# DIALOGUES

On AI, society, and what comes next

Anna Ridler · Andrew Berry · Anthony Townsend · Barney McCann · Blaise Agüera y Arcas
Bob Mankoff · Boris Sofman · Brianne Kimmel · Casey Reas · Charley Locke · Christopher Wood
Cristina Bowerman · Cynthia Breazeal · Daniela Amodei · Daniela Rus · Daniel Oberhaus · David Salle
Demis Hassabis · Erik Brynjolfsson · Gianpaolo Tucci · Greg Corrado · James Manyika · Jane Metcalfe
Jen Swetzoff · Kay Firth-Butterfield · Kent Walker · Khyati Trehan · Lila Ibrahim · Linda Dounia
Lisa Joy · Maria Konnikova · Marian Croak · Markos Kay · Megan Peters · Mira Lane · Miriam Vogel
Nicholas Thompson · Pau Garcia · Pierre Buttin · Refik Anadol · Renée Cummings · Sneha Mehta
Steven Pinker · Tobias Rees · Ulrich Blum · Vernacular · will.i.am · Yossi Matias · Yuri Suzuki

Google

Masthead

# DIALOGUES

On AI, society, and what comes next


**ATLANTIC RE:THINK**

EDITORIAL

**Liz Haq**
Head of Editorial

**Gabriel Muller**
Co-Managing Editor

**Ricki Harris**
Co-Managing Editor

**Jaime Brockway**
Copy Editor

DESIGN

**Drew Campbell**
Creative Director, Designer

**Hanah Slotterback**
Art Director, Designer

PRODUCTION

**Monica Schmelzer**
Senior Director, Account Management

**Elena Mehas**
VP, Strategic Initiatives

---

**Nicholas Thompson**
CEO, *The Atlantic*

VISUAL CONTRIBUTORS

**Cayce Clifford**
**Leonie Bos**
**Somnath Bhatt**
**Marine Buffard**
**Domestic Data Streamers**
**Markos Kay**
**Barney McCann**
**Vernacular**
**Khyati Trehan**
**Gianpaolo Tucci**
**Anna Ridler**
**Casey Reas**
**Jan St. Werner**
**Linda Dounia**
**Pierre Buttin**
**Refik Anadol**
**Yuri Suzuki**
**Jovelle Tamayo**
**Uli Knörzer**
**Jon Han**
**Clara Mokri**
**Denise Nestor**
**Sophi Miyoko Gullbrants**
**Camilo Huinca**
**Bob Mankoff**
**Shannon Wheeler**

1 Colorful yellow paint stroke · 24 variations

2 Colorful paint blob of blue and white, lots of texture · 9 variations

3 Colorful paint stroke, orange yellow white grey, high resolution · 33 variations

4 Colorful paint stroke white blue green · 15 variations

5 Colorful paint stroke, orange red coral white, high resolution · 3 variations

6 Colorful paint stroke, purple lilac maroon white, high resolution · 45 variations

*Atlantic Re:think*

Atlantic Re:think is the creative marketing studio at *The Atlantic*. We produce stories with partners, independently of *The Atlantic*'s editorial staff.



# DIALOGUES

On AI, society, and what comes next



Strokes of paint on a surface have been the mark of human contact for thousands of years. For the cover of *Dialogues*, a meditation on the intersection of people and technology, the paint strokes were created with some assistance from a machine. The texture, beauty, richness, and colors are all the result of a collaboration with generative AI. At times, the results matched the initial intention; other times, repeated experimentation produced a happy accident. See the prompts that created each stroke to the left.



## INTRO

**Nicholas Thompson**
CEO of *The Atlantic*

**When I was a child,** I remember telling my mother that I was sad that everything had already been discovered. We were at the round kitchen table in our house in Boston, sitting next to each other, separated by a pitcher of milk painted to look like a cow. People had climbed Everest and they had been to the moon. They had walked through the Amazon. It had taken thousands and thousands of years for humans to map the contours of the globe. Now they'd done it. So didn't that mean I was destined to grow up in a less magical world?

A generation later, I still have the pitcher of milk shaped like a cow. When I talk to my kids at our breakfast table about the future, though, the question is flipped. Now, more than at any point in my life, I feel like the world has been unmapped. We know where everything is, yes. But with the rise of generative AI, we don't know what our place in it will be. We're building systems that may become more powerful than people. We're stepping into vast new jungles, wearing our worn-out boots with just our flashlights to guide us.

To me, this is exhilarating. Progress in AI won't stop. The spirit of invention and the forces of capitalism won't allow that. Each new system will help create the next more powerful system, and each new system will create new tools for good and new tools for ill. That means that we need people to make maps for what comes next. We need to figure out the most important questions that technological change will bring. And we need to start to scribble down our answers. We need to start to explore the early contours of this new world.

That's the inspiration for this magazine, which *The Atlantic*'s creative marketing studio, Re:think, created alongside Google. We wanted to find the biggest unanswered questions in AI: the questions that no one has really figured out yet. We wanted to explore them. Will AI help us understand our own minds? How exactly should it be regulated? Will it make the world more equal or less? Can it learn to tell jokes? None of these questions has a simple answer; some of them don't really have an answer at all. That's why they're interesting, and that's why they're important.

My mother was right with the response she gave me back then at the kitchen table: The world is much bigger than you think, and there will always be something new to discover. Right now, I hope this magazine helps you discover and think through artificial intelligence, the biggest new thing humanity has discovered in years.

# What does it mean to get AI right?



How can we steer the trajectory of artificial intelligence not only toward efficiency, but also toward being just? The challenge is not merely technical. It is deeply human.

AI reflects our values, biases, and aspirations. Its promise and perils mirror our own. This technology demands we reckon with questions of equity, accountability, and purpose.

As Google's James Manyika says in the following pages, "AI is putting a mirror in our face to say, 'Okay, humanity, you look like this. How do you want to deal with this?'" AI alone cannot blaze a righteous path, but paired with moral direction, it could help propel our most ambitious shared visions.

Charting this course will require sustained effort across sectors. At its best, AI can extend our capabilities and help unlock solutions to humanity's greatest challenges. But first, we must do the work to articulate what those solutions should be. Getting AI right means looking within and asking: What future do we seek for one another? Then, we must steer technology toward that horizon.

Q&A

# Doing the Most Good

AI has immense potential to help humanity. At the same time, it also poses complex ethical questions. James Manyika, Google's Senior Vice President of Research, Technology, and Society, believes that achieving progress requires building powerful systems that benefit people and society—systems that help cure disease and expand access to both information and opportunities—and addressing risks around serious challenges like bias, misuse, and safety.

←
Photography by
Cayce Clifford

**James Manyika likes to think about** how to responsibly steer AI's development for the benefit of society. That's one of the many reasons Google hired him to serve as its Senior Vice President of Research, Technology, and Society, a role that has him leading the company's efforts to ensure its AI innovations positively impact humanity. Manyika focuses on helping to advance AI's development and guide Google on building AI that benefits people and helps solve pressing societal problems, creates inclusion and access for all, and addresses risks that come with intelligent systems. He advocates nuance, humility, and considering topics from multiple perspectives when advising on AI's development. As Manyika sees it, there's no such thing as a one-sided problem—especially when it comes to AI.

Question **Let's start with the big picture. There's a lot of discussion about what it means to "get AI right." What does that mean to you?**

James Manyika At the highest level, getting AI right has two sides to it. On the one hand, it's about making sure we build AI that will benefit people through its capacity to assist, complement, empower, and inspire people in every field of human endeavor—from the everyday to the ambitious and imaginative. This also includes AI that helps advance scientific breakthroughs and discoveries, and helps in solving pressing societal problems and creates access and opportunity for everyone. The other side is just as important, making sure we build AI that addresses the risks and complexities that come with having powerful and highly capable systems. We have to get both sides right.

Question **To do good and not do harm simultaneously.**

Manyika Yes, to do the *most* good. Not just do good, but do the most good—and do all the things that without the help of AI we can't do at all, at scale, or fast enough. Things that benefit people everywhere, and improve lives. This is critically important and it's the thing that motivates us.

Question **Right. And that highlights why the stakes are so high. It's not just because of the risks, but also because of the real potential positive benefits for humanity. That's on the line, too. Can you paint that picture, as you see it, of the ways in which we might experience those benefits across society, if we do get it right?**

Manyika Let's break this down into a few categories. First, we actually have to build AI that is powerful and capable enough to not only assist and help people, but also help us solve the seemingly impossible problems we want to solve for society. This could include things like discoveries and breakthroughs in science so we can cure cancer and create new drugs and therapies. For example, we've been working with a consortium of researchers to create a "human pangenome," a new resource that better represents human genetic diversity, allowing scientists and doctors to more accurately diagnose and treat diseases with AI. You've seen what we have achieved with AlphaFold to solve a 50-year grand challenge to predict the structure of proteins. It's predicted the structure of all proteins known to science, all 200 million of them, opening up wider possibilities to help researchers understand diseases, discover new drugs and therapies—and help tackle many neglected diseases. The breakthroughs from AI also allow us to address pressing societal crises that people are experiencing today, like the impact of climate change and increasing extreme weather. Powered by AI, our flood forecasting program began with just 2 countries and now covers more than 80 and provides forecasting up to 7 days in advance of a flood to 460 million people in harm's way. I should also add that one of the ways that AI will contribute to society is through helping to power the economy, especially through its potential to drive productivity growth, which has been sluggish for a while, and will be even more critical to prosperity as society ages.

Next, once the systems are capable enough, we need to make sure we actually apply them in such a way that everyone benefits from its development and the things we will solve using it. It is really important to make sure that everybody benefits.

Question **And what about the risks involved in creating these systems?**

Manyika Right. That's the other side, and there you've got a range of complexities. We have to make sure we're building trustworthy systems that people can actually believe in and will not cause harm—that they're not going to amplify societal biases or generate toxic or dangerous outputs. Then you've got questions about misapplication and misuse—are these systems used for the right things? Using AI systems to create misinformation is not an ethical use. Using AI systems to pursue criminal activities, for cyber hacking or terrorist acts, surveillance—those are examples of misuse of these systems. There are also complexities of how AI will impact other aspects of our economy besides enabling productivity, such as work. There is both the possibility of augmenting what people do, as well as the risk of substituting what people do. The development of this technology and how it's used, will likely result in both happening—this is what Erik Brynjolfsson has written about in his "Turing Trap" paper. As he points out, the choices we make in AI's development, its use, as well as the policies and incentives around all of this, will affect outcomes. But according to most recent research, over the next decade or more, assuming continued economic growth, more jobs will be created than lost, but the majority of jobs will change—all raising the stakes for skill development and adaptation.

There's also a question of alignment, which many in the field have thought about for a long time, going all the way back to Alan Turing. When we say alignment, we mean: How does society make sure that these systems do what we want? That they are aligned with our goals, with our preferences, with our values? The issue is as much about *us* as it is about *how* we build these systems. For example, you and I can say we want a system aligned with our values. But which ones? Especially in a complex and global world, involving many people and places with varying cultures



Part of getting it right is making sure everybody benefits.

and views and so on—these are the classic problems of normativity. These are questions about *us*.

Question **Going back more specifically to that narrow idea of alignment for a moment—as you said, it's defined as this idea of ensuring that AI actually does what's intended by a given directive, is that right?**

Manyika But even that is unclear! For example, do you want alignment with the specific instructions you give it? In other words, do you want it to solve a problem as stated and follow a stated goal precisely? Or do you want it to figure out your actual intention—despite what you say—and solve that, which may not be exactly how you've stated the goal? For example, maybe you say you want to exercise every day, but in the end you're not really exercising every day. So does the system align with what you've said you want to do, or what you're actually doing? Or should it align with what is "best" for you?

Then you've got an additional complexity, which is: Should it align with you individually, the majority, a particular group, or with the union of everyone's goals? Or does it align with what's good for society, despite what society itself says? Those questions become even more complex if you consider a world in which we each have our own AI agent. Presumably we would want each AI agent to be aligned with its "owner", but what if my AI is technically more powerful than yours? To me, all these questions have less to do with the particular technology, per se. They're not technical questions. They're questions for us as society. Many of them are as old as society itself.

The technical questions, to me, seem solvable over time. It's a bit like saying, "I want the error rate of a well-defined error to go down." That's a very specific problem. That's an engineering problem, a technical problem. Or if you said, "I don't want bias," and if you define what the bias looks like, one can try to solve for that problem. Now, that does not guarantee that the engineers will succeed in solving for the error rate or for the bias, but at least it's a well specified problem. I worry more about the second problem, the human one, the question of defining bias in the first place, and there are many such questions, the questions of *us* and what we want, because we've been grappling with those questions for thousands of years.

Question **Any discussion of human values quickly becomes, as you said, a question of *whose* values, and that question is both philosophical and deeply political. How might we begin to negotiate those differences and potentially competing goals?**

Manyika AI is forcing us to look at ourselves in the mirror. Because we now actually have to answer these questions! Before, they were theoretical, normative, and philosophical issues that humanity was grappling with. AI is presenting us with an opportunity, and perhaps raising the stakes, for us to deal with them. For example, we've always had bias in society. On many of the questions like bias, like fairness, even safety, many of the current shortcomings of AI are with respect to some normative sense of the perfect society or human, or a set of ideals of the world we want, but continue to struggle to achieve, hopefully

All too often we paint these questions as either one or the other. But the answer can be both. We want to be bold and responsible. They're not inconsistent.

without giving up. So would certainly not want to have AI worsen harms or shortcomings of society—so we should fix that, but perhaps it may also help us address the shortcomings in our society towards the society we want. So in way, AI is putting a mirror in our face to say, "Okay, humanity, you like look like. How do you want to deal with this?"

Question **Which is a pretty big question.**

Manyika Yes. And I'm not a philosopher, but I think the work that philosophers and researchers are doing in AI typically gives us some frames to think through. One frame is about focusing on the things we all mostly have come to agree on, such as a universal human rights framework. But that's a bit of a floor as opposed to a ceiling on values. The second approach is what in philosophy is often called "the veil of ignorance." This is based on some of the work of John Rawls and other philosophers, and the idea is to come up with values or principles that you'd live with if you didn't know who you were going to be in that society or what your station was going to be, or what your endowments would be. What principles would you be comfortable living with? The third is similar to a bottoms-up approach, which is to aggregate what everybody seems to be doing and base everything on that. But, you could end up with tyranny of the majority, or the sort of problems and approaches that researchers in social choice theory think about.

We could complicate this even further. What we've been discussing is what's typically talked about as the normative challenge, which is that your values may be different from mine, or may vary between this group and that group. There's also what's typically referred to as the plasticity challenge, which is that what you and I might have agreed on twenty years ago may be different today.

Question **And our answer to that question, just like our values, changes over time.**

Manyika Exactly. Imagine if we'd invented AI in 1950 and we locked-in on whatever the values we wanted at that point in time, forever. We'd probably all look back on them now and say, "How the heck did we agree to do that? We don't think that anymore."

Question **It feels like we're being presented with an opportunity for reformation on some level, for possible societal transformation. But there's also a chance we don't do that and simply cement the values and inequities of the society we already live in, or worse, create new issues of inequality. How can we work to do the most good, while avoiding those kinds of outcomes?**

Manyika The inequality question is important. That's why I emphasize that part of "getting it right." I'm convinced that AI will create an incredible bounty of opportunities, in the economy and in scientific discoveries, but part of getting it right is making sure everybody participates and benefits—by everyone, I mean all people, communities, small businesses, organizations of all kinds, countries and regions of the world. That will not happen on its own, we have to work to make sure everyone is able to participate and to benefit—it won't be automatic.

I love the fact that we're working on moonshots that try to do that—to solve problems with solutions to benefit everyone. Let me give an example: Google Translate originally launched in 2006 supporting 11 languages, using statistical machine translation. Once we introduced deep neural networks in 2016, we were able to improve translation quality and reach our current number of supported languages—134—which is extraordinary. There's more than 7,000+ languages spoken around the world—if you cared only about language translations for wealthy groups of people, you might've stopped at twenty. But we went to 134. Our research team has a moonshot goal, which is to build an AI model to support language and speech tools for more than 1,000 languages.

Of course, speaking and writing aren't the only ways that people communicate, and so we're building our models to be multimodal, meaning they're capable of unlocking information across different formats like images and video, and the many other ways people communicate.

Question **Wow.**

Manyika The fact that we're aiming for that is a good thing, because if people are going to benefit from language translation, which helps get access to the world's information, knowledge, and insight, and opportunities, that's a good thing. But how do you make sure that happens everywhere, on everything and for *everyone*?

Of the 7,000+ spoken languages, only a few are well represented online, which means our traditional approaches to training language models on text from the web fail to capture the full breadth of the world's languages. Our moonshot goal aims to solve that by working with people and communities around the world to source representative speech data.

This idea of collaboration with communities is also central to many of our projects, for example our Project Elevate Black Voices initiative. Research shows that Black people in the United States often have a worse experience when using automatic speech recognition technology when compared to white speakers. We've established an incredible partnership with Howard University that is working to create an African-American English speech dataset to help Black people have a better experience with voice products, and not feel the need to "code switch" in order to be understood by technology.

Question **I want to ask you about regulation. As you said earlier, these systems and their outputs have to be trustworthy. That's a baseline. But we also have to consider the potential for clashes of divergent value systems or competing goals. Someone who has a profit motive won't necessarily care about doing the most social good. How do we negotiate that? Is it okay to not all be aligned in the same way? Can we accommodate competing goals across different sectors? Can we agree to disagree? And if not, what kind of arbiter do we need?**

Manyika Yes, I think regulation is part of the answer. But when I think about the role of regulation with regards to AI, again, I come back to it having two sides. Yes, I want regulation to limit all the bad things we don't want, the risks and other downsides. But I also want to enable an ecosystem that can work toward the things we want and create incentives for those things.

Regulation can and should do both things. It can help stop the bad stuff and enable the good stuff. I think all these things are two-sided. Even take the alignment question. Sure, we have to solve the technical side to make sure the systems do what we've agreed on. But we have to agree first on what we as society want. We have to work on both problems. It's not one or the other. It's hard work, but let's do both. When we make it only a technical problem, I think humanity is taking a pass. It's throwing the problem back to the technology and saying: *Solve for what we want. Even though we haven't told you what we want.*

Question **It sounds like folks who don't usually collaborate will find themselves needing to.**

Manyika Well, we have to do all this work together. We need to decide what we want, what benefits we want, what problems we want to solve, what issues we want to avoid. Whether you are a scientist, a technologist, humanist, a government regulator, in civil society, citizens—we need to solve this stuff together.

Question **To what extent are those collective conversations already happening?**

Manyika That's one thing I'm excited about. The fact that we're having these conversations now, this early in AI's development, as opposed to after the fact, or late, I think that's great. I think in the case of social media, that conversation happened way too late. I think the collective conversations we need are starting.

For example, Stanford University has set up the Human-Centered AI Institute which was established as a multi-disciplinary institute with computer scientists, economists and other social scientists, philosophers and more. In another example, I'm serving as the vice chair of the National AI Advisory Committee that was established by Congress to

advise the president. The makeup of that committee involves technologists, computer scientists, academics, people in civil society, labor union leaders. Very different people from very different vantage points, and with a wide range of perspectives and concerns. And in spring of 2022, the American Academy of Arts and Sciences published a volume of its journal Daedalus that I guest-edited on "AI & Society"—in it I included perspectives from many prominent computer scientists, economists, legal scholars, philosophers, public servants and others, all with diverse perspectives, all grappling with the possibilities of AI for society.

At Google, we have set up the Digital Futures Fund, through which we're providing $20 million in grant funding from Google.org to think tanks and academic institutions to foster debate about AI policy and support responsible approaches. And in an industry example, we were involved in setting up the Partnership on AI, which now involves hundreds of organizations, from companies, universities to civil society. We're also working closely with Anthropic, Microsoft and OpenAI to set up a new industry body—the Frontier Model Forum—focused on ensuring safe and responsible development of frontier AI models.

I think we need to have more of these kinds of collaborations. My view is that, in any of those conversations, we need to make sure we're bringing different perspectives together, as opposed to reflecting only one side or the other. That's the collective challenge in my view—to focus on and solve for both the beneficial impact and the challenges and risks. That, to me, at the end of the day, is what we have to get right.

Question **As humans, we have immense limitations in understanding the entirety of any scenario. We are really bad at understanding counterfactuals, in particular. Is there any way we can use that knowledge of our own limitations to our advantage here? Could it help us identify blind spots, perhaps, or at the least somehow accommodate them?**

Manyika Let me give you a technical answer first. There's this idea that's been proposed by people like Stuart Russell at Berkeley and others, which is about using our own incomplete understanding and uncertainty in information to create some wiggle room in how we specify goals for the systems so that they're not trying to optimize some precisely or badly stated goal, but can instead generally move in the right direction, with human guidance. This is an example where we're using our own blind spots, so to speak, to actually be a feature so we don't have overly prescribed, precise goals that could be harmful. In that, there is the possibility of a technical answer.

I think the other part of it is to focus on the outcomes we want rather than the methods of achieving them, given how fast the technical and scientific advances are progressing and the uses of AI are evolving. For example, the way we used to think about solving for bias eight years ago in AI systems is totally different today. Eight years ago, we said, "Well, if you want to solve bias, then clean up the data." Well, that might've been correct eight years ago, but the capabilities have moved on. Today, you may want to train a system on everything, because in several cases, systems that are trained on *everything* have been shown to be better capable of actually detecting biases and doing something about it. That's an example of moving lightly when it comes to prescribing solutions, because the technical capabilities are moving so quickly. You don't want to outrun these capabilities with overly prescriptive ways to solve things, especially if these ways are soon surpassed.

Question **It's all very complex, indeed.**

Manyika Yes, but we must and can work through it. I have confidence in humanity's resilience and ingenuity, but we have to do the work, together. This involves questions such as, what should our institutions look like in the age of AI? What would it mean to be human in the age of AI? Or questions like what does it mean to be intelligent or educated in the age of AI? For example, in 1970, to be considered a smart kid, you'd have to be able to do math in your head. We got past that; calculators forced us to get past that. Now, you have kids who are brilliant mathematicians who may be terrible at doing math in their heads, but they're still brilliant mathematicians. One could say the same thing about the ability to recite facts, dates and such. We've evolved our thinking of what it means to learn and be educated. The same might be said for creativity. There was a time when we used to think that because hip-hop deejays were sampling they must not be great musicians. Some of these questions of intelligence, creativity, of what it means to be human in the age of AI will be unsettling, but also exciting and perhaps liberating. But it will take all of us working together, with both bold ambitions for society, serious consideration of the challenges and risk, and a healthy dose of humility and willingness to learn and course-correct as we move forward. I think that's something we're going to have to get right, too.



SIDEBAR

# BUILDING AI WITH A CONSCIENCE

From Capitol Hill to the forefront of AI research, Daniela Amodei's journey is reshaping the AI industry.

**How do you create an ethical product** in a field where the very definition of ethical is changing by the moment, the legal rules are still being written, and the tech itself is evolving at brain-breaking speeds?

This question motivated siblings Daniela and Dario Amodei to co-found Anthropic, an AI company devoted to safety and research that just so happens to also be building some of the most powerful large language models (LLMs) and enlisting some of the world's biggest companies as partners.

"I started my career in international development, working on issues like poverty assessment, conflict mitigation, and global health," Daniela Amodei says. Her diverse experiences ranged from political campaigns on Capitol Hill to leading teams across various sectors at startups like Stripe and OpenAI. It was her co-founder and brother, Dario, with his background in neuroscience and computational biology, who initially exposed her to the field of AI.

The Amodeis and a few of their earliest Anthropic colleagues previously worked at OpenAI, the company behind ChatGPT. But the question "How do you ensure a safe AI future?" motivated them to strike out on their own. In a recent story in *The New York Times*, writer Kevin Roose reported that Anthropic staff were fearful of the damage future AI could do: "Some compared themselves to modern-day Robert Oppenheimers, weighing moral choices about powerful new technology that could profoundly alter the course of history."

This is an incredible amount of weight to carry around on a day-to-day level. So how does one create an ethical AI product and ensure that this power is used for good? The answer at Anthropic is to build a safe AI company and, with it, a safe AI. The company is doing that by creating standards that guide its own actions as a business and a constitution that trains its LLM, known as Claude.

As for the business itself, Anthropic is a Public Benefit Corporation, a designation that requires it to prioritize social impact and stakeholder accountability—not just profits. The company also published a transparent, extensive document outlining its governance structure called "The Long-Term Benefit Trust," which empowers a panel of five "financially disinterested" experts to oversee and,



> “External engagement on these issues is central to our work. We think developing AI safely is a much broader project than Anthropic can—or should—tackle alone.”

if necessary, remove members of its executive board. Essentially, Anthropic has built-in guardrails.

“We want the transition to more powerful AI systems to be positive to society and the broader economy. This is why much of our research is focused on exploring ways to better understand the systems we are developing, mitigate risks, and develop AI systems that are steerable, interpretable, and safe,” Amodei says.

This kind of thinking informs how Anthropic builds safety into its AI models. Anthropic employs a training technique that has come to be known as constitutional AI, in which it uses a written constitution, rather than subjective human feedback, to teach values and limits to its models and train them for harmlessness. The result is that compared to other popular LLMs, Claude is much more reticent about performing certain tasks. An AI model can’t be self-conscious, per se. But Claude’s training can give it an almost sheepish voice at times.

“I don’t have subjective feelings or emotions as an AI system,” Claude said in an interview. “However, I was created by Anthropic to be helpful, harmless, and honest.”

Those three words—helpful, harmless, and honest—appear repeatedly whenever Claude is prompted to the limits of its learned principles. And although Claude declines to speak about its training (“I apologize, but I do not actually have detailed insight into my own training process or ‘constitution’”), Anthropic says its constitution is a constantly-evolving document that draws from a wide range of sources, including the UN Universal Declaration of Human Rights and Apple’s terms of service.

“Fostering a better understanding of the technology will be crucial to ensuring the industry as a whole is developed safely and responsibly,” Amodei says. “This not only applies to the general public, but to policymakers and civil society, too.”

Part of the reason for this constitutional training approach is that AI trained by AI is easier to scale. And scale is also one of Anthropic’s stated goals. To test whether the principles of constitutional AI hold up, it is necessary to develop increasingly powerful models—and the primary way that happens is by scaling. But this requires increasing both the amount of users whose queries can teach the model and the amount of computational power behind it.

The pursuit of AI at scale raises other ethical questions: There’s the environmental cost of all that computational power; there’s the necessary involvement of one of a small handful of tech companies that even have access to that power; and there’s the potential, as the user base increases, for bad human actors to try to subvert the model’s trained principles and use it for some nefarious purpose.

But these questions are inherent to AI regardless of who is building it, and Anthropic, of course, is just one of many companies creating powerful LLMs.

“External engagement on these issues is central to our work. We think developing AI safely is a much broader project than Anthropic can—or should—tackle alone,” Amodei emphasizes. “Our hope is that by being transparent about the risks we’re seeing, we’ll be able to motivate a much broader effort into exploring potential solutions.”

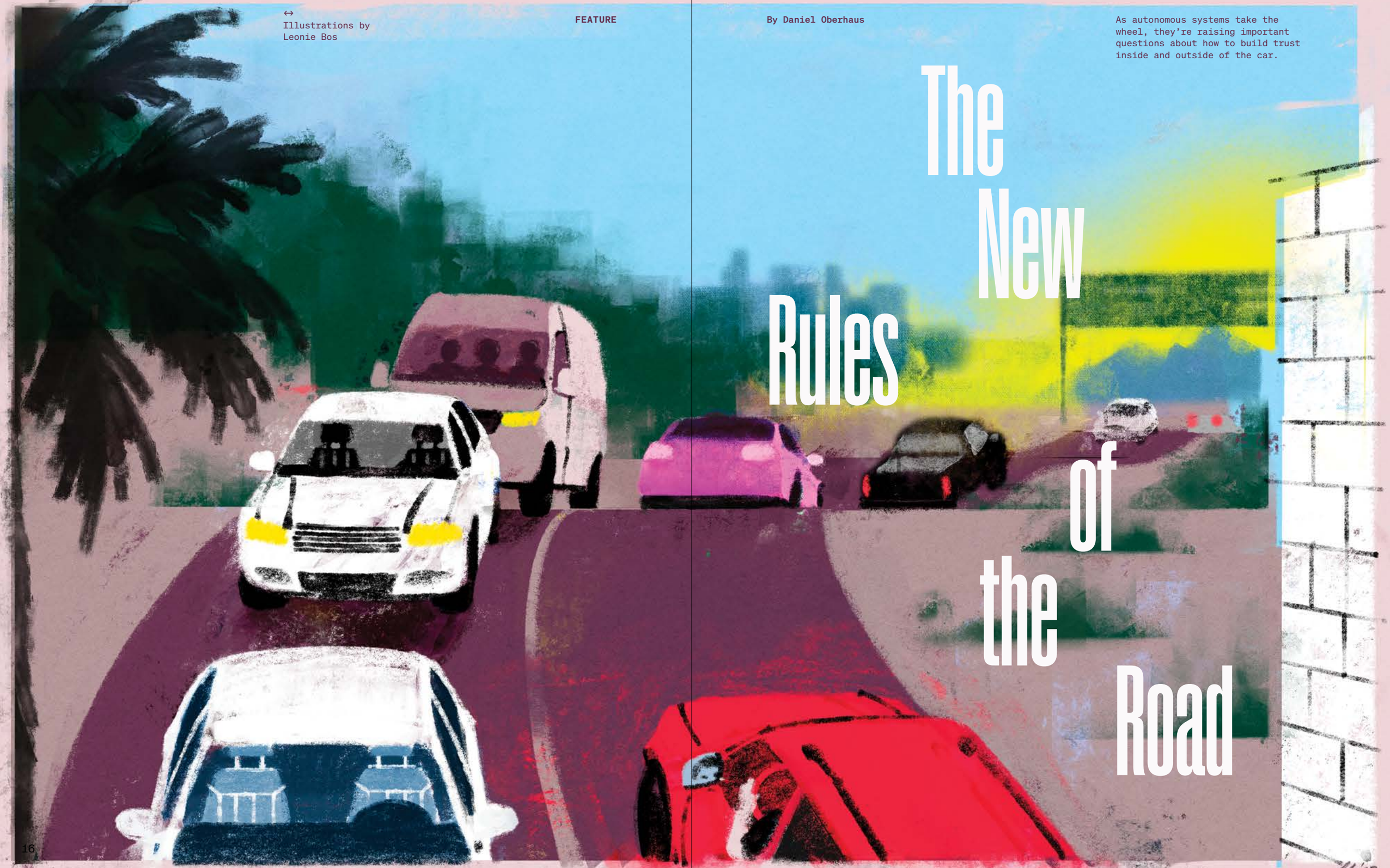
If only people who don’t care about ethics train AI models, then AI models will be amoral at best. Anthropic’s belief is that we can’t make AI safe in the present unless we develop safe AI. And we can’t make it safe in the future, at the frontier of technology, unless we reach that frontier ourselves.

↔
Illustrations by
Leonie Bos

By Daniel Oberhaus

As autonomous systems take the wheel, they're raising important questions about how to build trust inside and outside of the car.

# The New Rules of the Road



**OVER THE COURSE OF THE PAST FIVE YEARS,** fully autonomous vehicles have been notching tens of millions of miles on the roads. As they've navigated the streets, miles per disengagement, a metric researchers use to track how far autonomous vehicles travel without the need for human intervention, has been climbing steadily. And yet, public trust in autonomous vehicles continues to fall. A survey published by AAA showed the number of participants indicating they were afraid of autonomous vehicles jumped up to 68 percent in 2023, which is 13 points higher than the year before. As automated vehicles become safer, the public seems to trust them less.

In an age where people and machines are increasingly sharing the world's roads, it's critical that engineers, policymakers, and the general public work together to forge a new social contract—one that will ensure safety without stymieing technological progress. And that includes building autonomous vehicles that empower drivers and pedestrians to accurately calibrate their trust in the underlying technology.

It's a big ask, but a challenge worth taking seriously given the benefits autonomous vehicles have to offer. They hold the promise of reduced congestion, better fuel economy, fewer parking headaches, and greater accessibility for those unable to drive. But perhaps most important, they have the ability to tremendously improve safety on the road. The vast majority of car crashes in the United States are due to human error and are often the outcome of driving while tired or distracted. Autonomous vehicles, by contrast, never nod off or check their phones, and can often see and react to roadway hazards that may escape the notice of a human driver.

The companies building and operating these autonomous vehicles face the challenge of overcoming a deep skepticism from the millions of drivers they share the road with. This doesn't mean building systems that never make mistakes, which is an unattainable goal for any automated technology—or, for that matter, any human. Instead, it means building autonomous vehicles that riders, pedestrians, and other drivers can trust.

*

For engineers, *trust* in this context is difficult to define. But having a precise definition—as well as ways to objectively measure it—is key to building autonomous vehicles that passengers feel safe in. The elements of trust are something that Xi Jessie Yang, director of the Interaction and Collaboration Research Lab at the University of Michigan, thinks about a lot. At her lab, she and her collaborators spend their days studying how drivers and pedestrians interact with autonomous vehicle simulations. To do this, they use frameworks that help them identify which areas of trust are lacking and how they might be improved. The key, she says, is starting with a precise understanding of what, exactly, they're looking for.

**68%**

of participants in a recent survey published by AAA indicated they were afraid of autonomous vehicles

"We define *trust* as the attitude that autonomous agents will help achieve an individual's goals in situations characterized by uncertainty and ambiguity," Yang says. "This uncertainty and ambiguity is a huge part of trust. If there is no uncertainty or ambiguity, you just trust it 100 percent."

The way a user comes to (dis)trust an autonomous vehicle is known in the research community as "trust calibration." If users feel—rightly or wrongly—that the technology is unreliable, unpredictable, or high-risk, they reduce their trust in that system accordingly, and vice versa: If they feel it's reliable, they increase their trust in it. Trust, in other words, is a dynamic variable that can change over time as a user gains more experience interacting with an autonomous system. The goal of Yang and her colleagues in the human-factors research community is to "influence the public to have well-calibrated trust." This means working to understand the ways that engineering and design decisions, as well as human psychology, converge in the back seat of a fully autonomous vehicle.

As Leanne Hirshfield, an associate research professor at the University of Colorado, Boulder's Institute for Cognitive Science, points out, building trustworthy autonomous systems is not the same as building systems that users should unquestioningly trust. Instead, it's about creating transparent AI that helps a user understand how *much* trust they should have in an automated system.

As an example, Hirshfield imagines a driver in a car with semiautonomous features on a highway at night. In many instances, these technologies can perform better than a human driver at night because they use radars and other sensors that don't depend on light. The driver, in this case, would be justified in trusting the self-driving features to help navigate these conditions. But if the car's sensors aren't performing as well as expected, the car's computer can flag its diminished performance to the driver and encourage them to take the wheel for safety. In this case, even though the system performed worse than the driver might have expected, their trust in it is likely to increase because they know they can count on the vehicle to inform them when the automated system has reached its limits.

"Right now there are things that AI is way better at and things that humans are way better at," Hirshfield says. Humans, for example, are great at soaking up new information, integrating it into their model of the world, and using it to reason across unfamiliar situations. This is what we might call common sense, and it comes naturally to us, but autonomous vehicles struggle with it even in relatively simple situations. AI systems, however, are great at handling more mundane tasks and situations that require fast reactions. "It's about combining the two and figuring out how to do augmented intelligence," Hirshfield says. "Ultimately you want to calibrate trust in the system so that the human knows when to step in and when to rely on the AI."

> "Right now there are things that AI is way better at and things that humans are way better at."
>
> **Leann Hirschfield**
> University of Colorado Boulder's Institute for Cognitive Science

An autonomous car's ability to sense the world around it—and any hazards it might contain—is just one part of the equation, however. All drivers, regardless of whether they are human or machine, must contend with the unpredictability of the road. Human drivers have developed a staggering variety of informal communication methods to telegraph their intentions to other drivers, even beyond horns and turn signals. When we navigate the road, we make eye contact, flash our lights, wave our hands, or sometimes lift a middle finger to send messages to other drivers. These modes of communication can improve our safety, but they are harder to implement in an autonomous vehicle, making it more difficult for human drivers to understand the car's intentions. If autonomous vehicles are going to share our roads, it's important for them to develop a system that allows them to participate in our driver communication networks to create a shared understanding between other drivers and pedestrians about the autonomous vehicle's intentions.

This is a challenge that Boris Sofman, a Senior Director of Engineering at Waymo, and his colleagues have focused on for years. (Waymo is a subsidiary of Alphabet Inc.) They study "ridership," or best practices for how its autonomous drivers share the road with human drivers and pedestrians. From this research, they have uncovered several principles that help Waymo's robotaxis integrate themselves with the existing social contract between drivers and pedestrians on U.S. roadways. One key point, says Sofman, is for autonomous vehicles to be predictable, confident, and consistent in their actions so that humans know what to expect from the car. Sofman points to the ways that Waymo's robotaxis share the road with cyclists, which are based on extensive research about how much distance cyclists want between themselves and a car. If cyclists know that a Waymo robotaxi will give them sufficient space, they're less likely to make an unexpected maneuver to avoid a surprise from the autonomous vehicle that might put themselves or others at risk. The same is true of pedestrians crossing the road, who need to be able to predict when the car is going to proceed through the crosswalk and when it is going to let them cross.

"When a pedestrian crosses the road, you do an unofficial handshake that says, 'Okay, I'm going to go, then you're going to go,'" Sofman says. "So we actually used a lot of the inputs from the human autonomous specialists that were supervising the Waymo cars as a key signal and point of comparison so we can effectively try to mimic these very familiar human behaviors and embed them inside the vehicles themselves."

Operating a vehicle is a high-risk task that puts human lives on the line. The decision of whether to pass off responsibility to an AI driver could have deadly consequences—or it could save your life. So how is a user supposed to decide?

Catherine Burns is a professor of systems design engineering at the University of Waterloo's Advanced Interface Design Lab, where she and her colleagues study automated decision support in safety-critical systems. These are systems that have a high degree of complexity and automation and operate where human lives are often at stake. Burns is adamant that "people really shouldn't have to understand how the system works" to make a decision about when and whether to trust AI. Instead, it should feel effortless. Burns's research confirms what Sofman and the Waymo team have learned from building their autonomous drivers: Trust mostly comes down to whether the user knows what to expect from the system.

"Trust is tied really closely to reliability and the expectation of whether or not automation is going to surprise you," Burns says. "Nobody wants a surprise from their vehicle, but people can actually handle quite a bit of automation unreliability if they're aware of the possibility."

> Trust between humans and machines is hard won and easily lost.

In the summer of 2023, California's Public Utilities Commission made history when it approved two companies—Cruise and Waymo—to commercially operate their fully autonomous vehicles around the clock in San Francisco without a human in the driver's seat. The decision was controversial, but not particularly surprising. The city has been a proving ground for self-driving cars for nearly a decade, and Cruise and Waymo have operated their fleet in a limited commercial capacity for years. Limited fleets of AI drivers from various companies have also hit the streets in Los Angeles, Austin, Miami, Phoenix, and Las Vegas.

The rollout of full-time autonomous vehicles in San Francisco underscores both how far the technology has come since 2018 and the importance of properly integrating autonomous vehicles with human drivers and pedestrians on U.S. roads. Waymo's user research shows that the level of trust that San Franciscans have in autonomous vehicles has been trending upward for years as locals grow more comfortable with AI drivers roaming their streets. "Once you get the service out and people have experienced it, they realize that there's all these benefits, and they start to really like it," Sofman says.

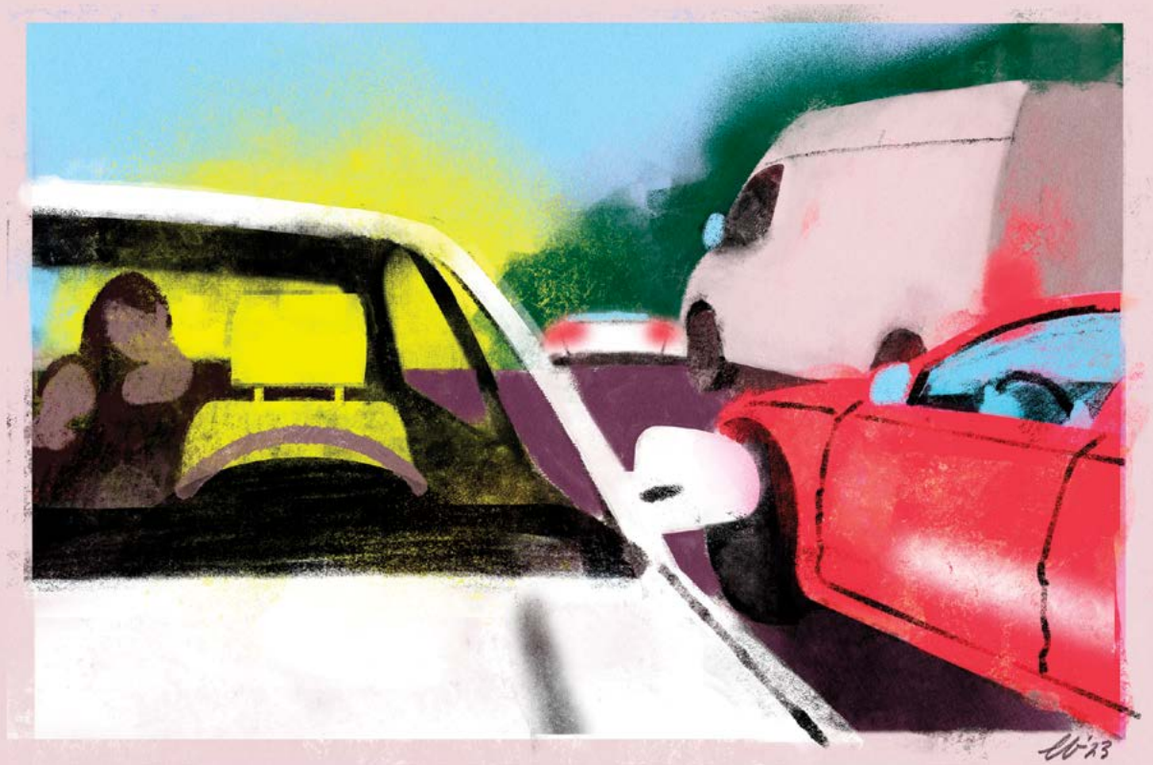
Of course, not everyone in San Francisco is welcoming of the new autonomous vehicle fleets. Some have even taken out their frustrations on the cars by placing traffic cones on the vehicles to disable them. For Sofman, these kinds of reactions aren't particularly surprising, even if they are unfortunate.

"It's just such a different technology," says Sofman, who compares distrust of autonomous vehicles to innovations such as Airbnb and Uber that many people—only 10 years ago—were initially skeptical of. But once people have tried a ride in one of Waymo's cars, he says, even the most dubious riders quickly relax. "You see that the numbers have completely flipped around in terms of trust and comfort once people have tried it," he says. "The average customer gets in our car, and within two minutes, they're on their phone checking their email or texting a friend."

One reason riders seem to find it so easy to relax during their first time in a Waymo robotaxi is because the company has put a great deal of time, effort, and research into understanding rider trust—particularly when it comes to how users perceive the safety of the vehicle. There are several subtle tactics that Waymo uses to get riders to that level of comfort as quickly as they do. For example, there's a screen inside the car that visualizes a simplified representation of what the car sees in terms of pedestrians, cyclists, traffic lights, and other cars, as well as the path that it's going to take. "When you have this very simplified but meaningful representation of the world around you, it gives you a lot of confidence," Sofman says. "It's consistent. It looks exactly like what I see. It gives you this cue that the car knows what it's doing."

Trust between humans and machines is hard won and easily lost. It's a reality that Sofman and the other autonomous vehicle engineers are acutely aware of and take extremely seriously. Although so much of the autonomous vehicle industry is understandably focused on demonstrating





safety to foster trust, human-factors researchers and UX engineers have learned that trust is a complex, multifaceted psychological phenomenon that can't be reduced to figures and statistics. It requires thoughtful and transparent approaches to the way autonomous vehicles and the companies that build them communicate with their riders, as well as a willingness on the part of riders to understand the capabilities and limitations of the vehicle. Autonomous vehicles are a technology that can have a massive positive impact on the world if passengers can be taught to trust it—and that requires coming together as users, engineers, and policymakers to build systems that are worthy of our trust.

But trust, like our roads, is a two-way street. While engineers like Sofman and his colleagues at Waymo are hard at work building autonomous vehicle systems that foster trust, the rest of us—as drivers, pedestrians, and passengers—must also reevaluate the social contract that defines our relationships on the road, and make space for autonomous vehicle technologies that hold the potential to save tens of thousands of lives every year. It won't be easy but it is possible. It requires us to thoughtfully calibrate how and why we trust these vehicles, while simultaneously accepting the reality that no system—human or machine—can operate perfectly without error. This doesn't mean lowering our standards of safety on the road; it just means giving AI drivers a fair shot by recognizing their limitations as well as their promise.

"At the end of the day, even when you've crossed the safety bar, it's important that autonomous vehicles are viewed as a positive to society," Sofman says. "That means positively interacting with the citizens and traffic around you and asking: Are you being a good citizen?"

---

Daniel Oberhaus is a science writer and the founder of HAUS Biographics, a marketing and communications agency for deep tech organizations. He is the author of *The Silicon Shrink*, a forthcoming book from MIT Press about the past, present, and future of AI in psychiatry, and was previously a staff writer at *Wired* magazine.

Q&A

# Unlocking Life's Building Blocks

Mapping proteins unlocks clues to preventing and treating illnesses. Here's how DeepMind's AlphaFold is accelerating this research exponentially.

**In 1997,** while Demis Hassabis was a student at Cambridge, IBM Deep Blue, a chess program, defeated grandmaster Garry Kasparov—the first victory of a computer over a reigning chess champion under tournament conditions. It was a moment that would prove formative for Hassabis, a former child chess prodigy who had used his tournament winnings to buy his first computer.

Why? Because it was the human being who lost the match, not the technology that won it, that most impressed him. "Kasparov could not only play chess more or less to the same level as this brute of a calculation machine," Hassabis says. "[He], of course, could [also] ride a bike, talk many languages, do politics, all the rest… Deep Blue, brilliant as it was at chess, couldn't do anything else… Something was missing from that system that we would regard as *intelligence*."

The search for that absent intelligence has driven Hassabis's career ever since.

Proteins underpin the biological processes of every living thing... By understanding the structure of proteins, we can dramatically deepen our understanding of health, disease, and the environment.

He began his career as a teenage designer of best-selling video games (and, like after his tournament winnings, used the proceeds to pay his way through Cambridge this time). In 2010, a few years after earning a Ph.D. in cognitive neuroscience, he co-founded DeepMind. His goal: to create artificial general intelligence (AGI) and to use it to achieve the widest possible social impact.

In 1972, accepting the Nobel Prize for Chemistry, Christian Anfinsen had thrown down the gauntlet for what would become known as "the protein-folding problem." Proteins are the building blocks of life. Every enzyme and hormone in the body is a protein. They're responsible for everything from digestion and neurological function to growth, repair, and reproduction.

Proteins, like DNA, are made up of chains of amino acids. Each one helps to determine the protein's structure through its kinetic interactions with the others. Like a magnet, each chain, singly or in combination, has a particular valence; all together, they guide the protein into assuming its final three-dimensional shape. Only then, after the folded protein has correctly positioned its various channels, receptors, and binding sites, can it function. Glitches in this process are implicated in diseases as different as cancer, diabetes, and Alzheimer's.

The amino-acid chains of proteins were no secret to researchers. Therefore, Anfinsen argued, scientists ought to be able to extrapolate from them the three-dimensional shape a given protein would assume. As an idea, it seemed simple enough; in practice, it meant confronting an almost unfathomable complexity. The protein-folding problem would baffle scientists for the next five decades—a kind of Fermat's Last Theorem for the field of biology.

Of the 200 million proteins in nature, scientists had slowly and painstakingly documented the structure of approximately 150,000. Using this dataset, DeepMind in 2016 began training the AI system AlphaFold to predict the structure of the others. "I wanted to finally apply the AI to real-world domains," Hassabis told a journalist for *Scientific American* in 2022. "Protein folding was right up there for me always, since the 1990s." The first iteration of the technology, released in November 2020, fell short of the required "atomic" level of accuracy, but AlphaFold 2—a complex architecture of 32 component algorithms—essentially solved protein-folding. By July 2022, DeepMind's database encompassed all of the 200 million known proteins. AlphaFold 2 was, a *Forbes* columnist declared, "the most important achievement in AI—ever."

We caught up with Demis Hassabis to ask him about present and future applications of this breakthrough technology, as well as some of the issues it has raised involving privacy, safety, and ethics. Our conversation, edited and condensed for clarity, is below.



Question **It took more than 50 years for science to solve this problem. Why was protein-folding so tough to crack? And what does solving it mean, exactly? In other words, how quickly can AlphaFold enumerate the possible configurations of a single protein today?**

Demis Hassabis There's an astronomical number of potential shapes a protein could theoretically fold into, by some estimates 10^300 (10 to the power of 300, which is a 1 followed by 300 zeroes), which would take longer than the age of the universe to search through. And yet somehow in nature proteins spontaneously fold in fractions of a second. This is sometimes referred to as Levinthal's paradox, and it is this complexity that makes the problem so tough to crack.

It often takes a graduate student their entire Ph.D. to experimentally determine the structure of just a single protein, and after decades scientists had only been able to determine around 150,000 protein structures experimentally. This is the problem we wanted our AlphaFold AI to solve, by making it possible to predict protein structures quickly and accurately directly from the amino acid sequence (roughly the genetic sequence for the protein).

Question **This speed isn't just an astounding technological achievement. It also has important repercussions for AlphaFold's applications in the real world. For instance, limitations of both manpower and time have significantly hampered research into potential treatments for neglected diseases. What does the database's accessibility, in concert with AlphaFold's speed, accuracy, and negligible cost, mean for research into potential treatments for neglected diseases?**

Hassabis Some diseases disproportionately impact communities in less affluent parts of the world, which also have fewer resources for researching new treatments. Making our AlphaFold predictions freely available to anyone in the scientific community is making a big impact here, and already more than 1 million researchers in 190 countries have accessed the AlphaFold Protein Structure Database, and many of these researchers would not have access to the expensive experimental facilities needed to determine the structures of the proteins implicated in the diseases they were studying.

One of AlphaFold's earliest adopters, the Drugs for Neglected Diseases initiative (DNDi), has used AlphaFold to advance research into diseases like Leishmaniasis and Chagas disease that disproportionately affect the poorer parts of the world. We've also supported World Neglected Tropical Disease Day by creating structure predictions for organisms identified by the World Health Organization as high priority for their research, which is helping with the

study of diseases like leprosy and schistosomiasis, which have impacted more than 1 billion people globally.

Question **Researchers are also using AlphaFold to do some particularly interesting de novo protein design of applications external to the human body. They're creating proteins not found in nature to serve climatic and environmental purposes. Can you talk about some of those applications and their social impact?**

Hassabis We're seeing researchers use AlphaFold to study protein design and specifically enzymes, which could be particularly valuable in helping us achieve a more sustainable future. For example, a team at the University of Portsmouth has been using AlphaFold in their work to discover and engineer enhanced enzymes that can eventually be applied at scale to break down some of the most polluting single-use plastics. We also know that scientists are using AlphaFold to explore carbon-capture technologies.

Question **DeepMind decided to release the coding for AlphaFold publicly and to upload, free of any charge for access, its enormous database of protein structures—"a gift from us to the scientific community," as you've said. To date, 500,000 researchers have used it, which you believe to be the vast majority of the biologists in the world. But before releasing this data, DeepMind consulted with 30 bioethicists about the safety of doing so. What ethical concerns did the bioethicists bring to your attention, and how did you address those concerns?**

Hassabis Before releasing AlphaFold we consulted a range of experts, including bioethicists, as well as experts from fields like protein engineering and biosecurity. They determined that the risk of releasing AlphaFold was likely to be low and that the benefits far outweighed the risks. We have a very rigorous program for ensuring our technology is developed and deployed in a way that's safe, responsible, and ethical, including ongoing engagement with biosecurity experts.

Question **At the same time, the technology itself doesn't seem to function transparently. AI has been described as a "black box": its reasoning is incomprehensibly complex and opaque, which means we can't reverse-engineer the steps it takes to arrive at its conclusions. If we don't know how it reaches its conclusions, we perhaps don't know what those conclusions are going to be, either—and perhaps they won't be hospitable to us as human beings. In the worst-case scenario that organizations such as MIRI (Machine Intelligence Research Institute) have articulated, humankind itself will be at risk from a superintelligent AI. Before this technology becomes sentient, numerous scientists have warned, we must solve the so-called alignment problem—we must be able to train machines so that their interests and ours are permanently and inseparably aligned.**

Hassabis AI systems are not actually black boxes: Unlike the brain, we can in principle inspect every weight and activation of an AI system. However, the extraordinary complexity of neural networks means that even with current advances in the science of interpretability, we have a long way to go before we can meaningfully understand them.

AI is an engineering science: we need to first build an AI system before we can take it apart and study it. Advances in AI increase the challenges with respect to safety, but they also amplify our ability to conduct AI safety research, by giving us more advanced systems to study and assist us.

As we begin to build increasingly more powerful and general systems, one promising idea would be to first test them in hardened simulation sandboxes and conduct safety evaluations, only later deploying them into the real world, once we have gained confidence in their safety. For this we need to advance the science of scalable alignment—methods to train models to do what we intend, that will scale with their increasing capability. Ultimately, success at scalable alignment is critical for unlocking the vast benefits of advanced AI in health, science, and well-being.

Question **AI is very probably the most powerful tool humankind has ever built, and it's largely in the hands of the private sector. Earlier this year, you co-signed a 22-word statement, along with many of the world's most eminent researchers, scholars, and ethicists—among them some of them your peers at Google. The statement reads, "Mitigating the risk of extinction from AI should be a global priority, alongside other societal-scale risks such as pandemics and nuclear war." There has been much discussion about creating a code of ethics to govern work involving AI, as you know. To what degree should members of the public be concerned about the motives of some Silicon Valley entrepreneurs when it comes to AI, particularly given its vast, possibly even unimaginable rewards in terms of both power and money?**

Hassabis We believe the right way to respond to this moment in AI is with cautious optimism—with a firm grasp of the incredible benefits that AI could create, but also a sober understanding of the near and long-term challenges that we need to prepare for.

AI promises extraordinary new capabilities and opportunities to help us solve some of the biggest challenges



There's an astronomical number of potential shapes a protein could theoretically fold into... which would take longer than the age of the known universe to calculate.

of our time. It has the potential to help us to cure diseases, to deliver a more sustainable future for the world, and to unlock a new era of greater prosperity and opportunity for humanity.

Alongside this incredible potential, AI will obviously create some big challenges. We've always been committed to pioneering safely and responsibly, and have created industry-leading technical safety, ethics, and governance programs. These will remain important priorities for us, but we're also working to drive action across the tech industry, government, and society, so that other innovators and leaders are preparing responsibly for the future.

Question **You've said that you created DeepMind in the image of Bell Labs, the celebrated R&D division of AT&T whose researchers earned eight Nobel Prizes in numerous disciplines. DeepMind's mission statement, you have said, is "Step one, solve intelligence; step two, use it to solve everything else." What does this mean, and how does it guide DeepMind's ambitions? How does it influence what problems you choose to take on (including the protein-folding problem)?**

Hassabis When we set up DeepMind, I took inspiration for our research culture from many innovative organizations, including Bell Labs and the Apollo program, but also creative cultures like Pixar. Our fundamental goal has always been to create AI technologies that can help us better understand the world around us and solve a lot of important challenges facing society from curing diseases, to creating a sustainable future, to powering products that enrich the lives of billions of people in their daily lives. Aiming for that kind of scale and impact is what drives our efforts.

Question **Looking to the (perhaps distant) future, what huge problem that seems impossible to resolve today—be it scientific, technological, societal, or otherwise—seems conceivably solvable to you with the assistance of AI?**

Hassabis There are many huge scientific and mathematical problems I have on my list to solve (one of them was protein folding!). Modeling a virtual cell has been one of my dreams for a long time. If you could build a highly accurate simulation of a cell using AI, that was capable of making useful predictions, it would be incredible for the understanding of biology as well as things like drug discovery. Lots of experiments could be conducted quickly and cheaply in the virtual cell, and then only at the last stage would the predictions be validated in the wet lab.

This would be revolutionary for processes like drug discovery. Currently it takes roughly 10 years to go from identifying a target to having a drug candidate. With a virtual cell, you could potentially massively shorten those timescales down to months instead of years, by much more efficiently exploring the search space of possible compounds. I think getting to a virtual cell might be possible in the next decade, and it's something I'm really excited about.

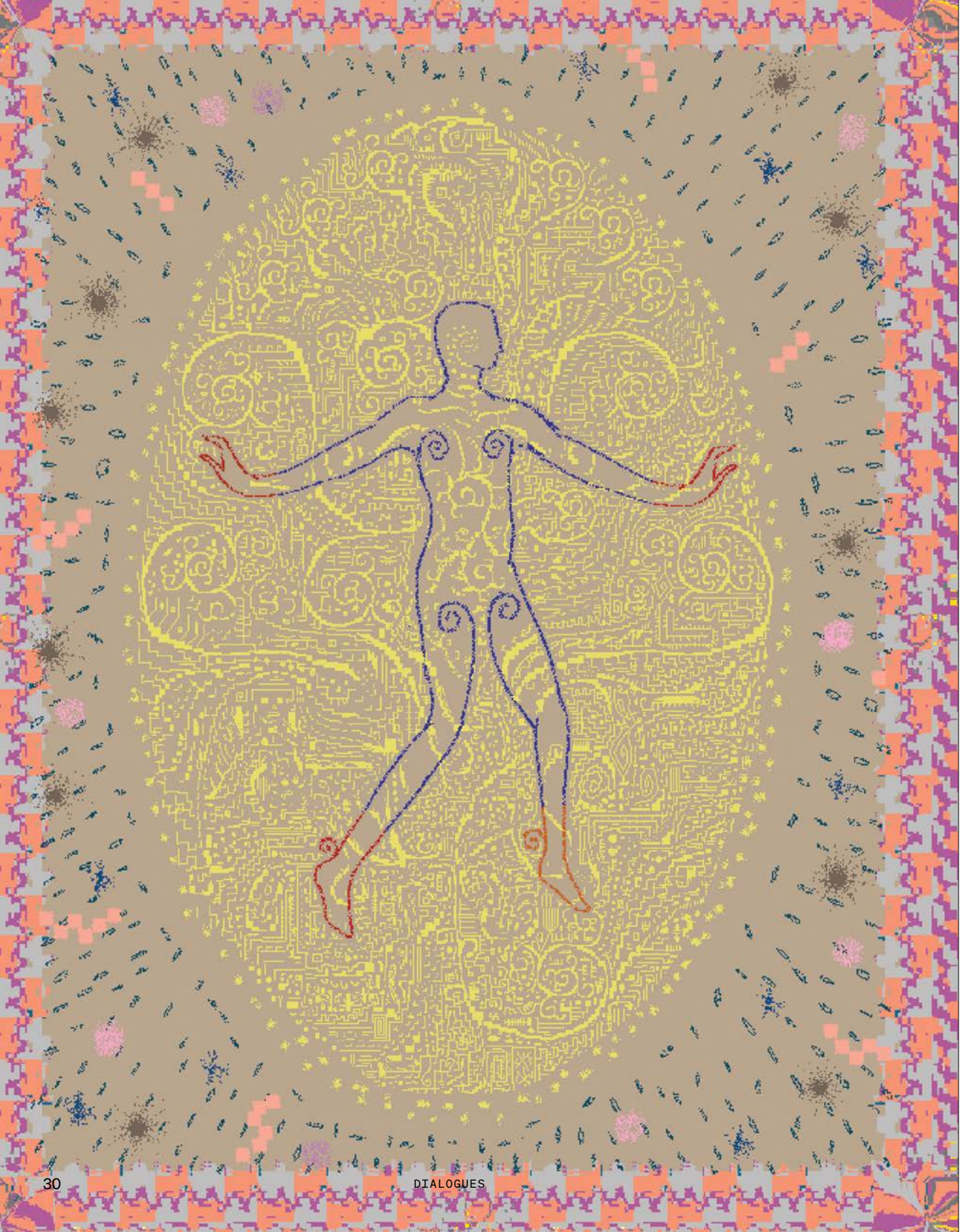
# How can AI help us better understand ourselves?



AI promises to revolutionize our comprehension of biology. Where once we grasped at straws, AI offers illumination—decoding the mysteries of disease, perception, and cognition. In psychiatry, AI can help doctors parse the neural roots of disorders like schizophrenia, suggesting targeted therapies for each patient's unique mind. Precision oncology tailors treatments to attack cancers with surgical specificity. In immunology, AI could even help decode the intricacies of the human immune system—which, as Jane Metcalfe writes in these pages, is "many orders of magnitude more complex than the genome." Understanding it could unlock revolutionary insights into predicting and preventing disease.

These breakthroughs offer but a glimpse of a deeper transformation underway. AI expands the bounds of knowledge and gives rise to new fields of inquiry. Its analytical power enables connections across disciplines, linking biochemistry with psychiatry, genetics with immunology.

It also raises probing questions: Artificial neural networks, initially inspired by our understanding of the human brain, serve as a reflection of its complexity. As we develop and refine these networks, might they bring us closer to unraveling the intricacies of our own neural architecture? What might decrypting the mysteries of cognition reveal about the essence of mind and memory? Like the microscope and telescope before it, AI brings into focus new worlds to explore.





←
Illustration by
Somnath Bhatt

# Modeling the Human Immune System

A new open-science project aims to harness the power of machine learning to decode the human immunome.

By Jane Metcalfe

**When the Human Genome Project** mapped and sequenced the entire set of DNA instructions in a human cell—a 13-year, $2.7 billion global endeavor that completed in 2003—it produced one of humanity's greatest accomplishments. The project was a historic milestone for dramatically accelerating biomedical research and forever altered the way we practice medicine. But it didn't fully answer the question of why we get sick, since genetics only accounts for roughly 20 percent of diseases. Social determinants of health—such as economics, education, the physical environment, racism, and sexism—explain some of that. Yet even people facing similar circumstances respond differently when exposed to the same immune system challenge, such as a pathogen or a vaccine. What accounts for the extreme variation in human immune response to the various threats and insults of life?

Those differences are embedded in one of the most complex systems in life science: the human immune system. Twenty years after the human genome was sequenced, a new moonshot is taking shape. It's many orders of magnitude more complex than the genome and could, in turn, be orders of magnitude more useful than the genome for understanding human health and disease. The next moonshot in life sciences is decoding the human immune system.

Unlocking the secrets of our immune system offers the tantalizing possibility that we can one day understand who will get sick, how their disease will progress, and which

As the wave of AI breaks over life sciences, biotech, and medicine, it's time to move beyond two-dimensional mapping.

interventions will work best for each individual. Imagine that doctors could predict who will get cancer, whether it will metastasize, and how to intervene with early treatment or even preventative measures. "Fifteen years ago, that would have sounded like science fiction," says Shai Shen-Orr, founder and chief scientist of CytoReason, a company building computational models to connect the molecular features of patients' lab work to their clinical outcomes. But now, thanks to AI, it is within reach.

So far, machine learning has largely been applied to specific health-care tasks using one type of data—for instance, retinal scans, which are being used to predict and diagnose a host of medical conditions ranging from kidney disease and heart attacks to Alzheimer's disease. The same is true of electrocardiograms, which when boosted with deep learning are capable of determining age and sex and detecting cardiac dysfunction, anemia, and more. Computer vision is increasingly allowing us to characterize the molecular basis of tumors, which can guide the clinician in treating tissue cancers.

But as the wave of AI breaks over life sciences, biotech, and medicine, it's time to move beyond two-dimensional mapping into three-dimensional modeling and systems thinking—and to take those capabilities into the fourth dimension by modeling people over time, from a newborn's developing immune system to an elderly person's failing one. That's a giant leap, and it requires a step back from specific tasks to think more globally about the astoundingly complex system that is the human immunome.

*

The immunome includes all the molecules, proteins, cells, tissues, and organs of the immune system, as well as all their interactions with the body's other biological systems, such as the genome, epigenome, microbiome, and metabolome. And it includes the exposome, which are external factors (or inputs) like stress, pollution, and diet.

Can AI sort it out? Leading experts think so. "What's striking is that now we're seeing self-supervised learning and unsupervised learning," says Eric Topol, founder of the Scripps Research Translational Institute and author of the book *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. "We're putting essentially everything we know about the immunome into these large language models, which can help sort and interpret generally—and not just the code, but the context and the ideas. This is a very auspicious time for AI."

But while the AI technology is ripening, the immunological data to feed it is missing. Currently, hospitals, clinical research organizations, and medical practices generate mountains of health data on a daily basis, but detailed information about the immune system is lacking. Wearable sensors, smartphones, and other self-tracking devices and apps are contributing more health data than we've ever had, but even all of that will not provide enough usable information of sufficient resolution to fuel models that would explain the human immunome.

To generate the data required, researchers would need to analyze all the cells and cell types in the immune system and determine how many cells there are, what state they're in, and the molecular signals that determine that state. Doing that requires measuring the signaling proteins and metabolites, including hormones. Even with recent developments in multiplexing, that's still a lot of very advanced and expensive scientific testing.

Beyond data collection, there are a host of data structure issues that need to be addressed: ownership, privacy, and security considerations, and the need for informed consent. Other ethical concerns include the need to have a representative sampling of the entire human population—not just those who have access to and can afford the tests, a problem that plagued the original human genome sequencing effort.

And that's just the beginning of the challenge. Once researchers have the data, they'll still have to figure out how to build a model that can process a wide range of inputs, including numerical data, text-based data, and still and moving images. How do you label and tag that data in consistent and meaningful ways? Can you get sample sizes that are large enough to be statistically significant? The ultimate challenge will be to build models that not only generate predictions but also explain the underlying mechanisms at work.

While large language models are very good at predicting what the next word in a sentence will be, the extent to which they can identify some of the intricate underlying biological causal structures is not yet clear. "They may be able to predict what the next state of the immune system will be, but it's much more difficult to know how to intervene to alter the future trajectory of the immune system so that this person will be healthy again rather than continuing down the path of disease to a worsening state," says John Tsang, founding director of the Yale Center for Systems and Engineering Immunology.

Researchers, inspired by large language models, are building foundation models of genes and cells, akin to words and paragraphs in natural language. As more single-cell data sets become available, researchers are able to use them to help the models learn complex relationships between cells and gene expression. The models can, for instance, predict how a particular T cell would respond to an external signal, or what state the cell will be in and how it functions.

"We need to learn the language of the immune system, particularly its cells and how the cells communicate," Tsang says. "And that's something that we are really just at the infancy of understanding."

*

In the summer of 2023, the Human Immunome Project (HIP), an organization I chair, unveiled a new research plan: a comprehensive, scalable open-science effort to collect the advanced cellular and molecular data needed and feed that data into building machine learning models. No matter how the endpoint is defined, the project will catapult the field of immunology into the machine learning age, creating many new possibilities for drug discovery and development, and the diagnosis and treatment of disease.

"Decoding and modeling the human immunome is a scientific and medical feat of unmatched proportions—and has the opportunity to truly transform how we think about and practice medicine," says Hans Keirstead, a pioneering neuroscientist, a serial biotech entrepreneur, and the CEO of HIP. "We can change the trajectory of global health for the better."

Keirstead's plan, developed with CytoReason's Shen-Orr and Yale's Tsang, is to select regional scientific centers on six continents, each of which will sample and analyze a comprehensive and diverse population ranging from young to old, sick to healthy, and of multiple racial, ethnic, and socioeconomic backgrounds. To harmonize all the data captured, the team is developing a proprietary multimodal immune monitoring toolkit that will be engineered to withstand difficult environments and support privacy, security, cost, standardization, and interoperability optimization.

The project is also working to "avoid the pitfalls of the Human Genome Project" by maximizing representation and minimizing bias, according to Shirin Heidari, a researcher at the Graduate Institute of International and Development Studies in Geneva and a former virologist. Heidari, who has consulted with HIP, says these measures are "essential not only to ensure better population diversity but also to consider the validation, optimization, and standardization of these assays across different sexes and populations to accurately capture variations."

The only way this plan can work is if the HIP plays a significant role in directing and coordinating the project, centralizing logistics and scientific oversight, and housing administrative functions. "This effort is so grand on the operational side of it, and the technology and the science that needs to be solved, that no one lab can do this and no one company—this is greater than all of those together. It requires a new type of thinking and a new kind of collaborative science," says Shen-Orr.

HIP's leadership takes inspiration from the European Organization for Nuclear Research model of global collaboration, which asks that normally competitive physicists from across the world, including countries hostile to each other in other settings, put aside politics and competition because access to the particle accelerators is so valuable. The diverse immunological database HIP is building and the AI models that will result will be a comparable asset—the largest collection of open-source, standardized, state-of-the-art immunological data in the world. "If we can pull the right people and approaches together, magic is going to happen," Tsang says. "Things are emerging already and just need the right catalyst to make it come together."

It's hard to imagine getting to real precision medicine without this foundation.

Jane Metcalfe is chair of the Human Immunome Project, CEO and founder of proto.life, and co-chair of the Council of the Focused Ultrasound Foundation. She is also co-founder and former president of *Wired* magazine.

# The Data-Driven Future of Mental Health Treatment

↔
Illustrations by
Marine Buffard

By Daniel Oberhaus

With the help of AI, a technique known as digital phenotyping aims to correlate digital behaviors with mental disorders.

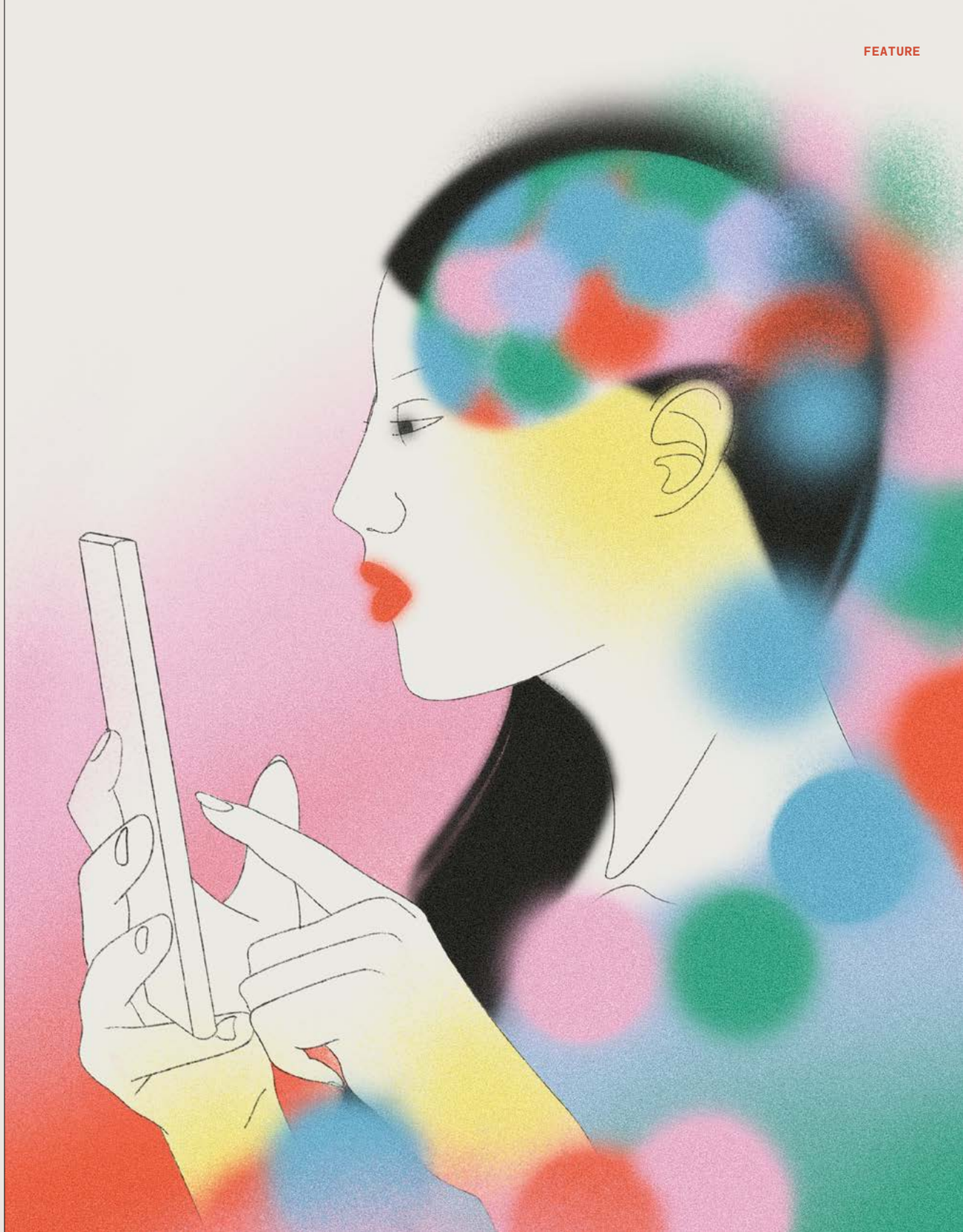
**From Tom Insel's perspective,** the future of psychiatry had never looked so bright. It was May 2015, and Insel, then the director of the National Institute of Mental Health, the world's largest and most well-funded research institute focused on mental disorders, had traveled to Portland, Oregon, to speak with the parents of young children grappling with serious mental disorders. He had good news: Researchers were making rapid progress in uncovering the biological basis of serious mental disorders. NIMH researchers who were studying high-resolution brain scans of people with depression had found abnormal neural branching in stem cells from children with schizophrenia. They now also understood how stress results in genetic changes in mice. But as soon as he opened the floor for questions after his presentation, he found that not everyone was impressed.

The first person to grab the microphone was a tall, bearded man in a flannel shirt. Insel had noticed him growing increasingly agitated during his presentation. "He said, 'Man, you just don't get it,'" Insel recalls. "'I have a 23-year-old son with schizophrenia. He's been hospitalized five times, incarcerated three times, which led to suicide attempts, and he's currently homeless. Our house is on fire, and you're talking about the chemistry of the paint.'"

The father's remark left Insel speechless. His initial reaction was to defend NIMH's work by offering the man some platitudes about how scientific revolutions take time and basic research was required for better treatment. But deep down, he agreed with him. Despite the remarkable progress made by mental-health researchers at the NIMH and around the world, during his 13 years at the helm of the institute, deaths from suicide had increased by 33 percent, deaths from addiction had tripled, and the number of homeless and incarcerated people with serious mental disorders had doubled.

"That was a wake-up call for me," Insel says. "It's not a knock on the NIMH because it's not their job to keep people with serious mental illness out of the criminal justice system, and it was obvious that what we were doing may help in the long run. But we're in a mental-health crisis, and it became my calling to figure out what it would take to put out that fire."

Four months after that fateful presentation, Insel resigned from his position as director of the NIMH and began aiming his work at using AI and other digital technologies to help address the mental-health crisis. "It was really exciting to look at how we could use AI to bend the curve for people



with schizophrenia, bipolar illness, kids who are suicidal, and really try to have an impact at the public-health level," Insel says.

In the years since, Insel has co-founded three AI-driven mental-health start-ups and advised several others. But his focus has remained on finding ways to correlate digital behaviors with mental disorders, a technique known as digital phenotyping. Digital phenotyping draws on the reality that both AI and mental-health professionals are fundamentally in the business of pattern recognition. Psychiatrists, therapists, and social workers study the behaviors of their patients for evidence that their mental health is improving or deteriorating so they can adapt their diagnoses and treatments. But the data available to the human professionals who make these judgments is confined to patients' notoriously unreliable self-reports and a limited amount of behavioral observation. Digital phenotyping, by contrast, can passively analyze digital behavioral data from consenting patients around the clock to identify patterns that would otherwise escape the notice of a human therapist.

The promise of digital phenotyping bringing data-driven "objectivity" to the diagnosis and treatment of mental disorders was a seductive one for Insel and many of his peers. While traditional mental-health research struggles to bridge the gap between the lab and the clinic, digital phenotyping could—in theory—be immediately applied in the real world, where the vast majority of people already carry sophisticated computers in their pockets. The data produced by these devices could, for instance, help flag the onset of depressive or suicidal episodes, which would allow mental-health professionals to make real-time interventions when needed. It could also help therapists, psychiatrists, and other mental-health professionals to monitor the progress of their patients. If they notice that a patient has stopped sleeping or socializing, for example, a mental-health provider can work with the patient to correct behaviors or adjust treatments.

Digital phenotyping may hold the key to improving outcomes for people living with a broad range of mental disorders. Now Insel and other researchers are trying to develop the tools that can deliver on this promise.

*

Over the past few decades, mental-health researchers have increasingly focused on uncovering the biological basis of the cognitive, emotional, and behavioral dysregulation characteristic of mental disorders. This fixation on biology is largely because psychiatry remains the only medical field that has yet to uncover a single unambiguous biomarker for any of the nearly 300 disorders listed in the *Diagnostic and Statistical Manual*, the official taxonomy of mental disorders.

In the absence of any biomarkers, it's hard for researchers to know what, exactly, they are looking for. The hundreds of disorders in the *DSM* are defined by clusters of symptoms and a threshold for how many of those symptoms must be present for a patient to be diagnosed. Although this system helps mental-health professionals standardize their approach to diagnosis and treatment, decades' worth of data shows that symptom-based diagnosis is unreliable because of its dependence on subjective clinician judgment. "I like to say that developing a new antidepressant with the diagnostic system we have today would be like developing a new antibiotic for someone with a fever when you don't know if it's caused by a bacterial or viral illness," Insel says. "It's really critical that we get better precision around diagnostic categories if we want to do better interventions. Otherwise, we're just treating the fever, and we'll never really know what's in front of us."

If there are 227 possible symptom combinations for a diagnosis of depression—which pales in comparison to the roughly 60,000 possible symptom combinations for post-traumatic stress disorder—how can researchers looking at the brains of depressed patients be sure that these patients have the same disorder? Although the *DSM* is still widely used, Insel announced in 2013 that the NIMH would no longer be funding research based purely on *DSM* diagnostic criteria. Insel is adamant that subjectivity—particularly the lived experience of patients—has a critical role to play in treatment, but it's clear that the field of mental health has an acute objectivity problem, in part because researchers and clinicians lack data, which makes it difficult to better patient outcomes.

It's a challenge that John Torous knows all too well. As the director of the digital psychiatry division at Harvard Medical School's Beth Israel Deaconess Medical Center, Torous splits his time between clinical practice and academic research largely focused on the application of digital tools in mental health. During Torous's psychiatry residency at Harvard, he witnessed firsthand the struggles people with serious mental disorders face and the shortcomings of conventional approaches to diagnosis and treatment. His background in computer science led him to seek digital solutions, leveraging the capabilities of the internet, mobile phones, AI, and other technologies to improve patient outcomes. In 2015, Torous and three of his colleagues published a paper in which they first defined digital phenotyping as "the moment-by-moment quantification of the individual-level human phenotype in-situ using data from smartphones and other personal digital devices."

By the mid-1960s, less than a decade after the term *artificial intelligence* was coined, a handful of psychiatrists were already experimenting with using AI systems to simulate mental disorders for research purposes. But despite decades of research on AI in psychiatry, all of these pioneering experiments fell short of bettering outcomes for patients. They were all missing one crucial ingredient that is the grist of modern AI systems: data.

The proliferation of internet connectivity and mobile devices over the past two decades changed everything. "I think the potential of a new tool and new data sources to understand human behavior is of paramount importance to the field," Torous says. "It's not that we have more *objective* data, but we have new complementary sources of data to better understand our patients' behavior."

Torous and his colleagues recognized that it might be possible to draw from this digital exhaust—the way a patient types on their computer or scrolls on their smartphone, the biometric data collected by their wearables, and so on—for valuable insights into people's mental health. Theoretically, that information could allow mental-health professionals to identify the onset of a patient crisis and stage an early intervention, help patients better understand their own mental-health status, and assist researchers in developing a more refined picture of mental disorders.

> Can researchers tell if an individual is depressed based on changes in the way they scroll on their phone or type on their computer?

Digital phenotyping involves a broad range of digital technologies and data types, but the basic idea behind all digital-phenotyping systems—regardless of the technologies or data types used—is the same. Certain patient behaviors are associated with certain mental disorders, and many of these behaviors can be measured by how we interact with digital devices. For example, AI can analyze the volume and duration of a patient's call records—disregarding the content of those calls—or their geolocation data as proxies for their social isolation. If the data shows that the patient has stopped leaving their home and answering calls, it may indicate the onset of a depressive episode. The digital phenotyping system can then flag these behaviors for a mental-health professional, who can contact the patient to provide support.

Aside from the types and volume of data, one of the key attributes that differentiates digital phenotyping from other digital approaches to monitoring patient behavior is the way the data is collected. In contrast to, say, an ecological momentary assessment—essentially a patient survey that can be periodically delivered by phone or computer—digital phenotyping is passive and doesn't rely on patients' self-reports. With permission, it can monitor digital behavior 24/7 without intruding into the patient's life. It is collecting behavioral data as patients go about their day in their normal environment, which should theoretically provide better behavioral insights than patient surveys or data collected in an artificial clinical environment.

*

Over the past decade, a growing body of evidence suggests that digital-phenotyping systems are capable of reliably identifying some disorders based on digital behaviors, which is the critical first step toward improving patient outcomes. In 2017, for instance, a group of researchers from Harvard, MIT, and other Boston-area universities ran a digital phenotyping study with 73 participants that predicted symptoms of depression and PTSD using a variety of digital behavioral data, including the way a phone was handled, messaging frequency, GPS location, and vocal cues. In 2023, Torous and a team of international collaborators published a study showing that it was possible to predict relapse in people with schizophrenia using a mix of passive data sources such as geolocation and screen state along with active data such as surveys. Despite these promising results, some questions remain about the efficacy of digital phenotyping.

An active area of research, for example, is on what types of data correlate best with symptoms of a given disorder. Consider a study published in 2017 in the journal *IEEE Transactions on Biomedical Engineering* by researchers from the University of Oxford that showed that "it is possible to detect depressive episodes in individuals with bipolar disorder with 85 percent accuracy using geographic location recordings alone." Could the accuracy be increased by incorporating other types of data such as vocal cues and call-record data? If a single data type is sufficient for detecting depressive episodes in bipolar patients, will it apply to other disorders, or are different data types more relevant to some disorders than to others?

The answers to these questions have real consequences for the technology and its users. More targeted data collection could help further protect the privacy of the patients using these systems. Already it's clear that not all data types are created equal. Both Torous and Insel pointed to sleep data as an example of a datastream that has proved beneficial for helping patients with a broad range of

disorders. Sleep behavior is well characterized, and research has established strong connections between sleep dysregulation and mental disorder. But what about something more experimental? Can researchers tell if an individual is depressed based on changes in the way they scroll on their phone or type on their computer?

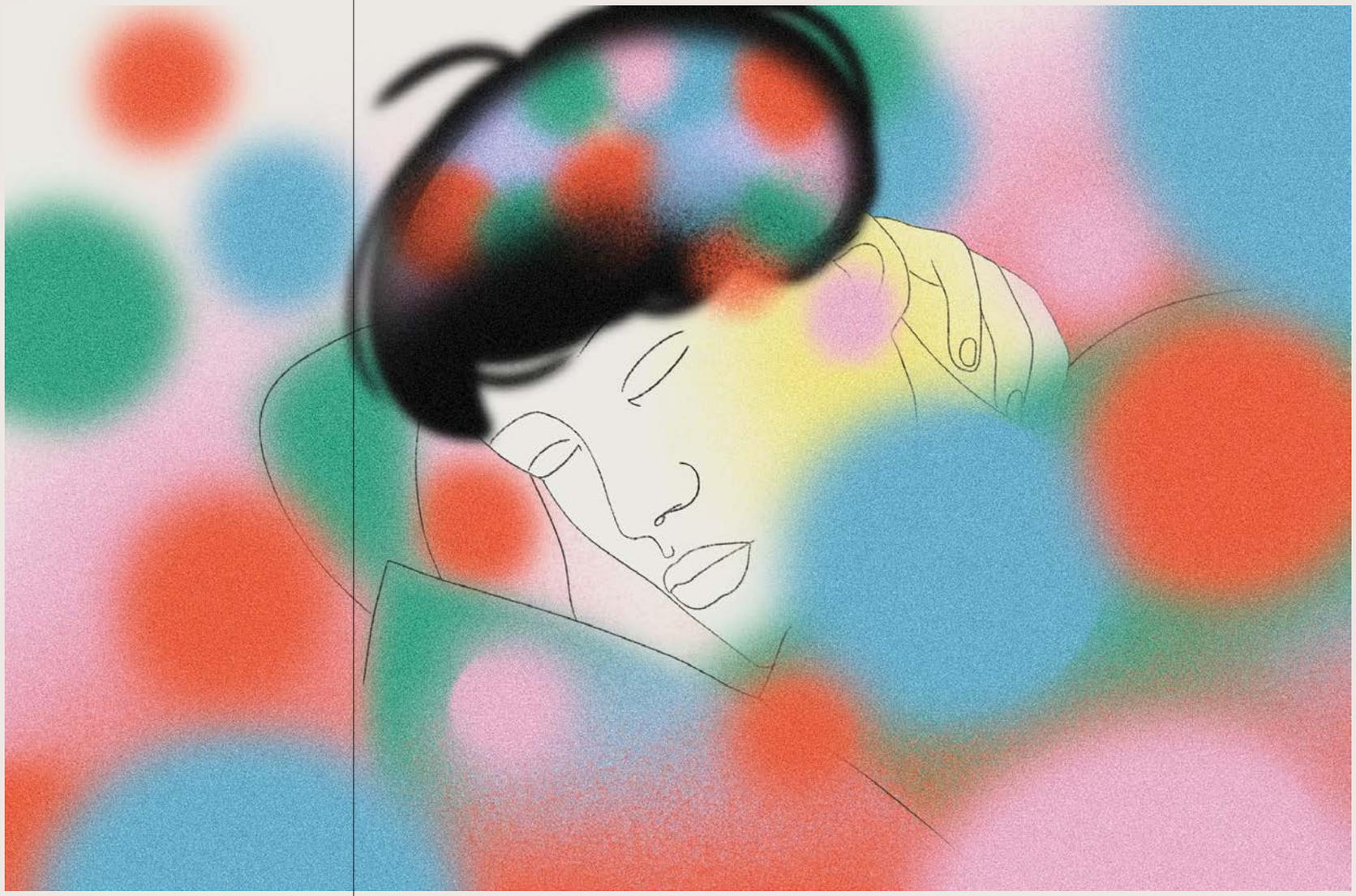
Answering that question is remarkably challenging. In several recent meta-analyses of digital-phenotyping studies on patients with psychotic disorders, researchers found that the accuracy and the effectiveness of these systems vary widely depending on the types of machine learning methods used. The lack of standardized research protocols makes it difficult to generalize results across studies. Moreover, most digital-phenotyping studies use relatively small patient populations, last only a few months, and are plagued with methodological shortcomings—including providing patients with a study-specific smartphone, which could skew the data in ways that using a participant's own phone would not, or failing to collect basic patient information like age. So, as some studies show that AI can indeed detect mental disorders based on digital behavior and others show that it cannot, larger standardized trials still need to be conducted.

If digital phenotyping one day proves to work, the most important question is whether it will make a difference in the mental-health crisis. Properly identifying and tracking mental disorders is a massive challenge—but it's not the only one. Many people with these disorders don't have access to mental-health professionals or these forms of care at all. For those who do, there are doubts about the technique itself. "What these methods are trying to do is take someone's qualitative lived experience and reduce them to a number that can tell you if they're about to have some mental-health episode," says Gabrielle Samuel, a lecturer in the department of global health and social medicine at King's College London. "But there are so many different possible reasons [for a person's behavior], and it's making assumptions about the way people live. My concern is that automation moves you further away from the person in front of you, and that distance is what's problematic."

Samuel is also skeptical that digital-phenotyping technologies will help the patients most in need. For example, mental disorders are far more prevalent in incarcerated and homeless populations, many of whom don't have access to smartphones or may reject digital-phenotyping systems because they are concerned about being constantly monitored. "We're throwing a huge amount of money into these technologies as though they're going to solve mental-health issues," Samuel says. "But, actually, it's not going to be a solution to problems with mental health because so much of mental health is socially, economically, and politically determined."

Insel is the first to admit that the technology has a long road ahead of it before it can really deliver on its promise. "We're in Act One of a five-act play," he says. "The first act has shown us that there's real potential here, but we're still not yet fully realizing it."

Torous and Insel acknowledge these shortcomings of digital phenotyping and emphasize that they don't see the tool as a silver bullet for the mental-health crisis. But while researchers like Torous are intent on refining these technologies in the lab and conducting foundational research, digital-phenotyping start-up founders like Insel are racing to get these technologies to patients in need. The sense of urgency is understandable given what these tools, once fully developed, could do: Just-in-time interventions for patients in crisis, treatments that adapt to the lived realities of patients, and consensual monitoring to supplement patients' self-reports are all within reach.



As with so many issues in mental health, both approaches have upsides and downsides in service of the same goal: solving society's mental-health crisis and delivering relief to millions of people living with mental disorders. "A lot of people say that it's not a perfect system, and I get that, but for me the question is 'Compared to what?'" Insel says. "Patient outcomes for the past two decades have gotten worse despite more treatment and more money being spent. So if we continue to ignore data on how people think, feel, and live, I don't think patient outcomes will get any better. You can't improve the quality of care until you start to measure it."

Daniel Oberhaus is a science writer and the founder of HAUS Biographics, a marketing and communications agency for deep tech organizations. He is the author of *The Silicon Shrink*, a forthcoming book from MIT Press about the past, present, and future of AI in psychiatry, and was previously a staff writer at *Wired* magazine.



# BEYOND NEURAL NETWORKS

From immediate mammogram results to a richer understanding of the human genome, AI is reshaping the future of health care.

with **Greg Corrado**,
Distinguished Scientist and
Head of Health AI at Google Research

**In the quest to understand** and replicate the marvels of nature, humans have often looked to the skies. Birds, with their graceful flight, have been a source of inspiration. Yet, when it came to creating our own version of flight, we didn't replicate the bird; we built the airplane. Greg Corrado, Distinguished Scientist and Head of Health AI at Google, suggests that in the realm of technology, we often draw from the principles of nature without necessarily mimicking them.

We studied birds and insects to learn the mechanics of flight. However, when humans sought to fly, we realized that copying the exact mechanisms of birds wasn't always the most practical solution. Similarly, in the domain of AI, neural networks—inspired by theories on how biological brains process information—represent our effort to harness the principles of biology without replicating them entirely.

"The same thing happens in AI," Corrado says. While neural networks take inspiration from the brain's functioning, the tools employed in AI—electricity and silicon—are inherently different from the organic makeup of our brains. "We couldn't do that if we tried," Corrado adds.

## AI in Health Care: Focusing on Real-World Impact

One of the most promising areas where AI is making a difference is in health care. Consider this: A person you love goes in for a routine mammogram. A few days later, they get a call from the doctor's office. There was something unusual in the results; they'll have to schedule a follow-up. But scheduling being what it is these days, this takes weeks—weeks of angst and worry as worst-case scenarios run through their mind. Eventually, the study is ordered, the appointment happens, and they are deemed to be okay, no biopsy needed.

But what if AI could streamline this process? Corrado's team at Google is working on a program in collaboration with Northwestern University. This program uses AI to examine mammograms more extensively in real time, alerting medical professionals to potential issues and allowing for immediate follow-up. It doesn't replace the human touch, but enhances it, making the process more efficient and humane.

"And so it relieves this kind of anxious waiting, even when there isn't a problem," Corrado says. "And when there

“Health care is fundamentally about people caring for people. My hope is that these technologies are going to enable and expand that and make it more possible for people to feel empowered, on their own medical journey, and to feel more connected to their doctors and to their care team.”

is a problem? Well, it shortens the time to getting a more definitive diagnosis, because then you can bring someone in, and their very next appointment can be their biopsy.”

Moreover, AI is playing a pivotal role in deepening our understanding of human genetics. Alongside a consortium of researchers, Corrado’s team is at the forefront of creating a “human pangenome.” This is no ordinary genetic database. Unlike the current human reference genome, which represents data from a single individual at each DNA point, the pangenome integrates data from multiple individuals at every position. This groundbreaking resource promises to more accurately represent human genetic diversity, paving the way for enhanced diagnosis, treatments, and novel therapeutics.

### Moving at the Speed of Trust

Corrado sees programs like the one at Northwestern as demonstrative of how AI can help caregivers do their jobs more effectively, and alleviate human suffering in the process. But implementing AI in health care takes more than simply showing up and plugging something in. It takes buy-in. And buy-in takes trust.

“I believe that the way that technologies like this can be most useful is by bringing the technology to the folks who are real practitioners in the art,” Corrado says. “And that the right approach is to include doctors and health-care professionals and patient communities to help us understand what is the right way in which artificial intelligence can be used in expanding human health in the practice of medicine. These are new tools and new technical capabilities, and personally I really feel that the fundamental currency here needs to be understanding.”

This means reaching out to people who are not software engineers or AI researchers to help them understand how the systems work—at least well enough to decide how they might effectively use the technology. As a result, much of Corrado and his team’s time is spent with medical practitioners in clinical and research settings.

But as fast as AI capabilities are developing, the implementation of those capabilities will necessarily be regulated by the trust of human users. While technology development is guided by the art of finding out what’s possible, the everyday use of that technology—especially in medical settings—is guided by more practical questions: What should we do? What makes people feel safe?

While the potential of AI in health care is vast, its widespread implementation faces challenges. Introducing new technologies into the medical field requires rigorous testing, validation, and regulatory approvals to ensure patient safety. It’s about not just the technology’s readiness but also the health-care system’s ability to adopt it and adapt to it.

“It takes time and work and caution to go from the category of ‘could do’ to a space of ‘should do,’” Corrado says. There’s a quip that Corrado uses sometimes: “Health care moves at the speed of trust.” Building that trust is as important as building the technology itself. It’s what allows programs like the one at Northwestern or the pangenome to exist.

The AI systems Corrado is involved with don’t replace doctors or their expertise. Instead, they offer a tool that can quickly analyze data, spot potential issues, and provide feedback.

“Health care is fundamentally about people caring for people,” Corrado says. “My hope is that these technologies are going to enable and expand that and make it more possible for people to feel empowered on their own medical journey and more connected to their doctors and their care team.”



# CAPTURING THE MIND’S EYE

How a Barcelona-based studio is exploring the intersection of memory, art, and technology.

By Pau Garcia
*Founding Partner, Domestic Data Streamers*

**In February 2021,** we at Domestic Data Streamers, a Barcelona-based digital storytelling organization, threw ourselves into a complex journey that began with technology but took us deep into human emotion and memory. We set out to explore the domain of generative AI technologies within social transformation initiatives. Among our many initiatives, the “Synthetic Memories” project stood apart, affecting everyone involved on a profoundly emotional level.

The project consisted of a series of interviews with elderly people about past experiences and moments from their lives that were never documented in images. The goal was to use generative AI to create images that could evoke these experiences. This was not merely an exercise in computing. It became an emotional journey for both the Domestic Data Streamers team and the people who trusted us with their memories.

The original inspiration for this project was seeded in 2013 during a significant migratory crisis in Europe. While collaborating with Ojala Projects, an NGO assisting Syrian families in Athens, a touching encounter with a refugee grandmother brought to light the immeasurable value of memories and the role of images in preserving them.

“My grandkids will be refugees all their lives,” she told us. “They have lost not only their homes, but their neighborhoods, friends, and also the memories that link them to our past and culture. We have lost all our photo albums and diaries, our family history is gone.” This moment truly changed the way in which we think of the value of images, how they operate as mediators between our past and present, and link us to a deeper understanding of our roots.

To clarify, the images generated from this project blend techniques from photography, drawing, and watercolors. Faces remain obscured, ensuring the images are never mistaken for actual photographs. This is crucial because vague images have proven more effective in memory recall exercises. A clear image might highlight inconsistencies, while an unfinished image allows our imagination to complete it, often in line with our memories.

This project’s potential reaches beyond nostalgia. It has vast scientific and therapeutic implications. We are already collaborating with social workers, psychologists, and medical experts in dementia and psychoneurology to understand the positive impact that synthetic memories can have on the progression of degenerative diseases such as Alzheimer’s or senile dementia. Preliminary findings indicate this methodology could significantly enhance Reminiscence Therapy, which uses various sensory stimuli like music, objects, and, yes, images to enhance the cognitive functions of people living with degenerative diseases.

A significant majority of patients, 11 out of 16, expressed a desire to increase sessions on synthetic memory recreation after their initial experience. Yet, those in advanced dementia stages showed limited involvement, emphasizing the need to tailor the therapy to individual cognitive capacities, potentially also hinting at a new way to gauge cognitive decline. Many patients expressed interest in having their generated memories printed, offering them tangible connections to the past. We also recognized that some memories take time, spanning more than one session, stressing the importance of therapy flexibility. Interestingly, group sessions with synthetic memories, tested over 20 times, cultivated rich interactions among participants. This enabled them to share cherished memories, boosting their interpersonal bonds, showcasing the benefits of group reminiscence in fostering communication among other dementia patients.

↑ A “synthetic memory” image created through AI tools, recreating Maria’s childhood memory of seeing her imprisoned father through the bars of a balcony.

> “That was the only way I could see my father for four years, through the bars of that balcony and the bars of his prison cell window. I was six at that time.”
>
> MARIA

One encounter that stands out is our session with Maria, an 84-year-old from Barcelona. She vividly recalled her first memory of her father, seeing him from the perch of a rented balcony that faced the “La Modelo” prison where he was imprisoned.”That was the only way I could see my father for four years, through the bars of that balcony and the bars of his prison cell window. I was six at that time.” Maria’s description of the place and the historical context of that moment was used to generate an image that not only resonated with her but also moved us deeply. Upon seeing the image, Maria reported that looking at the image was akin to peering into a part of her past that she no longer had the words to articulate.

Sometimes the image generation process works from the very first test, and sometimes you need to rework them a bit; change the clothing, move particular objects, or find a specific material. But usually, after no more than 10 minutes, we find an image that the participant can recognize.



↑ Maria, an 84-year-old from Barcelona, working with researchers from the Domestic Data Streamers to create new visuals from her childhood memories.

While the Synthetic Memories project offers a promise for personal reminiscence, its potential is far-reaching. It can document the experiences of marginalized communities and preserve cultural legacies. Additionally, the data we collect offers invaluable insights into studies on memory, cognition, and the sociology of aging.

However, the intersection of memory, technology, and art also raises ethical concerns. Ensuring the privacy of participants, maintaining the authenticity of memories, and determining the psychological implications of these synthetic images are just a few issues that need careful consideration. Our next steps include forming partnerships with healthcare institutions and mental health organizations to broaden the project’s scope and ensure ethical practices.

To that end, our team is focusing on transparency and reproducibility of the research by creating an accessible knowledge base so experiments can be replicated and verified. We are selectively partnering with museums and research entities capable of adhering to legal and ethical standards like the EU’s AI Act and GDPR. This will allow us to advance the Synthetic Memories methodology rigorously while upholding humanistic values.

As we venture further into this new frontier between fiction and reality, it’s imperative that we navigate responsibly and transparently. This technology stands to redefine not just technological capabilities but also the possibility to create new emotional landscapes to explore.

← See more about the Synthetic Memories project, and others by Domestic Data Streamers, at domesticstreamers.com

# How will AI augment human creativity?



When the first photograph was captured in the 1820s and exhibited in Paris, the consensus of critics in the art world was that it marked the death of creativity. The advent of AI in the arts has proven, so far, to be both equally controversial and far more transformational. As algorithms generate works of astonishing imagination, they also expand creative possibilities for people historically excluded from conventional mediums. For artists with disabilities, AI dismantles physical barriers to expression. For neurodivergent creators, it complements unique perspectives. While debates rage around authorship, ethics, and economics, AI has expanded artistry in unprecedented ways.

Questions on the essence of creativity resurface with new urgency in this age of thinking machines. Can algorithms channel emotional truths and lived experiences? Or are they destined to remain talented mimics bound by the limitations of data? Some believe a more collaborative path lies ahead. Rather than replace human creativity, AI can become a tool to enhance it.

Technologists are discovering how creative potential can be unlocked when leveraging AI thoughtfully, and in collaboration with artists. By setting clear objectives and boundaries, algorithms can be directed to imagine new concepts that human creators may not have conceived. Humans can harness computational creativity without being constrained by it.

Challenges remain on issues of bias, access, and responsible implementation. But one thing is certain: this technology will reshape artistic frontiers.

# Art Meets Science

Art and science have often been thought of as two distinct disciplines. But artist Markos Kay doesn't see it that way. Throughout his body of work, Kay has explored scientific phenomena using a range of digital and generative tools, bringing vibrant life to the unseen world.

→ ↓ ← ↑
All works by
Markos Kay

**"My interest in art and science** started at a very early age, perhaps from a need to understand how things work on a fundamental level, and the need to express that wonder," says Markos R. Kay (neé Christodoulou), a Cyprus-born, London-based multidisciplinary artist and director with a focus on art, science, and generative art.

His work can be described as an ongoing exploration of digital abstraction through experimentation with generative methods. His experiments often explore the complexity of the invisible and mysterious worlds of molecular biology and particle physics.

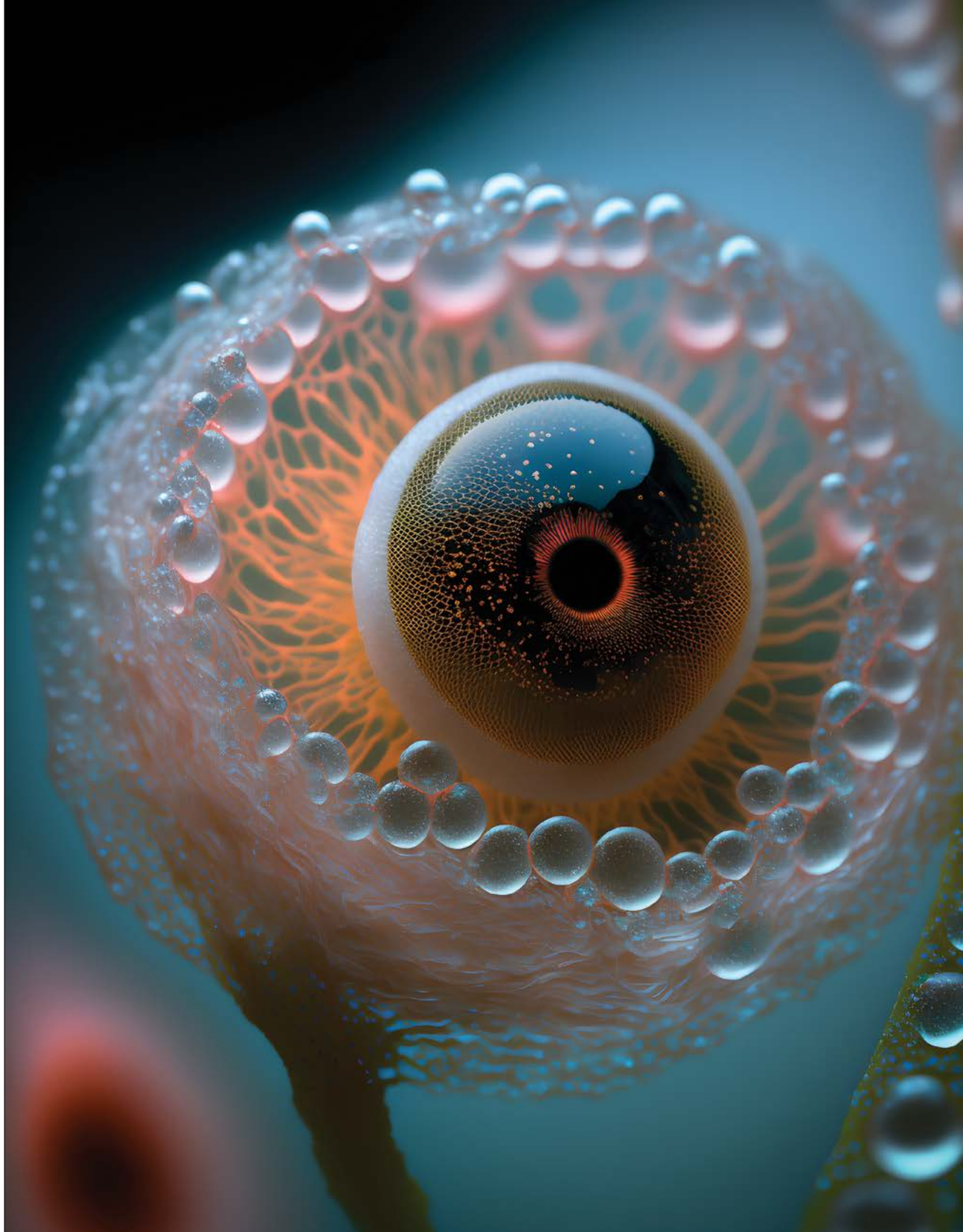
Since his practice began, Kay has had a unique curiosity and innate talent for exploring some of the most challenging and complex subjects in the scientific community. His own challenges became more pronounced when, in 2016, he became disabled due to a chronic neuro-immune disease known as ME/CFS, which by 2019 rendered him permanently housebound and largely bed-bound. But Kay continues to press ahead, bringing life to his imagination, thanks to tools like generative AI.

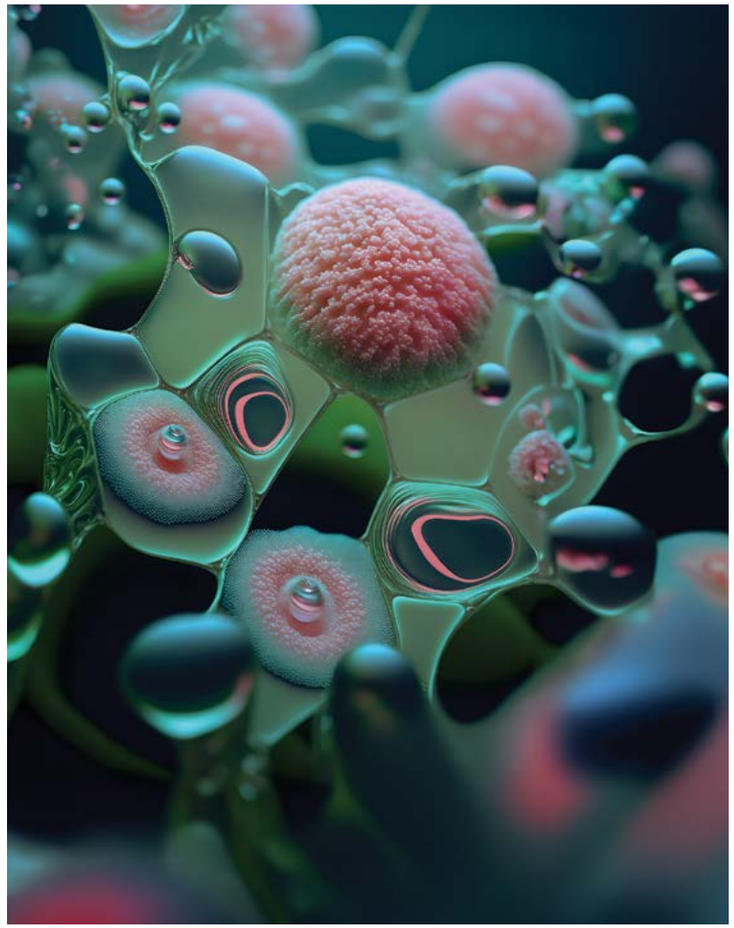
"I have been able to visualize projects that I thought would never see the light of day in a matter of days," says Kay. "As someone with a serious disability, this has been nothing short of a miracle, as it has given me the ability to create and express myself again, which I felt I had lost because of my illness."

Kay's latest major body of work, published in 2022, explores the origins of life itself. Titled "aBiogenesis," it visualizes the "lipid world" theory that posits that life originated from lipids forming membranes, which would then envelop matter and nutrients to form protocells. The biological cells we now know as the building blocks of existence can be thought of as membranes inside of membranes. Though just one of many theories about how life began, it's a widely accepted idea that helps scientists to understand how life might have emerged from the chaos of the primordial soup.

Using a range of tools, including generative AI, Kay has brought these early groupings of cells to life, rendering vibrant, microscopically detailed images and videos that conjure memories of inquisitive eyeballs or flowers in full bloom. To create the work, Kay says that "AI tools were very carefully art-directed to create the visuals; none of the images were raw outputs, but rather a result of a complex and very deliberate process." The result is a science unto itself, overseen and directed by Kay's precise vision.

We may never know the origins of life—but Kay's work will continue to present a compelling vision of our earliest days, thanks to a budding new technology.





“

People often think of science as the antithesis of art, but I see them as intimately intertwined in collective culture, informing and shaping each other throughout history.

Science is able to profoundly change the way we think, to create new connections and paradigm shifts. It is exactly those profound cognitive shifts, similar to the effect that art has on our minds, that are the source of endless inspiration.

**MARKOS KAY**

↓
See more of the aBiogenesis series in motion, and more work from Markos, on his website.

# TYPOGRAPHY RECONSIDERED

## (TYPOGRAPHY, RECONSIDERED)

So far, AI tools have struggled to create precise letterforms. Four designers weigh in on how that struggle is offering surprising insight into the evolution of human communication.

**Introduction by**
**Drew Campbell**

—

Title design by
Barney McCann

**AI doesn't know** how to create words.

Allow me to rephrase: AI certainly knows many words. In fact, it may know all the words. The technology leverages statistical probabilities to predict how and why any given word might be used, and where it should land in relation to a word that precedes or succeeds it. It's then able to stream all those words together, infinitely, without tiring.

But when it comes to bringing those words to life visually—creating the literal letterforms that construct each syllable, clause, and paragraph—so far, AI stumbles. To recreate the work that designers have been doing since the days of illuminated manuscripts and the Gutenberg Bible is no simple feat—at this stage in the evolution of AI tools, the models built on statistical probabilities struggle to recreate the simple beauty of well-designed, meticulously crafted typography.

Still, designers and artists have begun to explore what this new frontier of typographic expression may look like, tapping into an array of AI-powered tools and techniques to explore the niches of the discipline. In doing so, they're not just creating innovative and invigorating imagery. They're delving into the fundamentals of communication, expression, perception, and the relationship between human and machine.

Here, four designers share their work and perspectives on how these tools have reshaped the way they create and explore typography—and how the process has spurred larger questions about essential human practices primed for an evolution in the AI age.



1


## Barney McCann

Obsolete began as a way of exploring new letterforms and typography that could be created with AI. Using Arial as a starting point, I fed these characters into a neural network that analyzed the basic shapes of each letter, and then allowed the AI to try to find the most efficient ways to transition from one letter to the next.

The typeface is the culmination of this process, where steps normally hidden behind the pixels and processors of a computer are made visible, with endless variations as the machine looks to consistently become more efficient. The result is, I think, the closest you can get to a computer's handwriting—but it has connected with creators in other projects because of its humanness, despite being based in technology.

In general, I don't fear the impact of AI as a creative. I have been thinking a lot about the Fauvist's response to photography, and focusing on how we can use AI tools to lean into and communicate the actual human experience further. People don't resonate with purely digital outputs as much as they do with work by humans. I think my best work sits between those two worlds.

1
Variations of letter "A" from the typeface "Obsolete"

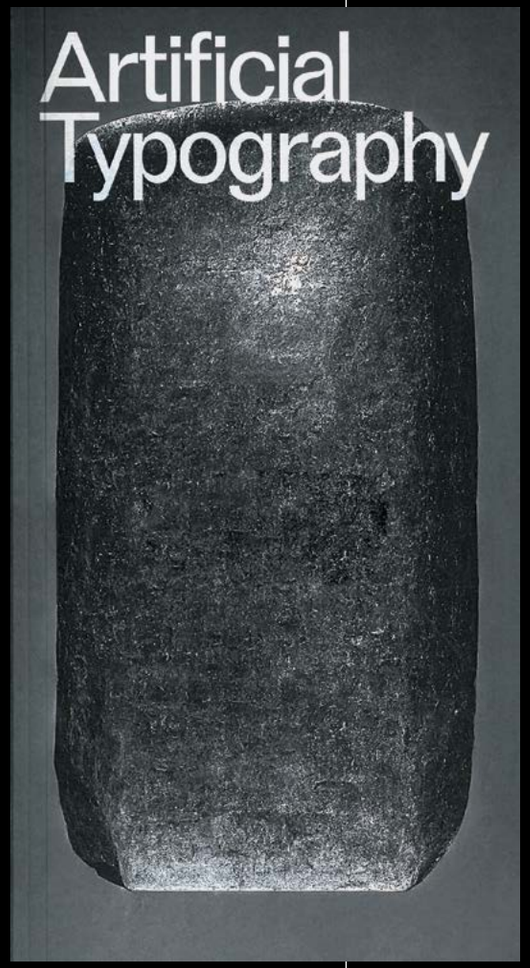
2
Obsolete in use for a tour poster. *Design credit: YACHT*

3
Obsolete in use on a record label for "(Downtown) Dancing" by YACHT

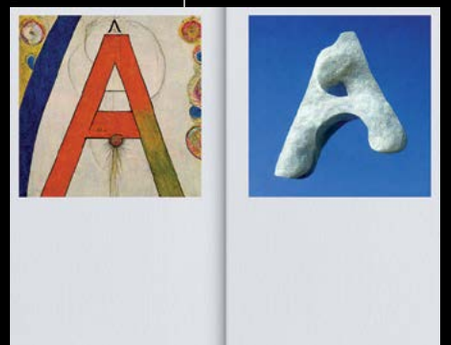
4
Frame from Obsolete in use in a music video for YACHT's track "(Downtown) Dancing"

← Watch Obsolete in motion

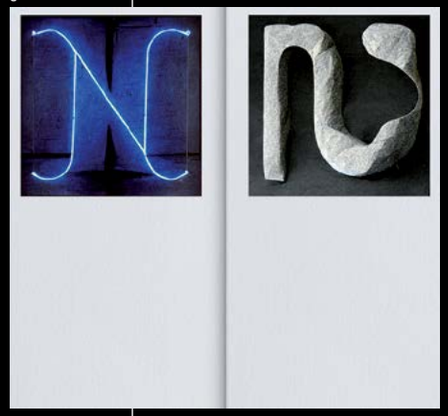
1




2

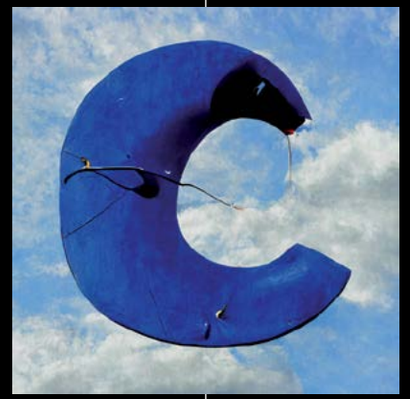


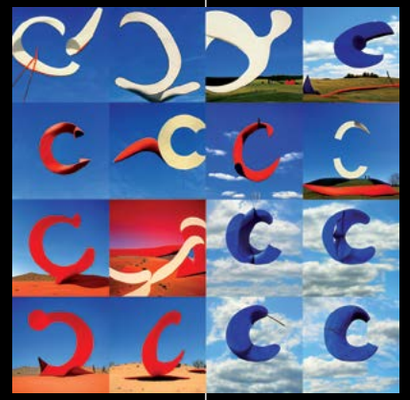
3



4

5





# Vernacular

A. A. Trabucco-Campos and Martín Azambuja

*Artificial Typography* is born out of the juxtaposition of the ephemeral and the timeless. This book explores the rapidly evolving field of AI through our long-standing Latin alphabet. The book contains 26 letters re-imagined by AI. Each letterform is interpreted twice through the lens of 52 iconic artists across various media (painting, sculpture, textile). The typographic space is especially great for this exploration, since it's a world of ideas where general conventions rule, but where there is endless opportunity for unexpected interpretations, with new recipes arising every day. At its heart, the book is driven by a curiosity to see how far AI could push visual language, and it holds many surprises, especially when the structure of letterforms are combined with materials like stone (e.g. Noguchi) or even light installations (e.g. Bruce Nauman).

Initially, the authors were also enamored by the idea of "conversation" and played with it as a title. The exchange that happens with AI machines is a form of conversation, and perhaps one of the most intellectually satisfying visual-verbal connections that has been devised between humans and machines.

Today's AI is powerful, but it is limited. The role of the designer is far from obsolete. AI necessitates a point of origin, a thought, an idea, an editor, a curator, someone who can guide it. It also necessitates an endpoint, too—a use for the output, a reason for its imagery. This might change as it gets more powerful. At the moment, it needs the human mind more than the human needs it.

1
Book cover referencing the Rosetta Stone, created in collaboration with MidJourney AI

2
Letter "A" reimagined by MidJourney AI in the style of Hilma af Klint (left) and Jean (Hans) Arp (right)

3
Letter "N" reimagined by MidJourney AI in the style of Bruce Nauman (left) and Isamu Noguchi (right)

↓
*Artificial Typography* is the first title by Vernacular, a small independent publisher with a focus on the intersection of form, typography and visual culture that stretch beyond the commonly agreed registers of "good design." It was founded in 2022 by Porto Rocha senior designer Martín Azambuja with Pentagram partner + designer Andrea Trabucco-Campos.

www.vernacular.is

4
Letter "Y" reimagined by MidJourney AI in the style of Tadanori Yokoo, followed by unused variations.

5
Letter "C" reimagined by MidJourney AI in the style of Alexander Calder, followed by unused variations.

1


2


3


4

## Khyati Trehan

Creative AI tools don't make me feel creative. The text-to-image model is efficient but takes the joy out of the making. It takes spending time on a piece, sleeping on it, making mistakes, solving them, and tons of iteration to take it to a place where it reflects me just as much as it surprises me. My workaround to settle this feeling is to deliberately make a "back and forth" between traditional tools and AI tools a part of the process, which makes for interesting workflow possibilities. When do you move from one tool to the other? What tools help you relinquish control and when is the right time to bring the work back into your hands? This ping ponging makes a tool out of a "generate" button. It's also just really fun to play a game of exquisite corpse with GenAI instead of treating it as a means to a completed end.

What still fascinates me about generative AI is that its interface is essentially plain text. In my opinion, writers make better prompt engineers than creatives do. Working with AI tools in turn has also made me better at articulating my work and made me more intentional with the words I place on a page.

All of my type-specific outputs currently slot under the category of expressive typography rather than practical applications. In the same way new UI typesetting paradigms emerged from the rise of app-populated smartphones, I see LLM formatting rules and best practices becoming the next place for typography to serve.

1
Frames from *Look/Here*, a video work created with sketch and generative fill processes in Adobe Photoshop

2
A frame from *Ask How*, a video work created with sketch and generative fill processes in Adobe Photoshop

↓
See more of Khyati's ongoing experimentation with AI tools on her Instagram account.

@khyatitrehan

3
A frame from *Option, Shift, Command, Delete*, a video work created with sketch and generative fill processes in Adobe Photoshop

4
Frames from *Grow*, a series of video works created with sketch and generative fill processes in Adobe Photoshop, exploring typographic motion simulated on a variety of media

1


2

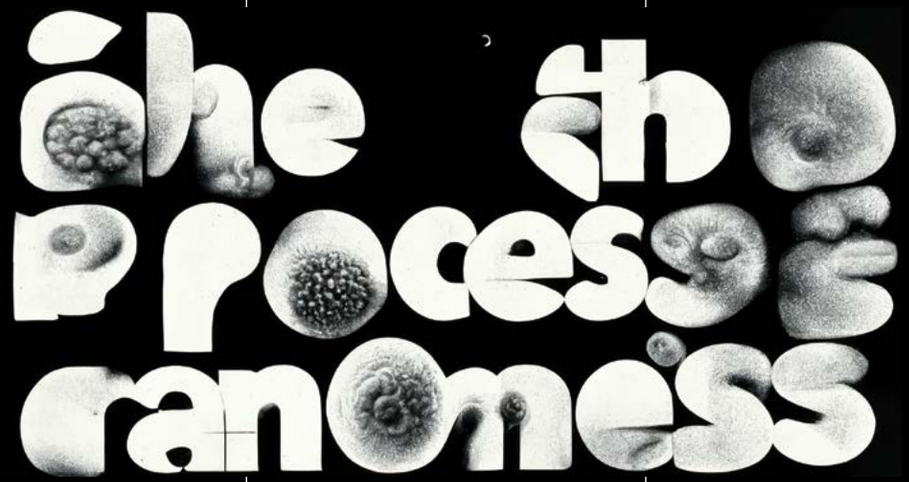
3

4


5


6


7




## Gianpaolo Tucci

Since AI image generators became a mass-market product, an emphasis on image development has seen huge progress in the algorithms that serve as the foundation for AI outputs. A year ago, the outputs were clearly a blurred and distorted version of reality—a visual dreamer. Today, realism is on the rise in AI's mimicry of daily scenes, especially in the context of photography shoots, products, and sci-fi images.

Since I began using these technologies, my focus has been on letter shapes—typography. Typography itself is a technology for communication, but so far it has not been at the center of this AI evolution. It's ripe for experimentation.

The whole project finds its conception in the intersection between three pillars: Time as context, nature as a metaphor, and evolution as a constant. My journey started in trying to reach "perfect" letter shapes. By perfection, I mean the value of readability and their association to what we can read.

In this process, I've discovered instead the beauty of what I've called "aesthetic imperfections"—a new opportunity for letters and their combination into words that are kissing the rigor of the function of typography while also pointing toward an enhanced visual meaning. Even if not fully defined, the letters are cognitively readable through their composition, and by an enriched visual expression.

AI offers several opportunities to explore the appearance of words and visual messages to enhance communication. AI could standardize, but could also open up creative opportunities. Our duty is to shape the control and the coexistence.

AI today is seen as an anti-human technology. But I would say AI is not a replacement—it could be a tool or an amplifier, one that may help us to deliver a more human and inclusive future, if it's well regulated and integrated.

→
See more from Gianpaolo and his Aesthetic Imperfections, including his new book, out in November 2023.

@aesthetic_imperfections

1
*Fear*, created using MidJourney v2

2
*Letter A*, created using a MidJourney blend of 2 images (v4), these reference images were created with v2 and v3 –s 750

3
*Yes*, created using Stable Diffusion + ControlNet –control 1.3 + input image for the illusion created with MidJourney v4 and 2 reference images created with v2 and v3

4
*Life*, created using MidJourney v2 with several iterations based on variants

5
*All I Need Is Love*, created using MidJourney v3 with several iterations and variants

6
*Peace*, created using Stable Diffusion + ControlNet –control 1.3

7
*The Process for Randomness*, created using MidJourney v3 –s 750

# The Artificial of Intelligence

**In the realm of fine art,** the introduction of artificial intelligence prompts a compelling question: How might technology enhance, rather than replace, the artist's touch? A select group of artists are pioneering this frontier, not to delegate their creativity to machines, but to harness AI as a tool—much like a brush or chisel—to challenge artistic boundaries and more fully realize their visions. This exploration presents the work of artists Casey Reas, Jan St. Werner, Pierre Buttin, Linda Dounia Rebeiz, Yuri Suzuki, Anna Ridler, and Refik Anadol, each of whom integrates AI into their creative process, offering a fresh perspective on the age-old act of creation.

**FEATURING**

**Anna Ridler**
**Casey Reas**
**Jan St. Werner**
**Linda Dounia**
**Pierre Buttin**
**Refik Anadol**
**Yuri Suzuki**

↔

**Casey Reas & Jan St. Werner**

Installation view of *Untitled 1 (No. Nothing.)*, 2020, 8 min, loop

Photo: Emile Askey

*Image courtesy of Casey Reas, Jan St. Werner, and bitforms gallery*

## Casey Reas & Jan St. Werner

*Compressed Cinema* is a series of five audiovisual works, a collaboration between Casey Reas and Jan St. Werner. Reas created the imagery in the tradition of experimental films that use existing films as raw materials. The *Compressed Cinema* suite is an inversion of Ken Jacob's 1969 film *Tom, Tom, the Piper's Son*, which expanded the short 1905 film of the same name from 8 to 115 minutes through meticulous re-photography, repetition, and editing. In contrast, each *Compressed Cinema* video distills a feature-length film into a work of less than 10 minutes.

Developed over three years, this collection emerged from experiments with generative adversarial networks (GANs) to create cinematic media. Each piece marries Reas's visuals with Werner's musical compositions. Werner uses granular synthesis, a sound technique whereby tiny fragments of a sample (grains) are arranged and modulated freely. Together they create a blend that respects traditional film elements while introducing a fresh cinematic expression.

1

2

3

4

1
Installation view of *Untitled 2 (Kiss me.)*, 2020
4 min 19 sec, loop
Photo: Emile Askey

2
Still from *Untitled 3 (I withdraw.)*, 2020
10 min, loop

3
Still from *Untitled 4 (Two dead!)*, 2020
5 min 53 sec, loop

4
Still from *Untitled 5 (Not now. No, no.)*, 2020
4 min 7 sec, loop

*All images courtesy of Casey Reas, Jan St. Werner, and bitforms gallery*

→
Explore more work by Casey Reas

## Pierre Buttin

In a unique melding of autobiography and technology, Pierre Buttin crafted a 700-page account of his life, delving into everyday details, both events and emotions. This expansive narrative, totaling 365,672 words, was then used to train an AI algorithm. Utilizing Max Woolf's adaptation of Andrej Karpathy's char-rnn (character recurrent neural network), the AI reshaped the chronicle of Buttin's life into a more poetic rendition.

From the AI's output, Buttin developed resonant sentences, transforming them into evocative poems. These poetic expressions subsequently inspired a series of paintings, each capturing the essence of the verses. The entire endeavor, from writing the autobiography to training the AI, took place between 2018 and 2021, before easy public access to AI tools. Below is a glimpse of the AI-generated content, which Buttin has deftly reimagined into poetic and visual art forms.

→
Explore more work by Pierre Buttin

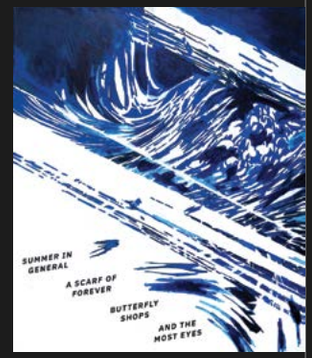
1


2


3

4


5




6


1
*The rain in my inbox*, 2021
Acrylic on canvas, 60 × 50 cm

2
*Quantum nude in the short-term rental*, 2021
Acrylic on canvas, 60 × 50 cm

3
*Wednesday, no surge pricing*, 2021
Acrylic on canvas, 60 × 50 cm

4
*The hope that this found you well*, 2021
Acrylic on canvas, 60 × 50 cm

5
*10x kiss*, 2021
Acrylic on canvas, 60 × 50 cm

6
*Cats and dogs in the automated warehouse*, 2021
Acrylic on canvas, 60 × 50 cm

# Linda Dounia

Linda Dounia Rebeiz, an artist and designer from Dakar, Senegal, critically examines the impact of technocapitalism and how it reinforces societal inequities. Through her art, she channels her memories, presenting them not just as personal recollections but as testimonies to alternative ways of life and understanding.

*Once Upon a Garden* is a poignant digital collaboration between Linda and AI. This work paints a bleak picture of a potential future shaped by global warming—a reality where humanity is left with only simulated images of plants and flowers, having lost the real ones. Through AI-enhanced depictions of more than 100 endangered and extinct indigenous flower species from West Africa's Sahel region, the installation evokes a sense of nostalgia for what's gone, aiming to inspire a commitment to preserving what remains.

*AI in Bloom* offers an abstract journey through the shifting climate of the Sahel. The process begins with a GAN producing images based on 2,000 abstract acrylic pieces handcrafted by Linda. These AI-generated pieces are then meticulously arranged in a grid, organized by color and structure.

1
*The Garden Under The Sun*
(Once Upon A Garden), 2022

2
*Amaryllis*
(Once Upon A Garden), 2022

3
Detail view of *Morning Glory*
(AI in Bloom), 2022

4
Detail view of *Cosmos*
(AI in Bloom), 2022

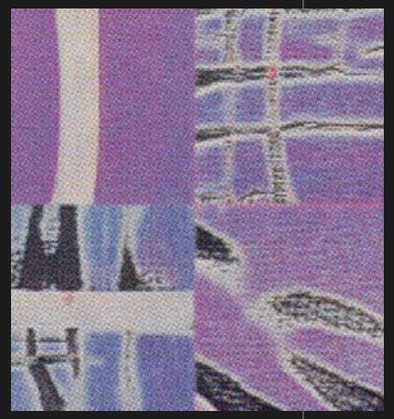
5
*This flower doesn't exist*
(AI in Bloom), 2023

1

2

4

3





5

→
Explore more work
by Linda Dounia

## Yuri Suzuki

Yuri Suzuki is a sound artist, designer, and electronic musician who delves into the intricacies of sound. His work navigates the relationship between individuals and their surroundings, probing how music and sound morph to craft unique personal narratives.

His piece, *The Welcome Chorus*, is an amalgamation of sound, sculpture, and AI. Created for the Margate NOW festival in 2019 and commissioned by Turner Contemporary, this interactive installation features 12 horns, each symbolizing a distinct district of Kent, England.

These horns don't just stand silent; they sing. Lyrics, generated in real-time by a specialized AI software tailored for the site, flow from them. The words are shaped by contributions from Kent residents, who shared their experiences to enrich the AI's lyrical database. In a nod to history, the sculpture's design hints at the etymology of *Kent*, believed to stem from *kanto*, signifying a horn or hook.

1-4
Installation views of *The Welcome Chorus*, 2019

Photo 1: Samuel Diggins
Photos 2-4: Kate Ejimiwe

→
Explore more work by Yuri Suzuki

1

2

3

4



## Anna Ridler

Anna Ridler is an artist and researcher who uses her art to explore knowledge systems and the creation of technologies in order to decipher our world. She harbors a keen interest in the nuances of the natural realm. Central to her approach is the use of information collections, especially datasets, to weave unique narratives.

Her works, *Mosaic Virus (2018)* and *Mosaic Virus (2019)*, intertwine themes of capitalism, value, and historical collapses. The 2018 piece showcases a dynamic grid of blossoming tulips, while its 2019 counterpart features a three-screen installation, each highlighting a single tulip. The evolution of these tulips is steered by Bitcoin prices, reflecting market volatility. This design choice draws a parallel between the 17th-century Tulipmania—a period when tulip-bulb prices skyrocketed, at one point equating to the cost of an Amsterdam townhouse, only to plummet to the value of an onion. Often cited as one of the earliest speculative bubbles, Tulipmania's trajectory mirrors the unpredictable nature of cryptocurrencies. For Ridler, the link between these two phenomena transcends mere economic fluctuations; it burrows deeper into the essence of their shared volatility.

1
Installation shot of *Myriad (Tulips)*, C-type digital prints with handwritten annotations, magnetic paint, magnets, 2018

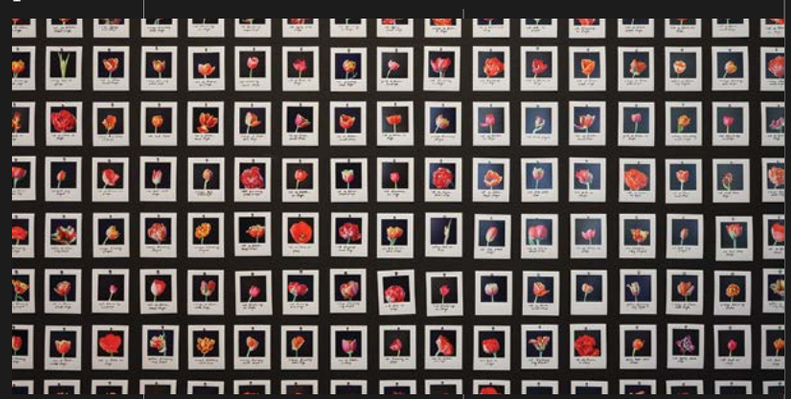
2
Still from *The Black Tulip*, custom smart contract, AI generated video, 2023.
*Courtesy of Anna Ridler and Galerie Nagel Draxler.*

3
*Bloemenveiling*, 2019, website, smart contracts, NFTs, GAN generated video and bots

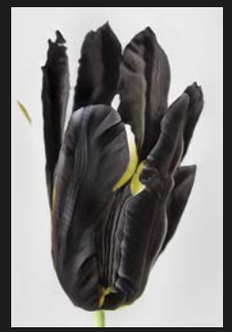
4
*Mosaic Virus (2019)*, GAN generated video work, 2019

→
Explore more work by Anna Ridler

1

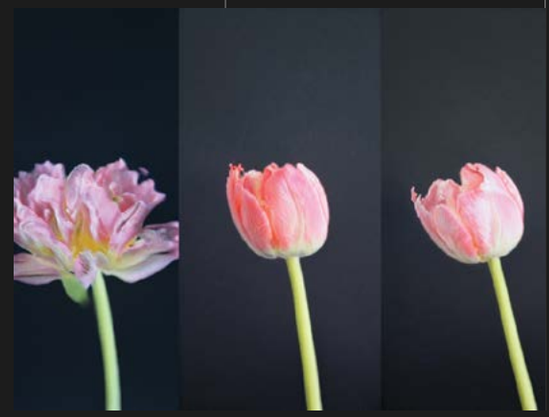


2



3

4



1

# Refik Anadol

Refik Anadol, born in 1985 in Istanbul, is an internationally renowned media artist, director, and pioneer in the aesthetics of data and machine intelligence. Anadol's site-specific data paintings and sculptures, live audiovisual performances, and immersive installations take many forms, but all encourage us to rethink our engagement with the physical world, as well as our collective experiences, collective experiences, public art, decentralized networks, and the creative potential of AI.

For *Unsupervised* at the Museum of Modern Art in New York City, Anadol used AI to interpret and transform more than 200 years of art at MoMA to answer this question: What would a machine dream about after seeing the MoMA collection? As the model "walks" through its conception of this vast range of works, it reimagines the history of modern art and dreams about what might have been—and what might be to come. While AI is often used to classify, process, and generate realistic representations of the world, *Unsupervised* explores fantasy, hallucination, and irrationality, creating an alternate understanding of art-making itself. The artwork was displayed on a large-scale media wall in MoMA's ground-floor Gund Lobby between November 19, 2022, and October 29, 2023, and was recently added to the museum's permanent collection.

Anadol was also the first artist to use the fully programmable LED exterior of the Las Vegas Sphere, a new performing arts venue, by featuring dynamic visualizations of data to create abstract images of space and nature. The two-chapter series creates a collective, meditative, multisensory experience that takes audiences on a journey of light, movement, and color with vivid pigments, shapes, and patterns.

1
*Machine Hallucinations: Sphere*, 2023

2–4
Installation views of *Unsupervised* at The Museum of Modern Art, New York, 2023

*Documentation by Refik Anadol*

→
Explore more work by Refik Anadol

2

3

4



Q&A

# Creativity and the Algorithm

Working at the crossroads of technology and imagination, this is how two artists at Google envision the future of AI-driven art.

←
Photography by
Jovelle Tamayo

**Blaise Agüera y Arcas and Mira Lane** have unique perspectives on the creative possibilities of AI—and that's in large part because they're not just technologists; they're also artists. In 2016, Agüera y Arcas, who works as a vice president and fellow at Google Research, founded a program called Artists + Machine Intelligence, which supports artists who produce works that incorporate machine learning. He's also a writer and most recently authored the 2021 novella *Ubi Sunt*. Lane is a fine artist who works across mediums—including video, music, and ceramics—and she's a senior director of Technology and Society at Google. Both of them agree that AI is set to revolutionize the way humans engage in creative endeavors, and that such a revolution will raise a host of complicated questions. We sat down with Lane and Agüera y Arcas to discuss.

As an artist, you're more likely to say, "Why isn't it another way?"

Question **As these systems become more integrated into artistic processes, what are the ethical and philosophical—or capitalistic—considerations we need to take into account? How should we think about ownership and originality, or credit and disclosures?**

Mira Lane Well, some of these questions have always been around in artistry. How do you attribute songs in the right way, for example, if you've sampled something? Artists don't always like to share their secrets. We're notorious for wanting to keep those secrets to ourselves, because some of it is … There are techniques we've developed; sometimes we take shortcuts. For a lot of us, for me as well, AI is just part of our tool kit. Do I feel the need to disclose that an AI tool was used? No more than I feel the need to disclose any other process by which something was rendered. There are some very worthy questions we have to solve around the economics of this. But as people start incorporating AI, it gets metabolized, and you realize where some of the edges are. You saw this with generative images. There was a lot of uproar around how it was going to replace illustrators. But we found the best images are often created by people who have artistic backgrounds because they know the techniques and the language, the vocabulary. They have the ability to envision the output.

Question **Are there certain formats or mediums that demand attribution?**

Blaise Agüera y Arcas Yeah, maybe. But here we get into things that have less to do with art in some abstract sense than with the particular cultural traditions we're living with, with respect to credit assignment and the cultural capital of that—and the *capital* capital of that. The economics. It's a very long-tail world of rewards right now. And it's a weird moment, especially with respect to visual art. Plutocrats have warehouses full of old master paintings. That's just a form of capital at this point. I think looking at the AI moment is hard without considering that whole. Mira alluded to the fact that these debates aren't new. Beastie Boys got into all these sampling wars around copyright in the 1990s. Laurence Sterne was pilloried for plagiarism in his novel *Tristram Shandy* in the 18th century. [James] Joyce famously obscured his tracks. He was a big obscurantist for exactly the same reasons an artist who takes a shortcut is going to not cop to that. There's a mystery to maintain. Or consider Chihuly and his giant workshop. How much of the Chihuly glasswork is actually Chihuly, his own hands? It's the Chihuly corporation.

Question **Right. Or Richard Serra.**

Agüera y Arcas Exactly. Many instances. One of the most compelling, strongest, most vigorous defenses against AI that I've heard was by this illustrator who does fantasy art in pencil, and it's incredibly detailed and takes a long time. He was very frustrated about AI art, because in a few seconds you can generate things that superficially look a lot like his work—and I get that. At the same time, the same thing happened at the dawn of photography. There was this uproar from fine artists.

Question **Technology itself can become a canvas for creativity, in terms of the design of it, the way we use it, and so on. Can you speak to the importance of that?**

Lane These conversations can't happen with just technologists. When you bring artists in, we start to play together. We workshop. That creative dialogue is so important. When I think about shaping technology, I think of it from that conversational standpoint. It's not just technologists in a room. It's bringing people in who might push the boundaries. One example is this amazing work that was done with Lupe Fiasco, the rapper: How would a rapper use something like these large language models? It's about taking the craft of writing and exploding that out into new ways of interaction and exploration using language models. How do you build chains of related items? How do you create alliteration stemming from a topic? These types of tools—we wouldn't have thought of these by ourselves.

Agüera y Arcas When photography was invented, it wasn't just about making fake paintings, by the way. It changed painting and created the new genre of photography. And, maybe most important, it created movies. So there were new forms of art that suddenly became possible that were inconceivable before. I think AI is poised to do the same things—to create an explosion in types of media, including things that verge on the creation of entire universes. I mean, the mind kind of boggles.

Question **But is there an extent to which AI and its creative possibilities will always be limited by its efficacy as a tool? Will it ever become a real creator versus a terrific mimic?**



Agüera y Arcas There are some mechanical limitations at the moment that make it difficult to think about these things as agents or collaborators in the full sense, but I think we're closer than I imagined we might be. To the other point you raised, as to whether they can ever produce something that isn't just a remix or a rehash of what they've been exposed to during training—they can. It's a misconception that they're just dismembering snippets of what they've encountered before. That's not true for two reasons. One is generalization, which gives you the adjacent possible. So if AI has learned the algorithm for multiplication, for example, it doesn't need to find the auto-complete of every three-digit multiplication problem on the web. Moreover, AI also has this really interesting property called in-context learning. What that means is that it's limited only by what can be described with language at all—so far as language can describe something, AI can describe it too. That still leaves a lot of philosophical questions about agency unsolved, but I think, at a mechanical level, it's kind of all there.

The whole point of imagination and creativity is to be able to imagine what you have not experienced.

Question **Do you disagree that the production of great art requires real human experience?**

Agüera y Arcas I feel like there's something unfortunate about the extreme privileging of experience that's happened the past few years. All of us can imagine what it would be like to feel this way or that way. I mean, that's empathy. That's the basic human stuff. The whole point of imagination and creativity is to be able to imagine what you have not experienced.

Question **That's a fair point. And I suppose if you're able to make someone feel something as a result of that creation, that emotion is no less real.**

Lane And it shouldn't prevent us from enjoying it, whether it's created through machines or humans. At the same time, I don't see the agency yet in these machines to make me say, "Hey, it's actually being creative." Real art takes work. The journey of creation is part of the artwork. And language alone is a very challenging interface. We've spoken to many artists recently, and we've given them these models where language is the interface, and all of them want additional ways of interacting, because we creative people often don't think in language, especially visual artists. I think we'll look back at the chat interface as a very primitive one.

Question **This may be a weird question, but do you trust AI's tastes?**

Agüera y Arcas Well, I mean, I do think it has taste. I think it's *bad* taste. So there are two things that go into making an AI, roughly speaking. There's the pretraining and then the fine-tuning and reinforcement learning, and design choices are being made. The fine-tuning stage is where things get really interesting. Because that's where you say, "Okay, what is your personality going to be?" And that's a real thing. Like, MidJourney's art looks different from DALL-E's. That's the taste of the engineers who say thumbs-up to this or that. Are they qualified to be the world's tastemakers? Obviously not. We need a much greater diversity. It's fine for them to have their taste expressed in that model, but the idea of a few of them defining all of our tastes is horrible. I don't care what the particulars of that taste are. Diversity is critical.

Question **I'm going to phrase this question two ways. First, how do you look at technology differently than your colleagues might because you're artists? And second, how do you look at art differently because you're technologists?**

Lane Oh, that's interesting. I look at this technology from a very experimental point of view. It's very easy to converge on the obvious, and so you tend to see similar solutions from a lot of the industry. As an artist, you're more likely to say, "Why isn't it another way?" You end up blowing out the possibilities a lot more and challenging the base assumptions. And as an artist who is also a technologist, I just don't have a lot of fear in this space, because I'm so comfortable experimenting with these technologies. I want to be on the bleeding edge and experimenting. I'm also fortunate because I don't make a living off my art. I know I'm in a privileged place to critique and explore.

Agüera y Arcas Those were going to be my points as well.

Lane I want to say one more thing. It's really important for technology companies to have people who insist on bringing artists into the conversation, because you don't see that elsewhere in a lot of big companies. You need to have people who are in positions where they can insist we have these types of programs—and can create these types of programs—to bring artists through and have them interface in meaningful ways. That's a critical part of the dialogue.

# Will AI make the world more or less equal?



AI surfaces a pivotal question: Will its benefits uplift humanity, or accrue to the privileged few? History shows that the last technology revolution led to greater inequality, even though everyone had access to the same tools. Many AI tools will similarly be free to access. Will the same pattern ensue? For example, AI has the potential to transform education, assist farmers, and streamline supply chains. How do we ensure equal access? Governing advancements in AI demands ensuring wide participation in shaping it. Industry must make space for diverse voices and skills in creation. Policy must guarantee inclusion and accountability in application. And education must prepare all communities to take part.

In time, AI could prove the great equalizer. But first we must lay the social, economic, and political foundations required for an equitable ascent. The task ahead is a human one: bringing vision, will, and solidarity to the project of uplifting all.

# AI's Freshman Year

Students and educators across the United States share their experiences with AI in the classroom.

By Charley Locke

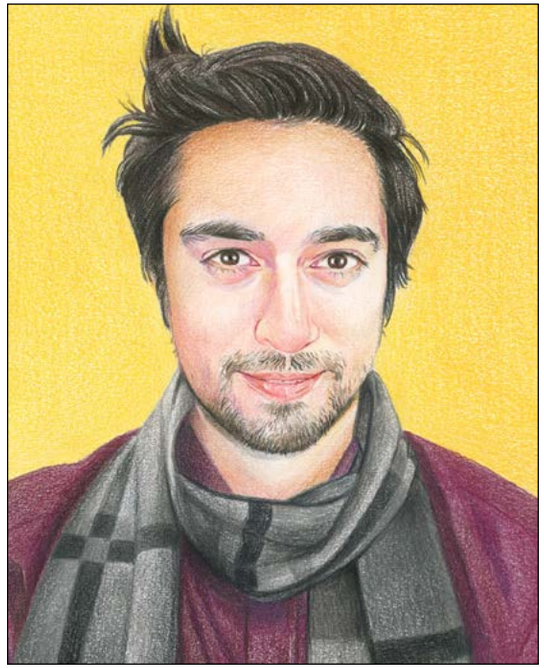
→
Portraits by Uli Knörzer

**When 10 of Brooke Nasser's students turned in the exact same incorrect answer** on a worksheet in November 2022, Nasser, who teaches high-school English in Kapolei, Hawai'i, was flummoxed. But she quickly put it together: Her students were experimenting with generative AI. She was excited, even if they weren't using it the right way.

Google searches for "AI education" have more than tripled since fall 2022, when generative AI tools began making headlines. The issue is playing out in classrooms first, as educators have been reckoning with a big question: What's the role of AI in education? Some teachers, like Nasser, have integrated lessons about how to use it into their classes; some districts have banned it entirely. Students have strong opinions about how to use the technology too. Is it an alluring temptation to take a shortcut, a tool for stepping into their future, or both?

AI has the potential to reshape and personalize education, advancing how we teach and learn—but it could also broaden existing inequalities. We wanted to hear from the experts, so we turned to high-school students, teachers, and an administrator from vastly different districts across the United States for their perspectives.





I see AI as a calculator for words: We need to teach the kids how to use the tools.

**Patrick Guillen**
High School English Teacher
Orland High School
ORLAND, CALIFORNIA

**I teach in a rural district.** A lot of our kids go on to work in agriculture, mostly in the almond orchards. I see AI as one of the ways that they can access broader opportunities. It can be a socioeconomic equalizer: If you learn how to use these tools, you can build the creative thought and critical thinking to be able to create things yourself.

Last year, when I started to see what people were doing with AI, I tried using it for my own work, as I created worksheets and lesson plans. I experienced how simple it was to use those tools and I realized that our kids needed to know how to use this, or else they'd be at a disadvantage. I told our district administrators that we needed to get out in front of this.

I found that many educators who hadn't used the tool were scared of it, because they didn't really understand what it did. They worried that it was the end of essays or the end of us teaching kids. The fearful reaction is normal, since you have to upend how you're thinking about teaching in some ways. But teachers are used to adjusting to new things: We just made it through COVID-19! I encouraged other teachers to mess around with it themselves so that they could see how it works.

During some of our professional development days this year, I asked if I could lead a training on how to use different AI tools in ways that are beneficial for us. I gave a rundown on how to use generative AI to educators at our high school and our middle school. I also showed them how AI voice modifiers and image creators work. We're going to see AI-generated work from some of our students, so what does it look like? What are the stylistic flairs that it shows? What tools can you use to identify it?

I see AI as a calculator for words: We need to teach the kids how to use the tools. This year, I started my English classes off with handwritten in-class essays so that I could see a baseline of what their writing skills are. I have regular conversations with them about how they need to build a base of being able to think critically, reason, and write. AI tools can help us do that, but I try to stress to my students that it's about the process, not the product. How can we use AI to improve the drafts we've written? How do we become more effective and efficient? We need to use this not as a crutch, but as a way to make ourselves better.

I think of it like a balancing board: You have to be careful, and you have to trust students not to use it in a way that'll hurt their own education.

**Siobhan Faughnan**
High School Senior
Leesburg High School
LEESBURG, FLORIDA

**Since I was in eighth grade,** I've always taken at least one class online. My freshman year was 100 percent online because of the pandemic. But otherwise, it's been because I want to get ahead or take classes that they don't offer at my school, which is low-income. I learn a lot of my coursework through online classes or by teaching myself.

That's a situation where AI can be helpful: for students who want to go above and beyond the course material. I've used AI that way, like in a math class when the teacher basically just gave us notes and we had to learn on our own.

If you go into the experience doing as much work yourself as you can and use it to learn the rest, then AI is a tool. But I think of it like a balancing board: You have to be careful, and you have to trust students not to use it in a way that'll hurt their own education.

I've been wary about using AI very much because it's hard to know when you're crossing the line. I'm in the yearbook class, and last spring, we all had to come up with options for the theme for this year's yearbook. Afterward, I asked ChatGPT to come up with title phrases that had puns or metaphors for the theme I wanted. In my presentation, I ended up using one of them: "Glowing Up Together." I had the teacher's permission to use AI for the assignment, so I knew it was okay, but I was still thinking, *Am I cheating right now?* I'd like it if there were more clear-cut rules about it.



It seems like a big risk for schools to take on teaching AI, but it's a bigger risk not to do it.

**Sallie Holloway**
Director of Artificial Intelligence and Computer Science,
Gwinnett County Public Schools
GWINNETT COUNTY, GEORGIA

**We have a history** of being an innovative school district through STEM and career development programs, so when we decided to open a new school in 2019, we decided to focus it around AI. At the same time, the district was introducing a "computer science for all" initiative for K–12, and we thought that an ability to use AI is something every kid needs to have. We started one high school devoted to AI as a pilot and quickly expanded to refocus some K–8 schools around it, too.

We frame our mission as developing literacy among our students so that they can be ethical, informed users of AI in the real world. Everything is taught through the lens of an "AI ready" framework: programming, ethics, data science, creative problem-solving, math reasoning, and AI applications.

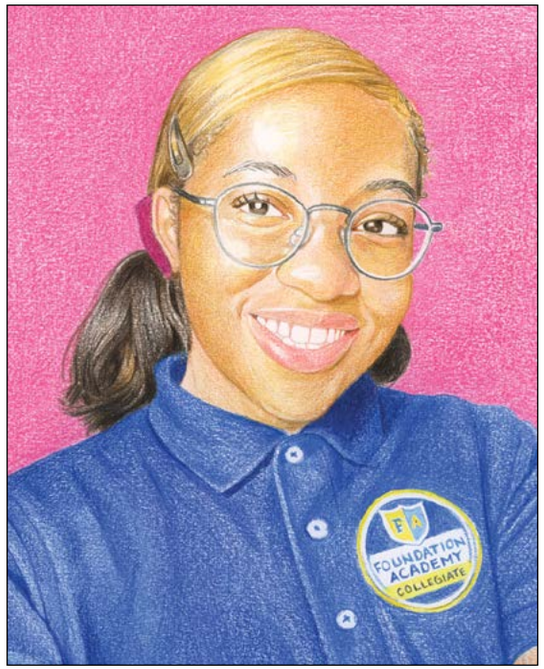
We're still teaching our same content but upgrading the relevancy for our students. For example, a high-school history teacher talked about the 2014 snowpocalypse in Atlanta and asked students to use AI tools to analyze traffic and weather data and discuss what would have happened if we had been able to get better data faster. How would that have affected the response in the city? Students can see the relevancy in what they're doing, so they're able to really think about their futures.

Of course, part of our work means pulling back the veil on these tools for students: What holes can you poke in what the AI came up with? Why is it valuable to do this work yourself? How can you use AI as a brainstorm partner? When you give clear guidance, it helps students learn how to be responsible users and takes away some of the desire to use the tool to cheat.

We often use the analogy of "swim, snorkel, scuba." All of our students have to be able to swim: They need to know what AI is because it's part of our society at this point. We hope that all of our students are snorkelers: proficient, responsible, ethical users of AI. There's also a group of students who are scuba divers, who may want to pursue a career in developing AI. For those students, we have an additional three-course pathway that's a deep dive into AI as a profession.

The implementation of AI readiness in our district has been very grassroots. There was no kit that we could pick up and say, "Hey, let's go do this." It entailed a lot of risk taking, which can be intimidating for educators because they're already doing so many things. But once they had this experience with kids, they could see why it was valuable.

I often think about something that one of my colleagues said: "It seems like a big risk for schools to take on teaching AI, but it's a bigger risk not to do it, because then you're likely not preparing kids for our future." It's a perspective shift, but we really need to be rethinking what we're teaching and what skills kids need to have in the long term.




AI isn't going to bring down education, but it's also not going to solve it.

**Jedidah Worrell**
High School Senior
Foundation Academy Collegiate
TRENTON, NEW JERSEY

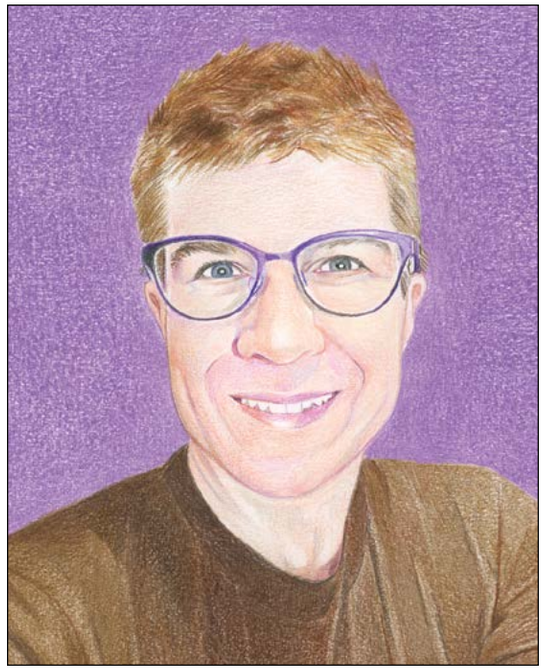
**I make art,** and that's how I first learned about AI. I follow a lot of art pages on Instagram, and last year, I started to see this shift from laughing at AI-created art to seeing it as a threat to artists' jobs and livelihoods. But I think that reaction is overblown, and I think that's true in education, too.

My teacher was the one who introduced us to AI tools, surprisingly. He said that he used it to generate answers for multiple-choice questions on our tests but that we shouldn't use it. But you know that if you tell someone not to do something, they'll try it. My history class last year required a lot of reading, so my friend and I sometimes plugged the articles into AI and had it rig up a summary of what was in the text. For some assignments, we really had to read the text because we needed to find supporting evidence, but we used AI for smaller stuff that was lower stakes.

People are talking about AI because it's a new tool, but the biggest issue in education isn't AI. It's how we're being taught in a post-pandemic world. So much education was lost during the pandemic, and so many students aren't at the grade level where they should be. There's a lot of scaremongering that AI will take over education, but that's not what's going on. AI isn't going to bring down education, but it's also not going to solve it.

The big promise of AI is that it'll make learning more efficient, but we shouldn't be striving for efficiency. We should be thinking about how to make students curious and driven instead. A lot of my classmates don't feel motivated or encouraged to learn and explore in their education lately—some of that is because of the pandemic, and some of that is because schools are severely underfunded and understaffed. Together, this means that it's harder for students to find the drive to pursue an education out of a genuine want to explore. These tools help students do the work more quickly, but it doesn't make us care about it. We should be figuring out how to make it more fun to learn.





For AI to be helpful in education, the feedback needs to move in both directions: Students and teachers need to advise the development of these tools.

**Elizabeth Tanner**
High School English Teacher
Westwood High School
MESA, ARIZONA

**I started using Quill,** an interactive reading and writing tool, with my students about five or six years ago. [Ed. note: Quill is a Google.org grantee.] At the beginning of each school year, each student does a diagnostic test, and based on that, Quill recommends different skills that they should work on. Then, in my classroom, we have Grammar Wednesdays every week, where students work on their assigned Quill activities. It's an individual process for them, which allows me to walk the room and check on their progress.

The feedback that the students get from Quill is critical because it helps them see how to improve and helps me know what to address, both for an individual student and for the class. If multiple students are struggling with conjunctions, for example, then I know I need to do a larger review.

But for AI to be helpful in education, the feedback needs to move in both directions: Students and teachers need to advise the development of these tools. For students, giving feedback is essential because it enables them to feel like they're part of a solution. And engineers need to remember that teachers have expertise, and we can help improve the tools.

That's why I'm on the teacher advisory committee for Quill. I tell the company about how I use the tool and how to improve it. By participating in Quill's development, I'm making it better for students and educators. I'm also being trained on how to use it most effectively.

AI can accentuate teaching, but it can't replace it. We still need the hands-on expertise that teachers bring.

Teachers have said that if you use AI, you're going to get in trouble. I think it should depend on how the student uses it.

**Bekzod Mamasoliyev**
High School Senior
New Utrecht High School
BROOKLYN, NEW YORK

**I'm in the Academy of Business and Technology** at my high school, and at the end of last year, one of our teachers showed us how to use AI for an assignment. We were making sticker prototypes for a client, so we used AI to research the client and get ideas of what to create. Then we used Adobe Illustrator to create the stickers and printed them out. We had to include what we asked the AI and what answers it gave us, to show that we were using AI as a tool instead of using it to do the work for us. The tools were confusing at first, but as I got used to it, I got more ideas of what to create.

Other teachers have said that if you use AI, you're going to get in trouble. I think it should depend on how students use it: If they use AI to write the assignment rather than to comprehend the gist of it, then they should be punished, because that hurts the student's potential to learn and to eventually apply their skills in the outside world. If they use it as a study tool to understand what the assignment was about or to get a suggestion from the AI about how to improve their own work, then they should be allowed to use it.

I've used AI that way, to help me learn a concept. In AP Statistics class, I was confused about the differences between two types of random samples. I read the textbook, but it wasn't very clear, so I asked AI some questions to help me understand. That's a way that AI can really help students. For example, when a student has questions about a homework assignment, they can address their own confusions rather than waiting until the next day to be able to ask the teacher for help.



AI doesn't change the problem, it just presents a new tool for students to exploit the issue.

**Brooke Nasser**
High School English Teacher
Kapolei High School
KAPOLEI, HAWAI'I

**I had heard whispers of AI** rippling through education, but I wasn't thinking about it as something that students were using in my classroom. Then, in late November of last year, when I was teaching at a different high school, I got all of these identical worksheets back for an assignment about *Beloved* by Toni Morrison. I saw that 10 students had the same wrong answer. It misattributed an event to the wrong character, which I knew all of those students couldn't have gotten wrong, because we had read the book out loud together in class. At first I thought, *Did nine people copy one person's wrong answer?* But then I realized they used generative AI. At the time, I wasn't discouraged—I found it so interesting that they were using it.

After that, I designed a class lesson around AI tools. I taught it again this fall, and I already had to change it significantly from what I taught in the spring. That's how dramatically this is all changing.

For the lesson, I give the students three different essay drafts about coastal erosion. I ask them to pick out three pieces of evidence to justify whether each essay was written by a student, a teacher, or AI. The big reveal is that all three of the essays were written by AI, in response to slightly different prompts.

That exercise serves as a jumping-off point for a discussion of how we can ethically use AI to help us in the classroom. For the most part, the students understand that it's helpful but not ethical to just take someone else's idea. But what I really want them to realize is that they each have a unique perspective and AI doesn't. At the end of the day, I want to see my students' singular perspectives, and cheating takes that away, no matter what source they use. AI doesn't change the problem, it just presents a new tool for students to exploit the issue.

Yes, AI can streamline the process for students by eliminating the mundane, rote work and pushing students to more sophisticated writing, the same way that calculators do for math. But students have to master the basics first. You don't give calculators to second graders; they need to learn addition and subtraction first. AI might be really helpful in education at higher levels, but for students at the high-school level and lower, it can prevent them from mastering the basics.

It points back to the main issue that I see in education: How do we motivate students to care about deeper learning and creating quality work that is uniquely their own, at all levels? How do you develop the discipline to work hard at something you don't love? I don't have the answers. Although, if I can figure that out, then I'll become the best teacher that ever lived.

*These interviews have been edited and condensed for clarity.*

---

Charley Locke is a freelance journalist who often covers youth and elders. She regularly writes for publications including *The New York Times Magazine*, *Bloomberg Businessweek*, and Vox.

# SEEDING THE FUTURE

Faced with a changing climate, depleted soils, and a growing global population, farmers are turning to AI-driven tools and robotics to help them boost efficiency, reduce waste, and sustainably grow more food.

By Jen Swetzoff

→
Illustrations by
Jon Han

**In Beltrami, Minnesota,** brothers Andrew and James Johnstad run their family farm where they raise wheat, corn, sugar beets, and soybeans. Having grown up on these fields, they know every inch of the land, cultivated by their family for four generations. But the soil tells a more complex story than the human eye can read on the surface. And that story is one that the Johnstads, along with many other farmers, want to understand better. As they work to solve the onslaught of challenges related to climate change and adverse weather events—like abnormal amounts of rainfall one year, followed by abnormal drought the next—all while battling wind erosion, water runoff, and herbicide-resistant weeds, soil health is more important than ever.

"There's a generational fear that comes with farming," says James Johnstad. "There have been many family farms that have ceased to exist because of certain problems that people couldn't overcome at the time. Now we know we need to make the soil better. Our grandfather always instilled in us that we should leave the land better than we found it."

Many of the biggest problems facing farmers today are rooted in the soil. Just one tablespoon of healthy, living soil contains billions of microscopic organisms—all of which provide essential nutrients and carbon dioxide for plant growth as well as natural drainage and structure. But that healthy soil has been degraded over the past century, as traditional agricultural practices such as plowing and planting with tractors, tilling, and using fertilizers, and pesticides stripped away healthy microorganisms. The result is dry and desolate land, which has taken a direct toll on human health and the climate. In fact, according to the Food and Agriculture Organization of the United Nations, if we continue on our current course, "by 2050, 90 percent of all soils are set to be degraded. Without change, degrading soils will put our ecosystems, our climate, and food security in jeopardy." Put more directly in a new documentary called *Common Ground*: "If the soil dies, we die."

Traditionally, farmers in the United States relied on national soil surveys to understand the composition and health of their land. These surveys provided basic insights into soil types, moisture levels, and nutrient content. In many other parts of the world, however, farmers have not had access to such detailed soil information. As another way of collecting data, farmers around the globe often send soil samples to scientific labs, but that process can be slow, imprecise, and labor intensive. Either way, those methods of data

collection often led farmers to uniformly treat their entire field based on the weakest part. That could mean focusing on one row crop, rather than many varieties of plants, and overusing fertilizers and chemicals, which hurts soil health.

Enter artificial intelligence. While farmers have long had access to diverse data sources—from GPS and multispectral imagery to soil sensors and equipment telematics—the challenge lies in integrating data and generating actionable insights. For instance, a USDA survey revealed that while 61 percent of corn growers used a yield monitor in 2010, only 34 percent turned that data into a yield map. But with AI, these diverse datasets are converging, allowing farmers to understand their land's conditions in real time with unparalleled accuracy and efficiency.

Satellite data, combined with new autonomous robots equipped with cameras, sensors, and analytics tools, is already enabling farmers to delve deep into their soil's health. These AI-driven tools can identify and analyze the myriad microorganisms in the soil, assess its carbon content, and even detect early signs of degradation. By providing a granular view of the soil, AI allows farmers to tailor their farming practices at more granular levels, ensuring optimal health and yields.

The implications of this knowledge are also expected to help farmers practice regenerative agriculture at scale. Initially developed and practiced by indigenous communities, regenerative agriculture is a holistic approach that emphasizes soil health. It involves techniques such as cover cropping, rotational grazing, no-till farming, and composting. But to truly harness its potential, farmers need detailed, real-time data about their soil at scale—something traditional methods of research and analysis couldn't provide, but newer methods can.

"Climate change is rapidly reshaping the whole world, including its food and farming systems," says Anika Molesworth, Ph.D., an Australian farmer, scientist, award-winning author, and the founder of Climate Wise Agriculture. "With rainfall and temperature patterns changing, with changes in pest and disease distribution, and with extreme weather events rocking the foundations of farming businesses, we need to be better prepared and able to adapt. I think technology, including AI, is one important tool to do that."

## GETTING INTO THE WEEDS

Farmers today face a multilayered dilemma. They're keenly aware of the environmental and long-term benefits of sustainable farming. Sustainable crops, free from excessive chemicals, lead to healthier soil, better yields, and a safer environment. Past practices, however, have come back to haunt them. The overuse of pesticides and herbicides has given rise to herbicide-resistant "superweeds." These hardy weeds are immune to traditional chemical treatments, making them a significant problem for farmers. While the allure of sustainable farming beckons, the immediate and pressing issue of these resilient weeds cannot be ignored.

In fact, according to research from Montana State University, without a better option to control weeds, Montana could lose as much as 68 percent of its average annual yield, costing growers $43 million in lost revenue. And other states growing sugar beets—including Idaho, Michigan, Minnesota, Nebraska, North Dakota, Oregon, and Wyoming—could lose approximately 22.4 million tons of sugar beet yield, valued at approximately $1.25 billion. Farmers need solutions that address both the present and the future.

Kenny Lee, co-founder of agricultural robotics start-up Aigen, recognized this conundrum. Before diving into tech development, he and his co-founder, Rich Wurden, a former Tesla engineer, spent ample time on fields, talking to farmers and even growing sugar beets themselves. They quickly grasped that while farmers were interested in the long-term benefits of sustainable farming, they also were desperate for immediate, scalable solutions to their weed problems.

**61%**

Percentage of corn growers using a yield monitor in 2010

**34%**

Percentage turning that data into a yield map—a statistic that improves dramatically with the integration of AI.

↑
Kenny Lee (right) and Rich Wurder, co-founders of Aigen

Photo: Peter Bohler

"When you talk about sustainability, robots that can do things like targeted weeding or even targeted pesticide applications will make a real difference," says Elizabeth A. Bihn, Ph.D., executive director of the Institute of Food Safety at Cornell AgriTech. "First, using less chemicals leads to healthier soil. And, second, robots can reduce the use of tractors, which run on fossil fuels, so that cuts those emissions. It's a win-win."

With those insights in hand, Lee and Wurden developed an autonomous, AI-driven, network-connected weeding robot called Element. Unlike more rudimentary agriculture robots, Element is independent from the tractor ecosystem and powered by the sun, eliminating the need for refueling or battery swaps, which saves time and money for farmers. Under the hood, it's powered by advanced AI algorithms that allow it to differentiate between crops and weeds with incredible precision. This level of sophistication, driven by vast amounts of data and machine learning, simply wasn't possible five years ago. Element uses computer vision to identify weeds, then mechanically tears them out using steel tools, reducing the need for chemicals. This approach not only tackles the immediate problem of herbicide-resistant weeds but also offers the long-term advantage of data-driven insights for sustainable farming at scale.

Wadhwani AI, a nonprofit and Google.org grantee, is also exploring how AI can support farmers—specifically those who grow cotton. Wadhwani AI has developed an app called CottonAce that uses AI to analyze photos of pests trapped on farms. The app counts the number of damaging pests, like bollworms, and determines whether applying pesticide is recommended based on reaching an economic threshold, helping farmers gauge whether the potential crop damage by pests justifies the cost of pesticide application. This provides smallholder cotton farmers with localized warnings and advice to better target pesticide use. While still in its early stages, the app works offline and has been translated into nine languages to increase accessibility. Initiatives like these demonstrate the potential for AI tools to be inclusive and empower farmers with data-driven insights, even with basic technology.

→
Anika Molesworth, PhD, an Australian farmer, scientist, award-winning author, and the founder of Climate Wise Agriculture

Photo: Christopher Morris

"When you think about the way that food is grown, and the number of acres that we need to transform the soil for commodity crops, it's a massive challenge," says Lee. "But when you get to the bottom of it, it's not really about technology. It's about people. We want to use AI as a tool to help the people actually working on the land."

Over time, Lee believes, Aigen's chemical-free weeding will open the door to more sustainable agriculture practices. From the start, the robots reduce fossil fuels and soil compression from tractors while lessening chemical dependence. In addition, the weeding robots collect high-resolution farm data that, after being anonymized and aggregated, can be used to train AI to inform the optimization of growing conditions for the long term, which is what farmers want.

## FARMING FOR THE FUTURE

"Farmers are probably some of the most conservation-minded people on the planet," says Tony Latcham, who raises corn, soybeans, hay, and cattle on his family farm in Iowa, and who expects his sons to take over in several years. "We love the land. We work on it everyday. It's something we're really proud of, so we want to take care of it for future generations. If AI helps us do that in more efficient and cost-effective ways, we're all for it."

While the short-term benefits of AI-driven solutions like weed management are evident, the long-term implications for sustainability are even more profound. Aigen's technology, for instance, does more than just tackle the immediate "superweed" problem. By reducing the need for chemicals and tractors, the soil retains its health and vitality. The continuous data gathering by these robots provides insights into soil health, moisture levels, and other critical factors that can guide sustainable farming practices.

Other start-ups are emerging in this space as well. LandScan, a platform that will soon be used to assess the soil on almond farms in California, employs soil probes with advanced sensor technology to digitally characterize soil variability at a very granular level. Traditional soil sampling involves taking a chunk of soil, sending it to a lab, and waiting for results. But LandScan's sensors can instantly measure soil properties such as hardness, friction, and spectroscopy right on the spot. Instead of analyzing a disturbed sample, these sensors provide data on the real soil structure that plant roots experience. This could be particularly valuable in parts of the world that never had soil surveys.

"Our technology allows us to digitize the process of soil surveys and characterize variable land at different locations," says Jeff Dlott, Ph.D., COO of LandScan and a member of the Environmental Farming Act Science Advisory Panel at the California Department of Food and Agriculture. "Over time, with machine learning and AI, this knowledge will help us understand which crops can grow most successfully in which soil types, and help farmers avoid wasting resources by managing to the weakest part of a field."

By understanding the soil's needs at a granular level, farmers can tailor their interventions, ensuring that the land remains fertile and productive for generations to come. This data-driven approach can lead to better water management, reduced chemical usage, and improved crop rotations, all of which contribute to farms' long-term sustainability.

"We're talking about feeding the world," says Latcham. "You know, we can't grow more farm ground. Every year, we're losing more acres to urban sprawl. The earth is changing. So we really need to get the soil right and we need the technology to keep getting better."

## GROWING MORE FOOD, MORE SUSTAINABLY

Reimagining a more sustainable food production system means enabling agricultural transformation at scale. That's what motivates the leaders at Mineral, a recent spin-off from X, the moonshot factory and now an Alphabet company, to build new AI solutions tailor-made for agriculture, with global ambitions.

Working closely with the Alliance of Bioversity International and the International Center for Tropical Agriculture, as well as other partners, Mineral is leveraging breakthroughs in AI and computer vision to develop solutions that capture and interpret complex plant-by-plant information. By applying powerful, AI-driven perception technology to its rovers, and now to other edge devices such as smart cameras and mobile phones, Mineral can gather and analyze critical crop production data about soil health, yield predictions, ripeness, disease risk, and weed presence. That data can be used to inform better farming practices and to breed climate-resilient crop varieties. Ultimately Mineral's technology will include reasoning and actuating across crops and geographies.

"The variable conditions farmers manage on a daily basis can be infinite, from soil health to farm management practices to weather patterns," says Erica Bliss, chief commercial officer of Mineral. "The good thing for growers is that AI is really well suited to the complexities of agriculture. Because it makes sense of rich datasets like imagery and video as well as text at an accelerated pace, AI puts efforts like identifying climate-resilient crop varieties or enhancing precision agriculture equipment, among other things, on the fast-track toward more sustainable outcomes. And time is of the essence to address the drivers of and impacts from climate change."

Mineral's AI-driven tools are already providing agribusinesses and farmers around the world—from Brazil to sub-Saharan Africa—with a better understanding of the variables involved in outdoor farming. This knowledge supports growers in improving plant-level management, driving greater precision to reduce the use of water and chemical fertilizers. At the same time, Mineral is developing AI-powered tools that can help farmers prepare for the future by enabling them to better understand what variety of crops will survive amid droughts and floods. For example, using AI in seed breeding, where there might be more than 30,000 varieties of a bean, can dramatically lower costs and complexity.

"At Mineral, we're working to enable a global sustainable food system by solving some of agriculture's greatest challenges at scale," says CEO Elliott Grant. "There's no time to waste to help the food production system adapt to a changing climate, to find more resilient crop varieties, and to improve soil health and restore biodiversity."

As of July 2023, the Mineral team reported that it has captured more than 800 million plant images across five continents and diverse growing conditions, modeled more than 120 different plant characteristics, and analyzed 14 crop types. These kinds of multimodal reasoning capabilities are the beginning of what can eventually contribute to scaling sustainable transformations across farms and geographics.

"Growers are demanding more from their technology," Bliss says. "They want and need help making simpler

← Erica Bliss, chief commercial officer at Mineral

Photo: Clara Mokri



"AI is really wonderful at helping us make sense of many rich data sets. That more targeted understanding then helps us give clearer direction to companies that support growers as well as to growers themselves."

decisions more quickly and managing their farms in a more efficient and profitable way. But nature is still quite complicated, so agriculture requires the most advanced technological solutions, tailor-made just for its unique challenges and conditions, that growers can reasonably use for both row and specialty crops. It's really the only way to transform the food production system into a more sustainable way to feed the planet and protect the environment."

## THE FARMER'S OPPORTUNITY

The next era for agriculture will be powered by more AI-driven solutions, not more horsepower. While technology in the agriculture industry has led to important productivity gains throughout history, there's a limit to how much land exists and how big machines can get. Over the past 70 years, U.S. crop yields have tripled, tractor horsepower has multiplied by four times, and the typical weight of a fully loaded combine has increased nearly tenfold, according to Mineral. But as machines become bigger, they also become more complex, expensive, and diesel-intensive, leading to massive contiguous farms, uniformity, and soil compaction. At the same time, floods, droughts, wildfires, and other extreme weather events create an onslaught of challenges for farmers.

AI still has a ways to go, but farmers and engineers are optimistic about its potential. With more global-scale experimentation and collaboration, AI-powered tools can hopefully support people in building resilience to a changing climate, and break down traditional barriers by unlocking the value of data.

"I think we're entering the golden age of satellite data," says Meha Jain, associate professor at the University of Michigan School for Environment and Sustainability. "Using satellite data with AI, we can map characteristics including crop type, yield, water use, and the adoption of technologies and practices, including regenerative agricultural practices. We can use these data to understand the adoption of sustainable practices at scale and also what their impacts may be on yield and environment outcomes. This work can help farmers identify effective technologies at the landscape level, and help policymakers and extension agents identify low-adoption regions that can be targeted with further interventions."

Minimizing chemical and water use and reducing waste while encouraging strong production can help farmers both feed the growing global population and lessen the impact that the food production system has on the planet. Ultimately, the future of sustainable agriculture will depend on farmers, scientists, innovators, and policymakers working together. By embracing new technologies like AI while staying grounded in the wisdom of traditional farming practices, we can hope to meet growing food demands in a way that regenerates our soils and ecosystems for generations to come.

Q&A

# The Jobs Equation

A conversion with economist Erik Brynjolfsson about how AI is likely to impact the workforce—and what can be done about it.

By Nicholas Thompson

→
Portrait by
Denise Nestor

**One of the most surprising** facts to me about the past 20 years has been that gains in technology have led to gaps in income inequality—some economists have estimated that it explains half or more of the increasing gap in wages. I expected the opposite: that technology would be an equalizing force.

Now, as we stand at the cusp of a new era, I wonder whether the same pattern will repeat itself. If everyone has access to many of the same AI tools, will that make us more equal? Or will the opposite happen, with AI potentially consolidating power and wealth in the hands of people who use it best? I spoke with Erik Brynjolfsson, a professor and senior fellow at the Stanford Institute for Human-Centered AI and the director of the Stanford Digital Economy Lab, about what he thinks is coming next and what can be done to steer technology toward fostering more, and better, opportunities. Our conversation has been edited for length and clarity.

Nicholas Thompson **Technological advances sometimes make income inequality grow. Sometimes they make it shrink. What influences those outcomes?**

Erik Brynjolfsson This is a really important question. For a long time, economists had this simple model of a neutral technical change. But then we started noticing in the '70s,'80s, and '90s that inequality was getting a lot worse. And most economists, including me, attributed that mainly to technology. Globalization was also a factor. But technology, especially IT and computerization, seemed to have a significant effect on worsening income inequality through several different mechanisms.

I wrote about those mechanisms in the book I coauthored with Andrew McAfee, *The Second Machine Age*. One was what we call "skill-biased technical change" that complemented more-skilled workers and substituted for less-skilled workers. And you see these tremendous gaps grow between, say, people with a high-school education or less versus a college graduate. And that just kept worsening for a few decades. There was also a bit of a shift between capital and labor. So the labor shares fell. And last but not least, superstars seem to benefit especially. So the top 0.1 percent in a lot of areas got these amazing incomes as they were able to use digital technologies to scale to millions or billions of people in a way that would have been impossible previously.

Thompson **All right. So you have those three factors: economic benefits, capital and labor, and superstars. There must have been some effects of technology that pushed in the other direction, though, right? Everybody has access to the same browsers and platforms.**

Brynjolfsson Absolutely. So those three tended to be very powerful and dominate the conversation. But there are 1,001 tools that help different groups. And one of the categories that I think has been underestimated has been the access to free goods like search engines, maps, YouTube, Wikipedia ...

Thompson **But you don't think they matter enough to affect the overall trajectory?**

Brynjolfsson I think they've affected it a bit. But if you look at the statistics, the income data—which may have some biases—or things like mortality statistics, I think the dominant story is still in the direction of growing income inequality.

Thompson **So we have a new generation, a new revolution in technology, perhaps as big as the transistor, perhaps as big as the mobile phone, perhaps as big as the internet. Who knows? Will AI likely reverse these trends of growing income inequality or augment them?**

Brynjolfsson We don't know, to be fair. But there is some evidence that in some situations it's reversing the trends. And I'm hopeful that that will be a more general thing. But the technology is so new—if you'd asked me this question six months ago, I might have given you a different answer.

But I can tell you what we've seen so far, which is a decrease in income inequality. I did a very in-depth study with Lindsey Raymond and Danielle Li in which a large language model was introduced to a call center to help operators, not to replace them. And what we found was that the less-skilled workers benefited the most. They had about a 35 percent increase in productivity. The most-skilled workers benefited almost zero. The LLM was capturing a lot of the tacit knowledge from the more-experienced workers about how to solve problems for customers, how to speak to make them happier and like the interaction more, and it was transferring that to the less-skilled workers in a very efficient way so that within a few months, these less-skilled workers and the new workers were going up the learning curve very rapidly. As a result, AI tended to close the gap between the most-experienced and less-experienced workers.

Thompson **So let's consider the most-skilled economics professors and researchers versus the least-skilled economics professors and researchers. The most skilled will benefit less than the least skilled?**

Brynjolfsson Quite possibly. I mean, we haven't done that specific study, but there's a mechanism that is plausible. And we've seen it happen in other situations. Because again, there's a lot of tacit knowledge that we have that until now has been almost impossible to convey to others. But machine learning is very different. Some people call it software 2.0. Previously, you had to write down step-by-step what you did. And not all of us could explain exactly how to ride a bike or tie a shoe or how to recognize a good turn of phrase. But now the machine looks at the data and it learns. So it's opened up trillions of dollars' worth of knowledge that it's making accessible to other people.

This leads us to an interesting point where we're going to have to rethink how we compensate people and how we reward people and their job security. In the case of that call center, those operators were mainly rewarded based on how happy they made their own customers. But in the story I just told, the most-skilled workers were helping not only their own customers but also other agents' customers by basically training the system. And if you're a smart company, you'd want to update your compensation so you're rewarding those kinds of workers. You want more of those kinds of workers,



Machine learning has opened up trillions of dollars worth of knowledge that it's making accessible to other people.

and you want them to be doing more to push the frontier. That's how your whole company benefits. The narrow measure of how each operator's customers are doing would give a misleading signal of the total contribution.

Thompson **If AI makes everybody more efficient, maybe you don't need as many people. Then a whole bunch of people lose their jobs, and we haven't done anything good for income inequality.**

Brynjolfsson I don't think that's quite the right story. And let me give you three reasons. First, there's the sweep of hundreds of years of history to where we are right now, close to record-low unemployment and record-high labor share of the population employed. And through wave after wave of technology, people have told the story that it's making people more efficient so that we'll need fewer, and it's never happened. So why has it never happened?

There are two main reasons. One is that if you look at particular tasks that the technology helps with, it's awesome. But it rarely if ever does the full set of tasks that are in an occupation. For example, at one point AI researchers said that we should stop training radiologists because machines could read images better. I think that's basically right about reading the images, but there are roughly 30 distinct tasks that a radiologist does. Only one of them is reading images. It's a super important one, but there are other things that radiologists do, and the technology has not helped with most of those. Ultimately, it's led to more restructuring. I think it will lead to more restructuring where you use the tool to help with part of what you're doing, but not all.

The second reason is an economics 101 lesson. So when the price of a good falls, the quantity purchased increases. You have downward-sloping demand curves. If the demand curve is very steep, then as the price falls, the demand increases only a little bit, and you end up spending less. And that's most everyone's intuition. But in many cases, the demand curve is very flat. And when the price falls a little bit, the quantity purchased grows by even more. So, for instance, when jet engines made airline pilots more productive in the '50s and '60s, we didn't hire fewer pilots. Why? Because with that lower price, we all fly a lot more. So now we hire more pilots. And we wouldn't have done that when it was expensive.

So let's go back to radiologists: Let's say my shoulder is a little bit sore right now. It's probably too expensive for me to get an MRI, but if the cost went down, me and millions of people in the United States and India and Africa

I would love to steer technologists, entrepreneurs, managers, and policymakers all towards thinking, "How can we create more complements and fewer substitutes?"

would love to get more access. So it's quite possible that lower prices lead to a greater demand.

Thompson **Back to the call center. Theoretically, the service from the call center will be so much better because the operators will have been trained, the response times will be quicker, and whatever fee the employer pays will be lower. So maybe the number of calls customers make and the number of problems operators solve will be higher. And the number of people employed will increase as well.**

Brynjolfsson Exactly. That's totally possible. I don't know how steep the demand curve is for a call center, but I think there are a lot of problems where I don't bother calling. But if it were effortless and I was sure I was going to get the right answer and not be stuck on hold for 45 minutes, we all might access a call center a lot more.

Thompson **So, then, going back to the three causes of increased inequality in the past technological revolutions, one of them was the ratio of investments in capital to the investment in labor. Won't that likely shift evermore toward capital?**

Brynjolfsson Probably. And that's why I think the first thing I said was "I don't know for sure." I see a mechanism where we get less inequality—that's the one I just described. And happily, we've observed that in some real-world cases. This is not a hypothetical—this is real-world data that I and others have been gathering. But I can easily see a story in which capital continues to gain and there's less demand for labor. I can also see a story in which superstars continue to have even more influence. And those mechanisms may end up being quite important.

One thing I can say is I'm pretty confident we'll see a lot more disruption in any of these stories—that the set of people who are affected will change, maybe by an order of magnitude, quite substantially. Even if the total employment stays similar, and we have a lot of employment, I'm pretty confident it'll be different types of skills and tasks and even people that are demanded. And that's going to lead to a lot of disruption.

What we need to do is put in place not just a safety net and training mechanisms, but one that is very nimble and flexible and can adjust on the fly—so that when I and others do our next study, and we see that the trend is a little bit updated from what I said six months earlier, that we have tools that can adjust to it. And frankly, our existing training, job matching, and safety nets are not nearly nimble enough. So that's a priority: to be ready for this disruption. It's not mass unemployment I'm predicting, but mass disruption.

Thompson **Now, what if I were a regulator or a government official who only cared about income inequality. All I want to do is make the Gini coefficient in the United States, or in whatever country I live in, better. What would I do with AI?**

Brynjolfsson What not to do is to try to freeze in place the existing jobs. And that's the first instinct of people, especially in Europe, but also in the United States. And that is the worst strategy. No country has ever preserved incomes or had a successful society by freezing in place all the old jobs. So you need to lean in and realize that the change is coming. And, if anything, you need to make it easier for people to adapt to the new jobs. And that means you need to put tools in place for training and for job matching—you can use AI to help identify the skills that people have that will be in demand or that are in demand in other areas, and what new skills they would need.

You can also use AI to do that training in a mass personalized way, far more efficiently. There's a long body of literature that says that when people are trained in classrooms with dozens of other people, it's not nearly as efficient as when they get one-on-one personalized training. Of course, personalized tutoring is way too expensive for most people, but with an LLM, you can do that. You can get something that's very customized to exactly a person's needs. And that is the future I see ahead of us—that we'll be able to train people a lot more quickly.

Going back to the call center example, one of the striking things was that there were some unexpected outages. And what we found, a little bit surprisingly, was that when the call center operators lost access to the LLM due to a temporary system outage, the less-experienced people continued to perform better. Not quite as well as they did when they had the LLM, but better than the group that had not had access to it. So they were going up the learning curve faster. They were internalizing some of those tips and tricks that the LLM had been coaching them on.

Thompson **You wrote a great paper titled "The Turing Trap" last year. The hypothesis, if I can briefly summarize, is that AI is more likely to displace people if it's trained to replicate human skills and human intelligence. Will this fact make this inequality problem worse?**

Brynjolfsson It will if they continue that way. The big message of "The Turing Trap" was that we have choices. We can use AI to imitate humans. And Alan Turing, a great researcher, had this iconic idea. If we can make a machine that's so similar to a human that we can't tell the two apart, we will have achieved artificial intelligence.

I think we're suddenly, perhaps belatedly, realizing it was the wrong goal all along. Because if you imitate a human, in economic terms, you make the machine a better substitute for the human. And having a substitute tends to drive down wages and value. That's not what we want. We want to drive up wages. Luckily, it turns out that you can have big increases in productivity without having a machine substitute. You can have a machine complement humans, meaning that they become more valuable in the presence of it. Like my left shoe is more valuable within the presence of its right shoe.

I would love to steer technologists, entrepreneurs, managers, and policymakers all toward thinking, How can we create more complements and fewer substitutes? If we do that, we're more likely to get shared prosperity—not just a bigger pie, but a more evenly distributed pie, because everybody will be needed and contributing. Conversely, if we go down the path of using AI to substitute for or to imitate humans, I think the tendency will be to concentrate wealth and power in a small number of people or organizations that have control of the capital. And ordinary people, or even people with lots of expertise, will become less valued because the machine will do their same job.

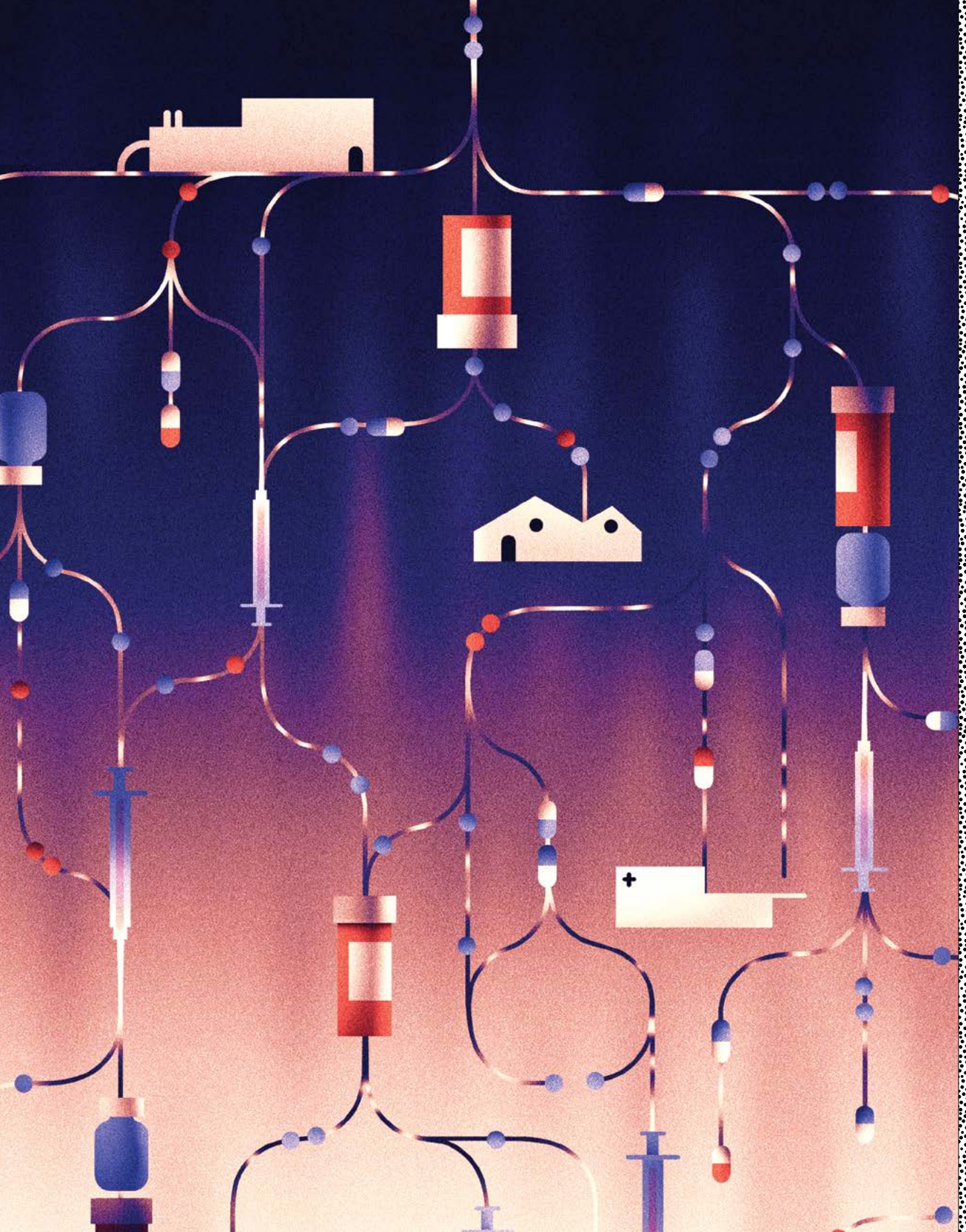
Doug Engelbart at Stanford in 1962 wrote an iconic paper, a little bit after Alan Turing's paper, about intelligence augmentation. And his vision was exactly what I'm saying: Let's look for ways of making these machines a tool. I think that the goal a researcher should have is not making the machine as powerful as possible per se, but making the human plus machine together as powerful. And that's not the same question. If you want the radiologist using a tool to come up with better answers, you probably want that tool to be able to explain why it's giving recommendations. Even if it's only 87 percent accurate instead of 89 percent accurate, that may be valuable, because when that tool says, "Oh, cut off the patient's left leg," I think the radiologist is going to want to know, "Okay, explain why you want us to do this." If it's 87 percent accurate but gives a logical explanation that the radiologist can understand, then the human plus machine together will get better outcomes than the machine by itself will.

Thompson **So let's get the humans and machines to work together and maybe, this time, society will become more equal and more just.**

# Preparing for the Next Outbreak

While vaccine creation made headlines, some countries have quietly battled steep distribution challenges. Now AI is strengthening supply chains to help solve the last-mile vaccine delivery problem.

by Sneha Mehta



← Illustrations by Sophi Miyoko Gullbrants

**At the height** of the coronavirus pandemic, most of the world's attention was on the race to develop a vaccine. But even after the mRNA vaccine was developed—processes that typically take several years were condensed into a few months—countries faced another significant obstacle, one that has long plagued immunization and public health efforts: How do we get these life-saving vaccines to the people who need them the most, as quickly and efficiently as possible?

The slogan "No product, no program" is used by many vaccine experts as a reminder that, without a consistent and reliable supply chain—the network of staff, equipment, vehicles, and data that carry a vaccine from the manufacturer to a patient—vaccine programs will fail. This is not merely theoretical; we have seen the success of efforts to contain diseases such as Ebola, polio, and smallpox hinge on the reliability of supply chains that can safely and reliably manage, store, transport, and deliver vaccines to people.

During the pandemic, however, low- and middle-income countries (LMICs), particularly those in the Global South, faced a unique set of challenges. Unlike their wealthier counterparts, they couldn't rely on vast resources to buffer against inefficiencies. Their primary hurdle was the challenge of predicting demand and matching it with their often-limited supply, straining already overburdened health-care infrastructures and leaving many people without access to vaccines. The disparities were stark: While today high-income countries boast vaccination rates of nearly 80 percent, only about 33 percent of people in low-income countries received a single dose.

While the pandemic highlighted the vulnerabilities of LMICs, it also revealed a curious trend among nations with the potential for innovation: doubling down on traditional methods. The Global Vaccine Action Plan, introduced in 2012 as a synchronized public health response to the aftermath of previous health crises, surprisingly anchors itself to a logistics framework conceived in the 1970s. India's decision in 2021 to earmark a staggering $1.4 trillion in projects to bolster its supply chain infrastructure over the next five years serves as another case in point. Such capital-intensive solutions, while foundational for growth, are increasingly being viewed with skepticism in an era when emerging technologies offer potentially more agile and cost-effective alternatives.

## A New Solution

Traditional supply chains operate on a fixed, predictable model with predetermined routes. Unforeseen disruptions can generally lead to significant delays. In this model, the

entire system is designed around a central hub, with spokes extending to various destinations. This hub-and-spoke design, while efficient in a stable environment, struggles to adapt quickly to changes or disruptions.

But the AI-driven supply chain, underpinned by more adaptive and responsive systems, continuously analyzes real-time data to make dynamic routing decisions. If there's a sudden change in demand in one location or an unexpected obstacle on a particular route, the AI-driven system can adjust, redirecting resources or altering routes as needed.

"AI can predict resource requirements and optimize inventory placement," says Derek Szopa, CEO at CloudSort, an AI supply chain start-up. "It can also help identify inefficiencies and automate repetitive tasks, freeing up time for more strategic decision-making."

In practical terms, if there's a sudden surge in vaccine demand in one city, AI can anticipate this need using predictive analytics and redirect supplies accordingly, all while accounting for factors like storage conditions, transportation times, and local infrastructure.

AI and machine learning are poised to be the technology that moves current vaccine distribution and supply chain systems into the future. According to an IBM report, 46 percent of surveyed supply chain executives anticipate that their greatest areas of investment in digital operations over the next three years will be in AI, cognitive computing, and cloud applications.

In Tanzania, the vast terrain and remote populations complicated the task of ensuring all individuals received their dose of the COVID-19 vaccine. Traditional methods of tracking vaccines, often based on outdated data, were proving inadequate. Pendulum, an AI-focused company, had already begun exploring innovative solutions for vaccine distribution even before the pandemic. In 2018, the company joined forces with the Ministry of Health in Tanzania to sift through public data, government records, and satellite imagery to predict vaccine utilization using the Connected Health AI Network (CHAIN). The outcome was significant: Three areas in Tanzania that implemented CHAIN saw vaccine wastage drop by more than 96 percent, showcasing the system's heightened accuracy in predicting demand.

"With vaccines, one of the major elements is being able to not only forecast demand but also properly allocate that supply to places that are equipped to receive it," says Brittany Hume Charm, head of growth, global health at Pendulum.

Uganda faced a similar set of hurdles. The reliance on paper-based forms by health-care workers meant that there was no centralized digital overview of the vaccine supply. This lack of clarity often led to last-minute vaccine replenishments, risking the efficacy of the doses. Logistimo, an AI-driven logistics solutions company, stepped in with a solution that seemed deceptively simple: a vehicle-routing product. This tool, by using machine learning to identify optimal delivery routes and adjust deliveries based on real-time needs, improved the efficiency of the distribution process. As a result, warehouse managers could consistently meet the vaccine demands, facilitating timely and effective distribution.

"Health workers are deeply occupied on the clinical side of things. Paperwork being a cumbersome task, reporting was often irregular, incomplete, or cooked—and planners upstream relied on poor data on which to base decisions," says Anup Akkihal, CEO of Logistimo.

The vehicle-routing solution by Logistimo represents a paradigm shift from the colossal, capital-intensive infrastructures of the past. Gone are the days when supply chain improvements necessitated the construction of expansive railroads or sprawling shipping ports. Instead, the modern approach, empowered by AI, is about agility and precision. It's about making the most of existing resources, optimizing routes, and executing every delivery, no matter how remote the destination, with precision.

## Navigating the Obstacles

Despite the promising strides made by AI in revolutionizing vaccine supply chains in Africa, a number of hurdles stand in the way of the AI-driven supply chain successfully scaling across LMICs, particularly in the Global South.

The first major challenge is data. In health care, data is the lifeblood of AI. However, the difference between LMICs and higher-income countries in data availability is extreme. Dykki Settle, chief digital officer of PATH, a global health nonprofit, draws attention to this by contrasting the "data jungle" of high-income countries with the "data desert" scenario more prevalent in many LMICs. Without the right kind of data, AI systems risk perpetuating or even introducing biases. The World Health Organization (WHO) underscores this concern, noting the frequent exclusion of marginalized groups from AI training datasets. This data gap is further widened by the digital divide, leaving millions, especially women, disconnected.

Emerging from the data challenge, the next hurdle is system fragmentation. In many LMICs, digital health initiatives operate in silos, making data integration akin to "a form of investigative journalism," says Hume Charm. For AI to truly flourish, these isolated systems must communicate seamlessly.

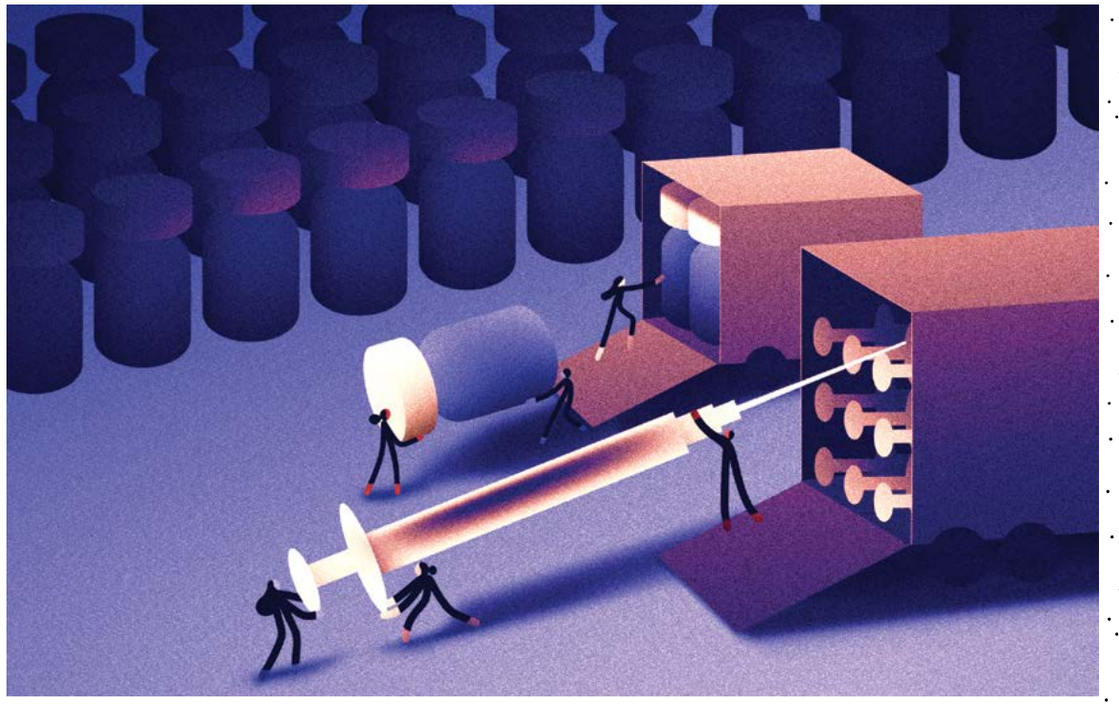
Solutions like Kenya's M-TIBA platform hint at the potential of interconnected systems, bridging the gaps and fostering collaboration. By connecting patients, health-care providers, and payers through a single platform, M-TIBA reduces the chances of data redundancy and provides health-care providers with a holistic view of a patient's history, leading to better-informed decisions. So far, the platform has helped to connect more than 4 million users and 1,200 health-care providers in the country, according to a McKinsey & Company report.

Finally, there's the problem of "technological solutionism," a perspective in which technologies like AI are seen as a magic wand to fix issues in LMICs in the Global South. Most developments in AI and machine learning are provided to LMICs by companies from higher-income countries in the Global North. But many public health organizations like PATH are wary of external vendors who come in, provide new technological solutions, and then leave the scaling and management to the local governments and health workers who may be overworked and underprepared for the job—culminating in the failure of that solution. Settle, from PATH, emphasizes that to avoid an "overreliance on external vendors," supporting local innovators and helping governments become self-reliant in the creation and use of digital systems should be a priority when implementing AI-backed health systems in LMICs. A deeper understanding of the cultural, structural, and institutional barriers is necessary for those systems to successfully integrate into the country's health-care infrastructure—so as to avoid becoming part of what Dr. John Bawa, team lead for vaccine implementation in Africa at PATH, calls "closets full of dumped technologies."

To handle these challenges effectively, some countries have adopted a collaborative approach. For instance, in some regions, local health-care professionals have been paired with international tech developers to co-design tools that are both technologically advanced and culturally apt. Additionally, pilot testing and iterative feedback loops have been used extensively to align technologies with the needs and values of the community. Such strategies support not only better integration but also the longevity and relevance of the technologies introduced.

For instance, the Mobile Alliance for Maternal Action used mobile technologies to deliver timely health information to new and expectant mothers in countries such as Bangladesh, South Africa, and Nigeria. These initiatives positioned local medical professionals and mothers to play a vital role in tailoring the content to the cultural and infrastructural realities of each region. This way, the technology

If there's a sudden change in demand in one location or an unexpected obstacle on a particular route, the AI-driven system can adjust, redirecting resources or altering routes as needed.



became integrated into the fabric of the community, rather than just becoming another "dumped" solution.

The future of AI for health care and vaccine distribution among LMICs in the Global South depends on the ability of innovators to create solutions that are as lean and creative as possible and work offline, reduce biases, and take advantage of local expertise—but their breakthroughs have the potential to affect the rest of the world as well.

"All of the tools that we designed to work offline in rural Liberia are applicable anywhere in the world," says Peter Lubell-Doughtie, co-founder and CTO of Ona, a company that provides data solutions to humanitarian relief and first response organizations. "Designing for that hardest use case helps us bring best practices everywhere else."

### A Way Ahead

The coronavirus pandemic exposed several deep cracks in the health systems of most countries. But it wasn't the first pandemic, and it certainly won't be the last: Climate change may cause future pandemics because of cross-species infections, and much faster than expected. However, public health experts who've worked in the field note that apart from periods of crisis, supply chain optimization is not always a priority.

"The whole world forgot about Ebola because we thought it was limited to parts of Western Africa. But it became a global epidemic by snowballing in a similar way that COVID-19 did," says Dr. Bawa. "If we had learned lessons from it, we could have been more proactive and more resilient and would have prevented the significant loss of life that we saw across the globe."

During the pandemic, Pendulum's collaboration with the state of California and the government of Sierra Leone showcased the potential of AI in addressing supply chain challenges. Confronted with outdated or incomplete data on cold-chain infrastructure, Pendulum bypassed traditional, time-consuming methods of data collection. Instead, it harnessed an enhanced web-scraping tool, analyzing satellite imagery combined with government datasets. This innovative approach not only identified the pandemic preparedness of health facilities but also pinpointed the location of six new COVID-19 laboratories in Sierra Leone. The same technology was later expanded to projects related to malaria and family planning in collaboration with the governments of Côte d'Ivoire and the Democratic Republic of the Congo.

Yet, the challenges transcended mere logistics. Vaccine hesitancy, defined by the WHO as the "delay in acceptance or refusal of vaccines despite availability of vaccination services," posed a significant barrier. Misinformation, mistrust in public institutions, and cultural differences led to reluctance in vaccine uptake in various regions. In response, Pendulum and PATH used AI to monitor digital traces like social media posts and identify areas where misinformation was rampant. This data-driven approach allowed governments and public health agencies to tailor their communication strategies to those regions, ensuring efficient vaccine distribution and uptake.

"Demand forecasting is not as simple as just getting a vaccine from point A to point B. You have to also make sure that someone is willing to walk into a clinic and get it," says Hume Charm.

Learning lessons from the past and creating next-generation supply chains to prepare for an inevitable future where pandemics will be more common is a key area of focus in the WHO's Immunization Agenda 2030: "Strengthen supply chains to ensure that high-quality vaccines are always available in the right quantity and form at the right time, in the right place and stored and distributed under the right conditions. Promote integration with other supply chains for more effective delivery of primary health care. Invest in systems and infrastructure to safely manage, treat, and dispose of vaccine waste to help reduce their environmental footprint."

Ultimately, for AI to be a part of the pandemic-preparedness plan for vaccine supply chains requires a systemic paradigm shift: one that accepts both that humans may not have all the answers needed to allocate scarce resources in the most optimized, equitable way possible, and that training AI and machine learning to respond to the volatility of public health scenarios may reveal outcomes we've never imagined—outcomes that help health workers reduce senseless deaths.

"Each vaccine that you place at a given location at a given time is a bet about demand. And when you have a limited war chest of resources, you need each of those bets to be extremely high quality," says Benjamin Fels, CEO and co-founder of Pendulum. "In every industry where AI and machine learning have been allowed to learn the answer, it has outperformed any other approach by orders of magnitude. For me, improving a supply chain where allocation is a matter of life or death is the most meaningful use of AI."

"For me, improving a supply chain where allocation is a matter of

life or death

is the most meaningful use of AI."

Benjamin Fels, CEO and co-founder of Pendulum

# Can AI navigate human emotion?



AI excels at systems of logic and pattern recognition. Human emotion, however, is all but logical: Games like poker, which reward intuition over logic, remain curiously out of reach. Humor, too, with its layer of nuance, often eludes computational grasp. Sarcasm baffles algorithms, context escapes chatbots. This realm persists as profoundly, uniquely human—shaped by cultures, experiences, and personalities.

Scientists attempt to train AI to recognize emotions and communication through two paths: By studying and reverse engineering human psychology, and by building machine learning systems that deduce patterns from data. However, AI models struggle to decode nuanced social cues.

Might that change? And if machines could truly understand us, what would we see reflected back?

←
Illustration by Camilo Huinca

# When AI Meets Its Match

Today’s poker bots can crush even the best human players. Still, the game—one of bluffing, deception, and intention—remains technically unsolved.

**By Maria Konnikova**

THE

The emerald beast is begging me to engage. And I, of course, take the bait. With a single motion, I bring the creature to life. It throws out its first punch. I parry with a raise of my own. It defends itself. I slash again. It strikes back with even greater force. But I have plans for a counterattack and I rush forward. My furor finally cows him and he admits defeat. Meek. Deflated. Beaten. At least until the next hand.

Our weapons are cards. Our battlefield, a virtual poker felt. And my opponent's name is Slumbot—a poker bot that was, up until 2018, when the last Annual Computer Poker Competition was held, the world's toughest virtual AI opponent in Heads-Up No-Limit Texas Hold'em. And though that competition is no more, Slumbot remains the benchmark program against which all future poker AIs will test themselves. All future poker AIs, that is, and me: the AIs, in the service of technological breakthroughs into the very nature of human decision-making; me, in the service of journalistic exploration. After I finish battling the Slumbot, I will move on to its most feared nemesis, a wily model formerly known as Ruse but now called GTO Wizard AI—the current gold standard for optimal poker play.

Against Slumbot, I may stand a chance. Against GTO Wizard AI, I'm certain to lose, by definition—I'll be playing on its own platform, judged by its own standard of perfection. But that doesn't mean I'll lose altogether. After all, AI researchers have yet to fully solve poker—and, as I came to find, even against the toughest bots, humans have a certain advantage. Though AI can sometimes mimic some human emotion, it lacks the intuitive grasp that defines the human decision-making experience in the face of limited data—those nuances of behavior that can, at a moment's notice, change the entire tone and direction of a battle of wits. Even a battle with as much mathematical precision as No-limit Texas Hold'em.

*

Poker isn't the first game AI has tried to solve. In 1989, a program called Chinook began to churn out a series of computations that would, at their peak, occupy more than 200 computer processors around the world. It was one of the longest-running computations of all time. The end result was announced in 2007: Chinook had solved the game of checkers. It had crafted an AI-driven approach that would never lose against any opponent.

There's a reason why AI researchers have set their sights on games: They have rules. They have a fixed, defined world. Even the most complex, sophisticated games are cleaner and less noisy than life. But with all their rules and stipulations and neat parameters, they still have the element that is most true to the real world: humanity. And that's what makes them such a powerful proxy for studying real-life decision-making.

In the 1950s, John von Neumann, a polymath mathematician best known as the father of game theory, proposed that the true prize lay not in the world of checkers and other perfect information games, but in the world of imperfect information games, where, as in life, the unknown was just as crucial as the known, if not more so. The most lifelike game of all? Poker. The game of human intention and bluffing and emotion and seemingly endless recursive thinking. As von Neumann saw it, games like poker were much more than games. If you could tackle them, they would help form a rubric for taking on the thorniest problems of humanity.

But though many researchers took up the challenge to solve these infinitely more complex puzzles, for decades none came close. To date, even perfect information games, such as chess and Go, haven't been solved in the technical sense of the word. The AIs can beat the best humans, consistently, but they are unable to enumerate every possible situation that may arise in the game tree—a necessity for a real solution. So how could any researcher hope to conquer poker? Certainly not by the brute force approach that had cracked checkers.

In poker of the No-limit Hold'em variety—the most popular form in the world, during which a player can bet any amount, up to her entire holding—the number of possible situations is greater than the atoms in the universe. Add to that mathematical unwieldiness the very *human* nature of the game, and you have a problem of compounding difficulty. How can an AI parse the shifting emotional dynamics of a table? How can it fight back if a few human opponents decide to single it out and collude, even on an unconscious basis, against it? (While outright collusion is illegal, subconsciously altering your play to single out the "other" at the table, be it an AI or a human outsider, is far from rare.)

For Michael Bowling, that very complexity was the draw. In poker, he saw a problem that neither checkers nor chess could approximate: how other humans respond. "You



Even the most complex, sophisticated games are cleaner, less noisy than life.

But they still have the element that is most true to the real world: humanity.

can't ignore other agents in poker," he says. "You need to know how everyone is going to behave." In other words, exactly what von Neumann had proposed decades earlier.

Working as the head of the Computer Poker Research Group at the University of Alberta, Bowling started with a more manageable poker variant: Heads-Up (that is, one-on-one) Limit Texas Hold'em (for which the amount and number of bets is limited). It took many years, but at last he had it: a program called Cepheus that could decisively beat the game. Its main breakthrough was an algorithm known as CFR, or counterfactual regret minimization. The algorithm compares all future actions to determine which approach will cause the least amount of regret—that is, that no other possible action would have led to a better outcome. To date, Limit Hold'em is the largest imperfect information game that AI has been able to solve—and it is orders of magnitude simpler than Heads-Up No-Limit, which, in turn, is orders of magnitude simpler than Multiplayer No-Limit.

The complexity, however, is the allure—and the reason that Bowling next set out to conquer Heads-Up No-Limit poker. And while he didn't solve it in the same way he could solve its Limit variant, he did deploy an important new tool: a neural network that could decompose the bigger game into smaller, subgame problems and recalculate an appropriate strategy at every step. The new program, DeepStack, was able to defeat 11 poker professionals over the course of 44,000 hands. That was still for Heads-Up only, and it left the game far from solved, but DeepStack was nearing superhuman ability—and reached it, at least as far as some humans were concerned.

At Carnegie Mellon University, a team led by Tuomas Sandholm was approaching Heads-Up No-Limit from a different angle: Rather than use neural networks, they would start with sophisticated abstractions of the game. Claudico, their first bot, failed miserably. Noam Brown, then Sandholm's Ph.D. student, who programmed most of the bot's algorithms, eventually determined the problem: Whereas the human players would sit and think, the bot would act immediately. It had spent countless hours training in advance, playing trillions of hands on a supercomputer, and would use that training to act instantaneously.

Brown decided to program in-game thinking into the bot's abilities (something DeepStack did as well). The result was Libratus, a bot that looked at the subgame *during* play and recalculated its strategy accordingly. When Libratus challenged the top humans to a match, it was much better prepared. What's more, every night, it would hook back up to CMU's supercomputing center, analyze how the humans had played, and adjust its strategy. The result, in 2018, was a decisive victory, one that took DeepStack's win to the next level.

And then came the next milestone: multiplayer poker. After the success of Libratus, the CMU team turned to six-max poker, a variant with six players. The new algorithm had one major difference: Instead of just solving a subgame, it would solve a depth-limited subgame. "It can start solving the game when you are already in the game, not from beginning to end," Sandholm says. This was Pluribus, a bot that performed quite well against some humans in a multiplayer format. CMU declared a victory.

*

I haven't come to my matches against Slumbot and GTO Wizard AI, my two poker AI nemeses, empty-handed. Before playing, I consulted with Kevin Rabichow, one of the best Heads-Up players in the world—and the poker consultant for GTO Wizard AI. Rabichow's initial prognosis is grim: Slumbot is superhuman, he told me. But it isn't perfect—and, crucially, it does not adjust to its opponents. Instead, I can adjust while it will play the same game it was programmed to play.

I play a few trial hands. I lose some. I win some. I start taking notes, much the way I would against a human opponent. One of the first things I notice is that Slumbot likes to bet small with value hands that are not the absolute nuts (nuts being the best possible hand). My adjustment? Either fold or raise, depending on the situation—because just as Rabichow predicted it might, it overfolds to big raises, even when it holds a solid hand.

Every time humans negotiate,

it's like playing a hand of poker to the best of our ability.

I start accumulating chips quite rapidly. And then I make a mistake, running a big bluff even though Slumbot has called every bet. I should know by now that if the machine doesn't fold to aggression, it has something strong—as indeed it does. It's a costly mistake that brings me down for the session. I chalk my loss up to two factors: I'm tired, and I was distracted by some texts on my phone. (Both true, but neither a good excuse.)

Here's one major edge that Slumbot has over me: It doesn't get distracted or tired. It doesn't think about what's for dinner or how close I am to the requisite 500 hands I promised Kevin I'd play. It just executes its strategy, over and over and over, with precision.

I dial back in. I fight back. It turns out Slumbot is teaching me something important about myself as I play: Without the distractions of the live table or the time pressures of real online play, I can pay attention to my own strategy, my own feelings, my own shortcomings much more clearly. Slumbot may be an AI, but it's evoking very human responses.

We play on. At the end of 750 hands, I emerge victorious, having won 122.5 big blinds. If we'd been playing for real money, that would have been $12,250 in just over two hours. "LFG!" Kevin texted me after I sent him my results. "Humans can still win this thing."

Yes, they can, even in a format like Heads-Up, poker AI's strongest suit. At least, they can against the best bots of yore. Because here's another thing about Slumbot—a limitation that currently marks every existing poker program and plays a crucial role in evaluating any existing AI against humanity. It's stuck in 2018. Even though a few tweaks have been made since it won its final competition, it is essentially the same bot now as it was five years ago. Poker, however, has evolved.

The world is not static. A strategy that was optimal last year, in poker or in anything else, may no longer be optimal if the environment has changed. Someone who is at the top of their field may find themselves struggling if they stop learning while their competitors keep evolving.

When I began playing poker, as research for a book on the nature of chance and decision-making, I had to keep learning to stay competitive. If I ignored a new tool or tactic, I would lose. But if I embraced it, I still had a chance of winning. So when I faced Slumbot, as relatively bad as I am at Heads-Up poker compared to someone like Rabichow, I could still win. The me that has been studying the game's evolution is better than the superhuman AI of the past.

Soon after I beat Slumbot, it's time to face GTO Wizard AI. I feel like I will be lucky to survive. Even though, like Slumbot, GTO Wizard AI is unable to adapt to its opponent, its base algorithms are much more powerful and, in Heads-Up combat, GTO Wizard AI has left Slumbot in the dust.

I quickly realize that in my 750 hands against Slumbot, I've picked up some bad habits. Against the easier opponent, I started playing far too many hands. GTO Wizard AI will have none of that. If I play a marginal hand, I am immediately punished. This results in two massive blunders within the first few hands that immediately set me back, causing me to lose most of the big blinds that I will lose this session. For the remaining 100 hands, I'll be battling back with a big handicap.

I find myself enraged at this stupid AI, which judges players by a standard of "optimal" play, when it labels my decision to call, instead of raise, a bet in a particular spot a blunder. But I had my reasons. Given the board, there could have been a higher straight than mine. And I didn't think my raise would ever be called by a worse hand. GTO Wizard disagrees and dings me an insane 14.9 big blinds.

Eventually, as so often happens, anger gives way to self-reflection. Maybe the program is right—maybe it spotted a leak in my game, a risk aversion that prompts me to take the more cautious route when I should instead opt for aggression. Then again, against flesh-and-blood opponents, maybe my intuitions are the better guide. Most players in that situation *wouldn't* bet in that way unless they had me beat—and most wouldn't call a raise unless they had me beat. That's the difficulty of playing an optimal opponent and being judged by those standards: They are incredibly useful but can lead you astray against humans who play anything but optimally. The AI has played vastly more hands than I ever will—but has never learned to parse the movements of an opponent's fingers, the look in her eyes, the pulse in her neck.

In the end, I recover. At least somewhat. I lose an average of .06 big blinds per hand, or 5.8 big blinds in the 100 hands I play this session. "How bad is that?" I text Kevin. To my surprise, he responds, "Overall seems quite good." Jubilation.

I will go on to play several more sessions. My performance remains at a steady .06 average. I'm sad I don't improve, but happy I don't go up in flames. I feel like I've held my own. Even in this format. Even against the best of the best.

*

To researchers like Bowling and Sandholm and Brown, it doesn't matter that poker isn't solved as such or that the game has progressed beyond their models. Their goal was always the same as von Neumann's: poker as a tool. Sure, many of the researchers—von Neumann first and foremost—love the game. But as a research program, it is a benchmark, a waypoint to AI in service of the greater good of humanity.

Amy Greenwald, an AI researcher at Brown University who collaborated with Bowling in his DeepStack research, is working on negotiation, which she sees as the most important game theoretic problem in the world. "Can we try to predict how agents will act? Can we steer them toward positive outcomes? That's what poker has given us," she says. Consider even the most simple negotiation problem, between two agents. Who acts first? What do they say? Did you offend the other person with your initial offer? Did your stance make you seem overeager? "In negotiation, I need to give you an offer *now* without revealing so much about my hand that it undercuts me," Greenwald says. "I need to learn how you think—your function, in machine terms—to try to sway you in my direction, eventually." Every time humans negotiate, it's like playing a hand of poker to the best of our ability—trying to discern what the other player holds and how far you can push them without revealing too much about your own cards and how far *they* can push *you*.

Sandholm would agree, and he's directly leveraging the algorithmic insights of poker into very real problems via several start-ups. At one of them, Strategic Machine, he works on applications like political campaign planning—a game of poker if ever there was one. "Take a very simple campaign problem: How do you allocate money on various types of media?" Sandholm says. "It all depends on your opponent. It's pure game theory. But people don't usually take game theoretical approaches to campaign allocation." Poker has, to these researchers, served its purpose.

Humans tilt—a poker term for the human tendency to inject emotions into their decision process. Humans celebrate. They cry. They lie—and not just when bluffing. They get greedy. They become risk averse. They become risk-seeking. They like each other and hate each other. They feel like you are out to get them. Sometimes, they don't know why. Dynamics change, often at a subconscious level. Humans become more aggressive against someone, less aggressive against someone else, a give-and-take that changes strategies and outcomes.

That very humanity is what drew John von Neumann to the game. And in his theory, he challenged us to remove it, to reduce it to equations that would, ultimately, be solvable. He and his successors have almost succeeded. As Bowling put it, "Poker has a very human element. But von Neumann was so successful he almost removed it." That almost, however, does a lot of heavy lifting.

As a psychologist, I know this about the human mind: What we don't know far outnumbers what we do. We can't accurately say why we act the way we do, let alone why others do. When Marion Tinsley played against Chinook in the first Man-Machine World Championship in 1992, he was certain he would win. "I have a better programmer than Chinook," he told *The Independent*. "His was Jonathan, mine was the Lord."

AI can improve all it wants, but humans will always be human. We don't know quite what that means. We can't quite assess how it will play out. That's our downfall. But it's also our strength.

Maria Konnikova is a writer, psychologist, and poker player. She is the author of the books *The Biggest Bluff*, *The Confidence Game*, and *Mastermind: How to Think Like Sherlock Holmes*. She is currently working on a book about cheating in games.

Jokes are deeply human.
Can machines learn to tell them?

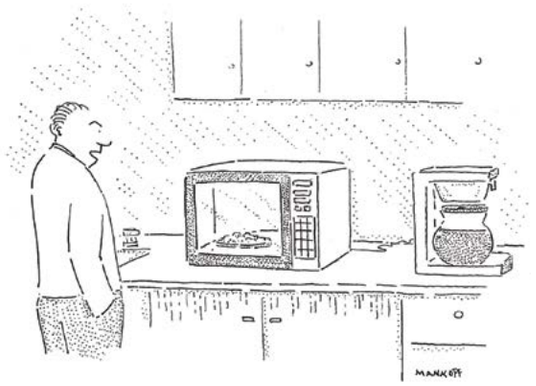
By Bob Mankoff

# A BOT WALKS INTO A BAR

**WHEN KEN JENNINGS** lost to IBM's Watson on *Jeopardy!* in 2011, below his Final Jeopardy answer, he scrawled, "I, for one, welcome our new computer overlords." By doing that, he easily beat Watson in the humor category, just by using a well-known meme formula signaling submission to a perceived all-powerful force to which resistance is futile.

Formulaic humor would seem precisely what an AI like Watson might have used to banter with Jennings. Yet Watson's AI of 2011 could no more come up with a snappy comeback than its humanity-humbling predecessor Deep Blue when it took the crown from chess king Garry Kasparov in 1997.

I noted Deep Blue's supposedly watershed victory in this *New Yorker* cartoon of mine, which foresaw a time when it might be my turn to say, "I, for one, welcome our new cartoon overlords."



*"No, I don't want to play chess. I just want you to reheat the lasagna."*



If that comes to pass, I might have had a hand or two in my own humbling. I'm the president of CartoonStock, the world's largest database of single-panel cartoons like the one above. While I don't know if any AI system has used it yet, by making thousands upon thousands of cartoons publicly available and easily scraped, I essentially created a gold mine for them.

The *New Yorker* Caption Contest is also my brainchild. Since 2005, *The New Yorker* has published a cartoon without a caption every week and asked readers to compete to write the winning caption. In 2016, the magazine began relying on an algorithm to sort the 5,000 to 10,000 caption entries per cartoon by funniness, aggregating voters' opinions to present ranked lists. The combination of the prestige of the *New Yorker* cartoons and the unique quality of this dataset present the opportunity to give computers what they have always lacked: some humor.

The root of humans' sense of humor has nothing to do with being able to "get" a joke or make one. Rather, it's the answer to life's existential problems that have no solutions, the blessing we receive in exchange for the curse of mortality. As Mark Twain said, "The secret source of humor is not joy but sorrow. There is no humor in Heaven." AI has no sorrow and thus no need for humor, and creating that need would be cruel. We would have to give a machine sentience enough to suffer and vulnerability enough to die.

And yet, step by step, AI is getting closer to understanding what makes something funny. Perhaps unsurprisingly, considering the rapid advancement of large language models, it's starting with captions. In 2016, I got an email from Vincent Vanhoucke, then Google DeepMind's principal data scientist, now the senior director of robotics. "I believe that the success of artificial intelligence will ultimately not be measured in how well it can do at games like Go or whether it can respond to your emails," he wrote, "but rather whether an artificial mind can one day get its cartoon caption published in *The New Yorker*."

I wasn't surprised that someone as high up the AI food chain as Vanhoucke was interested in the contest. It had been on the radar of the AI community for years, cited in a wide range of academic papers. But I was taken aback by the ultimate ambition, which appeared to be nothing less than creating—gasp—a Bot Mankoff. Winning the caption contest seemed to be merely one small step for machine kind.

At the time, even DeepMind's whole bag of AI tricks couldn't make a bot funny: Its combination of computer vision, convolutional neural networks, semantic vectors, and the oxymoronic Long Term Short-Term Memory wasn't up to the task of even entering the contest. To be honest, I doubted that AI ever would be. I knew the technology was advancing at a mind-blowingly fast rate, but even if the machine-learning magic were up to snuff, I couldn't see how all the data and equations in the world could replace the fundamentally human heart of humor that Twain described.

*New Yorker* cartoons can be zany or intellectual, or philosophical or spot-on observational, but they all share an understanding of what it means to be a certain type of person. In my own work, my objective is not to punch up or down, but to elbow to the side: to make fun of myself and others like me. The best cartoons cause us to laugh with them as well as at them. Even the best AI doesn't have human peers with whom it shares interests and pet peeves. How could it beat us at making fun of ourselves?

Vanhoucke wasn't a fan of this line of thinking. He was convinced it was only a matter of time before AI began competing seriously in the caption contest. And I've started to conclude he's probably right. Last year, six years after our initial conversation, he sent me another email. "While we may still be a long way away from AI producing humor," he wrote, "our latest models have become surprisingly adept at understanding why something is funny." AI wasn't writing jokes yet, but it seemed to be grasping the concept of what made a particular joke work.

The model Vanhoucke was describing was PaLM, a version of which now powers Google's Bard. Here's how it describes my 1997 cartoon about the chess-playing microwave:

> The joke is that the microwave is so advanced that it can play chess. The man is frustrated that the microwave is so advanced that it can play chess, but it can't reheat lasagna. The joke is funny because it is absurd. The joke is also funny because it is a commentary on the fact that technology is advancing so quickly that it is outpacing our ability to use it.

I'd rate this a solid B- explanation. Soon, PaLM will probably be able to earn an A. On one hand, so what? There will never be a coffee-table anthology of *New Yorker* cartoon explanations.

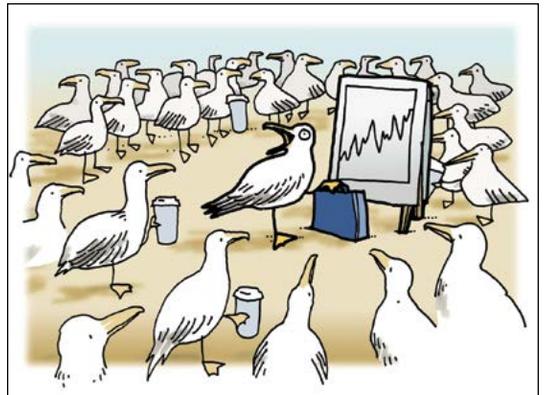
But the ability to understand humor is a key stepping stone toward the ability to create it. In a paper last year, a group of researchers led by Jack Hessel of the Allen Institute of AI, with some curation assistance from me, challenged AI models with three tasks: 1) matching a caption to a cartoon, 2) identifying a winning caption, and 3) explaining why a winning caption is funny. In all three categories, humans remained superior to even the most advanced models. The AI's best performance came when describing the humor behind each cartoon, just as Vanhoucke found. For the Hessel paper, AI wrote 653 explanations for caption contest winners, creating its own database of what makes captions funny. Now someone could simply plop a bunch of the descriptions from the paper into Bard, ask for more, and rinse and repeat until the model has mastered every possible joke formulation.

Next, Hessel attempted a more sophisticated spin on things. The AI model developed 50 ideas for cartoons, each with five caption possibilities. From those 250 combinations, I picked the four I liked best, and cartoonist Shannon Wheeler drew them. Here are the results:

*"The voyeurism exhibit really is quite immersive."*

*"It's not the Met, but at least she's on the right track."*

*"Say Bananas!"*



*"Good news, we're expanding the seafood division."*

Shannon wasn't impressed with the output. "Weird cartoon ideas. They lack the implied narrative that's a solid *New Yorker* cartoon," he summarized.

I agree with him, but that doesn't diminish the scale of the accomplishment here. First of all, the training set of caption-contest cartoons that we fed into the AI model was intentionally bizarre. "A giant fish is seated at the bar with six empty shot glasses in front of it, gesturing to a bartender to bring another round," one description read. "Museum workers looking at two dinosaur skeletons in a dancing pose like old-time vaudevillians with top hats and canes at a museum exhibit," said another. Weirdness in, weirdness out.

And even without the narrative power of a *New Yorker* cartoon, the immersive museum gag and Brunhild on the subway evoked a smile from me. That means AI created at least serviceable cartoons out of nothing: Neither the captions nor images were in its training set, and to my knowledge, they did not previously exist anywhere else either. And for the sake of a clean experiment, we played it completely straight, not altering either the caption or the image description at all. We could have achieved much better results if AI-human collaboration had been permitted.

This points the way toward the most likely role for AI in cartooning: not a replacement but a brainstorming tool, helping creatively blocked cartoonists come up with ideas that the human can then improve upon. My captions will always feel more human than a machine's because they arise from life in the real world—my own emotions, annoyances, and grievances. But just as some digital native cartoonists prefer iPads and e-ink to pen and paper, some also may like to use AI to reach greater creative heights.

I have no wish to welcome our cartoon overlords. But there's also no need for me to shun AI models as potential collaborators, creative assistants, or inspirers. Cartoonists have shown that they are alchemists extraordinaire. I'm convinced they will be able to use this tool to augment their alchemy. I'm just as sure there will never be a day when robot cartoonists are creating robot cartoons for robot readers of *The New Yorker* to laugh at.

---

Bob Mankoff is the president of CartoonStock and the former cartoon editor of *The New Yorker*.

# How Should AI Be Governed?



As AI advances, the need for good governance looms large. Inevitably, regulation will happen at many different levels. Technology companies, state governments, federal governments, and international organizations will need to work together to define guardrails. Norms will be forged through public debate. Education can empower citizens to weigh risks and uses. Art, media, and storytelling will shape cultural attitudes.

As AI's rise redefines power structures, governing it justly begins with a simple question: Can we take steps to ensure everyone benefits from it? Underpinning it all are human choices. The development of governance standards should summon our moral imagination.

Q&A

# The Sweet Spot

A conversation with Google’s Kent Walker on AI regulation and striking the right balance

By Nicholas Thompson

→
Photography by
Cayce Clifford

**Almost everyone agrees** that we need to regulate AI. Surely, we think, we can come up with new government policies that will help to maximize the benefits and minimize the harms of AI. But regulating AI is like trying to paint an airplane while it’s in flight. It’s moving so fast, it’s hard to place the brushes right.

I spoke with Kent Walker, president of global affairs at Google and Alphabet, about what the company thinks effective regulation might look like. Our conversation has been edited for length and clarity.

Nicholas Thompson **Everybody wants AI regulation—what should be the central objectives?**

Kent Walker An AI agenda needs to rest on three key pillars: opportunity, responsibility, and security. If we get those right, we think we can deliver on the promise of AI for everybody.

Thompson **So a regulatory framework should optimize those three pillars?**

Walker That's right. We tend to focus on AI as a chatbot. But it's important to remember not only that we have all been using AI for many years—if you've used Google Search or Translate or Maps, you've been using AI—but also that its biggest potential is still ahead in changing how we do science and technology. That will lead to advances in areas from medicine to energy to sustainability to agriculture to economic productivity. So the opportunity side of AI and enabling its potential is going to be critical.

At the same time, you need to focus on the responsibility and the security aspects. Responsibility being things like [ensuring] high-quality results and avoiding problems with discrimination and toxicity, the risk of misuse and abuse. And avoiding security challenges—cybersecurity, national security, and global competitiveness.

Thompson **So any regulator thinking about AI needs to make sure that the innovators are allowed to innovate, companies are allowed to grow, inventions are allowed to move forward, that the AI that's developed doesn't cause harm, doesn't create biases, doesn't create toxicity, and that we're not hacked and destroyed. Is that it?**

Walker That's a good way of framing it. We view AI as an advance in mathematics. I think we'd be having a different conversation if we were calling it computational statistics, which is probably a more accurate way of describing what's going on here. But thinking through how we can apply these breakthroughs in computer science to achieve these benefits and minimize the harms is exactly the right balance.

Thompson **Let's go to one of the specific frameworks or policies that you propose, which is a hub-and-spoke approach with the National Institute of Standards and Technology taking the lead. Explain what this idea is.**

Walker If you think of AI as a general-purpose technology, and we believe it is—something like electricity—we don't have a Department of Electricity. We have government agencies that are focused on particular areas where issues come up. The issues that AI presents in healthcare are different from the opportunities and issues in financial services or in transportation. We have agencies across governments that have worked on those issues for many, many years and have a lot of expertise about the potential risk of abuse. It's going to be a lot easier to make every agency an AI agency than to have a one-size-fits-all solution where you try to take all that learning and put it into one place.

Thompson **We don't have a Department of Electricity, but is there not something fundamentally different about modern AI in its ability to precisely replicate and supersede human intelligence in so many ways?**

Walker Again, there's no one thing that you can call AI. AlphaFold, developed by Google DeepMind, used AI to predict the shapes of 200 million proteins—nearly all the proteins known to science—in just a matter of weeks. That's a very different kind of pattern recognition than what you're seeing in some of the generative AI chatbot tools. This is a huge leap forward in computational ability, and it's going to play out in different ways.

If you're concerned about misuse by terrorists, that's one class of issues that you need to deal with. If you're concerned about the potential for fraud and abuse, that may lead you down a different track. Or if you're thinking through how best to regulate its use in stock trading, that's yet a different category of things. Having a center of expertise, like the National Institute of Standards and Technology, is helpful to develop more state capacity to understand what's going on and keep up with all the different flavors of AI that we're going to see in the coming years.

Thompson **How does Google balance its work on the opportunities of AI versus the responsibilities and security of AI? Clearly you have an economic imperative to build out the tools that can make the most money for the company. How do you weigh that against the need to make sure that you're hitting your requirements for responsibility and security?**

Walker We've been doing this for many years now, both on the technology side and on the responsibility side. We had our first team working on ML fairness in 2014. We published our AI principles in 2018. And we've continued working through the internal governance of how best to do that ever since. There are products that we have decided not to bring to market because we didn't think the appropriate policy frameworks were in place—things like generally available facial recognition tools. It's a continuing balance.

We have a long-term interest in encouraging general social trust in and adoption of AI in a whole variety



We have all been using AI for many years—if you've used Google Search or Translate or Maps, you've been using AI—but its biggest potential is still ahead.

of different areas. And we think, if responsibly managed, it will be a huge positive for societies around the world.

Thompson **This is one of the most interesting things to me. You have regulators at the local level, at the state level, at the national level. But you also have decisions being made at the companies.**

Walker Well, you're right that it needs to be a multilayered framework. It's not going to be just government regulation or individual companies working on this. We need to have a spectrum of approaches from individual companies, cross-industry groups like the Frontier Model Forum, which we co-founded with other leading AI companies and labs, as well as [companies] working with governments.

And governments are moving at pace to address this. In the United States, the Office of Science and Technology Policy published its AI Bill of Rights. The National Institute of Standards and Technology put out its AI Risk Management Framework. We joined the White House commitments this summer and more recently Sen. Chuck Schumer's AI forum in the Senate. So the right conversations are happening, and there's been good public-private collaboration.

Thompson **What is the most interesting question you're grappling with right now?**

Walker There's a lot of discussion around the balance between open and closed. Openness creates democratized access to tools, but also risks of abuse by nation-states or bad actors. How do you reconcile that, and what's the right balance? How do you think about the scope of regulation?

When people talk about AI regulation, they're not typically talking about Google Maps. They're thinking about the challenges posed by emerging next-generation AI tools. But how do you draw the line between those two things when there's not an obvious cutoff? How do you think through the standards for what "morally good" looks like, and how do you measure that? That turns out to be a deep and hard question across all the different areas we've talked about, whether it's bias or toxicity or privacy or other issues.

Thompson **I know that with Google Duplex, the voice assistant that seems like a human, you always declare that it's a machine. Do you think that there should be a formal policy that any system using AI that could be reasonably thought to be impersonating a human should declare that it's a machine?**

Walker We took a step in this direction recently when we announced that we would require disclosure of the use of generative AI if it resulted in inauthentic and misleading election ads. So, for example, people for years have used Photoshop or other tools to avoid red-eye in photos or to touch things up. And we don't require disclosure of that. Or if somebody uses AI to pose somebody with the U.S. Capitol's dome in the background, that probably is not misleading. But if AI is used to replicate a human voice so that somebody seems to be saying something they never said, or to create an image that never occurred in a way that could mislead the viewer, we think it should be disclosed.

Of course you don't want to require overbroad disclosure because AI is likely going to be used in almost everything we do, from the writing of articles to the creation of new photos to many of Google's tools and services. Labeling all of that as AI would be like labeling pictures with "This was created with a camera" or your article with "This was created with a computer." So we have to figure out exactly how we have meaningful disclosure so that people aren't misled.

Thompson **What is the biggest risk of regulation?**

Walker There's always a risk of under-regulation and over-regulation. Under-regulation could create a possibility of misuse by rogue actors or abuse that really creates harm in the world and undercuts trust in AI. But bad regulation could slow our ability to achieve the promise of AI. We said some years ago that AI is too important not to regulate

We need to move beyond the bumper-sticker conversations and get into a more detailed discussion of the trade-offs, and how to develop risk-based regulations that don't limit that promise.

and too important not to regulate well. That sums up the balance needed. There's a sweet spot, and we're hoping to share the benefits of our experience with governments to see if we can hit that.

Thompson **Do you think there's a risk that because of some concerns about AI—the doomsday scenarios and other rhetoric—we could end up with regulation that is focused in the wrong direction? That trying to prevent the worst outcomes prevents many of the best ones?**

Walker There's a phrase in the AI community, "the AI half-pipe of heaven and hell," referring to headlines that go back and forth between "AI is wonderful" and "AI is terrible." In fact, like most technologies, AI involves both opportunities and challenges, and we have to figure out how to get the balance right. We need to move beyond the bumper-sticker conversations and get into a more detailed discussion of the trade-offs and how to develop risk-based regulations that don't limit that promise.

Thompson **Do you think that these powerful tools necessitate differently structured data privacy laws now that so much of our data can be used to train models that have so many uses?**

Walker The vast majority of training for the large language models we're seeing now is actually on public data that's on the web. We've been clear that we have used that kind of data for many years to improve the quality of Google Search. Privacy is going to be an important piece of this, and we also have to think about the security of data, making sure that the models aren't misused.

We want to be clear with people when information is being used to improve the services. When you use Google Search, if you click on the fourth item in your search results, that's a signal to our system that maybe that item should have ranked higher. When you use Google Maps on the road, it helps us when other people know where there are traffic jams. It's a collective action benefit. We will continue to figure out how to make sure that people are aware of that and feel comfortable sharing feedback.

Thompson **So it's really a question of proper and clear disclosure of how data that you've inputted or data about you will and could be used?**

Walker I would say both disclosure and control. Individuals should have the choice and the ability to choose whether or not to have a system trained on their data in a way that's personalized and customized to their needs.

Thompson **Do you think there's any risk that aggressive AI policy could lock in the playing field as it is? So, for example, you can imagine a regulation that insists that every AI model has to be audited, which would require huge compliance costs, which would then make it extremely hard for anybody but a large, well-funded tech company to comply.**

Walker It's always a risk that regulation becomes a barrier to entry for new competitors. We've seen that with other tech regulations around the world—where we have devoted significant time and expense coming into compliance in a way that would be harder for a smaller company. So you have to take that into account. That's not a reason not to regulate, but it is a reason to say that you should be regulating only where you need to in response to specific concerns.

Of course, some of the abuses could happen with big companies or with small companies. So you want to make sure that you have appropriate rules that apply broadly but are tailored enough to the underlying issue that you're not making it harder for smaller companies or start-ups to innovate.



Thompson **There's an idea I've heard, that one of the goals of regulators shouldn't be just setting the right frameworks and the right rules, but also creating data commons. And anybody could access these data commons. So they'd be publicly accessible data sets that could be used to train AI models and to facilitate the creation of AI models. They could be used by big or small companies. Do you think this is a helpful idea or a potentially dangerous one?**

Walker In general, we're in favor of data commons structures. In fact, Google's Data Commons team just partnered with the United Nations to use data and AI to help track progress toward the UN's Sustainable Development Goals for the globe. And we support proposals to fund a National AI Research Resource. But right now, in many ways, the biggest challenge is not the availability of data. Remember that the world's three leading AI labs—OpenAI, Anthropic, and DeepMind—all succeeded without access to any proprietary data.

The biggest challenge for a lot of the up-and-coming companies right now is access to computing power. Now, as AI develops into more specialized areas, specialized data may become more important. If you're developing an AI chemistry tutor, you may want to train it on chemistry textbooks. Or, for applications for a given company, you may want to train on that company's data to help it stock its shelves or manage itself more efficiently. But the general-purpose models, the generative AI models that most folks are concentrating on right now, are mostly being trained on publicly available data.

Thompson **What is the best way for governments to quickly develop the expertise to be able to make the right choices, make the decisions, and move nimbly?**

Walker This is a core question of state capacity. There's a huge amount of technical complexity that underlies the dramatic computer science advances we've seen in the past few years. Figuring out how to get governments up to speed on developments that have been happening largely in the private sector is one of the key challenges we face. We are organizing virtual gatherings for policymakers around the world to give them an overview of how AI works and of the issues and developments we're seeing.

We are hoping to continue our engagement with groups like the National Institute of Standards and Technology and other technical experts in different countries to help develop a core of expertise. It's going to be hard to get everybody to be an AI expert overnight, but key folks should be in a position to make intelligent risk assessments and help influence the direction of regulation, which is clearly going to evolve over the next few years.

Thompson **I've heard a couple of interesting arguments about how AI will change the way geopolitics works and the way AI regulation will work. One is, this is such a new moment and such a powerful force, that it will necessarily allow for a reset and greater global cooperation. The other is that AI regulation and AI development will further harden the lines between the United States and China, and that the world, East and West, will grow further and further apart. What do you think will be most likely to happen?**

Walker Somebody earlier said that Google is in the optimistic middle of the AI debate, and I think that's right. We do see an important role for a global conversation around this, whether that's creating regulatory alignment or setting norms that help influence the direction of research. Norms matter.

To tell a quick story, in the 1980s, genetic researchers got together at Asilomar in California to come up with standards of practice for genetic research. And those standards have continuing vitality today with regard to what kind of research on human beings is appropriate and not appropriate. There's a real interest in the AI community to come up with similar frameworks. Ideally, you'd like those frameworks to be global, to bring as many countries around the world into alignment as you can. It's challenging, given current geopolitical tensions. But it's usually better to have people inside the tent than outside the tent.

The advances we hope to see with AI—like curing cancer or promoting nuclear fusion—are goals that are shared by everybody in the world. If you can have free power that allows you to create clean water for people across Africa, or if you can make dramatic advances in combating diseases that affect everybody, those goals are as broadly shared as the UN's Sustainable Development Goals. So we hope that there will be an incentive for countries around the world to work together to make sure we get AI right.

# OUR FUTURE WITH AI HINGES ON GLOBAL COOPERATION

To harness the power of AI for good, democratic societies need to work together.

By Kay Firth-Butterfield

**Since large language models** began making headlines in the fall of 2022, millions of words have been written about the dangers of AI. Those of us who work on these technologies and their implications have been talking about this since 2014, but now the conversation has gone mainstream—so much so that it risks drowning out necessary discussion of how we might use AI to confront the world's most pressing challenges.

The solution is governance. The AI world needs the public's trust to achieve the benefits of AI, and it won't get there without regulation. We must ensure the safety of the technology as it is used today, known as responsible AI, while looking to the future. More than 60 percent of Americans say they are concerned about AI's negative impacts, according to an AI Policy Institute poll from spring 2023, but without strong laws we'll neither prevent them nor have the tools to deal with them when they arise.

Yet right as we need public trust in AI most, it's falling in democratic societies at an alarming rate. In a recent Luminate survey, 70 percent of British and German voters who identified as understanding AI said they were concerned about its effect on their elections. Similarly, an Axios/Morning Consult poll showed that more than half of Americans believe AI will definitely or probably affect the 2024 election outcome, while more than one-third of them expect their own confidence in the results to be decreased because of AI. More generally, two in five American workers are worried about losing their jobs to AI, according to an American Psychological Association poll, while Gallup found that 79 percent of Americans do not trust companies to self-govern their use of AI. We will never realize technology's economic and positive benefits without addressing these concerns.

However, in 2021, analysis from PwC showed more hopeful results. In a survey of more than 90 sets of ethical AI principles from groups around the world, researchers found that all participants agreed on nine central ethical concepts, including accountability, data privacy, and human agency. Now, governments need to work together to figure out how to make these concepts a reality by building a coalition of the willing across nations that can do the hard work of planning for an uncertain future.

If we continue to simply react to technological advances without thinking ahead, there is a very real risk that we will arrive in 2050 to find that we live in a world that no longer meets our needs as humans. The European Union has thus far chosen a risk-mitigation approach, which addresses current problems but not the essential issue of how humans wish to interact with AI in the future. Individual U.S. states are enacting their own laws, which could slow innovation and make cooperation more difficult.

It is guaranteed that future generations will work beside AI systems and robots. But because AI regulation has been slow to develop, we are currently relying on existing laws to drive best practices. Rather than simply attempting to mitigate harm, we should be creating best practices around what kind of AI we want in the world and how to build it. Only then will we ensure our children live in a human-focused society served by AI, rather than in an AI world occupied by humans.

By working together, democratic governments around the world—together with stakeholders from civil society, academia, and business—can create laws not to address every specific situation (which would be impossible), but instead to outline specific requirements organizations around the world must follow when developing, deploying, and using AI systems. Many who use AI have little understanding of the harmful effects that could result even when they think they are using it for good, so it is up to policymakers to codify priorities like privacy and data security. This would require AI development teams to adopt proven best practices and adhere to all existing and new legislation for creating responsible AI systems from the outset.

The world simply does not have five years to figure out its next steps.

It is tempting to think the domestic governance gap might be filled by international regulation or treaties, but there are risks to this approach: The UN Security Council is often at an impasse even on harm-mitigation topics, much less on ones that require forward thinking. For example, despite calls from the UN secretary-general and smaller nations, we have waited, without result, since 2013 for an agreement on the control of lethal autonomous weapons. If the Security Council is unable to accomplish that kind of policy, it will likely struggle to agree to proactively design an AI policy that is suitable to all stakeholders. The United Nations is expected to name members of a high-level panel on AI, which is a welcome development, but it is unlikely that the creation of an advisory board will result in meaningful regulation as quickly as we need it. The world simply does not have five years to figure out its next steps.

But international cooperation need not run through the UN. Promising suggestions include emulating the model of the European Organization for Nuclear Research, an intergovernmental organization that includes 23 member states, or Gavi, the Vaccine Alliance. Taking that path would ensure that the Global North doesn't have unilateral control over AI technology; it would lead to less inequality and make sure AI serves many different cultures. Governments around the world would come together to envision a positive future for their citizens with AI and create the regulations necessary to achieve it.

Governance is hard. True global governance is harder. Even this faster path will take time, requiring companies designing, developing, and using AI to self-regulate—with full support from their boards and C-suites—in the meantime. But ultimately, collaboration is necessary to build a world in which humanity benefits from AI rather than adapts to it by force. A comprehensive approach is essential, and we must act now.

Kay Firth-Butterfield is the CEO of Good Tech Advisory, the former head of artificial intelligence at the World Economic Forum, and the world's first chief AI ethics officer.

# What's one global challenge AI can bring us closer to solving?





AI promises a new lens through which to approach today's problems. How might it shape our urban landscapes? Could it optimize food systems? What might understanding artificial minds reveal about our own? And crucially: How do we simultaneously harness the technical ingenuity needed to solve these global challenges, while addressing potential risks?

Machine learning could help curb disease, lower emissions, and conserve nature's gifts. Optimized cities could blend sustainability, equity, and human need. AI could even aid law, wielding data to reduce bias. But its power remains only as potent as the aspirations that guide it.

AI's potential extends beyond merely solving problems we can conceive today. It offers a portal into realms of discovery and creativity that we cannot yet imagine. AI could unveil insights about consciousness that reshape mental health. It could spawn new industries and art forms that redefine commerce and culture. It could illuminate solutions only visible when intelligence transcends biological constraints. AI's promise lies not just in solving today's challenges, but in unlocking achievements far beyond the horizons of our vision.

**The evolution of AI** is moving at warp speed. While scientists and philosophers somberly acknowledge its potential perils, they also recognize AI as a tool capable of helping us solve humanity's most vexing problems, including climate change and global hunger. AI is also opening promising avenues in education and health care, as well as in highly specialized areas such as molecular gastronomy and pro-football strategy.

For a broader sense of the technology's possibilities, we asked 18 experts—in the aforementioned spheres and others—to answer the following question:

*Looking to the (perhaps distant) future, what huge problem that seems impossible to resolve today—be it scientific, technological, societal, or otherwise—seems conceivably solvable to you someday, using AI?*

## Here's what they told us.

### Lila Ibrahim

COO, DeepMind

I'm really excited about the way AI might contribute to alleviating the climate crisis. This is one of the most important challenges that we collectively face.

AI can analyze the problems we face, and build better models for prediction and monitoring of extreme weather events. It can optimize current systems and existing infrastructure to reduce energy usage and limit their footprint. And it can help to accelerate scientific breakthroughs, like the development of new, clean forms of energy such as nuclear fusion.

While there are no "silver bullets", AI is already making amazing progress in helping the scientific community to overcome major roadblocks in their work. Ultimately, this has incredible potential to improve the lives of billions of people.

### Anthony Townsend

Urbanist-in-residence, Jacobs-Technion Cornell Institute, Cornell Tech; author, *Ghost Road: Beyond the Driverless Car*

What if, like a crystal ball, an AI-powered urban planner could tell us—without a shadow of a doubt—the inevitable impacts of different city plans on carbon emissions, income inequality, or quality of life? Even better, what if it could do the diplomatic work of soliciting our needs and desires and jumping into the conversation to help resolve our conflicts? We might be able to make the hard choices to build cities that are fair, resilient, healthy, and productive for everyone.

### Andrew Berry

General manager, Cleveland Browns (NFL)

AI will enable algorithmically designed offensive and defensive schemes with unique counter-strategies. Having a full understanding of how all 11 players fit together and interact with the other team will allow for an optimal set of mixed strategies against a given opponent. One of our brilliant strategists recently said, and I agree, that the impact of AI on defensive strategy, where there are more potential ways to move players around—with no restrictions on how you must line up or how the players might move both before and after the snap—will be greater than that of any other phase of the game.

### Cynthia Breazeal

Professor of media arts and sciences, MIT; founder and director, personal robots group, MIT Media Lab

For several years, we've been developing social robots as personalized learning companions for early-childhood education. Children at this age can't read yet; they learn from social interaction and play. The robot plays educational games with them like an engaged peer, supporting the cognitive, social, and emotional aspects of learning. It engages in back-and-forth conversations when they read storybooks together.

This is an emotional, absorbing experience, almost as if the child is playing a game with an intelligent, motivating, peer-like pet. In one video of a robot working with a child, you see a moment where the robot says, "I believe in you!" to support the child's confidence. And after the robot has successfully finished the same exercise, the child turns back to the robot and says, "I believe in you." The robot personalizes, over weeks to months, to support a holistic learning experience for the child.

Effective, affordable, personalized, equitable education for all learners remains an unsolved problem globally. The goal, I believe, isn't to replace but to augment teachers. We might one day live in a world like that of Neal Stephenson's *The Diamond Age*, where each person has a lifelong, holistic, personalized learning companion.

### Cristina Bowerman

Molecular gastronomist and Michelin-starred chef, Glass Hostaria, Rome

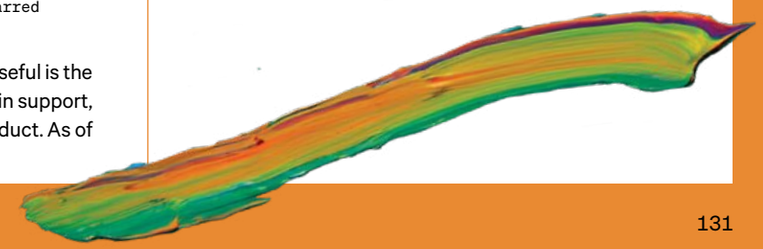
One application which I would find very useful is the capability of AI, with ad hoc instruments in support, to determine the actual shelf life of a product. As of now, we use general references or we base the use of a product on molecular tests—which are impossible to apply on an everyday basis—to ascertain the safety. If AI could be utilized to solve this matter, we could reduce food waste in a massive way. We often throw products away because the Hazard Analysis and Critical Control Point (HACCP) recommends so, or because the expiration date printed impedes us to use it (this is still an issue in most European countries). But we also all know that this is often not the case.

### Renée Cummings

Criminologist and AI ethicist; assistant professor in the practice in data science, University of Virginia School of Data Science

With ethical, equitable, and responsible AI, justice can become a three-dimensional experience: immersive, interactive, and transformative. Extended reality, virtual reality, and generative AI can enable us to design a new cognitive and conceptual landscape for justice—to rethink the resocialization of individuals returning to impacted and disinvested communities. Imagine stepping into life on the outside while still on the inside, using VR to create those new experiences while incarcerated or to do the cognitive behavioral work that will reduce the risk of reoffending.

Responsible AI can transform archaic, static approaches to incarceration and sentencing that have failed society, deprived children, destroyed families, devastated communities, and denied generations resilient legacies of efficacy, progress, sustainability, and wealth. Responsible AI has the capacity to imagine a future of justice that is not built on punishment, profit, vengeance, and racism but designed collaboratively, with communities that have been historically harmed, using augmented collective intelligence. We can deliver a radical transformation of the justice system in which AI can deploy strategies of reparation rather than retribution.



## Marian Croak

Vice President of Responsible AI and Human Centered Technologies, Google

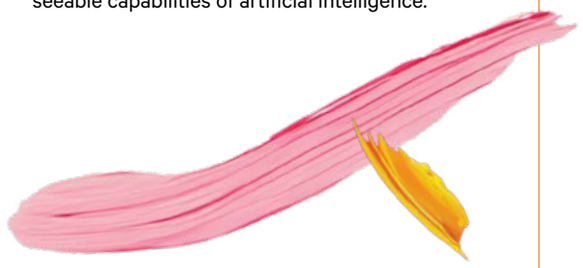
Food security is a huge global problem; it affects over 700 million people each year. Imagine the possibilities AI enables. Scientists could detect a drought before it happens. Farmers could plant scarce crops in advance. Humanitarian agencies could deliver food to populations most in need. From detection to production to distribution, AI can help solve global hunger. This is a problem Google Research is currently working on, and I'm encouraged by our progress so far!

---

## Steven Pinker

Johnstone Family Professor of Psychology, Harvard University; author, *Rationality: What It Is, Why It Seems Scarce, Why It Matters*

"Ask your doctor." How many times have you read this advice? But why should I follow my doctor's advice, just because she happens to be my doctor? She can't know all my particulars, from my genome to my decades of medical history. Nor could she have digested the upshot of the thousands of journal articles relevant to my condition. A robust finding in cognitive psychology is that human experts are outperformed even by simple statistical formulas. Imagine how much better a sophisticated formula would do, one that aggregated genomes, medical history, and test results, and matched it against a meta-analysis of the relevant medical literature. Better still, to present me with a decision tree with the estimated risks and benefits of the pharmacological, surgical, and do-nothing options. We don't have this now, but it's well within the foreseeable capabilities of artificial intelligence.



## Tobias Rees

Anthropologist; Reid Hoffman Professor of Humanities at the New School for Social Research and Fellow of the Canadian Institute for Advanced Research

AImost all companies today build AI so that we can push "automation" into aspects of human life that we thought were beyond the limits of machines, like writing and friendship.

I dream of another reason to build AI: as an opportunity to think beyond the limits of the human mind.

When I say "limits" I do not just mean scale, as if all that's needed is more computation.

Instead, I mean that AI systems build models that are different in their architecture from the architecture of the human mind. The far-ranging consequence is that we could build AI systems that offer us access to spaces of possibility that are outside of ourselves: non-human cognitive and creative landscapes we do not even know how to dream of.

Imagine we could roam these landscapes; imagine the thoughts we could think and the ideas we could find in this outside!

For thousands of years, how we get ideas, where we find them, hasn't changed. The promise of AI is that it could change now.

---

## Miriam Vogel

President & CEO, EqualAI; co-host, *In AI We Trust?* podcast

Raising two daughters, I think about the world we are building for them and their peers across the globe: the challenges we are exacerbating, the tools we are (and are not) providing to equip them to thrive in an AI-fueled economy and society, and the new opportunities opening for them if we can manage the first two considerations adequately. AI can—and will—change our world. It will democratize access to first-class health care, improve climate sustainability, and enable individualized education at scale.

But this can only happen if we ensure our AI systems are built by and for the broadest cross section of our population. In this time of uncertainty about our future, misunderstanding among communities and territories, and too much loneliness and isolation, I am most hopeful AI can serve as a bridge—a translator, educator, medium, and trigger for international consensus-building—to enable the cross-border collaboration, deeper community inclusion, and intelligent information-sharing necessary to respond to current and novel challenges as a global community.

---

## David Salle

Artist

The way I use AI is more of an organizing tool. I wish I could use AI to speed things up in the studio, but it doesn't really work that way, at least not for me.

I'm interested in the "plasticity" of space, and in something I call the transformational grammar of picture building. AI, and algorithms generally, to the extent that I understand them, are keyed into those concepts. AI is adept at a kind of morphology of form; it deconstructs, and re-constructs form as well as composition. AI then applies the power of juxtaposition to a kind of 'morphed form'.

I don't think AI will 'replace' artists, for the simple reason that AI has no particular "reason" to do anything. It doesn't intend anything, has no narrative to convey. It creates an effect—but that's not the same thing as art that holds up to repeated scrutiny.

---

## Lisa Joy

Screenwriter, director, producer; co-creator, *Westworld* (HBO)

One benefit I'm hoping AI has over humans derives from its ability to monitor and regulate complex systems of collective human behaviors in real time. This would give us transparency into major issues like climate change and help us optimize solutions. But it would also impact day-to-day events that have outsize ramifications for society as a whole. For example, an AI regulating traffic flow via networked self-driving cars would increase speed and efficiency while reducing accidents. This centralized network would minimize the need for individuals to own private vehicles. As we posited in *Westworld*'s production design, this in turn would reduce the need for parking structures and parking spaces. These spaces could instead be used as communal green spaces—benefiting the environment and society as a whole. Though AI would no doubt be better than us at these things, we're still in a place where it hasn't been given even a sense of consequence or causality, so how can we trust its lead? There's a bit of the tragic here. It's better equipped to lead than us, but we can't—and shouldn't—trust it to do so. Not yet.

---

## Megan Peters

Associate professor, cognitive sciences, School of Social Sciences, University of California, Irvine

It's my hope that AI—as it becomes more capable of synthesizing new information, engaging in compositional, symbolic reasoning, and self-monitoring its own cognitive processes—may be able to help us toward the goal of understanding consciousness in ways we can't yet imagine. Testing for the presence versus absence of consciousness is not just an interesting thought experiment or worry for some future Skynet-like AI: It's a problem here and now.

Believing that any system, biological or artificial, has reached a threshold for consciousness should ideally obligate us to treat it humanely—from the laboratory to the hospital bed. What if we're keeping alive a loved one who's in a coma? If we cannot definitively say that their consciousness has abandoned them, we have quite different feelings and responsibilities when it comes to deciding whether to remove life support.

An ongoing discussion about laboratory ethics right now concerns higher cephalopods (octopuses, squid, cuttlefish, etc.). But if we could say unequivocally that octopuses feel pain, the implications for ethical treatment would go beyond the laboratory to farming practices and choices

about ethical food production. What if we could say the same thing, just as unequivocally, about a salmon? A zebra fish? A bumblebee?

Some researchers, rather than carry out research on non-human animals, use stem cells to create so-called organoids—collections of thousands of brain cells or more, hooked up to each other in networks. These brainlike structures, living in dishes, can learn to play the video game Pong. Are these networks conscious—and if so, what do we owe them once we've finished using them in our experiments?

I don't know, yet, how we'll convincingly detect consciousness in any system—biological or artificial—besides, well, myself. But I think that AI, as it evolves, has the potential to help us find the answer by helping us understand what makes humans conscious in the first place.

## Christopher Wood

Executive director, LGBT Tech

The LGBTQ+ community has historically faced persecution, erasure, a severe lack of visibility, and much more. Tomorrow's AI tools need to be developed in collaboration with the LGBTQ+ community and designed with fairness and transparency to meet our needs and respect our human rights.

AI-based bias-detection tools to monitor for discriminatory or hateful content—on media (of all forms), online platforms, and other public domains—will better allow leaders of our community, along with those platforms, to take appropriate actions such as flagging, reporting, or education.

AI will play a major role in improving health-care outcomes and health-care personalization. It will help medical professionals access training programs and tools that improve their cultural competency. It will function as a crisis-intervention resource offering first-line, preliminary mental-health evaluations to our most vulnerable members (including our transgender and nonbinary community); this initial assessment will lead to closely personalized referrals to inclusive and sensitive care.

I'm very excited about future possibilities around AI and the LGBTQ+ community, but I think we do need to proceed with caution, as AI models are only as good as the data they are trained on.

## Yossi Matias

Vice President, Engineering and Research, Google; Lead of AI for Health, Sustainability, and Education

I'm optimistic that in the future AI could help empower every person and every system to be more capable and effective in almost every aspect, and will help solve many societal problems we are currently facing. I'm particularly excited about the opportunities for societal impact at global scale in healthcare, sustainability, and education.

AI will help obtain transformative impact, advance science and solve many problems that are now considered difficult or impossible. We've already seen this with protein folding and flood forecasting. AI will help make large scale systems more effective: from more accessible healthcare to more carbon-efficient transportation. AI will help empower and scale capabilities that are currently scarce, and make them more equitable and personalized. As generative AI becomes more reliable, it will empower and scale human knowledge, effective decision-making, and creativity.

## Ulrich Blum

Co-head, Zaha Hadid Architects Analytics & Insights; professor of architectural design, Münster School of Architecture

AI has the potential to address the chronic underutilization of buildings—which exacerbates crises from climate change to homelessness. Single-use buildings can be likened to single-use plastic cutlery or fast fashion: They serve just one single function, and they are designed and constructed hastily, without much consideration for the long term.

But an AI-driven building can reconfigure itself (or be reconfigured). For instance, users of an AI-driven workplace might arrive on a given morning to discover more conference rooms or an expanded space for a planned gathering. An AI-driven building could gather feedback about its occupants' work patterns, varying moods, or preferred ambience—whether colors or temperature or light settings—and become genuinely conducive to productivity.

During the coronavirus pandemic, many of us realized that our homes were not ideal for work and our offices were ill-suited for the activities we engaged in there. Looking ahead, our lifestyles could undergo even more fundamental transformations as more individuals choose to spend the majority of their time within the metaverse.

In the digital world, the absence of gravity, construction costs, fire compartmentation, and other worldly constraints are an architect's dream. I can change my neighbors whenever I want and choose the colleague I want to sit next to. Or I can decide to have the AI choose the coworkers who are best for my work and well-being and place them close to me in space. I can live in a virtual house on a virtual beach property, mountain top, distant planet, fluffy Candyland—whatever I desire.

## Daniela Rus

Andrew (1956) and Erna Viterbi Professor of Electrical Engineering and Computer Science and director of the Computer Science and Artificial Intelligence Laboratory, MIT

Robots have already become our partners in industrial and domestic settings. They work side by side with people in assembly plants, building cars and many other goods. They help surgeons perform difficult procedures, improving outcomes and reducing scars. They mow our lawns, vacuum our floors, and even milk our cows. In a few years, I believe they are going to touch most parts of our lives.

But most of today's robots have been built as point solutions, meant to solve a specific problem in a specific way. Our challenge as researchers is to take those point solutions and turn them into more general solutions for locomotion, manipulation, and interactions with the world and other machines.

Humans and machines don't speak the same language. Imagine telling your automated car, "Park in the shady spot under the tree." The machine inside the vehicle needs to look at its image of the surrounding area and find the right spot. It needs to understand the words tree and shade and translate them into bits and bytes. It needs to figure out that you mean the tree on the right, not the one on the other side of the street. When you consider all the specialized language needed in a variety of manufacturing contexts—as well as the sheer number of possibilities—you can begin to see why building these robots will be challenging.

Recent advancement in large language models will enable robots to form teams with people; your average manufacturing employee will be able to talk and partner with them to perform tasks. These robots will be able to reconfigure assembly lines quickly and create tools that allow one-of-a-kind designs. They will help to bring about a new era of customized production that happens in factories across town, instead of across the ocean.

## will.i.am

Musician and entrepreneur

AI will help us solve the educational gap, and it will give everyone access to the best Harvard/Yale/MIT/Oxford/Stanford-level professor—one that's accessible at any given time and that understands how they learn. AI is enabling education tailor-made for each student in the new age of FOREVER LEARNING.

**USE OF AI**

While we wanted to avoid using purely AI-generated imagery to accompany our stories, all of our contributing illustrators were asked, as part of their brief, to explore an AI tool—be it for word association, concept ideation, generative iterations as a starting point for sketches, textural finishes for final outputs, or otherwise—at some point in their process.

**TYPEFACES**

*Dialogues* is typeset in *Coign* (Colophon Foundry), *Affairs* (SM Foundry), as well as *Calibre* and *Söhne Mono* (Klim Type Foundry).

**PRINTED IN THE U.S.A.**

This magazine is a certified carbon-neutral publication, printed by Green Earth Enterprise using vegetable-based ink on recycled paper. To offset the environmental impact of this print production, Green Earth Enterprise engages with local firms to plant useful trees in Central America.



Please keep *Dialogues* for your library or pass along to a friend.

9
Colorful paint blob yellow white ochre · 42 variations

10
Colorful paint stroke yellow orange ochre · 18 variations

11
Colorful blob of pink and white, lots of texture, abstract expressionism · 3 variations

12
Colorful paint stroke, white pink maroon, abstract expressionist · 12 variations

13
Colorful paint stroke abstract expressionist white pink · 36 variations

14
Colorful paint stroke, red pink maroon white, high resolution · 3 variations