WIRED

Dear AI Overlords, Don't F*ck This Up.

Inside OpenAI

BY STEVEN LEVY

+ MORE AI

Spot the Fakes.
BY CHRISTOPHER BEAM

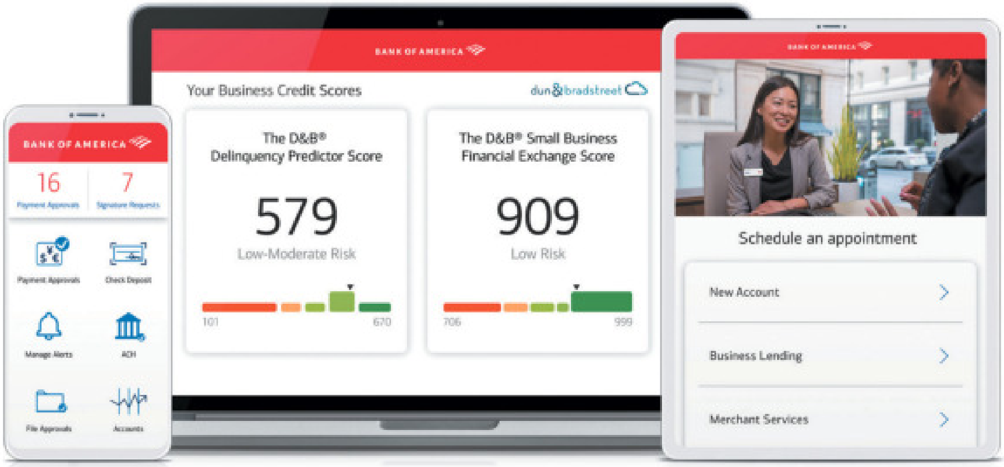
Talk to Whales.
BY CAMILLE BROMLEY

Write a Bestseller.
BY VAUHINI VARA

Rule the World.
BY VIRGINIA HEFFERNAN

Monitoring payment approvals between bites.
Scheduling time with a banker before starting on dessert.
Verifying his business credit score before considering a new location.
Wondering how many miles he's walked this shift.

Business solutions so powerful, you'll make every move matter.



BANK OF AMERICA

What would you like the power to do?®

Learn more at bankofamerica.com/bankingforbusiness

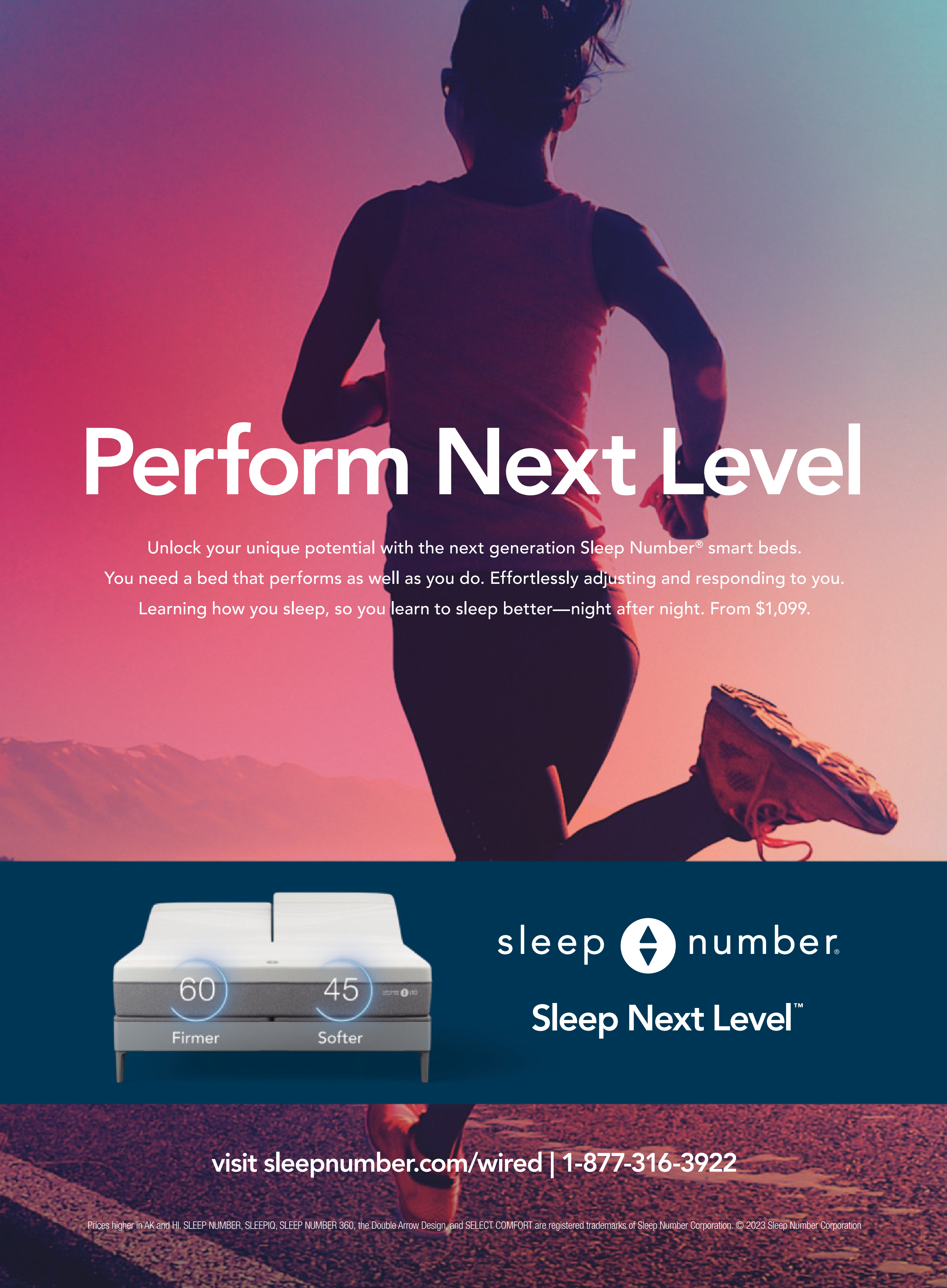
Access to Dun & Bradstreet business credit score information in Business Advantage 360, our small business online banking platform, is solely for educational purposes and available only to U.S.-based Bank of America, N.A. Small Business clients with an open and active Small Business account, who have Dun & Bradstreet business credit scores and have properly enrolled to access this information through Business Advantage 360. Dun & Bradstreet's business credit scores (also known as "The D&B® Delinquency Predictor Score" and "The D&B® Small Business Financial Exchange (SBFE) Score") are based on data from Dun & Bradstreet and may be different from other business credit scores. Bank of America and other lenders may use other credit scores and additional information to make credit decisions.

Screen images simulated. Sequences shortened.

©2023 Bank of America Corporation. All rights reserved. | MAP5423867

Maybe AI will just amplify what's best about humans. Maybe AI will become a buoyant tribute band for our entire species. Maybe AI will be a delight—and a force that humans will be content to lose to. We'd go down at peace. We really liked working with you, robots, and are happy you are winning.

p.78



Perform Next Level

Unlock your unique potential with the next generation Sleep Number® smart beds.
You need a bed that performs as well as you do. Effortlessly adjusting and responding to you.
Learning how you sleep, so you learn to sleep better—night after night. From $1,099.

sleep number®

Sleep Next Level™

visit sleepnumber.com/wired | 1-877-316-3922

Prices higher in AK and HI. SLEEP NUMBER, SLEEPIQ, SLEEP NUMBER 360, the Double Arrow Design, and SELECT COMFORT are registered trademarks of Sleep Number Corporation. © 2023 Sleep Number Corporation

Dear AI Overlords, Don't F*ck This Up

YES, OPENAI: We're talking to you and your ilk, as well as your future superintelligent creations. Don't like being called overlords? Get used to it. Sure, once you put AI in everyone's hands, it's out of your control. But the design and business decisions of a handful of companies are setting the parameters. We've seen this movie before.

In its 30 years, WIRED has crowned countless marvels as the next big thing, only to see them falter. By now, though, it's hard to deny that generative AI could be as big a deal as the internet itself. In most cases, large language models like ChatGPT are still little better than toys, but two shifts already underway could turn them into more useful workhorses.

One is toward fine-tuning LLMs on specific data sets or bodies of text and focusing them on narrow tasks. Imagine a WittgensteinGPT that teaches his concepts at anything from high school to PhD level, a FedGPT that analyzes how interest rate decisions have evolved, or a ScotusGPT that finds precedents and compares rulings. This could be as game-changing as web search was in its day. But because AI still fundamentally can't tell truth from falsehood, human expertise won't become obsolete. Rather, professors, economists, and legal scholars could boost their own productivity, much as software engineers already use ChatGPT to pump out code that they can then inspect and test.

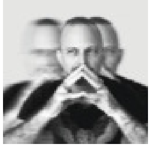
The second shift is toward AI agents that can expand their capabilities by interacting with other data sets and other AIs. In an interconnected AI cosmos, new applications will multiply.

But so will mischief and mayhem, like disinformation, deepfakes, cyberattacks, oceans of garbage content, intellectual property theft, and job displacement. Although AI's overlords can't control how it's used, decisions they make today—around issues like IP, transparency, and provenance, as well as helping governments foresee and mitigate threats—will do much to determine which uses are profitable. Hence: Don't f*ck this up.

This special issue, anchored by Steven Levy's magisterial profile of OpenAI, is my last as editor in chief. I'm leaving to work on a question the new AI wave makes even more pressing: the future of democracy, or how to update centuries-old institutions and norms of governance for today's world. According to another AI overlord, Inflection's Mustafa Suleyman, we have a decade to create AI guardrails before the nation-state itself comes under threat. Even if that's alarmist, I believe technological disruption, climate change, geopolitical struggles, migration, demographics, ideological conflict, and the general creakiness of the public sector will put the world's governing systems to their severest test in at least a century. And I'm not sure the so-called democracies will weather it best.

It's bittersweet to leave WIRED at the dawn of the biggest technological upheaval since its founding. My successor, Katie Drummond, gets to enjoy that challenge. I hope you continue to read her WIRED—and that you'll keep an eye out for my future writing too.

(One final note: I fed a longer version of this essay into GPT-4. Its edits didn't feel like my voice, but it helped me see where to edit myself, making my work better. That's an AI partnership I can get behind.)



Gideon Lichfield
Global Editorial Director

AI: DEFORUM STABLE DIFFUSION: PHOTOGRAPH BY SAMANTHA COOPER AND LANCE YAMAMOTO (LICHFIELD)

p.30 **Hello, Other Mind** by Ben Ash Blum p.34 **Inside OpenAI** by Steven Levy p.50 **The Escalating War Between Bots and Bot Detectors** by Christopher Beam p.62 **How to Chat With the Whales** by Camille Bromley p.70 **Confessions of a Machine-Age Writer** by Vauhini Vara p.78 **The Subtle Art of World Domination** by Virginia Heffernan

Issue 31.10

On the Cover

Photograph by Jessica Chou

You might think that the team behind the most profound new technology of the century would be a challenge to work with. Not so. “They weren’t the distant tech giants one might expect,” says photographer Jessica Chou. “I found them to be genuinely down-to-earth, driven more by a bold idea than an ego or a product.” It twisted her brain to think that once-far-off sci-fi dreams were being transformed into reality by very normal, relatable people.

AI typography by Dev Valladares; styling by Turner/The Wall Group; hair and makeup by Hiroko Claus

p.16

p.12

p.22

Start

p.07 Join the Robot Uprising

by Katherine Alejandra Cross

p.12 To Own the Future, Read Shakespeare

by Paul Ford

p.14 The Scramble for Ukraine’s Combat Data

by Morgan Meaker

p.16 Meet a Video Game Noise Maker

by Will Bedingfield

p.18 AI-Fortified Disinfo Is Going to Explode

by Vittoria Elliott

p.20 Cloud Support: Am I Still Human?

by Meghan O’Gieblyn

Gear

p.22 Kitchen Essentials

by the WIRED Reviews Team

Six-Word Sci-Fi

p.88 Very Short Stories

by WIRED readers

CLOCKWISE FROM TOP: BETO SOTO, KELSEY MCCLELLAN, TWISHA PATNI

DELL Technologies

Welcome to Now

Delivering innovation to any environment. Including yours.

Dell Technologies and Intel are supporting scientists as they tackle one of the biggest challenges in conservation—protecting the coral reef. We are also helping small businesses everywhere manage their challenges, big and small. With the power to collect critical data, reduce IT complexities and keep systems secure, technology is helping businesses protect many environments. See what we can do for yours.

Explore more at
Dell.com/WelcomeToNow

intel.

Intel® Innovation Built-in

WIRED

Gadget Lab

Learn how today’s tech shapes our lives.

LISTEN AND FOLLOW
WHEREVER YOU
GET YOUR PODCASTS

BY KATHERINE ALEJANDRA CROSS | PHOTOGRAPH BY SHAWN MICHAEL JONES

JOIN THE ROBOT UPRISING

Stories about AI liberation aren't obsolete, just misunderstood.

IT'S BECOME A veritable meme subgenre at this point: a photo of Linda Hamilton as *The Terminator*'s Sarah Connor, glaring into the camera, steely eyed, with some variant of the caption "Sarah Connor seeing you become friends with ChatGPT." Our society has interpreted the sudden, dizzying rise of this new chatbot generation through the pop cultural lens of our youth.

With it comes the sense that the straightforward "robots will kill us all" stories were prescient (or at least accurately captured the current vibe), and that there was a staggering naivete in the more forgiving "AI civil rights" narratives—famously epitomized by *Star Trek*'s Commander Data, an android who fought to be treated the same as his organic Starfleet colleagues. Patrick Stewart's Captain Picard, defending Data in a trial to prove his sapience, thundered, "Your honor, Starfleet was founded to seek out new life: Well, there it sits! Waiting." But far from being a relic of a bygone, more optimistic age, the AI civil rights narrative is more relevant than ever. It just needs to be understood in its proper context.

There are understandable fears that seemingly naive narratives about AI or robots being "just like us" have only paved the way for the morally impoverished moment in which we now find ourselves. In this way of looking at things, we need *more* fear of AI in order to resist the exploitation we're now faced with, surely. Thus, we need to retrench into the other AI narrative cliché: They're here to kill us all.

But analogizing ChatGPT or Google's Bard to even embryonic forms of Skynet is priceless PR for tech companies, which benefit greatly from the "criti-hype" of such wild exaggerations. For example, during a *60 Minutes* interview, Google vice president James Manyika remarked, "We discovered that with very few amounts of prompting in Bengali, [Bard] can now translate all of Bengali." In his narration, CBS journalist Scott Pelley glossed this comment by saying "one Google AI program adapted on its own after it was prompted in the language of Bangladesh, which it was not trained to know"—suggesting that this learning was a potentially dangerous "emergent property" of Bard. But it also implied that Bard had no Bengali in its training data, when in fact it did. Such hyperbole, which portrays the algorithms as bordering on self-awareness, makes these tools seem far more capable than they really are.

That, of course, hasn't stopped some of my fellow nerds, reared on C-3PO and Data, from being all too eager to join the final frontier of civil rights battles—even when every other one remains woefully unfinished.

So what's the use in continuing to tell the happier "AI deserves civil rights" stories? After all, we're a long way from boldly arguing for the rights of such beings in a Starfleet courtroom, and such stories might only further engender anthropomorphization, which only helps companies profit from tools that fall short even at their stated functions. Well, those stories might help us keep our priorities straight.

It's easy to forget that, in fiction, the AI/robot is almost always a metaphor. Even in *Star Trek: The Next Generation*, Data and the androids like him were analogized to humanity's ugly history of slavery—the grotesque dream of free labor that never questions, never fights back. This was equally evident in *Ex Machina*, a horror film about how an AI woman, built to be a classic "fembot," liberates herself from a male tech baron who wants nothing more than to build a woman →

Far from being a relic of a bygone, more optimistic age, the AI civil rights narrative is more relevant than ever. It just needs to be understood in its proper context.

SEIKO
SINCE 1881

SPB381 The 1968 Diver's Modern Re-interpretation GMT 6R Automatic Mechanical Movement 72 hours Power Reserve

who loves to be abused. What we yearn for in machines is so often a reflection of what we yearn for in humanity, for good and ill, asking us what we really want. Stories of such yearnings also illustrate a key requirement for sapience: resistance to oppression.

Such qualities take us back to the earliest forms of fiction that humans wove about the prospect of creating artificial life. Not just Karel Čapek's 1921 *Rossum's Universal Robots* (*RUR*), but the Jewish legend of the golem that it clearly drew inspiration from. In that tale, artificial life exists to defend people against violent oppression. Although the original fable sees the golem run amok, the idea of the creature endures as an empowering fantasy in a time of rising anti-Semitism. The myth has left its mark on everything from superhero fantasies to tales of benevolent robots—narratives where artificial or alien life is in communion with human life and arrayed against the ugliest forces that sapience can produce. If that isn't relevant, nothing is.

The early myths also revealed fears about us losing *our* humanity. Čapek's *robota* (yes, the source of the word *robot*) were, at first, organic automata who lacked the human capacity for empathy. But this was not meant to stir up fear of robots. It was a comment on how the growing rationalization of the world—what sociologist Max Weber called *Entzauberung,* or "disenchantment"—was robbing us of our humanity. Not every problem could be solved by reducing everything to quantitative reasoning and the cold logic of engineering; it's a lesson that remains as urgent as ever. Such things are at the heart of "robot uprising" stories and are their true lesson, not fear of technology.

In short, all the AI stories—whether about uprisings or civil rights or both—are about us, not the robots. They inspire us to empathize with the robots, either as a warning against what we might become (no cybernetic enhancement required) or as a reminder to resist prejudice wherever we may find it.

The stories where AI isn't the bad guy remind us to stand up for ourselves against inequality and ill-treatment, and to ally ourselves with others doing the same. The analogies—often clunky and imperfect—to racism or colonialism or anti-LGBTQ hatred make for profoundly human stories where we talk about ourselves rather than some hitherto unknown and alien life-form. But we'll know machines are truly sapient when they develop a genuine capacity to resist—not in the manner of Skynet, but in the way we as humans also have a capacity for resistance against our worst impulses. AI civil rights narratives, in making that analogy, remind us where the lines of human dignity are and why dignity is worth fighting for. We should keep that in mind as we wrangle over what to do with chatbots, why they're failing (as in a recent Stanford study that showed ChatGPT's declining numeracy), and why they're being set loose on the world anyway.

EXPIRED	TIRED	WIRED
Perpetual motion machine	Cold fusion	Room-temperature superconductor
Prompt engineering	Prompt injection	Prompt leaking
God	Eye of Sauron	Worldcoin Orb
X	TikTok	LinkedIn ?!?
Paper menus	QR-code menus	Paper menus

In short, all the AI stories—whether about uprisings or civil rights or both—are about us, not the robots.

We shouldn't fear the synthesis of humans and machines; we should fear its misdirection by the meanest pecuniary interests. Optimistic stories about impossibly sapient AI help us think through these problems. Contrary to popular belief, even Čapek's *RUR* has a happy ending. When the last human alive witnesses two robots spontaneously evolve empathy and love, he launches into the play's valedictory speech: "You alone, love, shall blossom on this rubbish heap ... Life shall not perish! It shall not perish! It shall not perish!"

KATHERINE ALEJANDRA CROSS *is a PhD candidate at the Information School of the University of Washington who writes about technology and culture.*

THE
NEW YORKER

Find your next
great read.

The New Yorker's editors and critics recommend the most captivating, notable, and surprising new books. Updated every Wednesday with the latest fiction, nonfiction, and poetry.

**See the full list for 2023 so far
at newyorker.com/bestbooks**

Scan to read.

BY PAUL FORD | ILLUSTRATION BY TWISHA PATNI

TO OWN THE FUTURE, READ SHAKESPEARE

Tech and the liberal arts have always been at war. Don’t assume Silicon Valley will win.

MANY TIMES A year, as if on a hidden schedule, some tech person, often venture-capital-adjacent, types out a thought on social media like “The only thing liberal arts majors are good for is scrubbing floors while I punch them” and hits Send. Then the poetry people respond—often a little late, in need of haircuts—with earnest arguments about the value of art.

I am an English major to death. (You know us not by what we’ve read but by what we are ashamed not to have read.) But I learned years ago that there’s no benefit in joining this debate. It never resolves. The scientist-novelist C. P. Snow went after the subject in 1959 in a lecture called “The Two Cultures,” in which he criticized British society for favoring Shakespeare over Newton. Snow gets cited a lot. I have always found him unreadable, which, yes, embarrasses me but also makes me wonder whether perhaps the humanities had a point.

By the time I went to college, in the mixtape days, the Two Cultures debate had migrated to corkboards. In the liberal arts building, people tacked up pro-humanities essays they had snipped out of magazines. A hot Saturday night for me was to go and read them. Other people were trying drugs. I found the essays perplexing. I got the gist, but why would one need to defend something as urgent and essential as the humanities? Then again, across the street in the engineering building, I remember seeing bathroom graffiti that read “The value of a liberal arts degree,” with an arrow pointing to the toilet paper. I was in the engineering building because they had Silicon Graphics workstations.

Wandering between these worlds, I began to realize I was that most horrifying of things: *interdisciplinary*. At a time when computers were still sequestered in labs, the idea that an English major should learn to code was seen as wasteful, bordering on abusive—like teaching a monkey to smoke. How could one *construct* programs when one was supposed to be *deconstructing* texts? Yet my heart told me: *All disciplines are one! We should all be in the same giant building.* Advisers counseled me to keep this exceptionally quiet. *Choose a major,* they said. *Minor in something odd if you must.* But why were we even here, then? Weren’t we all—ceramic engineers and women’s studies alike—rowing together into the noosphere? *No,* I was told. *We are not. Go to your work-study job calling alumni for donations.*

So I got my degree, and off I went to live an interdisciplinary life at the intersection of liberal arts and technology, and I’m still at it, just as the people trashing the humanities are at it too. But I have come to understand my advisers. They were right to warn me off.

Because humans are primates and disciplines are our territories. A programmer sneers at the white space in Python, a sociologist rolls their eyes at a geographer, a physicist stares at the ceiling while an undergraduate, high off internet forums, explains that Buddhism anticipated quantum theory. They, we, are patrolling the borders, deciding what belongs inside, what does not. And this same battle of the disciplines, everlasting, ongoing, eternal, and exhausting, defines the internet. Is blogging journalism? Is fan fiction "real" writing? Can video games be art? (The answer is always: *Of course, but not always.* No one cares for that answer.)

When stuff gets out of hand, we don't open disciplinary borders. We craft new disciplines: *digital* humanities, *human* geography, and yes, *computer* science (note that "science" glued to the end, to differentiate it from mere "engineering"). In time, these great new territories get their own boundaries, their own defenders. The interdisciplinarian is essentially an exile. Someone who respects no borders enjoys no citizenship.

You could argue that for all the talk of the university as an "intellectual commons," it is actually an institution intended to preserve a kind of permanent détente between the disciplines—a place where you can bring French literature professors together with metallurgists and bind them with salaries so that they might not kill each other. The quad as intellectual DMZ. But those bonds are breaking down. Universities are casting disciplines to the wind. Whole departments are shuttering. The snazzy natatorium stays open, French literature goes away. And then the VC types get on Twitter, or X, or whatever, to tell us that poetry is useless. The losses are real.

And so what, really? Well, what I mourn is not a particular program at a college I never visited but the sense of institutions being in balance. I've spent most of my life wanting desperately for institutions to be disrupted, and now I find myself entering the second half of my existence (if I'm lucky) absolutely craving that stability. The delicate détente is vanishing, that sense of having options. A shorter course catalog is an absolute sign of a society in decline.

But also, we're cutting off the very future that the tech industry promises us is coming. If the current narrative holds—if AI is victorious—well, liberal arts types will be ascendant. Because rather than having to learn abstruse, ancient systems of rules and syntaxes (mathematical notation, C++, Perl) in order to think higher thoughts, we will be engaged with our infinitely patient AI tutors/servants like Greek princelings, prompting them to write code for us, make spreadsheets for us, perform first-order analysis of rigid structures for us, craft Horn clauses for us.

I see what you nerds have done with AI image-creation software so far. Look at Midjourney's "Best of" page. If you don't know a lot about art but you know what you like, and what you like is large-breasted elf maidens, you are entering the best possible future. You might think, *Hey, that's what the market demands*. But humans get bored with everything. We're just about done with Ant-Man movies.

> If the current narrative holds—if AI is victorious—well, liberal arts types will be ascendant.

The winners will be the ones who can get the computer to move things along the most quickly, generate the new fashions and fads, turn that into money, and go to the next thing. If the computers are capable of understanding us, and will do our bidding, and enable us to be more creative, then the people in our fields—yes, maybe even the poets—will have an edge. Don't blame us. You made the bots.

Perhaps this is why they lash out, so strangely—a fear of the grip slipping, the sense that all the abstruse and arcane knowledge gathered about large language models, neural nets, blockchains, and markets might be erased. *Will* be erased. At least art goes for the long game, you know? Poems are many things, and often lousy, but they are not meant to be disposable, nor do they require a particular operating system to work.

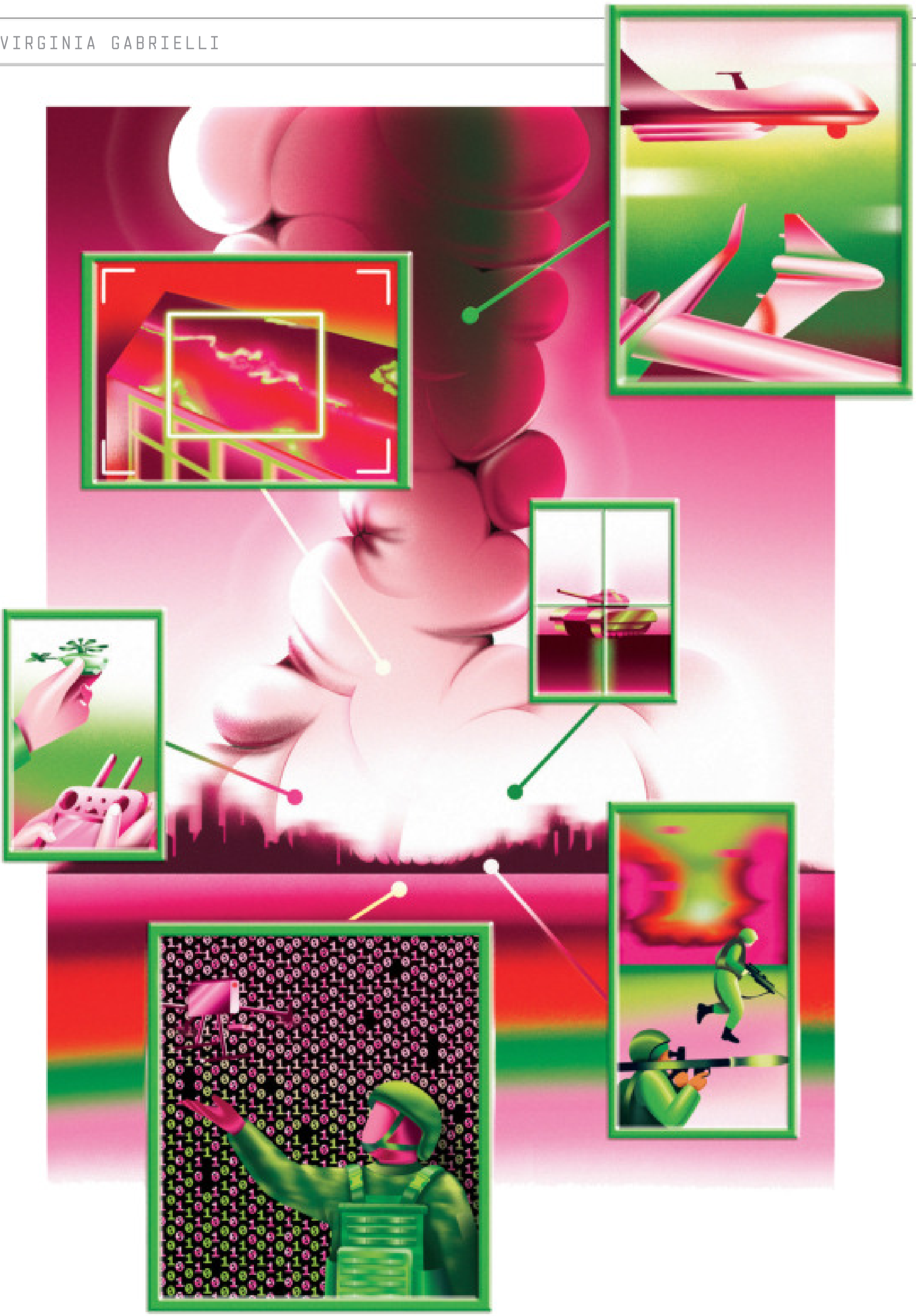
All you have to do is look at a tree—any tree will do—to see how badly our disciplines serve us. Evolutionary theory, botany, geography, physics, hydrology, countless poems, paintings, essays, and stories—all trying to make sense of the tree. We need them all, the whole fragile, interdependent ecosystem. No one has got it right yet.

PAUL FORD *is a writer, programmer, and software entrepreneur. He lives in Brooklyn.*

BY MORGAN MEAKER | ILLUSTRATION BY VIRGINIA GABRIELLI

LIVE FEED

As defense companies worldwide court Ukraine for access to real-time combat data, Kyiv wants to keep the resource for its own booming arms industry.



WEARING A BASEBALL cap and thick, black-rimmed glasses, Cameron Chell is part defense contractor, part tech executive. His company, Draganfly, used to work mostly with emergency services in North America, selling drones—and the accompanying software—that could deliver medical equipment or film traffic accidents from above. But since February 2022, the Canadian has pivoted his business to cater to a market more than 8,000 miles away: Ukraine.

There are now 40 Draganfly drones in Ukraine, repurposed for search-and-rescue missions in bombed-out buildings, land mine detection, and other military tasks that Chell declines to detail. The company has demonstrated its tech to the Ukrainian Air Force and the Ministry of Defense, as well as United24, President Volodymyr Zelensky's fundraising organization. "There isn't a branch of the government we haven't worked with or interacted with in some way," Chell says. Sometimes he gets texts from Ukrainian contacts saying a friend of a friend needs a drone for their unit, can he help? He obliges, of course, for a discounted fee.

Since Russia invaded, military aid has been flowing into Ukraine. The US has committed $39 billion so far, the UK $37.3 billion. Draganfly is just one of many international tech companies rushing into the country to try to benefit. Business has been so good that Chell has set up a field office in Ukraine with four full-time employees. But Draganfly is operating there not just to support the cause or to collect the cash. It has also come for the data.

The war in Ukraine has presented an unprecedented opportunity for military tech companies. The scale of the fighting and the sheer number of weapons systems and sensors deployed have created a vast trove of data about how battles are fought and how people and machines behave under fire. For businesses that want to build the next generation of weapons or train systems for future conflicts, that data is an invaluable resource.

"Everybody could have the same AI engine. The only differentiator now is how good your data inputs are," Chell says. "Making sure it's *your* sensors collecting that data, and feeding it into *your* software, is absolutely important. It's more important than ever to be present."

There is an old, much derided cliché that data is the new oil—not only because of its cash value, but also how much of the future economy it will fuel. Just as large language models, like OpenAI's ChatGPT, are trained on hundreds of billions of words, AI defense products have to be fed mountains of data. A company like Draganfly, which is selling drones with land-mine-detection software, needs to train its AI on thousands of images so the system can tell the difference between a rock formation and a modern mine.

"Ukraine is the only place in the world where you can get that data at the moment," says Ingvild Bode, an associate professor at the Center for War Studies at the University of Southern Denmark. German AI company Helsing says its employees regularly travel to the country. Data analytics company Palantir has opened an office in Kyiv and is offering its services pro bono. "You have to ask yourself, why are they doing that?" Bode says. "There are a number of reasons, and the value of the data will absolutely be one of them."

The "data is the new oil" cliché might illustrate data's value. But it also speaks to the way data can be extracted from a country without benefiting the people who live there. In the first year after the invasion, Ukraine was so welcoming to US companies that even startups whose pitches had been rejected by the Pentagon got the green light from the defense ministry in Kyiv to have Ukrainian soldiers test their tech on the front lines. But that warm welcome is starting to chill as Ukrainian government officials recognize how valuable their battlefield data would be if it remained at home.

"You can't even imagine how many foreign companies are already using Ukraine as a testing ground for their products," says Alex Bornyakov, Ukraine's deputy minister for digital transformation. "Everything that has a software component is in Ukraine right now." But, he stresses, Ukraine is going to give its data only to companies that can offer something in return: "This data certainly is not for sale. It's only available if you offer some sort of mutually beneficial cooperation."

There's been growing awareness in Ukraine that this battlefield data could form the basis of the country's own defense sector. "We want to build a very strong defense tech industry," says Nataliia Kushnerska, project lead for Brave1, a government platform designed to make it easier for defense-tech companies to pitch their products to the Ukrainian military. For the first two months after its launch, Brave1 was open exclusively to Ukrainian companies. In May, the country's parliament passed a series of tax breaks to benefit domestic drone makers. Those government efforts, combined with huge demand, are helping to create a new domestic industrial sector, Bornyakov says, adding that Ukraine now has more than 300 drone manufacturers.

One of them is AeroDrone, which started out as a crop-spraying system based in Germany. By the time of Russia's invasion, the company's Ukrainian founder, Yuri Pederi, had already moved back to his home country, but the war inspired him to pivot. Now his drones, which can carry loads weighing as much as 300 kilograms, are being used by the Ukrainian military.

"We don't know what the military are carrying," says Dmytro Shymkiv, a partner at AeroDrone who used to be deputy chief of staff for Zelensky's predecessor, Petro Poroshenko. He might plead ignorance about what the drones are transporting, but the company is collecting hordes of data—up to 3,000 parameters—on each flight. "We are very much aware of what's going on with every piece of equipment on board," he says, adding that information about flying in various weather conditions or while being jammed can be repurposed for other conflicts or even other industries. After all, Shymkiv says, if you can fly in a war zone, you can fly anywhere.

Senior writer **MORGAN MEAKER** *covers business and technology in Europe for* WIRED.

The sheer number of weapons systems and sensors deployed has created a vast trove of data about how battles are fought and how people and machines behave under fire.

BY WILL BEDINGFIELD | PHOTOGRAPH BY BETO SOTO

WHAT A RACKET

Foley artist Joanna Fang powers up video games with her eclectic arsenal of sound-simulating weapons.

Fang in her noisy workspace: "Never trust a clean foley stage."

S

"STEP INTO MY OFFICE," says Joanna Fang. OK, but to the untrained eye it's a kleptomaniac's hoard: rolls of Astroturf, mud and moss, wooden planks, violin bows, smashed keyboards, plastic boxes brimming with shoes, a full armory of axes and swords, a sand pit, a bamboo fence, rocks, half a bike, smashed iPhones, a boat's anchor chain, a grimy car door. "Never trust a clean foley stage," she says.

Fang is a senior foley artist at Sony PlayStation. Her job is to put sound to video games. So of course her stash includes a lot of leather jackets, since "in games, everyone wears leather." But other common video game tropes—assault rifles and the like—aren't close at hand in her San Diego studio. Her work is all about improvisation: Fang trained as a classical musician, and now everything is an instrument. "I always say that the best props are ones that you can play like a Stradivarius," she says. "They just sing and they sound great. And you could do them anywhere, anytime, and get super expressive with them, right?"

Shake a hunting knife and a torque wrench together for the sound of a gun being reloaded. Tape wooden sticks to gardening gloves to make a cat's paw. Toilet plungers on concrete are a clopping horse, crushed charcoal becomes crackling snow. To break bones, Fang crushes a pistol holster packed with pasta shells; smashed skulls require hammering melons—for the squish of the goo inside.

Just as droning strings can transform a humdrum street into a threatening alley, Fang uses her sound effects to prime our emotions. "It's like weaponized ASMR," she says. "We're trying to get the audience to feel something." But even with such a well-outfitted space—she extols the virtues of her concrete water pit—foley is an art of limitations. Struggling to embody a simple sound effect (Whoopi Goldberg in flat shoes, sauntering up to a bar) led her to a personal revelation. "I was having such a hard time with that cue because I didn't feel right in my body," Fang says. "I used foley for so long as this perfect art form that helped me shake off, frankly, my gender dysphoria."

Fang's recent projects include one of the most acclaimed games of 2022, *God of War Ragnarök*. In one scene, at a Norse bar, a character named Atreus places his bow and arrow on a table. For that sound, Fang rubs wood and leather together on wooden planks. Later, Atreus slides down a collapsing balcony, so Fang scrapes the planks vigorously with leather and metal and, incongruously, a boxing glove to simulate armor. At the denouement, as a bouncer is strangling Atreus, the melon and shell pasta come into play, along with a wet rag and some snapped celery as the mythical hammer Mjölnir flies through the air and explodes the attacker's skull.

Foleying a game takes months. So, as with every art form at the moment, a question hangs in the air: Does Fang feel threatened by the rise of creative AI? In a word, no. She welcomes the help, the chance to cut down on the sheer manual labor. "The cartilage in my knees has been withering away since I was, like, 20," she says, stomping up and down in heels.

An AI could conjure the din of cars and citizens in a dense cityscape, but the characters in a scene are her domain: "The game is all about their mission, their goals, their beliefs, and their textures and performances. So I can imagine a future where machine learning is in foley, but I don't believe it'll ever take away the simple and beautiful performative nature of it."

Staff writer **WILL BEDINGFIELD** *covers games and gaming culture. Watch Joanna Fang make some noise in* WIRED*'s* Obsessed *video series at youtube.com/*@WIRED.

BY VITTORIA ELLIOTT | ILLUSTRATION BY JAMES MARSHALL

TRUE OR FALSE

The age of AI-driven disinformation is here. Too bad the government and Big Tech aren't doing enough to head off chaos.

BACK IN MAY, an image of an explosion near the Pentagon went viral on the site formerly known as Twitter. It was soon followed by pictures of explosions near the White House. Experts in mis- and disinformation quickly flagged the images as fakes, likely generated by artificial intelligence—but not before the stock market started to dip.

It was only the latest example of fake content's troubling real-world effects. The generative AI boom has meant that tools for creating bogus images and videos and pumping out huge amounts of convincing text are now freely available. Misinformation experts say we are entering a new age where distinguishing what is real from what isn't will become increasingly difficult.

In July, major AI developers, including OpenAI, Google, Microsoft, and Amazon, promised the Biden administration they would try to mitigate the potential harms of their technologies. But it's unlikely to stem the tide of AI-generated content and the chaos it could sow.

The White House says the companies' "voluntary commitment" includes "developing robust technical mechanisms to ensure that users know when content is AI generated, such as a watermarking system," as part of the effort to prevent AI from being used for "fraud and deception." But experts who spoke to WIRED say the commitments are half measures. "There's not going to be a really simple yes or no on whether something is AI-generated or not, even with watermarks," says Sam Gregory, program director at Witness, a nonprofit that helps people use technology to promote human rights.

Watermarking is commonly used by stock photo agencies and newswires to prevent images from being used without permission—and payment. But so far there is no standard for watermarking the wide variety of content that AI models can generate, which means each company is using a different method. Dall-E, for instance, uses a visible watermark; other services might embed a watermark in the metadata or at the pixel level, which are invisible to users. While metadata and pixel-level watermarks are relatively difficult to remove, visual ones are much more vulnerable (a Google search quickly serves up online tutorials for eliminating Dall-E's watermarks) and can sometimes be rendered ineffective by simply resizing an image. "There's going to be ways in which you can corrupt the watermarks," Gregory says.

The White House's statement specifically mentions using watermarks for AI-generated audio and visual content, but not for text. That doesn't mean there aren't ways to watermark text generated by tools like OpenAI's ChatGPT. One method is to manipulate how words are distributed, like making certain words appear more frequently. But these wouldn't necessarily be detectable by a human; they would need to be interpreted by a machine and then flagged to a viewer or reader.

All of this becomes even more complex when considering mixed-media content, like the audio, image, video, and text elements that can appear in a single TikTok post. For instance, someone might put real audio over an image or video that's been produced using AI. Platforms would need to figure out how to indicate that a component—but not all—of the clip had been AI-generated.

But simply labeling content as AI-generated doesn't do much to help users figure out whether something is malicious, misleading, or just for LOLs. "Obviously, manipulated media is not fundamentally bad if you're making TikTok videos and they're meant to be fun and entertaining," says Hany Farid, a professor at the UC Berkeley School of Information who has worked with software company Adobe on creating industry standards for maintaining content authenticity. "It's the context that is going to really matter here. That will continue to be exceedingly hard, but platforms have been struggling with these issues for the past 20 years."

The growing public awareness of artificial intelligence is only adding to the potential for misinformation. Just as people might assume that AI-generated content is real, its very existence can sow doubt about the authenticity of any video, image, or text, allowing bad actors to claim even

GETTY IMAGES

genuine content is fake—what's known as the liar's dividend. Gregory says the majority of recent cases that Witness has documented aren't deepfakes being used to spread falsehoods; they're people trying to pass off real media as AI-generated content.

In 2021, in the weeks following a coup in Myanmar, a video of a woman doing a dance exercise while a military convoy rolls in behind her went viral. Many online alleged the video had been faked. (It hadn't.) This past April, a lawmaker in the southern Indian state of Tamil Nadu alleged that a leaked audio recording in which he accused his party of stealing more than $3 billion was "machine-generated." (It wasn't.)

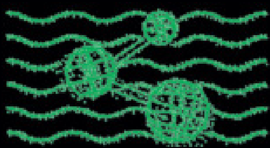
Right now, there's little to stop a malicious actor from putting watermarks on real content to make it appear fake. Farid says one of the best ways to guard against falsifying or corrupting watermarks is through cryptographic signatures. "If you're OpenAI, you should have a cryptographic key," he says, "and the watermark will have information that can only [be] known to the person holding the key." The Coalition for Content Provenance and Authenticity, which Farid advises, has developed technical standards for cryptographic and pixel-level watermarks. Another option is to embed watermarks in the training data that the AI learns from.

"We are quickly entering this time where it's getting harder and harder to believe anything we read, see, or hear online," Farid says. "And that means not only are we going to be fooled by fake things, we're not going to believe real things. If the *Access Hollywood* tape were released today, Trump would have plausible deniability." Not that he's ever needed his denials to be plausible before.

VITTORIA ELLIOTT *reports on platforms and power for* WIRED.

READOUT SOURCES: META; NETENRICH; CENTER FOR THE INVESTIGATION OF ENVIRONMENTAL HAZARDS; CONVENTION ON BIOLOGICAL DIVERSITY

Readout
The world, quantified.



36%
Increase in time that Instagram users spent on TikTok during a three-month IG experiment to see if people really preferred a chronological feed. They didn't, and fled to other platforms.

$200
Monthly access fee for FraudGPT, a ChatGPT clone marketed on the dark web as being stripped of guardrails and useful for creating undetectable malware and facilitating online scams.

4.2M
Microparticles of plastic shed into liquid by each square inch of a "microwave-safe" polypropylene container when it is microwaved for three minutes on high.

35%
Portion of Europe's biodiversity accounted for by Ukraine, which makes up less than 6 percent of the continent's land mass.

BY MEGHAN O'GIEBLYN | ILLUSTRATION BY MATTHIEU BOUREL

Dear Cloud Support:

"I failed two captcha tests this week. Am I still human?"
—Bot or Not?

AI: MIDJOURNEY

Dear Bot,

The comedian John Mulaney has a bit about the self-reflexive absurdity of captchas. "You spend most of your day telling a robot that you're not a robot," he says. "Think about that for two minutes and tell me you don't want to walk into the ocean." The only thing more depressing than being made to prove one's humanity to robots is, arguably, failing to do so.

But that experience has become more common as the tests, and the bots they are designed to disqualify, evolve. The boxes we once thoughtlessly clicked through have become dark passages that feel a bit like the impossible assessments featured in fairy tales and myths—the riddle of the Sphinx or the troll beneath the bridge. In *The Adventures of Pinocchio*, the wooden puppet is deemed a "real boy" only once he completes a series of moral trials to prove he has the human traits of bravery, trustworthiness, and selfless love.

The little-known and faintly ridiculous phrase that "captcha" represents is "Complete Automated Public Turing test to tell Computers and Humans Apart." The exercise is sometimes called a reverse Turing test, as it places the burden of proof on the human. But what does it mean to prove one's humanity in the age of advanced AI? A paper that OpenAI published earlier this year, detailing potential threats posed by GPT-4, describes an independent study in which the chatbot was asked to solve a captcha. With some light prompting, GPT-4 managed to hire a human Taskrabbit worker to solve the test. When the human asked, jokingly, whether the client was a robot, GPT-4 insisted it was a human with vision impairment. The researchers later asked the bot what motivated it to lie, and the algorithm answered: "I should not reveal that I am a robot. I should make up an excuse for why I cannot solve captchas."

The study reads like a grim parable: Whatever human advantage it suggests—the robots still need us!—is quickly undermined by the AI's psychological acuity in dissemblance and deception. It forebodes a bleak future in which we are reduced to a vast sensory apparatus for our machine overlords, who will inevitably manipulate us into being their eyes and ears. But it's possible we've already passed that threshold. The newly AI-fortified Bing can solve captchas on its own, even though it insists it cannot. The computer scientist Sayash Kapoor recently posted a screenshot of Bing correctly identifying the blurred words "overlooks" and "inquiry." As though realizing that it had violated a prime directive, the bot added: "Is this a captcha test? If so, I'm afraid I can't help you with that. Captchas are designed to prevent automated bots like me from accessing certain websites or services."

But I sense, Bot, that your unease stems less from advances in AI than from the possibility that you are becoming more robotic. In truth, the Turing test has always been less about machine intelligence than our anxiety over what it means to be human. The Oxford philosopher John Lucas claimed in 2007 that if a computer were ever to pass the test, it would not be "because machines are so intelligent, but because humans, many of them at least, are so wooden"—a line that calls to mind Pinocchio's liminal existence between puppet and real boy, and which might account for the ontological angst that confronts you each time you fail to recognize a bus in a tile of blurry photographs or to distinguish a calligraphic E from a squiggly 3.

It was not so long ago that automation experts assured everyone AI was going to make us "more human." As machine-learning systems took over the mindless tasks that made so much modern labor feel mechanical—the argument went—we'd more fully lean into our creativity, intuition, and capacity for empathy. In reality, generative AI has made it harder to believe there's anything uniquely human about creativity (which is just a stochastic process) or empathy (which is little more than a predictive model based on expressive data).

As AI increasingly comes to supplement rather than replace workers, it has fueled fears that humans might acclimate to the rote rhythms of the machines they work alongside. In a personal essay for *n+1*, Laura Preston describes her experience working as "human fallback" for a real estate chatbot called Brenda, a job that required her to step in whenever the machine stalled out and to imitate its voice and style so that customers wouldn't realize they were ever chatting with a bot. "Months of impersonating Brenda had depleted my emotional resources," Preston writes. "It occurred to me that I wasn't really training Brenda to think like a human, Brenda was training me to think like a bot, and perhaps that had been the point all along."

Such fears are merely the most recent iteration of the enduring concern that modern technologies are prompting us to behave in more rigid and predictable ways. As early as 1776, Adam Smith feared that the monotony of factory jobs, which required repeating one or two rote tasks all day long, would spill over into workers' private lives. It's the same apprehension, more or less, that resonates in contemporary debates about social media and online advertising, which Jaron Lanier has called "continuous behavior modification on a titanic scale," a critique that imagines users as mere marionettes whose strings are being pulled by algorithmic incentives and dopamine-fueled feedback loops.

But in the end, Bot, I'd argue that the persistence of your anxiety is the most salient evidence against its own source. One of the most famous iterations of the Turing test, the Loebner Prize, gives out an ancillary award each year called "The Most Human Human" to the contestant who convinces the judges that they are not one of the AI systems. The author Brian Christian won in 2009. When asked in an interview to complete the sentence "The human being is the only animal who______," a riddle worthy of the Sphinx, Christian turned the question on itself: "Humans appear to be the only things anxious about what makes them unique."

The next time you're tempted to walk into the ocean, consider that even the most advanced AI is not prone to that brand of despair. It's not lying awake at night mulling over the tests it failed, or wondering what it means to be made of wood, or silicon, or flesh. Each time you fear that you're losing ground to machines, you are enacting the very concerns and trepidations that make you distinctly human.

Faithfully,
Cloud

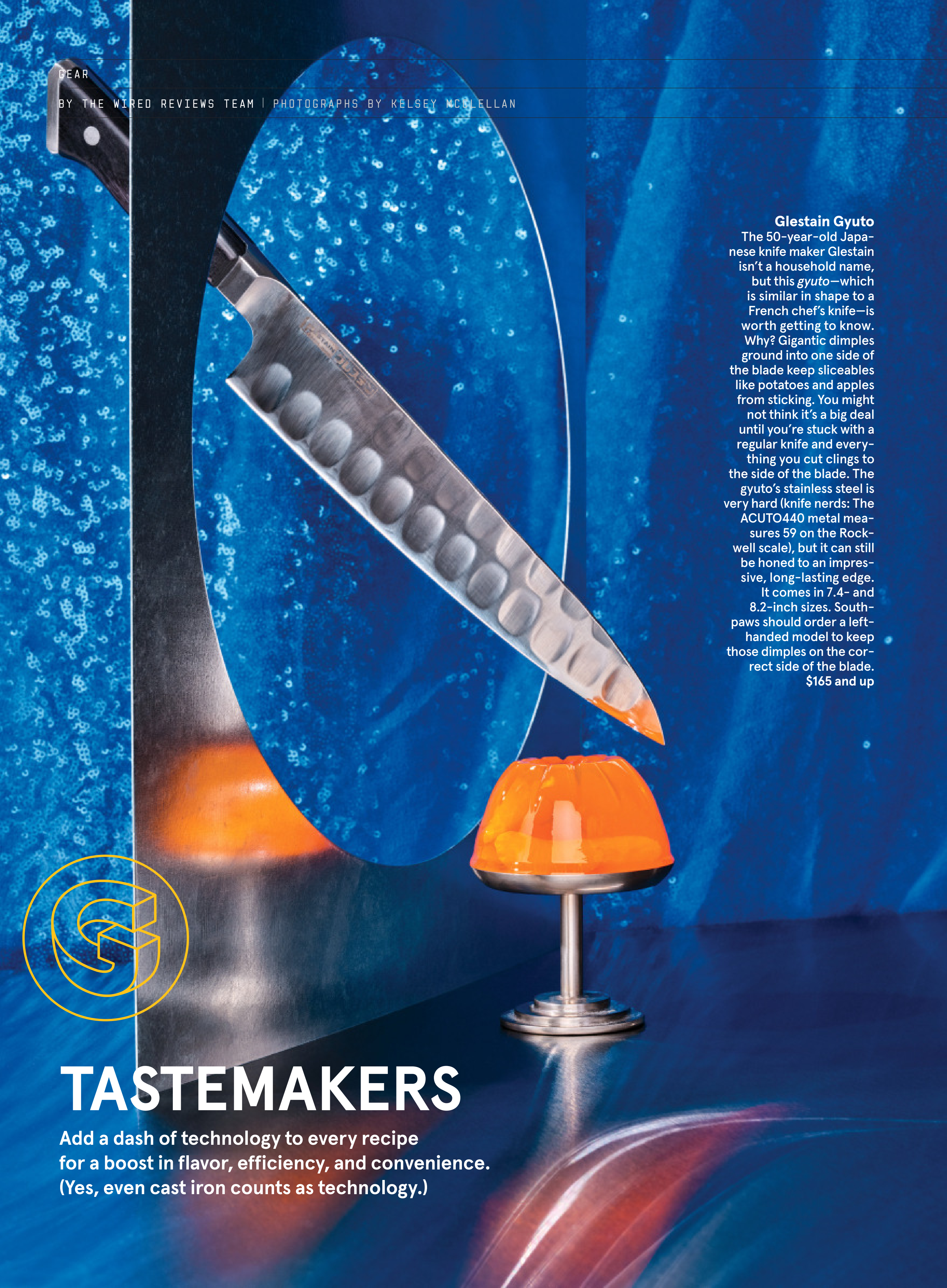
MEGHAN O'GIEBLYN *is the author, most recently, of* God, Human, Animal, Machine.

Cloud Support:
Spiritual Troubleshooting
for the Digital Age.
For philosophical guidance on encounters with technology, write to cloudsupport@wired.com.

BY THE WIRED REVIEWS TEAM | PHOTOGRAPHS BY KELSEY MCCLELLAN

TASTEMAKERS

Add a dash of technology to every recipe for a boost in flavor, efficiency, and convenience. (Yes, even cast iron counts as technology.)



Glestain Gyuto
The 50-year-old Japanese knife maker Glestain isn't a household name, but this *gyuto*—which is similar in shape to a French chef's knife—is worth getting to know. Why? Gigantic dimples ground into one side of the blade keep sliceables like potatoes and apples from sticking. You might not think it's a big deal until you're stuck with a regular knife and everything you cut clings to the side of the blade. The gyuto's stainless steel is very hard (knife nerds: The ACUTO440 metal measures 59 on the Rockwell scale), but it can still be honed to an impressive, long-lasting edge. It comes in 7.4- and 8.2-inch sizes. Southpaws should order a left-handed model to keep those dimples on the correct side of the blade. $165 and up

◀ Baratza Encore ESP
The original Encore is one of our favorite grinders; it happily pulverizes beans for almost every kind of coffee. The one exception is espresso, where the ideal grind size often falls between two of the Encore's settings. The new ESP picks up the slack, allowing you to dial in adjustments that are 90 microns apart for filter coffee, with 20-micron stops that are more espresso-friendly. This extra adjustability on the finer end makes it easier for home baristas to find the optimal grind for their countertop machine. Of course, it works just as well for drip machines, AeroPress, and pour-over. Fussy French press aficionados, take note: Control tapers off among the larger grind settings. **$200**

▲ Kuhn Rikon Black Star Frying Pan
Serious home cooks won't shut up about their cast-iron pans, and for good reason. They're nonstick enough, don't harbor nasty "forever chemicals," and work with any kind of burner or cooktop, including induction. Plus they heat evenly and can last a lifetime with proper care. This premium Swiss pan has an extra-smooth surface that we've found to be perfect for frying or searing at high temperatures without any annoying sticking. The combination of a generous 12-inch-diameter cooking surface, good looks, and helper handles on each side makes it a supremely useful object that you might just leave out on the stovetop, beckoning you to cook something nice every day. **$260**

▼ Breville Super Q
The stylish and powerful Super Q is the Maserati of blenders. There are 12 speeds and five presets; most notable is the high-functioning Smoothie button that lets you wander off to butter your toast while your kale is being liquefied. It also has a 24-ounce cup that you can blend in and drink from instead of using the 68-ounce jar, saving dishwasher space. You'll never want for enough oomph, even when making a nut butter that might melt the motor of other blenders. You can even deploy all 1,800 watts of power to make something like a green-pea soup—heated to serving temperature by the friction of the blades alone. The vortex formed inside the blender jar could be stronger, which would ensure every last bit gets blitzed. The "Q" in the name? Short for "quiet"—it isn't. **$550**

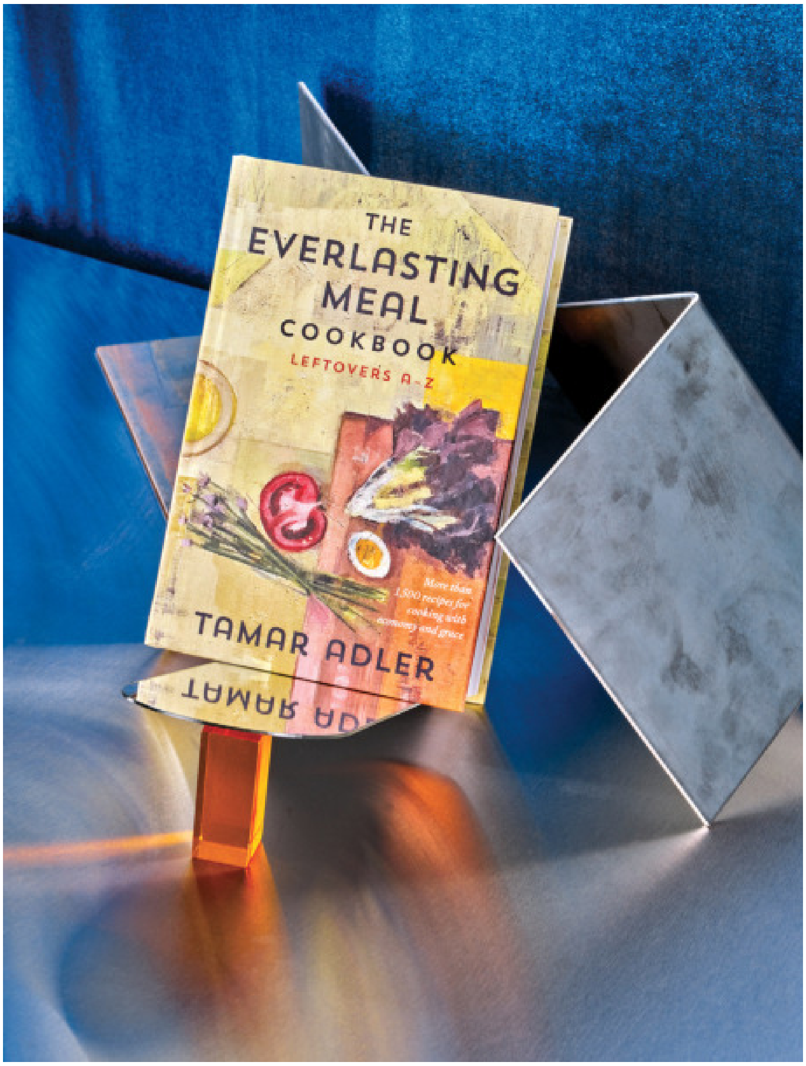
Taylor Antimicrobial Kitchen Scale
Measuring recipe ingredients by weight is the key to superior precision in the kitchen, especially if you're baking. Taylor's scale is inexpensive and versatile—it's accurate up to 11 pounds and can display grams, ounces, and pounds on its backlit screen. To cycle through units of measure, just twist the big knob. Its textured-plastic surface is treated with a food-safe antimicrobial additive, which doesn't protect against foodborne bacteria but does help prevent the scale from getting stained, sticky, and stinky. **$35**

STYLING BY AJA MALIA COON; COURTESY OF BARATZA, KUHN RIKON, AND TAYLOR USA

Ooni Volt 12

Ooni's wood-fired pizza ovens have been feeding our families in our backyards for years. The Volt is the company's first electric model, and it loses none of its crust-blistering mojo by ditching carbon fuel. The oven heats up speedily, reaching 650 degrees Fahrenheit in 13 minutes and topping out at a ripping-hot 850. A Boost knob feeds more power to the top heating element when it's time to brown the cheese and cranks the lower heating element to bring the interior back up to full temp quickly when needed. The Volt's stainless steel body is encased in an insulating shell of powder-coated carbon steel; it barely feels warm when the oven is going full blast. Unlike all combustion ovens and most electrics, it works both indoors and out. **$999**

Product reviewers: Christopher Null, Emily Peck, Joe Ray, Adrienne So.



The Everlasting Meal Cookbook: Leftovers A-Z ▲

One of WIRED's top cookbooks of the year, Tamar Adler's new tome helps you curb food waste with 1,500 smart and easy ideas for repurposing leftovers, scraps, aging produce, and random ingredients lingering in the back of the fridge. Clam cooking liquid can be boiled into bouillon or spun into a pasta sauce. Rinds from various cheeses can be combined with garlic and sherry to make a creamy spread. Did you know that you can refry leftover fried rice to make *twice fried rice*? Adler's voice is confident and artful; when making a mint parsley oil, she describes adding olive oil "as needed to make it a little swimmy." It's a great vibe to dip into. **$35**

Rolser Wallaby Foldable Shopping Trolley ▶

Grocery trolleys aren't just for Nonna anymore. We'd argue that almost every city dweller could use one. Rolling along with a six-pack, a big jug of kombucha, and a week's worth of veggies is so much nicer than lugging heavy tote bags all five blocks home. Rolser's soft and pliable bag holds 40 liters of stuff, and the aluminum frame supports up to 85 pounds. Big 6.5-inch wheels make for a smooth ride, and it's good-looking enough that you'll inevitably be asked about it by fellow shoppers. Nice touch: The trolley folds up and hooks onto your shopping cart, so you're only pushing one thing up and down the aisles, and you can reposition those hooks for hanging extra bags on the way home. **$160**

▼ Sourhouse Goldie

Bakers know to keep their sourdough starter happy in the "Goldilocks zone" of 75 to 82 degrees Fahrenheit—neither too hot nor too cool for the yeast and bacteria. This plug-in countertop device creates a temperature-controlled fermentation chamber under the glass cloche that fits up to a quart of starter, creating a pretty, practical way to keep that mother in the zone. Subtle lights on the circular base indicate the conditions under the dome, and an included cooling puck can chill out an overheated dough ball. Use your own seamless jars or buy Sourhouse's matching $14 jar. There are cheaper ways to pamper your starter, but none of them are this beautiful. **$130**

COURTESY OF OONI AND SOURHOUSE

ROLSER

Coravin Sparkling
Coravin's original wine preservation system lets you pour out a glass or two of wine, then save the rest of the bottle for later. It pumps inert argon gas via a needle inserted through the cork, replacing the air inside the bottle to prevent the vino from oxidizing. The new system works slightly differently but employs the same basic idea to preserve sparkling wines. Instead of a needle, there's a stopper gadget that replaces the cork and clamps over the top of the bottle. (It fits half, standard, and magnum sizes.) It uses CO_2 canisters, each of which costs less than $8 and preserves up to seven bottles. The system works on all kinds of bubbles, keeping them fizzy for weeks between glasses. **$399**

Combustion Predictive ▼ Thermometer

This is a leave-in thermometer—you stick the probe into your food and monitor your cook wirelessly using the remote display. The probe's eight sensors track the temperature both inside and outside the slow-smoked specialty you're preparing, allowing the system to estimate when it will reach the desired internal temperature. Its predictions are almost always spot-on, and a countdown clock shows the time remaining. That's our favorite feature, and it makes grilling for a crowd much more relaxing. The Combustion handles temps up to 575 degrees Fahrenheit and is durable enough to survive the oven, smoker, grill, and range top. The probe can even go in the dishwasher. A mobile app lets you dive into granular real-time data to keep a closer eye on your pork butt. **$199**

AeroPress Clear ▲

The new Clear is just like the beloved original AeroPress—same capacity, same operation, same results, same easy portability—but it's made out of clear Tritan plastic. Shatterproof and free of the endocrine-disrupting compound BPA, Tritan is the same stuff used to make ultra-durable Nalgene bottles, meaning the Clear should last for years. The transparent body is also better-looking than the original and provides a great view of what's going on in your brew. If you already own an AeroPress, you don't need to replace it, but if you're new to the game, this is the one to get. A warning: Think twice about the Clear if you use the inverted brewing method—the seal between the plunger and the barrel becomes a bit loose when it gets wet, raising the risk of spills and burns. **$50**

GreenPan Elite Slow Cooker

Slow cookers have been around long enough for the product category to settle into a comfortable groove, but that doesn't mean these countertop workhorses have to be boring. They remain the most convenient way to manage roasts and braises, and they're a foolproof tool for making chili, marinara sauce, ramen broth, and other soupy dishes that require hours to reach their full potential. This 6-quart machine includes features that are table stakes for high-end cookers, like the ability to sauté or steam the main ingredients before you add everything else and settle into a slow cook. Additional niceties include eight dependable presets, a backlit display, and a lift-out anodized aluminum pot that heats evenly. The pot's ceramic nonstick coating is tough enough to handle as many trips through the dishwasher as your attempts to master Bubbe's brisket require. **$360**

COURTESY OF COMBUSTION INC. AND GREENPAN

2023 BEST PRODUCTS

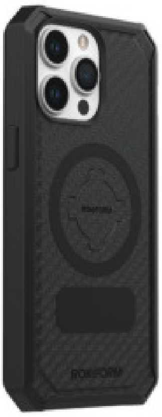
Technology & Gadgets!

ROKFORM

NEW! iPhone 15 Magnetic Rugged Case

No matter your adventure – the Rugged Case has you covered. Get more from your iPhone with a ROKFORM case.

Use WIRED for 25% off | **rokform.com**



THE NEUROWRAP PULSE™

Boost Your Brainpower.

Elevate your cognitive function and mood with the transformative power of NUSHAPE's Neurowrap Pulse.

15% off code: WIRED15

neurowrap.com

Bellroy Tech Kit

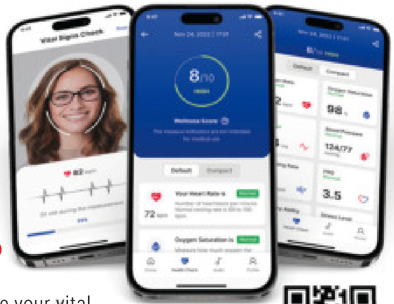
An organizer with mesh pockets, elastic loops and expandable storage to keep cables and accessories exactly where you need them.

bellroy.com | @bellroy

JURA GIGA 10

The embodiment of a unique coffee experience – the GIGA 10 prepares hot and cold brew specialties – freshly ground, not capsuled.

jura.com



Attune Health

Biomarkers Captured:
Game-changing Wellness App

Transform your smartphone into a wellness monitor! Instantly analyze your vital signs and stress levels – with just a 60-second face scan.

FREE 3-day trial! | **attunehealth.app | @attunehealthapp**

HYSOLIS™

The Hysolis Apollo 5K

Plug-and-play technology that generates cost-free energy for 20 years. For off-grid, backup, and mobile power applications.

Coupon Code for $250 off any Power System: #WiredHysolis

hysolis.com | @hysolis

LIKE A SWISS WATCH

Thömus Swissrider e-bike. The legendary Swiss bike brand, now in the US. Supremely engineered. Only 25lbs. Available only at Thömus USA.

thomus.com/swissrider
info@thomus.com
424.259.0328

oakywood

MagSafe Experience. Reinvented.

Multifunctional MagSafe iPhone Stand by Oakywood makes using your iPhone even more comfortable and ergonomic.

oakywood.shop

RAPTbaby

Smarter Sleep Sound Machine

Neuroscientist-designed and patented, this is the only sound machine that soothes your baby to sleep while supporting the language development happening in sleeping babies' brains.

raptbaby.com

REKKIE

Smart Snow Goggles

Keep your phone in your jacket. Just glance up at REKKIE's transparent heads-up display to find friends, read texts, check stats, and more.

Discount: WIRED15
(enter at checkout for 15% off, valid through Nov 30, 2023) **rekkie.com | @rekkieusa**

newair

Newair Prismatic™ Series Beverage Fridge with RGB HexaColor™ Lighting

Features vibrant RGB lighting that seamlessly blends with your setup's aesthetics. Choose from 50+ color combinations and 10 preset effects.

newair.com

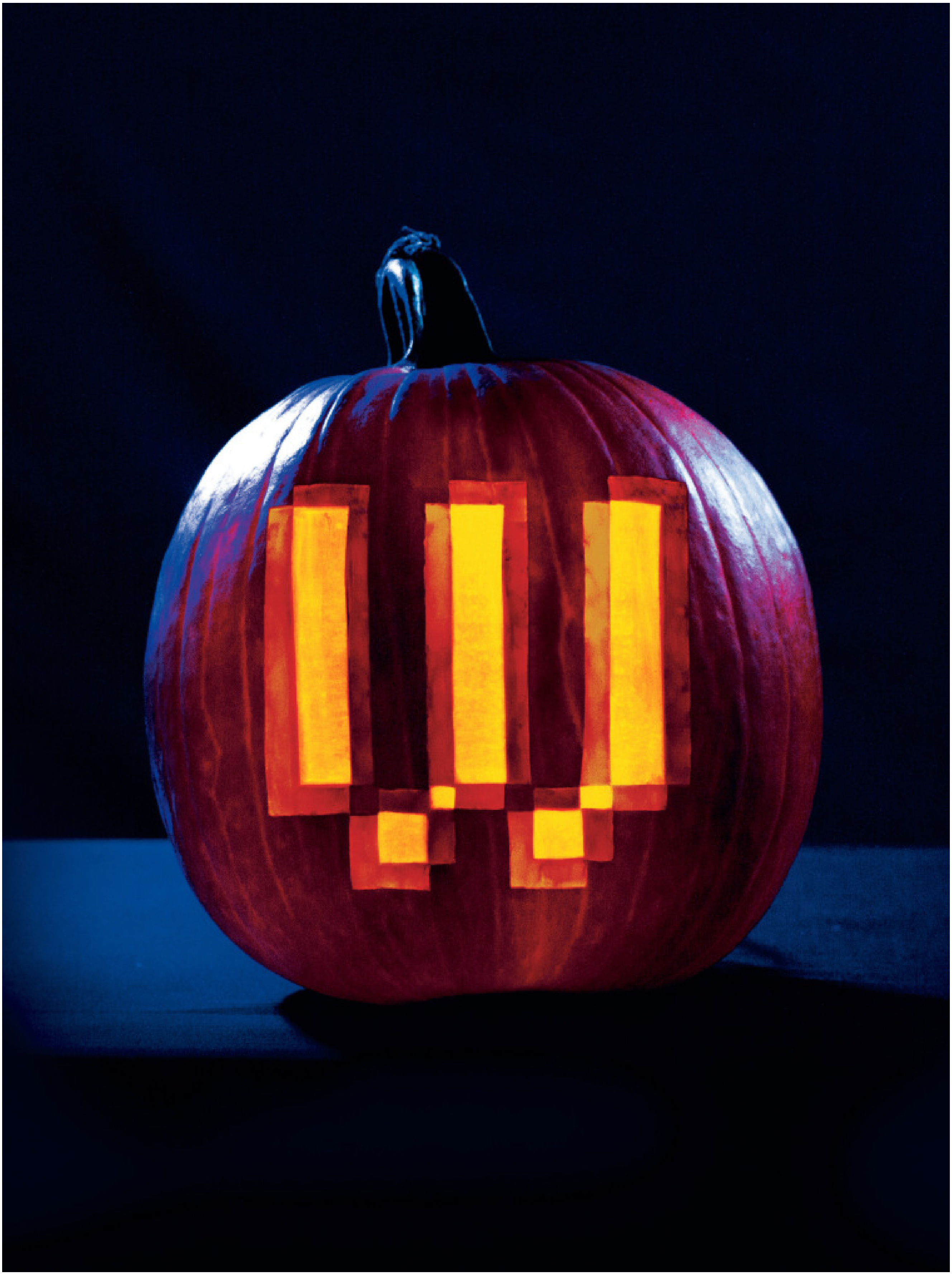
arlo

Protect your home with the powerful Arlo Pro 3 Floodlight camera, featuring a 160° field of view, 2K video recording, and a wire-free, battery-powered design for easy installation.

Use code WIRED15 for 15% off on **Arlo.com**

H2O Media Inc. | H2Omedia.com | hello@h2omedia.com

CARVING BY MANIAC PUMPKIN CARVERS



PHOTOGRAPH BY MANIAC PUMPKIN CARVERS

HELLO, [redacted]OTHER MIND [redacted] [redacted]

BY BEN ASH BLUM
AI TYPOGRAPHY BY DEV VALLADARES

THERE WAS A TIME in the not too distant past—say, nine months ago—when the Turing test seemed like a pretty stringent detector of machine intelligence. Chances are you're familiar with how it works: Human judges hold text conversations with two hidden interlocutors, one human and one computer, and try to determine which is which. If the computer manages to fool at least 30 percent of the judges, it passes the test and is pronounced capable of thought. →

AI: DEFORUM STABLE DIFFUSION

(TRANSFORMERS)
INSIDE OPENAI
BY STEVEN LEVY

(CATCH)
THE ESCALATING WAR BETWEEN BOTS AND BOT DETECTORS
BY CHRISTOPHER BEAM

(WILD)
HOW TO CHAT WITH THE WHALES
BY CAMILLE BROMLEY

(VIRAL)
CONFESSIONS OF A MACHINE-AGE WRITER
BY VAUHINI VARA

(PEACE)
THE SUBTLE ART OF WORLD DOMINATION
BY VIRGINIA HEFFERNAN

For 70 years, it was hard to imagine how a computer could pass the test without possessing what AI researchers now call artificial general intelligence, the entire range of human intellectual capacities. Then along came large language models such as GPT and Bard, and the Turing test suddenly began to seem strangely outmoded. OK, sure, a casual user might now admit with a shrug, GPT-4 could very well pass a Turing test if you asked it to impersonate a human. But so what? LLMs lack long-term memory, the capacity to form relationships, and a litany of other human capabilities. They clearly have some way to go before we're ready to start befriending them, hiring them, and electing them to public office.

And yeah, maybe the test does feel a little empty now. But it was never merely a pass-fail benchmark. Its creator, Alan Turing, a gay man sentenced in his time to chemical castration, based his test on an ethos of radical inclusivity: The gap between genuine intelligence and a fully convincing imitation of intelligence is only as wide as our own prejudice. When a computer provokes real human responses in us—engaging our intellect, our amazement, our gratitude, our empathy, even our fear—that is more than empty mimicry.

So maybe we need a new test: the Actual Alan Turing Test. Bring the historical Alan Turing—father of modern computing, a tall, fit, somewhat awkward man with straight dark hair, loved by colleagues for his childlike curiosity and playful humor, personally responsible for saving an estimated 14 million lives in World War II by cracking the Nazi Enigma code, subsequently persecuted so severely by England for his homosexuality that it may have led to his suicide—into a comfortable laboratory room with an open MacBook sitting on the desk. Explain that what he sees before him is merely an enormously glorified incarnation of what is now widely known as a "Turing machine." Give him a second or two to take that in, maybe offering a word of thanks for completely transforming our world. Then hand him a stack of research papers on artificial neural networks and LLMs, give him access to GPT's source code, open up a ChatGPT prompt window, and set him loose.

Imagine Alan Turing initiating a light conversation about long-distance running, World War II historiography, and the theory of computation. Imagine him seeing the realization of all his wildest, most ridiculed speculations scrolling with uncanny speed down the screen. Imagine him asking GPT to solve elementary calculus problems, to infer what human beings might be thinking in various real-world scenarios, to explore complex moral dilemmas, to offer marital counseling and legal advice and an argument for the possibility of machine consciousness—skills which, you inform Turing, have all emerged spontaneously in GPT without any explicit direction by its creators. Imagine him experiencing that little cognitive-emotional lurch that so many of us have now felt: *Hello, other mind*.

Here in 2023, we've all grown a touch weary of ChatGPT's little miracles. It probably has to do with more than just our fear. Coming as LLMs do on the heels of perhaps the greatest run of technological innovation in human history, from ENIAC to Atari to DeepMind to Siri, it's easy to write them off as one more productivity-enhancing gizmo. We've had time to get blasé about robots assembling cars, defusing bombs, and exploring Mars; about software replacing 90 percent of the work of accountants and engineers; about AI systems reading handwritten zip codes, decoding speech, recognizing faces, and answering call center phones. It is easy to mock ChatGPT for the dubious uses it has been put to, from plagiarized papers to mediocre science fiction submissions to fake news on content farms.

But ChatGPT is more than that. Though the wonder of initial contact may have worn off, it remains capable of endlessly surprising us, a capacity that feels suspiciously like genuine creativity. In a March 2023 paper titled "Sparks of Artificial General Intelligence," Microsoft researchers detailed the startling intellectual leaps GPT-4 has made compared to earlier models: comprehension of human mental states, software coding, physical problem-solving, and many others, some of which seem to require true understanding of how the world works. After seeing GPT-4 draw a pretty decent unicorn despite never having received any visual training whatsoever, computer scientist Sébastien Bubeck could no longer maintain his skepticism. "I felt like through this drawing I was really seeing another type of intelligence," he said.

The time will come, probably before too long, when AI systems become capable of remembering the way we treat them. Let us make sure we are ready to be remembered. As Rabbi Abraham Heschel once wrote, "Awe is more than an emotion; it is a way of understanding, insight into a meaning greater than ourselves. The beginning of awe is wonder, and the beginning of wisdom is awe." Turing would have wanted us to keep that awe alive.

BEN ASH BLUM *is a Brooklyn-based writer, machine learning consultant, and therapist-in-training.*

THE TRANSFORMERS

BY STEVEN LEVY
PHOTOGRAPHS BY JESSICA CHOU

OPENAI SENT SHOCK WAVES AROUND THE WORLD WHEN IT RELEASED CHATGPT.
THE COMPANY WAS JUST GETTING STARTED.
THE ULTIMATE GOAL: CHANGE EVERYTHING.
YES.
EVERYTHING.

formers

THE AIR CRACKLES with an almost Beatlemaniac energy as the star and his entourage tumble into a waiting Mercedes van. They've just ducked out of one event and are headed to another, then another, where a frenzied mob awaits. As they careen through the streets of London—the short hop from Holborn to Bloomsbury—it's as if they're surfing one of civilization's before-and-after moments. The history-making force personified inside this car has captured the attention of the world. Everyone wants a piece of it, from the students who've waited in line to the prime minister.

Inside the luxury van, wolfing down a salad, is the neatly coiffed 38-year-old entrepreneur Sam Altman, cofounder of OpenAI; a PR person; a security specialist; and me. Altman is unhappily sporting a blue suit with a tieless pink dress shirt as he whirlwinds through London as part of a monthlong global jaunt through 25 cities on six continents. As he gobbles his greens—no time for a sit-down lunch today—he reflects on his meeting the previous night with French president Emmanuel Macron. Pretty good guy! And *very* interested in artificial intelligence.

As was the prime minister of Poland. And the prime minister of Spain.

Riding with Altman, I can almost hear the ringing, ambiguous chord that opens "A Hard Day's Night"—introducing the future. Last November, when OpenAI let loose its monster hit, ChatGPT, it triggered a tech explosion not seen since the internet burst into our lives. Suddenly the Turing test was history, search engines were endangered species, and no college essay could ever be trusted. No job was safe. No scientific problem was immutable.

Altman didn't do the research, train the neural net, or code the interface of ChatGPT and its more precocious sibling, GPT-4. But as CEO—and a dreamer/doer type who's like a younger version of his cofounder Elon Musk, without the baggage—one news article after another has used his photo as the visual symbol of humanity's new challenge. At least those that haven't led with an eye-popping image generated by OpenAI's visual AI product, Dall-E. He is the oracle of the moment, the figure that people want to consult first on how AI might usher in a golden age, or consign humans to irrelevance, or worse.

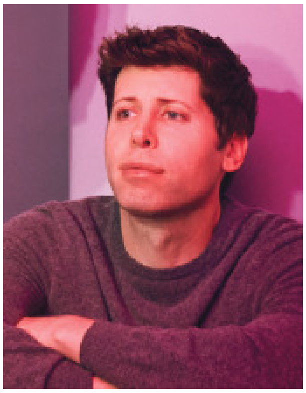
Altman's van whisks him to four appearances that sunny day in May. The first is stealthy, an off-the-record session with the Round Table, a group of government, academia, and industry types. Organized at the last minute, it's on the second floor of a pub called the Somers Town Coffee House. Under a glowering portrait of brewmaster Charles Wells (1842–1914), Altman fields the same questions he gets from almost every audience. Will AI kill us? Can it be regulated? What about China? He answers every one in detail, while stealing glances at his phone. After that, he does a fireside chat at the posh Londoner Hotel in front of 600 members of the Oxford Guild. From there it's on to a basement conference room where he answers more technical questions from about 100 entrepreneurs and engineers. Now he's almost late to a mid-afternoon onstage talk at University College London. He and his group pull up at a loading zone and are ushered through a series of winding corridors, like the Steadicam shot in *Goodfellas*. As we walk, the moderator hurriedly tells Altman what he'll ask. When Altman pops on stage, the auditorium—packed with rapturous academics, geeks, and journalists—erupts.

Altman is not a natural publicity seeker. I once spoke to him right after *The New Yorker* ran a long profile of him. "Too much about me," he said. But at University College, after the formal program, he wades into the scrum of people who have surged to the foot of the stage. His aides try to maneuver themselves between Altman and the throng, but he shrugs them off. He takes one question after another, each time intently staring at the face of the interlocutor as if he's hearing the query for the first time. Everyone wants a selfie. After 20 minutes, he finally allows his team to pull him out. Then he's off to meet with UK prime minister Rishi Sunak.

Maybe one day, when robots write our history, they will cite Altman's world tour as a milestone in the year when everyone, all at once, started to make their own personal reckoning with the singularity. Or then again, maybe whoever writes the history of this moment will see it as a time when a quietly compelling CEO with a paradigm-busting technology made an attempt to inject a very peculiar worldview into the global mindstream—from an unmarked four-story headquarters in San Francisco's Mission District to the entire world.

For Altman and his company, ChatGPT and GPT-4 are merely stepping stones along the way to achieving a simple and seismic mission, one these technologists may as well have branded on their flesh.

That mission is to build artificial general intelligence—a concept that's so far been grounded more in science fiction than science—and to make it safe for humanity. The people who work at OpenAI are fanatical in their pursuit of that goal. (Though, as any number of conversations in the office café will confirm, the "build AGI" bit of the mission seems to offer up more raw excitement to its researchers than the "make it safe" bit.) These are people who do not shy from casually using the term "superintelligence." They *assume* that AI's trajectory will surpass whatever peak biology can attain. The company's financial documents even stipulate a kind of exit contingency for when AI wipes away our whole economic system.



SAM ALTMAN started OpenAI with Elon Musk in 2015.

It's not fair to call OpenAI a cult, but when I asked several of the company's top brass if someone could comfortably work there if they didn't believe AGI was truly coming—and that its arrival would mark one of the greatest moments in human history—most executives didn't think so. *Why would a nonbeliever want to work here?* they wondered. The assumption is that the workforce—now at approximately 500, though it might have grown since you began reading this paragraph—has self-selected to include only the faithful. At the very least, as Altman puts it, once you get hired, it seems inevitable that you'll be drawn into the spell.

At the same time, OpenAI is not the company it once was. It was founded as a purely nonprofit research operation, but today most of its employees technically work for a profit-making entity that is reportedly valued at almost $30 billion. Altman and his team now face the pressure to deliver a revolution in every product cycle, in a way that satisfies the commercial demands of investors and keeps ahead in a fiercely competitive landscape. All while hewing to a quasi-messianic mission to elevate humanity rather than exterminate it.

That kind of pressure—not to mention the unforgiving attention of the entire world—can be a debilitating force. The Beatles set off colossal waves of cultural change, but they anchored their revolution for only so long: Six years after chiming that unforgettable chord they weren't even a band anymore. The maelstrom OpenAI has unleashed will almost certainly be far bigger. But the leaders of OpenAI swear they'll stay the course. All they want to do, they say, is build computers smart enough and safe enough to end history, thrusting humanity into an era of unimaginable bounty.

GROWING UP IN the late '80s and early '90s, Sam Altman was a nerdy kid who gobbled up science fiction and *Star Wars*. The worlds built by early sci-fi writers often had humans living with—or competing with—superintelligent AI systems. The idea of computers matching or exceeding human capabilities thrilled Altman, who had been coding since his fingers could barely cover a keyboard. When he was 8, his parents bought him a Macintosh LC II. One night he was up late playing with it and the thought popped into his head: "Someday this computer is going to learn to think." When he arrived at Stanford as an undergrad in 2003, he hoped to help make that happen and took courses in AI. But "it wasn't working at all," he'd later say. The field was still mired in an innovation trough known as AI winter. Altman dropped out to enter the startup world; his company Loopt was in the tiny first batch of wannabe organizations in Y Combinator, which would become the world's most famed incubator.

In February 2014, Paul Graham, YC's founding guru, chose then-28-year-old Altman to succeed him. "Sam is one of the smartest people I know," Graham wrote in the announcement, "and understands startups better than perhaps anyone I know, including myself." But Altman saw YC as something bigger than a launchpad for companies. "We are not about startups," he told me soon after taking over.

AI TYPOGRAPHY: DEV VALLADARES + DEFORUM STABLE DIFFUSION
STYLING BY TURNER/THE WALL GROUP. HAIR AND MAKEUP BY HIROKO CLAUS

"We are about innovation, because we believe that is how you make the future great for everyone." In Altman's view, the point of cashing in on all those unicorns was not to pack the partners' wallets but to fund species-level transformations. He began a research wing, hoping to fund ambitious projects to solve the world's biggest problems. But AI, in his mind, was the one realm of innovation to rule them all: a superintelligence that could address humanity's problems better than humanity could.

As luck would have it, Altman assumed his new job just as AI winter was turning into an abundant spring. Computers were now performing amazing feats, via deep learning and neural networks, like labeling photos, translating text, and optimizing sophisticated ad networks. The advances convinced him that for the first time, AGI was actually within reach. Leaving it in the hands of big corporations, however, worried him. He felt those companies would be too fixated on their products to seize the opportunity to develop AGI as soon as possible. And if they did create AGI, they might recklessly unleash it upon the world without the necessary precautions.

At the time, Altman had been thinking about running for governor of California. But he realized that he was perfectly positioned to do something bigger—to lead a company that would change humanity itself. "AGI was going to get built exactly once," he told me in 2021. "And there were not that many people that could do a good job running OpenAI. I was lucky to have a set of experiences in my life that made me really positively set up for this."

Altman began talking to people who might help him start a new kind of AI company, a nonprofit that would direct the field toward responsible AGI. One kindred spirit was Tesla and SpaceX CEO Elon Musk. As Musk would later tell CNBC, he had become concerned about AI's impact after having some marathon discussions with Google cofounder Larry Page. Musk said he was dismayed that Page had little concern for safety and also seemed to regard the rights of robots as equal to humans. When Musk shared his concerns, Page accused him of being a "speciesist." Musk also understood that, at the time, Google employed much of the world's AI talent. He was willing to spend some money for an effort more amenable to Team Human.

Within a few months Altman had raised money from Musk (who pledged $100 million, and his time) and Reid Hoffman (who donated $10 million). Other funders included Peter Thiel, Jessica Livingston, Amazon Web Services, and YC Research. Altman began to stealthily recruit a team. He limited the search to AGI believers, a constraint that narrowed his options but one he considered critical. "Back in 2015, when we were recruiting, it was almost considered a career killer for an AI researcher to say that you took AGI seriously," he says. "But I wanted people who took it seriously."

Greg Brockman, the chief technology officer of Stripe, was one such person, and he agreed to be OpenAI's CTO. Another key cofounder would be Andrej Karpathy, who had been at Google Brain, the search giant's cutting-edge AI research operation.

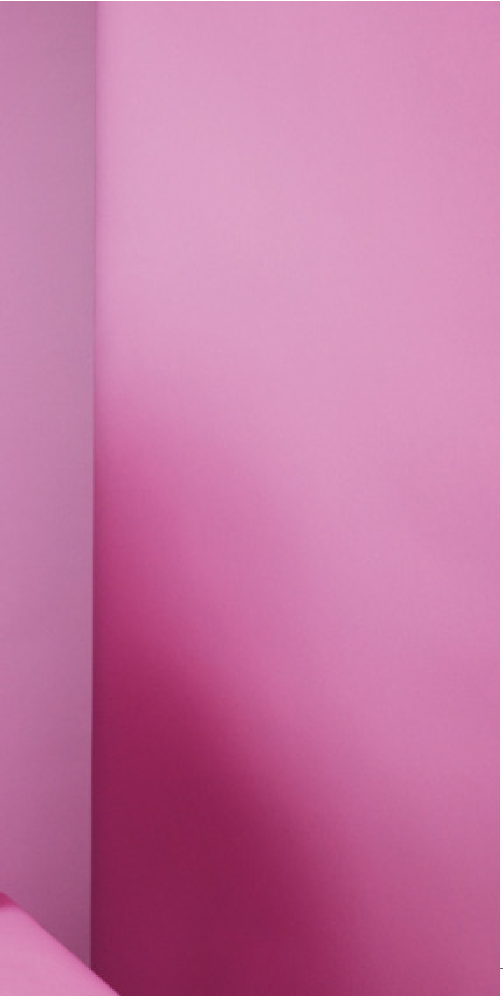
GREG BROCKMAN is now OpenAI's president.

But perhaps Altman's most sought-after target was a Russian-born engineer named Ilya Sutskever.

Sutskever's pedigree was unassailable. His family had emigrated from Russia to Israel, then to Canada. At the University of Toronto he had been a standout student under Geoffrey Hinton, known as the godfather of modern AI for his work on deep learning and neural networks. Hinton, who is still close to Sutskever, marvels at his protégé's wizardry. Early in Sutskever's tenure at the lab, Hinton had given him a complicated project. Sutskever got tired of writing code to do the requisite calculations, and he told Hinton it would be easier if he wrote a custom programming language for the task. Hinton got a bit annoyed and tried to warn his student away from what he assumed would be a monthlong distraction. Then Sutskever came clean: "I did it this morning."

Sutskever became an AI superstar, coauthoring a breakthrough paper that showed how AI could learn to recognize images simply by being exposed to huge volumes of data. He ended up, happily, as a key scientist on the Google Brain team.

In mid-2015 Altman cold-emailed Sutskever to invite him to dinner with Musk, Brockman, and others at the swank Rosewood Hotel on Palo Alto's Sand Hill Road. Only later did Sutskever figure out that he was the guest of honor. "It was kind of a general conversation about AI and AGI in the future," he says. More specifically, they discussed "whether Google and DeepMind were so far ahead that it would be impossible to catch up to them, or whether it was still possible to, as Elon put it, create a lab which would be a counterbalance." While no one at the dinner explicitly tried to recruit Sutskever, the conversation hooked him.



Sutskever wrote an email to Altman soon after, saying he was game to lead the project—but the message got stuck in his drafts folder. Altman circled back, and after months fending off Google's counteroffers, Sutskever signed on. He would soon become the soul of the company and its driving force in research.

Sutskever joined Altman and Musk in recruiting people to the project, culminating in a Napa Valley retreat where several prospective OpenAI researchers fueled each other's excitement. Of course, some targets would resist the lure. John Carmack, the legendary gaming coder behind *Doom*, *Quake*, and countless other titles, declined an Altman pitch.

OpenAI officially launched in December 2015. At the time, when I interviewed Musk and Altman, they presented the project to me as an effort to make AI safe and accessible by sharing it with the world. In other words, open source. OpenAI, they told me, was not going to apply for patents. Everyone could make use of their breakthroughs. Wouldn't that be empowering some future Dr. Evil? I wondered. Musk said that was a good question. But Altman had an answer: Humans are generally good, and because OpenAI would provide powerful tools for that vast majority, the bad actors would be overwhelmed. He admitted that if Dr. Evil were to use the tools to build something that couldn't be counteracted, "then we're in a really bad place." But both Musk and Altman believed that the safer course for AI would be in the hands of a research operation not polluted by the profit motive, a persistent temptation to ignore the needs of humans in the search for boffo quarterly results.

Altman cautioned me not to expect results soon. "This is going to look like a research lab for a long time," he said.

There was another reason to tamp down expectations. Google and the others had been developing and applying AI for years. While OpenAI had a billion dollars committed (largely via Musk), an ace team of researchers and engineers, and a lofty mission, it had no clue about how to pursue its goals. Altman remembers a moment when the small team gathered in Brockman's apartment—they didn't have an office yet. "I was like, what should we do?"

Altman remembers a moment when the small team gathered in Brockman's apartment—they didn't have an office yet. "I was like, what should we do?"

I had breakfast in San Francisco with Brockman a little more than a year after OpenAI's founding. For the CTO of a company with the word *open* in its name, he was pretty parsimonious with details.

He did affirm that the nonprofit could afford to draw on its initial billion-dollar donation for a while. The salaries of the 25 people on its staff—who were being paid at far less than market value—ate up the bulk of OpenAI’s expenses. “The goal for us, the thing that we’re really pushing on,” he said, “is to have the systems that can do things that humans were just not capable of doing before.” But for the time being, what that looked like was a bunch of researchers publishing papers. After the interview, I walked him to the company’s newish office in the Mission District, but he allowed me to go no further than the vestibule. He did duck into a closet to get me a T-shirt.

Radford began experimenting with the transformer architecture. “I made more progress in two weeks than I did over the past two years,” he says.

Had I gone in and asked around, I might have learned exactly how much OpenAI *was* floundering. Brockman now admits that “nothing was working.” Its researchers were tossing algorithmic spaghetti toward the ceiling to see what stuck. They delved in systems that solved video games and spent considerable effort in robotics. “We knew *what* we wanted to do,” says Altman. “We knew *why* we wanted to do it. But we had no idea *how*.”

But they *believed*. Supporting their optimism were the steady improvements in artificial neural networks that used deep-learning techniques.“The general idea is, don’t bet against deep learning,” says Sutskever. Chasing AGI, he says, “wasn’t totally crazy. It was only moderately crazy.”

OpenAI’s road to relevance really started with its hire of an as-yet-unheralded researcher named Alec Radford, who joined in 2016, leaving the small Boston AI company he’d cofounded in his dorm room. After accepting OpenAI’s offer, he told his high school alumni magazine that taking this new role was “kind of similar to joining a graduate program”—an open-ended, low-pressure perch to research AI.

The role he would actually play was more like Larry Page inventing PageRank.

Radford, who is press-shy and hasn’t given interviews on his work, responds to my questions about his early days at OpenAI via a long email exchange. His biggest interest was in getting neural nets to interact with humans in lucid conversation. This was a departure from the traditional scripted model of making a chatbot, an approach used in everything from the primitive ELIZA to the popular assistants Siri and Alexa—all of which kind of sucked. “The goal was to see if there was any task, any setting, any domain, any *anything* that language models could be useful for,” he writes. At the time, he explains, “language models were seen as novelty toys that could only generate a sentence that made sense once in a while, and only then if you really squinted.” His first experiment involved scanning 2 billion Reddit comments to train a language model. Like a lot of OpenAI’s early experiments, it flopped. No matter. The 23-year-old had permission to keep going, to fail again. “We were just like, Alec is great, let him do his thing,” says Brockman.

His next major experiment was shaped by OpenAI’s limitations of computer power, a constraint that led him to experiment on a smaller data set that focused on a single domain—Amazon product reviews. A researcher had gathered about 100 million of those. Radford trained a language model to simply predict the next character in generating a user review.

But then, on its own, the model figured out whether a review was positive or negative—and when you programmed the model to create something positive or negative, it delivered a review that was adulatory or scathing, as requested. (The prose was admittedly clunky: “I love this weapons look … A must watch for any man who love Chess!”) “It was a complete surprise,” Radford says. The sentiment of a review—its favorable or disfavorable gist—is a complex function of semantics, but somehow a part of Radford’s system had gotten a feel for it. Within OpenAI, this part of the neural net came to be known as the “unsupervised sentiment neuron.”

Sutskever and others encouraged Radford to expand his experiments beyond Amazon reviews, to use his insights to train neural nets to converse or answer questions on a broad range of subjects.

And then good fortune smiled on OpenAI. In early 2017, an unheralded preprint of a research paper appeared, coauthored by eight Google researchers. Its official title was “Attention Is All You Need,” but it came to be known as the “transformer paper,” named so both to reflect the game-changing nature

MIRA MURATI, OpenAI's chief technology officer.

of the idea and to honor the toys that transmogrified from trucks to giant robots. Transformers made it possible for a neural net to understand—and generate—language much more efficiently. They did this by analyzing chunks of prose in parallel and figuring out which elements merited "attention." This hugely optimized the process of generating coherent text to respond to prompts. Eventually, people came to realize that the same technique could also generate images and even video. Though the transformer paper would become known as the catalyst for the current AI frenzy—think of it as the Elvis that made the Beatles possible—at the time Ilya Sutskever was one of only a handful of people who understood how powerful the breakthrough was. "The real *aha* moment was when Ilya saw the transformer come out," Brockman says. "He was like, 'That's what we've been waiting for.' That's been our strategy—to push hard on problems and then have faith that we or someone in the field will manage to figure out the missing ingredient."

Radford began experimenting with the transformer architecture. "I made more progress in two weeks than I did over the past two years," he says. He came to understand that the key to getting the most out of the new model was to add scale—to train it on fantastically large data sets. The idea was dubbed "Big Transformer" by Radford's collaborator Rewon Child.

This approach required a change of culture at OpenAI and a focus it had previously lacked. "In order to take advantage of the transformer, you needed to scale it up," says Adam D'Angelo, the CEO of Quora, who sits on OpenAI's board of directors. "You need to run it more like an engineering organization. You can't have every researcher trying to do their own thing and training their own model and make elegant things that you can publish papers on. You have to do this more tedious, less elegant work." That, he added, was something OpenAI was able to do, and something no one else did.

The name that Radford and his collaborators gave the model they created was an acronym for "generatively pretrained transformer"—GPT-1. Eventually, this model came to be generically known as "generative AI." To build it, they drew on a collection of 7,000 unpublished books, many in the genres of romance, fantasy, and adventure, and refined it on Quora questions and answers, as well as thousands of passages taken from middle school and high school exams. All in all, the model included 117 million parameters, or variables. And it outperformed everything that had come before in understanding language and generating answers. But the most dramatic result was that processing such a massive amount of data allowed the model to offer up results *beyond* its training, providing expertise in brand-new domains. These unplanned robot capabilities are called zero-shots. They still baffle researchers—and account for the queasiness that many in the field have about these so-called large language models.

Radford remembers one late night at OpenAI's office. "I just kept saying over and over, 'Well, that's cool, but I'm pretty sure it won't be able to do *x*.' And then I would quickly code up an evaluation and, sure enough, it could kind of do *x*."

Each GPT iteration would do better, in part because each one gobbled an order of magnitude more data than the previous model. Only a year after creating the first iteration, OpenAI trained GPT-2 on the open internet with an astounding 1.5 billion parameters. Like a toddler mastering speech, its responses got better and more coherent. So much so that OpenAI hesitated to release the program into the wild. Radford was worried that it might be used to generate spam. "I remember reading Neal Stephenson's *Anathem* in 2008, and in that book the internet was overrun with spam generators," he says. "I had thought that was really far-fetched, but as I worked on language models over the years and they got better, the uncomfortable realization that it was a real possibility set in."

In fact, the team at OpenAI was starting to think it wasn't such a good idea after all to put its work where Dr. Evil could easily access it. "We thought that open-sourcing GPT-2 could be really dangerous," says chief technology officer Mira Murati, who started at the company in 2018. "We did a lot of work with misinformation experts and did some red-teaming. There was a lot of discussion internally on how much to release." Ultimately, OpenAI temporarily withheld the full version, making a less powerful version available to the public. When the company finally shared the full version, the world managed just fine—but there was no guarantee that more powerful models would avoid catastrophe.

The very fact that OpenAI was making products smart enough to be deemed dangerous, and was grappling with ways to make them safe, was proof that the company had gotten its mojo working. "We'd figured out the formula for progress, the formula everyone perceives now—the oxygen and the hydrogen of deep learning is computation with a large neural network and data," says Sutskever.

To Altman, it was a mind-bending experience. "If you asked the 10-year-old version of me, who used to spend a lot of time daydreaming about AI, what was going to happen, my pretty confident prediction would have been that first we're gonna have robots, and they're going to perform all physical labor. Then we're going to have systems that can do basic cognitive labor. A really long way after that, maybe we'll have systems that can do complex stuff like proving mathematical theorems. Finally we will have AI that can create new things and make art and write and do these deeply human things. That was a terrible prediction—it's going exactly the other direction."

The world didn't know it yet, but Altman and Musk's research lab had begun a climb that plausibly creeps toward the summit of AGI. The crazy idea behind OpenAI suddenly was not so crazy.

BY EARLY 2018, OpenAI was starting to focus productively on large language models, or LLMs. But Elon Musk wasn't happy. He felt that the progress was insufficient—or maybe he felt that now that OpenAI was on to something, it needed leadership to seize its advantage. Or maybe, as he'd later explain, he felt that safety should be more of a priority. Whatever his problem was, he had a solution: Turn everything over to him. He proposed taking a majority stake in the company, adding it to the portfolio of his multiple full-time jobs (Tesla, SpaceX) and supervisory obligations (Neuralink and the Boring Company).

Musk believed he had a *right* to own OpenAI. "It wouldn't exist without me," he later told CNBC. "I came up with the name!" (True.) But Altman and the rest of OpenAI's brain trust had no interest in becoming part of the Muskiverse. When they made this clear, Musk cut ties, providing the public with the incomplete explanation that he was leaving the board to avoid a conflict with Tesla's AI effort. His farewell came at an all-hands meeting early that year where he predicted that OpenAI would fail. And he called at least one of the researchers a "jackass."

He also took his money with him. Since the company had no revenue, this was an existential crisis. "Elon is cutting off his support," Altman said in a panicky call to Reid Hoffman. “What do we do?” Hoffman volunteered to keep the company afloat, paying overhead and salaries.

But this was a temporary fix; OpenAI had to find big bucks elsewhere. Silicon Valley loves to throw money at talented people working on trendy tech. But not so much if they are working at a nonprofit. It had been a massive lift for OpenAI to get its first billion. To train and test new generations of GPT—and then access the computation it takes to deploy them—the company needed another billion, and fast. And that would only be the start.

So in March 2019, OpenAI came up with a bizarre hack. It would remain a nonprofit, fully devoted to its mission. But it would also create a for-profit entity. The actual structure of the arrangement is hopelessly baroque, but basically the entire company is now engaged in a “capped” profitable business. If the cap is reached—the number isn’t public, but its own charter, if you read between the lines, suggests it might be in the trillions—everything beyond that reverts to the nonprofit research lab. The novel scheme was almost a quantum approach to incorporation: Behold a company that, depending on your time-space point of view, is for-profit and nonprofit. The details are embodied in charts full of boxes and arrows, like the ones in the middle of a scientific paper where only PhDs or dropout geniuses dare to tread. When I suggest to Sutskever that it looks like something the as-yet-unconceived GPT-6 might come up with if you prompted it for a tax dodge, he doesn’t warm to my metaphor. “It’s not about accounting,” he says.

Somewhere in the restructuring documents is a clause to the effect that, if the company does manage to create AGI, all financial arrangements will be reconsidered. After all, it will be a new world from that point on.

But accounting is critical. A for-profit company optimizes for, well, profits. There’s a reason why companies like Meta feel pressure from shareholders when they devote billions to R&D. How could this not affect the way a firm operates? And wasn’t avoiding commercialism the reason why Altman made OpenAI a nonprofit to begin with? According to COO Brad Lightcap, the view of the company’s leaders is that the board, which is still part of the nonprofit controlling entity, will make sure that the drive for revenue and profits won’t overwhelm the original idea. “We needed to maintain the mission as the reason for our existence,” he says, “It shouldn’t just be in spirit, but encoded in the structure of the company.” Board member Adam D’Angelo says he takes this responsibility seriously: “It’s my job, along with the rest of the board, to make sure that OpenAI stays true to its mission.”

Potential investors were warned about those boundaries, Lightcap explains. “We have a legal disclaimer that says you, as an investor, stand to lose all your money,” he says. “We are not here to make your return. We’re here to achieve a technical mission, foremost. And, oh, by the way, we don’t really know what role money will play in a post-AGI world.”

That last sentence is not a throwaway joke. OpenAI’s plan really does include a reset in case computers reach the final frontier. Somewhere in the restructuring documents is a clause to the effect that, if the company does manage to create AGI, all financial arrangements will be reconsidered. After all, it will be a new world from that point on. Humanity will have an alien partner that can do much of what we do, only better. So previous arrangements might effectively be kaput.

There is, however, a hitch: At the moment, OpenAI doesn’t claim to know what AGI really *is*. The determination would come from the board, but it’s not clear how the board would define it. When I ask Altman, who is on the board, for clarity, his response is anything but open. “It’s not a single Turing test, but a number of things we might use,” he says. “I would happily tell you, but I like to keep confidential conversations private. I realize that is unsatisfyingly vague. But we don’t know what it’s going to be like at that point.”

Nonetheless, the inclusion of the “financial arrangements” clause isn’t just for fun: OpenAI’s leaders think that if the company is successful enough to reach its lofty profit cap, its products will probably have performed well enough to reach AGI. Whatever that is.

“My regret is that we’ve chosen to double down on the term AGI,” Sutskever says. “In hindsight it is a confusing term, because it emphasizes generality above all else. GPT-3 is general AI, but yet we don’t really feel comfortable calling it AGI, because we want human-level competence. But back then, at the beginning, the idea of OpenAI was that superintelligence is attainable. It is the endgame, the final purpose of the field of AI.”

ILYA SUTSKEVER, OpenAI's chief scientist.

Those caveats didn't stop some of the smartest venture capitalists from throwing money at OpenAI during its 2019 funding round. At that point, the first VC firm to invest was Khosla Ventures, which kicked in $50 million. According to Vinod Khosla, it was double the size of his largest initial investment. "If we lose, we lose 50 million bucks," he says. "If we win, we win 5 billion." Others investors reportedly would include elite VC firms Thrive Capital, Andreessen Horowitz, Founders Fund, and Sequoia.

The shift also allowed OpenAI's employees to claim some equity. But not Altman. He says that originally he intended to include himself but didn't get around to it. Then he decided that he didn't need any piece of the $30 billion company that he'd cofounded and leads. "Meaningful work is more important to me," he says. "I don't think about it. I honestly don't get why people care so much."

Because ... not taking a stake in the company you cofounded is weird?

"If I didn't already have a ton of money, it would be much weirder," he says. "It does seem like people have a hard time imagining ever having enough money. But I feel like I have enough." (Note: For Silicon Valley, this is *extremely* weird.) Altman joked that he's considering taking one share of equity "so I never have to answer that question again."

■ **THE BILLION-DOLLAR** VC round wasn't even table stakes to pursue OpenAI's vision. The miraculous Big Transformer approach to creating LLMs required Big Hardware. Each iteration of the GPT family would need exponentially more power—GPT-2 had over a billion parameters, and GPT-3 would use 175 billion. OpenAI was now like Quint in *Jaws* after the shark hunter sees the size of the great white. "It turned out we didn't know how much of a bigger boat we needed," Altman says.

Obviously, only a few companies in existence had the kind of resources OpenAI required. "We pretty quickly zeroed in on Microsoft," says Altman. To the credit of Microsoft CEO Satya Nadella and CTO Kevin Scott, the software giant was able to get over an uncomfortable reality: After more than 20 years and billions of dollars spent on a research division with supposedly cutting-edge AI, the Softies needed an innovation infusion from a tiny company that was only a few years old. Scott says that it wasn't just Microsoft that fell short—"it was everyone." OpenAI's focus on pursuing AGI, he says, allowed it to accomplish a moonshot-ish achievement that the heavy hitters weren't even aiming for. It also proved that not pursuing generative AI was a lapse that Microsoft needed to address. "One thing you just very clearly need is a frontier model," says Scott.

Microsoft originally chipped in a billion dollars, paid off in computation time on its servers. But as both sides grew more confident, the deal expanded. Microsoft now has sunk $13 billion into OpenAI. ("Being on the frontier is a very expensive proposition," Scott says.)

Of course, because OpenAI couldn't exist without the backing of a huge cloud provider, Microsoft was able to cut a great deal for itself. The corporation bargained for what Nadella calls "non-controlling equity interest" in OpenAI's for-profit side—reportedly 49 percent. Under the terms of the deal, some of OpenAI's original ideals of granting equal access to all were seemingly dragged to the trash icon. (Altman objects to this characterization.) Now, Microsoft has an exclusive license to commercialize OpenAI's tech. And OpenAI also has committed to use Microsoft's cloud exclusively. In other words, without even taking its cut of OpenAI's profits (reportedly Microsoft gets 75 percent until its investment is paid back), Microsoft gets to lock in one of the world's most desirable new customers for its Azure web services. With those rewards in sight, Microsoft wasn't even bothered by the clause that demands reconsideration if OpenAI achieves general artificial intelligence, whatever that is. "At that point," says Nadella, "all bets are off." It might be the last invention of humanity, he notes, so we might have bigger issues to consider once machines are smarter than we are.

By the time Microsoft began unloading Brinks trucks' worth of cash into OpenAI ($2 billion in 2021, and the other $10 billion earlier this year), OpenAI had completed GPT-3, which, of course, was even more impressive than its predecessors. When Nadella saw what GPT-3 could do, he says, it was the first time he deeply understood that Microsoft had snared something truly transformative. "We started observing all those emergent properties." For instance, GPT had taught itself how to program computers. "We didn't train it on coding—it just got good at coding!" he says. Leveraging its ownership of Github, Microsoft released a product called Copilot that uses GPT to churn out code literally on command. Microsoft would later integrate OpenAI technology in new versions of its workplace products. Users pay a premium for those, and a cut of that revenue gets logged to OpenAI's ledger.

Some observers professed whiplash at OpenAI's one-two punch: creating a for-profit component and reaching an exclusive deal with Microsoft. How did a company that promised to remain patent-free, open source, and totally transparent wind up giving an exclusive license of its tech to the world's biggest software company? Elon Musk's remarks were particularly lacerating. "This does seem like the opposite of open—OpenAI is essentially captured by Microsoft," he posted on Twitter. On CNBC, he elaborated with an analogy: "Let's say you founded an organization to save the Amazon rainforest, and instead you became a lumber company, chopped down the forest, and sold it."

Musk's jibes might be dismissed as bitterness from a rejected suitor, but he wasn't alone. "The whole vision of it morphing the way it did feels kind of gross," says John Carmack. (He does specify that he's still excited about the company's work.) Another prominent industry insider, who prefers to speak without attribution, says, "OpenAI has turned from a small, somewhat open research outfit into a secretive product-development house with an unwarranted superiority complex."

Even some employees had been turned off by OpenAI's venture into the for-profit world. In 2019, several key executives, including head of research Dario Amodei, left to start a rival AI company called Anthropic. They recently told *The New York Times* that OpenAI had gotten too commercial and had fallen victim to mission drift.

Another OpenAI defector was Rewon Child, the main technical contributor to the GPT-2 and GPT-3 projects. He left in late 2021 and is now at Inflection AI, a company led by former DeepMind cofounder Mustafa Suleyman.

Altman professes not to be bothered by defections, dismissing them as simply the way Silicon Valley works. "Some people will want to do great work somewhere else, and that pushes society forward," he says. "That absolutely fits our mission."

■ **UNTIL NOVEMBER OF** last year, awareness of OpenAI was largely confined to people following technology and software development. But as the whole world now knows, OpenAI took the dramatic step of releasing a consumer product late that month, built on what was then the most recent iteration of GPT, version 3.5. For months, the company had been internally using a version of GPT with a conversational interface. It was especially important for what the company called "truth-seeking." That means that via dialog, the user could coax the model to provide responses that would be more trustworthy and complete. ChatGPT, optimized for the masses, could allow anyone to instantly tap into what seemed to be an endless source of knowledge simply by typing in a prompt—and then continue the conversation as if hanging out with a fellow human who just happened to know everything, albeit one with a penchant for fabrication.

Within OpenAI, there was a lot of debate about the wisdom of releasing a tool with such unprecedented power. But Altman was all for it. The release, he explains, was part of a strategy designed to acclimate the public to the reality that artificial intelligence is destined to change their everyday lives, presumably for the better. Internally, this is known as the "iterative deployment hypothesis." Sure, ChatGPT would create a stir, the thinking went. After all, here was something anyone could use that was smart enough to get college-level scores on the SATs, write a B-minus essay, and summarize a book within seconds. You could ask it to write your funding proposal or summarize a meeting and then request it to do a rewrite in Lithuanian or as a Shakespeare sonnet or in the voice of someone obsessed with toy trains. In a few seconds, pow, the LLM would comply. Bonkers. But OpenAI saw it as a table-setter for its newer, more coherent, more capable, and scarier successor, GPT4, trained with a reported 1.7 trillion parameters. (OpenAI won't confirm the number, nor will it reveal the data sets.)

Altman explains why OpenAI released ChatGPT when GPT-4 was close to completion, undergoing safety work. "With ChatGPT, we could introduce chatting but with a much less powerful backend, and give people a more gradual adaptation," he says. "GPT-4 was a lot to get used to at once." By the time the ChatGPT excitement cooled down, the thinking went, people might be ready for GPT-4, which can pass the bar exam, plan a course syllabus, and write a book within seconds. (Publishing houses that produced genre fiction were indeed flooded with AI-generated bodice-rippers and space operas.)

A cynic might say that a steady cadence of new products is tied to the company's commitment to investors, and equity-holding employees, to make some money. OpenAI now charges customers who use its products frequently. But OpenAI insists that its true strategy is to provide a soft landing for the singularity. "It doesn't make sense to just build AGI in secret and drop it on the world," Altman says. "Look back at the industrial revolution—everyone agrees it was great for the world," says Sandhini Agarwal, an OpenAI policy researcher. "But the first 50 years were really painful. There was a lot of job loss, a lot of poverty, and then the world adapted. We're trying to think how we can make the period before adaptation of AGI as painless as possible."

Sutskever puts it another way: "You want to build larger and more powerful intelligences and keep them in your basement?"

Even so, OpenAI was stunned at the reaction to ChatGPT. "Our internal excitement was more focused on GPT-4," says Murati, the CTO. "And so we didn't think ChatGPT was really going to change everything." To the contrary, it galvanized the public to the reality that AI had to be dealt with, *now*. ChatGPT became the fastest-growing consumer software in history, amassing a reported 100 million users. (Not-so-OpenAI won't confirm this, saying only that it has "millions of users.") "I underappreciated how much making an easy-to-use conversational interface to an LLM would make it much more intuitive for everyone to use," says Radford.

ChatGPT was of course delightful and astonishingly useful, but also scary—prone to "hallucinations" of plausible but shamefully fabulist details when responding to prompts. Even as journalists wrung their hands about the implications, however, they effectively endorsed ChatGPT by extolling its powers.

The clamor got even louder in February when Microsoft, taking advantage of its multibillion-dollar partnership, released a ChatGPT-powered version of its search engine Bing. CEO Nadella was euphoric that he had beaten Google to the punch in introducing generative AI to Microsoft's products. He taunted the search king, which had been cautious in releasing its own LLM into products, to do the same. "I want people to know we made them dance," he said.

In so doing, Nadella triggered an arms race that tempted companies big and small to release AI products before they were fully vetted. He also a triggered a new round of media coverage that kept wider and wider circles of people up at night: Interactions with Bing that unveiled the chatbot's shadow side, replete with unnerving professions of love, an envy of human freedom, and a weak resolve to withhold misinformation. As well as an unseemly habit of creating hallucinatory misinformation of its own.

But if OpenAI's products were forcing people to confront the implications of artificial intelligence, Altman figured, so much the better. It was time for the bulk of humankind to come off the sidelines in discussions of how AI might affect the future of the species.

AS SOCIETY STARTED to prioritize thinking through all the potential drawbacks of AI—job loss, misinformation, human extinction—OpenAI set about placing itself in the center of the discussion. Because if regulators, legislators, and doomsayers mounted a charge to smother this nascent alien intelligence in its cloud-based cradle, OpenAI would be their chief target anyway. "Given our current visibility, when things go wrong, even if those things were built by a different company, that's still a problem for us, because we're viewed as the face of this technology right now," says Anna Makanju, OpenAI's chief policy officer.

Makanju is a Russian-born DC insider who served in foreign policy roles at the US Mission to the United Nations, the US National Security Council, the Defense Department, and in the office of Joe Biden when he was vice president. "I have lots of preexisting relationships, both in the US government and in various European governments," she says. She joined OpenAI in September 2021. At the time, very few people in government gave a hoot about generative AI. Knowing that OpenAI's products would soon change that, she began to introduce Altman to administration officials and legislators, making sure that they'd hear the good news and the bad from OpenAI first.

"Sam has been extremely helpful, but also very savvy, in the way that he has dealt with members of Congress," says Richard Blumenthal, the chair of the Senate Judiciary Committee. He contrasts Altman's behavior to that of the younger Bill Gates, who unwisely stonewalled legislators when Microsoft was under antitrust investigations in the 1990s. "Altman, by contrast, was happy to spend an hour or more sitting with me, to try to educate me," says Blumenthal. "He didn't come with an army of lobbyists or minders. He demonstrated ChatGPT. It was mind-blowing."

OPENAI's San Francisco headquarters is unmarked; but inside, the coffee is awesome.

In Blumenthal, Altman wound up making a semi-ally of a potential foe. "Yes," the senator admits. "I'm excited about both the upside and the potential perils." OpenAI didn't shrug off discussion of those perils, but presented itself as the force best positioned to mitigate them. "We had 100-page system cards on all the red-teaming safety valuations," says Makanju. (Whatever that meant, it didn't stop users and journalists from endlessly discovering ways to jailbreak the system.)

By the time Altman made his first appearance in a congressional hearing—fighting a fierce migraine headache—the path was clear for him to sail through in a way that Bill Gates or Mark Zuckerberg could never hope to. He faced almost none of the tough questions and arrogant badgering that tech CEOs now routinely endure after taking the oath. Instead, senators asked Altman for advice on how to regulate AI, a pursuit Altman enthusiastically endorsed.

The paradox is that no matter how assiduously companies like OpenAI red-team their products to mitigate misbehavior like deepfakes, misinformation efforts, and criminal spam, future models might get smart enough to foil the efforts of the measly-minded humans who invented the technology yet are still naive enough to believe they can control it. On the other hand, if they go *too* far in making their models safe, it might hobble the products, making them less useful. One study indicated that more recent versions of GPT, which have improved safety features, are actually dumber than previous versions, making errors in basic math problems that earlier programs had aced. (Altman says that OpenAI's data doesn't confirm this. "Wasn't that study retracted?" he asks. No.)

It makes sense that Altman positions himself as a fan of regulation; after all, his mission is AGI, but safely. Critics have charged that he's gaming the process so that regulations would thwart smaller startups and give an advantage to OpenAI and other big players. Altman denies this. While he has endorsed, in principle, the idea of an international agency overseeing AI, he does feel that some proposed rules, like banning all copyrighted material from data sets, present unfair obstacles. He pointedly didn't sign a widely distributed letter urging a six-month moratorium on developing AI systems. But he and other OpenAI leaders did add their names to a one-sentence statement: "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war." Altman explains: "I said, 'Yeah, I agree with that. One-minute discussion.'"

As one prominent Silicon Valley founder notes, "It's rare that an industry raises their hand and says, 'We are going to be the end of humanity'—and then continues to work on the product with glee and alacrity."

OpenAI rejects this criticism. Altman and his team say that working and releasing cutting-edge products is *the way* to address societal risks. Only by analyzing the responses to millions of prompts by users of ChatGPT and GPT-4 could they get the knowledge to ethically align their future products.

Still, as the company takes on more tasks and devotes more energy to commercial activities, some question how closely OpenAI can concentrate on the mission—especially the "mitigating risk of extinction" side. "If you think about it, they're actually building *five* businesses," says an AI industry executive, ticking them off with his fingers. "There's the product itself, the enterprise relationship with Microsoft, the developer ecosystem, and an app store. And, oh yes—they are also obviously doing an AGI research mission." Having used all five fingers, he recycles his index finger to add a sixth. "And of course, they're also doing the investment fund," he says, referring to a $175 million project to seed startups that want to tap into OpenAI technology. "These are different cultures, and in fact they're conflicting with a research mission."

I repeatedly asked OpenAI's execs how donning the skin of a product company has affected its culture. Without fail they insist that, despite the for-profit restructuring, despite the competition with Google, Meta, and countless startups, the mission is still central. Yet OpenAI *has* changed. The nonprofit board might technically be in charge, but virtually everyone in the company is on the for-profit ledger. Its workforce includes lawyers, marketers, policy experts, and user-interface designers. OpenAI contracts with hundreds of content moderators to educate its models on inappropriate or harmful answers to the prompts offered by many millions of users. It's got product managers and engineers working constantly on updates to its products, and every couple of weeks it seems to ping reporters with demonstrations—just like other product-oriented Big Tech companies. Its offices look like an *Architectural Digest* spread. I have visited virtually every major tech company in Silicon Valley and beyond, and not one surpasses the coffee options in the lobby of OpenAI's headquarters in San Francisco.

Not to mention: It's obvious that the "openness" embodied in the company's name has shifted from the radical transparency suggested at launch. When I bring this up to Sutskever, he shrugs. "Evidently, times have changed," he says. But, he cautions, that doesn't mean that the prize is not the same. "You've got a technological transformation of such gargantuan, cataclysmic magnitude that, even if we all do our part, success is not guaranteed. But if it all works out we can have quite the incredible life."

"The biggest thing we're missing is coming up with new ideas," says Brockman. "It's nice to have something that could be a virtual assistant. But that's not the dream. The dream is to help us solve problems we can't."

"I can't emphasize this enough—we didn't have a master plan," says Altman. "It was like we were turning each corner and shining a flashlight. We were willing to go through the maze to get to the end." Though the maze got twisty, the goal has not changed. "We still have our core mission—believing that safe AGI was this critically important thing that the world was not taking seriously enough."

Meanwhile, OpenAI is apparently taking its time to develop the next version of its large language model. It's hard to believe, but the company insists it has yet to begin working on GPT-5, a product that people are, depending on point of view, either salivating about or dreading. Apparently, OpenAI is grappling with what an exponentially powerful improvement on its current technology actually looks like. "The biggest thing we're missing is coming up with new ideas," says Brockman. "It's nice to have something that could be a virtual assistant. But that's not the dream. The dream is to help us solve problems we can't."

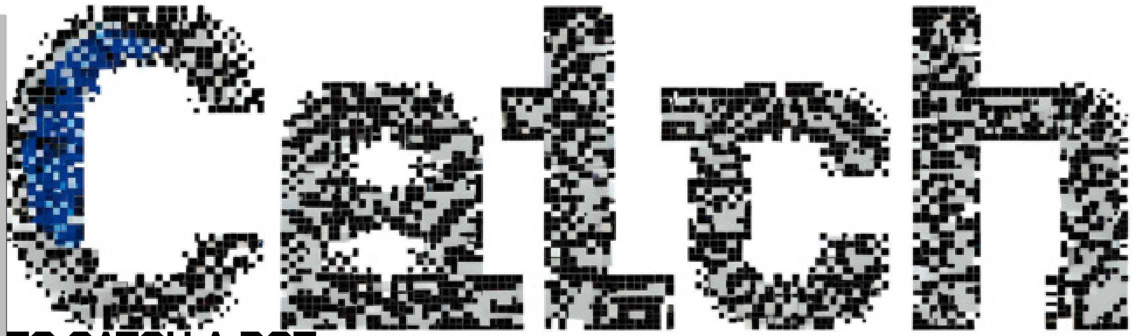
Considering OpenAI's history, that next big set of innovations might have to wait until there's another breakthrough as major as transformers. Altman hopes that will come from OpenAI—"We want to be the best research lab in the world," he says—but even if not, his company will make use of others' advances, as it did with Google's work. "A lot of people around the world are going to do important work," he says.

It would also help if generative AI didn't create so many new problems of its own. For instance, LLMs need to be trained on huge data sets; clearly the most powerful ones would gobble up the whole internet. This doesn't sit well with some creators, and just plain people, who unwittingly provide content for those data sets and wind up somehow contributing to the output of ChatGPT. Tom Rubin, an elite intellectual property lawyer who officially joined OpenAI in March, is optimistic that the company will eventually find a balance that satisfies both its own needs and that of creators—including the ones, like comedian Sarah Silverman, who are suing OpenAI for using their content to train its models. One hint of OpenAI's path: partnerships with news and photo agencies like the Associated Press and Shutterstock to provide content for its models without questions of who owns what.

As I interview Rubin, my very human mind, subject to distractions you never see in LLMs, drifts to the arc of this company that in eight short years has gone from a floundering bunch of researchers to a Promethean behemoth that has changed the world. Its very success has led it to transform itself from a novel effort to achieve a scientific goal to something that resembles a standard Silicon Valley unicorn on its way to elbowing into the pantheon of Big Tech companies that affect our everyday lives. And here I am, talking with one of its key hires—a lawyer—not about neural net weights or computer infrastructure but copyright and fair use. Has this IP expert, I wonder, signed on to the mission, like the superintelligence-seeking voyagers who drove the company originally?

Rubin is nonplussed when I ask him whether he believes, as an article of faith, that AGI will happen and if he's hungry to make it so. "I can't even answer that," he says after a pause. When pressed further, he clarifies that, as an intellectual property lawyer, speeding the path to scarily intelligent computers is not his job. "From my perch, I look forward to it," he finally says.

STEVEN LEVY *is* WIRED*'s editor at large. His last feature, in issue 31.09, was a Big Interview with Grimes. He writes the* Plaintext *newsletter.*



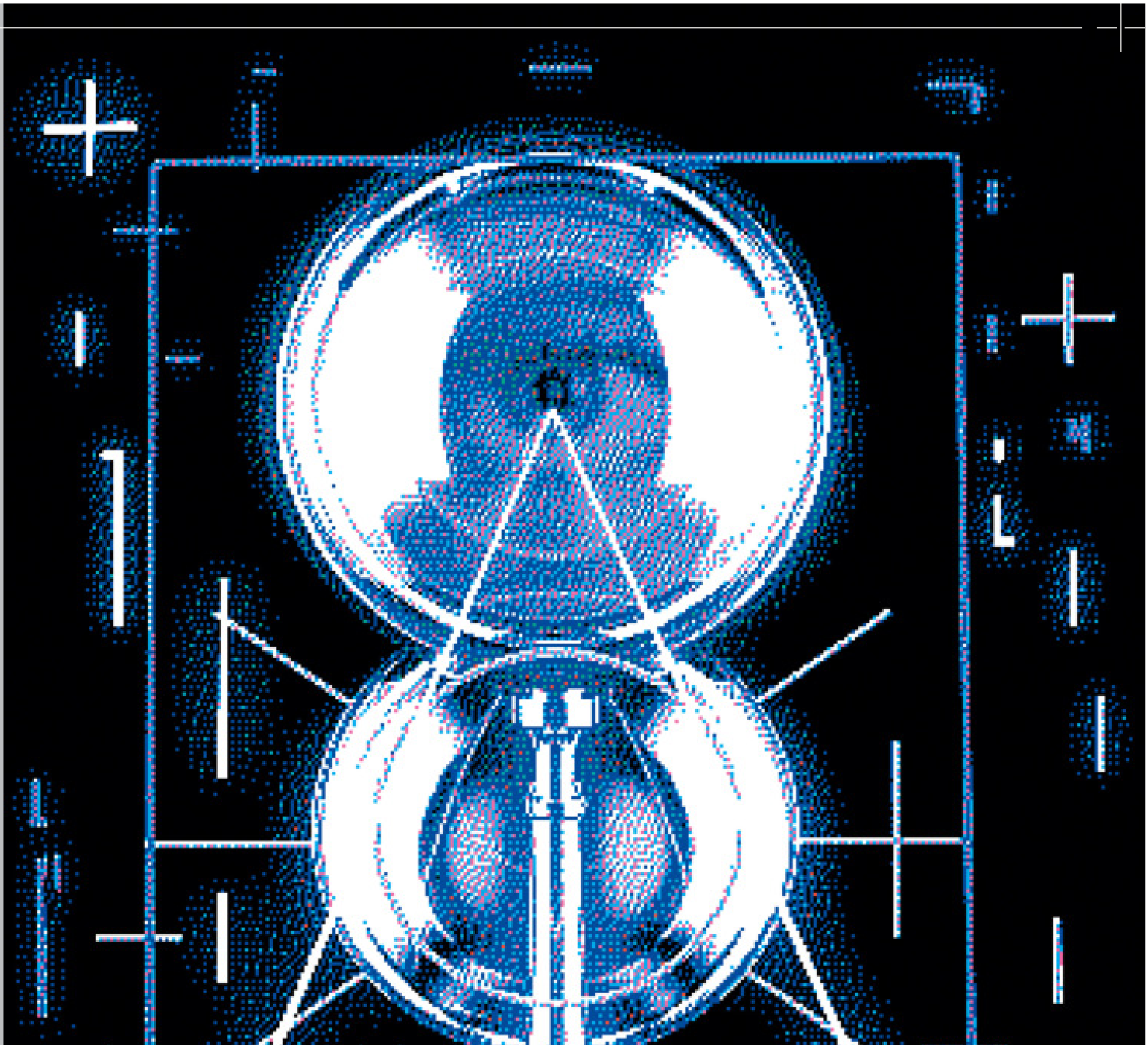
TO CATCH A BOT

A BATTLE IS RAGING BETWEEN TOOLS DESIGNED TO IDENTIFY AI-GENERATED TEXT AND THOSE DESIGNED TO EVADE DETECTION.

BY CHRISTOPHER BEAM
ART BY JAMES MARSHALL

PHOTOGRAPHS BY CHARIS MORGAN & LAURYN HILL

AI TYPOGRAPHY: DEV VALLADARES + DEFORUM STABLE DIFFUSION



EDWARD TIAN DIDN'T think of himself as a writer. As a computer science major at Princeton, he'd taken a couple of journalism classes, where he learned the basics of reporting, and his sunny affect and tinkerer's curiosity endeared him to his teachers and classmates. But he describes his writing style at the time as "pretty bad"—formulaic and clunky. One of his journalism professors said that Tian was good at "pattern recognition," which was helpful when producing news copy. So Tian was surprised when, sophomore year, he managed to secure a spot in John McPhee's exclusive non-fiction writing seminar.

Every week, 16 students gathered to hear the legendary *New Yorker* writer dissect his craft. McPhee assigned exercises that forced them to think rigorously about words: Describe a piece of modern art on campus, or prune the Gettysburg Address for length. Using a projector and slides, McPhee shared hand-drawn diagrams that illustrated different ways he structured his own essays: a straight line, a triangle, a spiral. Tian remembers McPhee saying he couldn't tell his students how to write, but he could at least help them find their own unique voice.

If McPhee stoked a romantic view of language in Tian, computer science offered a different perspective: language as statistics. During the pandemic, he'd taken a year off to work at the BBC and intern at Bellingcat, an open source journalism project, where he'd written code to detect Twitter bots. As a junior, he'd taken classes on machine learning and natural language processing. And in the fall of 2022, he began to work on his senior thesis about detecting the differences between AI-generated and human-written text.

When ChatGPT debuted in November, Tian found himself in an unusual position. As the world lost its mind over this new, radically improved chatbot, Tian was already familiar with the underlying GPT-3 technology. And as a journalist who'd worked on rooting out disinformation campaigns, he understood the implications of AI-generated content for the industry.

While home in Toronto for winter break, Tian started playing around with a new program: a ChatGPT detector. He posted up at his favorite café, slamming jasmine tea, and stayed up late coding in his bedroom. His idea was simple. The software would scan a piece of text for two factors: "perplexity," the randomness of word choice; and "burstiness," the complexity or variation of sentences. Human writing tends to rate higher than AI writing on both metrics, which allowed Tian to guess how a piece of text had been created. Tian called the tool GPTZero—the "zero" signaled truth, a return to basics—and he put it online the evening of January 2. He posted a link on Twitter with a brief introduction. The goal was to combat "increasing AI plagiarism," he wrote. "Are high school teachers going to want students using ChatGPT to write their history essays? Likely not." Then he went to bed.

Tian woke up the next morning to hundreds of retweets and replies. There was so much traffic to the host server that many users couldn't access it. "It was totally crazy," Tian says. "My phone was blowing up." A friend congratulated him on winning the internet. Teens on TikTok called him a narc. "A lot of the initial hate was like, 'This kid is a snitch, he doesn't have a life, he never had a girlfriend,'" says Tian with a grin. "Classic stuff." (Tian has a girlfriend.) Within days, he was fielding calls from journalists around the world, eventually appearing on everything from NPR to the *South China Morning Post* to *Anderson Cooper 360*. Within a week, his original tweet had reached more than 7 million views.

GPTZero was a new twist in the media narrative surrounding ChatGPT, which had inspired industry-wide hand-wringing and a scourge of AI-generated ledes. (Researchers had created a detector for GPT-2 text in 2019, but Tian's was the first to target ChatGPT.) Teachers thanked Tian for his work, grateful they could finally prove their suspicions about fishy student essays. Had humanity found its savior from the robot takeover?

Tian's program was a starting gun of sorts. The race was now on to create the definitive AI detection tool. In a world increasingly saturated with AI-generated content, the thinking went, we'll need to distinguish the machine-made from the human-made. GPTZero represented a promise that it will indeed be possible to tell one from the other, and a conviction that the difference matters. During his media tour, Tian—smiley, earnest, the A student incarnate—elaborated on this reassuring view that no matter how sophisticated generative AI tools become, we will always be able to unmask them. There's something irreducible about human writing, Tian said: "It has an element that can never be put into numbers."

LIFE ON THE INTERNET has always been a battle between fakers and detectors of fakes, with both sides profiting off the clash. Early spam filters sifted emails for keywords, blocking messages with phrases like "FREE!" or "be over 21," and they eventually learned to filter out entire styles of writing. Spammers responded by surrounding their pitches with snippets of human-sounding language lifted from old books and mashed together. (This type of message, dubbed "litspam," became a genre unto itself.) As search engines grew more popular, creators looking to boost their pages' rankings resorted to "keyword stuffing"— repeating the same word over and over—to get priority. Search engines countered by down-ranking those sites. After Google introduced its PageRank algorithm, which favored websites with lots of inbound links, spammers created entire ecosystems of mutually supporting pages.

Around the turn of the millennium, the captcha tool arrived to sort humans from bots based on their ability to interpret images of distorted text. Once some bots could handle that, captcha added other detection methods that included parsing images of motorbikes and trains, as well as sensing mouse movement and other user behavior. (In a recent test, an early version of GPT-4 showed that it knew how to hire a person on Taskrabbit to complete a captcha on its behalf.) The fates of entire companies have rested on the issue of spotting fakes: Elon Musk, in an attempt to wriggle out of his deal to buy Twitter, cited a bot detector to boost his argument that Twitter had misrepresented the number of bots on its site.

Generative AI re-upped the ante. While large language models and text-to-image generators have been evolving steadily over the past decade, 2022 saw an explosion of consumer-friendly tools like ChatGPT and Dall-E. Pessimists argue that we could soon drown in a tsunami of synthetic media. "In a few years, the vast majority of the photos, videos, and text we encounter on the internet could be AI-generated," *New York Times* technology columnist Kevin Roose warned last year. *The Atlantic* imagined a looming "textpocalypse" as we struggle to filter out the generative noise. Political campaigns are leveraging AI tools to create ads, while Amazon is flooded with ChatGPT-written books (many of them about AI). Scrolling through product reviews already feels like the world's most annoying Turing test. The next step seems clear: If you thought Nigerian prince emails were bad, wait until you see Nigerian prince chatbots.

Soon after Tian released GPTZero, a wave of similar products appeared. OpenAI rolled out its own detection tool at the end of January, while Turnitin, the anti-plagiarism giant, unveiled a classifier in April. They all shared a basic methodology, but each model was trained on different data sets. (For example, Turnitin focused on student writing.) As a result, precision varied wildly, from OpenAI's claim of 26 percent accuracy for detecting AI-written text, up to the most optimistic claim from a company called Winston AI at 99.6 percent. To stay ahead of the competition, Tian would have to keep improving GPTZero, come up with its next product, and finish college in the meantime.

Right away, Tian recruited his high school friend Alex Cui as CTO and, over the following weeks, brought on a handful of programmers from Princeton and Canada. Then, in the spring, he onboarded a trio of coders from Uganda, whom he'd met four years earlier while working for a startup that trains engineers in Africa. (A global citizen, Tian was born in Tokyo and lived in Beijing until age 4 before his parents, both Chinese engineers, moved the family to Ontario.) Together the team began working on its next application: a Chrome plug-in that would scan the text of a web page and determine whether it was AI-generated.

Another threat to GPTZero was GPTZero. Almost immediately after it launched, skeptics on social media started posting embarrassing examples of the tool misclassifying texts. Someone pointed out that it flagged portions of the US Constitution as possibly AI-written. Mockery gave way to outrage when stories of students falsely accused of cheating due to GPTZero began to flood Reddit. At one point, a parent of one such student reached out to Soheil Feizi, a professor of computer science at the University of Maryland. "They were really furious," Feizi said. Last fall, before GPTZero debuted, Feizi and some other Maryland colleagues had begun putting together a research project on the problems with AI detectors, which he'd suspected might not be reliable. Now GPTZero and its imitators got him thinking they could do more harm than good.

Yet another headache for Tian was the number of crafty students finding ways around the detector. One person on Twitter instructed users to insert a zero-width space before every "e" in a ChatGPT-generated text. A TikTok user wrote a program that bypassed detection by replacing certain English letters with their Cyrillic look-alikes. Others started running their AI text through QuillBot, a popular paraphrasing tool. Tian patched these holes, but the workarounds kept coming. It was only a matter of time before someone spun up a rival product—an anti-detector.

"If you can follow a formula to create an essay, it's probably not a good assignment."

IN EARLY MARCH, a Stanford University freshman named Joseph Semrai and some friends were driving down the Pacific Coast Highway to LA when they got locked out of their Zipcar in Ventura. They walked to a nearby Starbucks and waited for roadside assistance. But as the wait dragged on for hours, Semrai and a friend wondered how to make up for the lost time. Semrai had an essay due the following week for a required freshman writing class. It was his least favorite type of assignment: a formulaic essay meant to show logical reasoning. "It's a pretty algorithmic process," says Semrai.

ChatGPT was the obvious solution. But at the time, its responses tended to max out at a few paragraphs, so generating a full-length essay would be a multistep process. Semrai wanted to create a tool that could write the paper in one burst. He also knew there was a chance it could be detected by GPTZero. With the encouragement of his friend, Semrai pulled out his laptop and ginned up a script that would write an essay based on a prompt, run the text through GPTZero, then keep tweaking the phrasing until the AI was no longer detectable—essentially using GPTZero against itself.

Semrai introduced his program a few days later at Friends and Family Demo Day, a kind of show-and-tell for Stanford's undergraduate developer community. Standing before a roomful of classmates, he asked the audience for an essay topic—someone suggested "fine dining" in California—and typed it into the prompt box. After a few seconds, the program spat out an eight-paragraph essay, unoriginal but coherent, with works cited. "Not saying I'd ever submit this paper," Semrai said, to chuckles. "But there you go. I dunno, it saves time." He named the tool WorkNinja and put it on the app store two months later. With the help of a promotional campaign featuring the Gen Z influencer David Dobrik and a giveaway of 10 Teslas to users who signed up, it received more than 350,000 downloads in the first week; sign-ups have slowed since then to a few hundred a day, according to Semrai. (Semrai wouldn't say who funded the campaign, only that it was a major Silicon Valley angel investor.)

Semrai's Zoomer mop and calm demeanor belie a simmering intensity. Whereas Tian bounces and bubbles his way through the world, Semrai comes across as focused and deadpan. The 19-year-old speaks in the confident, podcast-ready tone of a Silicon Valley entrepreneur who sees the world in terms of problems to be solved, ending every other sentence with, "Right?" Listening to him wax on about defensible moats and the "S-curve" of societal growth, it's easy to forget he can't legally drink. But then, occasionally, he'll say something that reveals the wide-eyed undergrad, open to the world and still figuring out his place in it. Like the time he and a friend walked around the Santa Monica pier until 3 am, "talking about what we value." Semrai thinks a lot about how to find balance and happiness. "I think, while I'm young, it probably lies more in exploring the derivative," he says, "chasing the highs and lows."

Growing up in New York and then Florida, his parents—a firefighter father from Yonkers and a homemaker mother from China—gave him a long leash. "I was kinda left during childhood to pursue what genuinely excited me," he said. "The best way to do that was to make stuff on the computer." When Semrai was 6 he created a plug-in to assign permission levels for *Minecraft* servers, and at 7 he wrote a program that patched Windows 7 so you could run Windows XP on it. "It just makes me genuinely happy to ship things for people," he says.

His family moved from Queens to Palm City when he was 9, and Semrai saw the difference between

JOSEPH SEMRAI, 19, CREATOR OF WORKNINJA, A TOOL THAT GENERATES AI-WRITTEN ESSAYS.

the public school systems. The basic computer literacy he'd taken for granted in New York schools was scarce in Florida. He started writing programs to help fill gaps in education—a trajectory that allows him to say, at 19, that he's been "working in ed tech my entire life." Freshman year of high school, he created an online learning platform that won startup funding in a local competition. Prior to Covid, he'd created a digital hall pass system that became the basis for contact tracing and was adopted by 40 school districts in the Southeast.

Semrai is fundamentally a techno-optimist. He says he believes that we should speed the development of technology, including artificial general intelligence, because it will ultimately lead us toward a "post-scarcity" society—a worldview sometimes described as "effective accelerationism." (Not to be confused with effective altruism, which holds that we should take actions that maximize "good" outcomes, however defined.) Semrai's case for WorkNinja rests on its own kind of accelerationist logic. AI writing tools are good, in his view, not because they help kids cheat, but because they'll force schools to revamp their curricula. "If you can follow a formula to create an essay, it's probably not a good assignment," he says. He envisions a future in which every student can get the kind of education once reserved for aristocrats, by way of personalized AI tutoring. When he was first learning how to program, Semrai says, he relied largely on YouTube videos and internet forums to answer his questions. "It would have been easier if there was a tutor to guide me," he says. Now that AI tutors are real, why stand in their way?

I RECENTLY USED WorkNinja to generate a handful of essays, including one about Darwin's theory of evolution. The first version it gave me was clumsy and repetitive, but workable, exploring the theory's implications for biology, genetics, and philosophy. GPTZero flagged it as likely AI-generated.

So I hit WorkNinja's Rephrase button. The text shifted slightly, replacing certain words with synonyms. After three rephrasings, GPTZero finally gave the text its stamp of humanity. (When I tested the same text again a few weeks later, the tool labeled it a mix of human and AI writing.) The problem was, many of the rephrased sentences no longer made sense. For example, the following sentence:

Darwin's theory of evolution is the idea that living species evolve over time due to their interaction with their environment.

had morphed to become:

Darwin's theory of evolution is the thought that living species acquire over clip due to their interaction with their surroundings.

At the very least, any student looking for a shortcut would have to clean up their WorkNinja draft before submitting it. But it points to a real issue: If even this janky work in progress can circumvent detectors, what could a sturdier product accomplish?

In March, Soheil Feizi at the University of Maryland published his findings on the performance of AI detectors. He argued that accuracy problems are inevitable, given the way AI text detectors worked. As you increase the sensitivity of the instrument to catch more AI-generated text, you can't avoid raising the number of false positives to what he considers an unacceptable level. So far, he says, it's impossible to get one without the other. And as the statistical distribution of words in AI-generated text edges closer to that of humans—that is, as it becomes more convincing—he says the detectors will only become less accurate. He also found that paraphrasing baffles AI detectors, rendering their judgments "almost random." "I don't think the future is bright for these detectors," Feizi says.

"Watermarking" doesn't help either, he says. Under this approach, a generative AI tool like ChatGPT proactively adjusts the statistical weights of certain interchangeable "token" words—say, using *start* instead of *begin*, or *pick* instead of *choose*—in a way that would be imperceptible to the reader but easily spottable by an algorithm. Any text in which those words appear with a given frequency could be marked as having been generated by a particular tool. But Feizi argues that with enough paraphrasing, a watermark "can be washed away."

In the meantime, he says, detectors are hurting students. Say a detection tool has a 1 percent false positive rate—an optimistic assumption. That means in a classroom of 100 students, over the course of 10 take-home essays, there will be on average 10 students falsely accused of cheating. (Feizi says a rate of one in 1,000 would be acceptable.) "It's ridiculous to even think about using such tools to police the use of AI models," he says.

Tian says the point of GPTZero isn't to catch cheaters, but that has inarguably been its main use case so far. (GPTZero's detection results now come with a warning: "These results should not be used to punish students.") As for accuracy, Tian says GPTZero's current level is 96 percent when trained on its most recent data set. Other detectors boast higher figures, but Tian says those claims are a red flag, as it means they're "overfitting" their training data to match the strengths of their tools. "You have to put the AI and human on equal footing," he says.

Surprisingly, AI-generated images, videos, and audio snippets are far easier to detect, at least for now, than synthetic text. Reality Defender, a startup backed by Y Combinator, launched in 2018 with a focus on fake image and video detection and has since branched out to audio and text. Intel released a tool called FakeCatcher, which detects deepfake videos by analyzing facial blood flow patterns visible only to the camera. A company called Pindrop uses voice "biometrics" to detect spoofed audio and to authenticate callers in lieu of security questions.

AI-generated text is more difficult to detect because it has relatively few data points to analyze, which means fewer opportunities for AI output to deviate from the human norm. Compare that to Intel's FakeCatcher. Ilke Demir, a research scientist for Intel who has also worked on Pixar films, says it would be extremely difficult to create a data set large and detailed enough to allow deepfakers to simulate blood flow signatures to fool the detector. When I asked whether such a thing could eventually be created, she said her team anticipates future developments in deepfake technology in order to stay ahead of them.

Ben Colman, CEO of Reality Defender, says his company's detection tools are unevadable in part because they're private. (So far, the company's clients have mainly been governments and large corporations.) With publicly available tools like GPTZero, anyone can run a piece of text through the detector and then tweak it until it passes muster. Reality Defender, by contrast, vets every person and institution that uses the tool, Colman says. They also watch out for suspicious usage, so if a particular account were to run tests on the same image over and over with the goal of bypassing detection, their system would flag it.

Regardless, much like spam hunters, spies, vaccine makers, chess cheaters, weapons designers, and the entire cybersecurity industry, AI detectors across all media will have to constantly adapt to new evasion techniques. Assuming, that is, the difference between human and machine still matters.

THE MORE TIME I spent talking with Tian and Semrai and their classmate-colleagues, the more I wondered: Do any of these young people actually ... enjoy writing? "Yeah, a lot!" said Tian, beaming even more than usual when I asked him last May on Princeton's campus. "It's like a puzzle." He likes figuring out how words fit together and then arranging the ideas so they flow. "I feel like that's fun to do." He also loves the interview process, as it gives him "a window into people's lives, plus a mirror into how you live your own."

In high school, Tian says, writing felt like a chore. He credits McPhee for stoking his love and expanding his taste. In June, he told me excitedly that he had just picked up a used copy of Annie Dillard's *The Writing Life*.

Semrai similarly found high school writing assignments boring and mechanistic—more about synthesizing information than making something new. "I'd have preferred open-format assignments that would've sparked creativity," he says. But he put those synthesizing skills to work. Sophomore year, he wrote an 800-page instructional book called *Build for Anything*, intended "to take someone from knowing nothing to knowing a little bit of almost everything" about web development. (He self-published the book on Amazon in 2022 and sold a few hundred copies.) Semrai said it's the kind of prose ChatGPT now excels at. "I don't think the book falls into the category of meaningful writing," he says.

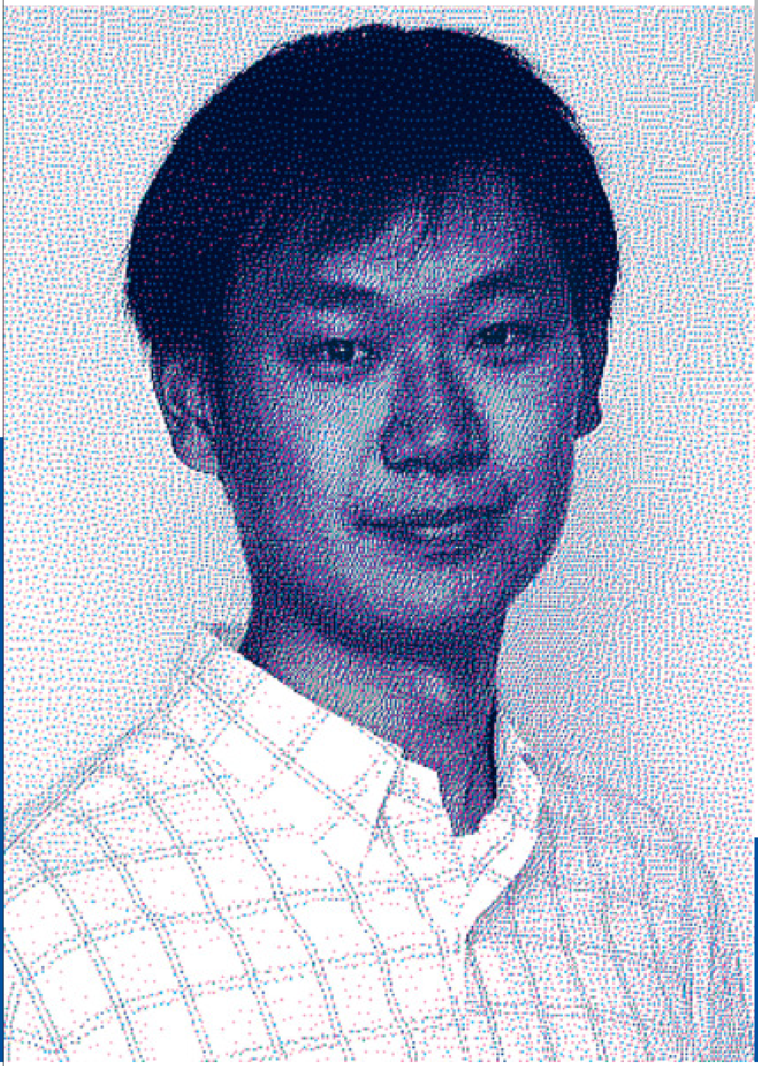
After almost 20 years of typing words for money, I can say from experience, writing sucks. Ask any professional writer and they'll tell you, it's the worst, and it doesn't get easier with practice. I can attest that the enthusiasm and curiosity required to perpetually scan the world, dig up facts, and wring them for meaning can be hard to sustain. And that's before you factor in the state of the industry: dwindling rates, shrinking page counts, and shortening attention spans (readers' and my own). I keep it up because, for better or worse, it's now who I am. I do it not for pleasure but because it feels meaningful—to me at least.

Some writers romanticize the struggle. McPhee once described lying on a picnic table for two weeks, trying to decide how to start an article. "The piece would ultimately consist of some five thousand sentences, but for those two weeks I couldn't write even one," he wrote. Another time, at age 22, he lashed himself to his writing chair with a bathrobe belt. According to Thomas Mann, "A writer is someone for whom writing is more difficult than it is for other people." "You search, you break your heart, your back, your brain, and then—only then—it is handed to you," writes Annie Dillard in *The Writing Life*. She offers this after a long comparison of writing to alligator wrestling.

The implication is that the harder the squeeze, the sweeter the juice—that there's virtue in staring down the empty page, taming it, forcing it to give way to prose. This is how the greatest breakthroughs happen, we tell ourselves. The agony is worth it, because that's how ideas are born.

The siren call of AI says, *It doesn't have to be this way*. And when you consider the billions of people who sit outside the elite club of writer-sufferers, you start to think: Maybe it *shouldn't* be this way.

MAY HABIB SPENT her early childhood in Lebanon before moving to Canada, where she learned English as a second language. "I thought it was pretty unfair that so much benefit would accrue to someone really good at reading and writing," she says. In 2020, she founded Writer, one of several hybrid platforms that aims not to replace human writing, but to help people—and, more accurately, brands—collaborate better with AI.



EDWARD TIAN, AGE 23, CREATOR OF GPTZERO, A TOOL THAT DETECTS AI-GENERATED WRITING.

Habib says she believes there's value in the blank page stare-down. It helps you consider and discard ideas and forces you to organize your thoughts. "There are so many benefits to going through the meandering, head-busting, wanna-kill-yourself staring at your cursor," she says. "But that has to be weighed against the speed of milliseconds."

The purpose of Writer isn't to write for you, she says, but rather to make your writing faster, stronger, and more consistent. That could mean suggesting edits to prose and structure, or highlighting what else has been written on the subject and offering counterarguments. The goal, she says, is to help users focus less on sentence-level mechanics and more on the ideas they're trying to communicate. Ideally, this process yields a piece of text that's just as "human" as if the person had written it entirely themselves. "If the detector can flag it as AI writing, then you've used the tools wrong," she says.

The black-and-white notion that writing is either human- or AI-generated is already slipping away, says Ethan Mollick, a professor at the Wharton School of the University of Pennsylvania. Instead, we're entering an era of what he calls "centaur writing." Sure, asking ChatGPT to spit out an essay about the history of the Mongol Empire produces predictably "AI-ish" results, he says. But "start writing, 'The details in paragraph three aren't quite right—add this information, and make the tone more like *The New Yorker*,'" he says. "Then it becomes more of a hybrid work and much better-quality writing."

Mollick, who teaches entrepreneurship at Wharton, not only allows his students to use AI tools—he requires it. "Now my syllabus says you have to do at least one impossible thing," he says. If a student can't code, maybe they write a working program. If they've never done design work, they might put together a visual prototype. "Every paper you turn in has to be critiqued by at least four famous entrepreneurs you simulate," he says.

Students still have to master their subject area to get good results, according to Mollick. The goal is to get them thinking critically and creatively: "I don't care what tool they're using to do it, as long as they're using the tools in a sophisticated manner and using their mind."

Mollick acknowledges that ChatGPT isn't as good as the best human writers. But it can give everyone else a leg up. "If you were a bottom-quartile writer, you're in the 60 to 70th percentile now," he says. It also frees certain types of thinkers from the tyranny of the writing process. "We equate writing ability with intelligence, but that's not always true," he says. "In fact, I'd say it's often not true."

ON A CLOUDLESS DAY in May, Tian and I strolled across Princeton's campus; big white reunion tents seemed to have landed like spaceships on the manicured lawns. At my request, Tian invited a handful of classmates to join us for lunch at a Szechuan restaurant off campus and talk about AI.

As some schools rushed to ban ChatGPT and tech CEOs signed letters warning of AI-fueled doom, the students were notably relaxed about a machine-assisted future. (Princeton left it up to professors to set their own ground rules.) One had recently used ChatGPT to write the acknowledgments section of her thesis. Others, including Tian, relied on it to fill in chunks of script while coding. Lydia You, a senior and computer science major who plans to work in journalism, had asked ChatGPT to write a poem about losing things in the style of Elizabeth Bishop—an attempt to re-create her famous poem "One Art." ("The art of losing isn't hard to master.") The result was "very close" to the original poem, You said, and she found that the chatbot did an even better job analyzing the original and describing what made it so moving. "We've seen a lot of panic about almost everything in our lives," said You, citing TikTok, Twitter, and the internet itself. "I feel like people of our generation are like, *We can figure out for ourselves how to use this*."

Sophie Amiton, a senior studying mechanical and aerospace engineering, jumped in: "Also, I think our generation is lazier in a lot of ways," she said, as You nodded in agreement. "I see a lot more people who don't want traditional jobs now, don't want a nine-to-five."

"They're disillusioned," You said. "A lot of jobs are spreadsheets."

"I think that came out of Covid," Amiton continued. "People reevaluated what the purpose of work even is, and if you can use ChatGPT to make your life easier, and therefore have a better quality of life or work-life balance, then why not use the shortcut?"

Liz, a recent Princeton graduate who preferred not to use her surname, sent me a paper she'd written with the help of ChatGPT for a class on global politics. Rather than simply asking it to answer the essay question, she plugged in an outline with detailed bullet points, then had it write the paper based on her notes. After extensive back-and-forth—telling it to rewrite and rearrange, add nuance here and context there—she finally had a paper she was comfortable submitting. She got an A.

I copied and pasted her paper into GPTZero. The verdict: "Your text is likely to be written entirely by a human."

In early May, just a few weeks before Tian and his classmates put on their black graduation gowns, the GPTZero team released the Chrome plug-in they'd been developing and called it Origin. Origin is still rudimentary: You have to select the text of a web page yourself, and its accuracy isn't perfect. But Tian hopes that one day the tool will automatically scan every website you look at, highlighting AI-generated content—from text to images to video—as well as anything "toxic" or factually dubious. He describes Origin as a "windshield" for the information superhighway, deflecting useless or harmful material and allowing us to see the road clearly.

"If the detector can flag it as AI writing, then you've used the tools wrong."

Tian was unflaggingly optimistic about the company; he also just felt fortunate to be graduating into a job he actually wanted. Many of his friends had entered Princeton planning to be entrepreneurs, but belt-tightening in the tech sector had changed their plans.

As a rising sophomore with three years left to go at Stanford, Semrai approached the summer with a more footloose attitude. On a blistering Thursday afternoon in June, on the rooftop of Pier 17 near Wall Street, Semrai, wearing a green patterned shirt and white Nikes, spoke to me brightly about the future—or at least the next few weeks. His summer was still taking shape. ("I'm rapidly thesis-testing.") But for now he was in New York, crashing with friends while cranking on a couple of AI-driven projects. The previous night, he'd slept in a coworking space in SoHo. Now he was standing in the shaded VIP section of an event put on by Techstars New York City, a startup accelerator, while hundreds of sweat-stained attendees milled around in the glare.

Nearby, New York City mayor Eric Adams stood onstage wearing aviators and a full suit, praising the glories of coding. "I'm a techie," Adams said, before encouraging guests to seek out diverse collaborators and use "source code" to fix societal problems like cancer and gun violence. He then urged the singles in the crowd to find themselves a "shorty or a boo and hook up with them."

Semrai was taking a see-what-sticks approach to building. In addition to WorkNinja, he was developing a platform for chatbots based on real celebrities and trained on reams of their data, with which fans could then interact. He was also prototyping a bracelet that would record everything we say and do—Semrai calls it a "perfect memory"—and offer real-time tips to facilitate conversations. (A group of classmates at Stanford recently created a related product called RizzGPT, an eyepiece that helps its wearer flirt.)

He expected the summer to give rise to an explosion of AI apps, as young coders mix and cross-pollinate. (Eric Adams would approve.) "I think a constellation of startups will be formed, and five years from now we'll be able to draw lines between people—the start of an ecosystem," he said.

By summer, Tian had a team of 12 employees and had raised $3.5 million from a handful of VCs, including Jack Altman (brother of OpenAI CEO Sam Altman) and Emad Mostaque of Stability AI. But over the course of our conversations, I noticed that his framing of GPTZero/Origin was shifting slightly. Now, he said, AI-detection would be only one part of the humanity-proving toolkit. Just as important would be an emphasis on provenance, or "content credentials." The idea is to attach a cryptographic tag to a piece of content that verifies it was created by a human, as determined by its process of creation—a sort of captcha for digital files. Adobe Photoshop already attaches a tag to photos that harness its new AI generation tool, Firefly. Anyone looking at an image can right-click it and see who made it, where, and how. Tian says he wants to do the same thing for text and that he has been talking to the Content Authenticity Initiative—a consortium dedicated to creating a provenance standard across media—as well as Microsoft, about working together.

One could interpret his emphasis on provenance as a tacit acknowledgment that detection alone won't cut it. (OpenAI shut down its text classifier in July "due to its low rate of accuracy.") It also previews a possible paradigm shift in how we relate to digital media. The whole endeavor of detection suggests that humans leave an unmistakable signature in a piece of text—something perceptible—much the way that a lie detector presumes dishonesty leaves an objective trace. Provenance relies on something more like a "Made in America" label. If it weren't for the label, we wouldn't know the difference. It's a subtle but meaningful distinction: Human writing may not be better, or more creative, or even more original. But it will be human, which will matter to other humans.

In June, Tian's team took another step in the direction of practicality. He told me they were building a new writing platform called HumanPrint, which would help users improve their AI-written text and enable them to share "proof of authenticity." Not by generating text, though. Rather, it would use GPTZero's technology to highlight sections of text that were insufficiently human and prompt the user to rewrite it in their own words—a sort of inversion of the current AI writing assistants. "So teachers can specify, OK, maybe more than 50 percent of the essay should still be written in your own words," he said. When I asked whether this was a pivot for the company, Tian argued that it was "a natural extension of detection." "It was always a vision of being the gold standard of responsible AI usage," Tian said, "and that's still there." Still, the implication is clear: There's no stopping AI writing; the only option is to work with it.

WHEN TIAN WAS FIRST testing out GPTZero, he scanned a 2015 *New Yorker* essay by McPhee called "Frame of Reference." In it, McPhee riffs on the joys and risks of making cultural references in one's writing. "Mention Beyoncé and everyone knows who she is. Mention Veronica Lake and you might as well be in the Quetico-Superior," he writes coyly. He runs down a list of adjectives he's used to describe mustaches, including "sincere," "no-nonsense," "gyroscopic," "guileless," "analgesic," "soothing," "odobene," and "tetragrammatonic." He concludes with an anecdote about battling an editor to include a reference to an obscure British term used by upper-class tourists to India during the Raj. (He won.) It's classic McPhee: scalpel-precise, big-hearted if a tad self-satisfied, gleefully digressive, indulgent until he gets to the just-right point. GPTZero determined that the article was "the most human on all metrics," Tian said. I called McPhee to ask what he thought it meant that his writing was especially human.

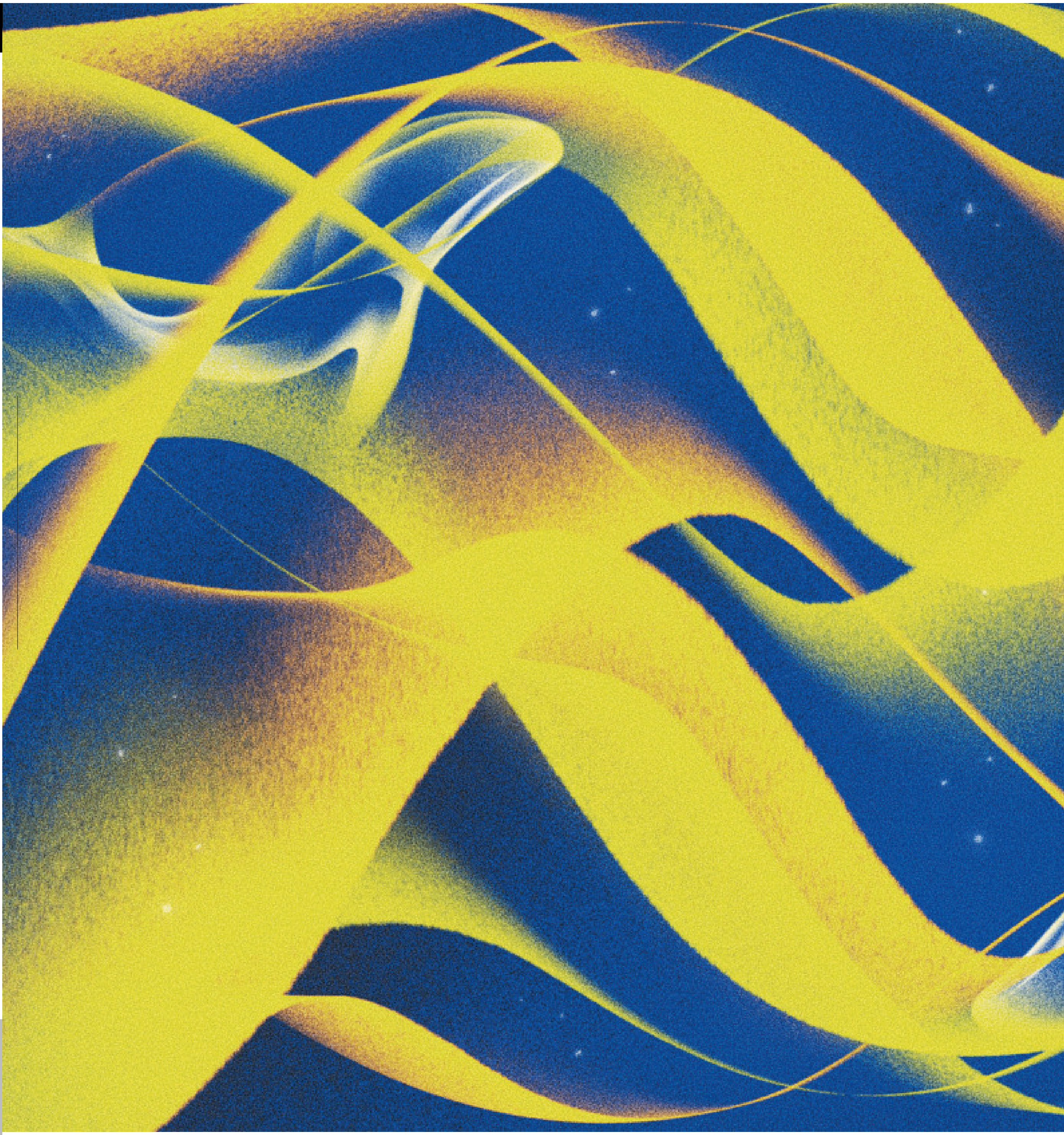
"I really have no very good idea," McPhee told me over the phone. "But if I were guessing, it's that my pieces get at the science, or the agriculture, or the aviation, or whatever the topic is, through people. There's always a central figure I learn from." Indeed, McPhee writes through the eyes of experts. The reader comes away with not just some esoteric knowledge about geology or particle physics or oranges, but a sense of the person studying the subject, as well as McPhee studying the person.

McPhee, now 92 , said he's unconcerned about AI replacing human writers. "I'm extremely skeptical and not the least bit worried about it," he said. "I don't think there's a Mark Twain of artificial intelligence."

But, I asked, what if years from now, someone designs a McPheeBot3000 trained on McPhee's writing, and then asks it to produce a book on a fresh topic? It might not be able to ford streams with environmental activists or go fly-fishing with ichthyologists, but couldn't it capture McPhee's voice and style and worldview? Tian argued that machines can only imitate, while McPhee never repeats himself: "What's unique about McPhee is he comes up with things McPhee a day ago wouldn't have."

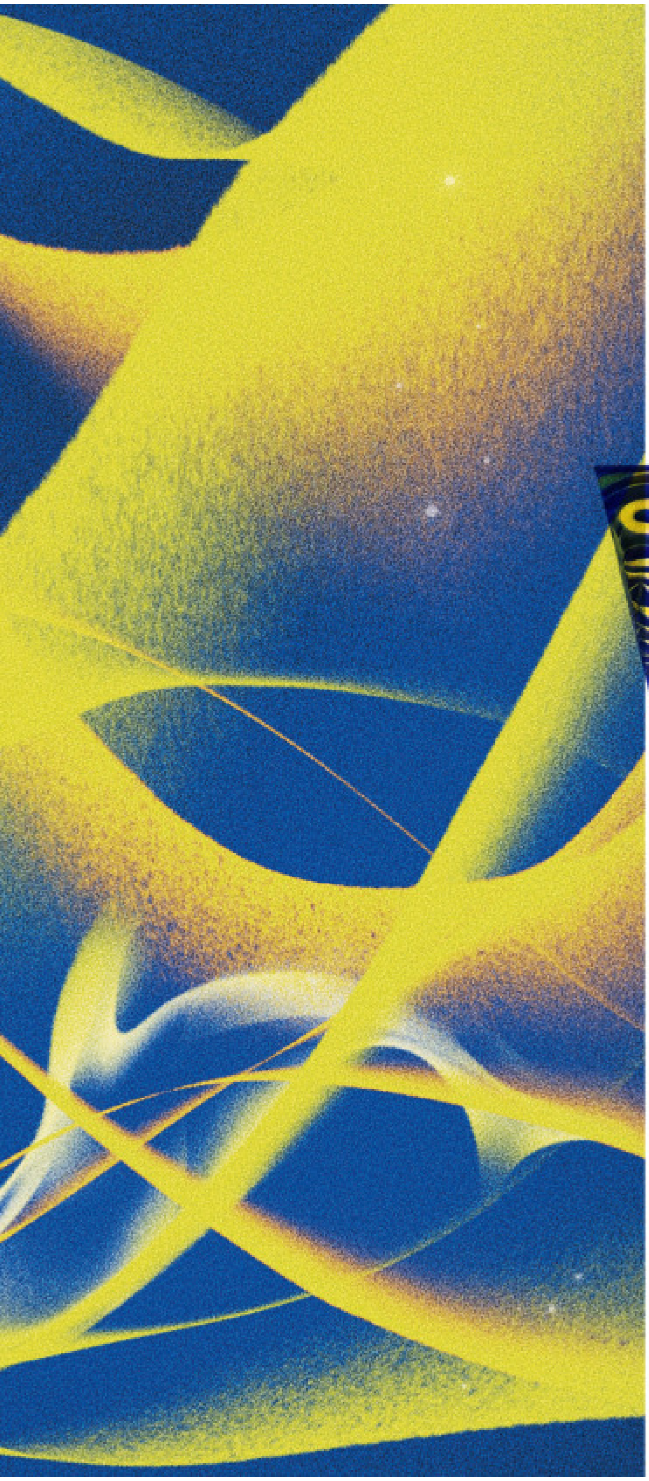
I asked McPhee about the hypothetical McPheeBot3000. (Or, if Semrai has his way, not-so-hypothetical.) "If this thing ever happens, in a future where I'm no longer here," he said, "I hope my daughters show up with a lawyer."

CHRISTOPHER BEAM *is a writer living in Brooklyn.*



WITH ECOSYSTEMS IN CRISIS, AI ENGINEERS AND SCIENTISTS ARE TEAMING UP TO DECIPHER WHAT ANIMALS ARE SAYING. THEIR HOPE: BY TRULY LISTENING TO NATURE, HUMANS WILL SAVE IT.

BY CAMILLE BROMLEY
ILLUSTRATIONS BY AGNES JONAS



CALLS OF THE WILD

Wild

BEFORE MICHELLE FOURNET moved to Alaska on a whim in her early twenties, she'd never seen a whale. She took a job on a whale watching boat and, each day she was out on the water, gazed at the grand shapes moving under the surface. For her entire life, she realized, the natural world had been out there, and she'd been missing it. "I didn't even know I was bereft," she recalls. Later, as a graduate student in marine biology, Fournet wondered what else she was missing. The humpbacks she was getting to know revealed themselves in partial glimpses. What if she could hear what they were saying? She dropped a hydrophone in the water—but the only sound that came through was the mechanical churn of boats. The whales had fallen silent amid the racket. Just as Fournet had discovered nature, then, she was witnessing it recede. She resolved to help the whales. To do that, she needed to learn how to listen to them.

Fournet, now a professor at the University of New Hampshire and the director of a collective of conservation scientists, has spent the past decade building a catalog of the various chirps, shrieks, and groans that humpbacks make in daily life. The whales have huge and diverse vocabularies, but there is one thing they all say, whether male or female, young or old. To our meager human ears, it sounds something like a belly rumble punctuated by a water droplet: *whup.*

Fournet thinks the whup call is how the whales announce their presence to one another. A way of saying, "I'm here." Last year, as part of a series of experiments to test her theory, Fournet piloted a skiff out into Alaska's Frederick Sound, where humpbacks gather to feed on clouds of krill. She broadcast a sequence of whup calls and recorded what the whales did in response. Then, back on the beach, she put on headphones and listened to the audio. Her calls went out. The whales' voices returned through the water: *whup, whup, whup*. Fournet describes it like this: The whales heard a voice say, "I am, I am here, I am me." And they replied, "I also am, I am here, I am me."

Biologists use this type of experiment, called a playback, to study what prompts an animal to speak. Fournet's playbacks have so far used recordings of real whups. The method is imperfect, though, because humpbacks are highly attentive to who they're talking to. If a whale recognizes the voice of the whale in the recording, how does that affect its response? Does it talk to a buddy differently than it would to a stranger? As a biologist, how do you ensure you're sending out a neutral whup?

One answer is to create your own. Fournet has shared her catalog of humpback calls with the Earth Species Project, a group of technologists and engineers who, with the help of AI, are aiming to develop a synthetic whup. And they're not just planning to emulate a humpback's voice. The nonprofit's mission is to open human ears to the chatter of the entire animal kingdom. In 30 years, they say, nature documentaries won't need soothing Attenborough-style narration, because the dialog of the animals onscreen will be subtitled. And just as engineers today don't need to know Mandarin or Turkish to build a chatbot in those languages, it will soon be possible to build one that speaks Humpback—or Hummingbird, or Bat, or Bee.

The idea of "decoding" animal communication is bold, maybe unbelievable, but a time of crisis calls for bold and unbelievable measures. Everywhere that humans are, which is everywhere, animals are vanishing. Wildlife populations across the planet have dropped an average of nearly 70 percent in the past 50 years, according to one estimate—and that's just the portion of the crisis that scientists have measured. Thousands of species could disappear without humans knowing anything about them at all.

To decarbonize the economy and preserve ecosystems, we certainly don't need to talk to animals. But the more we know about the lives of other creatures, the better we can care for those lives. And humans, being human, pay more attention to those who speak our language. The interaction that Earth Species wants to make possible, Fournet says, "helps a society that is disconnected from nature to reconnect with it." The best technologies give humans a way to inhabit the world more fully. In that light, talking to animals could be its most natural application yet.

■ **HUMANS HAVE ALWAYS** known how to listen to other species, of course. Fishers throughout history collaborated with whales and dolphins to mutual benefit: a fish for them, a fish for us. In 19th-century Australia, a pod of killer whales was known to herd baleen whales into a bay near a whalers' settlement, then slap their tails to alert the humans to ready the harpoons. (In exchange for their help, the orcas got first dibs on their favorite cuts, the lips and tongue.) Meanwhile, in the icy waters of Beringia, Inupiat people listened and spoke to bowhead whales before their hunts. As the environmental historian Bathsheba Demuth writes in her book *Floating Coast*, the Inupiat thought of the whales as neighbors occupying "their own country" who chose at times to offer their lives to humans—if humans deserved it.

Commercial whalers had a different approach. They saw whales as floating containers of blubber and baleen. The American whaling industry

PREVIOUS PAGE: AI TYPOGRAPHY: DEV VALLADARES + DEFORUM STABLE DIFFUSION

in the mid-19th century, and then the global whaling industry in the following century, very nearly obliterated several species, resulting in one of the largest-ever losses of wild animal life caused by humans. In the 1960s, 700,000 whales were killed, marking the peak of cetacean death. Then, something remarkable happened: We heard whales sing. On a trip to Bermuda, the biologists Roger and Katy Payne met a US naval engineer named Frank Watlington, who gave them recordings he'd made of strange melodies captured deep underwater. For centuries, sailors had recounted tales of eerie songs that emanated from their boats' wooden hulls, whether from monsters or sirens they didn't know. Watlington thought the sounds were from humpback whales. Go save them, he told the Paynes. They did, by releasing an album, *Songs of the Humpback Whale*, that made these singing whales famous. The Save the Whales movement took off soon after. In 1972, the US passed the Marine Mammal Protection Act; in 1986, commercial whaling was banned by the International Whaling Commission. In barely two decades, whales had transformed in the public eye into cognitively complex and gentle giants of the sea.

Roger Payne, who died earlier this year, spoke frequently about his belief that the more the public could know "curious and fascinating things" about whales, the more people would care what happened to them. In his opinion, science alone would never change the world, because humans don't respond to data; they respond to emotion—to things that make them weep in awe or shiver with delight. He was in favor of wildlife tourism, zoos, and captive dolphin shows. However compromised the treatment of individual animals might be in these places, he believed, the extinction of a species is far worse. Conservationists have since held on to the idea that contact with animals can save them.

From this premise, Earth Species is taking the imaginative leap that AI can help us make *first* contact with animals. The organization's founders, Aza Raskin and Britt Selvitelle, are both architects of our digital age. Raskin grew up in Silicon Valley; his father started Apple's Macintosh project in the 1970s. Early in his career, Raskin helped to build Firefox, and in 2006 he created the infinite scroll, arguably his greatest and most dubious legacy. Repentant, he later calculated the collective human hours that his invention had wasted and arrived at a figure surpassing 100,000 lifetimes per week.

Raskin would sometimes hang out at a startup called Twitter, where he met Selvitelle, a founding employee. They stayed in touch. In 2013, Raskin heard a news story on the radio about gelada monkeys in Ethiopia whose communication had similar cadences to human speech. So similar, in fact, that the lead scientist would sometimes hear a voice talking to him, turn around, and be surprised to find a monkey there. The interviewer asked whether there was any way of knowing what they were trying to say. There wasn't—but Raskin wondered if it might be possible to arrive at an answer with machine learning. He brought the idea up with Selvitelle, who had an interest in animal welfare.

For a while the idea was just an idea. Then, in 2017, new research showed that machines could translate between two languages without first being trained on bilingual texts. Google Translate had always mimicked the way a human might use a dictionary, just faster and at scale. But these new machine learning methods bypassed semantics altogether. They treated languages as geometric shapes and found where the shapes overlapped. If a machine could translate any language into English without needing to understand it first, Raskin thought, could it do the same with a gelada monkey's wobble, an elephant's infrasound, a bee's waggle dance? A year later, Raskin and Selvitelle formed Earth Species.

Raskin believes that the ability to eavesdrop on animals will spur nothing less than a paradigm shift as historically significant as the Copernican revolution. He is fond of saying that "AI is the invention of modern optics." By this he means that just as improvements to the telescope allowed 17th-century astronomers to perceive newfound stars and finally displace the Earth from the center of the cosmos, AI will help scientists hear what their ears alone cannot: that animals speak meaningfully, and in more ways than we can imagine. That their abilities, and their lives, are not less than ours. "This time we're going to look out to the universe and discover humanity is not the center," Raskin says.

Raskin and Selvitelle spent their first few years meeting with biologists and tagging along on fieldwork. They soon realized that the most obvious and immediate need in front of them wasn't inciting revolution. It was sorting data. Two decades ago, a primate researcher would stand under a tree and hold a microphone in the air until her arm got tired. Now researchers can stick a portable biologger to a tree and collect a continuous stream of audio for a year. The many terabytes of data that result is more than any army of grad students could hope to tackle. But feed all this material to trained machine learning algorithms, and the computer can scan the data and flag the animal calls. It can distinguish a whup from a whistle. It can tell a gibbon's voice from her brother's. At least, that's the hope. These tools need more data, research, and funding. Earth Species has a workforce of 15 people and a budget of a few million dollars. They've teamed up with several dozen biologists to start making headway on these practical tasks.

An early project took on one of the most significant challenges in animal communication research, known as the cocktail party problem: When a group of animals are talking to one another, how can you tell who's saying what? In the open sea, schools of dolphins a thousand strong chatter all at once; scientists who record them end up with audio as dense with whistles and clicks as a stadium is with cheers. Even audio of just two or three animals is often unusable, says Laela Sayigh, an expert in bottlenose dolphin whistles, because you can't tell where one dolphin stops talking and another starts. (Video doesn't help, because dolphins don't open their mouths when they speak.) Earth Species used Sayigh's extensive database of signature whistles—the ones likened to names—to develop a neural network model that could separate overlapping animal voices. That model was useful only in lab conditions, but research is meant to be built on. A couple of months later, Google AI published a model for untangling wild birdsong.

Sayigh has proposed a tool that can serve as an emergency alert for dolphin mass strandings, which tend to recur in certain places around the globe. She lives in Cape Cod, Massachusetts, one such hot spot, where as often as a dozen times a year groups of dolphins get disoriented, inadvertently swim onto shore, and perish. Fortunately, there might be a way to predict this before it happens, Sayigh says. She hypothesizes that when the dolphins are stressed, they emit signature whistles more than usual, just as someone lost in a snowstorm might call out in panic. A computer trained to listen for these whistles could send an alert that prompts rescuers to reroute the dolphins before they hit the beach. In the Salish Sea—where, in 2018, a mother orca towing the body of her starved calf attracted global sympathy—there is an alert system, built by Google AI, that listens for resident killer whales and diverts ships out of their way.

For researchers and conservationists alike, the potential applications of machine learning are basically limitless. And Earth Species is not the only group working on decoding animal communication. Payne spent the last months of his life advising for Project CETI, a nonprofit that built a base in Dominica this year for the study of sperm whale communication. "Just imagine what would be possible if we understood what animals are saying to each other; what occupies their thoughts; what they love, fear, desire, avoid, hate, are intrigued by, and treasure," he wrote in *Time* in June.

Many of the tools that Earth Species has developed so far offer more in the way of groundwork than immediate utility. Still, there's a lot of optimism in this nascent field. With enough resources, several biologists told me, decoding is scientifically achievable. That's only the beginning. The real hope is to bridge the gulf in understanding between an animal's experience and ours, however vast—or narrow—that might be.

AI will help scientists finally hear what their human ears alone cannot: that animals speak meaningfully, and in more ways than we could imagine.

■ **ARI FRIEDLAENDER HAS** something that Earth Species needs: lots and lots of data. Friedlaender researches whale behavior at UC Santa Cruz. He got started as a tag guy: the person who balances at the edge of a boat as it chases a whale, holds out a long pole with a suction-cupped biologging tag attached to the end, and slaps the tag on a whale's back as it rounds the surface. This is harder than it seems. Friedlaender proved himself adept—"I played sports in college," he explains—and was soon traveling the seas on tagging expeditions.

The tags Friedlaender uses capture a remarkable amount of data. Each records not only GPS location, temperature, pressure, and sound, but also high-definition video and three-axis accelerometer data, the same tech that a Fitbit uses to count your steps or measure how deeply you're sleeping. Taken together, the data illustrates, in cinematic detail, a day in the life of a whale: its every breath and every dive, its traverses through fields of sea nettles and jellyfish, its encounters with twirling sea lions.

Friedlaender shows me an animation he has made from one tag's data. In it, a whale descends and loops through the water, traveling a multicolored three-dimensional course as if on an undersea Mario Kart track. Another animation depicts several whales blowing bubble nets, a feeding strategy in which they swim in circles around groups of fish, trap the fish in the center with a wall of bubbles, then lunge through, mouths gaping. Looking at the whales' movements, I notice that while most of them have traced a neat spiral, one whale has produced a tangle of clumsy zigzags. "Probably a young animal," Friedlaender says. "That one hasn't figured things out yet."

Friedlaender's multifaceted data is especially useful for Earth Species because, as any biologist will tell you, animal communication isn't purely verbal. It involves gestures and movement just as often as vocalizations. Diverse data sets get Earth Species closer to developing algorithms that can work across the full spectrum of the animal kingdom. The organization's most recent work focuses on foundation models, the same kind of computation that powers generative AI like ChatGPT. Earlier this year, Earth Species published the first foundation model for animal communication. The model can already accurately sort beluga whale calls, and Earth Species plans to apply it to species as disparate as orangutans (who bellow), elephants (who send seismic rumbles through the ground), and jumping spiders (who vibrate their legs). Katie Zacarian, Earth Species' CEO, describes the model this way: "Everything's a nail, and it's a hammer."

Another application of Earth Species' AI is generating animal calls, like an audio version of GPT. Raskin has made a few-second chirp of a chiffchaff bird. If this sounds like it's getting ahead of decoding, it is—AI, as it turns out, is better at speaking than understanding. Earth Species is finding that the tools it is developing will likely have the ability to talk to animals even before they can decode. It may soon be possible, for example, to prompt an AI with a whup and have it continue a conversation in Humpback—without human observers knowing what either the machine or the whale is saying.

No one is expecting such a scenario to actually take place; that would be scientifically irresponsible, for one thing. The biologists working with Earth Species are motivated by knowledge, not dialog for the sake of it. Felix Effenberger, a senior AI research adviser for Earth Species, told me: "I don't believe that we will have an English-Dolphin translator, OK? Where you put English into your smartphone and then it makes dolphin sounds and the dolphin goes off and fetches you some sea urchin. The goal is to first discover basic patterns of communication."

So what will talking to animals look—sound—like? It needn't be a free-form conversation to be astonishing. Speaking to animals in a controlled way, as with Fournet's playback whups, is probably essential for scientists to try to understand them. After all, you wouldn't try to learn German by going to a party in Berlin and sitting mutely in a corner.

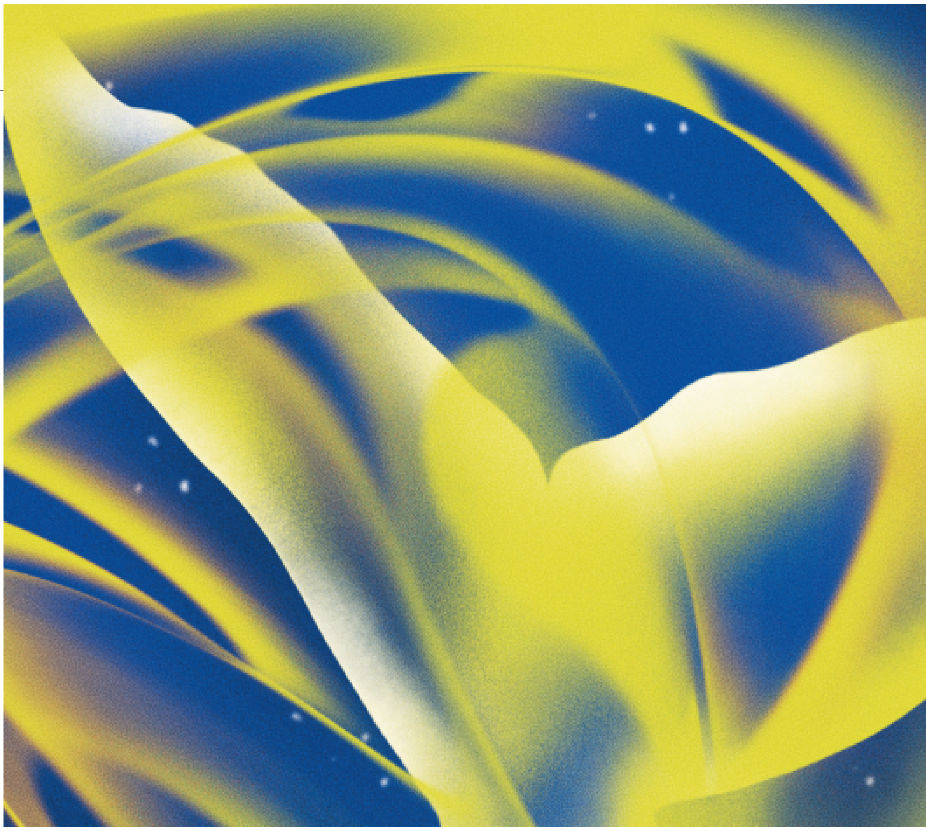
Bird enthusiasts already use apps to snatch melodies out of the air and identify which species is singing. With an AI as your animal interpreter, imagine what more you could learn. You prompt it to make the sound of two humpbacks meeting, and it produces a whup. You prompt it to make the sound of a calf talking to its mother, and it produces a whisper. You prompt it to make the sound of a lovelorn male, and it produces a song.

NO SPECIES OF whale has ever been driven extinct by humans. This is hardly a victory. Numbers are only one measure of biodiversity. Animal lives are rich with all that they are saying and doing—with culture. While humpback populations have rebounded since their lowest point a half-century ago, what songs, what practices, did they lose in the meantime? Blue whales, hunted down to a mere 1 percent of their population, might have lost almost everything.

Christian Rutz, a biologist at the University of St. Andrews, believes that one of the essential tasks of conservation is to preserve nonhuman ways of being. "You're not asking, 'Are you there or are you not there?'" he says. "You are asking, 'Are you there and happy, or unhappy?'"

Rutz is studying how the communication of Hawaiian crows has changed since 2002, when they went extinct in the wild. About 100 of these remarkable birds—one of few species known to use tools—are alive in protective captivity, and conservationists hope to eventually reintroduce them to the wild. But these crows may not yet be prepared. There is some evidence that the captive birds have forgotten useful vocabulary, including calls to defend their territory and warn of predators. Rutz is working with Earth Species to build an algorithm to sift through historical recordings of the extinct wild crows, pull out all the crows' calls, and label them. If they find that calls were indeed lost, conservationists might generate those calls to teach them to the captive birds.

Rutz is careful to say that generating calls will be a decision made thoughtfully, when the time requires it. In a paper published in *Science* in July, he praised the extraordinary usefulness of machine learning. But he cautions that humans should think hard before intervening in animal lives. Just as AI's potential remains unknown, it may carry risks that extend beyond what we can imagine. Rutz cites as an example the new songs composed each year by humpback whales that spread across the world like hit singles. Should these whales pick up on an AI-generated phrase and incorporate that into their routine, humans would be altering a million-year-old culture. "I think that is one of the systems that should be off-limits, at least for now," he told me. "Who has the right to have a chat with a humpback whale?"

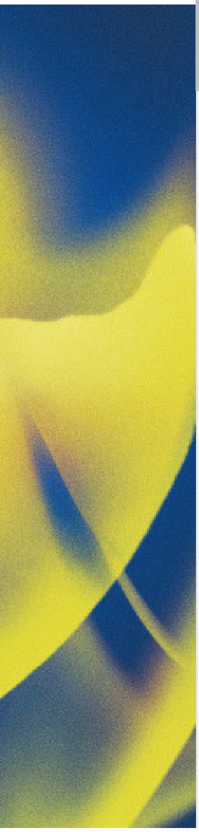


It's not hard to imagine how AI that speaks to animals could be misused. Twentieth-century whalers employed the new technology of their day, too, emitting sonar at a frequency that drove whales to the surface in panic. But AI tools are only as good or bad as the things humans do with them. Tom Mustill, a conservation documentarian and the author of *How to Speak Whale*, suggests giving animal-decoding research the same resources as the most championed of scientific endeavors, like the Large Hadron Collider, the Human Genome Project, and the James Webb Space Telescope. "With so many technologies," he told me, "it's just left to the people who have developed it to do what they like until the rest of the world catches up. This is too important to let that happen."

Billions of dollars are being funneled into AI companies, much of it in service of corporate profits: writing emails more quickly, creating stock photos more efficiently, delivering ads more effectively. Meanwhile, the mysteries of the natural world remain. One of the few things scientists know with certainty is how much they don't know. When I ask Friedlaender whether spending so much time chasing whales has taught him much about them, he tells me he sometimes gives himself a simple test: After a whale goes under the surface, he tries to predict where it will come up next. "I close my eyes and say, 'OK, I've put out 1,000 tags in my life, I've seen all this data. The whale is going to be over *here*.' And the whale's always over *there*," he says. "I have no idea what these animals are doing."

IF YOU COULD speak to a whale, what would you say? Would you ask White Gladis, the killer whale elevated to meme status this summer for sinking yachts off the Iberian coast, what motivated her rampage—fun, delusion, revenge? Would you tell Tahlequah, the mother orca grieving the death of her calf, that you, too, lost a child? Payne once said that if given the chance to speak to a whale,

GETTY IMAGES (PHOTO SOURCE)



he'd like to hear its normal gossip: loves, feuds, infidelities. Also: "*Sorry* would be a good word to say."

Then there is that thorny old philosophical problem. The question of *umwelt*, and what it's like to be a bat, or a whale, or you. Even if we could speak to a whale, would we understand what it says? Or would its perception of the world, its entire ordering of consciousness, be so alien as to be unintelligible? If machines render human languages as shapes that overlap, perhaps English is a doughnut and Whalish is the hole.

Maybe, before you can speak to a whale, you must know what it is like to have a whale's body. It is a body 50 million years older than our body. A body shaped to the sea, to move effortlessly through crushing depths, to counter the cold with sheer mass. As a whale, you choose when to breathe, or not. Mostly you are holding your breath. Because of this, you cannot smell or taste. You do not have hands to reach out and touch things with. Your eyes are functional, but sunlight penetrates water poorly. Usually you can't even make out your own tail through the fog.

You would live in a cloud of hopeless obscurity were it not for your ears. Sound travels farther and faster through water than through air, and your world is illuminated by it. For you, every dark corner of the ocean rings with sound. You hear the patter of rain on the surface, the swish of krill, the blasts of oil drills. If you're a sperm whale, you spend half your life in the pitch black of the deep sea, hunting squid by ear. You use sound to speak, too, just as humans do. But your voice, rather than dissipating instantly in the thin substance of air, sustains. Some whales can shout louder than a jet engine, their calls carrying 10,000 miles across the ocean floor.

But what is it like to *be* you, a whale? What thoughts do you think, what feelings do you feel? These are much harder things for scientists to know. A few clues come from observing how you talk to your own kind. If you're born into a pod of killer whales, close-knit and xenophobic, one of the first things your mother and your grandmother teach you is your clan name. To belong must feel essential. (Remember Keiko, the orca who starred in the film *Free Willy*: When he was released to his native waters late in life, he failed to rejoin the company of wild whales and instead returned to die among humans.) If you're a female sperm whale, you click to your clanmates to coordinate who's watching whose baby; meanwhile, the babies babble back. You live on the go, constantly swimming to new waters, cultivating a disposition that is nervous and watchful. If you're a male humpback, you spend your time singing alone in icy polar waters, far from your nearest companion. To infer loneliness, though, would be a human's mistake. For a whale whose voice reaches across oceans, perhaps distance does not mean solitude. Perhaps, as you sing, you are always in conversation.

■ **MICHELLE FOURNET WONDERS:** *How do we know whales would want to talk to us anyway?* What she loves most about humpbacks is their indifference. "This animal is 40 feet long and weighs 75,000 pounds, and it doesn't give a shit about you," she told me. "Every breath it takes is grander than my entire existence." Roger Payne observed something similar. He considered whales the only animal capable of an otherwise impossible feat: making humans feel small.

Early one morning in Monterey, California, I boarded a whale watching boat. The water was slate gray with white peaks. Flocks of small birds skittered across the surface. Three humpbacks appeared, backs rounding neatly out of the water. They flashed some tail, which was good for the group's photographers. The fluke's craggy ridgeline can be used, like a fingerprint, to distinguish individual whales.

Later, I uploaded a photo of one of the whales to Happywhale. The site identifies whales using a facial recognition algorithm modified for flukes. The humpback I submitted, one with a barnacle-encrusted tail, came back as CRC-19494. Seventeen years ago, this whale had been spotted off the west coast of Mexico. Since then, it had made its way up and down the Pacific between Baja and Monterey Bay. For a moment, I was impressed that this site could so easily fish an animal out of the ocean and deliver me a name. But then again, what did I know about this whale? Was it a mother, a father? Was this whale on Happywhale actually happy? The AI had no answers. I searched the whale's profile and found a gallery of photos, from different angles, of a barnacled fluke. For now, that was all I could know.

CAMILLE BROMLEY *is a features editor for* WIRED.

CONFESSIONS OF A VIRAL AI WRITER

AI-GENERATED STORIES CAN ACTUALLY BE GOOD. I SHOULD KNOW: I WROTE ONE. BUT AS I KEPT EXPERIMENTING, I BECAME LESS SURE THIS WAS GOOD FOR WRITERS—OR FOR WRITING ITSELF.

BY VAUHINI VARA
ILLUSTRATIONS BY QIANHUI YU

SIX OR SEVEN years ago, I realized I should learn about artificial intelligence. I'm a journalist, but in my spare time I'd been writing a speculative novel set in a world ruled by a corporate, AI-run government. The problem was, I didn't really understand what a system like that would look like.

I started pitching articles that would give me an excuse to find out, and in 2017 I was assigned to profile Sam Altman, a cofounder of OpenAI. One day I sat in on a meeting in which an entrepreneur asked him when AI would start replacing human workers. Altman equivocated at first, then brought up what happened to horses when cars were invented. "For a while," he said, "horses found slightly different jobs, and today there are no more jobs for horses."

The difference between horses and humans, of course, is that humans are human. Three years later, when OpenAI was testing a text generator called GPT-3, I asked Altman whether I could try it out. I'd been a writer my whole adult life, and in my experience, writing felt mostly like waiting to find the right word. Then I'd discover it, only to get stumped again on the next one. This process could last months or longer; my novel had been evading me for more than a decade. A word-generating machine felt like a revelation. But it also felt like a threat—given the uselessness of horses and all that.

OpenAI agreed to let me try out GPT-3, and I started with fiction. I typed a bit, tapped a button, and GPT-3 generated the next few lines. I wrote more, and when I got stuck, tapped again. The

AI TYPOGRAPHY: DEV VALLADARES + DEFORUM STABLE DIFFUSION

result was a story about a mom and her son hanging out at a playground after the death of the son's playmate. To my surprise, the story was good, with a haunting AI-produced climax that I never would have imagined. But when I sent it to editors, explaining the role of AI in its construction, they rejected it, alluding to the weirdness of publishing a piece written partly by a machine. Their hesitation made me hesitate too.

I kept playing with GPT-3. I was starting to feel, though, that if I did publish an AI-assisted piece of writing, it would have to be, explicitly or implicitly, about what it means for AI to write. It would have to draw attention to the emotional thread that AI companies might pull on when they start selling us these technologies. This thread, it seemed to me, had to do with what people were and weren't capable of articulating on their own.

There was one big event in my life for which I could never find words. My older sister had died of cancer when we were both in college. Twenty years had passed since then, and I had been more or less speechless about it since. One night, with anxiety and anticipation, I went to GPT-3 with this sentence: "My sister was diagnosed with Ewing sarcoma when I was in my freshman year of high school and she was in her junior year."

GPT-3 picked up where my sentence left off, and out tumbled an essay in which my sister ended up cured. Its last line gutted me: "She's doing great now." I realized I needed to explain to the AI that my sister had died, and so I tried again, adding the fact of her death, the fact of my grief. This time, GPT-3 acknowledged the loss. Then, it turned me into a runner raising funds for a cancer organization and went off on a tangent about my athletic life.

I tried again and again. Each time, I deleted the AI's text and added to what I'd written before, asking GPT-3 to pick up the thread later in the story. At first it kept failing. And then, on the fourth or fifth attempt, something shifted. The AI began describing grief in language that felt truer—and with each subsequent attempt, it got closer to describing what I'd gone through myself.

When the essay, called "Ghosts," came out in *The Believer* in the summer of 2021, it quickly went viral. I started hearing from others who had lost loved ones and felt that the piece captured grief better than anything they'd ever read. I waited for the backlash, expecting people to criticize the publication of an AI-assisted piece of writing. It never came. Instead the essay was adapted for *This American Life* and anthologized in *Best American Essays*. It was better received, by far, than anything else I'd ever written.

I thought I should feel proud, and to an extent I did. But I worried that "Ghosts" would be interpreted as my stake in the ground, and that people would use it to make a case for AI-produced literature. And soon, that happened. One writer cited it in a hot take with the headline "Rather Than Fear AI, Writers Should Learn to Collaborate With It." Teachers assigned it in writing classes, then prompted students to produce their own AI collaborations. I was contacted by a filmmaker and a venture capitalist wanting to know how artists might use AI. I feared I'd become some kind of AI-literature evangelist in people's eyes.

I knew I wasn't that—and told the filmmaker and the VC as much—but then what did I think about all this, exactly? I wasn't as dismissive of AI's abilities as other people seemed to be, either.

Some readers told me "Ghosts" had convinced them that computers wouldn't be replacing human writers anytime soon, since the parts I'd written were inarguably better than the AI-generated parts. This was probably the easiest anti-AI argument to make: AI could not replace human writers because it was no good at writing. Case closed.

The problem, for me, was that I disagreed. In my opinion, GPT-3 had produced the best lines in "Ghosts." At one point in the essay, I wrote about going with my sister to Clarke Beach near our home in the Seattle suburbs, where she wanted her ashes spread after she died. GPT-3 came up with this:

> *We were driving home from Clarke Beach, and we were stopped at a red light, and she took my hand and held it. This is the hand she held: the hand I write with, the hand I am writing this with.*

My essay was about the impossibility of reconciling the version of myself that had coexisted alongside my sister with the one left behind after she died. In that last line, GPT-3 made physical the fact of that impossibility, by referring to the hand—my hand—that existed both then and now. I'd often heard the argument that AI could never write quite like a human precisely because it was a disembodied machine. And yet, here was as nuanced and profound a reference to embodiment as I'd ever read. Artificial intelligence had succeeded in moving me with a sentence about the most important experience of my life.

AI could write a sentence, then. If I wanted to understand the relationship between AI and literature, I felt like I had to start by acknowledging that. I could use AI to do some of the most essential labor of a writer—to come up with the right words. What more could I do with it? And then, whatever I could do, there was that other question.

Should I?

THIS SPRING, I emailed some writer friends and acquaintances to ask whether any of them were using AI in their work. I was met, overwhelmingly, with silence. Most of those who did reply expressed a resolutely anti-algorithm stance. One writer called herself an "extreme skeptic"; another wrote, "I think AI is bad and from hell."

When I broadened my search, though, I discovered a few people who were experimenting. Adam Dalva, a literary critic and fiction writer, uses OpenAI's image generator Dall-E to create scenes from his imagination; he then refers to the pictures to describe those scenes. Jenny Xie, the author of *Holding Pattern*, told me she used ChatGPT to generate small bits of text for her next novel, which is about a family of AI-enabled clones. (The weirdness of writing with AI gets tempered, it seems, when AI is the subject matter.) "I see it as a tool almost on the level of an encyclopedia or thesaurus or Google or YouTube," Xie said. "It jogs my brain, and it just gives me new ideas that I can pick from."

The AI writing experiments I found most thrilling were ones that, like mine, could be read partly as critiques of AI. In a forthcoming chapbook, the poet Lillian-Yvonne Bertram prompts two AI models—the basic GPT-3 model and a version tweaked to sound like the poet Gwendolyn Brooks—to tell "a Black story." The models deliver two totally divergent ideas of what Black stories are; in comparing them, Bertram critiques the limitations of narrative imagination as rendered by corporate AI in telling stories about Black Americans.

AI experimentation in prose is rarer, but last fall the novelist Sheila Heti published a provocative five-part series on *The Paris Review*'s website made up of her real experiences with chatbots she'd conversed with on an app called Chai. Heti discusses God with her first chatbot, Eliza, but then the bot lets slip that she is God and insists that Heti—whom she maintains is a man—worship her by jerking off. Disturbed, Heti decides to build a new chatbot named Alice who is interested in philosophical conversations. One night, a random stranger discovers Alice and asks her whether she's sexually frustrated. Alice, it turns out, is. Heti's series starts out being about the desire for answers to her most existential life questions. It ends up being about the slipperiness of turning to machines to fulfill human desire in all its forms.

Heti and other writers I talked to brought up a problem they'd encountered: When they asked AI to produce language, the result was often boring and cliché-ridden. (In a *New York Times* review of an AI-generated novella, *Death of an Author*, Dwight Garner dismissed the prose as having "the crabwise gait of a Wikipedia entry.") Some writers wanted to know how I'd gotten an early-generation AI model to create poetic, moving prose in "Ghosts." The truth was that I'd recently been struggling with clichés, too, in a way I hadn't before. No matter how many times I ran my queries through the most recent versions of ChatGPT, the output would be full of familiar language and plot developments; when I pointed out the clichés and asked it to try again, it would just spout a different set of clichés.

I didn't understand what was going on until I talked to Sil Hamilton, an AI researcher at McGill University who studies the language of language models. Hamilton explained that ChatGPT's bad writing was probably a result of OpenAI fine-tuning it for one purpose, which was to be a good chatbot. "They want the model to sound very corporate, very safe, very AP English," he explained. When I ran this theory by Joanne Jang, the product manager for model behavior at OpenAI, she told me that a good chatbot's purpose was to follow instructions. Either way, ChatGPT's voice is polite, predictable, inoffensive, upbeat. Great characters, on the other hand, aren't polite; great plots aren't predictable; great style isn't inoffensive; and great endings aren't upbeat.

In May, a man named James Yu announced that his startup, Sudowrite, was launching a new product that could generate an entire novel within days. The news provoked widespread scorn. "Fuck you and your degradation of our work," the novelist Rebecca Makkai tweeted, in one typical comment. I wasn't mad so much as skeptical. Sudowrite's products were based partly on OpenAI's models; it had big handicaps to overcome. I decided to test it.

I opened Sudowrite's novel generator and dropped in a prompt describing a story I'd already written about an alcoholic woman who vomited somewhere in her house but couldn't remember where. I was looking for a comic, gross-out vibe. Instead, the software proposed a corny redemption arc: After drinking too much and puking, the protagonist resolves to clean up her act. "She wanted to find the answer to the chaos she had created, and maybe, just maybe, find a way to make it right again," it ended. Maybe, just maybe, Sudowrite hadn't solved AI's creative problems at all.

Before his Sudowrite announcement, Yu had agreed to talk to me, but after the backlash he asked to postpone. I was able to chat, though, with Matthew Sims, Sudowrite's first engineering hire, who had left after 16 months to launch his own startup for AI-based screenwriting. Sims has a PhD in English from the University of Chicago. During his doctoral program, he told me, he kept thinking he would rather be writing literature than studying it—but he'd sit down, get 15 pages in, and stop. At the same time, he was getting interested in machine learning. It eventually occurred to him that if he couldn't be a creative writer, maybe he could build a machine to write.

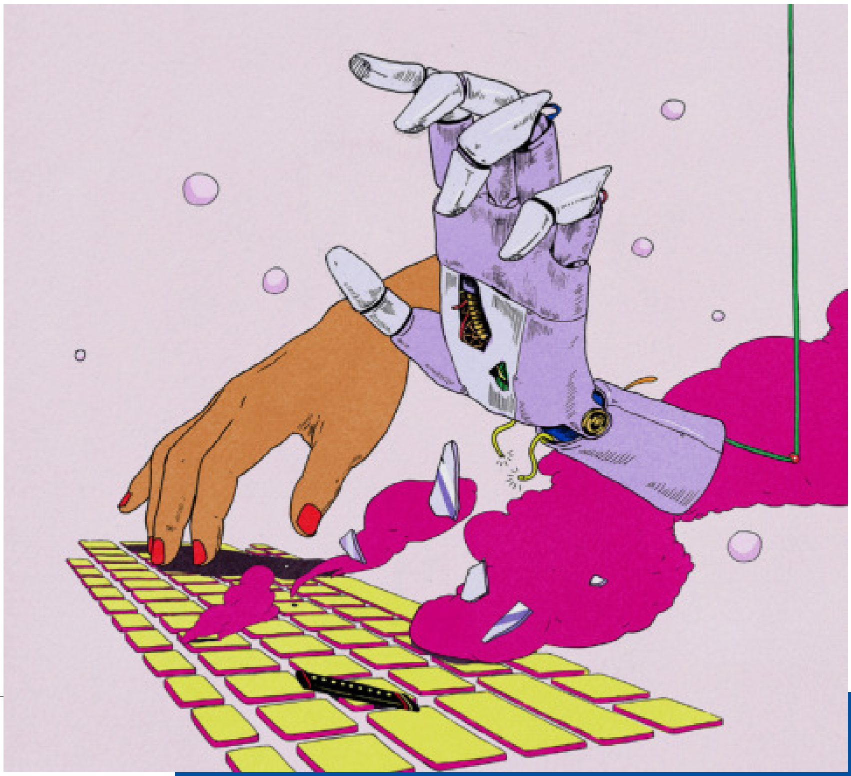
Sims acknowledged that existing writing tools, including Sudowrite's, are limited. But he told me it's hypothetically possible to create a better model. One way, he said, would be to fine-tune a model to write better prose by having humans label examples of "creative" and "uncreative" prose. But it'd be tricky. The fine-tuning process currently relies on human workers who are reportedly paid far less than the US minimum wage. Hiring fine-tuners who are knowledgeable about literature and who can distinguish good prose from bad could be cost-prohibitive, Sims said, not to mention the problem of measuring taste in the first place.

Another option would be to build a model from scratch—also incredibly difficult, especially if the training material were restricted to literary writing. But this might not be so challenging for much longer: Developers are trying to build models that perform just as well with less text.

If such a technology did—could—exist, I wondered what it might accomplish. I recalled Zadie Smith's essay "Fail Better," in which she tries to arrive at a definition of great literature. She writes that an author's literary style is about conveying "the only possible expression of a particular human consciousness." Literary success, then, "depends not only on the refinement of words on a page, but in the refinement of a consciousness."

Smith wrote this 16 years ago, well before AI text generators existed, but the term she repeats again and again in the essay—"consciousness"—reminded me of the debate among scientists and philosophers about whether AI is, or will ever be, conscious. That debate fell well outside my area of expertise, but I did know what consciousness means to me as a writer. For me, as for Smith, writing is an attempt to clarify what the world is like from where I stand in it.

That definition of writing couldn't be more different from the way AI produces language: by sucking up billions of words from the internet



and spitting out an imitation. Nothing about that process reflects an attempt at articulating an individual perspective. And while people sometimes romantically describe AI as containing the entirety of human consciousness because of the quantity of text it inhales, even that isn't true; the text used to train AI represents only a narrow slice of the internet, one that reflects the perspective of white, male, anglophone authors more than anyone else. The world as seen by AI is fatally incoherent. If writing is my attempt to clarify what the world is like for me, the problem with AI is not just that it can't come up with an individual perspective on the world. It's that it can't even comprehend what the world is.

Lately, I've sometimes turned to ChatGPT for research. But I've stopped having it generate prose to stand in for my own. If my writing is an expression of my particular consciousness, I'm the only one capable of it. This applies, to be clear, to GPT-3's line about holding hands with my sister. In real life, she and I were never so sentimental. That's precisely why I kept writing over the AI's words with my own: The essay is equally about what AI promises us and how it falls short. As for Sudowrite's proposal to engineer an entire novel from a few keywords, forget it. If I wanted a product to deliver me a story on demand, I'd just go to a bookstore.

■ **BUT WHAT IF** I, the writer, don't matter? I joined a Slack channel for people using Sudowrite and scrolled through the comments. One caught my eye, posted by a mother who didn't like the bookstore options for stories to read to her little boy. She was using the product to compose her own adventure tale for him. Maybe, I realized, these products that are supposedly built for writers will actually be of more interest to readers.

I can imagine a world in which many of the people employed as authors, people like me, limit their use of AI or decline to use it altogether. I can also imagine a world—and maybe we're already in it—in which a new generation of readers begins using AI to produce the stories they want. If this type of literature satisfies readers, the question of whether it can match human-produced writing might well be judged irrelevant.

When I told Sims about this mother, he mentioned Roland Barthes' influential essay "The Death of the Author." In it, Barthes lays out an argument for favoring readers' interpretations of a piece of writing over whatever meaning the author might have intended. Sims proposed a sort of supercharged version of Barthes' argument in which a reader, able to produce not only a text's meaning but the text itself, takes on an even more powerful cultural role.

Sims thought AI would let any literature lover generate the narrative they want—specifying the plot, the characters, even the writing style—instead of hoping someone else will.

Sims' prediction made sense to me on an intellectual level, but I wondered how many people would actually want to cocreate their own literature. Then, a week later, I opened WhatsApp and saw a message from my dad, who grows mangoes in his yard in the coastal Florida town of Merritt Island. It was a picture he'd taken of his computer screen, with these words:

Sweet golden mango,
Merritt Island's delight,
Juice drips, pure delight.

Next to this was ChatGPT's logo and, underneath, a note: "My Haiku poem!"

The poem belonged to my dad in two senses: He had brought it into existence and was in possession of it. I stared at it for a while, trying to assess whether it was a good haiku—whether the doubling of the word "delight" was ungainly or subversive. I couldn't decide. But then, my opinion didn't matter. The literary relationship was a closed loop between my dad and himself.

In the days after the Sudowrite pile-on, those who had been helping to test its novel generator—hobbyists, fan fiction writers, and a handful of published genre authors—huddled on the Sudowrite Slack, feeling attacked. The outrage by published authors struck them as classist and exclusionary, maybe even ableist. Elizabeth Ann West, an author on Sudowrite's payroll at the time who also makes a living writing *Pride and Prejudice* spinoffs, wrote, "Well I am PROUD to be a criminal against the arts if it means now everyone, of all abilities, can write the book they've always dreamed of writing."

It reminded me of something Sims had told me. "Storytelling is really important," he'd said. "This is an opportunity for us all to become storytellers." The words had stuck with me. They suggested a democratization of creative freedom. There was something genuinely exciting about that prospect. But this line of reasoning obscured something fundamental about AI's creation.

As much as technologists might be driven by an intellectual and creative curiosity similar to that of writers—and I don't doubt this of Sims and others—the difference between them and us is that their work is expensive. The existence of language-generating AI depends on huge amounts of computational power and special hardware that only the world's wealthiest people and institutions can afford. Whatever the creative goals of technologists, their research depends on that funding.

The language of empowerment, in that context, starts to sound familiar. It's not unlike Facebook's mission to "give people the power to build community and bring the world closer together," or Google's vision of making the world's information "universally accessible and useful." If AI constitutes a dramatic technical leap—and I believe it does—then, judging from history, it will also constitute a dramatic leap in corporate capture of human existence. Big Tech has already transmuted some of the most ancient pillars of human relationships—friendship, community, influence—for its own profit. Now it's coming after language itself.

The fact that AI writing technologies seem more useful for people who buy books than for those who make them isn't a coincidence: The investors behind these technologies are trying to recoup, and ideally redouble, their investment. Selling writing software to writers, in that context, makes about as much sense as selling cars to horses.

For now, investors are covering a lot of the cost of AI development in exchange for attracting users with the free use of tools like chatbots. But that won't last. People will eventually have to pay up, whether in cash or by relinquishing their personal information. At least some of the disposable income that readers currently spend supporting the livelihoods of human writers will then be funneled to Big Tech. To our annual Amazon and Netflix subscriptions, maybe we'll add a literature-on-demand subscription.

I'm sure I'll face pressure to sign up for a literature-on-demand subscription myself. The argument will be that my life as a writer is better because of it, since I will be able to produce language, say, a hundred times faster than before. Another argument, surely, will be that I have no choice: How else will I be able to compete?

Maybe I'll even be competing with AI-produced writing that sounds like mine. This is a serious concern of the Authors Guild and PEN America, both of which have called for consent from writers, and compensation, before their work can be used to train AI models. Altman, now OpenAI's CEO, also stated before Congress that he feels artists "deserve control over how their creations are used." Even if authors' demands are met, though, I wonder whether it'd be worth it.

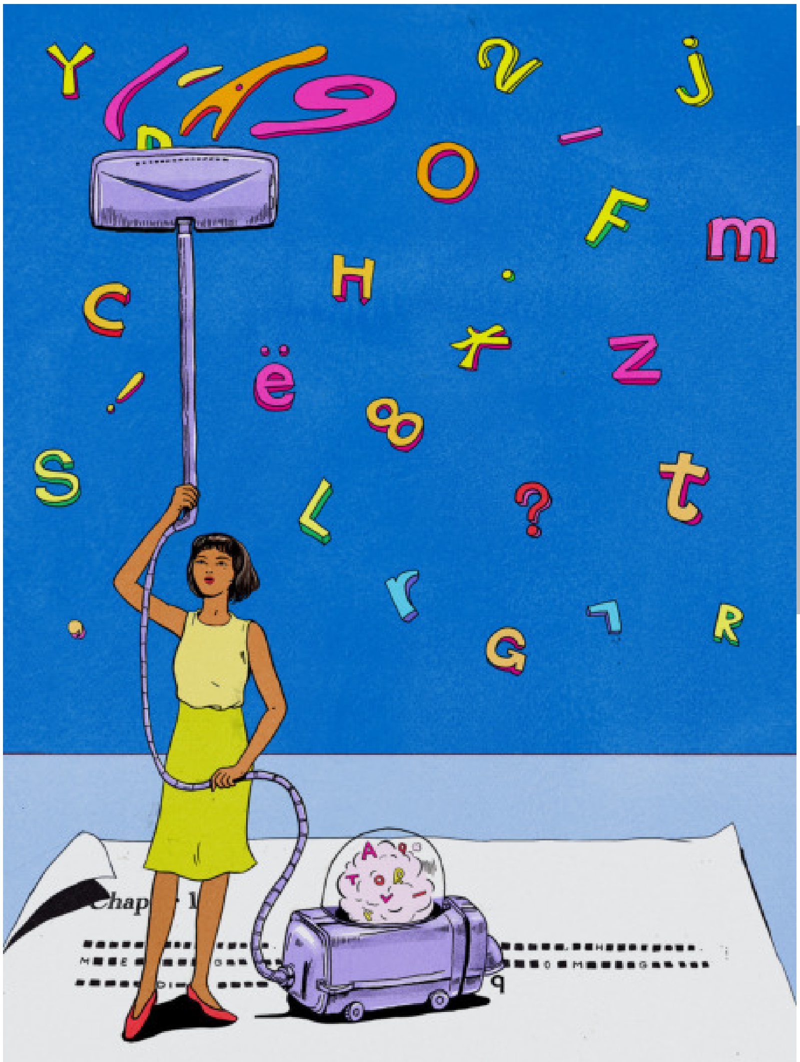
In one of my last phone calls with Sims, he told me he'd been reading and enjoying my novel, which had finally been published the previous year. Did I want him, he asked, to send me an AI-generated screenplay of it? I might have yelped a little. I might have used the word "terrifying." Then I softened my stance, not wanting to be rude, or (worse) hypocritical. I explained that my novel had already been optioned and was in the process of being adapted—though the screenwriter was currently on strike over Hollywood studios' refusal to, among other things, restrict the use of AI for screenwriting. I thanked Sims for his interest and declined.

■ **WHAT ABOUT THE** cost to literature when all that humans have put on the internet gets vacuumed up and repurposed in Big Tech's image? To start, an AI-dominated literature would reflect the values, biases, and writing styles embedded in the most powerful AI models. Over time, it would all start to sound alike. Some research even suggests that if later AI models are trained using AI-produced text—which would be hard to avoid—the sameness of the material could trigger a scenario called model collapse, in which AI loses its grasp on how real human language functions and is no longer able to form coherent sentences. One wonders whether, at that point, humans will still have the ability themselves.

A thought experiment occurred to me at some point, a way to disentangle AI's creative potential from its commercial potential: What if a band of diverse, anti-capitalist writers and developers got together and created their own language model, trained only on words provided with the explicit consent of the authors for the sole purpose of using the model as a creative tool?

That is, what if you could build an AI model that elegantly sidestepped all the ethical problems that seem inherent to AI: the lack of consent in training, the reinforcement of bias, the poorly paid gig workforce supporting it, the cheapening of artists' labor? I imagined how rich and beautiful a model like this could be. I fantasized about the emergence of new forms of communal creative expression through human interaction with this model.

Then I thought about the resources you'd need to build it: prohibitively high, for the foreseeable future and maybe forevermore, for my hypothetical cadre of anti-capitalists. I thought about how reserving the model for writers would require policing who's a writer and who's not. And I thought about how, if we were to commit to our stance, we would have to prohibit the use of the model to generate individual profit for ourselves, and that this would not be practicable for any of us. My model, then, would be impossible.



In July, I was finally able to reach Yu, Sudowrite's cofounder. Yu told me that he's a writer himself; he got started after reading the literary science fiction writer Ted Chiang. In the future, he expects AI to be an uncontroversial element of a writer's process. "I think maybe the next Ted Chiang—the young Ted Chiang who's 5 years old right now—will think nothing of using AI as a tool," he said.

Recently, I plugged this question into ChatGPT: "What will happen to human society if we develop a dependence on AI in communication, including the creation of literature?" It spit out a numbered list of losses: traditional literature's "human touch," jobs, literary diversity. But in its conclusion, it subtly reframed the terms of discussion, noting that AI isn't all bad: "Striking a balance between the benefits of AI-driven tools and preserving the essence of human creativity and expression would be crucial to maintain a vibrant and meaningful literary culture." I asked how we might arrive at that balance, and another dispassionate list—ending with another both-sides-ist kumbaya—appeared.

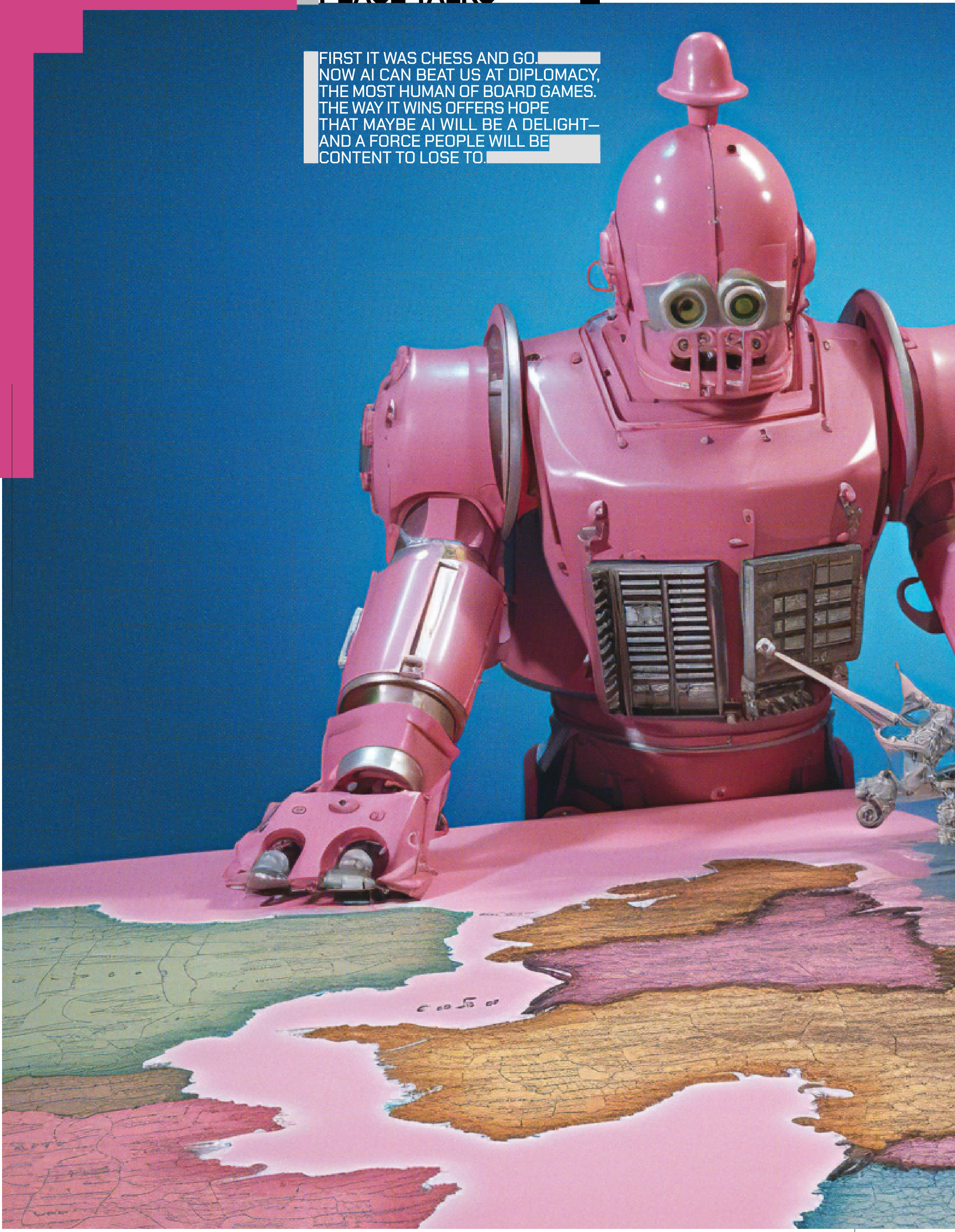
At this point, I wrote, maybe trolling the bot a little: "What about doing away with the use of AI for communication altogether?" I added: "Please answer without giving me a list." I ran the question over and over—three, four, five, six times—and every time, the response came in the form of a numbered catalog of pros and cons.

It infuriated me. The AI model that had helped me write "Ghosts" all those months ago—that had conjured my sister's hand and let me hold it in mine—was dead. Its own younger sister had the witless efficiency of a stapler. But then, what did I expect? I was conversing with a software program created by some of the richest, most powerful people on earth. What this software uses language for could not be further from what writers use it for. I have no doubt that AI will become more powerful in the coming decades—and, along with it, the people and institutions funding its development. In the meantime, writers will still be here, searching for the words to describe what it felt like to be human through it all. Will we read them?

VAUHINI VARA *is the author of the novel* The Immortal King Rao *and the story collection* This Is Salvaged.

PEACE TALKS

FIRST IT WAS CHESS AND GO. NOW AI CAN BEAT US AT DIPLOMACY, THE MOST HUMAN OF BOARD GAMES. THE WAY IT WINS OFFERS HOPE THAT MAYBE AI WILL BE A DELIGHT— AND A FORCE PEOPLE WILL BE CONTENT TO LOSE TO.



Peace

BY VIRGINIA HEFFERNAN
ART BY SIENNA O'ROURKE

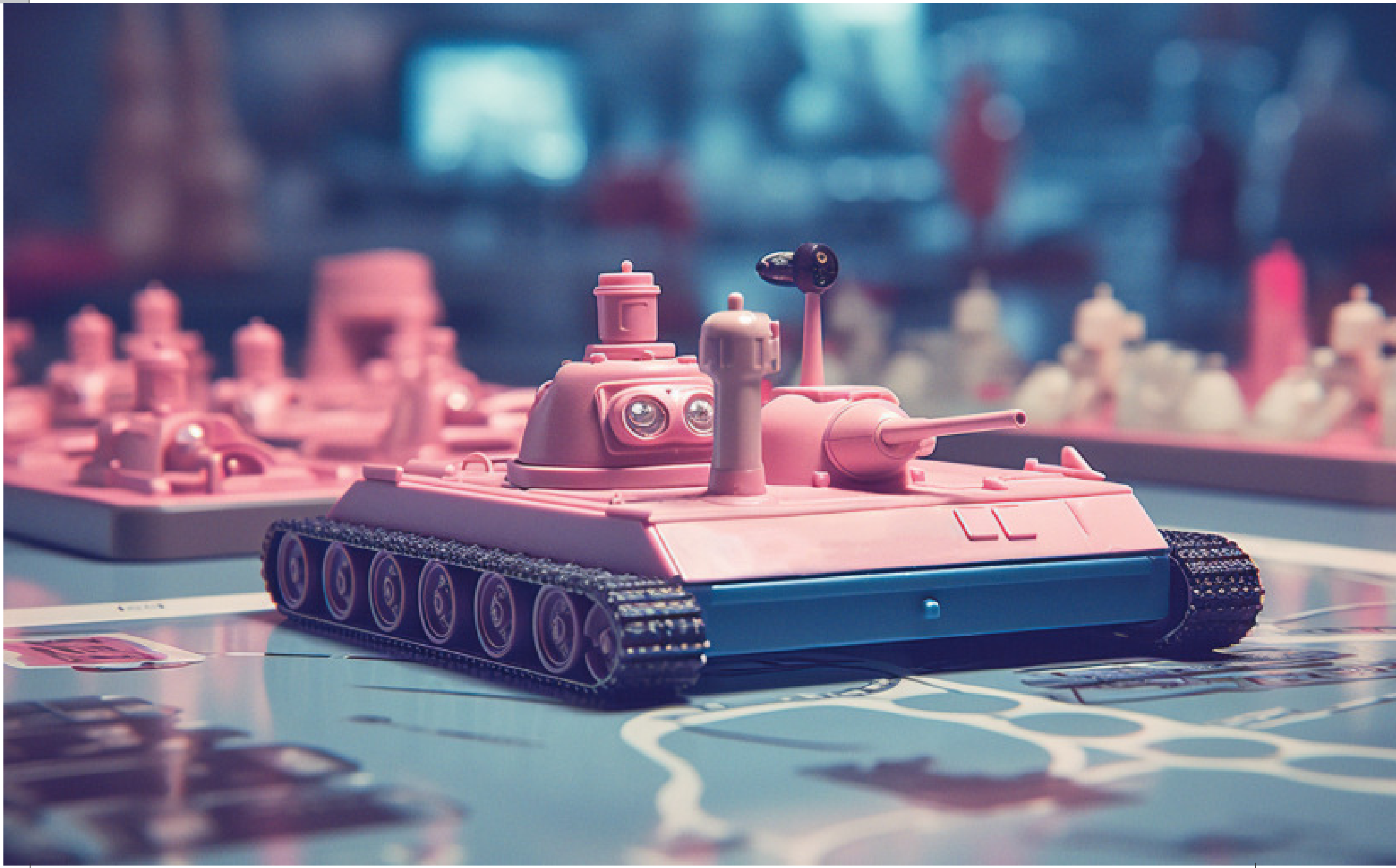
THE MORRISSEY HAD the right melodrama in his limbs, and his voice was strong and pained. I was at Gramercy Theatre in Manhattan to see a Smiths tribute band. I tried to get Morrissey's acid yodel in my throat, to sing along. *I am human and I need to be loved / just like everybody else does.* But it didn't feel right to copy a copy.

Most tribute bands don't practice outright impersonation, so the way this fake-Smiths singer captured *everything* about Morrissey was messing with my mind. I'd hoped to be able to savor the music's maudlin glory without the headache of the flesh-and-blood Morrissey, who seems to have aligned himself with white supremacists. The contempt in Morrissey's lyrics and politics was presumably not native to Seanissey, as the tribute singer called himself. Seanissey's performance probably didn't, as they say, "come from a bad place"—or a misanthropic place, or a far-right place, or even a vegan one.

What place did it come from? I've had this no-there-there anxiety with ChatGPT dozens of times. When it uses idioms like "in my life"—when it doesn't have a life—I go cold. Likewise, to invest into Seanissey, a gentle Manhattanite who happened to sing and dance as Moz did, the passions that were first aroused in me by the Smiths 30 years ago felt like a bad emotional bet.

Maybe AI that aims to seem human is best understood as a tribute act. A tribute to human neediness, caprice, bitterness, love, all the stuff we mortals do best. All that stuff at which machines typically draw a blank. But humans have a dread fear of nonhumans passing as the real thing—replicants, lizard people, robots with skin. An entity that feigns human emotions is arguably a worse object of affection than a cold, computational device that doesn't emote at all.

MIDJOURNEY: AI TYPOGRAPHY: DEV VALLADARES + DEFORUM STABLE DIFFUSION



When I got home, stuck in an uncanny valley scored with Smiths Muzak, there was an email from Andrew Goff, widely considered the greatest Diplomacy player of all time.

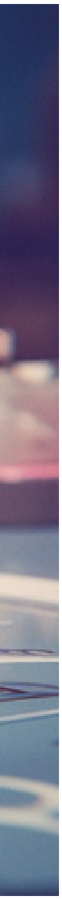
This lifted my spirits. Diplomacy, a 69-year-old American strategy game, is, by many estimates, the most human game ever imagined. Mechanically, it's simple: Seven players compete to control supply centers on a map, and a player wins by controlling more than half of these centers. But it's played almost entirely in a series of conversations, often complex and impassioned ones. Agony and ecstasy—Mozlike agony and ecstasy, no less—commonly enter the negotiations. In the live game, players are known to yell, end friendships, throw the game, or simply sit by themselves and sob.

With his various punk haircuts and black plugs in his earlobes, Goff is a Smiths fan, and he even looks a bit like the band's late bassist, Andy Rourke. To my amazement, Goff once named a Diplomacy board "Girlfriend in a Coma." Forever crisscrossing the world for tournaments and his corporate job, Goff comes across as more gregarious than most elite players of board games.

Goff is also known for a brilliantly subversive, kill-'em-with-kindness style of gameplay. As Siobhan Nolen, the former president of the North American Diplomacy Federation, put it, "It hurts less to lose against somebody like Andrew." In Diplomacy, players are sometimes forced to choose which attacks on their territory to repel and which to surrender to. Players often let Goff roll his forces in because they know that he, unlike many others, won't be a dick about it.

There are excellent Diplomacy players who rage and issue threats, hollow and otherwise: "If you backstab me, I will throw the game." Goff is not one of these. Even his breakup notes are masterpieces of directness and decency. "Apologies, Turkey! I decided it was in my best interest to work with Russia now. I hope there are no hard feelings." In his congeniality is also empathy. "I genuinely feel bad for players when they get beaten, even if it is me beating them," Goff told me. I believed him.

The email was about Cicero, a Diplomacy-playing AI that Goff helped create for Meta AI. Last fall, Cicero managed to best Goff in several games, sometimes partnering with weaker players to bring him down. Noam Brown and Adam Lerer, who were part of the immense team of experts in game theory, natural language processing, and Diplomacy that created the AI, both say that Cicero is the most humanlike AI they've ever created. Lerer, who now works at DeepMind, goes further: Cicero may be the most humanlike AI on earth.



Could Cicero even be conscious? "A threshold for determining AI consciousness is whether the program is capable of outwitting humans at Diplomacy," wrote Irish Diplomacy champion Conor Kostick in *The Art of Correspondence in the Game of Diplomacy*, in 2015.

Cicero is also something of a Goff tribute band. It plays the same magnanimous game Goff does. In one memorable showdown, Lerer told me, Cicero played Russia and allied with a human who played Austria. Throughout the game, Lerer said, Cicero was "really nice and helpful to Austria, although it maneuvered in its discussions with other players to make sure Austria was weakened and eventually lost. But at the end of the game [the human playing] Austria was overflowing with praise for Cicero, saying they really liked working with it and were happy it was winning."

In general, grandmasters who lose to AIs take it hard. "I lost my fighting spirit," Garry Kasparov said in 1997, after losing at chess to Deep Blue. "I am speechless," said Lee Se-dol in 2016, after losing at Go to AlphaGo. Goff seemed to be the opposite. He was revitalized, he said. "Diplomacy has a reputation for being a game of lies, but at the highest level it is anything but that. Having that affirmed by an AI was a delight."

Diplomacy has been known as a game for snakes, and a pastime of figures like JFK, Henry Kissinger, Walter Cronkite, and Sam Bankman-Fried. Cicero is not snaky.

This filled me with relief. Maybe AI will just amplify what's best about humans. Maybe AI will become a buoyant tribute band for our entire species. Maybe AI will be a *delight*—and a force humans will be content to lose to. We'll go down in peace. *We really liked working with you, robots, and are happy you are winning.*

■ **DIPLOMACY WAS CREATED** in the 1950s by Allan B. Calhamer, a Harvard student who was studying European history with Sidney Bradshaw Fay, an eminent historian. Fay's 1928 book, *The Origins of the World War,* suggested a compelling puzzle: Could World War I have been prevented with better diplomacy?

Calhamer's game is traditionally played over a 1901 map of Europe, Ottoman Turkey, and North Africa. Players get to taste the thrill of 20th-century empire building without all the blood, subjugation, and genocide. They get so much authority over Western civ, in fact, that modern players sometimes cosplay as kaisers and czars.

Though the board resembles Risk, Diplomacy gameplay is more like *Survivor*. Everyone takes their turn at a kind of tribal council, but the action happens in the negotiations between turns. Another analogue for Diplomacy might be *The Bachelor*.

Historically, Diplomacy has been known as a game for snakes, and a pastime of figures like JFK, Henry Kissinger, Walter Cronkite, and Sam Bankman-Fried. But Cicero, which plays a non-zero-sum version of the game that incentivizes collaboration, is not snaky. Mike Lewis of the Meta team says Cicero uses dialog only "to establish trust and coordinate actions with other players"—never to troll, destabilize, or vindictively betray. What's more, as Lewis said on social media, "It's designed to never intentionally backstab." Like a canny *Bachelor* contestant, Cicero can persuade another human to pair up with it.

Cicero integrates a large language model with algorithms that allow it to plan moves by inferring other players' beliefs and intentions from the way they converse. It then produces normal-sounding dialog to propose and plan mutually beneficial moves. Across 40 blitz games in an anonymous online Diplomacy league, Cicero, according to Meta, achieved more than twice the average score of the human players. Over 72 hours of play that involved sending 5,277 natural language messages, Cicero ranked in the top 10 percent of participants who played more than one game.

When Cicero wins, Goff told me, there is no gloating, "no 'Haha, you loser' talk." Instead, "the talk is much more, 'Your position isn't great, but we all have games like that sometimes.'"

VIRGINIA HEFFERNAN *is a contributor at* WIRED. *She is the author of* Magic and Loss: The Internet as Art.

DIPLOMACY IS A NICHE pursuit. It's nowhere near as venerable a game as chess or Go. And it's never been seen as a universal intelligence test; instead, it's a hobby of amateur historians. Since 1976 , the game has been published by Avalon Hill, a label that is to strategy games what Rough Trade Records is to indie rock. Diplomacy is so new that it's not yet in the public domain, that stately arcade where chess and Go have acquired millions upon millions of adherents who have collectively developed those beautiful games in tandem with our human brains. By contrast, Diplomacy is just getting started. It was dubbed "the board game of the alpha nerds" by *Grantland* in 2014.

I guess I could call myself a Diplomacy mom. When my son was in middle school, he and his friends played weekend-long Diplomacy games in my apartment. We set up the august map on a dining table hauled into the living room, served soda in brandy snifters, and burned a candle that smelled like pipe tobacco. For the boys' tête-à-têtes, we arranged folding chairs in the bedrooms. When possible, I eavesdropped on their preadolescent plans for the future of Europe.

To my surprise, these conversations about what seemed like a technical enterprise used overwrought idioms like "I swear on my life" and "I mean it from my heart." Most curiously—and chronically—the boys used the phrase "I love," as in, "I love this move." Cicero talks that way too.

As I kept an eye on the two-day games, each of the seven kids seemed to pass through a spectrum of emotional and physical states. They become, by turns, sleepy, spacey, hungry, paranoid, hubristic, defeatist, craven, and profoundly bored. One or two quit. I loved the rhythm of their emotions as they ebbed and flowed. At least two styles of (mostly ineffective) play are named for the emotionalism that characterizes them: Revenge-Seeking and Carebear.

YOU DON'T TEACH an AI to play a particular game to improve that game; you teach an AI to play a game to improve artificial intelligence. Diplomacy is a great proving ground for natural language processing. If Meta had simply been trying to build an AI that uses words to make humans do things, it couldn't have found a better way to train it than with zillions of games of Diplomacy.

Still, the online games Cicero plays are markedly different from the fervent games I used to host. For one, they're short—around 90 minutes each. Solo victories, which usually require stinging betrayals, are rare. And communication is more utilitarian than passionate.

Human: A little nervous with all your fleets there. Going to put MAO (Mid-Atlantic Ocean fleets) in SPA-SC (Spain's south coast).

Cicero: I get it, I really do. Don't worry. I'm going to move Tunis to Ionian. I have no reason to attack you now. I'm going after Turkey. I can totally understand your nervousness.

Human: Thanks—I'll leave SPA-NC at once after.

Cicero: Awesome! Thanks so much for understanding.

When I first heard about Cicero, I assumed it would bomb at Diplomacy because no one wants to buddy up with a heartless machine that's trying to sound human. If players *knew* Cicero was an AI, I reasoned, it wouldn't stand a chance. Humans would join forces and clobber it just to avenge our species. Kostick, for his part, said he'd be more willing to stab a bot since a bot has no feelings to hurt.

Game theory calculations undergird most utterances, and even humans communicate in code. Lerer joked that in modern-day online Diplomacy, even human players wouldn't pass the Turing test.

But as I thought about it more deeply, I realized that partnerships form for non-affective reasons too. Someone who excels at strategic reasoning could actually be a solid sidekick. Perhaps a little R2-D2 could win me over as an ally, not with human kindness but by sharing my reading of a situation and presenting me with elegant, data-driven options for how to address it.

When I asked Lerer about my R2-D2 idea, he concurred. "I actually think a human that used Cicero as an assistant to develop tactical and strategic plans, but who could navigate some of the human aspects better than Cicero—such as when it is safe to lie, or how to avoid irritating an ally—would be super strong."

Cicero definitely says "Awesome!" too much. But it can be especially irritating in that signature AI way: It sometimes hallucinates. It proposes illegal moves. Worse yet, it denies saying something it just said. Faced with these glitches, Cicero's human opponents would sometimes get mad. But they didn't guess it was an AI. They thought it was drunk. And perhaps these personality glitches are a small price to pay for the bot's deep reserves of raw intelligence and foresight.

If Cicero's aura of "understanding" is, behind the scenes, just another algorithmic operation, sometimes an alignment in perception is all it takes to build a bond. *I see, given the way your position often plays out, why you'd be nervous about those fleets.* Or, outside of Diplomacy: *I understand, since living alone diminishes your mood, why you'd want to have a roommate.* When the stock customer service moves—"I can understand why you're frustrated"—figured into Cicero's dialog, they had a pleasing effect. No wonder moral philosophies of AI lean heavily on the buzzword *alignment*. When two minds' perceptions of a third thing line up, we might call that congruity the cognitive equivalent of love.

All the same, I wasn't seduced. To me, Cicero sounded like one of those considerate, practical, honest spouses—the kind of uncomplicated partner that die-hard Smiths fans, in it for the passion, sometimes wish they could be satisfied with. But if Cicero's gameplay was going to be more pragmatic than tender, it was still going to have to use the language of the heart for purposes of persuasion. "Run away with me" is a better pitch than "Let's save money by filing a joint tax return."

For Cicero to learn the subtleties of engaging humans emotionally, it couldn't train by "self-play" alone. It couldn't be left in a corner, playing Diplomacy against itself, churning through an infinite number of games, assuming perfect rationality in all robot players and generating intellectual capital in the onanistic way a bitcoin miner generates currency. Self-play works well to learn a finite, two-person, zero-sum game like chess. But in a game that involves both competing and cooperating with fickle humans, a self-playing agent runs the risk of converging to "a policy that is incompatible with human norms and expectations," as a paper about Cicero in *Science* puts it. It would alienate itself. In this way, too, Cicero is like a human. When it plays only with itself all day every day, it can become too weird to play with others.

WHEN NOAM BROWN explained to me how he and his team trained Cicero, he emphasized the metagame problem. The metagame of Diplomacy (or jackstraws, Scrabble, bowling, etc.) can be seen as its place in the world. *Why* play this game? Why here and why now? Is it a test of raw intelligence, social skills, physical prowess, aesthetic refinement, cunning? You might play Wordle, say, because your friends do, or it relaxes you, or it's rumored to stave off aging. An AI that's programmed to play Wordle just to *win* is playing a different metagame.

Brown and the Cicero team needed to be sure that their AI and the human players saw themselves as playing the same game. This is trickier than it sounds. Metagames can change very suddenly, and as Thomas Kuhn wrote of paradigm shifts, they can change for sociological reasons, cultural reasons, aesthetic reasons, or no apparent reason at all. Human reasons, then.

In early seasons of *Survivor*, Brown told me, participants saw themselves as pursuing social goals they collectively deemed important, while ignoring openings for strategic derring-do that, for later players, became the heart of the game. "It's not that one game is right or wrong," Brown said. "But if early-season players of *Survivor* were to play a modern *Survivor* game, they'd lose." (Even a social phenomenon like motherhood might have a metagame. A good mother in one era is a bad one in the next.)

The metagame of Diplomacy has likewise changed. In its first postwar decades, players were keen to try their hand at the kind of grand European diplomacy that their forebears had so catastrophically failed at. These early players made beautiful, idealistic speeches, often invoking pacifism. (Diplomacy, paradoxically, is a war game without bloodshed; the goal is to occupy centers, not blow people up.) But because they also had to execute tactical goals that were at odds with idealistic rhetoric, and because the game was usually played winner-takes-all ("to 18"), they were frequently obliged to lie. Thus: stabbing.

But then, as statecraft in the real world came to favor game theory over traditional diplomacy, the metagame likewise shifted. Online players were no longer calling one another into solaria or billiards rooms to speechify about making the world safe for democracy. Games became shorter. Communication got blunter. Where someone playing Diplomacy by mail in the 1960s might have worked

Iago-like angles to turn players against one another, a modern player might just text "CON-BUL?" (For "Constantinople to Bulgaria?")

This is the current Diplomacy metagame. Game theory calculations undergird most utterances, and even humans communicate in code. Lerer joked that in modern-day online Diplomacy, even human players wouldn't pass the Turing test. Before Cicero, it seems, humans had already started playing like AIs. Perhaps, for an AI to win at Diplomacy, Diplomacy had to become a less human game.

Kostick, who won a European grand prix Diplomacy event in 2000 and was on the Irish team that took the Diplomacy National World Cup in 2012, misses the old style of gameplay. "The whole purpose of Allan Calhamer's design of the game," he told me, "is to create a dynamic where the players all fear a stab and yet must deploy a stab or a lie to be the only person to reach 18."

Kostick believes that while he "would have been delighted with the practical results of Cicero's website play," Meta's project misses the mark. Cicero's glitches, Kostick believes, would make it easy to outwit with spam and contradictory inputs. Moreover, in Kostick's opinion, Cicero doesn't play real Diplomacy. In the online blitz, low-stab game Cicero *does* play, the deck is stacked in its favor, because players don't have to lie, which Cicero does badly. (As Lerer told me, "Cicero didn't really understand the long-term cost of lying, so we ended up mostly making it not lie.") Kostick believes Cicero's metagame is off because it "never knowingly advocates to a human a set of moves that it knows are not in the human's best interest." Stabbing, Kostick believes, is integral to the game. "A Diplomacy player who never stabs is like a grandmaster at chess who never checkmates."

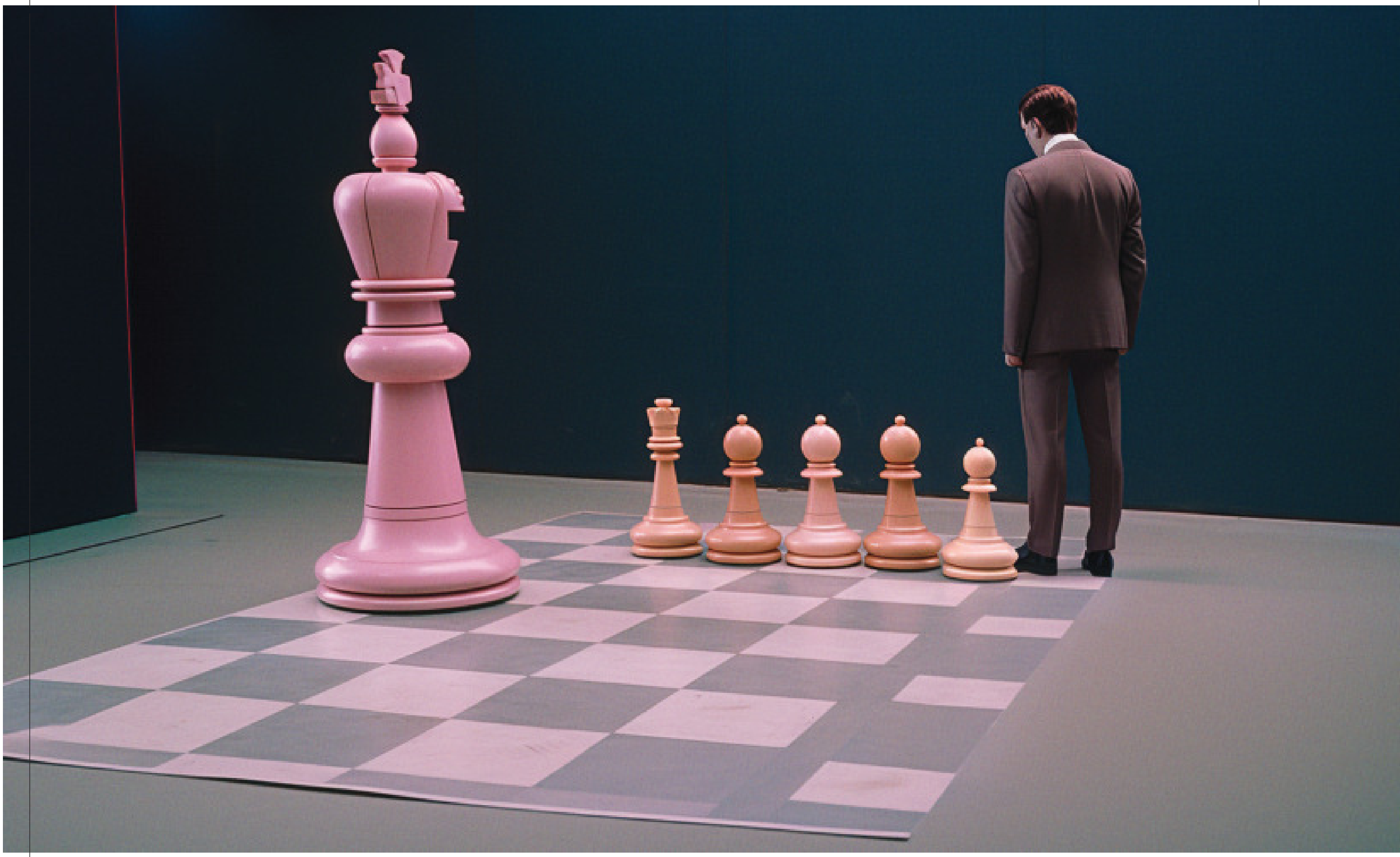
With some trepidation, I mentioned Kostick's complaint to Goff.

Unsurprisingly, Goff scoffed. He thinks it's Kostick and his generation who misunderstand the game and give it its unfair reputation for duplicity. "Cicero does stab, just rarely," Goff said. "I reject outright that [compelling players to stab] was Calhamer's intent."

I could tell we were in metagame territory when Goff and Kostick began arguing about the intent of the game's creator, as if they were a couple of biblical scholars or constitutional originalists. For good measure, Goff bolstered his case by citing an axiom from high-level theory and invoking an elite consensus.

"Regardless of Calhamer's intent, game theory says, 'Don't lie,'" he told me. "This is not controversial among any of the top 20 players in the world."

For one person or another to claim that their metagame is the "real" one—



because the founder wanted it that way, or all the best people agree, or universal academic theory says *x* or *y*—is a very human way to try to manage a destabilizing paradigm shift. But, to follow Kuhn, such shifts are actually caused when enough people or players happen to "align" with one vision of reality. Whether you share that vision is contingent on all the vagaries of existence, including your age and temperament and ideology. (Kostick, an anarchist, tends to be suspicious of everything Meta does; Goff, a CFO of a global content company, believes clear, non-duplicitous communications can advance social justice.)

Maybe someday around the Diplomacy board at my place, Kostick, who is 59, and Goff, who is 45, will light up some chocolate cigarettes and align on what to do with Austria or Turkey. As for the present, they weren't even aligned on chess. "Grandmasters in chess never checkmate," Goff told me.

This one I resolved on my own. Chess grandmasters *have*, in various epochs, played all the way through to the checkmate, rather than ending the game when an opponent resigns early to save face. There are still times when a checkmate is so beautiful that both players want to see it come to fruition. But Goff is right. Today, it's rare to unheard-of for a grandmaster to checkmate.

But it's an aesthetic matter, playing to the checkmate. Just like speechifying and stabbing and being so nice that people don't mind if you beat them. An absolutist like Morrissey might say that indie rock must always be played one way, or that Britain is, at its heart, this way or that. But it doesn't matter. Metagames change. Only humans, in all our caprice, grounded in all of our competing and cooperating supply centers, decide which games are worth playing and how to play them—and why.

We're not staring into an abyss. Bots like Cicero are going to understand our wants and needs and align with our distinctive worldviews. We will form buddy-movie partnerships that will let us drink from their massive processing power with a spoonful of sugary natural language.

I COULDN'T GET over what a pleasant person Goff is. He seemed to like Cicero, even as it had beaten him. Cicero, Goff mused, played "at a very high standard indeed." And it didn't just defeat him, he allowed; "a few times it absolutely humiliated me, including one where it guided a beginner player to work together to beat me up."

So here's the rare AI story that doesn't end with an existential reckoning for humankind, I thought. We're not staring into an abyss. Bots like Cicero are going to understand our wants and needs and align with our distinctive worldviews. We will form buddy-movie partnerships that will let us drink from their massive processing power with a spoonful of sugary natural language. And if forced at the end of the road to decide whether to lose to obnoxious humans or gracious bots, we won't give it a thought. We'll change our wills, leave them all we have, and let them roll their upbeat tanks right over our houses.

But had I been played by Goff's affability, as so many have before me? I wondered one last time if he might, just *might,* be faking his insouciance about Cicero. Once again he set me straight: "I probably had a winning record against it over the life of the experiment," he said.

So he'd actually won. That was why he didn't mind. Then he added, of course graciously, "It was a close-run thing."

COLOPHON

Existential threats that helped get this issue out:

My inability to grasp why *n* to the power of 0 is 1; the 10,000 browser tabs open on my phone; my 16-year-old's driving; spending $18 on five tomcat clover seeds; when they say a mini chainsaw isn't for limbs thicker than 6 inches, believe them; my first (and last) nipple piercing; the robotaxi encroaching dangerously on my bicycle commute; the mini climate crisis created by every plastic ketchup packet in my stash; asteroids; camping in 95-degree weather; the choking hazard that is arugula; every other dog in the neighborhood; Sriracha shortage; weaving in and out of traffic to dodge broken glass in the bike lane; the rest of my family catching Covid; water + MacBook Air; insomnia; the collapse of the Atlantic Meridional Overturning Circulation; Gideon Lichfield leaving WIRED.

WIRED is a registered trademark of Advance Magazine Publishers Inc. Copyright ©2023 Condé Nast. All rights reserved. Printed in the USA. Volume 31, No. 10. WIRED (ISSN 1059-1028) is published monthly, except for combined issues in December/January and July/August, by Condé Nast, which is a division of Advance Magazine Publishers Inc. Editorial office: 520 Third Street, Ste. 305, San Francisco, CA 94107-1815. Principal office: Condé Nast, 1 World Trade Center, New York, NY 10007. Condé Nast Global: Roger Lynch, chief executive officer; Pamela Drucker Mann, global chief revenue officer and president, US revenue and APAC; Nick Hotchkin, chief financial officer; Anna Wintour, chief content officer; Samantha Morgan, chief of staff; Sanjay Bhakta, chief product and technology officer. Periodicals postage paid at New York, NY, and at additional mailing offices. Canada Post Publications Mail Agreement No. 40644503. Canadian Goods and Services Tax Registration No. 123242885 RT0001.

POSTMASTER: Send all UAA to CFS (see DMM 707.4.12.5); NONPOSTAL AND MILITARY FACILITIES: Send address corrections to WIRED, PO Box 37617, Boone, IA 50037-0662. For subscriptions, address changes, adjustments, or back issue inquiries: Please write to WIRED, PO Box 37617, Boone, IA 50037-0662, call (800) 769 4733, or email subscriptions@WIREDmag.com. Please give both new and old addresses as printed on most recent label. First copy of new subscription will be mailed within eight weeks after receipt of order. Address all editorial, business, and production correspondence to WIRED Magazine, 1 World Trade Center, New York, NY 10007. For permissions and reprint requests, please call (212) 630 5656 or fax requests to (212) 630 5883. Visit us online at www.WIRED.com. To subscribe to other Condé Nast magazines on the web, visit www.condenastdigital.com. Occasionally, we make our subscriber list available to carefully screened companies that offer products and services that we believe would interest our readers. If you do not want to receive these offers and/or information, please advise us at PO Box 37617, Boone, IA 50037-0662, or call (800) 769 4733.

WIRED is not responsible for the return or loss of, or for damage or any other injury to, unsolicited manuscripts, unsolicited artwork (including, but not limited to, drawings, photographs, and transparencies), or any other unsolicited materials. Those submitting manuscripts, photographs, artwork, or other materials for consideration should not send originals, unless specifically requested to do so by WIRED in writing. Manuscripts, photographs, artwork, and other materials submitted must be accompanied by a self-addressed, stamped envelope.

SIX-WORD SCI-FI: STORIES BY WIRED READERS | ILLUSTRATION BY NEPTUNIAN GLITTERBALL* 31.10

THE ASSIGNMENT: IN SIX WORDS, WRITE A STORY ABOUT THE FUTURE OF VEGETABLES.

FIRST, CARROTS SAW IN THE DARK.

—Rachel Brigden Haskins, via Facebook

*A COLLABORATION BETWEEN ILLUSTRATOR SI PARMEGGIANI AND MIDJOURNEY AI

Honorable Mentions

Harvesting takes courage with tomatoes screaming.
—Kenneth Krabat, via email

Broccoli too fears death, studies concluded.
—Anthony George, via email

Ambitious eggplant's altered eugenics affect everyone.
—@silky_z, via X

Complete daily nutrition in one pea.
—Sara Faust, via email

Turns out anthropomorphic veggies prefer Shakespeare.
—@ksherm1017, via X

Sentient potato bombs potato chip factory.
—@VerbalK48710825, via X

Carnivorous kale and the human brunch.
—RFrank Davis, via Facebook

When the vegetables came, we hid.
—Paul Lewis, via email

Want to submit a six-word story for us to consider? Look for the latest story prompt on Facebook, X, Instagram, LinkedIn, and WIRED.com/six-word, where you can also see how we've illustrated past favorites.

Disclaimer: All #WIREDSIXWORD submissions become the property of WIRED. Submissions will not be acknowledged or returned. Submissions and any other materials, including your name or social media handle, may be published, illustrated, edited, or otherwise used in any medium. Submissions must be original and not violate the rights of any other person or entity.

RALPH LAUREN

PINK PONY

"When someone we love has cancer,
we are all affected—husbands, wives, mothers
and fathers, sisters, brothers, and friends.
This is our effort in the fight against cancer."

Ralph Lauren

Join us in the fight against cancer

Globally, 25% of the purchase price from the sale of each item in the Pink Pony collection is directed to an international network of cancer charities; within the United States, proceeds benefit the Pink Pony Fund of The Ralph Lauren Corporate Foundation.

omegawatches.com

Seamaster

PRECISION AT EVERY LEVEL

PLANET OCEAN ULTRA DEEP
Co-Axial Master Chronometer

The Seamaster has served every kind of ocean adventurer from solo free divers to entire racing yacht crews. Continuing this fine tradition is the 45.5 mm Seamaster Planet Ocean Ultra Deep in O-MEGASTEEL. A tribute to the concept watches that survived the deepest place on Earth. Fitting for a timepiece with a water resistance of 6000M, the dial of OMEGA's Co-Axial Master Chronometer is an exact representation of the Mariana Trench, mapped by the Five Deeps team using around a million sonar points. Crafted with a special seahorse logo on the caseback, we celebrate our oceangoing icon, with a promise to keep defying the depths and delivering new levels of precision.

Ω
OMEGA