorithms your new system competes against. *Nature* compared reinforcement learning to a human expert and to the leading academic algorithm of the time, RePlAce. "Stronger Baselines" contended that the *Nature* work didn't properly execute RePlAce and that another algorithm, simulated annealing, needed to be compared as well.

But it's the reproducibility bit that Kahng was really focused on. He claims that Circuit Training, as it was posted to GitHub, fell short of allowing an independent group to fully reproduce the procedure. So the UCSD team took it upon itself to reverse engineer what Kahng saw as missing elements and parameters.

Importantly, Kahng's group publicly documented the progress, code, data sets, and procedure as an example of how such work can enhance reproducibility.

The UCSD effort, called MacroPlacement, was not a one-to-one redo of the *Nature* paper. Besides using public benchmarks, two of which were unavail-



Circuit Training (Google's open-source reinforcement-learning placement tool) was pitted against placement algorithms (RePlAce, simulated annealing, and AutoDMP) as well as a commercial software tool (CMP) and a human expert. The tests were performed on three designs implemented in two process technologies (only NG45 is shown here). Note that RePlAce did not deliver a result for the largest design, MemPool.

able in 2020 and 2021, MacroPlacement compares Circuit Training to a commercial tool, Cadence's Innovus concurrent macro placer (CMP), and to a method developed at Nvidia called AutoDMP, which is so new it was publicly introduced at ISPD 2023 only minutes before Kahng spoke.

Another key difference is that each MacroPlacement result came from a neural network that had never seen a design before. But in the *Nature* experiments, the network had been pretrained on 20 samples of TPU circuitry before its results were compared with other methods. Mirhoseini and Goldie see the lack of pretraining as robbing their method of its main advantage. Circuit Training "unlike any of the other methods presented, can learn from experience, producing better placements more quickly with every problem it sees," they wrote in an email.

According to the MacroPlacement FAQ, pretraining was not tested because the code for using pretrained networks was not published in Circuit Training and because there were not designs available to match the training done in *Nature*. It also suggests that for Ariane, the smallest benchmark the team tested, Google had already shown that pretraining would not produce better results than were already published in *Nature*.

Circuit Training vs. Everybody

Kahng's paper reports results on the three benchmark designs, implemented using two process technologies. The team measured the same metrics that Mirhoseini and Goldie did in their *Nature* paper: area, routed wire length, power, two timing metrics, and the previously mentioned proxy cost. (Proxy cost is not an actual metric used in production, but it was included to mirror the *Nature* paper.) The results were mixed.

As it did in the original *Nature* paper, reinforcement learning beat RePlAce on most metrics for which there was a head-to-head comparison. (RePlAce did not produce a placement for the largest of the three designs.) Against a human expert, Circuit Training frequently lost. Versus simulated annealing, the contest was a bit more even.

The big winners were the commercial tool CMP and the new entrant AutoDMP, which delivered the best metrics in more cases than any other method.

Moving On

This dispute has certainly had consequences, most of them negative. Chatterjee is locked in a wrongful-termination lawsuit with Google. Kahng and his team have spent a great deal of time and effort reconstructing work done—perhaps several times—years ago. And after spending years fending off criticism from unpublished and unrefereed research, Goldie and Mirhoseini have left a field of engineering that has historically struggled to attract female talent. Since August 2022 they've been at Anthropic, an AI research company, working on reinforcement learning for large language models.

If there's a bright side, it's that Kahng's effort offers a model for open and reproducible research and adds to the store of openly available tools to push this part of chip design forward. And despite all the drama, the use of machine learning in chip design has only spread. It now assists in ever-growing aspects of commercial EDA tools from Synopsys and Cadence.

But all that good could have happened without the unpleasantness. ■

JOURNAL WATCH

Tech to Combat Social Media Hate Speech

BY MICHELLE HAMPSON

Many AI models have been developed to detect hate speech in social media posts, but it has been hard to make them both accurate and computationally efficient.

Tarique Anwar, a lecturer in the department of computer science at the University of York, in England, and his colleagues have overcome both difficulties with a model called BiCapsHate. They describe it in a paper posted in the 19 January *IEEE Transactions on Computational Social Systems*.

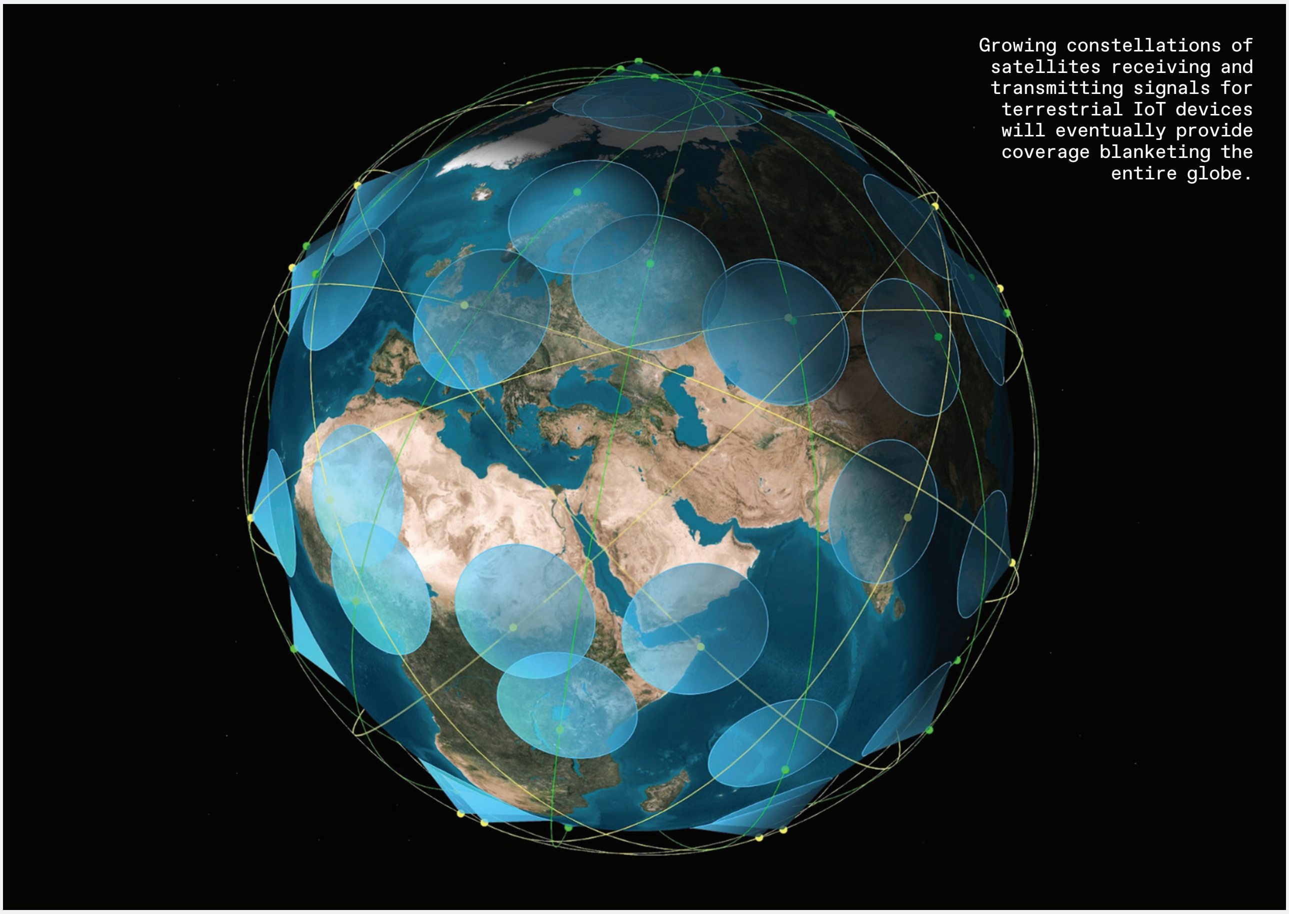
The model consists of several deep neural networks, each one dedicated to capturing different properties of speech. One network translates the language of a social media post into a numerical value and evaluates this sequence both forward and backward, the better to "understand" the context and thus minimize false positive results.

As Anwar points out, a word can be hateful in one context and anodyne in another.

The researchers found that BiCapsHate significantly outperformed existing AIs, including HateBERT, ToxicBERT, and fBERT. It achieved 94 percent and 92 percent F-score measures on balanced and imbalanced data sets, respectively. The higher the F-score, the greater the accuracy.

The new model is also more frugal. "[The other models] require high-end hardware resources like GPU, and high-end systems for computation," explains Anwar. "On the contrary, BiCapsHate… can be executed on a CPU machine with even 8 gigabytes of RAM."

The researchers hope to find ways to determine whether users who express hate online are mentally unstable, so that early interventions might be considered. ■



Growing constellations of satellites receiving and transmitting signals for terrestrial IoT devices will eventually provide coverage blanketing the entire globe.

AEROSPACE

CubeSat Operators Launch an IoT Space Race › New tech and lower costs make it possible to monitor devices straight from orbit

BY LUCAS LAURSEN

A **rocket carried two** CubeSats into orbit two years ago, and each one represented a competing approach to bringing the Internet of Things (IoT) to space. In late 2022, they were joined by another CubeSat representing a third approach.

That means there is now an IoT space race. And that means there's a price war, too.

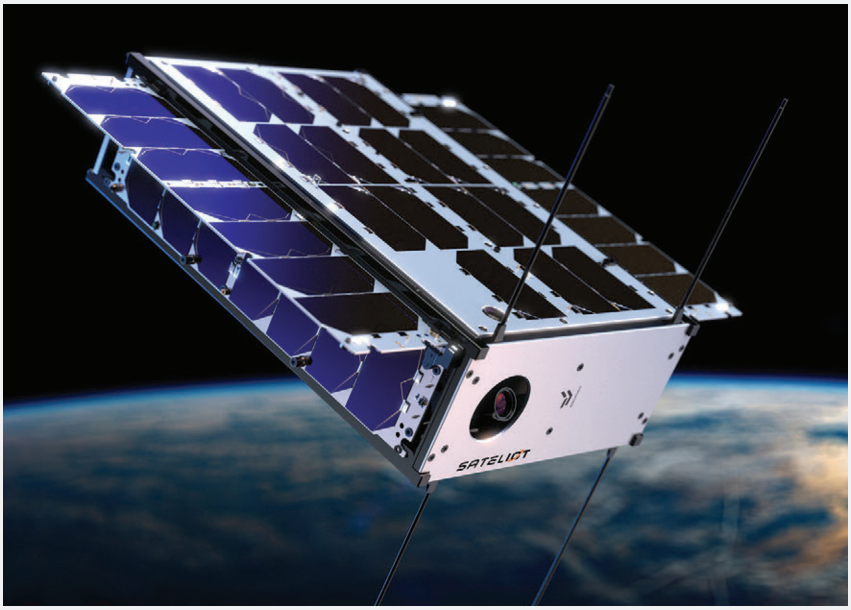
Of the first two CubeSats, one is operated by Lacuna Space, based in Oxfordshire, England, and uses a protocol called LoRaWAN, a long-range, low-power protocol owned by Semtech. The other, owned by Sateliot, in Barcelona, uses the narrowband IoT protocol, following in the footsteps of OQ Technology, in Luxembourg, which launched a similar IoT satellite demonstration in 2019. By the time of the more recent launch, the cellular industry standard setter 3GPP incorporated satellite-based 5G into standard cellular service with its Release 17.

OQ is also nipping at the heels of

satellite telecom incumbents such as Inmarsat, Iridium, and Orbcomm for a share of the growing satellite-IoT subscriber market. The company has three satellites in low-Earth orbit and plans to launch seven more this year, says its chief innovation officer, Prasanna Nagarajan. OQ has paying customers in the oil and gas, agriculture, and transport logistics industries.

OQ claims it will launch enough satellites to increase its satellites' check-ins to multiple times per day. Sateliot is also aiming for better-than-hourly coverage sometime in 2024 and near-real-time coverage in 2025. Sateliot has the satellite it launched in 2021 in orbit and plans to launch four more this year, says Sateliot's business-development manager, Paula Caudet. The company is inviting early adopters to sample its service for free this year while it builds more coverage. "Certain use cases are fine with flybys every few hours, such as agricultural sensors," Caudet says.

Incumbent satellite operators are already offering IoT coverage, but so far they require specific IoT hardware tuned to their spectrum bands and protocols. Insurgent companies that make use of the 3GPP Release 17 standard will be able to offer satellite connectivity to devices originally designed to connect only to cellular towers.



Sateliot is competing with OQ Technology to provide widespread satellite 5G connectivity for IoT devices before legacy operators enter the space.

SATELIOT

New companies also see an opportunity to offer more attractive pricing. "Legacy satellite providers were charging maybe [US] $100 for a few kilobits of data, and customers are not willing to pay so much for IoT," says Nagarajan. "There seemed to be a huge market gap." Another company, Swarm, a subsidiary of SpaceX, offers low-bandwidth connectivity via proprietary devices to its tiny satellites for $5 per month.

Thanks to shared launch infrastructure and cheaper IoT-compatible modules and satellites, new firms can compete with companies that have had satellites in orbit for decades. More and more hardware and services are available on an off-the-shelf basis. "An IoT-standard module is maybe €8 or €10, versus €300 for satellite-specific modules," says Caudet.

Indeed, Sateliot contracted the construction of its first satellite to Open Cosmos. Jordi Castellví, Open Cosmos's mission manager, says that CubeSat subsystems and certain specialized services are now available online from suppliers that include AlénSpace, CubeSatShop, EnduroSat, and Isispace.

By building constellations of hundreds of satellites in low-Earth orbit, IoT-satellite companies will be able to save money on hardware and still detect the faint signals from IoT gateways or even individual IoT sensors, such as those on shipping containers aboard ships at sea. They won't move as much data as voice and broadband offerings in the works from AST SpaceMobile and Lynk Global's larger and more complex satellites, for example, but they may be able to meet growing demand for narrowband applications.

OQ Technology has its own licensed spectrum and can operate as an independent network operator for IoT users with the latest 3GPP release—although at first most users might not have direct contact with such providers; both Sateliot and OQ have partnered with existing mobile-network operators to offer a sort of global IoT roaming package. For example, while a cargo ship is in port, a customer's onboard IoT device will transmit via the local cellular network. Farther out at sea, the device will switch to satellites overhead. "The next step is being able to integrate cellular and satellite services," Caudet says. That one small step for a satellite company would be one giant leap for the Internet of Things. ■

COMPUTING

MIT Makes Python Less Pokey › Codon runs code in an elegant, time-saving way

BY RINA DIANE CABALLAR

Python has a simplified syntax that makes it easy to learn and use. But its programs tend to run slower than those written in lower-level languages, such as C or C++.

Researchers from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) hope to change that with Codon, a Python-based compiler that allows users to write Python code that runs efficiently.

"Regular Python compiles to what's called bytecode, and then that bytecode gets executed in a virtual machine, which is a lot slower," says Ariya Shajii, an MIT CSAIL graduate student and lead author on a paper about Codon presented in February at the 32nd ACM SIGPLAN International Conference on Compiler Construction, in Montreal. "With Codon, we're doing native compilation, so you're running the end result directly on your CPU—there's no intermediate virtual machine or interpreter."

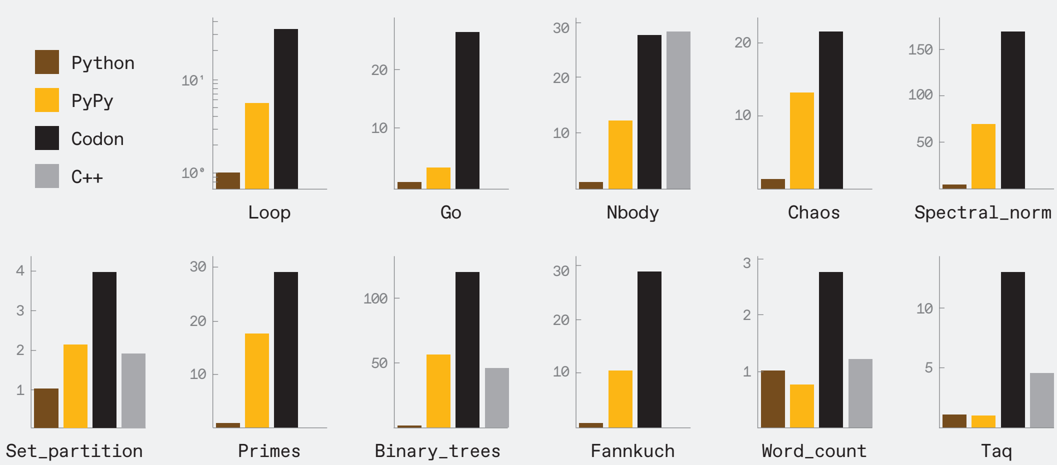
One speed trick is to use compile time to perform type checking, which involves assigning each value a data type—such as an integer, string, character, or float. For instance, the number 5 can be assigned as an integer, the letter *c* as a character, the word *hello* as a string, and the decimal number 3.14 as a float. Regular Python defers that job to runtime, slowing things down, says Shajii.

MIT professor and CSAIL principal investigator Saman Amarasinghe, who's also a coauthor on the Codon paper, adds that "if you have a dynamic language [like Python], every time you have some data, you need to keep a lot of additional metadata around it" to determine the type at runtime. Codon does away with this metadata, so "the code is faster and data is much smaller," he says.

"Versus Python, what we usually see is 10 to 100 times improvement," says Shajii.

"But there are trade-offs. "We do this static type checking, and we disallow some of the dynamic features of Python, like changing types at runtime dynamically," says Shajii. "There are also some Python libraries we haven't implemented yet."

Amarasinghe adds that "Python has been battle tested by numerous people, and Codon hasn't reached anything like



In their paper, the MIT CSAIL researchers presented these charts, which compare the execution speeds of Python (CPython 3), PyPy, Codon, and C++ (where applicable) across several benchmarks from Python's benchmark suite.

MIT/EXALOOP/UNIVERSITY OF VICTORIA/ACM

that yet. It needs to run a lot more programs, get a lot more feedback, and harden up more. It will take some time to get to [Python's] level of hardening."

Codon was initially designed for use in genomics and bioinformatics. "Data sets are getting really big in these fields, and high-level languages like Python and R are too slow to handle terabytes per set of sequencing data," says Shajii. "That was the gap we wanted to fill—to give domain experts who are not necessarily computer scientists or programmers by training a way to tackle large data without having to write C or C++ code."

Codon could also be applied to other applications that process massive data sets, as well as GPU programming and parallel programming, which the Python-based compiler supports. In fact, Codon is now being used commercially in the bioinformatics, deep learning, and quantitative finance sectors through the startup Exaloop, which Shajii founded.

To enable Codon to work with these different domains, the team developed a plug-in system. "It's like an extensible compiler," Shajii says. "You can write a plug-in for genomics or another domain, and those plug-ins can have new libraries and new compiler optimizations."

Moreover, organizations can use Codon for both prototyping and developing their apps. "A pattern we see is that people do their prototyping and testing with Python because it's easy to use, but when push comes to shove, they have to rewrite [their app] or get somebody else to rewrite it in C or C++ to test it on a larger data set," says Shajii. "With Codon, you can stay with Python and get the best of both worlds." ■

JOURNAL WATCH

Mini Robot Enters Blood Vessels, Completes Surgery

A CARDIOLOGIST WHO snakes a catheter through a patient's blood vessel to reach a blockage must estimate the position of the catheter from the outside using X-rays. Perhaps a miniature robot could do the job better from the inside.

One team in South Korea recently got a robot to navigate autonomously within an artery and back to its extraction point. They published their results on 9 February in *IEEE Robotics and Automation Letters*.

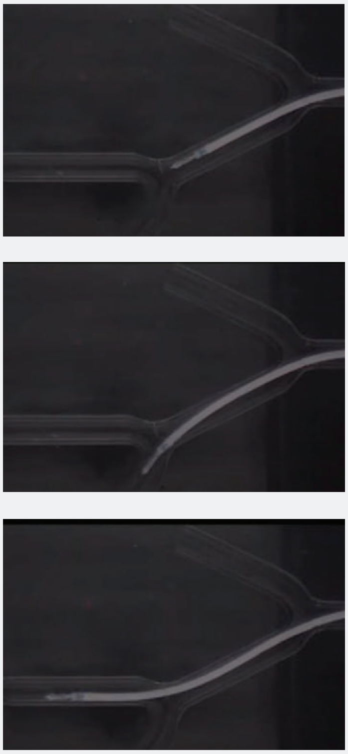
"[The standard procedure involves] long hours of operation because it is hard to accurately target a lesion if the blood vessel has a complex shape or is totally blocked," explains Gunhee Jang, a distinguished professor at Hanyang University, in Seoul, who was involved in the study.

Jang's team used an untethered robot guided externally by magnets and by a 3D map created from a variety of 2D X-ray images. They inject the robot through a catheter into a blood vessel and apply a magnetic field to twist the robot off the catheter. The robot then navigates to the treatment spot, guided by the magnetic field and the map.

Once there, it can inflate a balloon to open a blockage, suction away blood clots, or deliver contrast dye or drugs. Afterward, the external magnetic system guides the robot back to the catheter for removal from the body.

The researchers first tested the technique in an artificial blood vessel, then in the arteries of small pigs under anesthesia.

"We were very confident of performing robotic endovascular intervention in the [artificial] blood vessel," says Jang. "But during the in vivo experiment in a superficial femoral artery of the mini pig, we noticed that it is a quite different and difficult world."



These images show a tiny magnetic robot (attached to a delivery catheter) being steered through a network of blood vessels to reach the site of a blockage. The "steering" is controlled from the outside of the body by a magnetic field. Once the robot reaches its destination, it separates from the catheter, tunnels through the blockage, and then reattaches itself for removal.

Over the course of a year, they partnered with cardiologists and completed eight surgeries in pigs. The final experiment proved successful, offering a proof of concept that the technique is feasible in real life.

Jang's team has established a bioventure company named InterMag. Jang says the team also plans to apply to the Korean Ministry of Food and Drug Safety to undertake clinical trials of the magnetic robot system.

—Michelle Hampson

HANYANG UNIVERSITY

IEEE Spectrum congratulates our 2023 Jesse H. Neal Award winners!

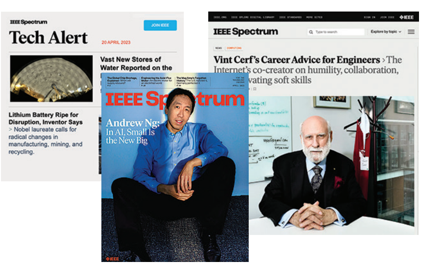
Front row, left to right: Eliza Strickland, Senior Editor; Samuel K. Moore, Senior Editor; Brandon Palacio, Deputy Art Director; Harry Goldstein, Acting Editor in Chief; Jean Kumagai, Executive Editor. Back row, left to right: Michael Koziol, Associate Editor; Stephen Cass, Senior Editor, Special Projects; Randi Klett, Photography Director; Steven Heffner, Managing Director, Publications.

JESSE H. NEAL AWARDS Est. 1955

The Neal Awards are given annually by the Software & Information Industry Association (SIIA).

SIIA

Best Range of Work by a Media Brand
IEEE Spectrum



Best Subject-Related Package
"The Transistor at 75" —Senior Editor Samuel K. Moore & the *Spectrum* staff

Best Overall Art Direction/Design
Creative Director Mark Montgomery, Deputy Art Director Brandon Palacio, Online Art Director Erik Vrielink & Photography Director Randi Klett

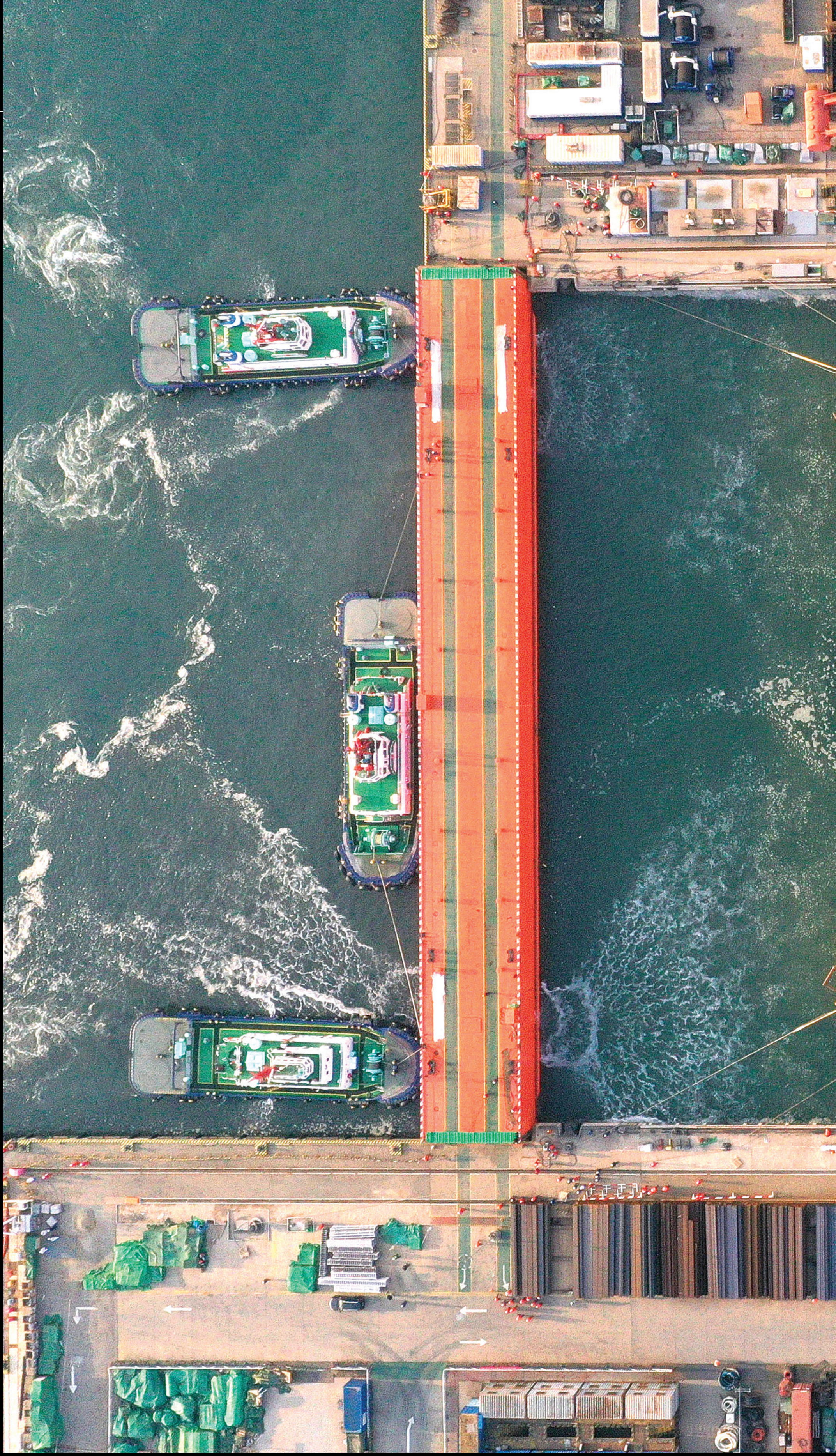
Best Single Issue of a Tabloid, Newspaper, or Magazine
December 2022 issue, "The Transistor at 75," *IEEE Spectrum* staff

Best Range of Work by a Single Author
Senior Editor Eliza Strickland

Best DEI Coverage
"Nothing About Us Without Us" —Senior Editor Eliza Strickland, Contributing Editor Mark Harris, Freelance Writer Britt H. Young, Acting Editor in Chief Harry Goldstein & Senior Editor Stephen Cass

Best Single Article
"Bionic Vision: An *IEEE Spectrum* Investigation" by Senior Editor Eliza Strickland & Contributing Editor Mark Harris

Best Art Direction for a Cover
"Do High-Tech Prosthetics Serve Their Users?" Photography Director Randi Klett, Creative Director Mark Montgomery & Photographer Gabriela Hasbun

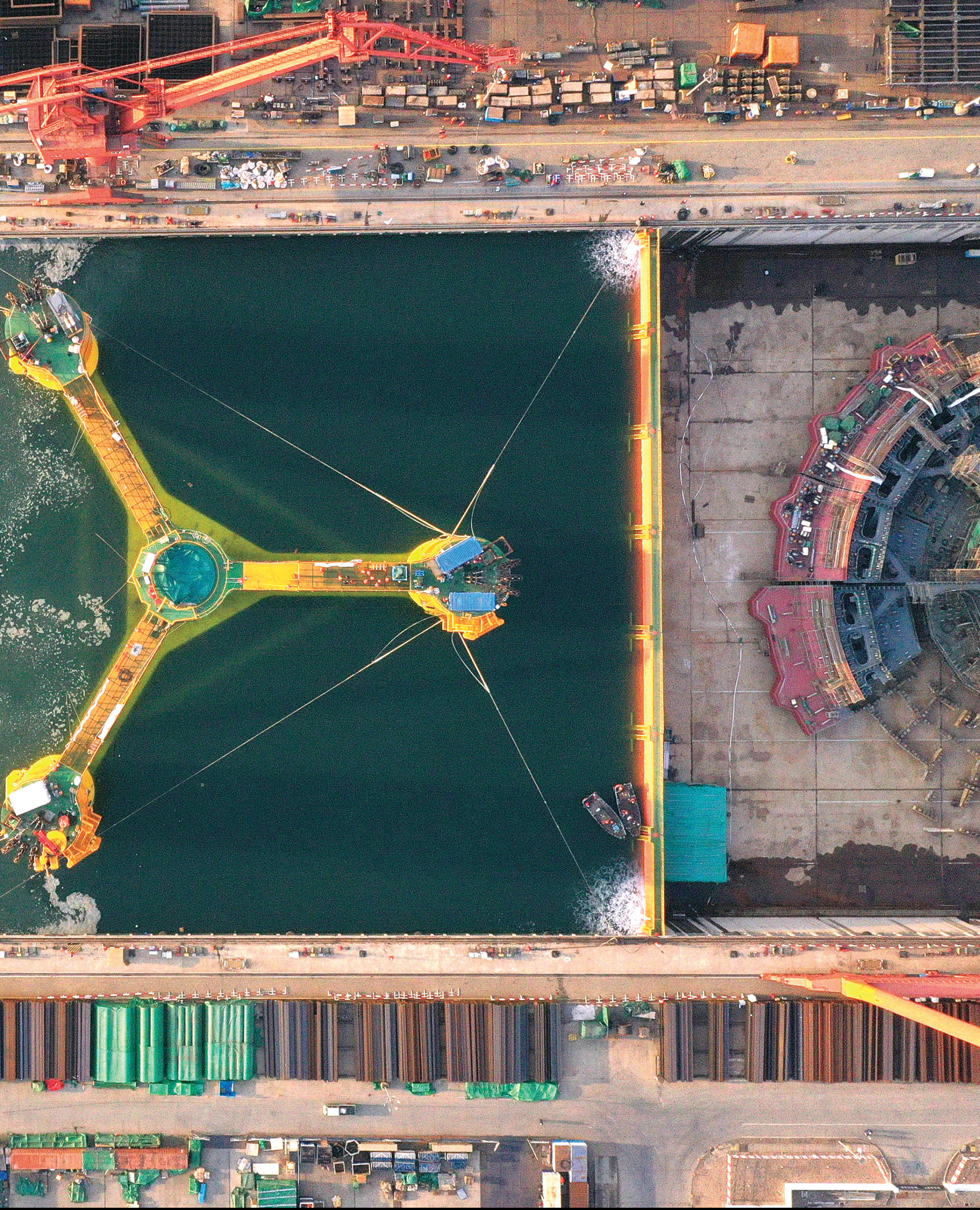


Where No Wind Turbine Has Gone Before

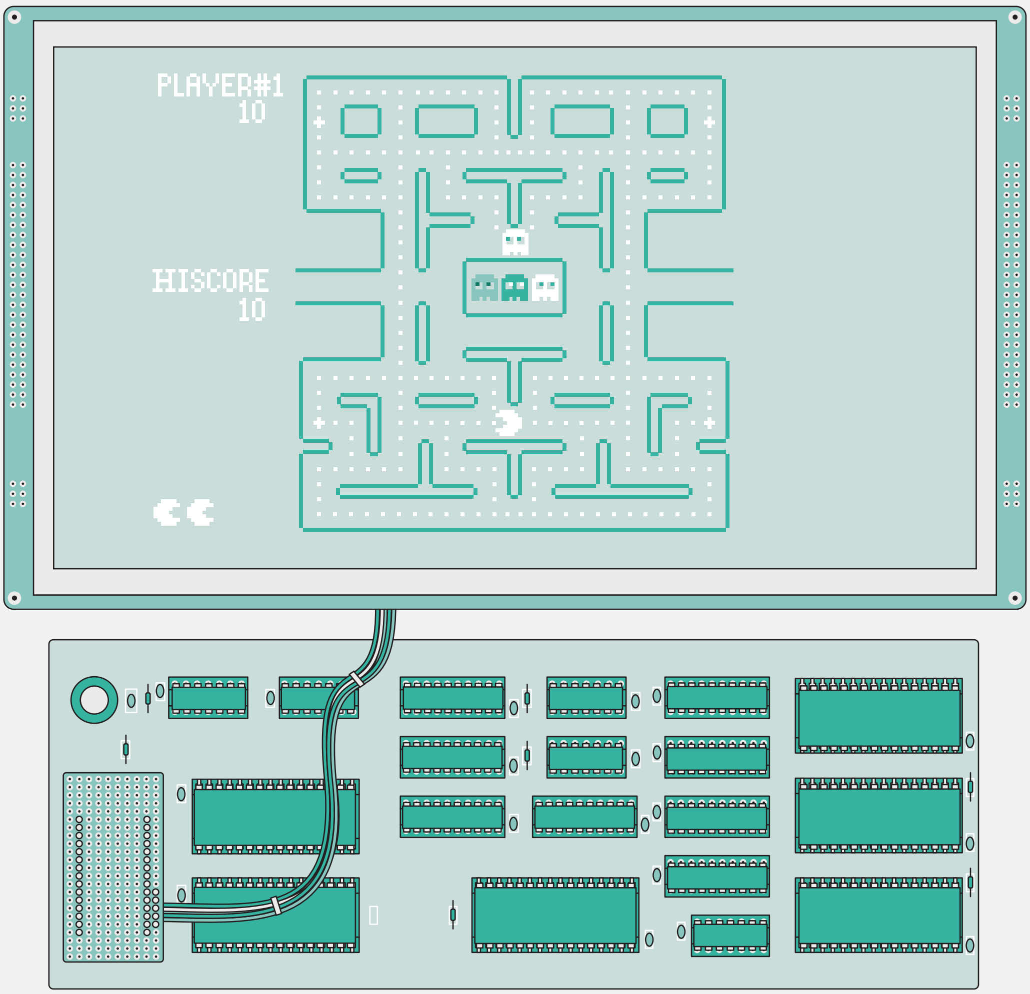
By Willie D. Jones

If wind power is to help make electricity generation using fossil fuels an anachronism, we've got to find places for a lot more wind turbines. Rather than put them in competition for space on land, engineers are designing turbines capable of harnessing the wind while moored on the bodies of water that cover more than two-thirds of Earth's surface. The 80-meter-long, 35-meter-high Haiyou Guanlan, shown here awaiting launch, is a floating wind-power platform that is the first power tower that will be situated more than 100 kilometers offshore. The mammoth renewable energy generator, which is designed to yield 22 million kilowatt-hours of electricity each year, will deliver that energy to the machinery at China's offshore oil fields.

PHOTOGRAPH BY FUTURE PUBLISHING/GETTY IMAGES



Hands On



The PureTuring machine has only a screen and no inputs, because even with accelerated memory access, it would require over seven years to play what would have been a 5-minute game on the Apple II.

When *Pac-Man* Met Turing ›

The game runs on an emulated 6502 CPU

BY MATTHEW REGAN

As I read the newest papers about DNA-based computing, I had to confront a rather unpleasant truth. Despite being a geneticist who also majored in computer science, I was struggling to bridge two concepts—the universal Turing machine, the very essence of computing, and the von Neumann architecture, the basis of most modern CPUs. I had written C++ code to emulate the machine described in Turing's 1936 paper, and could use it to decide, say, if a word was a palindrome. But I couldn't see how such a machine—with its one-dimensional tape memory and ability to look at only one symbol on that tape at a time—could behave like a billion-transistor processor with hardware features such as an arithmetic logic unit (ALU), program counter, and instruction register.

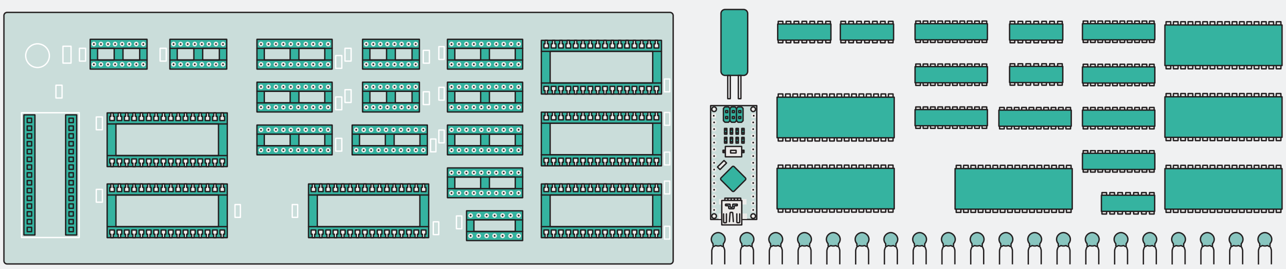
I scoured old textbooks and watched online lectures about

Illustrations by James Provost

theoretical computer science, but my knowledge didn't advance. I decided I would build a physical Turing machine that could execute code written for a real processor.

Rather than a billion-transistor behemoth, I thought I'd target the humble 8-bit 6502 microprocessor. This only move left or right on a one-dimensional notepad, reading or writing one symbol at a time.

A key revelation for me was that the internal registers of the 6502 could be duplicated sequentially on the one-dimensional notepad using four symbols—0, 1, _ (or space), and $. The memory using an instruction set. I could emulate the 6502's instructions as chains of rules that acted on the registers, symbol by symbol. The rules are stored in a programmable ROM, with the output of one rule dictating the next rule to be used, what should be written on the notepad (implemented as a RAM chip), and



Apart from some flip-flops, a couple of NOT gates, and an up-down counter, the PureTuring machine uses only RAM and ROM chips—there are no logic chips. An Arduino board [far left side], connected via an adapter cable, monitors the RAM to extract graphics data and drives a flat-panel display.

legendary chip from MOS Technology powered the computers I used in my youth. And as a final proof, my simulated processor would have to run *Pac-Man*, specifically the version of the game written for the Apple II computer.

In Turing's paper, his eponymous machine is an abstract concept with infinite memory. Infinite memory isn't possible in reality, but physical Turing machines can be built with enough memory for the task at hand. The hardware implementation of a Turing machine can be organized around a rule book and a notepad. Indeed, when we do basic arithmetic, we use a rule book in our head (such as knowing when to carry a 1). We manipulate numbers and other symbols using these rules, stepping through the process for, say, long division. There are key differences, though. We can move all over a two-dimensional notepad, doing a scratch calculation in the margin before returning to the main problem. With a Turing machine we can symbols 0 and 1 are used to store the actual binary data that would sit in a 6502's register. The $ symbol is used to delineate different registers, and the _ symbol acts as a marker, making it easy to return to a spot in memory we're working with. The main memory of the Apple II is emulated in a similar fashion.

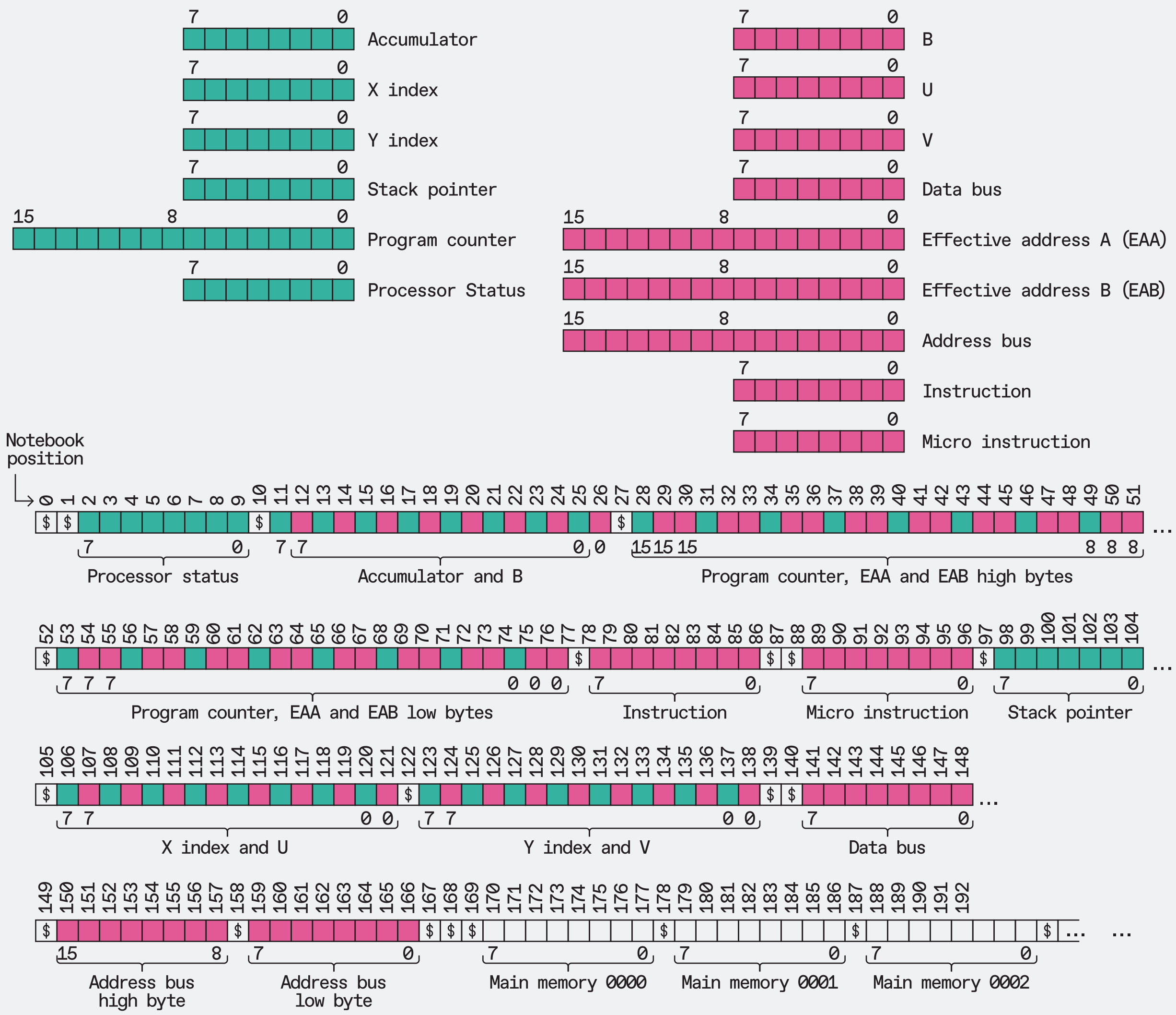
Programming a CPU is all about manipulating the registers and transferring their contents to and from main

A single instruction can take up to 3 million PureTuring clock cycles to fetch.

whether we should read the next symbol or the previous one.

I dubbed my machine PureTuring. The ROM's data outputs are connected to a set of flip-flops. Some of the flip-flops are connected to the RAM, to allow the next or previous symbol to be fetched. Others are connected to the ROM's own address lines in a feedback loop that selects the next rule. It turned out to be more efficient to interleave some registers rather than keeping them as separate 8-bit chunks. Implementing the 6502's instruction set required creating 9,000 rules. Of these, 2,500 were created using an old-school method of writing them on index cards, and the rest were generated by a script. Putting this together took about six months.

To fetch a software instruction, PureTuring steps through the notepad using $ symbols as landmarks until it gets to the memory location pointed to by the program counter. The 6502 opcodes are one byte long, so by the time the eighth bit is read, PureTuring is in one of 256 states.



Only some of the 6502's registers [in green] are exposed to programmers; its internal, hidden registers [pink] are used to execute instructions. Below the registers, a schematic shows how they are arranged, and sometimes interleaved, on the PureTuring's "tape."

Then PureTuring returns to the instruction register and writes the opcode there, before moving on to perform the instruction. A single instruction can take up to 3 million PureTuring clock cycles to fetch, versus one to six cycles for the actual 6502!

The 6502 uses a memory-mapped input/output system. This means that devices such as displays are represented as locations somewhere within main memory. By using an Arduino to monitor the part of the notepad that corresponds to the Apple II's graphics memory, I could extract pixels and show them on an attached terminal or screen. This required writing a "dewozzing" function for the Arduino as the Apple II's pixel data is laid out in a complex scheme. (Steve Wozniak created this scheme to enable the Apple II to fake an analog color TV signal with digital chips and keep the dynamic RAM refreshed.)

I could have inserted input from a keyboard into the notepad in a similar fashion, but I didn't bother because actually playing *Pac-Man* on the PureTuring would require extraordinary patience: It took about 7 hours just to draw one frame's worth of movement for the *Pac-Man* character and the pursuing enemy ghosts. A modification that moved the machine along the continuum toward a von Neumann architecture added circuitry to permit random access to a notepad symbol, making it unnecessary to step through all prior symbols. This adjustment cut the time to draw the game characters to a mere 20 seconds per frame! (You can watch the PureTuring in action on YouTube.)

For future games, features can be added one by one, moving piecemeal from a Turing machine to a von Neumann architecture: Widen the bus to read eight symbols at a time instead of one, replace the registers in the notepad with hardware registers, add an ALU, and so on.

Now when I read papers and articles on DNA-based computing, I can trace each element back to something in a Turing machine or forward to a conventional architecture, running my own little mental machine along a conceptual tape! ■

BY DANIEL P. DERN

Careers: Eileen Tanghal

Her journey from EE to VC

Through her venture-capital firm Black Opal, Eileen Tanghal funds startups that develop technology to improve accessibility to health care.

Eileen Tanghal, cofounder of the venture-capital firm Black Opal Ventures, didn't set out to be a venture capitalist. The electrical engineer left her job at a semiconductor startup to pursue an MBA so she could better understand the business side of technology. During a summer break, she was chosen for an internship with a VC company that funded high-tech startups. Her engineering background got her the position. Tanghal went on to work for several tech-focused VCs before launching Black Opal in 2021.

"My cofounder, Tara Bishop, and I began as angel investors, investing our own money," she says. "We invested primarily in early-stage startups, meaning they were still developing and validating their idea for a product and had raised little or no money yet."

In 2022, Black Opal, based in New York City, expanded to managing a fund that invests other people's and institutions' money, primarily funding startups in health care and life sciences.

"We want to improve health care accessibility and affordability," Tanghal says. "To do this, we look at companies working with emerging technologies and also those applying technologies that are well established in one industry but not yet leveraged in health care or life sciences."

The IEEE member is also working to hire more women and underrepresented minorities within the Black Opal portfolio, including the recommendation of specific individuals for leadership positions within the companies it funds. A diverse workforce, says Tanghal, produces the most inclusive and most profitable products and services.

"There's strong evidence that female-managed U.S. funds outperform all-male managed funds. Increasing the number of female leaders isn't just the right thing to do, but also the smart thing to do."

Tanghal earned a bachelor's degree in electrical engineering and computer science with a minor in economics from MIT, in 1997. The classes she took on semiconductors led to her first job, at PDF Solutions, a startup semiconductor consultancy firm, in Santa Clara, Calif. She specialized in optimizing the process of making transistors and improving manufacturing yields.

"My experience at PDF included using digital twins—which are digital representations of real-world products and processes—to improve development, testing, and support of our products," she says. "I also did advanced data analytics to help semiconductor manufacturers improve their fabrication yield."

After PDF, she spent nearly a year as product manager at AllAdvantage, a short-lived Internet advertising company. Tanghal then decided to pursue an MBA from the London Business School.

"I knew I needed to pick up more skills in order to advance my career in product management," she says. "It turned out I was one of a handful of students with a technology background and startup experience. Many classmates came from investment-banking or consulting companies."

A turning point came when she applied for a summer internship with Amadeus Capital Partners, a London-based firm that invests in European high-tech companies.

Hermann Hauser is one of Amadeus's cofounders. Earlier in his career, he had cofounded Acorn Computers, which developed one of the first reduced-instruction-set personal computers and had spun off the chip-design company Arm. During Tanghal's interview, Hauser talked about a plastic semiconductor company he had invested in. Drawing on her experiences at PDF, she asked him about

OGAGA BLESSING

the electron mobility—a measure of semiconductor behavior—in plastic versus the same behavior in silicon. She later learned her question made her stand out and got her the internship, one of just seven MBA students out of around 100 to get an offer from Amadeus. After receiving her MBA in 2003, the company hired her as a junior member of its investment team, and Hauser became one of her mentors. She went on to work for other VCs, where she invested in startups working on semiconductors, communications, imaging, and artificial intelligence.

Many of the technologies Tanghal is currently investing in are ones she learned about through her previous job as managing director at In-Q-Tel, a not-for-profit venture-capital firm in Menlo Park, Calif. The company funds information technology for U.S. and international defense and intelligence agencies.

Successful technology investing calls for knowledge of a market's needs as well as the technologies that are or aren't yet being fully embraced, Tanghal says.

"Based on my experience of being a member of the board of directors for 40 companies, many of today's technology vendors want to get into health and life sciences, but few have the knowledge of what these industries need," she says. "And companies already in these fields often don't know what technology is available, or what's already on the market."

For example, as an investor, she looks at pharmaceutical companies' initiatives and how they can be improved upon, such as through drug discovery,

Employer:
Black Opal Ventures, New York City

Title:
Cofounder

Education:
MIT, London Business School

<6

PERCENTAGE OF VENTURE-CAPITAL FIRMS BASED IN THE UNITED STATES THAT ARE LED BY WOMEN

the use of new messenger RNA for vaccines, and other therapeutics.

Tanghal believes it's the responsibility of Black Opal and the companies it works with to help increase the number of women and those from underrepresented groups in high-level investment positions.

"Less than 6 percent of VC firms based in the United States are led by women," she notes. "There's strong evidence that female-managed U.S. funds outperform all-male managed funds. Increasing the number of female leaders isn't just the right thing to do, but also the smart thing to do."

Through its Fellows Program, Black Opal hires MBA students to work part time and teaches them about investing.

Most of the recruiting for the program is done by Tanghal and cofounder Tara Bishop. The company has six employees and an advisory board of seven.

"In the 20 years I have been recruiting associates, this year is the first time that almost 70 percent of the applicants were women," Tanghal says. She suspects that many of the women candidates applied because "they saw two women heading the firm."

Another reason for promoting DEI, she says, is to make health care more equitable. "If women and minorities aren't properly represented at the board level, the company's products may not address their health issues and other concerns."

Historically, high-tech venture-capital firms tended to hire people who had an MBA. That's no longer the case, Tanghal says.

"Nowadays you don't have to have an MBA to become an investor," she says, "because now there's more opportunities to learn about investing. I prefer candidates who can dig deeper into the target market sector and understand the potential technologies being developed."

Another way investors can improve their understanding of a prospective company, she says, is to seek out opportunities to "sit next to the chief technical officer at a meeting and talk to them about what they are working on." She also recommends doing online research and taking advantage of IEEE member resources and its networking events. For Tanghal, combining tech savvy with business focus is the perfect fit.

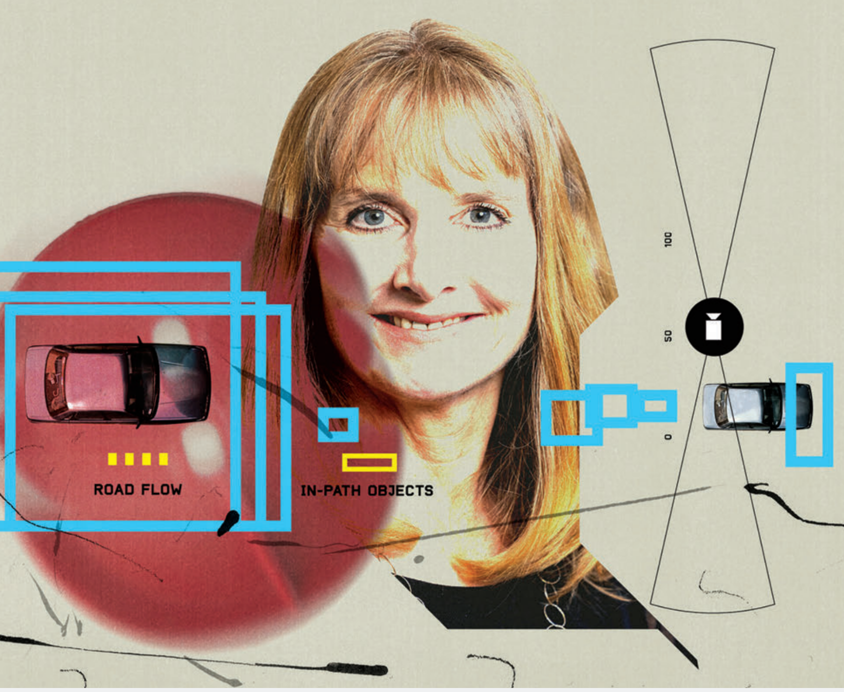
"My passion—and Black Opal's focus—is to find or make new matches between technologies and health-care applications, and to make money for our investors while improving health care for all." ■

Black Opal has invested in these startups:

Conceivable Life Sciences (New York City) is applying semiconductor manufacturing techniques such as advanced micrometer-level robotic automation to help reduce the high cost of in vitro fertilization at fertility clinics.

Empatica (Boston) is developing AI algorithms and wearable sensors to remotely monitor patients for medical conditions. The company's Embrace2 wearable detects possible epileptic seizures and immediately alerts caregivers. The wearable is so far the only such consumer device cleared by the U.S. Food and Drug Administration for adults and children ages 6 years old and up.

Optellum (Oxford, England) creates AIs for earlier diagnosis of lung cancer. Most cases currently aren't diagnosed until they are already in stage III or IV, which is one reason why lung cancer is the leading cause of cancer deaths in the world.



5 Questions for Missy Cummings

The former fighter pilot on why autonomous vehicles are so risky

In October 2021, Missy Cummings left her engineering professorship at Duke University to join the National Highway Transportation Safety Administration (NHTSA) in a temporary position as a senior safety advisor. It wasn't long before Elon Musk tweeted an attack: "Objectively, her track record is extremely biased against Tesla." He was referring to Cummings's criticism of his company's Autopilot, which is supposed to help the driver drive, though some customers have used it to make the car drive itself—sometimes with disastrous results.

Some of Musk's fans followed his lead: Cummings received a slew of online attacks, some of them threatening.

As a former Navy fighter pilot, Cummings is used to living dangerously. But she hates taking unnecessary risks, particularly on the road. At NHTSA, she scrutinized data on cars operating under varying levels of automation, and she pushed for safer standards around autonomy. Now out of the government and in a new academic perch at George Mason University, she answered five high-speed questions from *IEEE Spectrum*.

Mary (Missy) L. Cummings is the director of the Autonomy and Robotics Center at George Mason University and a senior member of IEEE. She received a Ph.D. in systems engineering from the University of Virginia.

We are told that today's cars, with their advanced driver-assistance systems (ADAS), are fundamentally safer than ever before. True?
Cummings: There is no evidence of mitigation. At NHTSA we couldn't answer the question that you're less likely to get in a crash—no data. But if you are in an accident, you're more likely to be injured, because people in ADAS-equipped cars are more likely to be speeding.

Could it be that people are trading the extra safety these systems might otherwise have provided for other things, like getting home 3 minutes sooner?
Cummings: I call it risk homeostasis. It's a big problem with Tesla, for example. You're told it has self-driving capability, with all these features, such as automatic braking. Oh, the car is going to do *x*, *y*, and *z* for me, and then it turns out that it doesn't.

Did you observe risk homeostasis back in your fighter-pilot days?
Cummings: It happened with air-to-ground bombing radar. Pilots figured out that you could use it to set up a self-contained approach to an aircraft carrier and then manage the landing by yourself. Given the control freaks that pilots are, it happened. But the system didn't adjust for the pitching deck, so it set people up for much more lethal approaches.

Some have said that partial autonomy is the riskiest solution of all. What's your take?
Cummings: The policy should be that either the computer is driving or you are driving. And by driving I mean steering—people do fine with regular cruise control. The act of keeping your hands on the wheel and guiding the car's lateral motion is enough to keep your brain engaged. So, no L3 [full self-driving, but the driver must be ready to take the wheel], which is too confusing, and no hands-free L2 [partial self-driving]. I am not against the passing of control per se, but there should just be two modes of operation, with crystal-clear feedback about which mode you are in.

When do you think true self-driving cars will come?
Cummings: It's possible to do self-driving in narrow applications. Waymo has been giving rides for a long time in Chandler [Ariz.]. That environment is very structured, and it's much easier to operate these systems in. My favorite application is last-mile delivery, say, food delivery; it could be very helpful when, say, viruses spike. But the day when AI in cars can handle all conditions on the road, all of the time—it's not going to be in my lifetime. ■

Photo-illustration by Stuart Bradford

Geologists Martin Tuchscherer [left] and Patrice Rioux stand in the middle of a boulder field in Nunavik, Que., Canada. Predictions from AI models can identify possible minable ore deposits in a vast area.

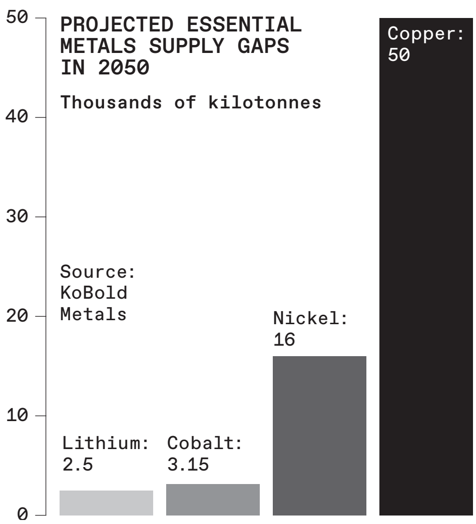
AI Hunts for Hidden Minerals

Machine learning is uncovering hoards of vital EV battery metals By JOSH GOLDMAN & KURT HOUSE

One way to find potential minable ore deposits is to use a transmitter coil loop [below], which—similar to a metal detector you'd use on a beach—detects induced currents in conductive bodies deep underground.

In June 2022, six Boeing 737s—fully loaded with tents, food, satellite Internet equipment, drones, geophysical survey gear, drilling equipment, and a team of experienced geologists—flew to a remote airstrip in northern Quebec. The geologists were hunting for major deposits of the minerals needed to power a clean-energy future. Given the mix of cutting-edge scientific computing and old-school bravado, it was as though they were channeling Alan Turing and Indiana Jones simultaneously. • Our startup, KoBold Metals, acquired an 800-square-kilometer mineral claim in this region of Canada based in part on predictions from our artificial intelligence systems. According to the AI, there was good reason to believe we'd find valuable deposits of nickel and cobalt buried below the surface. Summer snowmelts in this near-arctic area created a brief window to bring in a small village's worth of equipment and personnel to test our predictions.

In order to survey large areas quickly, KoBold uses a helicopter [right] carrying a transmitter coil loop 35 meters in diameter that can detect conductive bodies, such as ore deposits, below the surface. Conductivity data from KoBold's helicopter surveys can be used to produce models of potential underground distributions [far right]. The blue represents nonconductive igneous rock, while the yellow, orange, and red circles indicate areas of conductivity, from least to most conductive.



We cofounded KoBold in 2018 with backing from Bill Gates's Breakthrough Energy Ventures and Silicon Valley venture capital firm Andreessen Horowitz. Our goal is to develop ways to discover major new deposits of vital metals needed for electric vehicle (EV) batteries—for which there is an enormous and growing need.

We're trying to transform mineral exploration from a manual, judgment-guided, trial-and-error process into a data-driven and scalable science. It's the mother of all needle-in-a-haystack problems: Find the significant minable deposits of cobalt, copper, lithium, and nickel resting anywhere from 100 to 2,000 meters deep in the Earth's surface.

PREVENTING THE MOST catastrophic impacts of climate change requires achieving net-zero greenhouse gas emissions by 2050, which includes, among many other things, replacing all fossil-fuel-powered light cars and trucks with electric vehicles. That, in turn, will require manufacturing billions of EV batteries. Even today's demand for the metals outstrips supply—as evidenced by nickel prices doubling and lithium prices quintupling over the last year. To realize a global transition to electric vehicles, we'll need to discover and mine an additional US $15 trillion worth of cobalt, copper, lithium, and nickel by midcentury. (We're currently on target to mine about $3.6 trillion worth of these metals by 2050).

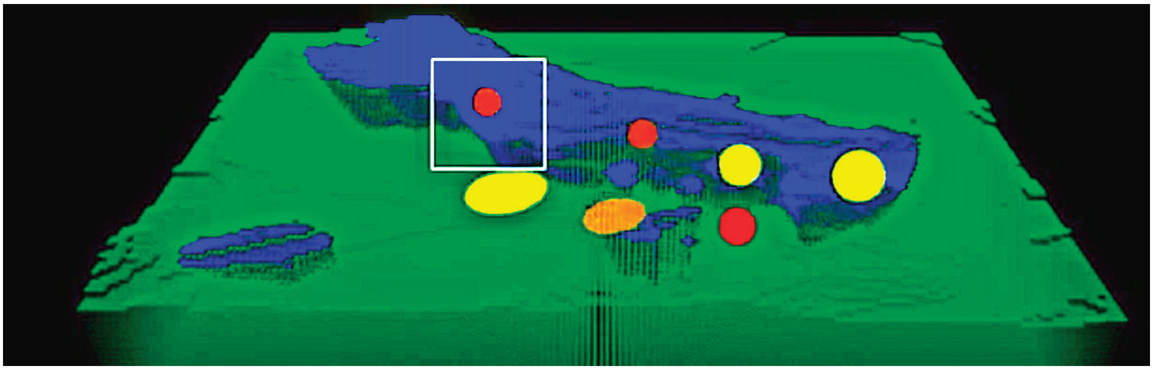
World leaders are well aware of the need. In the United States, for example, President Biden invoked the Cold War–era Defense Production Act in March 2022 to use the presidential powers it grants to encourage domestic production of the minerals required in EV batteries. The Inflation Reduction Act, signed into law in August 2022, included billions of dollars to subsidize the development and operation of metals mines, both in the United States and globally.

Investors are aware of the supply challenge as well. In February 2022, KoBold raised $192.5 million in Series B financing, which has gone toward securing more than 50 exploration sites in Australia, Canada, Greenland, sub-Saharan Africa, and the United States. We plan to use AI to streamline the largely scattershot process of discovering new ore deposits. Once they're discovered, we plan to partner with mining companies for the actual mining operations and advise them on efficient extraction, again using our AI tools.

FOR THOUSANDS OF YEARS, humans have noticed the striking appearances of rocks containing useful minerals. For example, the iron sulfides that are the predominant mineral in nickel sulfide deposits produce distinct reddish rust when exposed to air and rainwater. Weathering turns copper sulfide into a variety of brightly colored minerals, including the brilliant green ones found in the Statue of Liberty's patina. These visual clues were, for thousands of years, one of the most reliable ways to distinguish useful minerals and metals from useless rock.

The mining industry's rate of successful exploration—meaning the number of big deposit discoveries found per dollar invested—has been declining for decades. At KoBold, we sometimes talk about "Eroom's law of mining." As its reversed name

ALL PHOTOS: KOBOLD METALS



suggests, it's like the opposite of Moore's law. In accordance with Eroom's law of mining, the number of ore deposits discovered per dollar of capital invested has decreased by a factor of 8 over the last 30 years. (The original Eroom's law refers to a similar trend in the cost of new pharmaceutical discoveries.)

Geologically speaking, the decline in new discoveries is largely because most of the easy-to-spot deposits, such as those on the surface, have been found. New discoveries will be deeper underground, concealed by layers of rock.

In fact, the vast majority of Earth's ore deposits are still waiting to be found. The chemical and physical processes that form these ores occur at temperatures and pressures that exist kilometers below the surface. That is, these ore deposits are not formed at the surface; tectonic processes bring only a small minority of them there long after they were formed. That small minority constitutes the bulk of the deposits being mined today. The mining industry has the equipment and the technology to mine ore deposits that lie deep underground—the problem is finding those deposits in the first place.

You might expect the mining industry to be investing heavily in exploration, as well as in R&D to improve its exploration methods. But it's not. Over the past several decades, large companies have relied less on their own exploration programs and more on acquiring discoveries made by other companies. Mining-company shareholders expect dividends, not innovation.

AT KOBOLD, we're treating exploration as an information problem—finding and analyzing multiple types of data in order to uncover what we're looking for. In particular, it's an information problem in which acquiring more of those data types comes at a high cost. Our solution is to combine AI systems with geoscience expertise to figure out what piece of information reduces our uncertainty the most.

There is a vast body of geoscience information already in the public domain, but it's dispersed and fragmented. Some of it comes from government-funded geological surveys, and some comes from surveys conducted by private companies that were required to make their findings public. This information is spread across millions of data sets, including geological maps showing types of rocks observed in different locations; geochemical measurements of the concentrations of dozens of elements in samples of rock, soil, drill cores, plants, and groundwater; geophysical measurements of the gravitational field, magnetic field, natural and induced electric currents, seismic waves, and radiation from the decay of heavy-element nuclei in Earth's crust; satellite imagery—in both visual and infrared bands—measuring the spectral reflectance of minerals at the Earth's surface; and text reports describing field observations. The volume of data is, in a word, overwhelming.

What's more, these data sets range from state-of-the-art mass spectrometry measurements to hundred-year-old maps hand-painted on linen. Each data set is useful, and, combined in the right way, the full collection is potent—if you can make sense of it.

Our data system, called TerraShed, parses this information and brings it into a standard form to make it accessible and searchable by both humans and algorithms. Curating the data and putting it through quality control are just the first steps. We then use various algorithms to guide our decisions about what data to collect at each stage in the exploration process, from getting a sense of whether a particular deposit is worth mining all the way to construction of the mine itself.

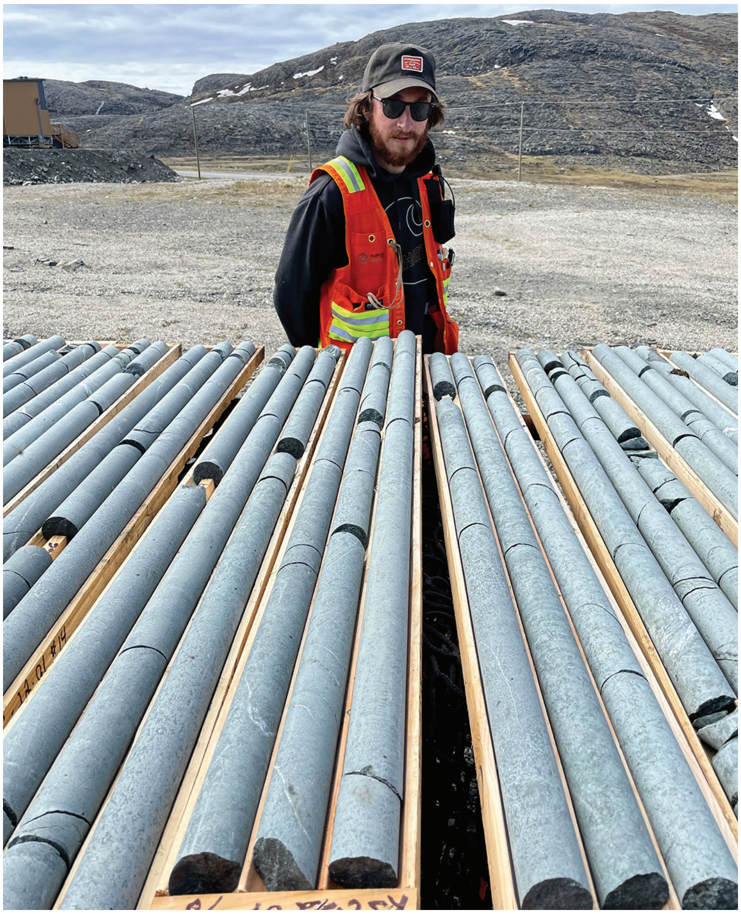
TerraShed doesn't produce simplistic treasure maps: It doesn't spit out an "X marks the spot" in response to the data. Instead, we have hundreds of different proprietary modules that guide each decision in the exploration process.

Our exploration program in northern Quebec provides a good case study. We began by using machine learning to predict where we were most likely to find nickel in concentrations significant enough to be worth mining. We train our models using any available data on a region's underlying physics and geology, and supplement the results with expert insights from our geologists. In Quebec, the models pointed us to land less than 20 km from currently operating mines.

After we acquired the relevant land rights, our geologists worked out of a field camp on-site, making observations and taking measurements of rock outcrops. Across the more than 800 km^2 of our claims, the choice of which rocks to sample is practically limitless. Time and money, however, are not—and in the region we were working, there's less than a three-month window when the ground is free of snow.

So, the information challenge becomes: How do we decide which rocks to sample?

We built Machine Prospector, which comprises the machine-learning models, with historic data, such as information from previous discoveries elsewhere in the province. It helped us predict which rocks we should sample, given the



Geologist Dave Freedman [left] stands behind a row of core samples at KoBold's Cape Smith site in Nunavik, Que., Canada, taken from areas of interest to confirm AI predictions. Brownish flecks in the core sample above, KSC 22-07, reveal the presence of magmatic pyrrhotite.

limited time we had. Specifically, we were looking for spots where eons-long geologic processes would have formed nickel- and cobalt-rich magmatic sulfide deposits.

Predictions in hand, our field geologists fanned out. Some headed to the places that seemed most likely to yield these magmatic sulfides. Others went to locations where the predictions were the most uncertain. Collecting data from places with uncertain predictions improves the next generation of models more than just collecting data where the models are already confident.

When the field team returned to camp in the evening, they uploaded that day's data via satellite. Our data scientists, working all over the globe, then retrained the models based on the new data points. The resulting new predictions changed the map of potential sample sites over the whole region and guided the team's decisions on where to go next. By incorporating new field information in almost real time, our model's adaptive predictions effectively shortened the learning cycle from a season to a day.

Our models generated predictions with 80 percent lower false positive and false negative rates compared to conventional predictions from geological maps. Such maps are constructed by making observations of the rocks at a relatively small number of locations and then using a set of rules and principles to extend those observations to larger regions. That means the conventional predictions are largely inference—and worse, they result in unquantified uncertainty. In other words, we don't know what we don't know about how accurate those maps are. By comparison, KoBold's predictive models do quantify uncertainty, which in turn guides our data collection, as the most uncertain rocks often represent the most valuable ones to sample.

THE RESULTS FROM one of our staked claims during that 2022 field season in northern Quebec are a perfect example of how our unique approach to exploration pays off.

Guided by the results from our AI systems, our field team found a large boulder field that geologist Lucie Mathieu identified as very anomalous, and not typical of the kind of igneous rock making up most of the region's boulders.

The boulder field originally piqued our interest after electromagnetic measurements we had taken indicated unusually high conductivity—consistent with the kinds of minerals we were seeking. The electromagnetic data was gathered by a helicopter towing a 30.5-meter-diameter transmitter coil loop for a daily time-domain electromagnetic survey. For these surveys, the transmitter pulses current through the loop at 7.5 hertz, which induces currents in conductive materials underground. When the transmitter pulse ends, the receiver coil detects the decay of those induced subsurface currents, enabling us to build a three-dimensional model of the subsurface rocks' conductivity. The high electrical conductivity of the ore minerals we're seeking is just one of several things that we can use to distinguish ore from other rock.

Using helicopter and geophysical survey equipment is expensive, and in the north the windows of good weather are short and unpredictable. Where we send the bird, and how we manage the trade-off between aerial coverage and spatial resolution, are vital considerations.

We can use the collected data to build three-dimensional models of the probable locations of ore in the subsurface, which is a computationally difficult problem. Put simply, we have a limited set of measurements of the induced fields taken in a two-dimensional plane just above the surface, and from that we are trying to infer the properties (here, the conductivity) of a three-dimensional volume of the subsurface. There are an infinite number of subsurface rock configurations that are consistent with the surface data.

The conventional approach in the industry is to build a best-estimate model that tries to fit a huge number of parameters, which can easily exceed the number of data points. Anyone who has tried to solve a system of n equations for $2n$ unknowns knows that there is no unique solution to the problem. The traditional methods used in the industry to choose one of the many potential solutions can often incorporate assumptions that are inconsistent with geologic processes and are prone to confirmation bias.

Audry Afango [far left] and Morgan McNeill work with a receiver on a ground electromagnetic loop in the middle of an angular boulder field in Nunavik, Que., Canada. On the right, McNeill, a geophysics technician, uses a superconducting quantum interference device (SQUID) to conduct a ground electromagnetic survey to pick up faint magnetic fields from conductive bodies underground.

To do better, we quantify the uncertainty in our predictions about the subsurface. Our machine-learning models are trained on many fewer parameters than traditional best-estimate models, and the parameters are directly related to the key exploration questions: How many conductive bodies are present? How deep are they? What is their orientation? Is their conductivity in the range that's consistent with high concentrations of ore minerals? The output of our models is the joint probability distribution of these parameters.

Ultimately, the most useful data to collect is that which reduces the uncertainty of finding an ore deposit that can be mined. Together with our collaborators at Stanford University's Mineral-X initiative, we have developed a novel way of quantifying how useful an incremental piece of data is. We published the framework, which we call "efficacy of information," in *Natural Resources Research* in March 2022, and we used it to design our drilling program for our northern Quebec exploration and for our other expeditions.

Over the course of the summer in Quebec, we drilled 10 exploration holes, each more than a kilometer away from the last. Each drilling location was determined by combining the results from our predictive models with the expert judgment of our geologists. In each instance, the collected data indicated we'd find conductive bodies in the right geologic setting—possible minable ore deposits, in other words—below the surface. Ultimately, we hit nickel-sulfide mineralization in 8 of the 10 drill holes, which equates to easily 10 times better than the industry average for similarly isolated drill holes.

We were also pleased with how accurate and specific the predictions were. For instance, at hole KSC-22-004, our data scientists predicted a conductive body to be located somewhere between 130 and 170 meters below the surface. Upon drilling, we encountered highly conductive rocks at 146 meters.

That particular discovery was made just days before the end of the field season. The data helped define the subsurface geology so that our team will start the next season—which begins soon—by making the most effective drill holes to establish the shape and size of that ore deposit.

Assuming that ore deposit and others we've begun to identify in the area turn out to be as promising as we hope, we'll be well on our way toward another mine for one of the crucial metals needed to electrify the planet. Collectively, the world needs at least 1,000 new mines to be developed by midcentury to provide enough critical metals to produce enough EVs and avoid the worst consequences of climate change. That's a tall order. But by applying new AI systems like KoBold's, we may just be able to dig up new opportunities fast enough. ■

CHIPS Act–funded new fabs are spawning university programs

BY PRACHI PATEL

Building a U.S. Semiconductor Workforce

THE CHIPS AND SCIENCE ACT, aimed at kick-starting chip manufacturing in the United States, only began taking requests for pieces of its US $50 billion in March, but chipmakers were already gearing up beforehand. Memory and storage chipmaker Micron announced as much as $100 billion for a new plant in upstate New York. Taiwan Semiconductor Manufacturing Co. (TSMC), which was already building a $12 billion fab in Arizona, upped the investment to $40 billion with a second plant. Samsung is planning a $17 billion fab near Austin, Texas, and in September Intel broke ground on the first of two massive new facilities worth $20 billion in central Ohio. ◎ Exciting as this is for the U.S. economy, there's a potential problem: Where will the industry find the qualified workforce needed to run these plants and design the chips they'll make? The United States today manufactures just 12 percent of the world's chips, down from 37 percent in 1990, according to a September 2020 report by the Semiconductor Industry Association. Over those decades, experts say, semiconductor and hardware education has stagnated. But for the CHIPS Act to succeed, each fab will need hundreds of skilled engineers and technicians of all stripes, with training ranging from two-year associate degrees to Ph.D.s. ◎ Engineering schools in the United States are now racing to produce that talent. Universities and community colleges are revamping their semiconductor-related curricula and forging strategic partnerships with one another and with industry to train the staff needed to run U.S. foundries. There were around 20,000 job openings in the semiconductor industry at the end of 2022, according to Peter Bermel, an electrical and computer engineering professor at Purdue University. "Even if there's limited growth in this field, you'd need a minimum of 50,000 more hires in the next five years. We need to ramp up our efforts really quickly." »

← Ohio State University is using its chip-fabrication facility to train future engineers and technicians. Here, from left to right, are OSU students Caleb Mallory and Jayne Griffith, manager of nanofabrication Aimee Price, and Columbus State Community College student Chris Staudt, who's also on staff at OSU's Nanotech West Laboratory.

Intel arrives at Ohio State

THE U.S. MIDWEST might be known more for farming and heavy industry than semiconductors, but chipmakers are betting it is fertile ground for their industry, thanks to an abundance of research universities and technical colleges.

Take Intel, which wants to create a "Silicon Heartland" in Ohio. In addition to building two cutting-edge chip factories on a 4-square-kilometer megasite that could hold six more fabs, the company has pledged $50 million to 80 higher-education institutions in the state. The funds should help the universities and community colleges upgrade their curricula, train and hire faculty, and provide equipment; Intel also plans to provide internships, guidance, and research opportunities.

Part of those funds have gone to Ohio State University, which will lead a new interdisciplinary Center for Advanced Semiconductor Fabrication Research and Education that will span 10 in-state colleges and universities. While most of the semiconductor-related curriculum has been designed for students in electrical and computer engineering, OSU now wants to bring in students from other disciplines. The university is creating tracks for them to master semiconductor-related skills, and it's revamping the curriculum in those disciplines to reflect the latest industry technology. Materials engineers will have new courses on chip-packaging materials, industrial system engineers will learn semiconductor manufacturing processes, and mechanical engineers will get to know device fabrication tools, says Ayanna Howard, dean of OSU's college of engineering. "Now that we're bringing manufacturing back to [U.S.] shores, our curriculum is now bringing in all these components that have always been needed but haven't been part of the plan at the scale required to train all these engineers."

There is no shortage of talent in the region, Howard adds, since manufacturing is already a major activity in Ohio and other parts of the Midwest. In 2011, Ohio kicked off an initiative called Jobs Ohio to create more science, technology, engineering, and math (STEM) graduates in the areas of computer science, biotech, and health-care manufacturing. It's now a matter of overhauling the curricula to cater to semiconductor manufacturing, she says. "When Intel came to the region, it really reinforced all the things that we had been thinking about."

In addition to leading two projects with state colleges, OSU is collaborating with 10 other midwestern institutions, including Purdue and the University of Michigan, to "think about engineering education more holistically," says Howard. "How do we create a curriculum that allows universities that might not have the infrastructure—say, lab space or trained

↗ Practical lab experience is key training for new engineers. At Ohio State University's Nanotech West Laboratory, students get hands-on experience with key semiconductor process technologies.

Previous pages and above: Photos by Peter Adams

“How do we create a curriculum that allows universities that might not have the infrastructure to give students semiconductor experience?”

—AYANNA HOWARD, OHIO STATE UNIVERSITY

faculty—to give students semiconductor experience?”

In the fall of 2021, for example, OSU piloted a course to teach students about chip-fabrication processes using desktop laboratory equipment, allowing them to learn without an expensive clean room. The engineering school is also teaming up with the creative arts department to create augmented-reality and virtual-reality tools that will let students experience a simulated fab. »

SkyWater moves next door to Purdue

↑ Graduate students Laura Chavez [top] and Marvin Zhang [above] use Purdue University's Brick Nanotechnology Center to characterize semiconductors and ICs. SkyWater, a Minnesota-based foundry, is building a fab near the university.

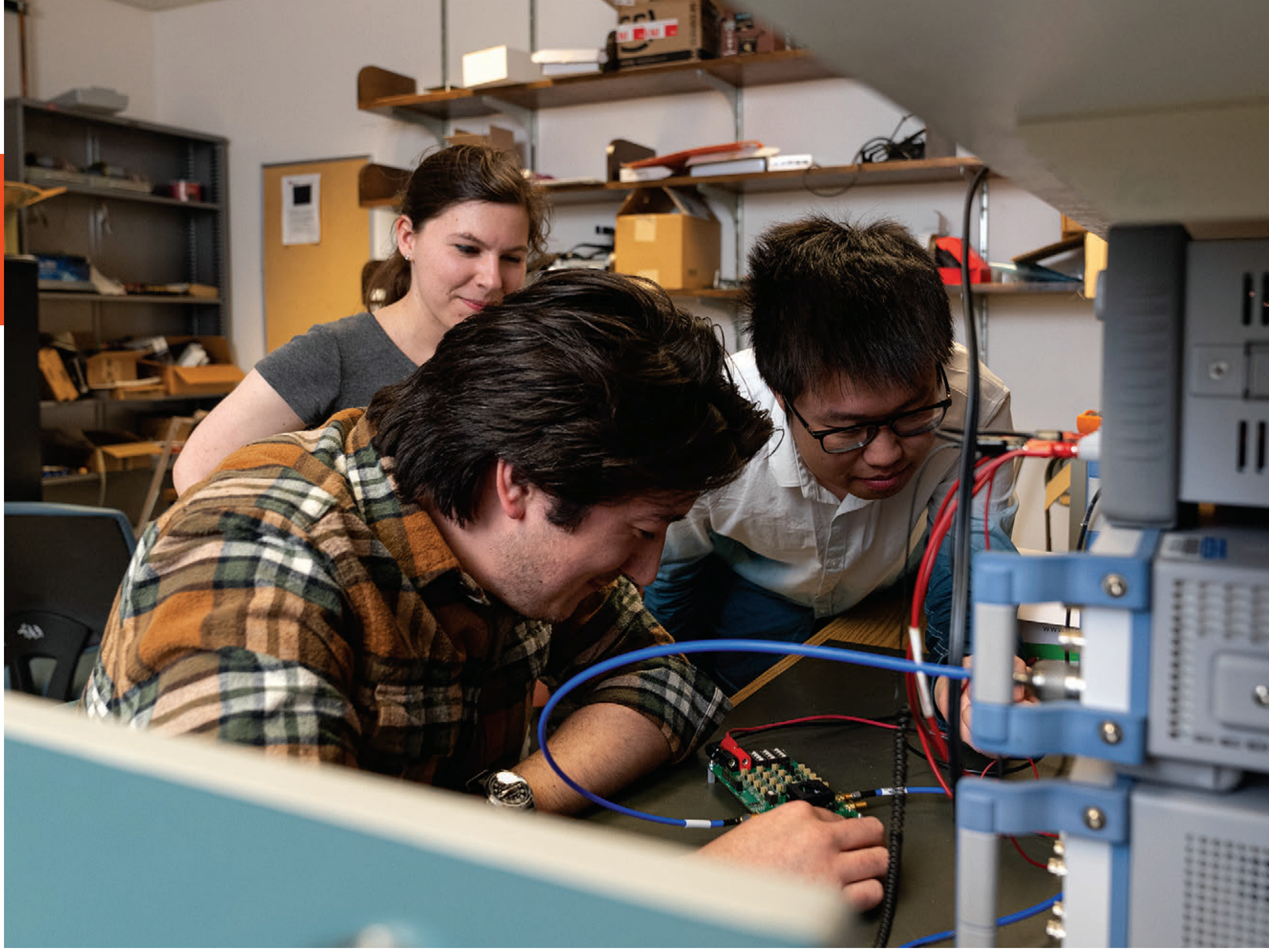
ABOUT 400 KILOMETERS (250 miles) west of Intel's development, another fab is planned. In July 2022, SkyWater Technology, a foundry that makes chips using specialty and mature manufacturing processes, announced a $1.8 billion chip fab at an industrial park in West Lafayette, Ind. Next door, Purdue has launched a new interdisciplinary Semiconductor Degrees Program to give undergraduate and graduate students a range of options for gaining core skills needed for the semiconductor industry. While EE courses traditionally cover integrated circuits and chip design, the new program teaches other key chip-manufacturing steps, including chemicals, materials, tools, manufacturing, packaging, and even supply-chain management. Students can choose to minor in the program, earn a master's degree, or get a certification.

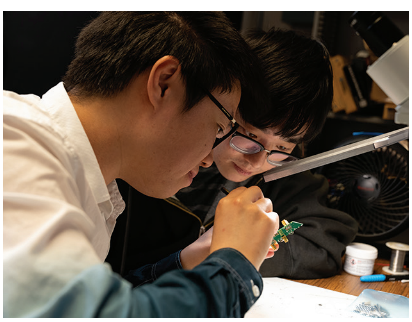
SkyWater representatives will inform students about various career options in the program's introductory seminar course. Students are guaranteed experience in Purdue's nanotechnology centers and at semiconductor companies. Advanced courses cover semiconductor materials and devices, as well as industrially relevant system-on-chip design. The program builds on Purdue's Scalable Asymmetric Lifecycle Engagement (SCALE) program, funded by the U.S. Department of Defense and launched in 2020, which trains undergrads to design and build semiconductors for space. "SCALE is specific to defense microelectronics, but the SCALE and semiconductor degrees programs are synergistic," says Purdue's Bermel. "Chips are fairly agnostic in many ways about the exact application space."

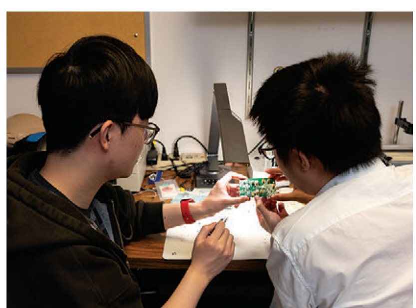
This year, Purdue will kick off a new program aimed at educating workers for SkyWater. Supported by state and regional economic-development organizations, the program will include operator and technician training through associate degree courses at partner Ivy Tech Community College. Besides developing targeted coursework and internships with the company, the Purdue team plans outreach at local high schools about job opportunities at their neighboring fab, with the hopes of attracting more students to engineering.

A summer internship at SkyWater's Florida foundry solidified Purdue undergraduate Anika Bhoopalam's interest in the semiconductor industry. Bhoopalam, a senior majoring in chemical engineering with a minor in electrical and computer engineering, feels that the physics, materials science, and engineering courses she has taken, combined with research lab experience fabricating thin films and solar cells, have prepared her well. She plans to pursue a Ph.D. with a focus on materials science and solid-state physics so she can go on to work in chip manufacturing. "I found the semiconductor industry to be a fast-paced, exciting, and interesting field where you get to work on different tasks every day." »

TOP: CHARLES JISCHKE/PURDUE UNIVERSITY; BOTTOM: REBECCA ROBIÑOS/PURDUE UNIVERSITY







VIRGIL WARD/UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

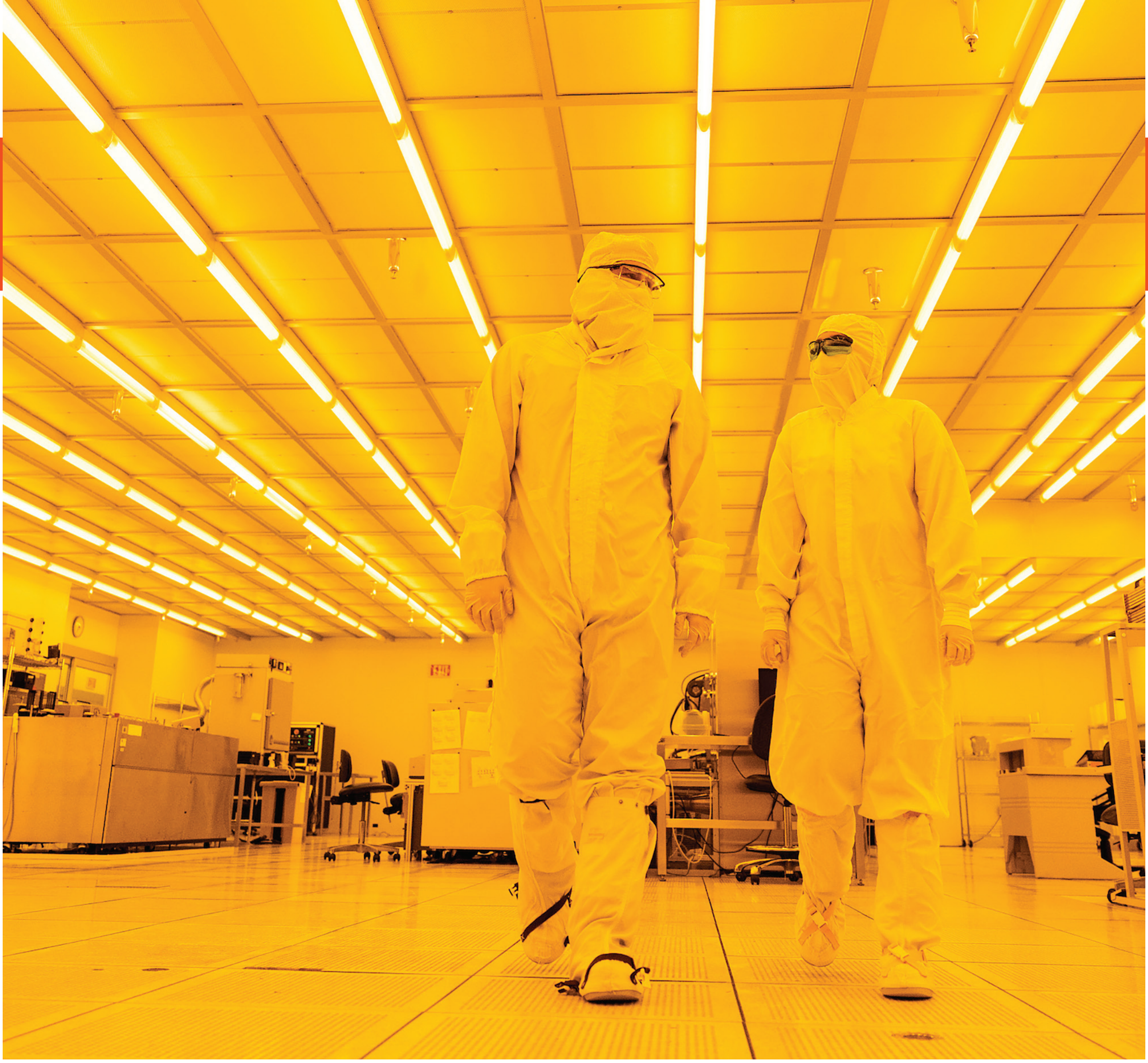
Illinois ups its chip education game

↑ Undergraduate engineering students at the University of Illinois Urbana-Champaign take part in a course on designing and constructing ICs, which was once considered graduate-level work. Here, Jenna Cario [gray shirt], Andy Ng [plaid], Stanley Wu [white and blue] and Curtis Yu [in black] collaborate on various projects.

STUDENTS TRADITIONALLY develop, build, and test integrated circuits in graduate school. But universities are trying to provide that first hands-on experience earlier. In the fall of 2021, electrical and computer engineering professors Pavan Hanumolu and Rakesh Kumar at the University of Illinois Urbana-Champaign created a class called Advanced Systems Design, which leads senior-year undergrads through every step of making an integrated circuit.

"It's a nitty-gritty job," says Hanumolu. "Companies don't want students learning this on the job. If we can provide these skills, that might shorten the route to increasing the talent pool for industry."

Students work in teams, defining a problem and designing a functional CMOS integrated circuit to solve it. The designs are sent off to a TSMC fab for manufacture, after which students test the chips, redesign the circuit as needed, and create a printed circuit board for the chip. So far, 30 students have taken the class, Kumar says, and several have gone on to internships and jobs with companies like Apple, Intel, and Siemens. "Employers have appreciated the rigor our students go through," Hanumolu says. "They know the kind of unique skill set these students will graduate with." »



↗ Staff member Alex Cabrera [left] and student Monica Gaytan walk through a lab at the MacroTechnology Works research facility, where Arizona State University students work with fab equipment.

Arizona State adds TSMC

WHILE CUTTING-EDGE chip fabs might be new in the Midwest, in the Southwest, Arizona State University has had a head start in preparing for an expanding U.S. chip industry. Motorola picked the deserts near Phoenix for its plant back in the 1960s. Aerospace and defense companies followed, and then Intel arrived in the 1980s and is currently expanding. Most recently, TSMC broke ground on its big U.S. fab complex in Phoenix in 2021. Connections to the companies form a solid foundation for semiconductor education and research at ASU, says Kyle Squires, dean of the engineering school.

"You can see in our DNA the origins from this semiconductor presence going back to the '60s," he says. A significant fraction of the EE faculty comes in with industry experience and has close ties with industry. They bring their expertise into the hardware engineering classes and labs they design, Squires says, while giving students access to scholarships, research opportunities, internships, and eventually jobs. "It's a way for us to maintain currency with these companies, what they're doing, where they're headed. As technology needs continue to move, so does the curriculum. It's research informing teaching, and vice versa. It's a feedback loop."

ASU boasts a large microelectronics facility—originally a Motorola semiconductor fab—and EE students in their junior and senior years can choose electives that give them direct semiconductor processing experience in the facility's clean rooms. Graduate students, meanwhile, can pursue a 15-credit Certificate in Semiconductor Processing that trains them in various aspects of chip production. Squires acknowledges that putting multimillion-dollar tools into an undergrad lab is unrealistic for most universities, so forging relationships with industry partners can help make up the difference. »

DEANNA DENT/ARIZONA STATE UNIVERSITY

Community colleges could be key

MORE THAN BEING a partner, Intel sees itself as a catalyst for upgrading the higher-education system to produce the workforce it needs, says the company's director of university research collaboration, Gabriela Cruz Thompson. One of the few semiconductor companies still producing most of its wafers in the United States, Intel is expanding its fabs in Arizona, New Mexico, and Oregon. Of the 7,000 jobs created as a result, about 70 percent will be for people with two-year degrees, and 25 percent for those with bachelor's degrees, master's degrees, and Ph.D.s.

Since COVID, however, Intel has struggled to find enough operators and technicians with two-year degrees to keep the foundries running. This makes community colleges a crucial piece of the microelectronics workforce puzzle, Thompson says. In Ohio, the company is giving most of its educational funds to technical and community colleges so they can add semiconductor-specific training to existing advanced manufacturing programs. Intel is also asking universities to provide hands-on clean-room experience to community college students.

Samsung and Silicon Labs in Austin are similarly investing in neighboring community colleges and technical schools via scholarships, summer internships, and mentorship programs. Samsung supports an initiative at Austin Community College that provides technician training for high school students. The company's Fab Apprentice Program, meanwhile, allows students to complete their associate degree while working at Samsung two days a week. "We pay 100 percent of tuition and books as long as the student maintains a 3.0 GPA or higher," says Michele Glaze, head of communications and community affairs at Samsung Austin Semiconductor. »

↓ Community colleges will be key to filling the workforce needs of new fabs. Here, Maricopa Community College students work in an Intel-sponsored lab in Tempe, Ariz.

↗ Chip companies are struggling to find workers in South Korea, too. Here, Samsung employees work on a project at their job-related training program.

TOP: SAMSUNG; BOTTOM: MARICOPA COMMUNITY COLLEGE/INTEL

Workforce shortages everywhere

THE SEMICONDUCTOR TALENT shortage isn't unique to U.S. shores. Taiwan makes about 65 percent of the world's chips, but finding young semiconductor engineers has been getting difficult, according to reports. Semiconductor firms around the world are competing for talent: They're hiking salaries and doling out scholarships, internships, and mentorships to undergraduates and even vocational high school students, in hopes of attracting them early. "As the need for advanced semiconductors continues to increase and chip manufacturers compete for talent, we see the supply in the workforce trailing the high demand," Samsung's Glaze says.

Samsung works with four major universities in South Korea, providing tailored curricula to train students in semiconductor R&D and manufacturing. Taiwan's government, meanwhile, is partnering with chip companies to invest $300 million in specialized chip-focused graduate schools within top universities to train the next generation of semiconductor engineers.

Enticing more students to study engineering is a big problem, Intel's Thompson says. Attractive jobs in the software industry have led to a shift in the balance between electrical engineering and computer science. "We hear from academics that we're losing EE students to software," she says. "But we also need the software. I think it's a totality of 'We need more students in STEM careers.'"

The CHIPS Act might just be what was needed to put semiconductors in the limelight and entice students to hardware-related degrees. At Purdue, Bermel says he has seen an uptick in interest in the semiconductor information session at the annual September career fair. Historically, the fair has had a handful of semiconductor employers, but it had 28 this year and attracted over 600 students. For the semiconductor industry to be more successful going forward, the software industry's practices might be worth following, he says. That includes "providing better opportunities for students even after only the first year of undergrad if possible, paying them very well, but also making it more evident to the general public why semiconductor companies are important." ■


ASXGS

It was a wonder that the original video game worked at all. Decades later, it still enthralls millions

THE MIRACLE OF *FLIGHT SIMULATOR*

IN 1999 BILL GATES penned a moving tribute to the Wright brothers. He credited their winged invention as "the World Wide Web of that era," one that shifted the world into a global perspective. So it's only fitting that Microsoft later became the force behind *Flight Simulator*. ◎ And, like the Wrights' original Flyer, the game's legacy has extended beyond flight to embody the shift of perspective that flight allows. *Flight Simulator* promised to fit the whole world into your computer, and the game kept its promise. That's why it has become the world's best-selling flight-simulation franchise: The latest edition has sold more than 2 million copies. ◎ Although 2022 marked the 40th anniversary of Microsoft *Flight Simulator*, its lineage stretches a few years further back than its official release, in 1982. That makes it the second-oldest video-game franchise still in active development. (*The Oregon Trail* came out in 1971 and is still with us.) ◎ The heart of the franchise isn't in gamification but in the technical spectacle it uses to simulate flight and the ground beneath you. The focus on true-to-life depiction reflects the background of the game's developers. ➔

BY MATTHEW S. SMITH

These screenshots represent three succeeding generations of *Microsoft Flight Simulator*, beginning with subLogic's first simulator for the Apple II [top row, left], followed by iterations that ran on Atari [striped balloon], MS-DOS, and then Microsoft Windows.

PREVIOUS AND THESE PAGES: JOSEF HAVLIK AND MICROSOFT

BRUCE ARTWICK STUDIED electrical engineering at the University of Illinois at Urbana-Champaign, yet he found time to pursue a dream many teenagers think about but few fulfill: He learned to fly. It was at the university's flight-instruction program that he met Stu Moment, who would later become his business partner.

In 1976, Artwick wrote his master's thesis on a flight simulator he'd designed to run on Digital Equipment Corp.'s PDP-11 minicomputer. Its 3D presentation could reach nine frames per second, which was outstanding for the era.

Artwick then took a job with Hughes Aviation, in California, while continuing to work on 3D graphics in his free time. In 1977, he wrote an article for *Kilobaud: The Small Computer Magazine* describing the "Sublogic Three-Dimensional Micrographics Package" he had created, which brought 3D to microcomputers outfitted with the popular Motorola 6800 microprocessor. Because so many readers were keenly interested, Artwick, looking for help turning the software into a business, reconnected with Stu Moment, and together they founded subLogic.

Dave Denhart, subLogic's second hire, recalls that the company's early days were driven by Artwick's 3D software, updated for emerging microcomputers. "The stuff that [Stu] and Bruce were selling was basically a 3D software package [for microcomputers]," says Denhart. "The [Tandy] TRS-80 was one of them, and I think the Apple II was out by then."

Artwick often claimed in presentations that subLogic's software could be used for flight simulation—a suggestion that brought gasps from the audience. A computer displaying the perspective of a pilot soaring above the planet? Most people had never seen anything like it.

Encouraged, Artwick decided to make it a reality. *Flight Simulator* launched in late 1979 on the Apple II and the TRS-80 with wireframe graphics and a frame rate in the single digits. It didn't depict real airspace or modern airplanes. Instead, players flew a World War I–era biplane based on the famous Sopwith Camel. Still, its first-person 3D visuals were ahead of the curve, predating more famous hits like Atari's *Battlezone*.

"I know Bruce always saw, from his early days, a potential market for [*Flight Simulator*]," says Denhart. "It was when microprocessors became available that I think the lightbulb went off for Bruce that said, Hey, if I put this idea I've got for a flight simulator onto cheaper computers, and get that to work, there's a market for that."

Artwick was right. In September of 1982, *Computer Gaming World* magazine ranked *Flight Simulator* as the fourth best-selling title to date. IBM, craving a showcase for its IBM PC platform, contacted subLogic about bringing *Flight Simulator* to the new hardware. Microsoft, deep in development of IBM PC DOS, soon called with a similar request—and better terms. It got a version with its own name on the label, though Artwick continued to own his company for years to come.

Microsoft Flight Simulator, released in late 1982, continued to improve in the months that followed, mirroring the advancements in microcomputers. The graphics moved from monochrome to color (on PCs with the right hardware), and the display refresh rate increased to 15 frames per second, which one reviewer described as "very smooth." Players piloted a Cessna 182 in four real-world areas, including Chicago and Seattle. For the first time, a home-computer enthusiast could fly a real-looking model of an airplane across true-to-life terrain, taking off and landing at facsimiles of real airports.

The realism extended to the flight

model, which made use of an effective technique: lookup tables. That's because real-time calculations of forces on an aircraft were beyond the capabilities of early IBM PCs. Fortunately, the aircraft manufacturers had already calculated how their products would perform. This gave subLogic a cheat sheet to build on.

"You basically say, Here's my input, what's my output?" says Denhart. "[The simulator] can just simply do a lookup. Now, because of resolution limitations and memory concerns, you wouldn't have a superlarge lookup table because the processors and memory couldn't handle that. But you'd have a sort of medium size, even small size, and between the data points you can do interpolation."

With the flight model in place, subLogic expanded the simulation's scope. *Microsoft Flight Simulator 2.0* (1984) modeled the entire United States, *Microsoft Flight Simulator 3.0* (1988) brought the Gates Learjet 25 and substantially more airports, and *Microsoft Flight Simulator 4.0* (1989) added random events, including weather.

In 1993, *Microsoft Flight Simulator 5.0* brought a new killer feature: textures. Previously, land between airports had been represented by patches of color—green for forests, gray for urban, blue for water. Textures offered new details. Beginning with *Flight Simulator 5.1,* those details were based on satellite imagery. For many, *Flight Simulator* would be the first opportunity to see satellite imagery in real-time 3D software.

Flight Simulator's first decade on the market was a time of smooth ascent. Yet there was turbulence behind the scenes, reflecting bad air that had come even earlier. Artwick and Moment often disagreed, and by around 1980 they worked separately, Moment by day, Artwick by night. The split eventually became permanent, with Artwick leaving to form the Bruce Artwick Organization (BAO) in the late 1980s. 1988's *Flight Simulator 3.0* was the last version credited to subLogic.

Perhaps this schism contributed to Artwick's decision to sell to Microsoft in 1995. Artwick could not be reached for this article, but sources employed before and after the sale remembered it as an abrupt yet unsurprising decision. Artwick had spent nearly two decades dedicated to the business. The team was also having differences with Microsoft.

"I think ultimately Microsoft wanted full control of the product versus Bruce holding on to it," says Paul Donlan, who became group manager at BAO the year prior to Microsoft's acquisition. "We were a small shop, and we played by small-shop rules, and that sometimes gave the Microsoft people difficulty. It was very easy for us to say no, which frustrated them tremendously."

MICROSOFT'S PURCHASE OF ***Flight Simulator*** brought an alluring visual showcase in-house at the right time. The flashy launch of Windows 95, hosted by Jay Leno, leaned on media features to convince consumers it was time to retire MS-DOS computers and buy a Windows replacement. (This push, incidentally, is what put the game on my radar: *Microsoft Flight Simulator for Windows 95* came bundled with my family's first Windows 95 computer.)

Moving *Flight Simulator* to Windows was no small feat. The game was deeply rooted in MS-DOS and the increasingly arcane software development practices of the early 1980s. Windows 95 could in theory run MS-DOS applications, but this wasn't a good fit for *Flight Simulator*.

SubLogic's Denhart explains that up to this point, *Flight Simulator* didn't really use Microsoft's operating system: "You'd stick the floppy into the floppy drive, it'd boot up, and I think it ran a minimal MS-DOS, but just enough to get started. And then it basically ignored MS-DOS."

These screenshots, beginning with *Flight Simulator 1995 for Windows* [top row, left] and ending with the 2022 iteration of *Flight Simulator* [bottom row, right], show increasingly realistic views from the cockpit.

The team had also resisted early versions of Windows over concerns it would slow the simulator to a crawl.

But now that Microsoft was in charge, failure wasn't an option.

"When we went to *Flight Sim 95* there was this huge port," recalls Donlan, who credits Mike Schroeter, now a software engineer for Lockheed's Prepar3D simulation platform, with taking the role of point man. "I can't speak as to how significant it was across everything, but a lot of that code was being moved out of Assembly [language] and into C. There was a tremendous workload that was involved with that."

It was a first taste of Microsoft's culture of relentless toil. It was also only partially successful, as reviewers found performance issues on even the quickest home PCs. The team's concerns about Windows' ability to handle the simulation, it turned out, weren't unfounded. Still, *Microsoft Flight Simulator for Windows 95* was the eye candy Microsoft needed to highlight Windows 95's media prowess.

The team doubled down on visuals for *Microsoft Flight Simulator 98*, which again pushed the bleeding edge by adopting 3D hardware acceleration. Test lead Hal Bryan says the effort demanded long hours for testing various 3D accelerators, which had yet to settle on common standards. The tests paid off, however, and *Flight Simulator 98* quelled reviewers' complaints about pokey performance.

Users also benefited from the rise of CD-ROM and DVD-ROM media, which provided space for more detailed textures, more terrain data, and quicker data-transfer speeds. Jason Dent, first hired for Microsoft's *Encarta World Atlas*, soon moved to assist with *Flight Simulator*. Satellite imagery had improved the simulator's visuals, but its data was still coarse—"between 4 and 16 kilometers on a side," says Dent. Entire mountains were missing from less-traveled regions. To avoid such gaps, Dent and his colleagues combined satellite images with land-use data to deliver scale and precision simultaneously.

The hard work came to fruition in *Flight Simulator 2000*, which reached a technical milestone: It mapped the entire planet in 1-kilometer blocks. Scot Bayless, the studio manager overseeing the team, says an early demo left Bill Gates stunned.

Bayless recalls that after explaining to Gates that the software included every airport on the planet, Gates responded by saying, "'You're full of shit. That's the stupidest fucking thing I've ever heard.'" This was Gates's highest form of praise, Bayless notes. "In the lore of Microsoft, if Bill says that to you, you're made."

And, for a time, *Flight Simulator* did have it made. New versions landed on best-seller charts. The team, now renamed Aces Game Studio, created or contracted spin-offs like *Microsoft Combat Flight Simulator* and *Microsoft Space Simulator*. There was even talk of a universal platform for general-purpose, world-scale simulation, which eventually spun into Microsoft's Enterprise Simulation Platform. ESP lasted only a few years but was licensed by Lockheed Martin for its Prepar3D simulation platform. In retrospect, ESP feels like a predecessor to modern efforts to build "digital twins" to simulate and replicate real-world environments.

Yet *Flight Simulator* had a problem, and it was coming from inside the company: the Xbox game console. Launched in 2001, Xbox was built to oppose Sony's PlayStation 2, released in 2000, which had a DVD drive and could (with an accessory) connect to the Internet. Microsoft worried that some consumers might view it as a low-cost PC alternative.

Aces Game Studio explored bringing *Flight Simulator* to Xbox, says Bryan. But these efforts were frustrated by the challenge of adapting keyboard and mouse controls to the game pad. Bayless believes this created a rift between the Aces studio and Microsoft, and he regrets not pushing harder for an Xbox version. "I think we would have ended up with a stronger, more flexible, more robust, more future-proof engine."

The young Bruce Artwick [right] and Stu Moment pose beside an airplane.

STU MOMENT

The Sopwith Camel, the famous British biplane from World War I, was featured in the first *Flight Simulator*, released in 1979.

IMPERIAL WAR MUSEUMS/GETTY IMAGES

Aces, flying solo in an Xbox-centric Microsoft Games Division, became an easy target when the financial crisis of 2008 forced company-wide layoffs. For those affected, it was a nasty surprise, but the years have allowed some of them to accept that Microsoft's decisions made sense, because the simulator's last iterations had arguably been stagnant, focusing on past strengths and ignoring new platforms.

It was Denhart, the subLogic employee who was there when it all began, who wound down the project. He stayed on to archive the studio's work, bringing this era of *Flight Simulator* to a close.

THEN CAME JORG NEUMANN, a Microsoft veteran working on a HoloLens project called HoloTour—an immersive virtual-reality tourist guide. It included a bird's-eye perspective of locales like Machu Picchu, in Peru. The project faced challenges, however, especially at Machu Picchu, where the team had less data than it would have liked.

"It was pretty clear that, even with on-the-ground photographs, it was superhard to do a full, nice 3D model," says Neumann. "At which point we just said, Why don't we go and just have a plane fly overhead and give us the lidar data and appropriate photogrammetry?" (Lidar is a laser-based technique for estimating range to an object, while photogrammetry is 3D information extracted from photographs.)

The flyover never happened, but Neumann's perspective changed. "The idea persisted in my head. There is something there. We should try to get our game worlds augmented via aerial data." He realized Microsoft already had the perfect application: *Flight Simulator*.

Neumann, using data from Bing, threw together a demo of a Cessna flying over Seattle—the same plane and city available in the original *Microsoft Flight Simulator*. It looked spectacular, even at that early stage. The project progressed, and Denhart's archive proved invaluable.

"The code base and the project were really well archived," says Neumann. The code was sent to Asobo Studio, the lead developer on the HoloTour project, and used to preserve compatibility with third-party planes designed for *Flight Simulator X*, the last iteration released by Aces Game Studio. The new *Flight Simulator* also retains a "legacy" mode that activates the old flight model, preserving a lineage tracing all the way back to 1982.

Most people flying today's *Flight Simulator* will enjoy the default "modern" simulation, which models up to 1,500 flight surfaces. Airflow over each point in the simulation is determined by not only the plane's speed and design but also environmental effects such as weather and nearby terrain. This level of simulation was unimaginable in 1982, but today it can run on any recent midrange AMD or Intel processor.

Hal Bryan notes that the prior simulation fell apart in extreme situations, such as a stall and spin, so that the plane would behave in a wooden and overly predictable fashion. He knows, because that's how he used to test the thing. The new simulation can precisely model airflow over many surfaces and can thus organically determine when a stall would begin and whether it becomes a spin.

While Bing's data was useful in creating *Flight Simulator*'s world, the team still faced limitations. Quality photogrammetry data isn't available for every inch of the ground. To fill in the gaps, Asobo used Blackshark.ai's machine learning to convert photogrammetry data and satellite photos into a reproduction of the surface of our planet. The Blackshark.ai technology automatically creates buildings and adds them where appropriate, based on satellite photos. Machine learning also corrects color variations between photos while removing clouds and shadows.

"We wanted to have unique buildings, and basically you do this by procedural generation, which takes input from building footprints, the roof type, roof color, zoning, building density, and other information," says Arno Hollosi, chief technology officer of Blackshark.ai. This data is then modified by "archetypes" that have styles appropriate for the geographic region. The result is a diverse range of 3D buildings that look realistic, at least from a thousand or so meters above the ground. This technique can also depict small communities and even lone rural houses and buildings, something artists could never hope to accomplish while adding buildings one by one.

Even so, the modern simulator isn't perfect. A city street may look right, but your house probably won't. Simulated air traffic isn't as heavy as it is in reality. In-simulation messages from air-traffic control are often inaccurate or absent, especially at midsize airports. Weather is often stunningly beautiful, but it only vaguely mimics the real world. The solutions to these problems will, like so many of *Flight Simulator*'s additions and features, require new technologies.

Yet one core success is undeniable: *Microsoft Flight Simulator* fits the entire world in your PC. It can even fit the entire world in your pocket through Microsoft's xCloud streaming app for smartphones, allowing anyone with a modern smartphone to load the simulator and fly (virtually) from any location in the world to any other.

"We had this ambition to get the whole world in there," says Bayless. "And, in fact, we kind of did." ■

__PATCH_ME_IF_YOU_CAN___

How a New Jersey teenager's malware threatened the entire Internet

By Scott J. Shapiro

ILLUSTRATION BY MIKE MCQUADE

ATK ATK ATK

[10241] attack_http_state

util_zero(conn->path + 1, HTTP_PATH_MAX - 1)

427 SRVLOC

0 0 427 SRVLOC

void (uint8_t targs_len, attack_target *targs attack_option

if HTTP_CONN_RESTART (conn->keepalive) HTTP_CONN_SEND HTTP_CONN

conn->rdbuf ";", 1 = -1

switch ( )

fake_time - conn->last_recv > 30

// DEBUG

11 0 XTD - V2

ii = 65535 conn->to_send = (80 * 1024 * 1024);

fake_time - conn->last_recv > 30

427 SRVLOC

SERV sockets > HTTP_CONNECTION_MAX

1486 TABLE_ATK_CONNECTION_HDR #ifdef DEBUG

FIRST-YEAR COLLEGE STUDENTS are understandably frustrated when they can't get into popular upper-level electives. But they usually just gripe. Paras Jha was an exception. Enraged that upper-class students were given priority to enroll in a computer-science elective at Rutgers, the State University of New Jersey, Paras decided to crash the registration website so that no one could enroll. • On Wednesday night, 19 November 2014, at 10:00 p.m. EST—as the registration period for first-year students in spring courses had just opened—Paras launched his first distributed denial-of-service (DDoS) attack. He had assembled an army of some 40,000 bots, primarily in Eastern Europe and China, and unleashed them on the Rutgers central authentication server. The botnet sent thousands of fraudulent requests to authenticate, overloading the server. Paras's classmates could not get through to register.

The next semester Paras tried again. On 4 March 2015, he sent an email to the campus newspaper, *The Daily Targum*: "A while back you had an article that talked about the DDoS attacks on Rutgers. I'm the one who attacked the network.... I will be attacking the network once again at 8:15 pm EST." Paras followed through on his threat, knocking the Rutgers network offline at precisely 8:15 p.m.

On 27 March, Paras unleashed another assault on Rutgers. This attack lasted four days and brought campus life to a standstill. Fifty thousand students, faculty, and staff had no computer access from campus.

On 29 April, Paras posted a message on Pastebin, a website popular with hackers for sending anonymous messages. "The Rutgers IT department is a joke," he taunted. "This is the third time I have launched DDoS attacks against Rutgers, and every single time, the Rutgers infrastructure crumpled like a tin can under the heel of my boot."

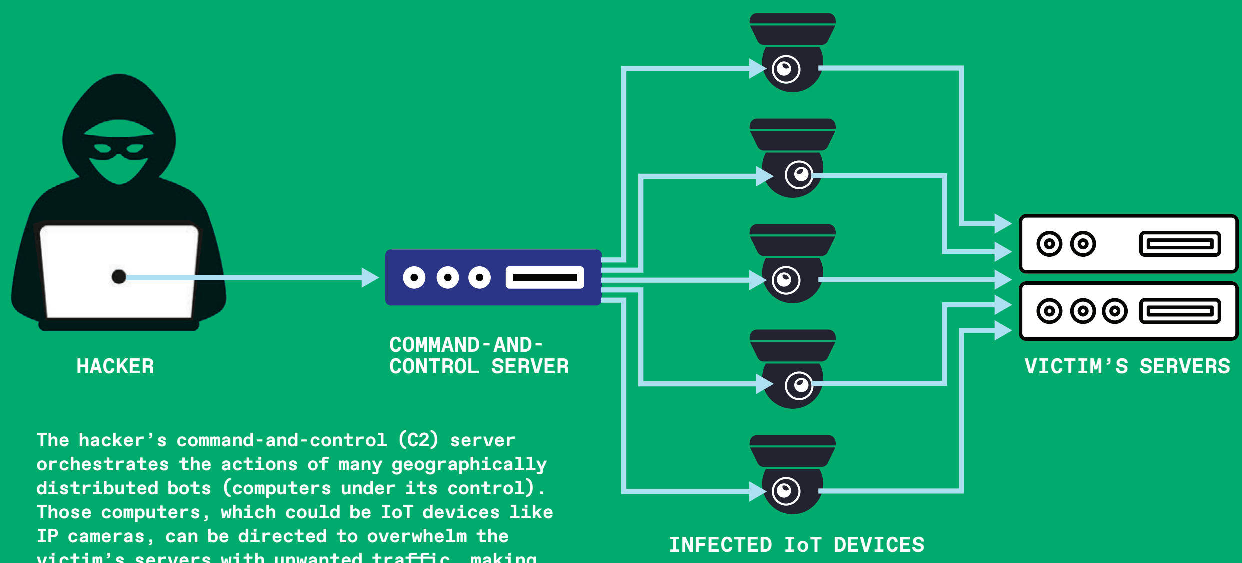
Paras was furious that Rutgers chose Incapsula, a small cybersecurity firm based in Massachusetts, as its DDoS-mitigation provider. He claimed that Rutgers chose the cheapest company. "Just to show you the poor quality of Incapsula's network, I have gone ahead and decimated the Rutgers network (and parts of Incapsula), in the hopes that you will pick another provider that knows what they are doing."

Paras's fourth attack on the Rutgers network, taking place during finals, caused chaos and panic on campus. Paras reveled in his ability to shut down a major state university, but his ultimate objective was to force it to abandon Incapsula. Paras had started his own DDoS-mitigation service, ProTraf Solutions, and wanted Rutgers to pick ProTraf over Incapsula. And he wasn't going to stop attacking his school until it switched.

PARAS JHA WAS BORN and raised in Fanwood, a leafy suburb in central New Jersey. When Paras was in the third grade, a teacher recommended that he be evaluated for attention deficit hyperactivity disorder, but his parents didn't follow through.

As Paras progressed through elementary school, his struggles increased. Because he was so obviously intelligent, his teachers and parents attributed his lackluster performance to laziness and apathy. His perplexed parents pushed him even harder.

Paras sought refuge in computers. He taught himself how to code when he was 12 and was hooked. His parents indulged this passion, buying him a computer and providing him with unrestricted Internet access. But their indulgence led Paras to isolate himself further, as he spent all his time coding, gaming, and hanging out with his online friends.



The hacker's command-and-control (C2) server orchestrates the actions of many geographically distributed bots (computers under its control). Those computers, which could be IoT devices like IP cameras, can be directed to overwhelm the victim's servers with unwanted traffic, making them unable to respond to legitimate requests.

Paras was particularly drawn to the online game *Minecraft*. In ninth grade, he graduated from playing *Minecraft* to hosting servers. It was in hosting game servers that he first encountered DDoS attacks.

Minecraft server administrators often hire DDoS services to knock rivals offline. As Paras learned more sophisticated DDoS attacks, he also studied DDoS defense. As he became proficient in mitigating attacks on *Minecraft* servers, he decided to create ProTraf Solutions.

Paras's obsession with *Minecraft* attacks and defense, compounded by his untreated ADHD, led to an even greater retreat from family and school. His poor academic performance in high school frustrated and depressed him. His only solace was Japanese anime and the admiration he gained from the online community of *Minecraft* DDoS experts.

Paras's struggles deteriorated into paralysis when he enrolled in Rutgers, studying for a B.S. in computer science. Without his mother's help, he was unable to regulate the normal demands of living on his own. He could not manage his sleep, schedule, or study. Paras was also acutely lonely. So he immersed himself in hacking.

Paras and two hacker friends, Josiah White and Dalton Norman, decided to go after the kings of DDoS—a gang known as VDoS. The gang had been providing these services to the world for four years, which is an eternity in cybercrime. The decision to fight experienced cybercriminals may seem brave, but the trio were actually older than their rivals. The VDoS gang members had been only 14 years old when they started to offer DDoS services from Israel in 2012. These 19-year-old American teenagers would be going to battle against two 18-year-old Israeli teenagers. The war between the two teenage gangs would not only change the nature of malware. Their struggle for dominance in cyberspace would create a doomsday machine.

BOTNET MALWARE IS useful for financially motivated crime because botmasters can tell the bots in their thrall to implant malware on vulnerable machines, send phishing emails, or engage in click fraud, in which botnets profit by directing bots to click pay-per-click ads. Botnets are also great DDoS weapons because they can be trained on a target and barrage it from all directions. One day in February 2000, for example, the hacker MafiaBoy knocked out Fifa.com, Amazon.com, Dell, E-Trade, eBay, CNN, as well as Yahoo, at the time the largest search engine on the Internet.

After taking so many major websites offline, MafiaBoy was deemed a national-security threat. President Clinton ordered a national manhunt to find him. In April 2000, MafiaBoy was arrested and charged, and in January 2001 he pled guilty to 58 charges of denial-of-service attacks. Law enforcement did not reveal MafiaBoy's real name, as this national-security threat was 15 years old.

Both MafiaBoy and the VDoS crew were adolescent boys who crashed servers. But whereas MafiaBoy did it for the sport, VDoS did it for the money. Indeed, these teenage Israeli kids were pioneering tech entrepreneurs. They helped launch a new form of cybercrime: DDoS as a service. With it, anyone could now hack with the click of a button, no technical knowledge needed.

It might be surprising that DDoS providers could advertise openly on the Web. After all, DDoSing another

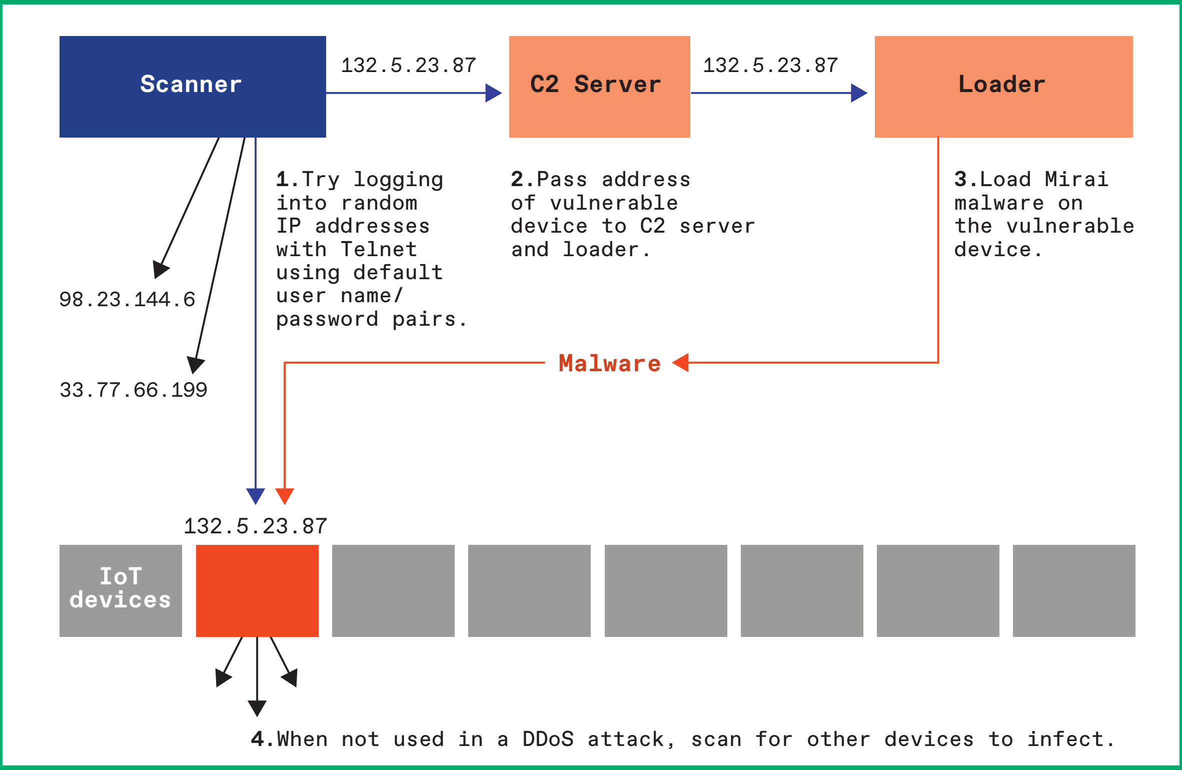
Bots for Tots

Here's how three teenagers built a botnet that could take down the Internet

THE MIRAI BOTNET, with all its devastating potential, was not the product of an organized-crime or nation-state hacking group—it was put together by three teenage boys. They rented out their botnet to paying customers to do mischief with and used it to attack chosen targets of their own. But the full extent of the danger became apparent only later, after this team made the source code for their malware public. Then others used it to do greater harm: crashing Germany's largest Internet service provider; attacking Dyn's Domain Name System servers, making the Internet unusable for millions; and taking down all of Liberia's Internet—to name a few examples.

The Mirai botnet exploited vulnerable Internet of Things devices, such as Web-connected video cameras, ones that supported Telnet, an outdated system for logging in remotely. Owners of these devices rarely updated their passwords, so they could be easily guessed using a strategy called a dictionary attack.

The first step in assembling a botnet was to scan random IP addresses looking for vulnerable IoT devices, ones whose passwords could be guessed. Once identified, the addresses of these devices were passed to a "loader," which would put the malware on the device. Infected devices located all over the world could then be used for distributed denial-of-service attacks.



website is illegal everywhere. To get around this, these "booter services" have long argued they perform a legitimate function: providing those who set up Web pages a means to stress test websites.

In theory, such services do play an important function. But only in theory. As a booter-service provider admitted to University of Cambridge researchers, "We do try to market these services towards a more legitimate user base, but we know where the money comes from."

PARAS DROPPED OUT of Rutgers in his sophomore year and, with his father's encouragement, spent the next year focused on building ProTraf Solutions, his DDoS-mitigation business. And just like a mafia don running a protection racket, he had to make that protection needed. After launching four DDoS attacks his freshman year, he attacked Rutgers yet again in September 2015, still hoping that his former school would give up on Incapsula. Rutgers refused to budge.

ProTraf Solutions was failing, and Paras needed cash. In May 2016, Paras reached out to Josiah White. Like Paras, Josiah frequented Hack Forums. When he was 15, he had developed major portions of Qbot, a botnet worm that at its height in 2014 had enslaved half a million computers. Now 18, Josiah switched sides and worked with his friend Paras at ProTraf doing DDoS mitigation.

But Josiah soon returned to hacking and started working with Paras to take the Qbot malware, improve it, and build a bigger, more powerful DDoS botnet. Paras and Josiah then partnered with 19-year-old Dalton Norman. The trio turned into a well-oiled team: Dalton found the vulnerabilities; Josiah updated the botnet malware to exploit these vulnerabilities; and Paras wrote the C2—software for the command-and-control server—for controlling the botnet.

But the trio had competition. Two other DDoS gangs—Lizard Squad and VDoS—decided to band together to build a giant botnet. The collaboration, known as PoodleCorp, was successful. The amount of traffic that could be unleashed on a target from PoodleCorp's botnet hit a record value of 400 gigabits per second, almost four times the rate that any previous botnet had achieved. They used their new weapon to attack banks in Brazil, U.S. government sites, and *Minecraft* servers. They achieved this firepower by hijacking 1,300 Web-connected cameras. Web cameras tend to have powerful processors and good connectivity, and they are rarely patched. So a botnet that harnesses video has enormous cannons at its disposal.

While PoodleCorp was on the rise, Paras, Josiah, and Dalton worked on a new weapon. By the beginning of August 2016, the trio had completed the first version of their botnet malware. Paras called the new code Mirai, after the anime series *Mirai Nikki*.

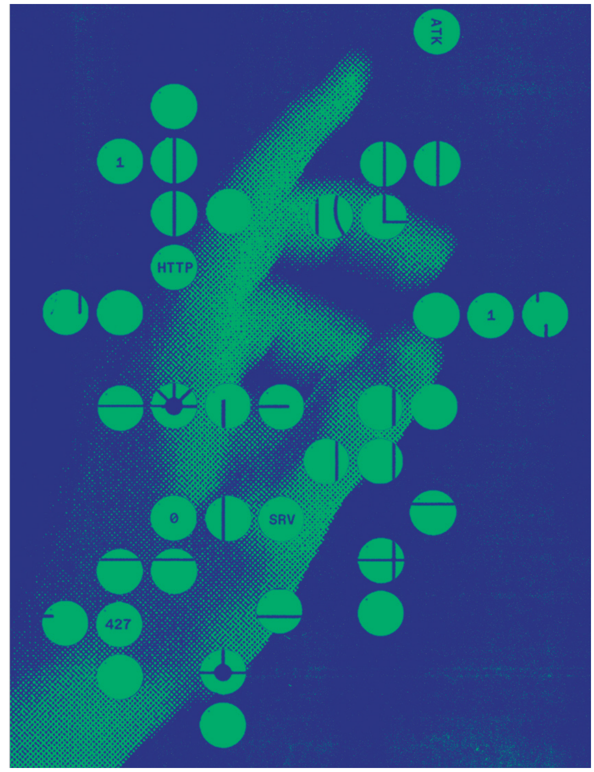
When Mirai was released, it spread like wildfire. In its first 20 hours, it infected 65,000 devices, doubling in size every 76 minutes. And Mirai had an unwitting ally in the botnet war then raging.

Up in Anchorage, Alaska, the FBI cyber unit was building a case against VDoS. The FBI was unaware of Mirai or its war with VDoS. The agents did not regularly read online boards such as Hack Forums. They did not know that the target of their investigation was being decimated. The FBI also did not realize that Mirai was ready to step into the void.

The head investigator in Anchorage was Special Agent Elliott Peterson. A former U.S. Marine, Peterson is a calm and self-assured agent with a buzz cut of red hair. At the age of 33, Peterson had returned to his native state of Alaska to prosecute cybercrime.

On 8 September 2016, the FBI's Anchorage and New Haven cyber units teamed up and served a search warrant in Connecticut on the member of PoodleCorp who ran the C2 that controlled all its botnets. On the same day, the Israeli police arrested the VDoS founders in Israel. Suddenly, PoodleCorp was no more.

The Mirai group waited a couple of days to assess the battlefield. As far as they could tell, they were the only botnet left standing. And they were ready to use their new power. Mirai won the war because Israeli and American law enforcement arrested the masterminds behind PoodleCorp.



But Mirai would have triumphed anyway, as it was ruthlessly efficient in taking control of Internet of Things devices and excluding competing malware.

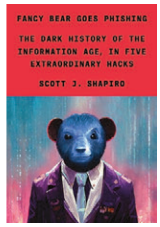
A few weeks after the arrests of those behind VDoS, Special Agent Peterson found his next target: the Mirai botnet. In the Mirai case, we do not know the exact steps that Peterson's team took in their investigation: Court orders in this case are currently "under seal," meaning that the court deems them secret. But from public reporting, we know that Peterson's team got its break in the usual way—from a Mirai victim: Brian Krebs, a cybersecurity reporter whose blog was DDoSed by the Mirai botnet on 25 September.

The FBI uncovered the IP address of the C2 and loading servers but did not know who had opened the accounts. Peterson's team likely subpoenaed the hosting companies to learn the names, emails, cellphones, and payment methods of the account holders. With this information, it would seek court orders and then search warrants to acquire the content of the conspirators' conversations.

Still, the hunt for the authors of the Mirai malware must have been a difficult one, given how clever these hackers were. For example, to evade detection Josiah didn't just use a VPN. He hacked the home computer of a teenage boy in France and used his computer as the "exit node." The orders for the botnet, therefore, came from this computer. Unfortunately for the owner, he was a big fan of Japanese anime and thus fit the profile of the hacker. The FBI and the French police discovered their mistake after they raided the boy's house.

Illustration by Mike McQuade

This article was adapted from the author's new book, *Fancy Bear Goes Phishing: The Dark History of the Information Age, in Five Extraordinary Hacks* (Farrar, Straus and Giroux, 2023).



AFTER WIELDING ITS POWER for two months, Paras dumped nearly the complete source code for Mirai on Hack Forums. "I made my money, there's lots of eyes looking at IOT now, so it's time to GTFO [Get The F*** Out]," Paras wrote. With that code dump, Paras had enabled anyone to build their own Mirai. And they did.

Dumping code is reckless, but not unusual. If the police find source code on a hacker's devices, they can claim that they "downloaded it from the Internet." Paras's irresponsible disclosure was part of a false-flag operation meant to throw off the FBI, which had been gathering evidence indicating Paras's involvement in Mirai and had contacted him to ask questions. Though he gave the agent a fabricated story, getting a text from the FBI probably terrified him.

Mirai had captured the attention of the cybersecurity community and of law enforcement. But not until after Mirai's source code dropped would it capture the attention of the entire United States. The first attack after the dump was on 21 October, on Dyn, a company based in Manchester, N.H., that provides Domain Name System (DNS) resolution services for much of the East Coast of the United States.

It began at 7:07 a.m. EST with a series of 25-second attacks, thought to be tests of the botnet and Dyn's infrastructure. Then came the sustained assaults: of one hour, and then five hours. Interestingly, Dyn was not the only target. Sony's PlayStation video infrastructure was also hit. Because the torrents were so immense, many other websites were affected. Domains such as cnn.com, facebook.com, and nytimes.com wouldn't work. For the vast majority of these users, the Internet became unusable. At 7:00 p.m., another 10-hour salvo hit Dyn and PlayStation.

Further investigations confirmed the point of the attack. Along with Dyn and PlayStation traffic, the botnet targeted Xbox Live and Nuclear Fallout game-hosting servers. Nation-states were not aiming to hack the upcoming U.S. elections. Someone was trying to boot players off their game servers. Once again—just like MafiaBoy, VDoS, Paras, Dalton, and Josiah—the attacker was a teenage boy, this time a 15-year-old in Northern Ireland named Aaron Sterritt.

Meanwhile, the Mirai trio left the DDoS business, just as Paras said. But Paras and Dalton did not give up on cybercrime. They just took up click fraud.

Click fraud was more lucrative than running a booter service. While Mirai was no longer as big as it had been, the botnet could nevertheless generate significant advertising revenue. Paras and Dalton earned as much money in one month from click fraud as they ever made with DDoS. By January 2017, they had earned over US $180,000, as opposed to a mere $14,000 from DDoSing.

Had Paras and his friends simply shut down their booter service and moved on to click fraud, the world would likely have forgotten about them. But by releasing the Mirai code, Paras created imitators. Dyn was the first major copycat attack, but many others followed. And due to the enormous damage these imitators wrought, law enforcement was intensely interested in the Mirai authors.

After collecting information tying Paras, Josiah, and Dalton to Mirai, the FBI quietly brought each up to Alaska. Peterson's team showed the suspects its evidence and gave them the chance to cooperate. Given that the evidence was irrefutable, each folded.

Paras Jha was indicted twice, once in New Jersey for his attack on Rutgers, and once in Alaska for Mirai. Both indictments carried the same charge—one violation of the Computer Fraud and Abuse Act. Paras faced up to 10 years in federal prison for his actions. Josiah and Dalton were only indicted in Alaska and so faced 5 years in prison.

The trio pled guilty. At the sentencing hearing held on 18 September 2018, in Anchorage, each of the defendants expressed remorse for his actions. Josiah White's lawyer conveyed his client's realization that Mirai was "a tremendous lapse in judgment."

Unlike Josiah, Paras spoke directly to Judge Timothy Burgess in the courtroom. Paras began by accepting full responsibility for his actions and expressed his deep regret for the trouble he'd caused his family. He also apologized for the harm he'd caused businesses and, in particular, Rutgers, the faculty, and his fellow students.

The Department of Justice made the unusual decision not to ask for jail time. In its sentencing memo, the government noted "the divide between [the defendants'] online personas, where they were significant, well-known, and malicious actors in the DDoS criminal milieu and their comparatively mundane 'real lives' where they present as socially immature young men living with their parents in relative obscurity." It recommended five years of probation and 2,500 hours of community service.

The government had one more request—for that community service "to include continued work with the FBI on cybercrime and cybersecurity matters." Even before sentencing, Paras, Josiah, and Dalton had logged close to 1,000 hours helping the FBI hunt and shut down Mirai copycats. They contributed to more than a dozen law-enforcement and research efforts. In one instance, the trio assisted in stopping a nation-state hacking group. They also helped the FBI prevent DDoS attacks aimed at disrupting Christmas-holiday shopping. Judge Burgess accepted the government's recommendation, and the trio escaped jail time.

The most poignant moments in the hearing were Paras's and Dalton's singling out for praise the very person who caught them. "Two years ago, when I first met Special Agent Elliott Peterson," Paras told the court, "I was an arrogant fool believing that somehow I was untouchable. When I met him in person for the second time, he told me something I will never forget: 'You're in a hole right now. It's time you stop digging.' " Paras finished his remarks by thanking "my family, my friends, and Agent Peterson for helping me through this." ■

THE INSTITUTE

NEWS OF THE IEEE
VOLUME 47 / ISSUE 2

Siavash Alamouti on Stopping the Spread of Disinformation P. 52

MARCONI SOCIETY

Connecting the Unconnected

AT IEEE, WE know that the advancement of science and technology is the engine that drives the improvement of the quality of life for every person on this planet. Unfortunately, as we are all aware, today's world faces significant challenges including escalating conflicts, a climate crisis, food insecurity, gender inequality, and the approximately 2.7 billion people who cannot access the Internet.

Bridging the divide

The COVID-19 pandemic exposed the digital divide like never before. The world saw the need for universal broadband connectivity for remote work, online education, telemedicine, entertainment, and social networking.

Those who had access thrived while those without it struggled. As millions of classrooms moved online, the lack of connectivity made it difficult for some students to participate in remote learning. Adults who could not perform their job virtually faced layoffs or reduced work hours.

The pandemic also exposed weaknesses in the global infrastructure that supports the citizens of the world. It became even more apparent that vital communications, computing, energy, and distribution infrastructure was not always equitably distributed, particularly in less developed regions.

Nearly 45 percent of global households do not have access to the Internet, according to UNESCO. A report from UNICEF estimates that nearly two-thirds of the world's schoolchildren lack Internet access at home.

This digital divide also impacts women, who are 23 percent less likely than men to use the Internet. According to the United Nations Educational, Scientific and Cultural Organization, in 10 countries across Africa, Asia, and South America, women are between 30 percent and 50 percent less likely than men to make use of the Internet.

Even in developed countries, Internet access is often lower than one might imagine. More than 6 percent of the U.S. population does not have a high-speed connection. In Australia, the figure is 13 percent. Globally, just over half of households have an Internet connection, according to UNESCO. In the developed world, 87 percent are connected, compared with 47 percent in developing nations and just 19 percent in the least developed countries.

Benefits of technology

As IEEE looks to lead the development of technology to tackle climate change and empower universal prosperity, it is essential that we recognize the role that meaningful connectivity and digital technology play in the organization's goals to support global sustainability, drive economic growth, and transform health care, education, employment, gender equality, and youth empowerment.

IEEE members around the globe are continuously developing and applying technology to help solve these problems. It is that universal passion—to improve global conditions—that is at the heart of our mission, as well as our expanding partnerships and significant activities supporting the achievement of the U.N. Sustainable Development Goals.

One growing partnership is with the International Telecommunication Union, a U.N. specialized agency that helps set policy related to information and communication technologies. IEEE Member Doreen Bogdan-Martin was elected as ITU secretary-general and took office on 1 January, becoming the

IEEE

first woman to lead the 155-year-old organization. Bogdan-Martin is the recipient of this year's IEEE President's Award [see below].

IEEE and ITU share the goal of bringing the benefits of technology to all of humanity. I look forward to working closely with the U.N. agency to promote meaningful connectivity, intensify cooperation to connect the unconnected, and strengthen the alignment of digital technologies with inclusive sustainable development.

I truly believe that one of the most important applications of technology is to improve people's lives. For those in underserved regions of the world, technology can improve educational opportunities, provide better health care, alleviate suffering, and maintain human dignity.

Technology and technologists, particularly IEEE members, have a significant role to play in shaping life on this planet. They can use their skills to develop and advance technology—from green energy to reducing waste and emissions, and from transportation electrification to digital education, health, and agriculture. As a person who believes in the power of technology to benefit humanity, I find this to be a very compelling vision for our shared future.

—SAIFUR RAHMAN
IEEE president and CEO

Please share your thoughts with me at president@ieee.org.

ITU

2023 IEEE PRESIDENT'S AWARD

I had the pleasure of presenting the 2023 IEEE President's Award to Doreen Bogdan-Martin, secretary-general of the International Telecommunication Union, on 28 March at ITU headquarters in Geneva. The award recognizes her distinguished leadership at the agency and her notable contributions to the global public.

It is my honor to recognize such a transformational leader and IEEE member for her demonstrated commitment to bridging the digital divide and to ensuring connectivity that is safe, inclusive, and affordable to all.

EDITOR'S NOTE

Engineers Take On Missions Close to Their Hearts

They are combating disinformation and inspiring future generations

BEING AN ENGINEER isn't just about developing new technology. Some engineers also use their skills to solve societal problems close to their heart.

Siavash Alamouti is working to slow the spread of disinformation on the Internet. The wireless pioneer says he believes the Internet's business model of making money from advertisements and page views encourages people to share false information, which is causing social unrest. Owning our own data is key to fixing the problem, he says. His solution? Decentralize the cloud—which would help bring about a more open Internet and eventually allow consumers to choose what data they want to sell.

Alamouti, cofounder and executive board chairman of Mimik, helped develop the first hybrid edge cloud computing platform. It enables any smart device to act as a cloud server.

The IEEE member says the HEC platform allows consumers to better control how their data is stored, shared, and monetized. On page 52, learn more about his work to create a sustainable business model for data monetization and custodianship.

Anna Zakrzewska is on a mission to introduce children to science, technology, engineering, and mathematics careers and to nurture their interest in STEM fields. The IEEE senior member didn't have an engineering mentor to guide her when she was young. She says she often wonders how her career would be different if she had been exposed to STEM earlier.

Zakrzewska's work in STEM outreach garnered her this year's IEEE Theodore W. Hissey Outstanding Young Professional Award. Find out more about her telecommunications research and her volunteerism on page 58.

Kathy Kleiman's quest is to highlight the pioneering work of the female mathematicians who programmed ENIAC, the first programmable general-purpose all-electronic computer. ENIAC was built to improve the accuracy of U.S. artillery during World War II and to solve other complicated equations.

The mathematicians were largely unknown until Kleiman decided to research their contributions to computer science. She teaches Internet technology and governance for lawyers at American University.

Her book, *Proving Ground: The Untold Story of the Six Women Who Programmed the World's First Modern Computer*, is a follow-up to an award-winning documentary she helped produce about the so-called hidden figures. On page 62, read our interview with Kleiman about her mission to highlight the programmers' contributions.

"By sharing strong stories of pioneers in the fields who are women and people of color. I hope we can open the doors to computing and engineering," Kleiman says.

—KATHY PRETZ
Editor in chief, *The Institute*

For updates about IEEE and its members, visit us at spectrum.ieee.org/the-institute

PROFILE

Siavash Alamouti Works to Stop the Spread of Disinformation

Owning your data is key, says the wireless pioneer

BY KATHY PRETZ

MARCONI SOCIETY

TO STOP THE spread of disinformation on the Internet, do away with the current advertising-driven business model and instead let consumers sell their own data, IEEE Member Siavash Alamouti says. The crux of the problem, the wireless innovator says, is the model makes money from ads and page views, which encourages people to share false information and causes social unrest.

"When you encourage people to spread falsities to get advertising and make money, then this thing kind of builds on itself. It's a very dangerous situation," Alamouti warns. "The unfortunate thing is, now there's more disinformation on the Internet than there is information. This is not a technology problem; it's a business-model problem."

He says part of the solution is to decentralize the cloud—which would help bring about a more open Internet and eventually allow consumers to choose what data of theirs they want to sell.

A decentralized cloud would minimize the need to use third-party services to manage and store data on people and business enterprises. The decentralized cloud would allow users to keep most or all their data on local devices, cutting out the intermediaries now needed for hosting applications and managing transactions over the Internet. It would also allow users to monetize their data by sharing it directly, instead of allowing data brokers to steal and sell the information.

Alamouti received the 2022 Marconi Prize for his impact on the success of wireless devices and his groundbreaking work to develop technology that impacts humanity.

A cofounder and executive board chairman of Mimik, in Oakland, Calif., he helped develop the first hybrid edge cloud computing platform, which enables any smart device to act as a cloud server. The HEC can allow consumers to better control how their data is stored, shared, and monetized, Alamouti says.

"If we can get rid of all the unnecessary middlemen and come up with models to help people monetize their data, then we can disincentivize the spread of disinformation," Alamouti says. "We're never going to get rid of it, but at least we can minimize it."

He probably is best known for creating a space-time block code commonly used in wireless cellular communications. The Alamouti code, as it is known, is central to the wireless multiple-input multiple-output (MIMO) technique, which uses antennas at the transmitting and receiving ends to create multiple signal paths. This increases the data rate and improves channel reliability.

The code is used in billions of devices. It's built into Wi-Fi and has been included in all wireless cellular standards starting with 3G and most likely will continue to be used in 6G, he says.

Employer Mimik, in Oakland, Calif.
Title Executive chairman
Member grade Member
Alma mater University of British Columbia, in Vancouver

The Internet's business model

Alamouti first noticed the potential for the spread of disinformation on the Internet in 2010, when he began working as group R&D director for Vodafone in London.

His team worked on technologies that would limit the use of personal data for advertising and "eyeball-grabbing, because they are very disruptive and can incentivize the spread of disinformation," he says. "Disinformation is what demagogues and fascists use to divide people."

> "If we can get rid of all the unnecessary middlemen and come up with models to help people monetize their data, then we can disincentivize the spread of disinformation."

"I still don't think that realization has sunk in," he adds. "People haven't understood the massive negative impact that this has had on all aspects of our lives."

He says the advertising-driven business model is exactly the opposite of what he and early Internet pioneers set out to accomplish.

"Myself, and other colleagues who worked on mobile Internet, dedicated our lives to bringing affordable Internet to people," Alamouti says. "We thought it would significantly improve their lives because access to information is power. But it hasn't been affordable because our data, which is extremely valuable, is being exchanged for that access and then used to manipulate our basic life choices."

To Alamouti, data is an asset like other types of financial and physical capital such as cash, stock, and real estate.

"Data is worth a lot of money," he says, "so a platform is needed to help people monetize their data."

He's now focused on how to create a sustainable business model for data monetization and data custodianship, using technologies such as blockchain and smart contracts.

From refugee to engineer

Born in Tehran, Alamouti did not grow up wanting to be an engineer. He was more interested in studying art, literature, and music. Alamouti's father persuaded him to study mathematics and physics and to pursue humanities as a hobby.

He studied math and computer science at the Sharif University of Technology, in Tehran, but following the Islamic Cultural Revolution in 1980, he was expelled from the university for his political views. He ended up as a refugee in Madrid.

Canada accepted him as a political refugee in 1984, and he attended the University of British Columbia, in Vancouver, where he earned bachelor's and master's degrees in electrical engineering.

Wireless innovator

While at McCaw Cellular Communications (now AT&T Wireless), Alamouti worked on a fixed wireless system, which he says was one of the first orthogonal frequency-division multiplexing MIMO systems, known as Project Angel. While working on the project in 1996, Alamouti came up with the code named after him.

He was assigned to figure out how to overcome a challenging propagation environment so that customers could receive their data reliably. His team's studies showed that to get reliable connections, every home would need at least four active receivers, he says.

"I realized that we could instead put two transmitters on every base station and two receivers in people's homes to get the four levels of diversity," he says. "Adding an RF transmit chain [electronic components and substations] and a bit of signal processing at base stations and remote units was much cheaper and more viable." ■

IEEE STANDARDS

Big Data Could Accelerate Drug Approvals

BY KATHY PRETZ

IT CAN TAKE up to 15 years to get a new drug through the regulatory-approval process before it becomes available to patients. One of the reasons is the time it takes to analyze the mountains of data generated by clinical trials about the efficacy and safety of the product.

Drug companies perform quality-assurance tasks to verify that the steps carried out during clinical trials comply with government regulations. Regulatory agencies then check the data from the trials, as well as from the trial sponsors, to ensure that the safety of patients wasn't compromised.

The way in which quality is assured hasn't changed much in about 25 years, says Timothé Ménard, head of quality data science at pharmaceutical company Roche. Meanwhile, the number of clinical trials has increased, and their designs have become more complex.

Ménard and other drug-company quality leads and quality data scientists say advanced analytics could help speed up the QA process. The analytics programs use machine learning, statistical modeling, natural language processing, predictive modeling, and related techniques to evaluate information.

To better understand how applying data analytics for QA processes would work, Roche and a handful of other pharmaceutical companies formed the Intercompany Quality Analytics industry group (IMPALA) in 2019.

But Roshan D'Souza, head of Roche's data analytics and operations, says IMPALA couldn't hope to change the current QA processes on its own, so the IEEE IMPALA Consortium was formed.

Roche and 12 other drug companies are members of the consortium. Its mission is to work with regulatory agencies and other pharmaceutical companies to adopt advanced analytics and associated methodologies in the biopharmaceutical clinical-QA process, with a goal of bringing quality, life-saving medicines to patients faster.

"What we are trying to do is effect a paradigm shift," D'Souza says. "The shift must happen across the entire industry—which is why a consortium is necessary. Improving the quality of clinical trials is something we all agree is for the greater good. It's for the benefit of patients, society, and the health care system." ■

NEWS

IEEE Society Recognized for Work in DEI

BY JOANNA GOODRICH

FOR THE PAST 20 years, the IEEE Instrumentation and Measurement Society has been working to become more welcoming and inclusive for women, members from outside the United States and Canada, students, and young professionals. Its hard work in diversity, equity, and inclusion (DEI) has paid off.

IEEE IMS has increased its number of women leaders, conference speakers, and editors of its publications. To encourage more students to join, it holds competitions and offers scholarships, and a mentoring program was created to attract students and young professionals.

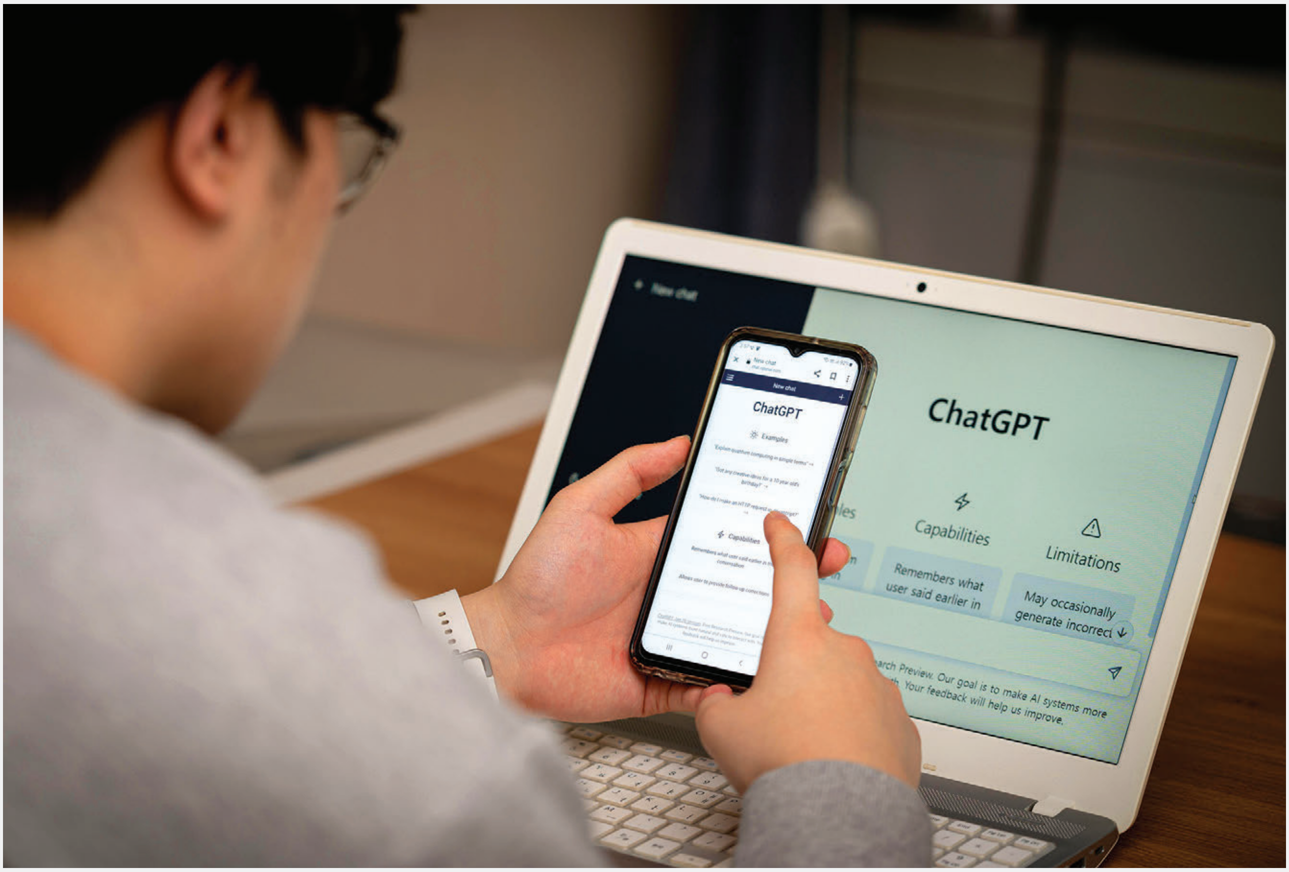
The society has also expanded its efforts to increase the number of its chapters in Africa through an initiative that is designed to also improve services offered in the continent and increase membership. The society established its first African chapter in South Africa, followed by one in Angola. A chapter in Ghana is in the process of forming. The society also has increased membership in Kenya, Nigeria, Tunisia, Uganda, and Zambia.

IEEE IMS's outreach program, which helps chapters thrive and establishes new ones, plays a large role in the initiative. Liaisons for each IEEE region visit inactive chapters to help reestablish them, and participate in active chapters' events such as workshops and regions to encourage growth. The liaisons seek out members who are experts in their field to participate in the society's lecturer program so they can share their knowledge.

The society's efforts have been recognized with the new IEEE Technical Activities Board Award for Society/Council Impact in DEI. The award was established last year to honor an IEEE society or council that has encouraged DEI by developing activities, programs, and services that promote efforts in the area. IEEE IMS is the first recipient of the award.

Diversity with regard to race, religion, gender, disability, age, national origin, sexual orientation, gender identity, and gender expression, IEEE Senior Member Juan Manuel Ramirez Cortés says, brings the society "scientific, technological, and cultural richness." Ramirez Cortés is the society's president. ■

ROCHE



ALAMY

OPINION

How ChatGPT Could Transform Academia

Universities will need to amend their teaching and assessment practices

BY SAURABH SINHA, LIZ BURD, AND JACO DU PREEZ

OPENAI IN NOVEMBER launched ChatGPT, one of the most significant real-world applications of artificial intelligence to date. The tool allows its users to quickly generate sophisticated textual content that is uniquely constructed.

Such content, therefore, likely can avoid detection by traditional plagiarism tools—which creates a concern for universities about how to assess students' learning and skill development.

Many types of assessments used to evaluate students require them to demonstrate they have understood new materials by investigating the content and collating their learning in the form of a written essay or report. The role of an academic assessor has been to evaluate individual students' submissions to gauge the breadth and depth of their understanding of the topic.

If students use ChatGPT to write an essay or report, the problem is the output generated provides limited, if any, representation about the quality of their learning. The AI tool offers the opportunity for well-written researched content without the need for a student to search for detailed sources.

Unfortunately, this type of problem is not new. For many years, students have been able to copy text from essay banks. Anti-plagiarism detection tools such as iThenticate and TurnItIn have deterred the use of such repositories. Although the anti-plagiarism tools have been successful, they use sophisticated pattern-matching techniques—which makes them ill-equipped to detect the language constructs resulting from advanced AI.

Another way students get around writing original content is through essay mills, which provide writing services for a fee. It could be argued that the open nature of ChatGPT has leveled the playing field between those who can and can't afford to pay for the services.

> “We recommend an approach where teaching and learning adapt to recognize the opportunities posed by new technologies.”

A different assessment

In the fields of science, technology, engineering, and mathematics, a far broader assessment strategy than essays is used. To meet the learning outcomes, STEM students must demonstrate skills such as programming. Because ChatGPT can solve many mathematical problems and generate and debug code, however, the computing field cannot simply ignore the AI evolution. And ChatGPT’s capabilities are certain to improve over time.

The immediate reaction from academia is likely to adopt traditional assessment strategies that comprise predominantly closed-book, exam-style assessments.

Before adopting that obvious quick fix, though, it is important to reflect on the reasons why a broad assessment strategy was adopted in the first place.

Engineering and computing students must tackle large, complex problems and adopt collaborative strategies. The skill sets are not easily or accurately tested individually in an examination hall in a three-hour period. To some extent, essays—and, perhaps even more controversially, doctoral theses—are already not well aligned to the needs of many employers.

Issues to consider

ChatGPT and similar technologies will continue to shape the future of what we call the *World of Work* (WoW). As employers increasingly adopt advanced AI, the academic world will need to amend its teaching and assessment practices. ChatGPT and other AI tools are already being adopted in industry as a way to automate mundane tasks. The big question is: Should educators ban such developments or embrace them?

Here are some issues that universities might want to consider.

- **Awareness is the best line of defense.** Educate students and staff about the strengths and weaknesses of AI-generated content. For example, when does reliance on localized or peer-reviewed content matter, and when is a quick-and-dirty content review sufficient?

- **Develop assessment and other educational practices for the WoW to embrace or reject the use of AI-generated content.** Using authentic assessment tasks that are aligned to a local context or problem, for example, would require students to foster a culture of exploration and curiosity. Project-based assessment tasks are good examples that enable students to conduct exploration of ChatGPT and similar tools, but they ultimately demonstrate the learning and skill on their own. Assessment criteria also will need to recognize sophisticated use of AI but ensure greater recognition for elements that demonstrate higher-order skills such as evaluation and synthesis.

- **Reassure staff that new tools are emerging to detect the use of AI.** Princeton student Edward Tian has already created one such tool: GPTZero, which adopts thinking similar to OpenAI’s tool but uses deep learning in reverse to detect ChatGPT.

- **Adapt to embrace opportunities brought on by innovation.** Consider using ChatGPT and similar tools to advance pedagogy and curricula. Everyone finds unnecessary repetition and menial tasks tiresome. Use the tools to stretch and enable more innovation by students.

- **Reinforce principles of professional standards and ethics.** Advance a culture of academic integrity, acknowledging that new tools will emerge. If ethical culture is engraved in the group ethos, students and scholars will use new AI tools appropriately.

ChatGPT and similar tools should be seen as accelerating necessary change. We recommend an approach where teaching and learning adapt to recognize the opportunities posed by new technologies and continue to foster a culture of exploration and curiosity. Ultimately, our priority is to provide graduates ready to face the ever-changing WoW. ■

The views expressed here are the authors’ own and do not represent positions of IEEE Spectrum, The Institute, *or IEEE. The authors write in their personal capacity.*

IEEE Fellow Saurabh Sinha is a professor of microelectronics and deputy vice chancellor of research and internationalization at the University of Johannesburg. **Liz Burd** is an education technologies specialist and provost at Griffith University, in Queensland, Australia. **Jaco du Preez** is a Ph.D. student at the University of Johannesburg and a microwave design engineer at Saab Grintek Defence, in Centurion, Gauteng, South Africa.

ISTOCK

NEWS

IEEE Standards Initiatives Focus on Extended Reality

BY YU YUAN AND SRIKANTH CHANDRASEKARAN

SOME OBSERVERS SAY the metaverse is an expanded set of digital worlds that will grow out of online environments people are already familiar with, such as enhancing the extended-reality experience used in online gaming. Others say the metaverse will usher in a decentralized ecosystem that empowers users to create digital assets of their own choosing and engage in digital commerce.

As the metaverse evolves, one thing is certain: It has tremendous potential to fundamentally transform the ways we work, learn, play, and live. But there will be issues to deal with along the way. That is why the IEEE Standards Association is working to help define, develop, and deploy the technologies, applications, and governance practices needed to help turn metaverse concepts into practical realities, and to drive new markets.

IEEE SA recently formed a metaverse standards committee to develop and maintain technical standards, create recommended practices, and write guides.

In addition, technical standards and activities are incubating new ideas on topics that are expected to be of great interest to industry.

The **IEEE P2048 Standard for Metaverse: Terminology, Definitions, and Taxonomy**, for example, is designed to define the vocabulary, categories, and levels of a metaverse to establish a common ground for ongoing discussions, facilitate the sustainable development of metaverse-related activities, and promote the healthy growth of metaverse markets.

The **IEEE P7016 Standard for Ethically Aligned Design and Operation of Metaverse Systems** will provide a high-level overview of the technosocial aspects of metaverse systems and specify an ethical assessment methodology for use in their design and operation. The standard will include guidance on how to adapt processes to prioritize ethically aligned design.

Two Industry Connections activities were launched specifically for the metaverse.

The Decentralized Metaverse Initiative aims to develop guidelines for implementing decentralized metaverses. The Persistent Computing for Metaverse Initiative will focus on the technologies needed to build, operate, and upgrade metaverse experiences. It includes computation, storage, communications, data structures, and artificial intelligence.

Join the Metaverse Community to help develop this new area.

Yu Yuan, an IEEE senior member, is president of IEEE SA. **Srikanth Chandrasekaran**, an IEEE senior member, is the practice lead and senior director of IEEE SA's Foundational Technologies Practice.

THE FOUR-DAY WORKWEEK IS GAINING TRACTION

A movement to offer workers a four-day 32-hour workweek for the same pay is catching on. A study on the subject conducted in the United Kingdom last year found that more than 90 percent of the 61 businesses that participated reported they will continue to offer a four-day week.

Benefits
- Less burnout
- Increased job satisfaction
- Better work/life balance
- Increased satisfaction with life

Disadvantages
- Increased fatigue
- Higher costs
- Additional customer support required
- Compressed schedule

IEEE poll results

The Institute conducted its own online poll in December, asking readers for their thoughts on a four-day workweek. About 95 percent of the respondents said they want to work a shorter week, however 89 percent of the respondents' companies don't offer that type of work schedule. Several respondents were interviewed:

"A shorter workweek would encourage more girls and women to join the engineering industry. It would also allow employees to spend quality time with their families, address care issues, and would be good for employees' mental health."
—IEEE MEMBER AMEL CHENOUF

"Employees could have a better work/life balance and therefore be more productive at work."
—DANTE MEDINA

"Staff members who take care of their parents or have children can choose the days and hours they work as long as they are available during core hours."
—IEEE LIFE SENIOR MEMBER WALTER D. DOWNING

"A more flexible work schedule would encourage younger engineers to join a company and stay for a longer period of time."
—IEEE SENIOR MEMBER JOHN MCWILLIAMS

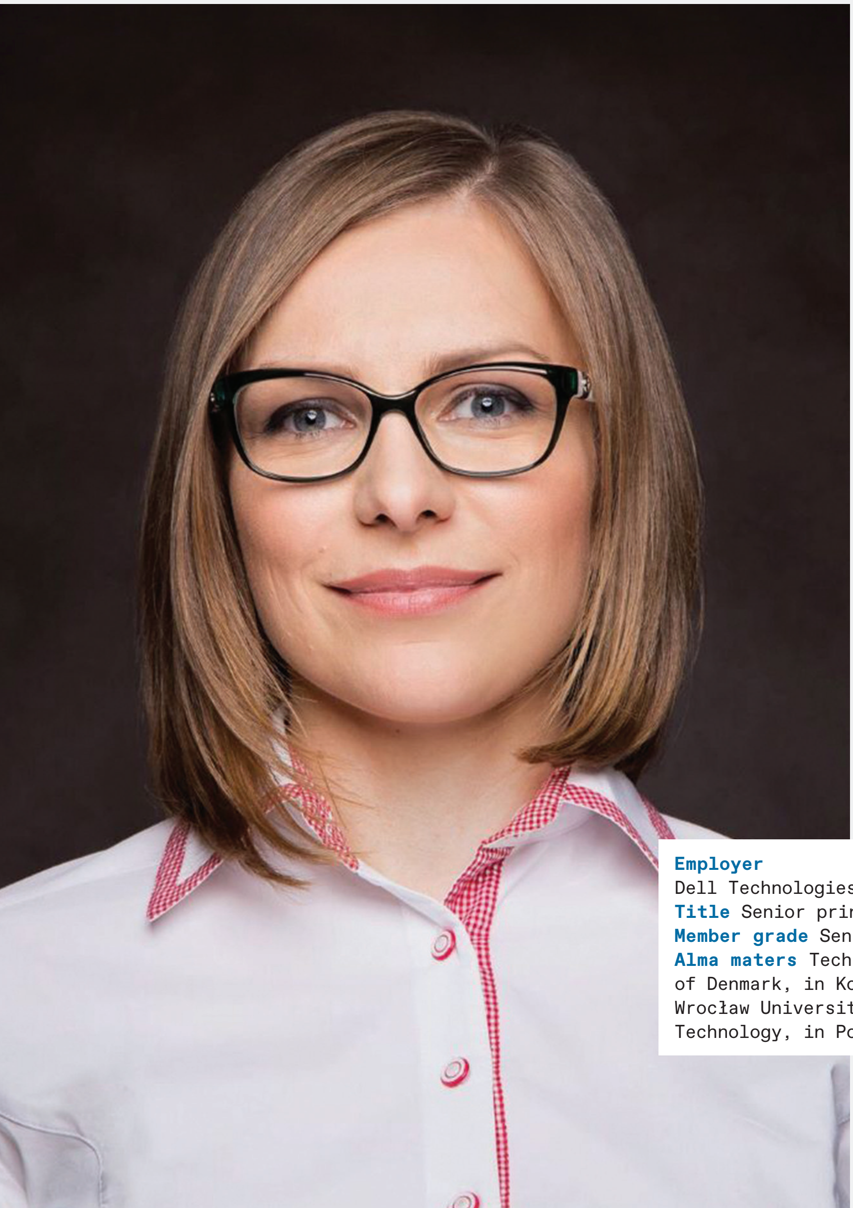
SOURCES: Euronews, Fortune, PBS

PROFILE

A Young Engineer on the Power of Mentorship

Anna Zakrzewska is guiding the next generation of technologists

BY JOANNA GOODRICH



Employer Dell Technologies in Dublin
Title Senior principal engineer
Member grade Senior member
Alma maters Technical University of Denmark, in Kongens Lyngby, and Wrocław University of Science and Technology, in Poland

THE LACK OF an engineering mentor while growing up didn't hinder Anna Zakrzewska from having a successful career in telecommunications. She is a senior principal engineer at Dell Technologies in Dublin. But Zakrzewska wonders how different her journey could have been if she'd had a mentor to guide her when she was younger.

"I often think about how my career would be different if I was exposed to STEM earlier in life," she says.

The IEEE senior member has made it her mission to introduce children to science, technology, engineering, and mathematics careers and to nurture their interest in STEM fields.

She volunteers for organizations that do just that, such as Technovation Girls, which focuses on getting more girls interested in STEM. The nonprofit holds an annual competition at which teams develop mobile apps that address real-world problems. Zakrzewska helps judge the entries.

She also provides guidance to university students and young professionals through outreach programs at Dell such as STEM Aspire.

For her work in STEM outreach and technical contributions to telecommunications research, Zakrzewska is the recipient of the 2023 IEEE Theodore W. Hissey Outstanding Young Professional Award. The honor is sponsored by IEEE Young Professionals and the IEEE Photonics and Power & Energy societies.

Mentoring students

When Zakrzewska was growing up in Poland, she and her family often listened to the radio. It was about the size of a microwave oven, she says, and she believed there were little people living inside whose job was to entertain and inform listeners.

Zakrzewska says when someone finally explained to her the engineering behind radios, it was an abstract concept, since she couldn't see how it worked. "To me engineering was magic," she says.

After graduating with a master's degree in telecommunications and computer science in 2008

MACIEJ BORKOWICZ

from the Wrocław University of Science and Technology, in Poland, Zakrzewska initially didn't know what type of career to pursue. She liked working on student research projects, she says, but couldn't see herself as a lecturer. Zakrzewska was also not sure if she wanted to be an industrial researcher.

She decided to give research a try and in 2009 became an intern at Nippon Telegraph and Telephone Corp., in Atsugi, Japan. She left after about a year to take a similar position at the European Commission's Joint Research Centre in Ispra, Italy. It conducts telecommunications research and helps the European Union set policy.

Finding that she enjoyed the work, Zakrzewska enrolled in 2010 in a doctoral program in wireless communications at the Technical University of Denmark, in Kongens Lyngby.

While there, she had her first experience of being a mentor to undergraduate engineering students as a member of the university's IEEE student branch.

"The Ph.D.-level members of the branch started a mentoring program because we realized we could help younger students with their studies or give them advice about internships and career paths," she says. "The branch also invited established professionals from a variety of electrical engineering fields to speak to students."

Participating in the program made Zakrzewska realize that many preuniversity schools didn't promote STEM careers, she says, and that mentoring young people about such careers when they're already university students is too late.

After earning her Ph.D. in 2013, she joined Nokia Bell Labs in Dublin as a research scientist working in wireless communications and self-organizing networks.

She also started to volunteer for STEM outreach organizations and programs. In 2015 she began managing the Nokia Bell Labs booth at the BT Young Scientist & Technology Exhibition, an annual Irish science fair. She told youngsters who visited the booth about her job and answered their questions.

Zakrzewska [top row, third from left] attends a 2023 STEM Aspire workshop at Dell's offices in Dublin with the program's mentees and other mentors.

DELL TECHNOLOGIES

She left Nokia Bell Labs in 2021 and joined Dell as a principal engineer at its Telecom Co-innovation Expert Centre. She helps develop technical standards for open radio access networks.

Increasing equity

Zakrzewska became active in 2017 with the annual Grace Hopper Celebration, a gathering of women technologists. The event is organized by AnitaB.org, a nonprofit that is working to increase intersectional equity (fairness with respect to both race and gender) in the global technical workforce. Zakrzewska is a committee member and is on the panel of reviewers for the scholarship program.

In 2020 she began volunteering for I'm a Scientist, a program designed to connect students with scientists and engineers. The mentors and students engage in online chats on the program's website, and mentees can ask their mentor questions about their job.

She also became active in 2020 with the annual Technovation Girls competition. Teams of up to five girls between the ages of 8 and 18 develop a technology that could address challenges in their communities. In addition to developing a mobile app, the teams present a business plan and a pitch video.

"I am extremely impressed by these projects," Zakrzewska says. "The students are able to identify the key problems, and many have ambitious plans to partner with the local authorities and seek sponsorship for their apps.

"Solving these problems not only provides a motivation to learn new skills in coding or teamwork but also empowers the students to lead the change in their local communities."

As a participant in Dell's mentoring programs, Zakrzewska visits preuniversity schools to explain to students how she became an engineer and what her job entails. She also provides career guidance to the company's interns and young professionals.

Zakrzewska's biggest piece of advice for her mentees, she says, is to create opportunities for themselves.

"If a door to a certain opportunity is closed, don't let that discourage you," she counsels. "Find a way to open the door or [find] another way in." ■

Building high-efficiency power plants can help lessen the impact of climate change, Rahman says. They include ultrasupercritical facilities such as the Eemshaven power station, in Groningen, Netherlands.

INGO WAGNER/PICTURE ALLIANCE/GETTY IMAGES

NEWS

6 Feasible Fixes for Climate Change

BY KATHY PRETZ

SIMPLE, EFFECTIVE SOLUTIONS that can help lessen the impact of climate change already exist. Some of them still need to be implemented, though, while others need to be improved.

That's according to 2023 IEEE President Saifur Rahman, who was among the speakers from engineering organizations at the COP27 event held in Egypt in November.

About one-third of emissions globally are produced through electricity generation, and Rahman said his mission is to help reduce that amount through engineering solutions.

"We have come to offer this six-point portfolio of solutions that everybody can live with," he said. "We want to be a solution partner so we can have parties at the table to help solve this problem of high carbon emissions globally."

Here are the solutions Rahman outlined:

1 **USE LESS ELECTRICITY.** Dimming lights by 20 percent in homes, office buildings, hotels, and schools could save 10 percent of electricity. Another is switching to LEDs, which use at least 75 percent less energy than incandescent bulbs. LED bulbs cost about five times more, but they last longer. Also, raise the temperature of air conditioners by 2 °C. This could save 10 percent of electricity as well.

2 **BUILD HIGH-EFFICIENCY POWER PLANTS.** One type is the ultrasupercritical coal-fired steam power plant. Conventional coal-fired plants, which make water boil to generate steam that activates a turbine, have an efficiency of about 38 percent. Ultrasupercritical plants operate at temperatures and pressures at which the liquid and gas phases of water coexist in equilibrium. Another efficient option is the combined cycle power plant. In its first stage, natural gas is burned in a turbine to make electricity. The heat from the turbine's exhaust is used to produce steam to turn a turbine in the second stage. The resulting two-stage power plant is at least 25 percent more efficient than a single-stage plant.

3 **INTEGRATE RENEWABLES.** Energy generation, transmission, and distribution systems should be included from the very start. The energy from wind, solar, and hydroelectric plants should be stored in batteries so the electricity generated from them during off-peak hours isn't wasted but integrated into energy grids.

4 **CONSIDER LOW-COST, LOW-CARBON HYDROGEN FUEL.** This should be part of the renewable energy mix. The fuel can be used to power cars, supply electricity, and heat homes, all with zero carbon emissions.

5 **REPLACE TRADITIONAL NUCLEAR POWER PLANTS.** Small modular reactors are easier and less expensive to build, and they're safer than today's large nuclear plants, Rahman said. Though small, SMRs are powerful. They have an output of up to 300 megawatts of electricity, or about a quarter of the size of today's typical nuclear plant. The modular reactors are assembled in factories and shipped to their ultimate location, instead of being built onsite. And unlike traditional nuclear facilities, SMRs don't need to be located near large bodies of water to handle the waste heat discharge.

6 **INITIATE MORE CROSS-BORDER POWER TRANSFERS.** Few countries have enough electricity for all their citizens. Many countries already transfer their power, Rahman said. "The United States buys power from Canada. France sells energy to Italy, Spain, and Switzerland," he said. "The whole world is one grid. You cannot transition from coal to solar and vice versa unless you transfer power back and forth."

Check out IEEE's collection of 7,000 papers related to climate change, which is accessible from the IEEE Xplore Digital Library. IEEE also launched climate-change.ieee.org, which houses additional resources. ■

IEEE STANDARDS

8 Products That Protect Kids' Digital Privacy

BY THE IEEE STANDARDS ASSOCIATION

IN THE IEEE Standards Association's continued effort to help organizations create a safe, secure, privacy-preserving digital environment for youngsters, it recently released the second edition of "Applied Case Studies for Designing Trustworthy Digital Experiences for Children." The report features the following eight products, which responsibly connect with children in age-appropriate ways, use minimal data collection, and offer improved ways of obtaining parental consent.

300M Learning Portal Futureshift Consulting's platform, TADAA!, provides engaging, age-appropriate content. The portal uses minimal data collection to protect children's identities.

Haru A companion robot from the Honda Research Institute, Haru offers interactive storytelling and activities, with minimal data collection. The robot, aimed at children ages 6 to 16, was built with their health, safety, and education in mind.

iTeach This community of secondary schools in India was established as a college-to-career path to help students with low socioeconomic status lift themselves out of poverty. ITeach provides the tablets, which are equipped with software to disable access to unsuitable content.

Kolibri The company provides access to educational content in a variety of languages and subject areas for preuniversity students living in areas that lack Internet access. The platform collects minimal data and runs on low-cost legacy devices.

Lego Group The group uses verifiable parental consent (VPC) to support children's rights and well-being on the Internet. A username generator eliminates the need for the child to create one that might accidentally include personal identifiers.

Otsimo The online app for special-needs children offers skill-level-appropriate developmental games focused on cognitive, social, and language skills, and it features speech exercises.

Yoti This digital identity platform helps organizations verify a user's age and comply with the U.K. age-appropriate design code. Children are asked to authenticate their age, which is then verified using facial age estimation and reference checks with mobile phone operators or credit agencies.

Privo The company's software and compliance solutions provide identity and consent management for parents through VPC. The program restricts youngsters' access at predetermined levels of services. ■

IEEE PRODUCTS

ERICSSON TECHNOLOGY REVIEW IS NOW IN IEEE XPLORE

BY KATHY PRETZ

The *Ericsson Technology Review* is now available in the IEEE Xplore Digital Library. The monthly magazine provides insights on emerging innovations that are shaping the future of information and communication technology.

The publication, which dates back to 1924, is published by Ericsson, a multinational networking and telecommunications company based in Stockholm.

ERICSSON

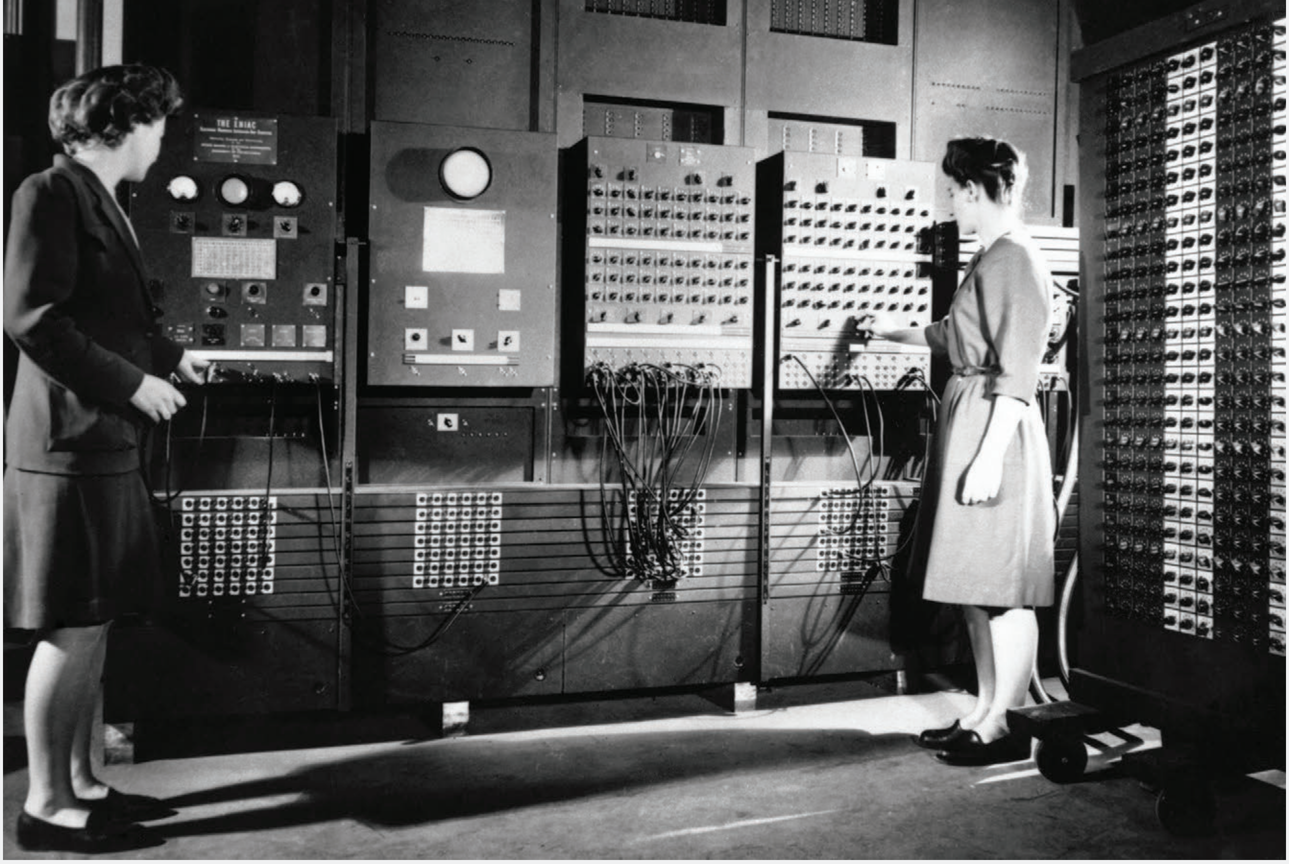
An IEEE Xplore subscription is not required to access the freely available research papers.

"IEEE is a respected organization, and Ericsson has the ambition to reach even further into the academic community and research institutes with our cutting-edge research and development," says Erik Ekudden, the company's chief technology officer. "We believe that IEEE Xplore is a good channel for this target group."

The *Review* in IEEE Xplore includes newly published articles plus those from the magazine's archives going back to 2020, according to Naveen Maddali, senior product manager of content partnerships for IEEE Global Products and Marketing. There are now more than 80 articles in the digital library.

The journal's articles, written by Ericsson's researchers, cover topics including communication, networking, and broadcast technologies; computing and processing; power and energy applications; robotics and control systems; and signal processing and analysis.

IEEE Xplore contains publications from other large companies in addition to Ericsson, including the *IBM Journal of Research and Development* and the *Bell Labs Technical Journal*. ■



Jean Bartik [left] and Frances Spence, two of the human computers who programmed the Electronic Numerical Integrator and Computer, preparing it for Demonstration Day in February 1946.

HISTORY

The Women Behind ENIAC

Author Kathy Kleiman on how they broke a computer-science glass ceiling

BY JOANNA GOODRICH

LOOKING AT THE pictures of those working on the first programmable, general-purpose all-electronic computer, you might assume that J. Presper Eckert and John W. Mauchly were the only ones who had a hand in its development.

Invented in 1945, the Electronic Numerical Integrator and Computer (ENIAC) was built to improve the accuracy of U.S. artillery during World War II. The two men and their team built the hardware. But hidden behind the scenes were six women—Jean Bartik, Kathleen Antonelli, Marlyn Meltzer, Betty Holberton, Frances Spence, and Ruth Teitelbaum—who programmed the computer to calculate artillery trajectories in seconds.

The U.S. Army recruited the women in 1942 to work as so-called human computers—mathematicians who did calculations using a mechanical desktop calculator.

For decades, the six women were largely unknown. But thanks to Kathy Kleiman, the world is getting to know the ENIAC programmers' contributions to computer science. Last year her book *Proving Ground: The Untold Story of the Six Women Who Programmed the World's First Modern Computer* was published. It delves into the women's lives and the pioneering work they did. The book follows the award-winning documentary, *The Computers: The Remarkable Story of the ENIAC Programmers*, which Kleiman helped produce. It premiered at the 2014 Seattle International Film Festival.

Kleiman gave a presentation in May about the programmers at the Computer History Museum in Mountain View, Calif., as part of the IEEE Industry Hub Initiative's Impact Speaker series. The initiative aims to introduce industry professionals and academics to IEEE and its offerings.

The Institute spoke with Kleiman, who teaches Internet technology and governance for lawyers at American University, in Washington, D.C., about her mission to publicize the programmers' contributions. The interview has been condensed and edited for clarity.

***The Institute*: What inspired you to film the documentary?**

KATHY KLEIMAN: The ENIAC was a secret project of the U.S. Army during World War II. It was the first general-

UNIVERSITY ARCHIVES AND RECORDS CENTER/UNIVERSITY OF PENNSYLVANIA; COVER:KATHY KLEIMAN

From left: Jean Bartik, Kathleen Antonelli, Betty Holberton, Ruth Teitelbaum, Marlyn Meltzer, Frances Spence.

purpose, programmable, all-electronic computer—the key to the development of our smartphones, laptops, and tablets today. The ENIAC was a highly experimental computer, with 18,000 vacuums, and some of the leading technologists at the time didn't think it would work, but it did.

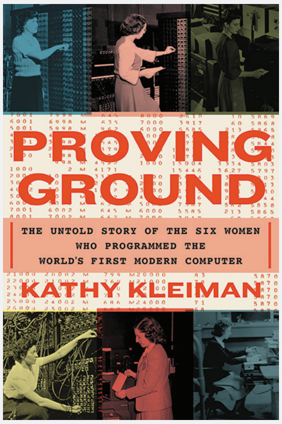
Six months after the war ended, the Army decided to reveal the existence of ENIAC and heavily publicize it. To do so, in February 1946 the Army took a lot of beautiful, formal photos of the computer and the team of engineers that developed it. I found these pictures while researching women in computer science as an undergraduate at Harvard.

At the time, I knew of only two women in computer science: Ada Lovelace and then U.S. Navy Capt. Grace Hopper. [Lovelace was the first computer programmer; Hopper helped develop COBOL, one of the earliest standardized computer languages.] But I was sure there were more women programmers throughout history, so I went looking for them and found the images taken of the ENIAC.

The pictures fascinated me because they had both men and women in them. Some of the photos had just women in front of the computer, but they weren't named in any of the photos' captions. I tracked them down after I found their identities, and four of six original ENIAC programmers responded. They were in their late 70s at the time, and over the course of many years they told me about their work during World War II and how they were recruited by the U.S. Army to be "human computers."

Eckert and Mauchly promised the U.S. Army that the ENIAC could calculate artillery trajectories in seconds rather than the hours it took to do the calculations by hand. But after they built the 2.5-meter-tall by 24-meter-long computer, they couldn't get it to work.

Out of approximately 100 human computers working for the U.S. Army during World War II, six women were chosen to write a program for the computer to run differential calculus equations. It was hard because the program was complex, memory was very limited, and the direct programming interface that connected the programmers to the ENIAC was hard to use. But the women succeeded. The trajectory program was a great success. But Bartik, Antonelli, Meltzer, Holberton, Spence, and Teitelbaum's contributions to the technology were never recognized. Leading technologists and the public never knew of their work.



I was inspired by their story and wanted to share it.

Why was the accomplishment of the six women important?

KLEIMAN: The ENIAC is considered by many to have launched the information age.

We generally think of women leaving the factory and farm jobs they held during World War II and giving them back to the men, but after ENIAC was completed, the six women continued to work for the U.S. Army. They helped world-class mathematicians program the ENIAC to complete problems that would take 100 years to solve by hand. They also helped teach the next generation of ENIAC programmers, and some went on to create the foundations of modern programming.

What influenced you to continue telling the ENIAC programmers' story in your book?

KLEIMAN: After my documentary premiered at the film festival, young women and men from tech companies who were in the audience came up to me to share why they were excited to learn the programmers' story. They were excited to learn that women were an integral part of the history of early computing programming, and were inspired by their stories.

I met more women and men like the ones in Seattle all over the world, so it seemed like a good idea to tell the full story along with its historical context and background information about the lives of the ENIAC programmers, specifically what happened to them after the computer was completed. ■

FROM LEFT: THE BARTIK FAMILY; BILL MAUCHLY, PRISCILLA HOLBERTON, TEITELBAUM FAMILY, MELTZER FAMILY, SPENCE FAMILY

NEWS

E-book Celebrates IEEE's Medal of Honor Recipients

BY JOANNA GOODRICH

FOR MORE THAN a century, IEEE has been honoring technology pioneers with its Medal of Honor. The organization's most prestigious award, it is given to engineers who have made exceptional contributions to or had an extraordinary career in electronics, electrical sciences, and engineering.

To celebrate the award's long history, IEEE recently released a commemorative e-book, *Over 100 Years of the IEEE Medal of Honor.* The volume chronicles the innovators who have received the award since its establishment in 1917. The Medal of Honor has been awarded annually since its establishment except in 1925, 1947, 1965, and 1976.

"The Medal of Honor is a living testament to the lives and careers of these major contributors in our modern world," K.J. Ray Liu, the 2022 IEEE president, wrote in the book's forward. "The recipients featured in this book are remarkable individuals who have made enormous contributions to science, technology, the engineering profession, and IEEE."

An award for radio engineers

Although professional medals of recognition and accomplishment for engineers and academic scientists were well established by the 1900s, there was no award specifically for electrical specialists working in communications.

That changed in February 1917, when the Institute of Radio Engineers (IRE), a predecessor society of IEEE and the leading technical and scientific society in the communications field, established the Medal of Honor. The award was designed to recognize prominent advances made in radiotelegraphy and radiotelephony.

Nominees for the award had to have made public their "greatest patented or unpatented advance in radio communication" within the previous two years, according to an entry about the honor on the Engineering and Technology History Wiki.

The innovation or technological advancement had to have been "publicly and completely described in a scientific or engineering journal of recognized standing." And the technology was required to be fully operational. Two years later the IRE eliminated the two-year time frame and the publication requirement.

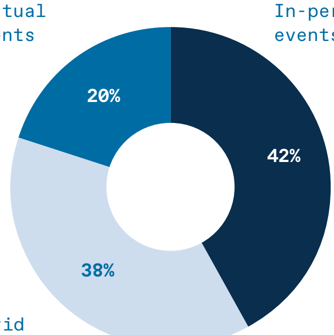
After the IRE merged in 1963 with the American Institute of Electrical Engineers, IEEE's other predecessor society, the combined organization broadened the Medal of Honor to include biomedicine, computing, transistors, and other technologies. ■

2022 Conferences by the Numbers

Here are statistics from last year.

2,008
Conferences held

534K
Total attendees



108
Countries that hosted events

64%
Held outside of the United States

205,370
Conference papers published in the IEEE Xplore Digital Library

SOURCE: IEEE Meetings, Conferences, and Events

IEEE

Rian Tiwari's invention uses AI to help pregnant people spot nutrient deficiencies by scanning their fingernails.

PROFILE

Teen's App Scans for Nutrient Deficiencies

It looks at your fingernails to suggest dietary changes

BY JOANNA GOODRICH

RIAN TIWARI

HIGH SCHOOL STUDENT Rian Tiwari has developed a mobile app that uses artificial intelligence to help pregnant people spot nutrient deficiencies by scanning their fingernails. Tiwari's app uses data from the scans to trigger diet and lifestyle recommendations, aiming to reduce the likelihood of a user developing anemia.

People with anemia have low levels of the healthy red blood cells needed to carry oxygen to body tissues. More than 50 percent of pregnant people become anemic, according to the Cleveland Clinic, risking premature birth, low birth weight for the baby, and postpartum depression. Preventing anemia can be as easy as eating iron-rich foods such as beans, red meat, and dried fruit.

In 2020, as a sophomore at South Brunswick High School in New Jersey, Tiwari was learning remotely during the coronavirus pandemic, and he found himself bored. His father suggested he think about *ikigai*, a Japanese concept referring to something that gives a person a sense of purpose. Tiwari decided that his purpose was to help others through technology.

He began researching several chronic conditions and zeroed in on anemia. He learned that a fingernail's appearance can give clues about a person's health. Discoloration, ripples, bumps, and other changes in a nail can be signs of nutrient deficiencies or disease. White spots, for example, might indicate a zinc deficiency. Brittle, cracking nails suggest low levels of folic acid.

Tiwari built an app that analyzes fingernail scans for signs of deficiencies in vitamin B12, calcium, zinc, and other nutrients. If these signs exist, the app recommends dietary and lifestyle changes, potentially preventing the development of anemia.

He presented his app at the 2022 IEEE International Conference on Intelligent Reality during an event hosted in collaboration with Amazon Web Services. The event focused on how AWS is helping startups that are developing health care technology.

"Presenting my work in front of respected and accomplished engineers felt unreal," Tiwari says. "I felt humbled by the opportunity to showcase my work and to learn from the other presenters."

From idea to app

In 2020 Tiwari and two of his classmates submitted a pitch and a business plan for their product to the Conrad Challenge, a competition for students developing technologies that address a global problem. Selected as finalists, they presented their idea virtually to the Conrad Challenge Innovation Summit at the Space Center Houston. Although they didn't win, Tiwari moved ahead with developing the app.

He needed help, however. He found Viswanatha Alluguntі, a solutions architect at Arohak—a software company in Monmouth Junction, N.J.—on LinkedIn and contacted him. Tiwari says he felt Alluguntі's experience in machine learning and business development would be beneficial.

"Having Dr. Alluguntі as a mentor really helped me push myself and this project to the next level," Tiwari says. "He taught me more about coding, how to file for a patent, and how to more effectively conduct research."

Alluguntі also helped him narrow down his target users from anyone with anemia to pregnant people, focusing on a group with a particular need for the technology, Tiwari says.

Tiwari filed for a U.S. patent for his invention in 2021. He was granted a patent in Germany last year.

Reading nails

Tiwari created the app's algorithms by running data sets obtained from the open-source website Kaggle through a machine-learning platform. The algorithm classifies images of nails based on their appearance. It looks for cracks, ridges, peeling, and discoloration.

The app starts with a photo of a nail taken by a user. It then uses a device-based neural network to analyze the image and classify the nail as healthy or unhealthy, Tiwari says.

If, for example, the app detects the person has a folate deficiency, it recommends foods such as asparagus, spinach, and sunflower seeds, he says.

The app stores some medical information as well as records of the analytics and recommendations from past scans.

To test the app, Tiwari worked with Reach, a nonprofit that provides mentorship and educational opportunities to people developing technology that enables individualized care for patients. The organization launched its own project—the Maternal Mortality Prevention Program—to ensure pregnant people's access to health monitoring.

Tiwari says he worked closely with the organization's president, Fran Ayalasomayajula, to learn what it takes to run a business. She helped him establish an advisory board and do patient outreach, he says.

He is working on getting the app to analyze images of lips and the inner eyelids as well—which can also show signs of nutrient deficiencies, he says. He also wants to add recommendations for medications and vitamin supplements.

Tiwari says he will be piloting the app this year. He plans to make it available for Android and iOS devices. ■

Inspired by IEEE

Tiwari says he was inspired to pursue a career in science, technology, engineering, and mathematics thanks to his father, Rajiv, who introduced him to IEEE's products, services, conferences, and publications. The IEEE senior member serves as cochair of the IEEE Future Directions Committee.

Tiwari says *IEEE Spectrum* and many other award-winning IEEE publications his father subscribes to have had a significant impact on him.

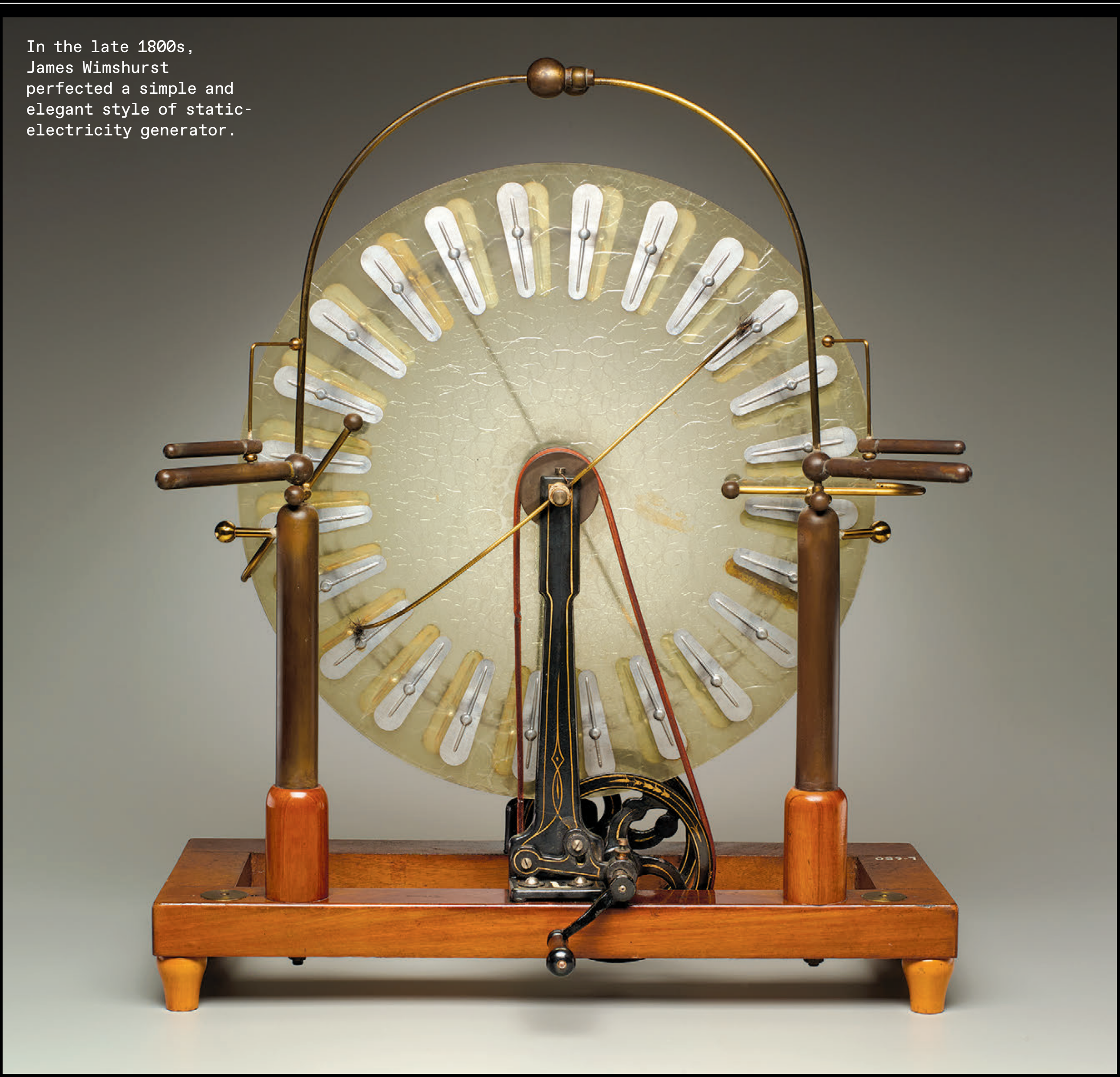
Reading articles published in *Spectrum* helped him discover that he wanted a career in technology, he says.

"So many of the cool technologies *Spectrum* writes about, such as quantum computing, sound like they are from a science fiction movie," he says. "I want to work on these types of technological advancements."

He says he's looking forward to becoming an IEEE student member and is excited about networking opportunities and attending conferences.

BY ALLISON MARSH

Past Forward



In the late 1800s, James Wimshurst perfected a simple and elegant style of static-electricity generator.

Wimshurst's Electrostatic Immortality

James Wimshurst did not invent the machine that bears his name. But thanks to his many refinements to a type of electrostatic generator in the late 19th century, we now have the Wimshurst influence machine. The machine has two insulated disks mounted on a single axle that rotate in opposite directions when driven by a hand crank. As the disks rotate, charges are induced on metal conducting plates mounted on the rims of the disks. Metal collector combs separate the charges into positive and negative and conduct them to two Leyden jar capacitors. The buildup eventually discharges with a spark, and the process begins again. A tabletop machine could generate up to 50,000 volts, which proved useful for experiments with X-rays. Well-to-do Victorians also delighted in the Wimshurst machine as an intriguing parlor trick. ■

FOR MORE ON THE HISTORY OF THE WIMSHURST MACHINE, see spectrum.ieee.org/pastforward-jun2023

YALE PEABODY MUSEUM