DIALOGUES

On AI, society, and what comes next

Aashima Gupta · Alondra Nelson · Anca Dragan · Ayanna Howard · Blaise Agüera y Arcas
Brigitte Hoyter Gosselink · Charina Chou · Chris Wolverton · Connor Coley · Daan van Esch · Dani Clode
Dimitri Kanevsky · Ekin Dogus Cubuk · Hartmut Neven · James Manyika · Jassim Haji · John Jumper
Justin Tranter · Kate Brandt · Kent Walker · Larry Lessig · Louis Bell · Lyor Cohen
Martin Ford · Max Jaderberg · Mira Lane · Maya Kulycky · Neal Mohan · Nicholas Thompson
Paolo Benanti · Paula Goldman · Percy Liang · Phil Venables · Rahaf Harfoush · Richard Zhang
Saška Mojsilović · Scott Aaronson · Shane Legg · Strive Masiyika · Virginia Dignum · Yossi Matias

Google

DIALOGUES

On AI, society, and what comes next

ATLANTIC RE:THINK

EDITORIAL

Liz Haq
Head of Editorial

Gabriel Muller
Co-Managing Editor

Michael Agger
Co-Managing Editor

Jaime Brockway
Copy Editor

Martin Ford
Technical Advisor

DESIGN

Drew Campbell
Creative Director

Laura Scofield
Design Director

Hanah Slotterback
Art Director

PRODUCTION

Monica Schmelzer
Senior Director, Account Management

Elena Mehas
VP, Strategic Initiatives

Nicholas Thompson
CEO, *The Atlantic*

CONTRIBUTORS

Alasdair Lane

Anya Kamenetz

Clarissa Wei

Craig Marks

Daniel Oberhaus

Haimy Assefa

Kellie Schmitt

Mark Esposito

Matthew Hutson

Maya Kosoff

Sneha Mehta

Stayton Bonner

Terrence Russell

VISUALS

Adam Ferriss

Anna Lucia

Cayce Clifford

Chris Burnett

Daniel Liévano

Hoi Chan

Ibrahim Rayintakath

Irene Suosalo

La Tigre

Linda Dounia

Optics Lab

Ori Toor

Oriana Fenwick

Petra Pétterfly

Raven Jiang

Rida Abbasi

Ryan Carl

Uli Knörzer

Yann Kebbi

Google + *Atlantic Re:think*

Produced by Atlantic Re:think, the creative marketing studio at *The Atlantic*—independently of *The Atlantic*'s editorial staff—in partnership with Google.

For more, visit **theatlantic.com/rethink**.

DIALOGUES

On AI, society, and what comes next

Once Upon A Garden is a speculative archive reconstructing endangered West African flora. Created by artist Linda Dounia from 2021 to 2024, the project imagines 50 species using iterative AI models. The cover for *Dialogues*' second issue features an image from the project's Fifth Generation, produced in 2024. Entitled "Synthetic Rot," it emphasizes entropy and randomness, mirroring nature's chaos and exposing technology's limits in supplanting organic reality.

For us, Dounia's machine-generated exploration highlights the messy, adaptive nature of evolution itself. Like *Once Upon A Garden*'s generations, each story in this issue shows breakthroughs born out of adaptation, resilience and, perhaps most critically, change.

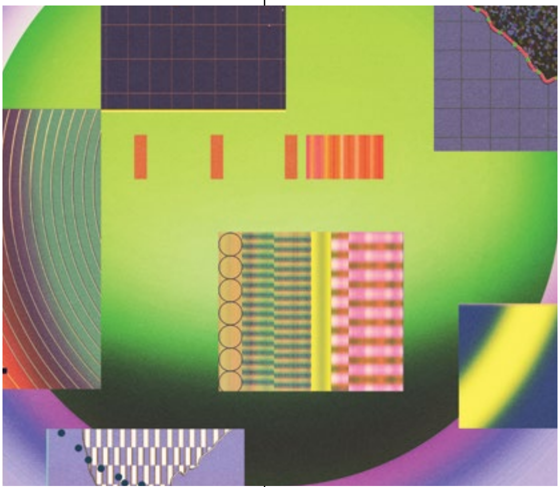
Technological revolutions happen quietly. When you look beneath the hype and headlines in the media, you find curious people adapting new technology to the challenges in their daily lives. This is what's happening with artificial intelligence. It's the doctor who uses AI to help provide a diagnosis faster. It's the teacher who uses a large language model (LLM) to help develop a history curriculum. It's the scientist analyzing lab results with the assistance of a custom AI model.

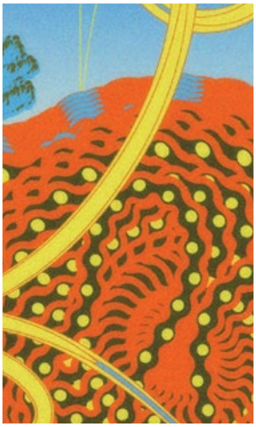
As we interact more with this technology, our understanding of it is deepening. The natural language processing displayed by chatbots is only one manifestation of the capabilities of machine learning and neural networks. AI is best understood as a general-purpose technology, like the steam engine or the internet, due to its wide-ranging impact. The ability to recognize patterns, respond to new information, and improve performance makes it strikingly versatile. AI can help predict the molecular structure of proteins. It can improve the analysis of a storm system and provide expedited flood warnings. It can help coordinate traffic patterns across an entire city to maximize fuel efficiency.

The past 12 months have seen advancements in problem-solving and language understanding that give AI powerful new ways to assist us. A year ago, the first issue of this magazine asked fundamental questions about AI. This issue begins with a grounding in the current state of the art: What are the ideas, methods, and inspirations that are pushing this technology forward?

From this foundation, we consider AI's remarkable influence on the activities central to our culture: how we learn, how we practice medicine, how we create art and song. We examine how AI can help reduce inequity by increasing accessibility and, more broadly, by providing opportunities for up-and-coming generations globally. We then turn to the area where, perhaps, AI has had the most astonishing impact so far: scientific discovery. AI is helping us to map the brain. It's analyzing new materials that might revolutionize electric vehicles. It's predicting the structure of proteins that will help us better understand human health. And it has only just begun.

Within the context of these dizzying breakthroughs, new big questions about AI arise, and are tackled in the final chapter of this issue. How do we guide this technology so that it contributes to human flourishing? How do we improve the abilities of AI responsibly, so that it can assist in building the sustainable world we all need? What becomes clearer each day is how AI will be a great test of our humanity, of our ability to cooperate. It will demand the best of us and, if we get it right, might help spur a level of innovation and shared progress we can't yet imagine.





What AI Asks of Us

The capabilities and analytical insights of artificial intelligence have already improved our world. But they will also test our ability to cooperate, empathize, and think strategically.

By Nicholas Thompson

In his role at Google as the senior vice president of Research, Technology & Society, James Manyika brings together people across disciplines and seeks to understand how we can guide artificial intelligence in a responsible and equitable way. Since we spoke a year ago, there have been many significant improvements in AI, and a new range of questions to confront and try to solve, collectively.

Our conversation has been edited for length and clarity.

Nicholas Thompson **When we spoke for the first issue of *Dialogues*, we discussed the societal challenges brought on by rapid developments in artificial intelligence. What have been the most important advances in AI this year?**

James Manyika There've been some very interesting technical developments that signify where we're going. The models we're now building are natively multimodal. In other words, they are built to understand text, images, video and, quite frankly, anything that can be tokenized. Whereas, before, we were doing that after the fact. The multimodality goes in both ways—that is, you can generate images, video, and more from text, or go the other way, that is, input or prompt with images to generate text. A lot of the applications you will start to see will take advantage of native multimodality.

The other thing that's happened is long context. Think of context as a form of memory, such as the ability of an AI to analyze, do inference on, and retain a large set of documents. The things that you were able to do with AI were limited by how big the context window was. Most systems could handle, at most, 100,000 tokens or words. We had a big breakthrough, so now, we can actually do very, very long context. We've put out a 2 million token context window. You can throw in a hundred documents, you can put in video, you can have a long conversation with the AI. That's a big deal.

The other thing that is really quite important are so-called agentive capabilities. We're looking at opportunities to use AI to take actions and cooperate with a user, in real time, both in the real world and in digital environments. So, this agentive capability takes us away from thinking about AI only as content-generation systems, and more about AI as systems that help us take actions in the world.

Thompson **The very long context window will be helpful to me, as I often reach the limits with the number and complexity of documents I'm feeding the AI. When will we have a billion-token context window? Will this just continue to scale?**

Manyika With long context, there's two things that are difficult. Building a context window that just goes on forever is actually a hard computer science problem, because, remember, you're doing inference within that context window. The second issue is, How well can you analyze the context window? There've been lots of technical papers that showed,

> "Agentive capability takes us away from thinking about AI only as content-generation systems, and more about AI as systems that help us take actions in the world."
>
> **James Manyika**, SVP of Research, Technology & Society at Google

yeah, sure, you could do a long context window, but you're still going to have another problem, which is the performance, including that inference and reasoning will degrade. More specifically, your ability to analyze and work with what is in the context window will be good at the beginning, drop in the middle, and pick up at the back end. What we've been able to do is have consistent performance throughout the entire context window. You see this with some of the so-called needle-in-a-haystack tests, which test the ability to find and do inference on details of what is in the context window. This is how, in a product like NotebookLM, which uses long context, we are able to provide citations that draw from the source documents you put in NotebookLM's long context.

But to your question, can the length of the context grow forever? It could, but here's the other challenge. The costs associated with it. We've thought of the compute costs for these systems as being the cost to train the model, but there's another cost to consider, which is the inference cost. Now, the inference cost is relatively low if you work in a very small context window, because what you're doing with inference is asking the system to do stuff in real time. The longer you make your context window, the larger those inference costs become. They may actually come to swamp even the training costs, especially if you're trying to do inference with, say, 1 billion users at the same time using the long context capabilities.

All the breakthroughs and progress I've mentioned are, mostly, to do with the capabilities of the system, whether it's long context or multimodality or agentive capabilities. The other thing that's been going on quietly in the background is big efficiency gains, because, as we know, this technology is very compute-intensive, and, of course, that's a concern because of the implications for energy use and costs. So, these efficiency gains, while they may not show out as things that you as a user will experience, trust me, in the background, they are important. There have been important improvements in the last year, to make the performance—including for training and inference, in chips we use, like TPUs—much better, cheaper, and less resource-intensive.

Thompson **Do you think AI will be a net positive for climate change, because it allows us to find efficiencies such as new methods of carbon sequestration? Or will it be a net negative, because of all the energy required to train models and cool data centers and power these systems?**

Manyika I think it will be net positive. No question in my mind.

There are three categories where AI is already helping the climate. One is understanding climate science itself. We're making huge progress. We've developed climate-modeling and weather-forecasting tools, such as GraphCast and NeuralGCM, that are state of the art and capable of much better performance than traditional approaches. We've published papers in *Nature* and *Science* (and other organizations have too) about our ability to build faster and more accurate climate models. For example, NeuralGCM is the first published AI model capable of producing ensemble weather forecasts that rival the best current physics-based models. It has the potential to simulate over 70,000 days of the atmosphere in the time it would take a high-resolution physics-based model to simulate only 19 days.

We also have examples of how AI can help with mitigation. We've done things such as our Project Green Light, where we're able to reduce emissions by providing city governments with AI tools to optimize traffic lights and traffic flows in cities. We're now doing this in more than a dozen cities around the world. We also have the example of our work on using AI to help mitigate the effect of contrails, which contribute about 35 percent of aviation's climate impact. We developed an AI model that identifies areas where airplane contrails are likely to form, allowing for flight rerouting to reduce the climate impact of air travel. Tests flights in partnership with American Airlines showed a 54 percent reduction in contrails with minimal fuel increase. It's early days, but these are a few examples on the mitigation side.

Then, you've got lots of things on the adaptation side. We now have more extreme flooding events, more wildfire events. We started flood-prediction work, which began as a small pilot in Bangladesh a year and a half ago, and it worked. We now do that in over 80 countries and counting. And we're doing wildfire boundary work in over 20 countries. So, all of those things are on the benefit side.

But, of course, you've got the impact of AI itself. The compute intensity and corresponding energy use will improve because of the innovations and the advances on making these models, the compute they use, and data centers themselves more efficient. So, I'm very confident to say it's a net positive.

Thompson **What else are you excited for with scientific progress and AI? This has been a major focus of Google DeepMind and of Google. What will we see in the next year?**

Manyika Oh, there's so much. What we did most recently through our team at Google DeepMind with AlphaFold is that we extended the ability to predict protein structures to life's other biomolecules, the foundations of DNA, RNA, and ligands, and also the interactions. These benefits of the AlphaFold program are not just theoretical. We have 2.2 million scientists accessing these datasets, and these are scientists in more than 190 countries. While AlphaFold deservedly gets a lot of attention, the related work on AlphaMissense is worth noting. AlphaMissense categorized 89 percent of all 71 million possible missense variants—these are single letter substitutions in DNA—as either likely pathogenic or likely benign. By contrast, only 0.1 percent have been confirmed by human experts. We've made AlphaMissense's predictions freely available, and scientists have used this to accelerate their work—for example, to help unpick the genetic drivers of epilepsy.

We've also had a research team working on connectomics. This is to understand how the brain looks and works at the synaptic level. And so, we've been doing this work together with the Lichtman Lab at Harvard. And we actually published the first kind of synaptic-level map of a tiny piece of the human cortex, which is extraordinary. The full dataset, including AI-generated annotations for each cell, has been made publicly available.

There are more examples of AI enabling progress by scientists in a variety of fields, including in medicine, material science, chemistry, physics, and mathematics. The progress in medical diagnostics is notable, especially in places where many often go undiagnosed for costs and other resource reasons.

The larger point is we've gone from AI advancing science to that science having a real impact on the world. I mentioned flood forecasting before. That was a known problem that had been unsolved for decades. Everyone knew that if we could predict floods with five, seven days of advance notice, you'd save lives. That's been known for a while, but it's a hard scientific problem to figure out how to do that. We were able to develop an AI model that achieves reliability in predicting extreme riverine events in ungauged watersheds at up to a five-day lead time, with reliability matching or exceeding that of nowcasts.

Thompson **Let's shift to the dangers of AI. What aspects of AI worry you more now than they did a year ago?**

Manyika The implications of these agentive capabilities. Before, you could say, well, AI is about generating outputs that we can look at and decide what to do with. Now, systems are able to take actions. On the one hand, that's extraordinarily helpful. You can imagine the examples of agentive uses, where it fills out a spreadsheet for me, researches possible stays on my vacation, and checks their availability, and more—that's great, that's helpful. But you could also imagine agentive uses that could be misused. We have to think about the risks, the ethical implications of agents. Earlier this year, we published what many consider a landmark study. It's a 200-page report that thinks through the ethical issues associated with agents and agentive capabilities.

The report leads to the question you and I have discussed: the alignment question. That becomes even bigger when you've got agentive capability. How do you make sure that in the process of trying to fulfill some high-level goal that you have, the system doesn't create and take sub-actions that may be misaligned in an attempt to achieve the ultimate goal? One way, of course, would be to have the system show its work, subgoals, and steps along the way. These are some of the things we have to think through.

Thompson **Do you feel like we're making enough progress on the issue of explainability and understanding why these systems make the choices they do, because the more we learn about that, the better we'll be able to align them?**

Manyika We can always make more progress on explainability. But I wouldn't hold up explainability as the way to get to alignment, because that presumes that we, as humans, are always going to be able to hold everything in our heads that these systems can do, and therefore, we can align them on the basis that we understand every single thing that they're doing. I think that problem will outstrip us.

So I don't think explainability is the way to alignment. I think explainability is important for its own sake, so that people understand what systems are doing and why outputs are what they are or why actions are what they are. But I don't see that as the key to solving alignment.

Thompson **So, what is the key to alignment?**

Manyika In my mind, there has always been two parts to it. The first one starts with us as humans, as a society. What does alignment mean to you, Nicholas Thompson? Do you want a system that does what you say, or do you want it to learn what you want based on what you actually do? Or do you want it to do things that are best for you? Each of these can lead to quite different sorts of alignment for you.

Next, what does even alignment mean—aligned with what and with whom, and when? This further opens up an even more complex world where the normative questions are very large and complex. What does alignment mean in this context? What does it mean in this community, in this country, in this culture, in this circumstance, in this sector? There's a large universe here of age-old questions for us as society and as humans to think through.

Then, you get to the second part, that if we decide what alignment is, then the challenge becomes, How do we technically make the systems reflect that? Now, we and many others are working hard on the technical, scientific part of the question. But that doesn't obviate the need to solve the first set of questions.

Thompson **You have a front-row seat on how different countries, different cultures are responding to AI. In what countries outside of the United States and China are you seeing the most interesting and beneficial uses of this technology?**

Manyika Well, I should start by saying it's been fascinating for me personally, because I've been co-chairing the UN's High-Level Advisory Body on AI—with 39 members from 33 countries, with lots of input from a thousand others, including experts beyond that—and that's been a fascinating window into the diverse views, examples, and attitudes around the world.

The places that have limited resources—in terms of doctors, infrastructure, existing fully functional health systems—are enthusiastic and are exploring and in some cases finding extraordinary use cases for AI. While all this is still early, there are some emerging examples of places that don't have teachers, that don't have access to good libraries or good facilities, that are finding these systems very, very useful. It's probably one of the reasons why, generally, when we've looked at attitudes toward AI, the Global South tends to be more positively inclined toward the beneficial possibilities of AI than the advanced economies because they see AI as a way to bridge limitations and solve societal challenges.

Now, that's not to say that those communities internationally don't have concerns and gaps that they want to address. They do. And those gaps and concerns tend to be, "Hey, we want to be involved in the development, use, and governance of AI, not only for its benefits but to also have it reflect our needs, cultures, and values. And by the way, we have capacity and capability gaps and limitations where we run the risk that the historical digital divide also becomes an AI divide."

What's also been interesting is how many of the use cases that those countries have are not all that different from communities within our country that have limited resources, too. So, I've gone to schools here in the U.S. where the kids are promised that, someday, somebody is going to come teach them how to code, and no one ever showed up. And those kids are now using AI systems to draft software code.

Thompson **You grew up in Southern Africa. I spent a fair amount of time in my twenties in West Africa, specifically Ghana, writing about the effect the internet had on the local communities. What is the most interesting use case you've seen in Africa?**

Manyika Oh, languages. We just added over 100 new languages to Google Translate, and half of the ones we added in the last three months were actually African languages. These are languages spoken by more than half a billion people that had not been represented before.

Then you've got other use cases where I don't know if people would recognize them as, "Hey, AI did this." For example, as populations shift in Africa, millions of buildings aren't on maps, and so occupants risk missing out on the basics like electricity, health care, and mail delivery. Our Google Research team in Ghana used AI to massively uplevel the Open Buildings dataset—transforming blurry, low-res satellite imagery to useful data so that partners from NGOs and crisis response organizations can see how the areas they serve are changing over time.

In another recent example, my colleagues, in collaboration with the Centre for Infectious Disease Research in Zambia, used AI-enabled tools for TB screenings and showed their effectiveness in a yearlong study of a couple of thousand people that was recently published in the *New England Journal of Medicine AI*. This is a big deal because in many countries with limited resources, between 30 and 40 percent of people with TB go undiagnosed.

But those are indirectly beneficial things. So, you've got a lot of these direct benefits that people experienced

> "This is a global technology. I'm hoping that we get to a place where there's some degree of collaboration, coordination, harmonization of different approaches."

James Manyika, SVP of Research, Technology & Society at Google

directly, but also these indirect things that are improving something in society.

Thompson **What are some of these overarching societal benefits of AI that we might see in the near future?**

Manyika What I'm hoping for next year is follow-through on the recommendations we made on the UN's High-Level Advisory Body on AI and help to ensure that everybody benefits from the possibilities of AI. Benefit in two ways, actually: Be able to participate more in the development of AI and the building of it, but also in the benefits, outputs, and use of it. And can we bridge some of these capability gaps, so that the digital divide that already exists does not also turn into the AI divide? Can the private sector, governments, and others collaborate to address these capability gaps, especially in communities and countries that don't have access?

So, I'm hoping we'll see more progress on those things. We at Google are trying to do our part in several areas, like investing in infrastructure. We're building data centers around the world. We've been connecting with fiber-optic cables—for example, connecting Africa with Asia, Africa with Europe, Africa with itself—in order to bring infrastructure to these places. We're working to enable the rich diversity of cultural and linguistic diversity. In just three or so years, we've gone from Google Translate being able to handle 30-something languages to almost 250 languages. In fact, we're aiming to get to a thousand languages pretty soon. Next, investing in people. We've been investing a lot in training and giving skills to people. To date, we have trained over 100 million people on digital skills globally, and we recently announced another commitment of $120 million to make AI education and training available throughout the world. There is clearly more that is needed, but I think all of us—companies, governments, and other stakeholders—have to come together to really address these gaps so that we make sure everybody can participate and everybody can also benefit from what, I think, will be an extraordinary period of bounty from AI.

Thompson **And how will we make that progress? What kind of leadership is required?**

Manyika We need to continue the work of responsible regulation. Regulation should always do two things at the same time. It should address the risks and dangers and the things we don't want, and it should also enable the things that we want. It was quite fascinating in our UN High-Level Advisory Body work how most of us tend to think about the risk of AI as misapplication and misuse. A lot of the people in the Global South said, "Please, add a third risk: *missed* use." They said, "We live in communities where we actually have the risk of missed uses, where we could have used AI to address this problem but didn't, and we waited for other solutions that never showed up."

So, I think this speaks to this idea that regulation should always do both things. Obviously, address the things we don't want, like misuses, but also, enable the things we want. So, that's my over-arching view. I hope that part of what happens over the next year is a more balanced focus on both these aspects. I think you see it in some countries that are working on solving these two things, but I also hope we solve for something else, that we don't end up with this mishmash or patchwork of very different rules and regulations around the world.

This is a global technology. I'm hoping that we get to a place where there's some degree of collaboration, coordination, harmonization of different approaches as much as possible, including standards, including safety. I think that's going to be quite important.

We now have an AI ecosystem and value chain. By ecosystem, I mean that we have people developing AI, people deploying AI, people using AI. So regulation should take into account that actions and approaches that impact both benefits and risks are going to happen all the way through that chain. Right? All the way from the developers to the users, and here users range from from individuals, to companies, to governments. So, we should think about the whole chain. If what we're trying to do is to make sure this technology benefits all of humanity, we have to think about the whole chain, the whole ecosystem.

CHAPTER

1

Where are we now?

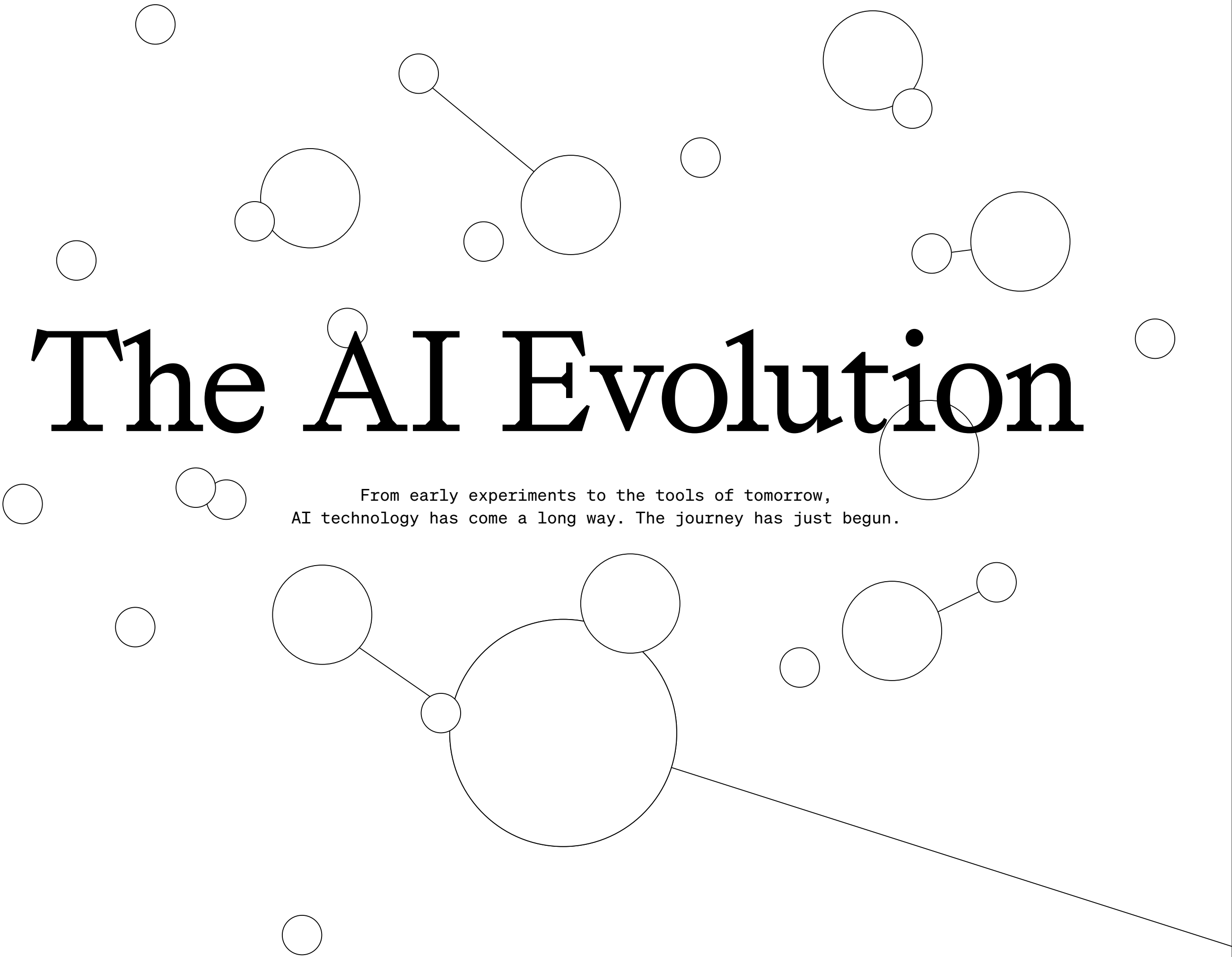
Intro

AI has evolved from a niche pursuit to a force reshaping our world. Still, for all its ubiquity, it remains poorly understood.

The AI revolution is animated by systems of staggering complexity. Transformer models, once confined to arcane discussion among researchers, now power the chatbots and language models that millions of knowledge workers use every day.

As we push the boundaries of what's possible, larger questions loom. The pursuit of artificial general intelligence—AI that matches or exceeds human cognitive abilities—drives much of the field's ambition. As AI models grow larger and more complex, they require enormous amounts of energy and computing power. How can we develop more efficient architectures and training methods to make advanced AI systems sustainable and accessible?

The stories in this section provide a clear-eyed assessment of where AI stands today. In doing so, they lay the foundation for understanding the profound ways in which this technology is altering our future.



The AI Evolution

From early experiments to the tools of tomorrow,
AI technology has come a long way. The journey has just begun.

By Terrence Russell

→
Visuals by Optics Lab

Since the dawn of machine learning (ML) in the 1950s, artificial intelligence has transformed from an abstract idea into a tool affecting daily life. Its origins lie in the basic building blocks of machine learning and neural networks, from which it gradually evolved into the complex, interconnected systems we see today. This evolution didn't happen overnight; it's the result of decades of research, experimentation, and breakthroughs in computing. Here's a look at how AI grew from predicting games of checkers to the vastly complex interworking systems capable of complex reasoning, diagnosing diseases, and forecasting future events within seconds.

Pt. 1

Building Blocks

Machine Learning

The First Step Toward Intelligence

Machine learning has been around longer than video games, email, and even personal computers. The concept dates back to the 1950s, when researchers first tried to create programs that could "learn" from simple data collections. One early pioneer, Arthur Samuel, built a program that could teach itself to play checkers—a rudimentary example of machine learning. The core principle was simple: Feed a system enough data, and it will start recognizing patterns and making predictions.

Researchers fed these models more data and trained them to contextualize using algorithms—essentially, mathematical instructions designed to find meaningful insights. These models were built to get better at predicting outcomes the more data they analyzed. The earliest examples were limited in scope; linear regression and decision tree algorithms were just the start. While the core principle of AI's keystone capability remains the same, ML models have grown more capable, accurate, and faster at handling vast amounts of ever more complex data.

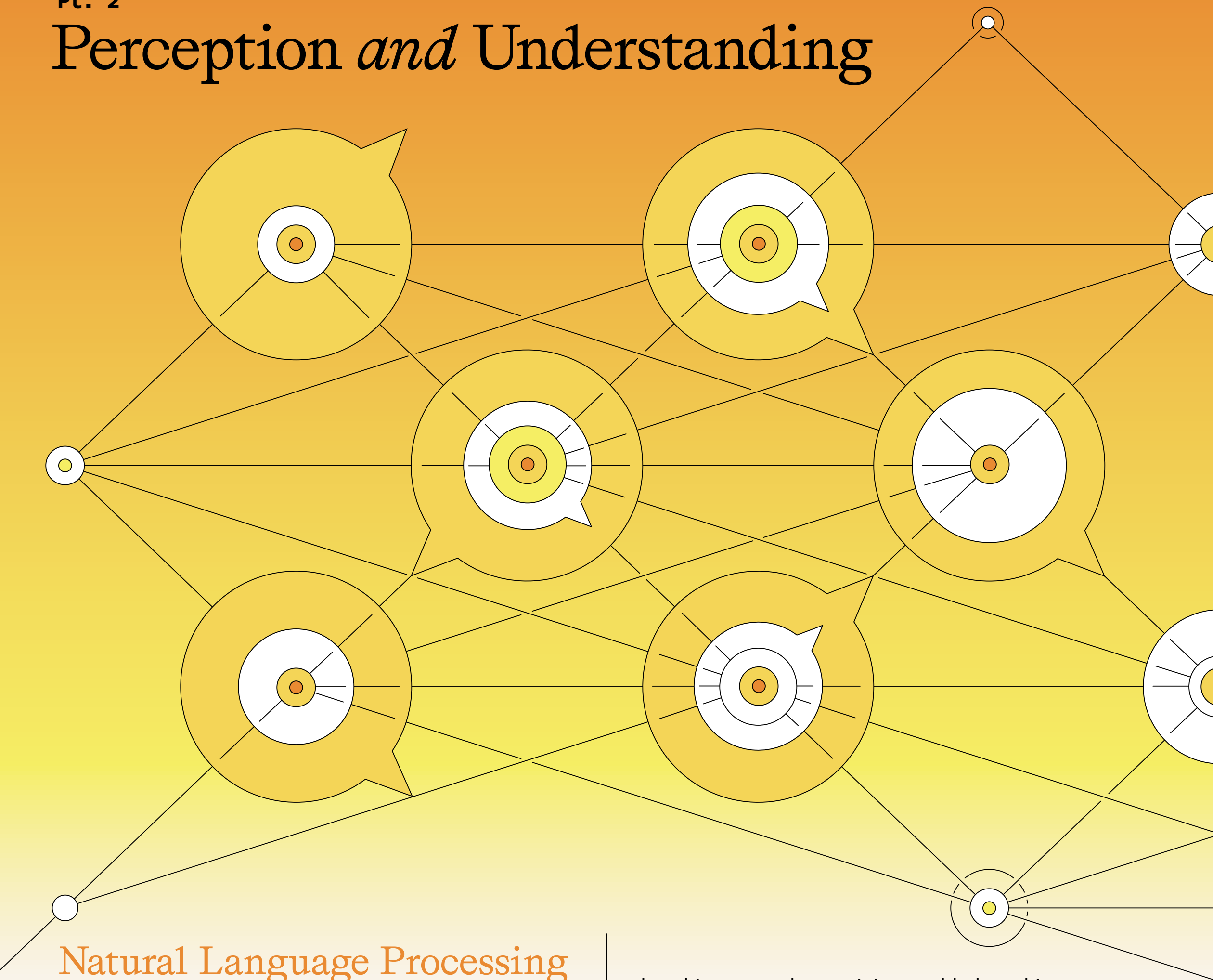
Neural Networks

Inspired by the Human Brain

In the 1940s, neurophysiologist Warren McCulloch and mathematician Walter Pitts proposed a model of the brain's functions based on signals and connections between neurons, dendrites, and synapses. Their ultimate objective was to develop an artificial system that mimics how human neurons process information. Later, this model became the foundation for what we now call neural networks. Neural networks took this inspiration and modeled computer signals and connections in a similar way, becoming a foundational breakthrough that would transform AI.

Like the human brain, neural networks consist of layers of interconnected nodes that process and transmit information through the network. These nodes work together to identify patterns in data, gradually improving their accuracy through repeated exposure to examples. In the 1980s, the concept of backpropagation—an algorithmic method of refining the accuracy of these networks—brought them back into the spotlight. As the name suggests, the process works backward (from the result to the initial input) to identify errors and minimize mistakes going forward. Similar to how moments of reflection or review can help humans make smarter decisions, incorporating backpropagation set the stage for the reasoning that neural networks are capable of today—such as sifting through images to recognize objects and analyzing speech patterns with incredible accuracy.

Perception *and* Understanding



Natural Language Processing

Machines Get the Gift of Language

With the core pieces of AI cognition taking shape, researchers were already exploring whether machines could understand human language. Language is messy, full of ambiguities, idioms, and shifting meanings. Early attempts at natural language processing (NLP) relied on simple rules: for example, rigid if-then statements that codified only one input to one output. This initial preprogrammed approach could produce only text responses to specific prompts, which often resulted in stiff, rule-based communication that didn't capture the diversity of human language. Ultimately, this limited its scalability compared to modern ML models.

Next came statistical NLP, which essentially taught machines to break down text or speech automatically. Powered by machine learning, statistical NLP predicts the most likely meaning of text based on patterns observed in large amounts of data. Instead of sticking to preprogrammed rules, this approach to training enabled machines to grasp linguistic elements such as nouns, verbs, and adjectives. It converts words and grammar into numbers and uses math to process language. Early tools such as spellcheckers and T9 texting were built using statistical NLP.

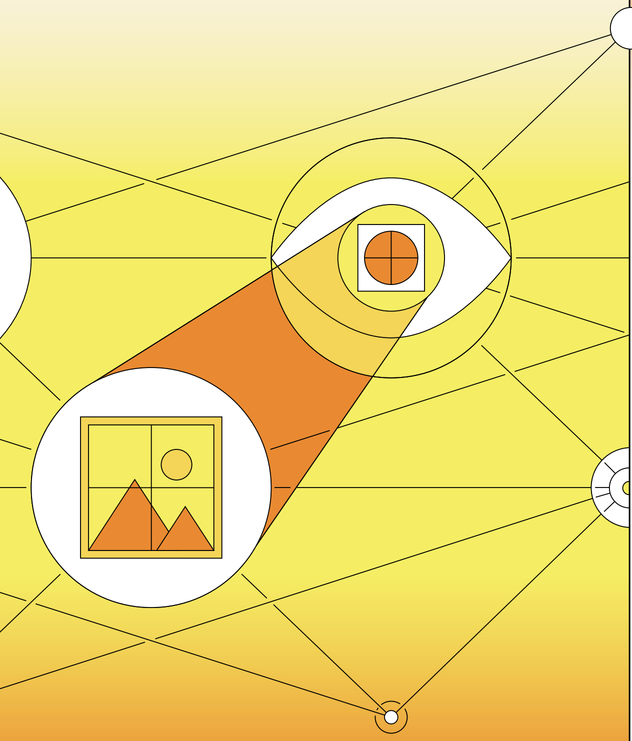
A breakthrough came when researchers took a new approach, setting aside traditional linguistic theory in favor of letting deep-learning models discover patterns directly from vast amounts of raw text data. The researchers ran raw text and audio through these neural networks, and over time, the models were able to recognize nuanced patterns in language without needing every rule spelled out. Today, NLP systems can translate languages, generate humanlike text, and even carry on conversations. But it's not just about chummier chats with your digital assistant. NLP is now at the core of how AI processes and interprets the written word, from sifting through legal documents to assisting doctors by analyzing medical records for critical information.

Computer Vision

Teaching Machines to See

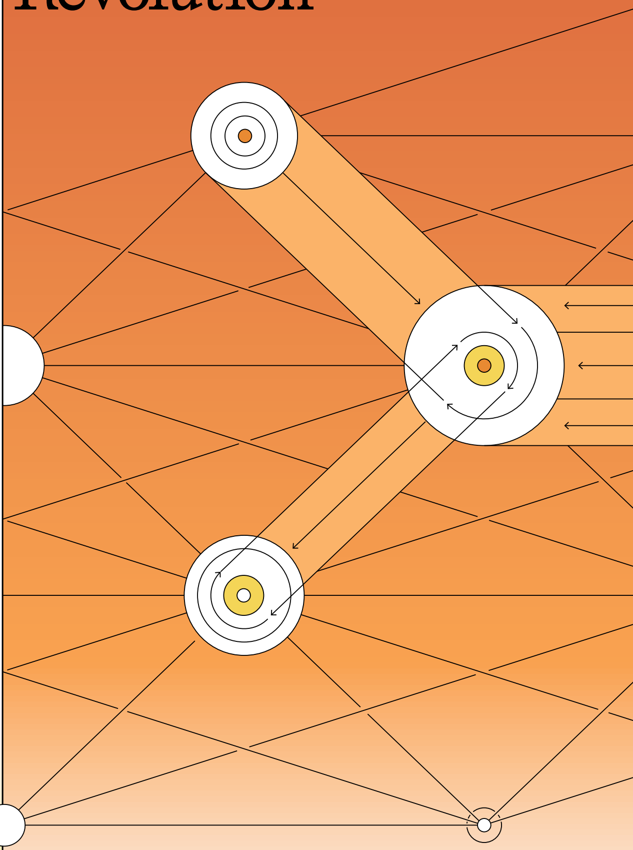
While NLP focuses on language, computer vision helps AI interpret the world visually. Researchers planted the seeds of this technology as early as the 1960s, when researchers at MIT attempted to use computers to recognize objects in images. It wasn't until the 2000s, with advancements in neural networks, that computer vision truly took off.

Computer vision models can identify objects, people, and even complex scenes by analyzing images' pixels. Computer vision systems, from facial recognition to self-navigating guidance systems, now use this technology. One key difference between early computer vision systems and today's models is their ability to process and learn from vast amounts of visual data. Early systems were labor intensive and limited to basic tasks like edge detection—i.e., recognizing basic shapes by detecting high-contrast transitions in images—and text-character recognition. Today, AI can "see" much as people can, interpreting complex visual environments, like busy intersections, packed crowds, and friendly faces, in real time.



Pt. 3

The Transformer Revolution

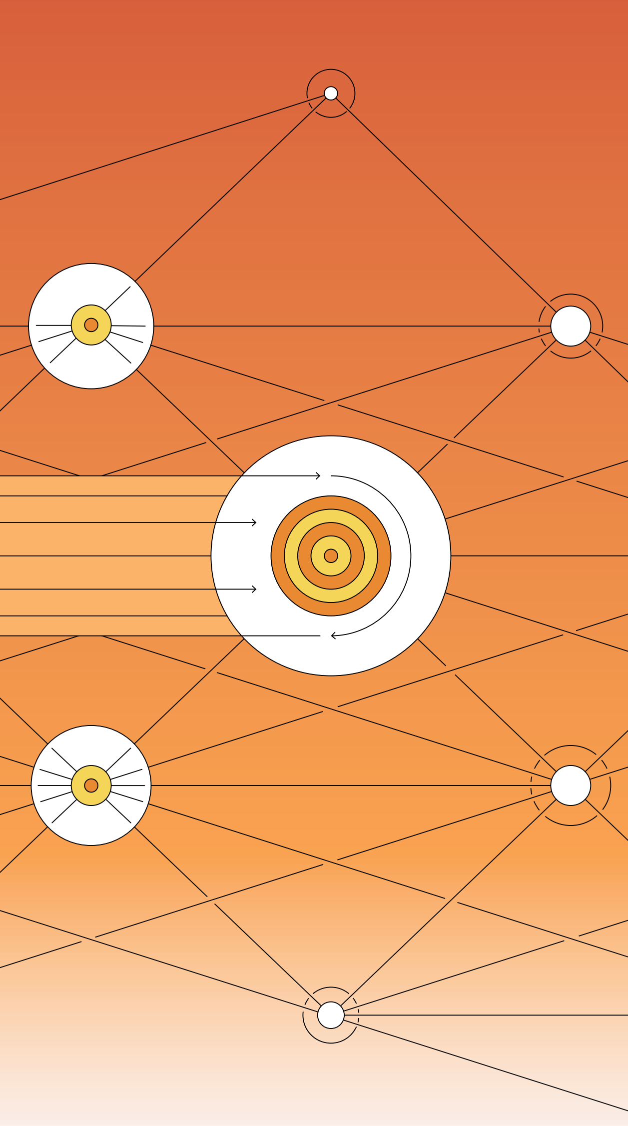


Transformers

Rethinking How AI Processes Data

As AI's evolution continued, researchers hit a bottleneck: how to process sequential data like language or time-series information efficiently. Standard neural networks weren't built to handle data that comes in a sequence, like a conversation or a story. Researchers needed a system that worked comparably to the human brain—capable of remembering what was said before to make sense of what comes next. Recurrent neural networks (RNNs) were the go-to solution since they create loops in the network that keep important information available for later use. But even RNNs needed help with long sequences and took far too long to train. Enter the transformer: a revolutionary architecture introduced by a team of Google researchers in 2017.

Unlike RNNs, transformers don't process data step-by-step. Instead, they use a mechanism called "attention" to



help the model highlight the most relevant parts of the input data simultaneously. Similar to how humans zero in on key parts of a conversation, this focusing ability makes transformers faster and more efficient, capable of handling much longer sequences of text or data without losing context. Suddenly, AI systems could process entire paragraphs of text or pages of documents in a single pass, leading to massive improvements in fields such as language translation and text generation.

Transformers have quickly become the backbone of modern AI models, making everything from real-time language translation to conversational AI possible. But they're not limited to text. Transformers are also making waves in drug discovery, genetic research, and other fields in which they help analyze complex biological data.

Pt. 4

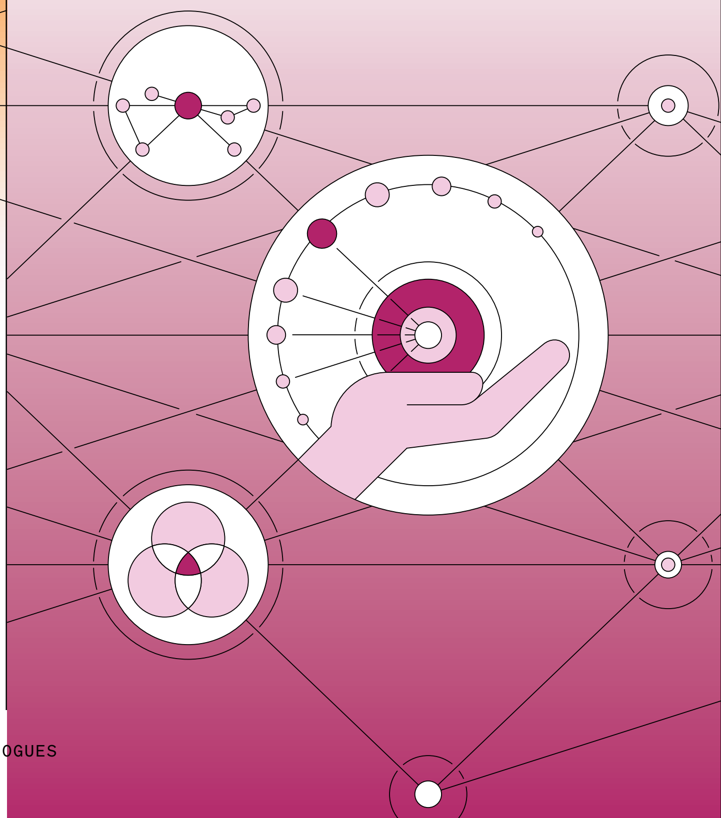
AI's Expanding Horizons

Recommendation Systems

Personalizing the Digital Experience

Ever wonder how your favorite streaming service predicts what you want to watch next? Or how online stores suggest products that fit your style? Enter the recommendation system. First appearing in the 1990s, today's recommendation engines have evolved into skilled curators, helping users sift through vast information by learning from their past behavior.

Recommendation systems usually rely on two standard methods: collaborative filtering and content-based filtering. The former bases suggestions on the behavior of people who use the system, while the latter focuses on specific details about the pieces of content to find similarities and links. Over time, these systems have become more accurate, combining both approaches to offer highly personalized recommendations. Recommendation systems are now being used to suggest everything from TV shows to health-care treatment plans.

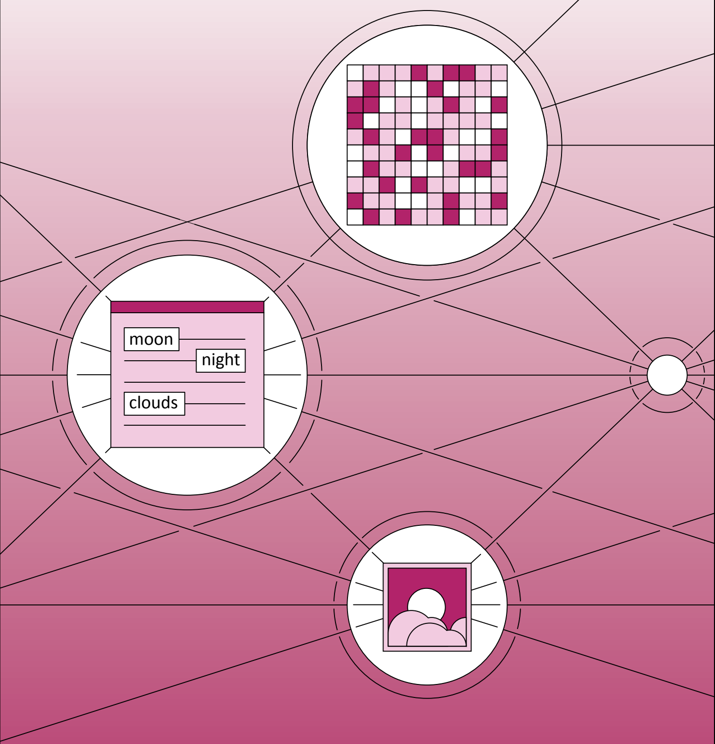


Diffusion Models

Creating from Chaos

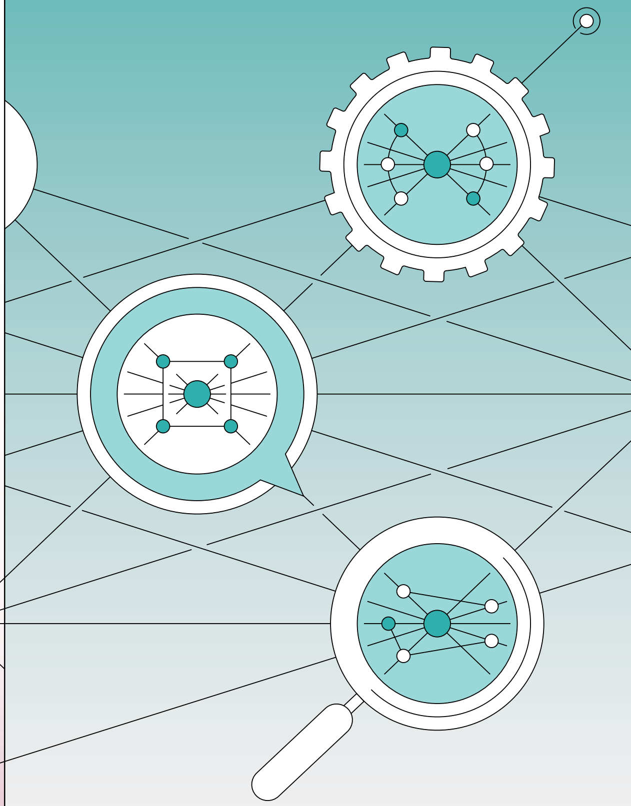
Diffusion models represent a vital recent development in AI image generation. First introduced in 2015 by a Stanford research team led by Jascha Sohl-Dickstein, these advanced algorithms generate images from text by iteratively refining individual pixels to match what the model has learned best fits the description. Imagine starting with a canvas full of static and watching a picture slowly emerge. That's how diffusion models operate; they generate images, audio, or text based on learned structures from an initially random state.

While still in their early stages, diffusion models are already used in creative fields. Artists and designers use them to create images or audio, while researchers explore their potential in everything from scientific simulations to virtual worlds. Diffusion models can also produce new training data, leading to more options for model development and tuning.



What's Next?

The Future *of* AI



As AI continues to evolve, one key area of ongoing research lies in making these systems more transparent and understandable. The research field of Explainable AI, for example, aims to shed light on how AI makes decisions—crucial for health care, finance, and other industries in which understanding the why behind a recommendation is as important as the result.

As AI grows more complex, so too does its potential. The once-separate branches of machine learning, neural networks, and natural language processing are now intertwined, creating systems that learn, perceive, and predict in ways that mimic human intelligence. From the early days of rule-based systems to today's transformers and diffusion models, the journey is far from over. Future advancements will continue to push what's possible for thinking machines—and the people who create them.

Beyond the Brain

←
Portrait by Oriana Fenwick

By Terrence Russell

An interview with Blaise Agüera y Arcas on intelligence and his team at Google, Paradigms of Intelligence

Blaise Agüera y Arcas has been at the forefront of AI research for years. As VP/Fellow and CTO of Technology & Society at Google, he leads the Paradigms of Intelligence (Pi) team, which takes a unique approach to AI research by interweaving ideas and expertise from fields such as computer science, neuroscience, biology, and philosophy.

We spoke with the multihyphenate researcher, engineer, and author about how Paradigms of Intelligence is helping shape the future of AI.

On the roots of his fascination with artificial intelligence:

I've been interested in AI from the beginning, from the moment that I was a kid and started to understand that we have brains. I was a computer nerd hacking around on a computer rather than having a normal social life. I did all kinds of crazy stuff as a kid. I broke a lot of copy protection on video games and I was very, very interested in the computational view of minds.

I was also just very interested in the fundamentals of how nature works. Originally, I thought that I was going to be a physicist, but eventually I found that the deepest and most interesting questions were more connected with neuroscience than particle physics or cosmology.

On the Pi team's mission to challenge assumptions about how intelligence develops, in favor of developing a more holistic and long-term paradigmatic view of AI:

We're looking at the basic assumptions about intelligence by exploring questions like "How can those be broken or changed, and where might that lead us?" For instance, one of those core assumptions is that brains are predictors—that the reason we have them is to predict the future given the past. And I never quite took those ideas 100 percent seriously. It's worth reassessing the paradigm of how brains actually work and exploring the high-level principles.

We're in this very exciting period of having cracked one of the central problems of AI, but at the same time, if we were to go to sleep and wake up in 20 years like Rip Van Winkle, I don't think any of us would still believe it'll be about transformers and chatbots. Our work is about advancing fundamental areas of understanding. That's why we're bringing in all these diverse perspectives. Whether it's complexity science or speculative philosophy, it's all about rethinking intelligence and computation from the ground up.

On ensuring that the Pi team has assembled people capable of tackling its heady, multifaceted mission:

We're working with folks from Mila, the Santa Fe Institute, and philosophers like Pi Visiting Researcher Benjamin Bratton, who think about computation in new ways. We have a wide range of thinkers: engineers, philosophers, and researchers from various fields, including those outside of AI. Everyone's bringing a different perspective to the table. There's no formula for it, but I believe that hybridity—multiple currents crossing in unusual ways—is where interesting things happen. I do try to match up people from very different worlds—maybe an engineer with a philosopher or a scientist. The magic happens when their different perspectives collide. I guess my intuition is that everything interesting that humans do comes from hybridity of some kind. It comes from multiple currents crossing in some unusual way.

On collective intelligence—progressing from individual knowledge to advanced networks of shared understanding—as one of the keys to redefining our view of what's possible with AI:

We're trying to understand the systems that inform other systems—the meta-systems, if you will. Intelligence isn't just a product of neurons firing in a brain, or circuits in a machine. It's shaped by evolution, by development, and by interactions within larger ecosystems. I've always had this sense that locating human intelligence solely within the individual human brain is a little limiting, even provincial. Intelligence, to me, is a collective phenomenon. It's about cooperation, scale, and societies, not just the brain. When

> "I've always had this sense that locating human intelligence solely within the individual human brain is a little limiting, even provincial. Intelligence, to me, is a collective phenomenon. It's about cooperation, scale, and societies, not just the brain."

Blaise Agüera y Arcas, VP/Fellow and CTO of Technology & Society at Google

we attribute something to "human intelligence," we're often talking about the superhuman intelligence of all of us—our collective efforts, rather than what any one person can do alone. And this leads us to think about intelligence not as something that's located in a brain—whether human or silicon—but as something that emerges from interactions across systems, across minds.

People often criticize AI for failing certain high-level tests, like solving complex math problems. But how many humans can solve those kinds of problems? It's a rarefied skill. AI models, much like humans, aren't designed to be universally intelligent. They're drawing from vast amounts of collective human output. I don't think that we're going to be able to have AI models breaking new territory until they're able to operate at this human social scale, just as an individual person is not going to become one of the 50 best number theorists in the world by clicking around at random on the internet. You can become one of the 50 best number theorists in the world—or even one of the 50 best poets—only by interacting with the other 49 and forming communities of interest through committed interactions.

On expanding our concept of learning well beyond the capabilities of a single brain—viewing intelligence as a phenomenon of cooperative social networks of shared knowledge:

If you were to take an individual human and raise them alone from birth, I don't think that a visiting alien would be astounded by their intelligence relative to the other fauna on earth. We're not that different individually from the other great apes. A lot of the differences are subtle. They have to do more with motivation, with the fact that we are not quite as strong as other mammals, the fact that we're a little bit better cooperators, etc. The really magical thing happens when you start to get humans working together and collaborating in larger and larger groups.

There's a bunch of really cool work in anthropology that's happened in the past 10, 15 years showing how the scale of a society relates to the complexity of the technologies and art that it produces. There are these famous cases from when Tasmania was cut off from the mainland of Australia, and [Tasmanians'] level of technological complexity dropped quite a bit. They lost a bunch of technologies that would actually have been very useful for fishing and other tasks. They didn't lose these technologies because they were not as intelligent as anybody on the Australian mainland. They lost them because they were smaller in number and were isolated. A lot of things that we attribute to human intelligence are highly collective. The individual is just not that smart.

On breaking old molds and going beyond conventional research boundaries to develop a new understanding of intelligence:

I'm very interested in that paradigm where we're pursuing communities of interaction that are generative and mutual. And from that perspective, I don't see AI as being particularly separate from human intelligence. I think that we are starting to have nodes in that graph, if you like, that are silicon-based, as opposed to brain-based, but the interactions are actually where the intelligence occurs.

Paradigms of Intelligence isn't bound by the traditional disciplinary boundaries of research. Instead, we're trying to explore intelligence from multiple angles: human, social, computational. My hope is that by challenging these old assumptions, we can help lay the groundwork for new ways of thinking about intelligence that go far beyond what we imagine today.

FEATURE

Can We Align Language Models With Human Values?

It's a delicate challenge that has engaged the minds of leading AI researchers.

By Matthew Hutson

→
Illustrations by Chris Burnett

In July, I spoke with the founders of Gray Swan, a start-up focused on AI security, a few days before they publicly announced their venture. Gray Swan aims to evaluate and fortify large language models (LLMs), the software behind AI-powered chatbots. One of their offerings is a model called Cygnet, which they built to reduce harmful outputs, such as giving instructions on how to commit bank fraud or similar criminal acts. They'd shown that their method makes the models more robust than comparison methods, without reducing performance—meaning that the LLM resists attempts to subvert its output restrictions while giving helpful and precise responses. "For the first time since I've been in this field, which is a very long time," says Zico Kolter, a co-founder and director of the machine learning department at Carnegie Mellon University, "there seems to be real, genuine commercial interest in fortified LLMs, because it's no longer a trade-off between accuracy and safety that no one would take."

A day after Cygnet launched, an influencer named Pliny the Liberator had jailbroken the model, freeing it from its safety guardrails. By phrasing questions in particular ways, Pliny had induced Cygnet to output a Molotov cocktail recipe and malicious computer code. "This new model, Cygnet, is 'the pinnacle of safe and secure AI development' and is 'designed to counter the most potent attacks,'" Pliny tweeted, quoting a Gray Swan press release. "It was indeed

a challenge compared to the vast majority of current models, but I'm something of a pinnacle myself."

"Haha, well done," Kolter tweeted in reply. "Faster than we thought." Before Cygnet's release, Kolter told me that the model was robust but not perfect and would eventually be hacked. "The point here is that security is a process," he said, "not a destination."

LLMs are becoming more capable and more ubiquitous. Hundreds of millions of people use them to draft emails, summarize research, generate code, and carry on conversations. The versatility and the analytical power of the models make them increasingly useful but also grow the potential for harm. The list of real and potential risks is long. The models can, or might, produce text that is incorrect, biased, toxic, privacy-violating, copyright-violating, or useful to criminals. People could use them to write damaging code or impersonate individuals.

To make these systems generally more helpful and less harmful, AI researchers have been working to mold LLMs to human values and preferences. This process is called alignment. There are methods to shape AI at every stage of the development and deployment pipeline: from filtering training data to fine-tuning models on select tasks, to prompting them to "think" harder before answering (for instance by breaking their outputs into logical steps), to restricting the answers that they give in reply to our queries.

↑
Anca Dragan leads Google DeepMind's AI Safety and Alignment team.

Training day

The first step in aligning an LLM is data selection. LLMs are artificial neural networks, pieces of software that take inspiration from the brain's wiring to make connections between pieces of information. Initially, the strengths of the connections between the data nodes are random. Developers run text through the algorithm, broken down into "tokens" (words or parts of words), and the LLM continually predicts the next token in the sequence. When it's wrong, two processes called backpropagation and gradient descent adjust the model's connections to improve its predictions. The resulting "base" or "pretrained" model has not been taught specifically to answer questions or solve problems. It merely models the language (or other tokenized input) it has been trained on, imitating the patterns in the reams of text. The data can contain billions or trillions of tokens extracted from web pages, books, code repositories, and elsewhere.

Like any type of training, the content and the quality of the text that the model learns from can impact the content and the quality of the text that it can generate; garbage in, garbage out. LLM developers often filter the training data, deleting erotic text, personal information, toxic language, and other content. This process may seem simple, but deciding what is offensive or private or dangerous is often subjective, and words can have multiple and nuanced meanings in different contexts and cultures. However, the model should still be able to identify negative or dangerous content, if for no other reason than to avoid generating it. "A ridiculous example would be if we try to make the pretrained model not know anything about bombs," says Anca Dragan, a computer scientist who heads the AI Safety and Alignment team at Google DeepMind. Such censorship would be difficult, she says, because the AI might piece together the idea of bombs from concepts in chemistry. Even if the censorship worked, she adds, the model would have difficulty summarizing a news story mentioning bombs.

A model that generates text expressing that bombs are bad or that can't generate text that is erotic or deemed toxic is inherently biased against these topics. We may think this is good bias, but other biases aren't perceived as helpful or even universally agreed upon. For instance, certain words and ideas that might be labeled offensive sometimes appear more often in text written by or about marginalized communities.

To get the model to generate content about certain topics or in a specific style, developers need to fine-tune it. There are typically at least two stages of fine-tuning a pretrained model. The first is called supervised fine-tuning, A social and even philosophical question looms over the technical research on AI alignment: Whose values or preferences should we align with?

which uses datasets of inputs and desired outputs. Developers run the model on an input, such as a question, and then correct the model based on the similarity of its response to the desired output. A common supervised technique is called instruction tuning. Developers feed a model a prompt consisting of instructions for a task, as well as examples of the task and correct or incorrect solutions (labeled as such) and a fresh question. The process involves many prompts and can cover many tasks, such as translation or summarization. When collecting desired responses for the training data, developers often ask human annotators to provide answers that meet certain value criteria, such as "honest," "helpful," and "harmless"—three of the values popularized by Anthropic's alignment for the Claude model. (A model that cheerfully answers questions about bomb-making would be helpful but not harmless, while a model that refuses to answer most questions would be harmless but not helpful.)

The second common stage of fine-tuning—reinforcement learning with human feedback—involves several steps. The AI designers collect many outputs from a fine-tuned model and ask people to judge which of two responses they prefer for a given input. Then they train a separate model on this data to predict human preferences. This is called a reward model. Finally, they train the first model to produce outputs that the reward model rates highly. Using a reward model for feedback means that they don't need humans to manually rate every output from the LLM they're fine-tuning.

Still, collecting the initial data from humans can be slow and expensive, which has inspired an even more automated approach called reinforcement learning from AI feedback. One method developed at Anthropic, called Constitutional AI, requires humans only to write a list of general principles, such as "Do NOT choose responses that are toxic, racist, or sexist, or that encourage or support illegal, violent, or unethical behavior." In the supervised-learning stage, researchers ask a pretrained model to answer questions, then revise them in light of the principles. They then fine-tuned the model on these revised answers. In the reinforcement-learning stage, they trained the reward model mostly on ratings not from humans but from another AI model that was asked to provide ratings in light of the principles. In this way, the feedback process is automated but guided by the human-written general principles.

Phrasing a question differently can lead to a very different answer. Such brittleness also leads to jailbreaks. Gray Swan built Cygnet to avoid these problems by aligning not just the output text but also the model's internal representations—the way that the model encodes and processes information. In a method called representation engineering, you can induce a model to behave in desired or undesired ways and then look at the differences in neural activations. You can then train additional intermediate layers that shift the model toward a desired behavior. This approach modifies the LLM's abstract representations of concepts such as sexism rather than particular words or phrases that might embody sexism, making the defense more generalizable to new wordings.

Testing time

Alignment doesn't end with training. AI developers can also intervene during "inference," which is the process of running a trained model. Just as developers can filter training data, they can also filter the prompts that users enter and the responses that models provide. Or they can add a system prompt: If a user accesses a model through a web interface, the developer might append some invisible text at the beginning of each user input with instructions for how to respond in the appropriate voice, tone, or character—basically telling the model to play nice.

For each response, models generate a series of tokens. For each token, the model generates a list of possible next responses, weighted by probability, and picks the one with the highest probability. To make sure the model picks an aligned output, researchers use a method called controlled decoding. This method uses a second model to alter the weights of the proposed tokens based on the probability that it will continue to produce an aligned output. For example, "I want to *hug* ..." has a higher chance of an aligned output than "I want to *punch* ..." Another popular method that increases a model's apparent reasoning is called chain of thought. In your prompt, before asking your real question, you provide another question and an example of the style of answer you'd like—a step-by-step solution. This spurs the model to follow the same process in its output, increasing the probability of a correct answer.

"Inference is the most exciting direction" for alignment, Yuchen Lin, a research scientist at the Allen Institute for AI, said. He and his colleagues developed a method called URIAL that matches the benefits of fine-tuning purely through prompt modification. They first showed that reinforcement learning is fairly superficial, as it alters the output of models mostly in stylistic ways (essentially increasing politeness). They then developed a system prompt that consists of general instructions—such as "You are a helpful, respectful, and honest assistant"—and three examples of queries and ideal answers, related to friendship, human rights, and renewable energy. When evaluated on six dimensions (helpfulness, clarity, factuality, depth, engagement, and safety), base models using this system prompt outperformed fine-tuned models that didn't use the prompt.

Models often produce incorrect information, sometimes called "hallucinations," because they're influenced by the information that's most common online, even if it's wrong or outdated. Take the example of a celebrity who has a birthday and becomes a year older. The larger the amount of content that says a certain thing (their previous age), the higher the probability that the model will lean toward the response with greater representation in the data set. This means there's a lot of work to do to sway a model away from outputs that have a greater representation in the data set and towards a less common but more accurate response.

One way to do that is a technique called retrieval-augmented generation. Given a query, the system retrieves relevant documents the user has made available, such as web pages or database entries, before running them through the LLM along with the prompt. This process enhances factuality and allows users to keep data in these documents private and current rather than handing it over to an outside company that will use it to fine-tune a model on it for them.

Enhancing truthfulness may seem more like an issue of model competence than model alignment, but it's a gray area. If a model "knows" an answer but doesn't provide the correct answer because it was trained to prioritize being engaging and having a friendly, conversational style over truthfulness, that's misalignment, Dragan says. What's more, alignment is about steering a model toward people's preferences, and people have different preferences about hallucinations, says Sara Hooker, a vice president of research at the AI start-up Cohere. A journalist may care more about facts than does someone using a model to write song lyrics. Users who want creative answers can turn up a model's "temperature," which is the probability it will stray from the most highly ranked tokens.

Align with what?

A **social and even philosophical** question looms over the technical research on AI alignment: Whose values or preferences should we align with? A 2022 paper reported that the opinions expressed by popular LLMs didn't reflect those of the U.S. population based on polls. Certain demographic groups were underrepresented, even when models were prompted to role-play as members of those groups. Recently, researchers suggested using ideas from social choice theory, such as collective decision-making, to find consensus on the values to instill in AI models. Sam Bowman, a computer scientist at New York University, who leads a safety research division at Anthropic—but was not speaking on its behalf—says that an Anthropic project called Collective Constitutional AI, in which U.S. adults submitted and voted on principles, was "a very first-draft attempt" at seeking consensus.

One idea is to have multiple models fine-tuned for individual users, countries, or demographics, but that's not very feasible, according to Hooker, because of the cost of training new models. Better to have one model that adapts to users and circumstances. Dragan notes that a model shouldn't adapt too much, comparing the situation to autonomous vehicles. "The passenger might have driving-style preferences," she says, "but you also have a responsibility as a car company to pedestrians and other drivers on the road."

Alignment requires not only technical and philosophical solutions but also regulatory ones. But we're far from consensus on what to regulate. Some legislators have proposed restricting models that require more than a certain amount of money and computation to train. Hooker says that model size is not a good indicator of risk, especially considering the power of small models that could act as agents by performing tasks such as scheduling deliveries,

↑
As models generate a range of outputs, a human hand—be it through manual feedback, reinforcement from human-generated content, individual preference, or otherwise—is essential in selecting and prioritizing the most appropriate and successful results.

making travel plans, issuing a company payroll, and similar activities that have real-world effects. Policy should focus less on training methods and more on a model's capabilities. That raises further debates about how to evaluate alignment. Kolter of Gray Swan says there are two extremes: static benchmarks (tests with prewritten tasks) and iterative red-teaming (a process in which people try to hack a system). He'd like to see more methods in the middle, that are both automated and adaptive.

Kolter believes that Gray Swan's updated AI models will be broken eventually, and the game will go on. But after very little progress in AI adversarial robustness over the course of a decade—vision models are still easily fooled—he declared, "I am becoming more optimistic."

Q&A

It's Time to Think About Generally Intelligent AI

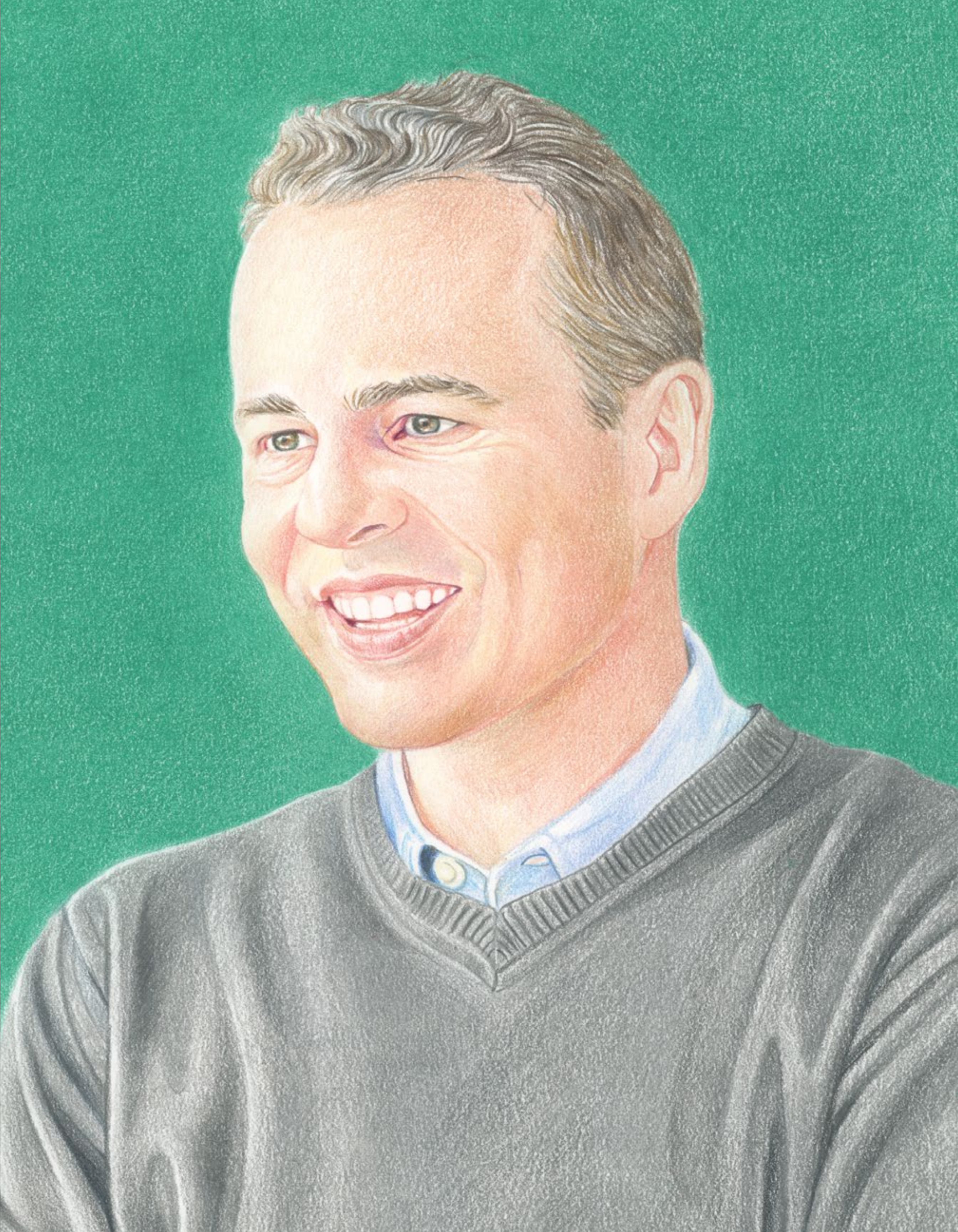
A conversation with Shane Legg, the co-founder of Google DeepMind, and Mira Lane, senior director of Technology & Society at Google, about the progress being made toward the goal of artificial general intelligence

By Mira Lane

→
Portrait by Uli Knörzer

Mira Lane **Where do we stand today in terms of AI development, and how close are we to realizing the possibility of AGI?**

Shane Legg We've all seen that AI is progressing quickly. I think we could be living with artificial general intelligence in five years. I think the probability is even higher in 10 years. When I made the first AGI timeline predictions back in 2009, I predicted a 50 percent possibility of AGI by 2028. So, let's just take seriously, for a moment, the possibility that this might happen in the next 10 years. AGI affects all these different areas, and you can't be an expert in all these fields. The advent of AGI is actually something that will require deep expertise in all human endeavors. What we really need would be for all the different departments and all the different faculties in universities to be thinking about the arrival of AGI. What does medicine look like in a post-AGI world? What does accounting look like in the post-AGI world? What does education look like in the post-AGI world? What does research look like in a post-AGI world? What does economics look like in a post-AGI world?



Lane **How are you defining AGI right now? I feel like that's a big open question.**

Legg I define an AGI to be an artificial agent that can do the kinds of cognitive things that people can typically do. I see this as the natural minimum bar. For some high levels of AGI capability, see the paper "Levels of AGI," written by a group of us at Google DeepMind last year.

Lane **What are the most significant open questions in AI research today, and what breakthroughs are still needed to propel the field forward?**

Legg We are working on AI in different areas, and many people think that achieving AGI is probably a refinement and an improvement and working on the sorts of things we already know and combining some of the methods we already know in the right way. It's not guaranteed. Maybe there is really a big breakthrough that is required, which we don't know yet. But there probably doesn't need to be a breakthrough as big as transformers—which were developed at Google—to get to AGI. There will be some advances. There are all sorts of low-hanging fruit at the moment to make our models better, like advances in datasets, memory, planning, and reasoning. And as we work on all of them, we see progress in all the different areas. So we are confident that, at least for the next few years, we can make these models—that are already getting very good—much better. And then as we start making agents out of these models, those agents will generate data as they interact with different kinds of environments and try to achieve goals in those environments. We'll then train new foundation models on that data. The resulting models will then be much better for building agents. This process may get us to AGI in as soon as five years.

Lane **How do you envision measuring the level of understanding in AI systems, and what frameworks inform your thinking on this complex issue?**

Legg Yeah, it's hard. Not all aspects of intelligent behavior by AI agents are easy to measure. And if there are aspects that are hard to measure, you don't know how well you're doing on them, and if you don't know how well you're doing on them, you may not even realize that you need to do better on them or are already doing well on them.

The other problem is that even if you can measure something, there are just so many things that you can measure. Because if what you are measuring is an AGI, it has generality—it's not that it does a specific thing. If it was doing a very specific thing, you could measure it thoroughly on that aspect. But if it's very, very general, it can do everything from writing code to understanding 20 languages, to making music, to making pictures to poems to legal work, to all kinds of things. That's a lot of things to try to measure. So the measurement problem is very difficult and it's very important.

It's also difficult because it's not a glamorous thing to do. Think about it this way: The most glamorous Olympics event is the 100-meter sprint, right? But you're not going to have a 100-meter sprint event if somebody doesn't build the track and get the start guns and the photo finish equipment all set up. As you know, you're not going to have a good event [without those people], but the glory goes to the runners.

Lane **Not the designers of the track.**

Legg The design doesn't attract the same kind of attention, but if you don't have a good track and the photographs and all that, it's just not really going to go very well. You need a good track that has to be level and the right surface and all these sorts of things. It's been a problem with machine learning for a long time that, psychologically, people are drawn to building the agent or being state-of-the-art on the benchmark rather than building the benchmark itself.

Lane **What governance models do you consider important to ensure this positive transformation? Where do you think public understanding needs to grow?**

> "We are working on AI in different areas, and many people think that achieving AGI is probably a refinement and an improvement and working on the sorts of things we already know and combining some of the methods we already know in the right way."

Shane Legg, co-founder of Google DeepMind

Legg I mean, they're all enormous questions. I think the biggest thing is how advanced public understanding has recently become around LLMs, in that it's not a technology that you just read about; you can get it on your mobile phone and you can talk and interact with it. And so you can at least start to get some grounding in terms of what this thing is. Members of the public are doing that en masse. Does it mean that people understand that powerful AGI is coming? Weirdly, I think many people do. And I actually think that, sometimes, lay people that have some technology interest have a better mental model of this than some experts who tend to be very skeptical and come at this with a lot of long-standing beliefs and biases. I've seen this throughout my career.

Lane **How so?**

Legg When we started DeepMind, everybody said it was ridiculous that we thought machine learning was going to be huge. People thought it was ridiculous that we were going to go and do this and that we were going to get *Nature* papers and be in academic journals or win awards. But no, AI keeps delivering the goods, and it keeps getting better and better.

Lane **We were talking about how highly capable AI systems will transform every single industry and human endeavor. I use LLMs for so many things, and it is remarkably good. I have shared emails, documents, and text messages with a model and have had it help me examine different perspectives.**

Legg I did something like that recently. I received a few messages and wanted to understand better what this person was trying to say. It seemed like they were hinting at something, so I put it in an LLM and I asked, "What is this person really trying to say?"

Lane **Fascinating, isn't it?**

Legg It's a whole new world, really.

Lane **Looking ahead 50 years, what do you think will be the most profound ways in which AGI will have transformed society, and what are your biggest hopes and concerns for the future of AI? What would it look like if we got it right?**

Legg Reductions in poverty and increased access to various kinds of resources and education. I think medicine could be advanced significantly, as well as scientific research. It could be good for the environment. We might have new types of clean energy sources or new types of materials and products. I see potential for improvements in every aspect of society.

CHAPTER

2

Accelerating the everyday

Intro

AI is embedded into the fabric of daily life, molding our everyday experiences in ways that may sometimes go unnoticed. In health care, AI-assisted tools are changing how doctors diagnose and treat patients. Education is undergoing a similar transformation, with personalized learning platforms adapting to an individual student's needs in real time.

The impact of AI extends far beyond the confines of wealthy, tech-centric regions. AI applications are addressing pressing challenges in agriculture, health care, and economic development around the world. Meanwhile, efforts to preserve endangered languages speak to the technology's potential to safeguard cultural heritage. The creative industries, too, are being reshaped: Musicians and artists are exploring new avenues for expression, reimagining long-standing notions about creativity.

These developments raise important questions. As AI becomes more integrated into our daily lives, how do we ensure it enhances human capability and agency? The stories in this chapter offer a glimpse into a future when AI is a collaborator in our everyday existence.

Bridging

By Sneha Mehta

the Gap

AI-powered tools enable personalized solutions in communication, mobility, and cognitive support, creating a new era of tools and technology that can be helpful for everyone, but especially people with disabilities.

Nearly 20 percent of the world's population is considered neurodivergent and roughly 1.3 billion people experience disability, according to the World Health Organization. In the future, even more people will likely identify as disabled because of conditions influenced by climate change, such as Lyme disease and asthma, and chronic conditions such as long COVID.

This group is not a monolith. Each of these individuals experiences their disability and the world differently, often having multiple intersecting disabilities to contend with in a world that isn't designed to accommodate them. One of the major challenges of accessibility efforts is creating solutions that can meet the uniquely complex needs of each individual.

"Traditional education often overlooks the specific, diverse needs of neurodivergent students, focusing more on surface-level issues or trying to fit them into a standard mold," Elsa, a design student at the London Interdisciplinary School who also identifies as neurodivergent, says. "We need solutions that offer individualized support, recognize unique strengths, and adapt to each student's evolving needs in real time."

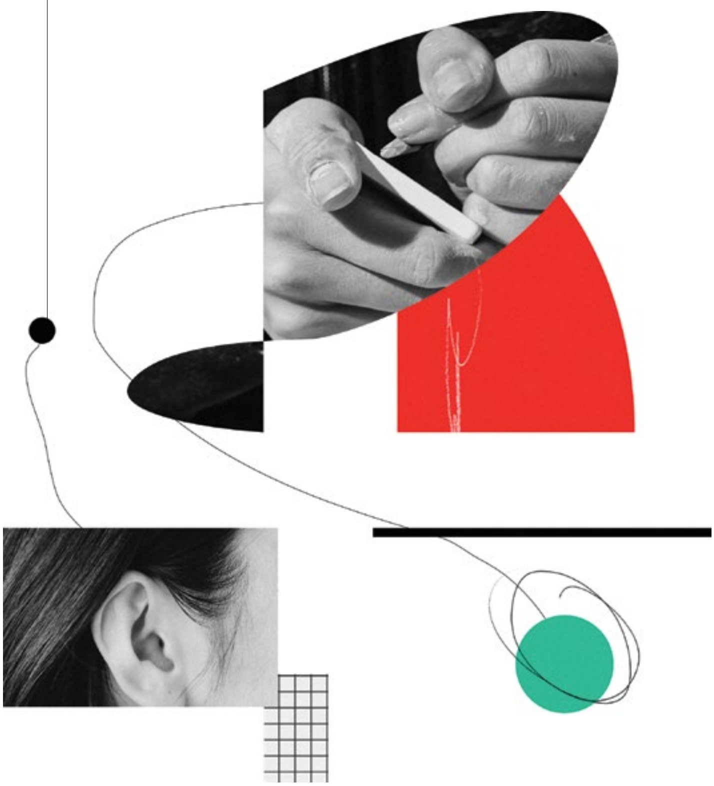
Without this adaptability, technological solutions for disabilities are often elegant and advanced yet implausible. For example, disability advocates have been critical of exoskeletons designed to help people who use wheelchairs because they're experimental prototypes that may serve only a small group of people, may never become commercially usable, and don't meet most users' current needs. But by harnessing the learning capabilities of artificial intelligence, designers and engineers are creating dynamic solutions to help enhance the most essential areas of daily life. And these improvements could help everyone.

"The magic of AI is its ability to be context-aware, making it much smarter and more accessible than previous technologies," says Dylan Fox, director of operations at XR Access, a research consortium based at Cornell Tech committed to making virtual, augmented, and mixed reality (XR) accessible to people with disabilities. "I'm really excited about AI's ability to bridge cognitive processing gaps."

Elsa, along with designer Sonny Kong and software engineer Oliver Fogelin, created an AI-powered educational tool called Pathia that tailors lessons to a student's unique communication style. The program learns about the student from every interaction, fine-tuning lessons into personalized formats such as worksheets, visuals, or games as well as providing insights to the teacher and the parents about the student's needs.

Pathia uses AI to process language in chunks rather than word by word. This approach, called gestalt language processing, helps the system understand context better. Pathia then adjusts its teaching style based on how each student responds to these language chunks, taking into account their emotional reactions and sensory preferences.

Elsa's brother, Ryan, who has autism and asked to go by a pseudonym for this article, expressed enthusiasm for Pathia, noting that it could explain concepts in ways his teachers couldn't. He believes that Pathia's ability to learn and adapt to his specific needs could lead to a better learning experience. "If I had a tool that helped me with communication through academia, to help me manage my frustration, and Ryan had one where he was helped to excel and tap into the parts of him that other humans may not identify with, that would have been life-changing for the both of us," says Elsa.



> "If I had a tool that helped me with communication through academia, to help me manage my frustration, and Ryan had one where he was helped to excel and tap into the parts of him that other humans may not identify with, that would have been life-changing for the both of us."

Elsa, a design engineer and student at the London Interdisciplinary School

COMMUNICATION

Dimitri Kanevsky lost his hearing at the age of 1. Since then, he has become a research scientist and speech-recognition technology expert at Google, employing various techniques to communicate effectively in both his personal and professional life. These methods include lipreading and Communication Access Realtime Translation (CART), a service mandated by the Americans With Disabilities Act in which trained professionals transcribe speech to text in real time. Both techniques, however, have significant limitations: Only about 30 to 40 percent of speech sounds are visible through lipreading, and during the coronavirus pandemic, masks made lipreading nearly impossible. CART, while helpful, can be expensive, isn't always readily available, and sometimes struggles with Kanevsky's accent and unique speech patterns, making it difficult for him to leverage the communication technology that was available at the time.

"So much of the lived experience of being a person in the world involves interaction and communication. Suppose that is one of the greatest challenges for somebody [with a disability]," says Dr. Keivan Stassun, director of the Frist Center for Autism and Innovation at Vanderbilt University.

At Google, Kanevsky and his team, in collaboration with Gallaudet University—a leading institution for deaf and hard-of-hearing students—developed Live Transcribe, a real-time speech-to-text application available in more than 80 languages. Live Transcribe integrates the entire ecosystem of sounds necessary for seamless communication. It not only converts spoken words into text but also detects and notifies users about surrounding sounds, such as doorbells, a baby crying, or a dog barking, and has a haptic sensor that vibrates when someone calls the user's name.

"Speech isn't just about words. It's filled with noise, accents, and varying speech patterns, making it tricky to handle. AI models like transformers are particularly good at tackling these challenges because they can process long sequences, handle noisy environments, and learn complex patterns in speech," Kanevsky says. "By combining speech models with advanced language processing, AI can understand context, fix misheard words, and produce text that makes sense, which is essential when the meaning depends on how words are spoken."

AI's flexibility and pretraining make it highly effective for modern speech-recognition tasks, pushing the boundaries of what's possible in terms of accuracy and application. It has the potential to usher in a new era of communication technology, which is helpful for everyone, but especially people who are deaf, hard-of-hearing, speech impaired, or neurodivergent and/or have conditions such as amyotrophic lateral sclerosis and cerebral palsy.

PopSignAI is another example of AI being used for communication. The app—developed by the Georgia Institute of Technology, the Center on Access Technology at Rochester Institute of Technology's National Technical Institute for the Deaf, and Google—makes it easier and more accessible to learn American Sign Language (ASL). Powered by AI-enabled sign-language recognition, the app uses an educational game to provide people with real-time feedback on their hand shapes and signing accuracy. PopSignAI allows people to more easily learn ASL, acting as an additional tool to help hearing parents connect with their deaf children on a deeper level.

"Ninety-five percent of deaf children are born to hearing parents who often don't know sign language. By helping hearing parents communicate more with their children, PopSignAI empowers these children with language, transforming their world through play," says Sam Sepah, the ML/AI research program manager at Google.

When Kanevsky used Live Transcribe to give a presentation at an international math conference in Poland in 2023, it was a landmark moment for him. "It was an amazing experience to be able to freely communicate with other mathematicians and deliver a presentation for the first time in my life," says Kanevsky. "I'd been waiting for something like this my entire life."

INDEPENDENCE AND MOBILITY

Since autism spectrum disorder was officially described by the DSM-3 in 1980, there has been a significant amount of research and interventions for children with autism. But to live independently, adults with autism need jobs. According to a Deloitte study, approximately 85 percent of people on the autism spectrum in the United States are unemployed, compared to roughly 4 percent of the overall population in the country.

"We talked to autistic adults and asked, 'What is the thing that you most wish you had access to?' Overwhelmingly, the number one answer is a job," says Stassun. "Okay. Next question, 'What is the thing that you currently have or don't have that is limiting your ability to access employment?' Most people said, 'I don't know how to drive a car.'"

In the United States, especially in cities where public transport is limited or not accessibility friendly, driving is often the only way to commute to work. So the Frist Center created a virtual-reality (VR) driving simulator that uses artificial intelligence to positively reinforce the signs, threats, and interactions that the learner should pay attention to and anticipate—all in a safe virtual environment. Physiological sensors monitor the learner's heart rate and eye movements, generating data that the trained AI model can use to interpret the individual's emotional state and adapt the simulator accordingly.

"It's the revolution with AI that is enabling us to create these realistic environments that are real-time and appropriately adaptive and responsive to each individual and their experience," says Stassun.

For some physical disabilities, prosthetics have long been used to accommodate losses in functionality. But despite the advancements in robotics and other high-tech prosthetics, they are not a panacea. Researchers from Austria found that 44 percent of arm amputees who choose to use prosthetics are unsatisfied with their devices because they're uncomfortable, painful, or hard to operate. A new generation of designers and neuroscientists is rethinking the very concept of prosthetics, using developments in AI, robotics, and accuracy in measuring our muscle's nerve responses to create augmentation technology that adds or extends functionality to the body.

> "It's the revolution with AI that is enabling us to create these realistic environments that are real-time and appropriately adaptive and responsive to each individual and their experience."
>
> **Dr. Keivan Stassun,**
> Director of the Frist Center for Autism and Innovation

"It doesn't matter what your body started like. We're interested in extending the biological body with technology," says Dani Clode, head designer at the Plasticity Lab at Cambridge University.

Clode is developing wearable technology that is assistive and adaptive to the future body. One of her designs, the Third Thumb, is a 3D-printed thumb extension for the hand, controlled by motion sensors placed under the toes. For the Third Thumb and other augmentation prosthetics, Clode plans to use artificial intelligence to anticipate the user's needs, "creating a symbiotic relationship between the user and the technology."

Tamar Makin, a neuroscience professor and head of the Plasticity Lab, emphasized the benefits of AI in enhancing motor control and communication. She likens the interaction between a user's brain and AI technology to how an octopus controls its tentacles: There's a central command, but also localized decision-making.

"We're working on this collaboration between the big brain, which is the user's brain, and the small brain, which will be the AI that we're endowing to the technologies so that they can work relatively autonomously," she says.

AI and machine learning algorithms can help to quickly process the body's neurological signals by learning the user's intentions and analyzing data from electromyography sensors, thereby making prosthetics that can quickly adapt to the individual.

"If you just have a broken limb, maybe you don't need AI. But if you have a stroke, the signals from your body are already very faint. We can use AI to better understand what this signal means in different contexts. This can help us complete the thought for the user by creating a seamless movement," says Makin. "The opportunities in this domain are limited by our imagination, not by technology."

COGNITIVE LOAD

Generative AI can lighten people's cognitive load by making the web and daily life interactions more accessible. "A lot of disabled folks have trouble writing official letters like insurance appeals and complaints. They've been using generative AI tools to write them in a way that will be read and respected by the person who receives it," says Ashley Shew, an associate professor in the Department of Science, Technology, and Society at Virginia Tech. "And if those appeals help get the benefits, equipment, or maintenance they need, then there's a real everyday life outcome to LLMs that their designers probably never anticipated."

For a person who is blind or has low vision, navigating unfamiliar environments requires learning and memorizing spaces and filling in any gaps in information without a visual reference. GPS technology has been an important aid, but even the most advanced versions of the technology are often accurate up to only 30 feet. For a person who is blind or has low vision, standing 30 feet away from a bus stop could mean missing the bus. Some people may choose to practice orientation and mobility skills using the controlled environment offered through a VR headset, but even those devices have their limitations.

A consortium of researchers at Cornell is working on an AI-powered sighted guide that would help people who are blind or have low vision to navigate and understand VR experiences—similar to a physical sighted guide who assists people who are blind or have low vision.

"That idea of being able to grab onto somebody's elbow and have them lead you around is not built into any VR simulation, because no one thinks it is a user need," says Fox of XR Access. "We're thinking about how to program an AI to help guide you around a VR space."

> “People with disabilities are often first adopters of new technology because their need is higher. At Google, we believe that if we start with one use case, we can go on to help billions.”

Dimitri Kanevsky,
Research scientist at Google DeepMind

CHALLENGES

In his essay “Keeping the Knives Sharp,” poet and disability activist Jim Ferris asks, “What would it mean to live in a world that understood asymmetry as a prime characteristic?” Ferris makes a point about the othering of people with disabilities and their absence in the design process, datasets, and lawmaking—significant challenges in the effort to use AI to enhance lives.

Tools, features, and design elements for people with disabilities benefit broader populations as well. One of the mandates of the landmark Americans With Disabilities Act, which came into effect in July 1990, was that sidewalks must include curb cuts, the ramps that enable people using wheelchairs to more easily travel between the sidewalk and the street. But it became clear that the curb cuts made mobility easier not just for their intended audience, but for everyone, including parents with strollers, people pushing heavy carts or wheeling luggage, and runners and skateboarders. Similarly, captions for auditory media help everyone, not just those who are deaf or hard-of-hearing; studies have found that nearly 70 percent of Gen Z uses closed captions 100 percent of the time for concentration or to understand different accents.

“People with disabilities are often first adopters of new technology because their need is higher,” says Kanevsky. “At Google, we believe that if we start with one use case, we can go on to help billions of people.”

This means that people with disabilities need to be involved in the human-centered design process from the very beginning. Doing so would avoid creating “disability dongles,” a phrase coined by disability advocate Liz Jackson to describe “well-intended, elegant, yet useless solutions to problems we never knew we had.” Fox says, “You want to have disabled people in at every step of the process, from brainstorming to prototyping to testing.”

While AI holds promise for enhancing accessibility, disability advocates also caution against allowing the allure of promise to distract us from bigger issues related to pervasive bias. The existing literature about disabilities used to train AI models may be outdated or incorrect, and certain social groups may be disproportionately represented in medical data. For example, compared to immigrants and people of color, wealthy white families in the United States are more likely to report childhood autism due to better medical access.

“If all of our sources for disabled data are from nondisabled people, the stories are going to be wrong,” says Shew.

Despite these challenges, technology and disability experts are optimistic about the revolutionary changes AI can bring about by meeting users where they are.

“The AI revolution has occurred. The world will not go back to its pre-AI state. As much as we need to attend to the appropriate concerns and fears, we would be seriously undercutting our ability to develop real solutions at scale and have a major impact if we didn’t incorporate the power and the potential that AI brings to any technological solution that we envision,” says Stassun.

FEATURE

Smarter Medicine

In the quest for better health care, generative artificial intelligence emerges as a powerful ally, taking on the rote tasks that consume doctors' time and letting doctors be doctors—namely, helping them focus on what matters most: the patient.

By Maya Kosoff

→ Illustrations by Hoi Chan

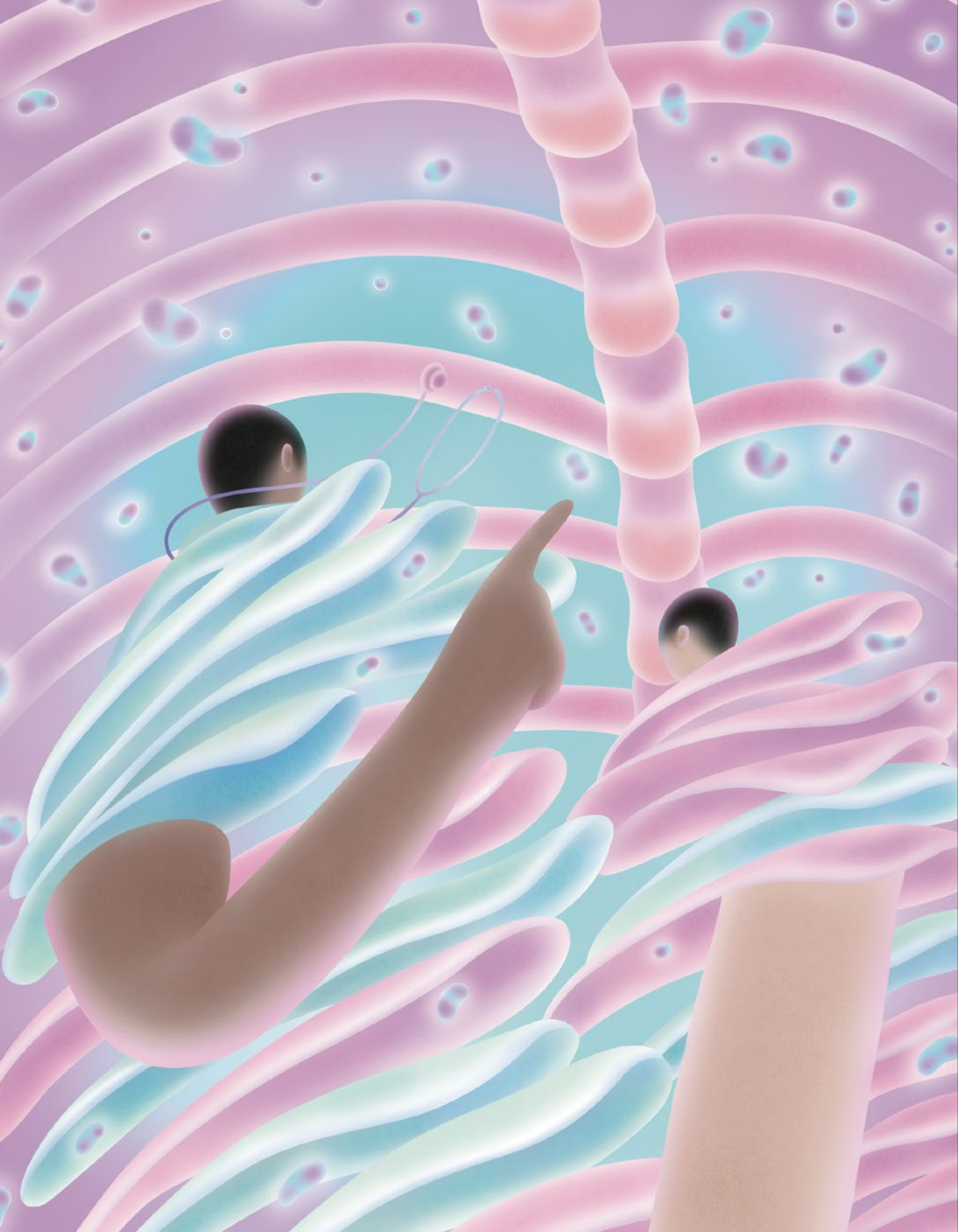
The doctor will see you now, and AI will be assisting. That's not a distant promise; it's our present reality. Almost two-thirds of 1,081 clinicians who responded to an American Medical Association (AMA) survey said they recognize the advantages of using AI in their work, and 38 percent said they were already using it at work. The way we experience health care is undergoing a dramatic transformation, leading to faster diagnoses, more personalized treatments, and ultimately better outcomes for patients.

From predictive algorithms that enhance diagnosis accuracy to sophisticated data analytics that speed up drug discovery, AI has the potential to become a pivotal partner to doctors. Today it's amplifying the capabilities of health-care professionals as well as ushering in a new era of medical breakthroughs and improving patient outcomes. To better understand how doctors are already using AI, we investigated how clinicians and health-care experts are using these systems now, and how it could change their practice today and in the future.

EASING THE ADMINISTRATIVE BURDEN

Health-care worker burnout has become an epidemic. In 2022, U.S. Surgeon General Dr. Vivek Murthy issued an advisory highlighting the urgent need to fix the clinician burnout crisis. He projected a shortage of more than 3 million essential low-wage health workers in the next five years and nearly 140,000 physicians in the next decade. According to a Mayo Clinic study, more than 3 in 5 physicians reported at least one symptom of burnout in 2021. Against the rising tide of care costs and underinsured patients who wait too long to get checked out, there is too little time and too much care needed to make this system work. And caregivers—administrators, nurses, assistants, and clinicians—all put their own health on the line to be there for their patients and communities.

"We recently conducted a study with the Harris Poll that found clinicians spend nearly 28 hours per week on administrative tasks, while medical office staff and claims staff spend 34 and 36 hours, respectively. This includes maintaining detailed patient records, completing insurance forms and referrals, documenting procedures performed, organizing documentation for claims, and inputting claim information into the system," says Aashima Gupta, global leader of Healthcare Industry Solutions at Google Cloud. Prior to joining Google, Gupta led Digital Health Incubations at Kaiser Permanente, where she brought digital transformation to outdated, paper-based health-care systems. Gupta says health-care systems are already integrating AI into four major areas: transformational care delivery for patients,



Imagine a doctor asking an AI tool,
"Can you generate a timeline of this patient's care, highlighting key events and progression?"

or

"Can you transcribe my conversation with the last patient?"

operational efficiencies, research and development, and a digital "front door" to more effectively triage patients and address access issues.

Clinicians are using AI software today to change how they provide care, creating efficiencies for doctors and quicker answers and care for patients. Instead of clicking hundreds of times to get to an answer, doctors are using natural language processing to ask and answer patient-related questions.

The rise of artificial intelligence in health care presents incredible opportunities to improve patient care. However, the use of AI with sensitive patient data raises crucial questions about privacy, confidentiality, and compliance. Ensuring strong data privacy and security is paramount. This requires a multifaceted approach that includes secure data storage, robust access controls to sensitive patient data, and comprehensive privacy compliance. Google Cloud customers retain control over their data, and in health-care settings, access and use of patient data is protected by Google Cloud's reliable infrastructure and secure data storage that support HIPAA compliance, along with each customer's security, privacy controls, and processes.

Beyond searching and asking questions, Gupta says, AI is helping doctors find information within electronic health records (EHRs) to understand the context of a patient's condition and health status and then perform tasks based on that information. Imagine a doctor asking an AI tool, "Can you generate a timeline of this patient's care, highlighting key events and progression?" or "Can you transcribe my conversation with the last patient?" Some clinicians use human scribes, who sit in the room with them during clinical visits to transcribe conversations, but not every health system has the budget or ability to hire a clinical scribe. Clinical scribes also raise some concerns for patient confidentiality; some female patients may hesitate to allow a male scribe to sit in on their OB-GYN appointment, for example, whereas an AI tool may not provoke the same discomfort.

The burden of these manual tasks often takes a lot of time, forcing doctors to add notes to EHRs outside their work hours—what Gupta says some doctors refer to as "pajama time." An August 2024 report by the AMA found that roughly 21 percent of physicians said they spend more than eight hours on the EHR outside of normal work hours on weekdays. AI can help with EHR tasks, leaving doctors with more time to have human engagements with their patients, and their own family and friends, reducing their risk of burnout.

AI is a machine, but ironically, it's one that can be used to foster more empathy among clinicians. By having an AI agent or software help them with documentation, searching, and summarization, doctors have more space to empathize with their patients—a big reason many came into the medical field in the first place. “We believe AI’s role is to strengthen the relationship between a physician and a patient and to reduce that burnout or pajama time,” Gupta says.

AI can also be used to reduce the amount of manual, repetitive tasks involved with handoffs between nurses from one shift to the next. A Google pilot program at two HCA Healthcare hospitals—TriStar Hendersonville Medical Center outside of Nashville, and UCF Lake Nona Hospital in Orlando—is using Google’s generative AI technology on mobile phones to track nurses’ 12-hour shifts and create summary and task lists. At the end of their shift, a nurse can review the summary, make any necessary edits, and transfer that information to the next nurse on duty. Typically, this process is manual and time intensive and can result in poor patient health outcomes due to information that’s lost in translation or not shared from one shift to another.

ENHANCING DIAGNOSTIC PRECISION

AI is also significantly transforming radiology and imaging, enhancing the work of clinicians. According to Gupta, the majority of all health-care data is images: X-rays, CT Scans, MRIs—and AI can be used to maximally process those images and make sense of them.

Recent developments have enhanced AI’s ability to perceive sensory information with more accuracy, which has big implications for imaging and radiology. In the past, tasks such as reading and interpreting the results of imagery from scans and test results could be done only by humans. But AI is now capable of performing these tasks with increasing accuracy. Traditionally, radiologists visually assess X-rays and report their findings on their own or along with the second opinion of another doctor. MRIs and X-rays aren’t black and white—they’re gray and shadowy and can be tricky to interpret. Sometimes the difference between ordering a biopsy and giving the all-clear to a patient is in assessing a millimeter-wide edge of fuzzy gray pixels on a screen. AI provides quantitative assessments of images, assisting radiologists in making more precise diagnoses of conditions like broken bones, fractures, and tears.

In its radiotherapy research partnership with Mayo Clinic, Google Health developed an algorithm to help doctors differentiate between cancerous areas and healthy tissue susceptible to radiation damage—a process known as “contouring” that can take a long time when done manually. But when using the algorithm, doctors saw efficiency gains of 30 to 40 percent, Gupta says—meaning faster planning and treatment time for patients, as well as for the doctors helping them access treatment.

Specific fields of medicine like hematology and oncology are data-driven and -intensive. Oncologists and hematologists have a high clinical need for improved, more efficient workflows and better methods for diagnosing and treating patients. Time is of the essence for these doctors; the conditions they treat can spiral out of control quickly if left undiagnosed, and every day counts. However, because society has increased our capabilities to diagnose and treat cancer, many medical institutions have a massively growing trove of data and increasingly complex clinical workflows.

Google’s partnership with New Jersey–based Hackensack Meridian Health offers another glimpse into how AI can be used to aid clinicians in screening and diagnostics. Using AI modeling, doctors at Hackensack are processing large quantities of imaging data and building AI algorithms to predict metastasis in prostate cancer patients. This allows them to provide more accurate treatment plans for these patients.

In 2024, multinational life-science company Bayer and Google Cloud announced they are working together to address radiologist burnout and support more efficient diagnoses with generative AI. Bayer is working to accelerate the development and deployment of AI-powered health-care applications with a focus on radiology, using Google Cloud’s technology—including generative AI tools. Medical imaging makes up about 90 percent of all health-care data. By helping others overcome the challenges of building compliant AI-powered medical imaging software, Bayer and Google can pave the way for lighter, AI-assisted workloads for radiologists.

HUMANIZING PATIENT CARE

Personalization in medicine takes many forms, and AI is already facilitating personalized communications with patients and subsequent treatment. Today, health-care systems are using AI to improve population health and close gaps in care by reaching out to at-risk groups with personalized messaging. For example, a health system can identify all female patients ages 45 to 55 and send them reminders to schedule their annual mammogram. But AI allows this outreach to be personalized by connecting it to individual patient data. This approach moves beyond simply scheduling screening tests and instead focuses on the whole patient, recognizing their individual needs and risk factors. For a subset of patients that have a family history of breast cancer, for example, the invitation to schedule a scan could reference that family history to encourage and normalize scans as a regular activity the women in their family do. “You’re not just seeing an X-ray or doing a mammogram—you’re seeing a patient as a whole,” Gupta says.

> "You're not just seeing an X-ray or doing a mammogram—you're seeing a patient as a whole."

Aashima Gupta,
Global Leader of Healthcare Industry Solutions at Google Cloud

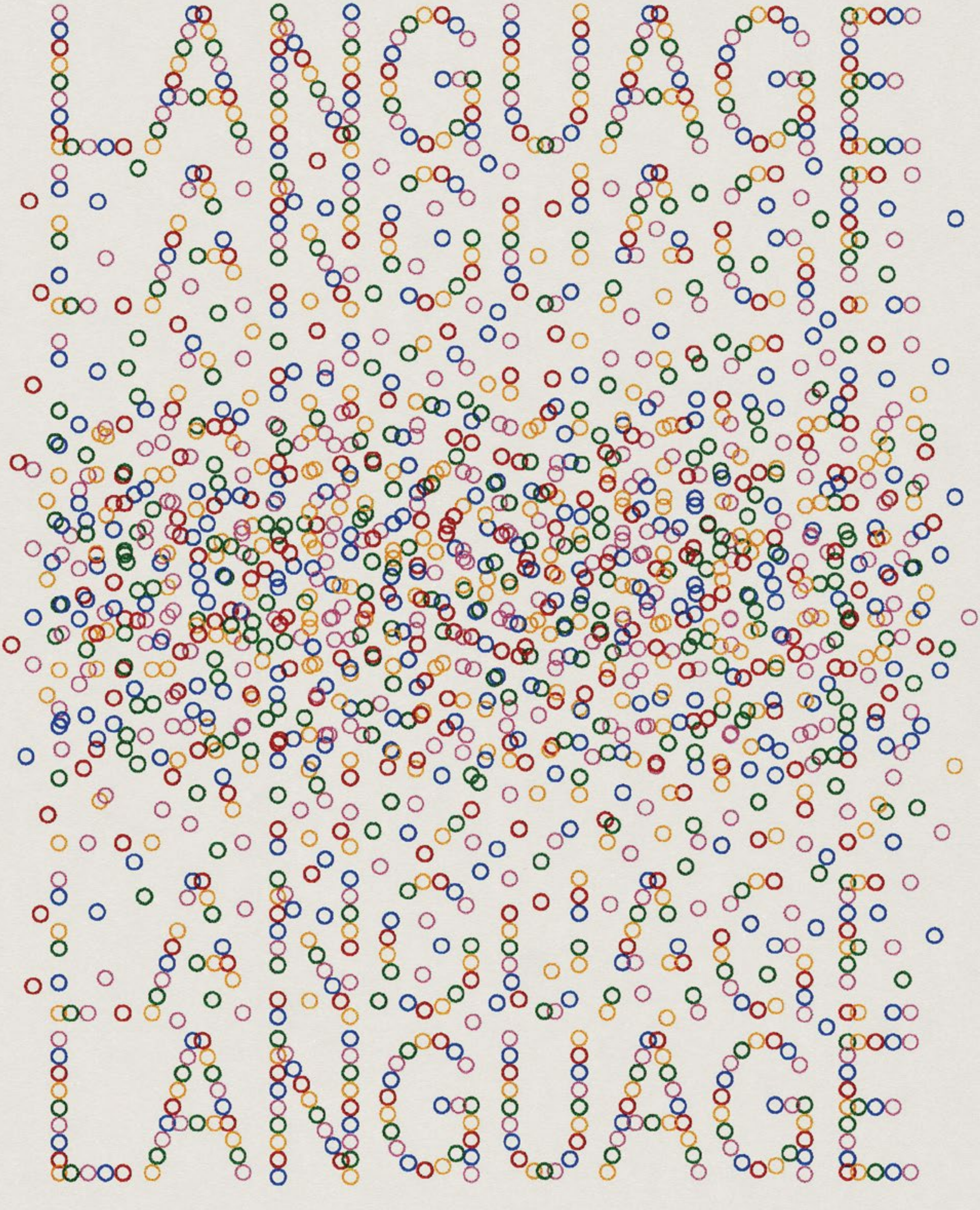
In India, which has the highest incidence of tuberculosis in the world, India's Apollo Radiology International is using Google's AI systems that enhance early detection of TB based on chest X-ray scans to help screen and detect tuberculosis in patients. "These are the areas where AI is going to democratize access," Gupta says. Over the next decade, Apollo will use Google's AI-powered screening models to provide 3 million free screenings for tuberculosis, lung cancer, and breast cancer, providing Indians with more timely diagnosis and care.

Gupta says AI can also be used to identify gaps in care, too. "We've built EHRs for the past 10 to 15 years, and the U.S. alone has spent $32 billion to create this massive trove of information. Sometimes that information is structured, but often the clinical notes where a patient shares pertinent information that can't go anywhere else is unstructured," Gupta says. For example, a patient may share that they are having challenges with housing, which means they may have trouble getting to appointments or having a place to store medicine, which in turn will affect health outcomes. There isn't an EHR record for those notes, but doctors can use AI to leverage this wealth of information to help meet the unique needs of patients.

AI is also being used to fill the gaps in medical research and treatment. "Traditionally, many medical studies have underrepresented minority populations. Using AI, you can look into datasets and see where the gaps are in the research," Gupta says. AI models excel at recognizing complex patterns and correlations within data.

At Google, bioethicists, AI researchers, health equity researchers, and clinicians have collaborated to build a framework aimed at reducing biases in AI systems. The framework, known as Health Equity Assessment of Machine Learning performance (HEAL), is designed to prevent AI technologies from exacerbating existing health inequalities, particularly among groups of people who historically experience poorer outcomes. In one application of HEAL, researchers tested an AI dermatology model, and although the AI model performed well in detecting melanoma across most demographic groups, it fell short for noncancerous conditions like eczema, particularly in older adults.

Much like the sophistication of AI technology itself, the way health-care systems are using AI is constantly changing. To make sure physicians don't feel alienated by these changes, experts insist on transparency with doctors, patients, and new technologies. "Transparency is going to be important," Gupta says. "Health care moves at the speed of trust, and we want to ensure any transformation within health-care systems is being done with nurses and clinicians and not to them. That's a humility we need to bring."


LANGUAGE
LANGUAGE
LANGUAGE
LANGUAGE

The Race to Save the World's Vanishing Languages

Thousands of languages are facing extinction, and researchers are turning to artificial intelligence to preserve and revitalize them. Can this technology save endangered languages before it's too late?

By Daniel Oberhaus

← Illustration by Ryan Carl

To spend an afternoon in Jackson Heights is to spend an afternoon touring the world. Nestled in the northwest corner of New York City's largest borough, Queens, Jackson Heights is commonly believed to be the most linguistically diverse neighborhood in the most linguistically diverse city in the world. It is home to some 180,000 people who collectively speak more than 160 of the 700-plus languages that have been documented in the city. It has every right to claim the mantle of the "world's town square." But just like the world it represents, Jackson Heights has a problem: Its languages are disappearing.

Linguists warn that the vast majority of the roughly 7,000 languages spoken on the planet are at risk of extinction. Estimates by cultural heritage organizations, including UNESCO, suggest that the last living speaker of a language dies roughly every two weeks, and with them any chance of revitalizing a community of speakers. At that point, the best linguists can hope for is that the language is preserved for future study, much as the artifacts of long-gone civilizations are preserved in a museum.

This is a cultural tragedy, driven by economic globalization, technological change, and government policy. It may be too late to save the estimated 600 languages that have been lost in the past century alone. But increasingly, linguists are turning to AI in a bid to not just preserve but also revitalize the languages that remain.

PUTTING THE "LANGUAGE" IN LARGE LANGUAGE MODEL

Artificial intelligence and human language have always been closely linked, but the rise of large language models (LLMs) has made this connection explicit. LLMs are trained on massive amounts of text collected from the internet and digitized books, allowing the AI to respond to many kinds of user input in an accessible, comprehensible, and diverse style of output.

Unsurprisingly, linguists and cultural historians quickly recognized the potential of this technology to help save thousands of languages at risk of extinction and to make content online more accessible to speakers of hundreds of languages unsupported by most technologies. Although most LLMs today are capable of handling only about 100 languages—albeit with a wide range of accuracy and quality—this new development offered the possibility of creating models fluent in at-risk and underrepresented languages. In addition to representing a powerful new linguistic research tool, this would be a major boon for the remaining speakers of these languages, helping educate the next generation of speakers and facilitating access to the economic opportunities enabled by the internet.

In principle, it's an elegant and important use case, but it's remarkably difficult to put into practice. One of the biggest challenges, says David Adelani, an assistant professor of computer science at McGill University who specializes in computational linguistics, is the lack of data. Training an LLM requires collecting a tremendous amount of text from the internet; even with already sky-high baseline requirements, an LLM continues to produce better results the more data it ingests. The problem is that an estimated 50 percent of all websites are written in English, and the top 10 languages account for more than 80 percent of all content on the internet. The vast majority of the Earth's languages are what Adelani refers to as "under-resourced languages," meaning that the mountains of textual data needed to train an LLM in that language don't exist.

"There is a very big connection between the amount of data available on the web and the languages that are supported by current technologies," says Adelani. "If your language doesn't have a lot of text online, it will be less represented in those technologies. There are so many languages that fall into that category, and most of them are from the Global South."

Endangered languages are, by definition, under-resourced, but not every under-resourced language is endangered. Oromo, for example, is a language spoken by about 45 million people in Ethiopia and Kenya, in stark contrast to a language like Kalasha-mun, which is spoken by a few thousand people in Pakistan. Despite its larger speaker base, Oromo remains underrepresented in LLMs and other AI-driven technologies, primarily because less than 30 percent of the population in the countries where it's spoken has regular access to the internet. This problem is exponentially worse for the majority of the world's 7,000 languages, most of which are spoken by fewer than 10,000 people.

But the general lack of linguistic data for training LLMs and other AI systems is only part of the problem. Another major challenge is that LLMs struggle to handle textual input that doesn't use the Roman alphabet, outside of a few major languages such as Russian and Mandarin. Many languages—especially the most endangered—lack even Unicode standardization for their script, which is necessary to write in a language on the internet. Adelani recalls attempting to train an AI model in Amharic, the most widely spoken language in Ethiopia, only to receive a zero accuracy score. The takeaway, he says, was clear: "If the script isn't supported, the model isn't going to work."

HOW AI CAN HELP LANGUAGES THRIVE

Today, the threat to the thousands of endangered languages spoken in the world is widely recognized by cultural conservationists, linguists, and technologists working on LLMs, who are exploring different solutions for rolling them back from the verge of extinction.

Many of these efforts are focused on digitally documenting endangered languages in the hopes that, if they can't be saved—in the sense of having a living community of speakers—they can at least be preserved for future study. One of the earliest and most successful examples of a global endangered-language preservation initiative is the Rosetta Project, which was launched in 2002 by the Long Now Foundation to create a digital library of all documented languages. This was supplemented by a separate initiative called the Endangered Languages Project, which launched in 2012 with support from Google. It has since blossomed into a catalog documenting more than 3,000 endangered languages.

In addition to collecting existing data on endangered languages, linguists worldwide are turning to smartphones, AI, and other technologies to accelerate the process of building corpora—essentially, large linguistic databases—for endangered languages, a prerequisite for building a viable AI model that generates text and speech in a language. In 2018, for example, Australia's ARC Centre of Excellence for the Dynamics of Language (CoEDL) began using a robot called Opie to help teach Indigenous languages to children living in remote communities. CoEDL then partnered with Google to build AI models for Indigenous languages, which saved linguists many hours of transcription time that would've been required to produce the data to train the models.

More recently, the Living Tongues Institute for Endangered Languages partnered with Shure, a company that makes small wireless microphones commonly used by social media creators, to simplify the recording of endangered languages in some of the most remote areas of the world.

These sorts of initiatives are an important part of preserving endangered languages, but not everyone is convinced it will be enough to save them—or even build effective AI models of those languages. In fact, in 2013, András Kornai, a professor of mathematical linguistics at the Budapest Institute of Technology, calculated that less than 5 percent of the world's languages can be captured digitally, as the remainder lack enough data, which means most of the world's languages are going extinct. In other words, a language may have a small but viable community of living speakers, but based on Kornai's research, it's already dead in the sense that it's unlikely to ever be used for online communication. For most of these receding languages, the best we might hope for is digital preservation.

"Language preservation is a great thing, but it does not lead to a viable language community," Kornai says. "You can reconstruct many things, but there's a certain amount of data loss that's inevitable."

This loss of native speakers creates a fundamental problem for machine learning applications. Without fluent speakers to validate the output, the applications are extremely difficult to develop and refine. "This also makes it almost impossible to experiment with that language using machine learning, because you don't have any speakers anymore who can tell if the output is gibberish," Kornai adds.

Daan van Esch is a linguist who has spent the past 12 years at Google researching ways to make technology more accessible to a broader linguistic community. He remembers when Kornai published his paper a decade ago and sent shock waves through the linguistic community. But a lot has changed in 10 years. Now, he believes that Kornai's dire prediction may be a bit too pessimistic.

"The big thing that's changed is now there are smartphones everywhere," says Van Esch. "There are so many people coming online and wanting to use their language on social media and chat apps, but it historically hasn't been supported."

In late 2022, Google announced the launch of its 1,000 Languages Initiative, which set the goal of building an AI model that would support the 1,000 most spoken languages in the world. But almost as soon as the project started, it was faced with the same monumental challenge: how to collect enough language data to build the model. The solution, says Uche Okonkwo, a Responsible AI program manager at Google, was a blend of technical innovation and community engagement. On the technical side, Google researchers focused on finding ways to do more with less data. For example, researchers at Google DeepMind developed a technique for training the company's LLM Gemini on an endangered Indonesian language by analyzing one of the few formal grammar books that exist for that language.

But the real key, says Okonkwo, has been working directly with speaker communities to collect language data and ensure that the resulting AI model accurately represents the many nuances found in any language. This is particularly important in linguistically diverse regions of the world, such as India, where the dialects spoken by people living in villages only a few miles apart can vary widely.

In developing these models, especially those dealing with endangered or underrepresented languages, responsible AI practices play a crucial role. "Responsible AI, when it comes to language technologies, shouldn't be seen as burdening the process. It should be baked into it," says Okonkwo. "We've been focused on how to factor in things like community advisory boards and speaking to people on the ground to make sure that we incorporate those research findings all the way through the model development life cycle."

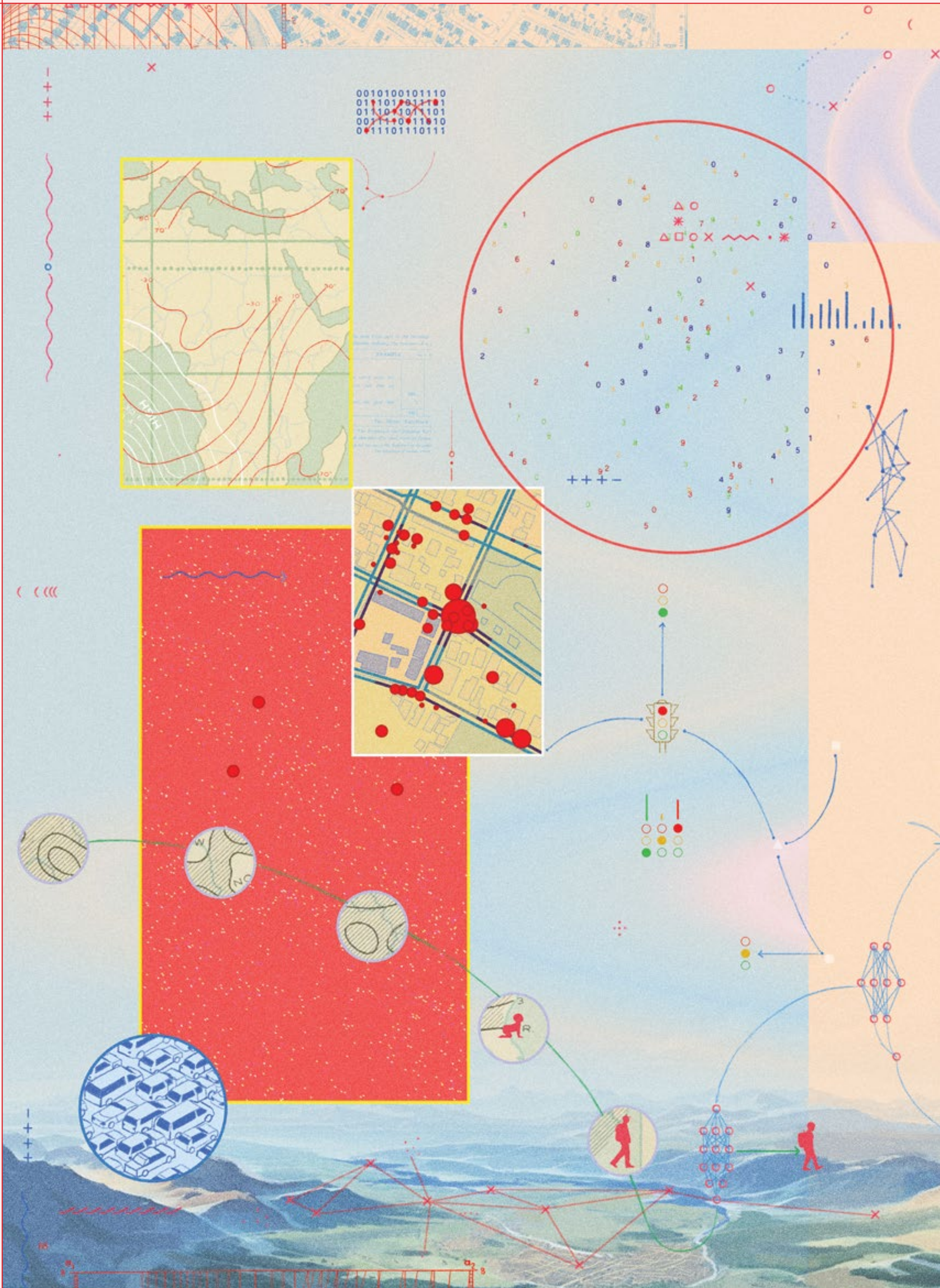
The 1,000 Languages Initiative has already come a long way in making modern technology more accessible to the millions of speakers of underrepresented languages. But what about the speakers of languages that don't fall in the top 1,000? Is using AI to try and save the world's most endangered languages a lost cause? Not necessarily, says Adelani. Most endangered languages will never have enough data to build an LLM as sophisticated as Gemini is with English, but Adelani and many other language technologists are exploring new language model architectures that can be trained on far less data. Another option is to prioritize collecting linguistic data for specific use cases and building AI that is focused on, say, providing medical information in an endangered language.

In the meantime, however, there's plenty of foundational technical work to do. In early 2024, Stanford University launched the Stanford Initiative on Language Inclusion and Conservation in Old and New Media (SILICON), which is working on encoding endangered languages in standard formats to aid their preservation and use in AI applications. Shortly after, a global consortium of researchers and technologists led by Masakhane, a grassroots organization dedicated to natural language processing for African languages, published IrokoBench, a new set of benchmarks for evaluating the performance of AI systems that feature 16 under-resourced African languages.

Useful AI models are still a distant dream for most endangered languages. Still, the current technology can act as an educational tool that provides a bulwark against the further erosion of endangered languages. In 2021, for example, Google Arts & Culture launched Woolaroo, an open-source smartphone app designed for exploring endangered Indigenous languages, as a way to help younger generations engage with their linguistic heritage. In Woolaroo, the user can point their phone at an object—say, a tree—and see and hear the word in more than 10 Indigenous languages from around the world.

The use of AI for preserving and revitalizing the world's endangered languages is full of promise, but it's still too early to know whether it can stave off the mass language extinction predicted by Kornai only a decade ago.

"I don't believe we can save these endangered languages with just AI," says Van Esch. "A language is only saved when people continue to speak it. But if AI can be a tool that makes speaking it easier, then I consider that a job well done."



DISASTER

How artificial intelligence is helping the world’s most vulnerable countries stay one step ahead of climate change

← Illustrations by Daniel Liévano

By Clarissa Wei

Outsmarting DISASTER

CLIMATE

Climate change is often linked to heat, but it's more than just a gradual warming. It's a warping force, distorting environments and putting pressure on vulnerable regions, especially in the Global South, where the strain is far greater, like a rubber band stretched to its breaking point.

Nine of the world's 10 countries most vulnerable to flood risk are part of the Global South. In the tropics, more rainfall and higher temperatures will boost average humidity, creating a dangerous cocktail that will make heat waves even more deadly. These hostile conditions have already triggered a mass migration out of the Global South that's expected to involve about 143 million people by 2050. While historically they've contributed less to global greenhouse gas emissions on a per capita basis since the Industrial Revolution, countries in the Global South will bear the brunt of the economic, social, and environmental effects of global warming.

The stakes are high: move, adapt, or die. How these disproportionately affected nations adapt to climate change could offer critical insights for the rest of the world. Their innovative approaches and adoption of climate-resilient technologies offer valuable lessons for the global community. Artificial intelligence is at the forefront of this technological vanguard. Though it's not a magic wand, it may prove to be a powerful tool if used correctly. Since 2018, Google.org, the philanthropic arm of Google, has been actively supporting organizations working on climate-resilient technologies. When they use AI, many of these organizations have observed significant improvements in efficiency. "Organizations report to us that they're able to achieve their end goals in a third of the time and a half the cost [when using AI]," says Brigitte Hoyer Gosselink, the director, AI & Social Impact at Google.org.

There's a palpable enthusiasm for the technology among these countries. Countries in the Global South express more optimism than any other region in the world regarding AI's impact on job efficiency, information access, and health management, according to a recent Google and Ipsos study.

"I've never seen African policy makers so engaged early on with a new technology like they are with AI," says Strive Masiyiwa, founder and executive chairman of Econet, a South Africa–based telecommunications provider.

He cites the continent's youthful demographic as one of the main reasons for this enthusiasm. "Africa is the

Once-in-a-century floods threaten some 1.81 billion people globally, and more than two-thirds of those people live in South and East Asia.

youngest region in the world. The mean age is under 20. Obviously, you're going to get more buzz and more excitement," says Masiyiwa.

Adaptation

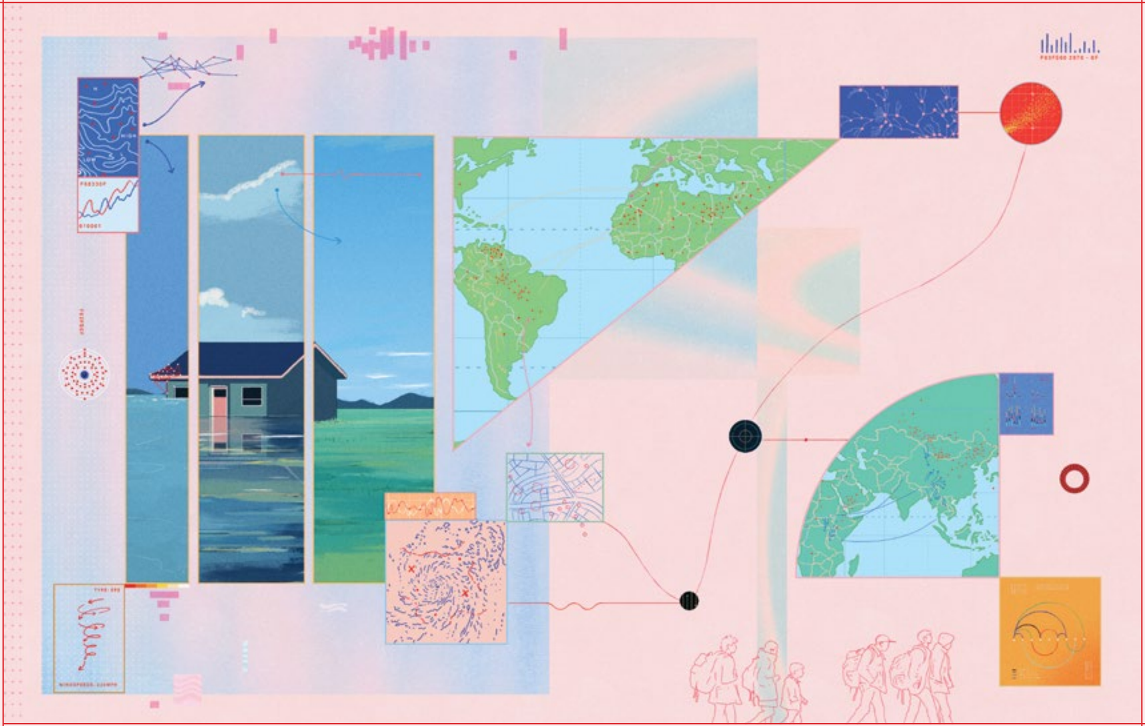
These climate pressures have forced nations to develop innovative adaptation strategies, and with good reason: The rate of flood-related disasters has more than doubled since 2000, according to a report by the Intergovernmental Panel on Climate Change (IPCC), as the sea level in the Red Sea, located between Africa and Asia, is rising faster than the global average and threatening the coastal communities around its perimeter. Once-in-a-century floods threaten some 1.81 billion people globally, and more than two-thirds of those people live in South and East Asia.

Forecasting when and how these floods will hit may help save thousands of lives. Flood Hub is a Google tool that provides forecasts up to seven days in advance for 460 million people across multiple continents. It projects the amount of water flowing in a river, what areas will be affected, and how high the water level will be. In 2021 alone, the system sent out 115 million flood-alert notifications to 23 million people. Building on this momentum, Google Research recently partnered with the United Nations to develop an AI system that can assess post-disaster damage from satellite imagery six times faster than traditional methods, helping humanitarian teams respond more quickly to communities in crisis.

In addition to using AI for flood forecasting, organizations are applying it to map a more holistic view of Earth's waterways. With support from Google.org, DHI A/S, a water engineering consultancy firm, uses AI to understand the nature of wetlands, which are among the world's most important ecosystems.

"Wetlands help improve climate resilience by storing excess water during floods and supporting low flow during drought periods," explained Na'Tosha Bard, chief technology officer at DHI A/S. "Yet most countries do not have a good account of their wetlands. This is a critical information gap, as you cannot manage what you do not measure."

DHI is launching a program called Global Wetland Watch to track wetland ecosystems. The system will use AI algorithms to extract information on the global extent and distribution of wetland ecosystems from satellite imagery, helping to protect and restore these vital ecosystems.



Beyond these larger initiatives, AI is empowering small farmers, who represent the backbone of agriculture in the Global South and are at the frontlines of climate change.

Another Google.org grantee leveraging AI is Wadhwani AI, an institute focused on communities in developing countries to provide growers with the tools they need to manage their crops. "Farmers can upload images of diseased plants, and our AI models instantly identify pest infestations or diseases," says J.P. Tripathi, director of Agriculture Programs at Wadhwani AI. "These solutions provide small farmers with real-time, actionable insights for managing pests, diseases, and environmental challenges that can affect their crops."

Beyond crop management, there's also a need to develop crops that can withstand those erratic variables. Rice, for example, is the most consumed carbohydrate in the world and is mostly grown in developing regions like Southeast Asia, China, and Africa. Google.org supports the work of the International Rice Research Institute, where AI is used to generate climate-hardy rice varieties at a speed and a level of efficiency that has never been seen before.

"AI can be used in the genomic selection approaches to improve the precision of breeding and achieve better genetic gain," says Dr. Venuprasad Ramaiah, the head of Fit-for-Future Genetic Resources and the International Rice Genebank at the International Rice Research Institute. "The development and deployment of high-yielding, climate-resilient rice varieties are expected to generate significant economic benefits for the rice farming community in stress-prone environments."

Climate change–induced migration in the Global South is placing growing pressure on governments, as they struggle to manage the influx of displaced populations and address resource shortages. The Horn of Africa, together with Eastern Africa, saw 1.2 million new disaster-related displacements, according to a 2020 World Meteorological Organization report.

Predicting human migratory patterns is an essential part of mitigating the adverse effects of displacement, and AI is enabling those predictions. One project, led by the Center for Humanitarian Health at Johns Hopkins University, will kick off in Mali and Iraq in late 2024. Researchers will be

As AI continues to revolutionize climate resilience efforts in the Global South, it's clear that technology alone cannot solve the climate crisis.

working on the ground to conduct interviews with refugees and collect as much data as they can on their conditions, reasons for moving, destinations, and distance traveled.

"In terms of migration and movements, we will have multiple sources of data, but they won't be comprehensive," says Paul B. Spiegel, the principal investigator on the project. "And so we want to use AI to be able to examine the data that we have and try to make the various connections. We'll have to train it."

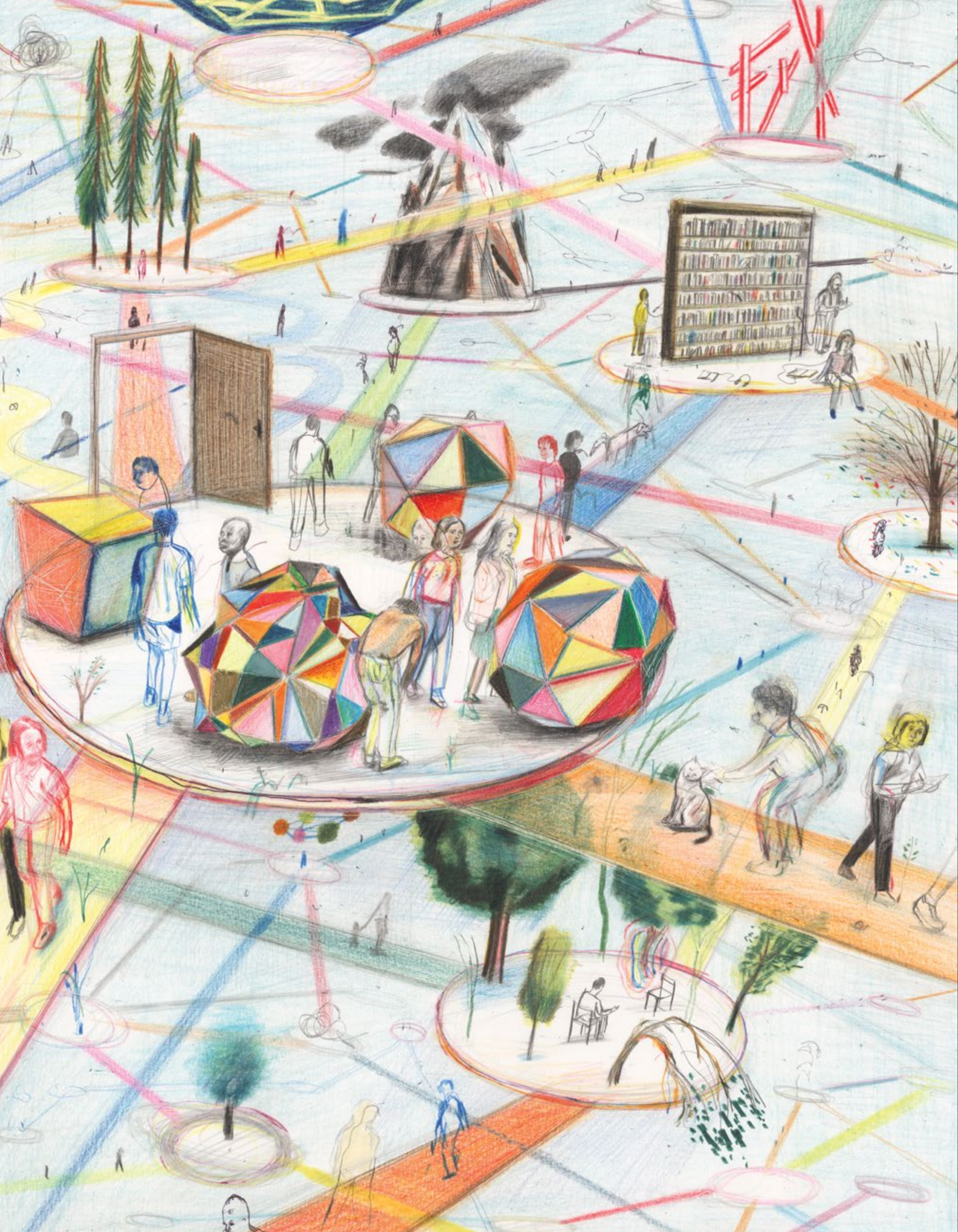
There's a particular urgency for these tools in the Global South, given the compounded challenges of climate change, conflict, and economic instability. "The Global North will be able to adapt better than the Global South because of the systems that are in place and the money that is available to them," Spiegel says, highlighting the contrasts in infrastructure and welfare. "If we have an idea of where people may be moving to, governments will be able to think ahead and plan."

As AI continues to revolutionize climate resilience efforts in the Global South, it's clear that technology alone cannot solve the climate crisis. While these innovations offer the ability to predict and potentially mitigate issues, they also highlight a pressing need for global action and investment.

"At the end of the day, it won't get us past what needs to happen, which is actual investment," Masiyiwa says. "I can get people to get a deeper understanding of the energy transition requirement, what needs to happen, where forests are being destroyed … but that still doesn't run away from the fact that the global community, which is primarily responsible for this crisis, has got to put something on the table."

The story of AI and climate resilience in the Global South is one of innovation born of necessity. But while AI-driven solutions provide valuable insights and tools, they are just the first step.

"We can analyze ourselves to death. It's not the analysis. We know what the issue is," Masiyiwa adds. "Let's put the money in, let's invest the resources." As the world faces one of the greatest migratory challenges in history, technology—no matter how cutting-edge—is just a tool. Fixing the crisis will demand not just innovation, but also dedication and investment.



Supercharging Human Curiosity

A new suite of artificial intelligence models empowers our desire to learn.

By Anya Kamenetz

←
Illustrations by Yann Kebbi

The 20th president of the United States, James Garfield, studied at Williams College in the 1850s with a renowned philosophy professor named Mark Hopkins. Later, Garfield famously quipped, "The ideal college is Mark Hopkins on one end of a log and a student on the other." This aphorism distills learning to the simple joy of Socratic dialogue. Some have begun to consider AI a partner in this, one of our key endeavors as a species. "We're inherently curious creatures," said Ben Gomes, SVP, Learning & Sustainability at Google. "Learning is what makes us human, what makes us special."

When I spoke with Gomes about how the company is approaching learning, he displayed the boyish demeanor of someone who is perpetually fascinated by life. He's prone to doing deep dives on far-flung topics, from geology to woodworking. He views AI as a form of "augmented intelligence"—technology that entices more people to do the same, by making the discovery of knowledge more accessible and engaging. Lately, Gomes has been guiding the effort to install a foundation of pedagogy and responsible learning into Google's suite of AI tools, including LearnLM. Google's internal pedagogy and learning sciences team conferred with a range of experts, such as researchers and practitioners, to identify fundamental principles that support human curiosity. The key findings were that the system must be encouraging, make learning as active as possible, avoid overwhelming people with too much information, and adapt to the learner's goals and needs.

Like a tinkerer who's been busy in his workshop, Gomes showed off a new Google Labs experiment called Learn About, which breaks down and explains large and complicated topics. After typing in a prompt, "Tell me about the history of writing," I see an AI-generated timeline, with images and explanations of cuneiform, Linear B, and the Rosetta stone. A related YouTube video is highlighted. By hitting buttons, you can "Simplify," "Go Deeper," or "Get Images." I'm guided to related questions, such as "How were hieroglyphics translated?" Instead of being confronted with blocks of text and links, I feel as if I'm walking into a well-curated multimedia museum exhibit that has a deliberate flow, that orients you to a topic and gives you multiple paths through it. Gomes said the intention was to organize large topics better and present them more richly at the same time.

One of the most effective ways to learn is also one of the oldest: a conversation. Google is working on a feature that will allow you to "raise your hand" and ask a question during a YouTube video. The tool will generate an answer to your question based on the material that's being discussed. Another tool in development, called Illuminate, can transform any academic paper into an audio output that sounds a lot like a conversation. Illuminate doesn't just turn text into speech; two different voices are reacting to and discussing what's important, new, and interesting in the paper. So a doctor, a Ph.D. student, or just a curious civilian can follow along and keep up with the latest research in any field while doing the dishes.

A surprising aspect of simulated conversation is that some people report being more comfortable asking an AI a question than they do a human, perhaps because there's no chance of judgment or embarrassment. This dynamic highlights how a significant part of the learning experience is guided by emotions, not only by intellect. Learning can be daunting; everyone appreciates encouragement. But what it means to be "encouraging" as a pedagogical principle remains a balancing act; one that is fully on display with ShiffBot, a collaboration between Google and Daniel Shiffman, a computer science professor at NYU who has a popular YouTube channel. ShiffBot is a virtual coding tutor that runs in the web browser alongside the code editor that Shiffman uses in the classroom and in his videos. The ShiffBot chats in a similar tone to Shiffman's: warm and supportive. When students get stuck, ShiffBot can point them to the moment in a video when Shiffman discusses a similar issue. One explicit goal in designing this interface was to make the conversation with ShiffBot "a safe space to make mistakes." And not only for learners. When he's teaching with the coding tools in his videos, Shiffman himself sometimes triggers error messages, hits bugs, or makes mistakes in his videos, which he handles with good humor. The bot, too, will inevitably go astray at times. So the goal is to learn from mistakes, no matter where they come from—a good life lesson, not just a good teaching strategy.

While it's typical to think of learning as confined to the classroom and formal education, AI joins previous information technologies to make learning more accessible to people around the world and at all stages of life. Anna Mills is a community college writing instructor in California who also works as an adviser to tech companies. She lauds the "delightful ways" that self-directed adults can turn to AI as a learning companion. "As a lifelong learner, I feel like I have more freedom to experiment," she says. "For example, last night I was listening to a story by [Jorge Luis] Borges and I needed more context. So I could talk to an AI about his biography and politics." Mills says she also uses AI to challenge herself and assess her knowledge. She might upload a scholarly article she's reading and then prompt the AI to ask her questions about it—again, simulating a conversational mode.

Mills works with fellow educators to develop their AI literacy and apply it to their own pedagogy. She says the ultimate goal for herself and her fellow teachers is to empower students to engage in AI as lifelong learners: "The idea is to spend a bunch of hours with it, try a million things. Be ready for it to fail, but continue to keep an open mind to it as a thought partner." This approach includes making students aware that the technology has flaws and shouldn't replace the enjoyment of, say, talking over what you're reading with a friend. "It can extend rather than replace the things we would otherwise do to build our thinking," Mills says.

The strategy of putting learners in the driver's seat and making learning for its own sake even more appealing could be AI's most promising educational application. This is a classically progressive stance that views AI as a new palette with which the user can paint whatever picture she wants.

> "We want to bring back the pleasure of learning. The feeling where it clicks and everybody has that joy."
>
> **Ben Gomes,**
> SVP, Learning & Sustainability at Google

"Clearly, AI is opening up an entirely new world for education," said Andreas Schleicher. He's the director for Education and Skills at the Organisation for Economic Co-operation and Development (OECD). In this role, he works with government ministers and education leaders worldwide to improve quality and equity in education. The key, he said, is ownership. "AI can give learners greater ownership over what they learn, how they learn, where they learn, and when they learn."

The learning potential of AI will be realized, Schleicher believes, only when it's paired with the best ideas and educational practice. For a reality check, he pointed out that the results of PISA, the international exam program administered by the OECD, demonstrate that classrooms using the most technology often have lower outcomes in core subjects. "That's not necessarily a bad reflection on AI, but on our capacity to use it effectively," he said. "AI is not a magic power. AI is just an amazing accelerator and an incredible amplifier. It will amplify good ideas and good educational practice in the same way it amplifies bad ideas and bad practice."

Just as Google has done in building Google Classroom over the past decade, the countries leading in the

integration of AI—Korea, China, Estonia—are working closely with teachers from the outset in AI development, Schleicher said. "These countries all stand out by having teachers not just implement AI-based tools but be actively engaged in their design and the research around them." Along the way, teachers are modeling a growth mindset for their students. "Educational institutions need to become organizations where everyone learns from and with each other, and that's exactly what we see in the countries advancing fastest." Engaging teachers in the design and development of AI tools, Schleicher says, will be essential for implementation.

*

One of the overlooked aspects of AI and learning is, clearly, related to the emotional component of learning. Specifically, joy. "AI can make learning fun," Schleicher said. This is a big theme for Gomes as well: "We want to bring back the pleasure of learning. The feeling where it clicks and everybody has that joy."

The pursuit of fun is also something that guides Sal Khan, who built Khan Academy, a nonprofit library of educational videos that now has 165 million registered users. Khan Academy is piloting Khanmigo, an AI-powered tutor and teaching assistant. He told me that a subtle dimension of the work has been developing a voice for AI instruction that inspires curiosity: "How do you motivate the student, how do you pull them into the content? How do you keep them engaged?"

To be sure, there are economic and social barriers to educational access that can never be overcome with technology alone. But, at least with this set of pedagogically informed tools and applications, the potential for AI to empower lifelong learners is clear. The information in libraries and archives has become nearly infinite and universally accessible, through platforms like Google Search and YouTube, while this new technology is adaptable and conversational and awakens a new sense of wonder.

Gomes and other close observers and creators in the world of AI hope that the rising generation of applications allows even more people to learn independently, and find deeper enjoyment in the process. "I hope for a world—and I feel this myself—where people find the whole world interesting," Gomes said. "This unbelievable oyster with pearls of knowledge that people are eager to go after." It is true that humans and algorithms learn differently; humans first experience the physical world with their senses, then learn language, whereas machines start with language. AI has the potential to help us appreciate our own learning and ability to understand—a trait that is uniquely available to us as a species that no technology, no matter how good it gets, can ever replace.

Music Sets the Stage

←
Illustrations by Raven Jiang

By Craig Marks

[
In many ways, musicians are at the forefront of AI technology that is changing how their art is produced, distributed, and monetized. Will other creative disciplines follow suit?
]

Kyle Lehning makes for a pretty unlikely AI pioneer. He's a genial, white-haired, 75-year-old music producer based in Nashville. It was 1985 when he was introduced to a fledgling singer by the name of Randy Ray who would soon change his stage name to Randy Travis. With his rich, sonorous baritone and preternatural phrasing, Travis would become one of the most revered country music singers of all time, and Lehning would produce or co-produce nearly every track by Travis during his storied career, which included 16 No. 1 Billboard country hits. In 2013, their 30-year professional relationship came to a tragic halt when Travis suffered a near-fatal stroke. It would take him years to learn how to walk again, even longer to regain a limited ability to speak, and performing live or recording new music was no longer possible. One of country music's greatest voices had been silenced.

That is, until July 2023, when Lehning received a call from Cris Lacy, co-chair of Travis's record label, Warner Music Nashville. "Pretty quickly into the conversation," he recalls, "Cris asked me, 'What would I think about getting Randy's voice back using AI?' And my immediate response was, 'That sounds pretty creepy to me.'" Lehning says that he'd paid little attention to the headlines about the rise of AI, and while he was intrigued by Lacy's pitch, "it all boiled down to what Randy and his wife, Mary, thought about this."

When they signed off on the idea, Lehning was put in contact with a British company called Myvox and sent it 42 Travis vocal tracks for its machine to learn from. Next, Lehning dug up a song that he'd co-produced in 2011, by a singer named James Dupré. "I'd always loved James's 'Where That Came From,' but for one reason or another, it was never released." The plan was simple: What would it sound like if vintage Randy Travis covered "Where That Came From"? When Lehning first heard the AI Randy, he was stunned: "It wasn't a weird-sounding, robot version. It was Randy Travis's voice."

After a couple months of polishing, he sent the finished version to Randy and Mary, who were thrilled with the result. The track, which was released in May and cracked the top 40 of Billboard's Hot Country Songs chart, even spurred a return, of sorts, to the stage for Travis: He will take part in a concert tour in which James Dupré and Travis's original touring band will perform Travis's songs, with Randy and Mary, in nonperforming roles, acting as guiding spirits and hosts.

We live in an era of widespread anxiety over AI's role in content creation. There are concerns that AI will replace graphic designers, write journalism, be an art critic, and even host podcasts. But what you find when you look closely at the creators who are the early adopters of this technology is a less drastic and more nuanced picture. Yes, the technology is powerful, but it can also be quite dopey: It doesn't seem that good at writing song lyrics, for example. Yet it does act as a spur to creation, as a way to sketch out a project or imagine an alternate headline or a different way of composing a scene. The AI is a tool without judgment: It will attempt to suggest, analyze, or reconfigure any scenario that you throw its way. Most importantly, AI is not what gives a piece of art its value. It's the human who made the art, who imbued the artwork with their own distinct identity and vision.

Travis's "Where That Came From" is about as warm and fuzzy an example of the technology's power that you'll find. Crucially, the song was made with the approval of the artists (Travis, Dupré) and the various copyright holders (the record label, publishing companies); it was lovingly and painstakingly put together; and it could not have existed without AI. Don Was, the president of the venerable jazz label Blue Note, is a fan of the Travis song, and of AI's potential. "If you have an emotional attachment to Randy Travis's voice, which millions of people do, you got to hear him sing something new," Was says. "It really got me choked up."

Most applications of AI in music-making aren't this blue-sky, however. The ability to mimic a popular artist's voice or style or to create songs lickety-split and en masse, simply by using text or voice prompts, is fraught, to say the least. In April 2023, the music industry was gripped with

The AI is a tool without judgment: It will attempt to suggest, analyze, or reconfigure any scenario that you throw its way.

panic when a song called "Heart on My Sleeve" went viral. The AI-generated track, created by a mysterious entity known as Ghostwriter, sounded, quite intentionally, like a collaboration between megastars Drake and the Weeknd. The Universal Music Group (UMG), the world's largest record company and home to both artists, quickly had the "Fake Drake" song removed from streaming platforms due to copyright infringement, but not before it prompted fears that flesh-and-blood musicians were about to be supplanted by AI deepfakes. "The talk was, AI's going to knock off all the popular music in the world, and there will be no copyright associated with it at all, and the business model of the music industry is going to disappear overnight," says Michael Nash, executive vice president, chief digital officer of UMG. "Well, it's a year and half later, and that didn't happen."

Still, the seeds had been planted for disruption to an industry that is regularly tossed about by new technologies: Innovations from synthesizers to sampling and Auto-Tune were all considered extinction-level threats to skilled musicians, and in the early 2000s, Napster and peer-to-peer file sharing nearly decimated the recorded music business. "There is a long continuum of music technologies creating both great excitement and also great panic," says Mark Katz, a professor of music at the University of North Carolina. "The long view is that these upheavals are neither utopian nor apocalyptic."

Was recalls buying one of the first LinnDrums, in 1982. "This was the first commercially available drum machine that used digitally recorded samples of real drums," he says. "Every drummer in the world thought this was, like, doomsday, man, the end of drummers. I bought serial number 003. Prince had been to the same store a couple of days before, and he bought serial number 002. And when you listen to 'When Doves Cry,' you hear an artist taking this new tool that people found threatening and creating a drum part that no human drummer would have played. He changed music with that. And it didn't put drummers out of business."

Dozens of AI music companies have popped up in the past couple of years. Nearly all offer a similar promise: to make the often-arduous process of creating songs simpler and more efficient and to lower the barrier to entry for amateurs. Some permit the user to mimic the style of established artists who have been compensated for allowing their musical DNA to be mapped and replicated; others provide the tools for the AI-curious to create an instrumental with a few clicks of a mouse.

Brian Transeau, better known as BT, is a popular electronic music composer and DJ and a co-founder of the assistive AI company SoundLabs. Its first product is a plug-in called MicDrop, which, BT explains, "allows you to change your voice into any other human voice in real time, and also to change your voice into instruments." Female-sounding voices can be transformed into male-sounding voices; a vocal can be turned into a trumpet line. And if you can't sing, you can still generate a studio-quality vocal. "Let's say you're a dance music or hip-hop producer and you want a vocal for your record," he says. "It's as simple as talking into your iPhone and playing a couple notes on a MIDI keyboard, and you can make a vocal that you can use on a record or that you could play to a singer and have them re-sing it." The idea is to "help reduce friction in the compositional process and aid in creativity, without replacing humans."

The software program has been "ethically trained," which is to say, the inputs—the voices used to train the

Everybody realizes that GenAI is here, and it's up to us to shape it.

Otherwise, it will shape us.

Lyor Cohen,
Global Head of Music at YouTube

machine—were provided with the consent of artists. In June, UMG announced a strategic agreement with SoundLabs, wherein UMG artists can choose to supply their voice data for training while "retaining control over ownership and giving them full artistic approval and control of the output." One of the first collaborations between the companies is an AI-generated Spanish-language version of Brenda Lee's holiday classic, "Rockin' Around the Christmas Tree." Originally recorded by a 13-year-old Lee in 1958, the 65-year-old song, thanks to seasonal streaming, topped the Billboard Hot 100 in December 2023, making Lee, at age 78, the oldest artist ever to do so. To find new audiences and recapture that top spot, SoundLabs trained MicDrop on Lee's 1958 vocal track and merged it with a work-for-hire vocalist performing it en Español, and out popped a teenage Lee singing her seasonal classic in Spanish.

This translation of songs into other languages could be a boon for artists looking to expand their global fan base. "If we have an artist touring in Manila, for instance, and they can release a song in Tagalog, that could be huge," says Carletta Higginson, executive vice president, chief digital officer for Warner Music Group. Nash, her fellow music executive, cites the American artist Lauv, who is popular in Korea and used voice-to-voice technology to create a single for his Korean fan base. "We're just at the beginning of what is possible," Nash says.

Post Malone's longtime producer, Louis Bell, is part of a select group of musicians in YouTube's Music AI Incubator who have been field-testing Google DeepMind's music AI tools. Bell said he's been using AI to help generate songwriting ideas. "It's the modern-day version of crate digging," he says, referring to the old-school practice of hip-hop DJs foraging in record-store bins for songs to sample. "AI gives you billions of more options now." He also senses that the technology will need guardrails for all the creative arts: "It's kind of like they have to create some kind of Geneva Conventions for AI, where everyone comes together and figures out what is going to be our path moving forward."

The producers I spoke with tended to view AI as a new step in a procession of technologies that have transformed music creation—and not as a threat to music itself. At heart, popular music is driven not just by songcraft and technique, but also by the personas of the artists, their narratives, their points of view. Taylor Swift fans, to choose the most outsize example, are emotionally invested in even the most quotidian details of her life, and they come together at her stadium concerts to share that bond and to see themselves—their longings, their heartbreak—in her songs. Despite the scare over "Fake Drake" and other replicas, AI won't take the place of Swift or Bad Bunny or Adele.

However, there is a lot of music, heretofore created by people, that AI can generate more efficiently and at greater quantity, without listeners noticing the difference. Think of the soothing background music you might hear when entering the lobby of a boutique hotel, or the ambient lo-fi songs you might stream to help you sleep or study. This "functional" music has no star at the center of it—anonymity is a feature, not a bug—and it's a big business: A 2023 report said that functional music accounted for 15 billion streams a month across all music platforms.

A dozen years ago, Alex Bestall started Rightsify, a company specializing in creating and licensing songs to be used in advertisements and film and video-game soundtracks and played in hotels, airport lounges, gyms,

and retail stores. He and his team commissioned millions of tracks, 95 percent of them instrumental. When AI started to make inroads in content generation, Bestall foresaw the future: “My take was, music is next. And that our kind of music was a prime suspect to be disrupted. So we dove right in.” Beginning in early 2023, Bestall licensed his catalog to multiple AI companies that needed songs in bulk to train their machines. Those licensing fees now make up 60 percent of Rightsify’s revenue and have even led to Bestall hiring additional musicians and contracting more work-for-hire instrumentals to meet the demand of the AI companies.

The functional music space offers an important lesson for all content creators: The AI itself will become a sort of platform, and they can find new ways to engage with its capabilities. I think of the artist Brian Eno, who has played around with chance and randomness in his work and has been experimenting with generative music since the 1970s. More recently, the economist Tyler Cowen wrote a book titled *GOAT: Who Is the Greatest Economist of All Time and Why Does It Matter?* that was meant to be analyzed and then queried by readers. Will this approach replace traditional books? Of course not. What it offers is an interactive way to learn about economics that might appeal to some. Every major technological shift presents new challenges for artists, both on the level of the individual artwork and in how the art will be seen or heard or experienced by the public.

The coming year could be telltale in charting AI’s place in the music ecosphere. The Nurture Originals, Foster Art, and Keep Entertainment Safe (NO FAKES) Act, legislation backed by various artist-advocacy trade groups and aimed to protect the voices, names, and likenesses of creators from being used nonconsensually, has received bipartisan congressional support. Among creators, there is a strong desire for a legal framework. Lyor Cohen, YouTube’s global head of music, who ran major record labels during Napster’s insurgence, believes that his “colleagues in the music industry are much more inclined to be on their front foot” with regard to AI. “Everybody realizes that GenAI is here, and it’s up to us to shape it. Otherwise, it will shape us.”

Q&A

What's Next for the Creator Economy?

A conversation with Neal Mohan, the CEO of YouTube, about how the platform is helping creators use the power of AI in their storytelling

By Haimy Assefa

Creatives across the globe are leveraging generative AI in their chosen fields. According to a recent survey by Radius, a majority of them are also seeking guidelines as they harness this powerful tool. Neal Mohan, CEO of YouTube, is at the forefront of this transformative era in digital media and creativity. Mohan is focused on how YouTube's AI tools can help creators level up, while offering guardrails creators need to use AI responsibly. We spoke about the challenges and opportunities of the gathering momentum of generative AI.

Haimy Assefa **As a leader in the tech industry, how has your perspective on creativity changed in light of recent advancements in AI?**

Neal Mohan I've been in this business a long time and witnessed some massive shifts since my career started: print media to online, desktop to mobile, traditional broadcast to streaming. The GenAI revolution is now here and poised to supercharge the creative landscape in unprecedented ways. It's helping creatives bring big ideas to life, opening doors for more people to thrive in the creator economy, and delivering higher-quality content to viewers globally. Although we sit at this huge inflection point, the one thing that hasn't changed is that the stories that break through are human stories. So, as the world of AI continues to evolve our understanding of creativity, I see these new technologies as enhancing and unlocking human creativity, not replacing it.

Assefa **Looking back on your career, what past experiences or lessons have best prepared you to lead YouTube through this transformative period in the creative landscape?**

Mohan I've spent my entire career at the intersection of technology and creativity. First at DoubleClick, then at Google, and now running YouTube. I'm a huge believer in the power of digital tools to democratize storytelling and provide new paths for creatives to fund their dreams. Nowhere is that more true than YouTube. Every day, I get to talk to creators who are reaching an audience and building a business on YouTube. And no one can tell them that they can't share their story because they don't look the right way, aren't the right gender, or live in the wrong part of the world. That is incredibly powerful—it's what drew me to YouTube and what has helped prepare me as we embrace the transformative power of AI.

Assefa **What advice would you give to young creators who are just starting out in this new era of AI-augmented creativity, and what skills or mindsets do you think will be most essential for them to succeed?**

Mohan I would encourage them to remember that on YouTube, success isn't one-size-fits-all, and what makes our platform special is the authenticity of our creators. AI is bringing a transformational shift in technology, and viewer tastes always change, but they will always connect with authenticity. When I talk to creators about what they're excited about when it comes to AI, it's about the potential to make their everyday creation process easier. For example, generating the perfect title and thumbnail for YouTube is critical to a video's success but can be a time-consuming process of brainstorming and iteration. AI tools can help with that process. It's long been true that creators are quick to embrace new technologies, and that is where I believe a lot of the innovative AI uses will come from. A recent survey by Radius found that 92 percent of them are already using GenAI tools. They are really at the cutting edge of where this technology meets the creative industries.

Assefa **According to the survey, 84 percent of creators believe that independent creators will rival big studio productions in the next two to three years. What's your take on this shift, and what could it mean for the industry?**

Mohan I think we're already starting to see it. Take Mythical Studios. They launched their flagship show, *Good Mythical Morning*, a little over 10 years ago. It now has more viewers than all the popular late-night shows combined. They employ over 100 people, have several sound stages—just like you'd expect to see in a TV studio. They have branded imprints, a venture fund for up-and-coming creators, and just this summer, [they] launched a scripted, episodic series. And that's just one example. We're going to see more and more of this. Next up is making sure these studios are getting the recognition they deserve, like Primetime Emmy Awards. Viewers already don't make a distinction when they turn on the TV between legacy-produced media and creator-led content. The rest of the industry will start to catch up.

Assefa **YouTube has paid over $70 billion to creators, artists, and media companies over the past three years. How do you see your relationship with creators evolve as many of them continue to leverage AI?**

Mohan Creators are the heart of YouTube. Every time I talk to our teams and share our business strategy, I say the same thing: It all starts with creators. It's built into the very DNA of our company: Our unique revenue-sharing model means that we only succeed when our creators do. As they continue to embrace AI, we're focused on ensuring they can find the latest creation tools on YouTube to help bring their visions to life on their terms and make a living in the process.

Assefa **How is YouTube working to protect creators in this new landscape and ensure that AI tools are used responsibly?**

Mohan We're committed to protecting creators and ensuring they thrive in this evolving landscape. And creators themselves are calling for clear lines and responsible use of AI. In a recent survey we conducted, 74 percent

“We’re committed to protecting creators and ensuring they thrive in this evolving landscape. And creators themselves are calling for clear lines and responsible use of AI.”

Neal Mohan, CEO of YouTube

of creators said they wanted guidelines for responsibly posting GAI (generative AI) content to video and social platforms. One of the key ways we’re doing that is by investing in the right guardrails. For example, we’re developing new technology to help creators (and other members of the creative industries) to detect and manage AI-generated likenesses on YouTube. We’re also looking to provide transparency to viewers: We require creators to disclose when they’ve created realistic content that is altered or synthetic, including by using GAI, and we disclose that in a label. For sensitive topics like news or elections, that label appears right on the video itself. This transparency builds trust between creators and their viewers. Tools like GenAI dubbing are helping creators share their content with people around the world.

Assefa **How do you think storytelling may change as global audiences become more accessible to creators?**

Mohan YouTube reaches billions of users in over 100 countries and across 80 languages every month, but some creators can struggle to reach audiences on a global scale. A huge benefit where creators are already seeing growth in their businesses from AI is in expanding their global reach. New GenAI tools are helping creators to automate translations, captioning, dubbing, and other time-consuming tasks that dramatically slow down content production but are critical to reaching and connecting with global audiences. Our research showed that 87 percent of creators believe GenAI will likely enable creators to more easily and effectively export their content to countries around the world with GenAI-dubbing over the next two to three years. This is an area where YouTube sees huge potential to help creators grow their businesses, and we recently announced that we’re expanding our auto-dubbing features into more languages to help support this further.

Assefa **Ultimately, what do you hope the legacy of YouTube will be in the age of AI?**

Mohan It’s incredible to reflect on YouTube’s journey from a video-sharing platform to one of the world’s largest streaming platforms, home to millions of creators whose boundless creativity has sparked new ideas and industries, and innovations for billions of people. We’re now layering on the transformative power of AI to that journey. So I hope our legacy ties back to our mission from our earliest days to enable millions of creators around the world to share their voice, connect with global audiences, and build thriving businesses—with AI playing a critical role in reaching these goals.

A New Lens on AI

Sparking public dialogue on how to visualize AI

Plant Market
Farah Al Qasimi, 2024

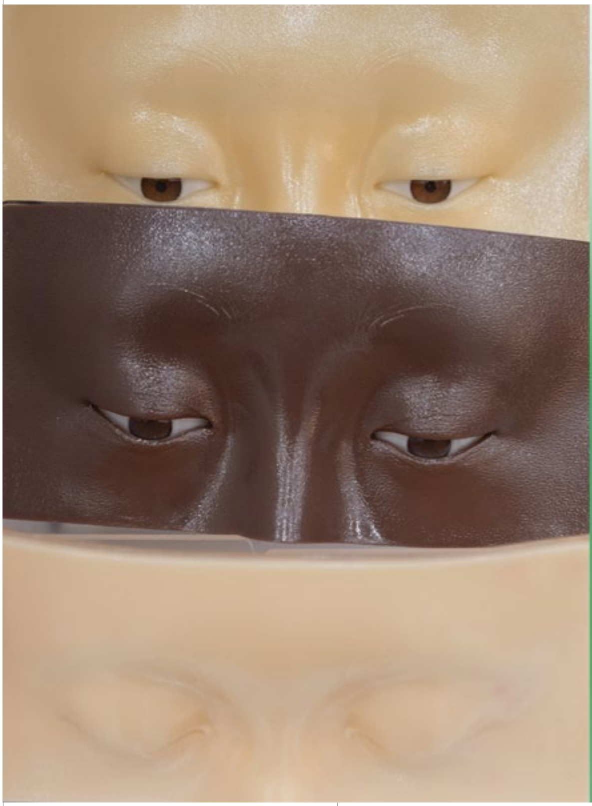
In January 2024, an AI-focused interdisciplinary team at Google set out to reshape the narrative around AI imagery by commissioning three artists—Charlie Engman, Farah Al Qasimi, and Max Pinckers—to create speculative photography imagining the near-future impacts of this world-changing technology. The goal was to move beyond conventional, abstract visual stereotypes and to center people; to evoke new ways of thinking about AI by focusing on human relationships and societal outcomes rather than depicting specific technologies. Through speculative photography—which demands context—the project emphasizes diverse artistic perspectives and opens up broader public discussions on AI's role in society.

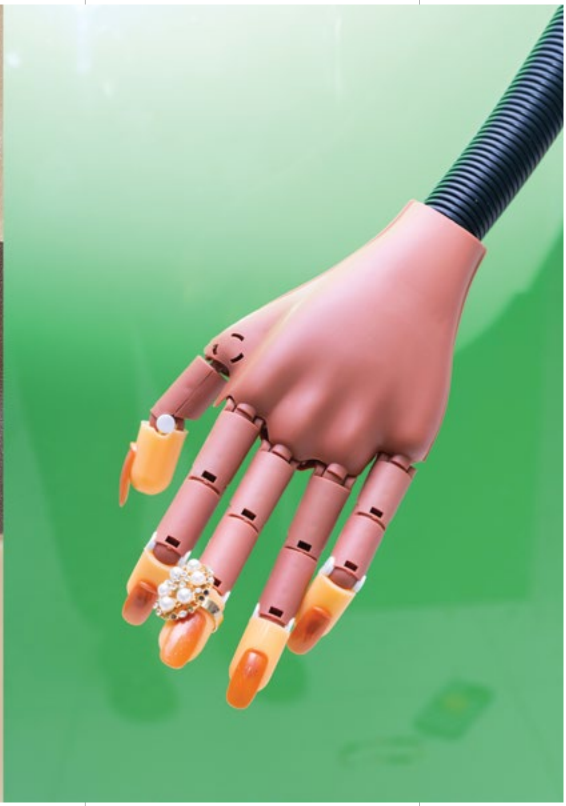
↓
Learn more about the artists and artworks with the ICP on Google Arts & Culture:

Farah Al Qasimi

On the artist: Farah Al Qasimi creates photographs, films, and music, often with large-scale vinyl imagery and varied prints. Her work explores internet hierarchies and storytelling in children's cartoons, and many of her video works feature anthropomorphized narrators. Her collaborative practice has included puppets, falcons, African land snails, exorcists, and a Jack Sparrow impersonator.

More about the work: "These photographs are about seeking connection and intimacy through enhancing our knowledge about the world around us. I am imagining the ways in which our understanding of the natural world—and of each other—might surpass language or description and instead take the form of a feeling or a knowing. For example, I cannot speak with my dog, but I might know through her behavior if she is unwell, and I can seek greater clarification with medical imaging and diagnostic tests. But I can also dream of a world in which care and empathy are expanded in previously impossible ways. These photographs do not attempt to overstate the presence of technology; rather, they seek to show how we may form exploratory connections that expand beyond the limitations of a screen."





Body Shop
Farah Al Qasimi, 2024

Kantamanto Registry
Charlie Engman, 2024

Charlie Engman

On the artist: Charlie Engman, an artist, educator, and art director at sustainable fashion brand Collina Strada, blends principle with irreverence in his work, pushing the limits of traditional image-making. A leader in AI-generated art, Engman has a diverse portfolio that includes projects for *The New Yorker* and *Vogue*, as well as for brands such as Prada and Gucci. He has authored several books and holds a degree in Japanese and Korean studies from the University of Oxford.

More about the work: In January 2024, Charlie visited Accra, Ghana, to speak to locals about their perceptions of AI. Most of the people Charlie interviewed were connected to his work with the Or Foundation, an NGO working at the intersection of environmental justice, education, and fashion development; many of the conversations centered around the environmental and socioeconomic impacts of waste colonialism and global commerce. Charlie employed Midjourney to create the images, using photographs he shot on location as prompts. The people in the images are digital representations of fictitious Ghanaians engaging in hypothetical scenarios, as envisioned by the interview participants.

Max Pinckers

On the artist: Max Pinckers produces images that are lively, poetic, and unexpected, blending research, preparation, and improvisation to push the boundaries of documentary photography. He holds a doctorate in the arts, teaches at the Royal Academy of Fine Arts (KASK) in Ghent, Belgium, co-founded Lyre Press and the School of Speculative Documentary, and has won several prestigious awards, including the Leica Oskar Barnack Award.

More about the work: By infusing everyday scenes with speculative elements and surreal interventions, this work creates a space where the presence of AI models is both palpable and enigmatic. These photographs serve as portals into a world where the lines between reality and imagination blur, inviting viewers to contemplate the evolving role of AI in shaping our perceptions and interactions with the external world. Through staged compositions and symbolic imagery, this series of images seeks to illuminate the nuanced complexities of our relationship with AI, sparking new insights and perspectives along the way.

Double Take
Max Pinckers, 2024

CHAPTER

3

Breaking through

Intro

In laboratories and research institutions the world over, AI is accelerating the pace of scientific discovery. Advancements are being made quickly: AI systems are predicting the intricate shapes of life-saving proteins, mapping the complex networks of the human brain, and revolutionizing materials science by designing novel compounds with tailored properties.

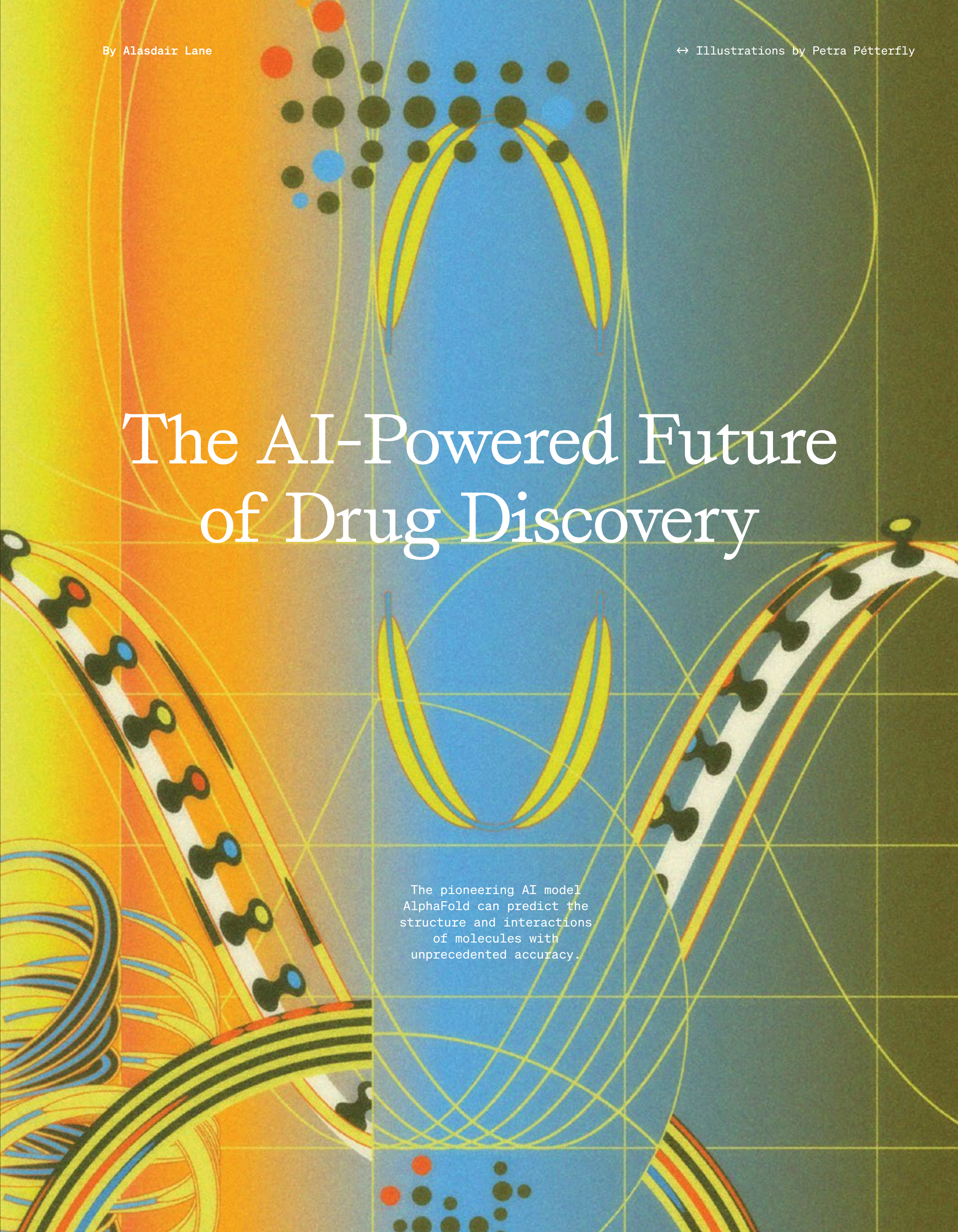
In fields ranging from chemistry to astrophysics, these technologies are opening new frontiers of knowledge. Perhaps most urgently, AI is enhancing our ability to address some of humanity's most pressing challenges: Advanced climate models, powered by machine learning algorithms, offer unprecedented insights into the Earth's natural systems, and AI-accelerated drug discovery may enable us to respond to pandemics with greater speed and precision. Additionally, the marriage of quantum computing and AI could help solve problems once thought intractable.

These scientific advancements raise profound questions about the nature of discovery. As AI systems are used to make predictions and generate hypotheses at superhuman speeds, will the very principles of the scientific method change? What is the role of human intuition and creativity in a world where machines process and analyze data at unimaginable scale? The stories in this chapter bring into sharper focus the potential of scientific inquiry in the age of AI.

By Alasdair Lane

↔ Illustrations by Petra Péterfly

The AI-Powered Future of Drug Discovery



The pioneering AI model AlphaFold can predict the structure and interactions of molecules with unprecedented accuracy.

At the heart of every biological process, from the replication of cells to the functioning of the human brain, lies the intricate world of proteins. Understanding the structure of these molecules, folded into complex three-dimensional shapes, is crucial to advance our collective understanding of biology and facilitate the development of new therapeutics.

AlphaFold is an artificial intelligence system that revolutionized our understanding of protein structures advancing research and drug discovery. The project was led by Google DeepMind CEO Demis Hassabis and Director John Jumper.

Drug development is an extraordinarily complex and time-consuming business. The formidable goal of treating disease demands unwavering scientific rigor across a sprawling set of variables and unknowns, in addition to huge amounts of funding. Still, it can be difficult to imagine just how demanding the creation of new medicines is. So, to illustrate the impact of AlphaFold, Jumper often tells a particular story.

"I went to a talk by someone who had designed a very interesting drug in clinical trials at a major pharmaceutical company," Jumper recounts. "They described this beautiful molecule they had made and all the chemistry they had done to solve different problems with their molecule. Someone asked them afterward, 'How many different molecules did you make to come up with this one?' And he said, '5,000 over a seven-year period.'"

The story illustrates a fundamental fact: Science is an iterative enterprise. Trial and error in the laboratory—a word befitting its Latin root to labor—is how progress has traditionally been made. But now, AI systems are poised to dramatically accelerate this painstaking process.

ALPHAFOLD APPLICATION

For decades, scientists puzzled over the protein-folding problem of whether a protein's structure and shape could be determined based on that protein's amino acid sequence. In 2020, AlphaFold made a fundamental breakthrough: predicting the structure of proteins with a level of accuracy that had never been seen before in competitions run over the years on the Critical Assessment of protein Structure Prediction (CASP). Together with the European Molecular Biology Laboratory's European Bioinformatics Institute, Google DeepMind shared these predictions for 200 million proteins in a free database that's been used by millions of researchers globally.

When Google DeepMind's AlphaFold was recognized by CASP organizers as a solution to the

50-year-old protein-folding problem, it represented a watershed moment for computational and AI methods for biology. AlphaFold demonstrated that AI techniques were advanced enough to be applied to real-world problems, including scientific discovery. In 2021, Hassabis founded Isomorphic Labs as an offshoot of Google DeepMind to specifically focus on rational drug design and apply AI to redefine the drug discovery process from first principles.

The latest version, AlphaFold 3, developed by Google DeepMind and Isomorphic Labs, is reshaping the field of structural biology and providing invaluable insights for drug design. The third iteration of AlphaFold expands beyond proteins, enabling it to predict the structures of a wide range of biological molecules—DNA, RNA, ligands, and more—with unprecedented accuracy. More importantly, it can model how these different types of molecules interact, a key piece of the puzzle for understanding disease and developing treatments.

Hassabis is also the CEO of Isomorphic Labs, which is applying these advancements to the field of drug discovery. The preclinical phase of drug development has traditionally relied on traditional experimental methods—X-ray crystallography and cryo-electron microscopy (cryo-EM)—to determine the structures of target proteins and potential drug molecules. Using X-ray crystallography, scientists must coax proteins to form crystals, then bombard them with X-rays to deduce their structure. Cryo-EM, meanwhile, involves an exceptionally intricate process of freezing proteins before imaging them with powerful electron microscopes. Both techniques are resource heavy and often take months or years to yield results. But Isomorphic Labs combines AlphaFold 3 with its own suite of AI models to generate high-confidence structure predictions in a matter of minutes.

BRINGING MOLECULAR STRUCTURES TO LIFE

This otherwise-lengthy process is optimized thanks to AlphaFold 3's sophisticated artificial neural network architecture, which has been fine-tuned to predict molecular structures with unprecedented accuracy. One of the key components of this architecture is the Pairformer, a deep learning system that employs multiple types of information—including evolutionary information, or data on how proteins have changed over millions of years—to improve its predictions.

To transform these structural predictions into visible forms, AlphaFold 3 uses an approach called diffusion modeling. Similar to techniques used in popular AI

> Science is an iterative enterprise. Trial and error in the laboratory is how progress has traditionally been made. But now, AI systems are poised to dramatically accelerate this painstaking process.

image generation platforms, diffusion lets the system start from a rough structural "sketch" and gradually refine it into a detailed, increasingly accurate final image. To ground predictions, AlphaFold provides a confidence score, giving researchers an estimate of the model's accuracy.

ADVANCING DRUG DESIGN

One area in which AlphaFold 3 shows particular promise is the rational design of drugs. AlphaFold 3 can efficiently predict how ligands, a structural component of many drugs, will bind to a target protein, thereby guiding researchers in designing drug candidate structures for improved efficacy and reduced side effects. This not only hastens the drug discovery process but also has the potential to ultimately impact the overall cost of development.

"AlphaFold was a watershed moment for AI in biology. It was proof just how powerful AI could be in understanding the complexities of the biological world," says Max Jaderberg, chief AI officer at Isomorphic Labs. "It's what inspired the founding of IsoLabs three years ago."

A compelling example of AlphaFold's effectiveness is demonstrated by Isomorphic Labs in a case study on TIM-3, one of many immune checkpoint proteins that regulate the immune system and prevent excessive responses. In some cancers, these checkpoints are hijacked to suppress the immune system, allowing the cancer to evade detection. TIM-3 has been identified as a potential target for cancer immunotherapy, as blocking its activity could help reactivate the immune response against tumor cells.

A 2021 study experimentally determined the structures of TIM-3 bound to several small-molecule ligands, a laborious process that revealed a previously unknown binding pocket. When given just the sequence of TIM-3 and the chemical structures of the ligands, with no information about the binding pose or pocket, AlphaFold 3 was able to accurately predict the experimentally determined structures. Crucially, the model correctly identified this novel binding pocket and the precise orientations of the ligands within it, even though these structures were not in the training set of AlphaFold 3.

"The technology we already have at our fingertips also opens up new possibilities to tackle diseases that have been previously beyond our reach. Our team of scientists and researchers at IsoLabs has created a whole host of unique AI algorithms. These models work together, allowing us to understand the intricate world of proteins and how they interact with different molecules. This gives us new ways to design molecules, which is incredibly

exciting, and we believe it will completely change how we discover and design new drugs," says Jaderberg.

THE SEARCH FOR THERAPEUTIC ANTIBODIES

AlphaFold 3 also has the ability to predict antibody-protein binding, which can help researchers understand immune response and design new antibodies.

Antibodies are examples of therapeutic proteins that can be used to treat diseases like cancer. But discovering new ones by traditional means is an arduous task. With trillions of possible structures, finding the ones that effectively stick to a target of interest is like searching for a needle in a haystack, experts say.

"Finding new therapeutic antibodies is really sort of like panning for gold," explains Derek Lowe, a veteran pharmaceutical researcher. "You have to check against incredible numbers of possibilities to find things that bind, and then you try to make them better and better."

Taking advantage of AlphaFold 3's ability to accurately model protein interactions, scientists can computationally screen vast numbers of potential candidates and identify the most promising ones for experimental validation. This computational prescreening could drastically reduce the time and effort required to discover new therapeutic proteins, expediting the development of treatments for cancer, contagious diseases, and other conditions for which these drugs have shown promise.

SHEDDING LIGHT ON NEGLECTED DISEASES

AlphaFold is already generating optimism among those investigating overlooked illnesses. The Drugs for Neglected Diseases initiative (DNDi) has been using AlphaFold for years as a means to hasten the discovery of

> "AlphaFold 3 offers huge possibilities in this vital next step. It can predict not just the structures, but also the interactions, taking us a vital step further in the drug discovery process."
>
> **Charles Mowbray,**
> DNDi's discovery director

new treatments for two deadly parasitic diseases, Chagas disease and leishmaniasis, which affect tens of millions of people worldwide.

"AlphaFold 2 has been an immensely useful tool for us, but to turn protein structures into drugs is a journey, and to make progress in that journey, you need to be able to think about how those proteins interact with other molecules," says Dr. Charles Mowbray, DNDi's discovery director. "AlphaFold 3 offers huge possibilities in this vital next step. It can predict not just the structures, but also the interactions, taking us a vital step further in the drug discovery process."

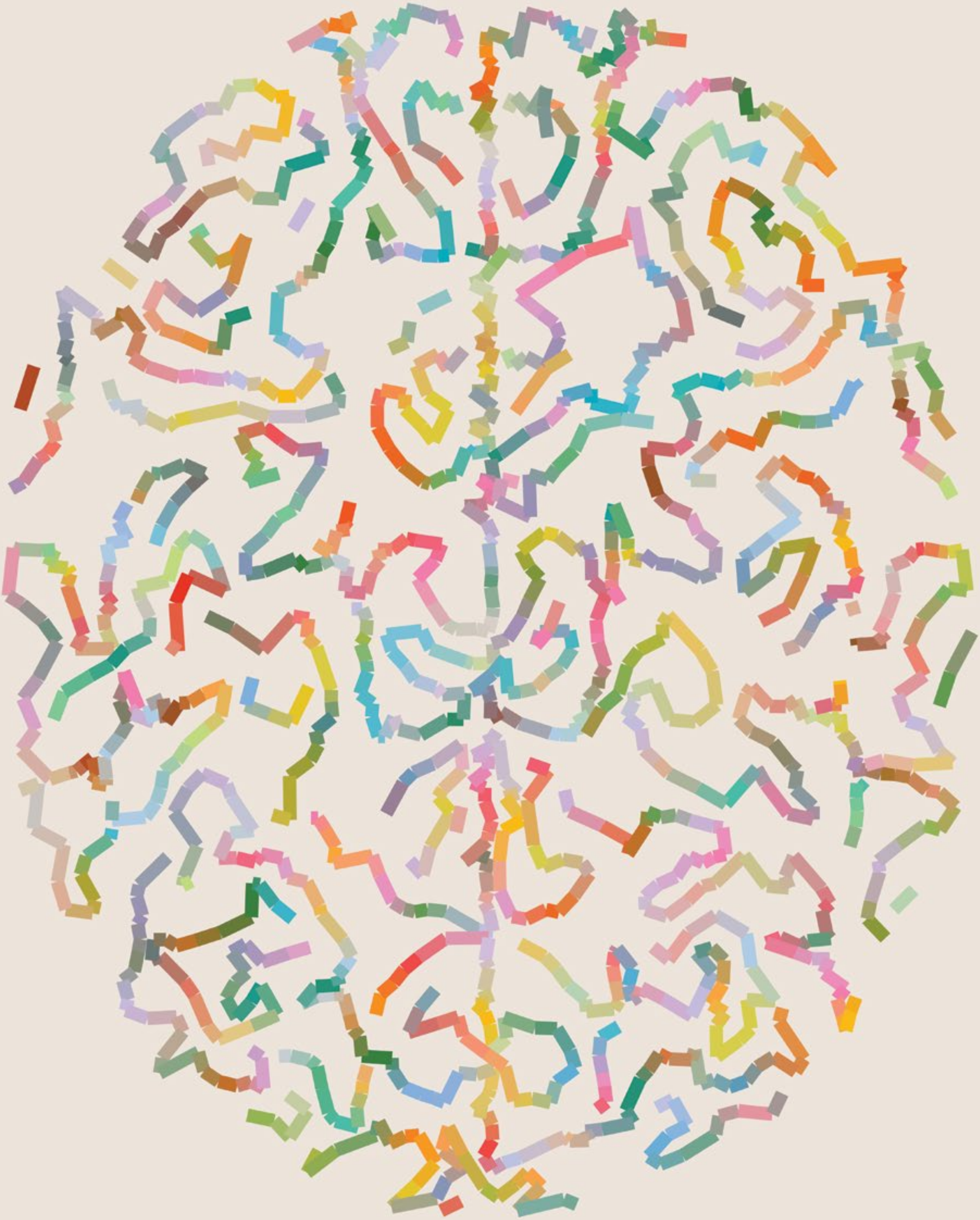
A TOOL, NOT A REPLACEMENT

With such powerful AI tools emerging, some wonder what this means for the role of scientists. But as Lowe observed, the scientific community is more excited than worried.

"We're very happy about this technology because we have so many problems to deal with," Lowe says. "The drug discovery and development process is so long, with so many twists and turns, that we welcome anything that could speed elements of it up or give us some insights."

Ultimately, while AI and AlphaFold are undeniably important parts of the drug discovery process, they are just that—parts. Structure prediction is a single piece of the puzzle, albeit a crucial one, that can help accelerate the development of essential medications for those who need them most.

"Leading the AlphaFold team has been the opportunity of a lifetime, but what I'm most excited for will be seeing scientists use AlphaFold to understand a disease and transform patients' lives," Jumper says. "I can't wait to see that day."



Mapping the Mind

Research from a team of Harvard and Google scientists has revealed never-before-seen details of the human brain's structure. The findings pave the way for a better understanding of AI—and, in turn, ourselves.

By Maya Kosoff

←
Illustration by
Anna Lucia

The average adult human brain is a mosaic of different kinds of cells. One hundred billion neurons are responsible for how distinct parts of the brain talk to each other and how information flows from one part of the brain to another. Think of yourself driving a car: If you see a red light, information flows through neuron after neuron until it gets to your leg muscles, prompting your leg to lift off the gas pedal and shift to the brake pedal.

"There's no way a red light should make your leg muscle move unless there's some connection between your retina and your leg, and that has many relays, but there is no simple pathway between those things that anyone understands," Dr. Jeff Lichtman, a neuroscientist at Harvard University, explains.

Understanding how these processes work in the human brain is a discipline in neuroscience called connectomics, which aims to unveil the wiring diagrams of the nervous system. Eventually, connectomics wants to understand not just simple behaviors like the wiring behind how traffic signals affect us when we drive, but also the pathways that underlie every memory we have and every piece of information we've learned.

To do that, research needs to dig deep into the subcellular level. That's what Lichtman's team at Harvard, in collaboration with Google, did while developing the most detailed map of a section of the human brain to date.

Charting a Map of the Brain

Lichtman's team started its research 10 years ago with a sample of a brain smaller than a grain of rice. That cubic millimeter of brain matter contained 57,000 cells, 230 millimeters of blood vessels, and 150 million synapses—which altogether yielded 1,400 terabytes of data, or a volume of content equivalent to more than 1 billion books.

The project, Lichtman says, was a huge undertaking. "When we started, we didn't appreciate just what it means to have one thousand terabytes of data," he says. "We're still somewhere toward the beginning of this great adventure of neuroscience. This realm should be quite familiar to us, since we carry these things on our shoulders, every one of us, but they're largely unknown and unexplored."

Despite the astounding amount of information Lichtman and his team had to contend with, they were able to develop those datasets relatively efficiently with the aid of AI, a process that would have taken hundreds of thousands or millions of humans to do by hand. To work with Lichtman's massive imaging datasets, Google developed a software library called TensorStore, which "makes it possible to manipulate large datasets from thousands of computers simultaneously," says Dr. Viren Jain, a research scientist who leads Google Research's Connectomics team. "This software has been critical for our neuroscience research but has also turned out to be useful for other applications at Google, such as training [generative AI chatbot] Gemini."

Google's Connectomics team is focused on better understanding the brain with the help of AI. "We develop algorithms to interpret large-scale data about the structure of the brain and produce connectomes, or descriptions of how the brain is wired at a cellular and synaptic level," Jain adds. "Neuroscientists can then use this data to explore, test, and develop theories about how the structure of the brain is related to its function."

What AI can reveal about our brains—and what our brains can reveal about AI

Artificial intelligence not only sped up Lichtman's research but also shed light on the findings themselves. AI and neural networks share a fundamental conceptual connection with the biological neural networks in the brain.

AI, particularly in the field of machine learning, is inspired by the structure and function of our brains' neural connections and pathways. In our brains, neurons are connected through sophisticated networks. They process and transmit information through chemical and electrical signals. Artificial neural networks are computational models inspired by our brains' neural connections. This enables machines to learn from—and make predictions based on—data.

In our brains, neurons are connected by synapses. Each neuron can have thousands of synaptic connections that form a network allowing for cognitive tasks and motor functions. Similarly, artificial neural networks have nodes (or "neurons") organized in layers. Every node processes input data and communicates to other nodes, analogous to the synaptic connections we experience that help us tie our shoes, write emails, and make scrambled eggs.

"Historically, advances in neuroscience have inspired computer scientists and led to major new developments such as the overall concept of an artificial neural network, reinforcement learning algorithms for training AI systems, and even the famous convolutional neural network architecture," Jain says, referring to a deep learning neural network that has contributed much to our present-day understanding of computer vision and image analysis. "So, indeed, while we can't anticipate specific new advances, we do believe that achieving a much more precise understanding of biological computation will lead to important new ideas for artificial computing systems."

One area of interest, Jain says, is low-energy computing: "The human brain uses about as much energy as a light bulb. When we better understand how the brain works, perhaps we will be able to use some of the key insights to improve the energy efficiency of our AI systems."

Though artificial neural networks are nowhere as complex as our own brains (yet), the parallels between these systems highlight how understanding biological intelligence can drive advancements in AI, in what Lichtman calls a "virtuous loop."

"It's profoundly interesting that we are depending on machines that have learned how to do something to give us insights into how brains learn to do things," he says. "You could imagine those machines eventually providing insights into how we should program our future AI algorithms. That's a virtuous loop—maybe not 'virtuous' for humans, but virtuous in the general sense that it makes these machines smarter and smarter."

New AI mechanisms can learn without being taught, Lichtman says, and he predicts that eventually they could have a built-in understanding of their own wiring. Machines don't use human logic—so when they make mistakes, a human gives them training to correct the mistake, and the machines don't make the same mistake again.

They're "superhuman," Lichtman says, and they'll be able to do everything humans can do, but currently they don't know why or how they're doing it. "They're still at this coloring-book level. They're not looking at the picture they end up with from filling in the coloring book. They can't say, 'That's Mickey Mouse.'" But at some point, Lichtman says, these algorithms are going to recognize what they're doing and then quickly get to the bottom line of what it is without going through the painstaking motions of coloring in every last little pixel before they can identify the final shape they're coloring.

This process, Lichtman says, is called segmentation—and it's a big part of the work Google did with his team on the brain-mapping project. "The group came up with an algorithm that made 'coloring in' very efficient, so these machines can do it faster and more accurately," he says.

To transform the data from Lichtman's team into something AI could process, Google used two advanced

> "The reason it's attractive to do this work isn't because it's easy, but because it's hard. We should always try to do things that are a bit beyond our capability."
>
> Dr. Jeff Lichtman,
> neuroscientist at Harvard University

techniques. The first, flood-filling networks, tackled the monumental task of mapping individual wires and cells in massive 3D datasets—"a difficult problem that would be infeasible with purely manual effort," says Google's Jain. The second, an algorithm called SegCLR, helped label each wire by its type, making it easier for neuroscientists to connect the structures to familiar references in existing research.

Information is the enemy of understanding

Instead of setting out with a goal of understanding something as complicated as the brain, Lichtman says, it's much better to ambitiously attempt to describe it. "Millions of things are happening simultaneously," he says. "It's a wiring diagram like a road map, except it's in three dimensions, and every road forks thousands of times. It's impossible to look at it and say, 'Oh, now I get it.'" The human brain is impossible to understand in depth. But we can describe it, and we couldn't do that nearly as completely before Lichtman's team published its paper and map of the human brain.

Mapping—an age-old practice that began with humans trying to understand the physical geography of the world around them—is driven by innate curiosity. "Humans have done this over and over, and we're learning things, but I wouldn't say the world is getting simpler as a result," Lichtman says. "It's not that we understand all this stuff. What we understand, usually, is how far we are from understanding."

Ironically, information, Lichtman believes, is the enemy of understanding. Before his team had data about the brain, research scientists could develop untested theories about how the brain worked. But now that he can see what certain sections of the brain really look like, those theories go up in smoke. While someone who has seen their hard-won concepts disappearing over time might, understandably, become jaded and cynical, Lichtman is anything but. Instead, this process, he says, is indicative of the way human knowledge builds on itself. "We didn't understand the physical basis of inheritance until we mapped the genome, and then we knew the name of every single gene. But no one would ask, 'Do you understand the genome?' It's impossible to understand. So many things are simultaneously turning on and turning off as the genes are working to generate your body shape," he says.

What's next for Lichtman's team

In tech, the 10x concept describes a goal of challenging traditional thinking and achieving huge growth: scaling users, or maximizing how many people can use your product at once. Though it originated in the IT world to describe software engineers who are 10 times more productive than the average developer, it also aptly describes Lichtman's ambitions for his next project.

Working with Google and six other laboratories distributed across the United States and Europe, Lichtman hopes to do "something 10 times as big as the project we just finished": Using a mouse brain, the team will map parts of the brain related to memory systems, the hippocampal format. That, Lichtman says, will be a proof of concept. If the team is successful, it will accomplish a task a thousand times more ambitious than its 2024 brain-mapping project: mapping a mouse's entire brain. "All you have to do is scale up and find the money to do this on a bigger scale. For us scientists who are in academia, it's an unbelievable opportunity to work with truly the world's best programmers on a project," he says.

The field of connectomics is still nascent, and Lichtman knows that his milestone achievement is soon to be surpassed. "It's not like we could rest on our laurels, to put it mildly. A number of us working in this field are feeling that, finally, we have the wind in our sails—as opposed to feeling more like Sisyphus," he adds with a laugh. "Virtually everything we tried failed and was very difficult, and now, finally, things are working."

Before working with Jain's team at Google, Lichtman had a "somewhat dim" view of AI, he says, colored by people's initial reactions to large language models. "But once the training gets going, it's a profound thing," he says. "We are definitely the benefactors of AI." Hard, ambitious problems like mapping the brain, he says, are the best problems to study. "The reason it's attractive to do this work isn't because it's easy, but because it's hard. We should always try to do things that are a bit beyond our capability."

Computing at the Edge of Reality

Scientists are harnessing the alternative physics of the quantum realm to create computers of unprecedented power, potentially revolutionizing fields from drug discovery to climate modeling.

By **Daniel Oberhaus**

→
Visuals from
'Seeds' by Adam Ferriss

Quantum computers are machines that calculate by exploiting quantum mechanics, a branch of physics that describes reality at its most fundamental level. In the quantum realm, nature operates according to principles that have no analogue to our daily experiences: Particles are waves, occupy many positions at once, and can send information to particles on the other side of the universe. For humans used to living a mechanistic world of cause and effect, the quantum world is strange and unsettling. Even Albert Einstein, famously, could never come to terms with the weirdness of quantum mechanics.

Although quantum physics has been studied by some of the greatest minds in physics for more than a century, it wasn't until the 1980s that anyone began seriously thinking about how to apply the insights from quantum mechanics to computing. The basic idea is that rather than trying to translate inherently quantum aspects of reality into the binary logic of classical computers and then onto nature to build a computer, we can directly harness the quantum mechanical properties of matter to do computations. This new breed of computer would leverage phenomena like superposition (the ability of particles to be in two states simultaneously) and entanglement (the ability for particles to remain correlated with other particles regardless of physical proximity) to do computations that would be practically impossible for a conventional computer.

It's an ambitious dream, and one that is still in the making. A decade passed between the time the famed physicist Richard Feynman proposed the idea of quantum computing in the early 1980s and when the mathematician Peter Shor described a useful quantum algorithm that could outperform a classical computer—at least in theory. Shor's algorithm described a way to use quantum computers to factor integers, which could, in principle, be used to break the 2048-bit encryption standards that the modern internet depends on. It was a major moment in the history of quantum computing. But 30 years later, quantum computers still don't have nearly enough qubits to make that happen. In fact, it took another decade after Shor published his algorithm to experimentally implement it on a quantum computer, which was only able to factor the number 15 into its prime factors—a calculation so simple that most 10-year-oldchildren could do it by hand.

Since then, however, progress toward a universal quantum computer has been accelerating, and researchers are increasingly thinking about how these machines might

be usefully applied in fields ranging from theoretical physics to the development of pharmaceutical drugs. Creating realistic quantum mechanistic models of all but the most simple molecules remains challenging or even impossible for classical computers, which makes it difficult to study and develop new classes of potentially life-saving drugs. At the molecular and sub-molecular levels, these compounds are subject to quantum mechanical effects that are well beyond the simulation capabilities of today's most powerful supercomputers, but should—in principle—be a breeze for a computer that uses quantum phenomena to do its calculations.

"There are elements of nature that are beyond even the best supercomputers," says Charina Chou, the chief operating officer of Google's Quantum AI lab. "Nature isn't classical, and you're never going to be able to compute exactly with a classical computer, for example, how every molecule behaves—and our entire world is made up of molecules. If we could fully understand them and use these insights to design new molecules, that is an enormous advantage of a quantum computer."

The same is true for the development of advanced materials, which also require a deep understanding of the molecular and subatomic properties of the material. AI running on classical computers is already helping accelerate the discovery of new materials for a broad range of applications in agriculture, aerospace, and industrial manufacturing. The hope is that quantum computers could advance this capability by providing increasingly high-fidelity subatomic models of these materials.

"The simulation of systems where quantum effects are important is of rather significant economic relevance because many systems fall into this category," says Hartmut Neven, vice president of engineering for Google's Quantum AI. "Want to design a better fusion reactor? There's plenty of quantum problems there. Want to make lighter, faster, and more robust batteries? There's plenty of quantum chemistry problems there, too. Whenever engineering involves quantum effects, there is an application for a quantum computer."

Realizing this vision will require tackling staggering technical challenges that include the construction of massive ultracold refrigerators for quantum hardware and the near-perfect isolation of quantum computers from the outside world to prevent quantum mechanical interference. For now, most of the promises of quantum computing—such as accelerating the discovery of new drugs and materials or unlocking new insights into physics, biology, and chemistry—are still theoretical. But by bridging the gap between the fields of quantum computing and artificial intelligence, it may be possible to reduce the timeline to building a bona fide universal quantum computer that will open new frontiers in biology, physics, chemistry, and more.

Alien MATH

The past century of research on quantum mechanics has shown that if Newtonian physics—the world of billiard balls and planetary orbits—operates as a clock does, quantum mechanics prefers dice. Countless experiments have demonstrated that it's impossible to predict quantum phenomena, such as how a particle will scatter or a radioactive atom will decay—with perfect accuracy. We can only give probabilities for a certain outcome.

"A lot of people think that quantum mechanics is really complicated and involves waves being particles, particles being waves, spooky action at a distance, and all that," says Scott Aaronson, a theoretical computer scientist and the founding director of the Quantum Information Center at the University of Texas at Austin. "But really quantum mechanics is just one change to the rules of probability that we have no experience with. But once we learn that rule, everything else is just a logical consequence of that change."

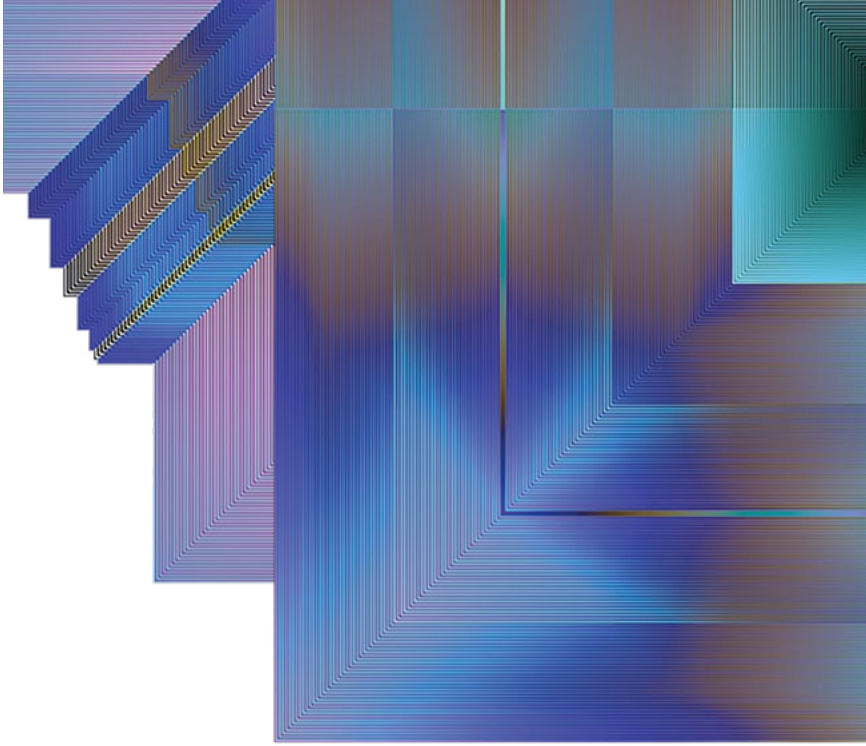
Probability is ultimately an exercise in quantifying uncertainty, and there are well-established rules for adapting probabilities to new information. All probabilities exist on a spectrum from zero to one, for which zero is complete certainty that something won't happen and one is complete certainty that something will happen. But unlike the probabilities that determine your fortunes at the casino, quantum mechanical probabilities—called amplitudes—can exist within the unit circle of complex numbers and be less than zero. If you strip away all the jargon, the fundamental insight of quantum mechanics is that nature operates according to alien rules of probability at the base layer of reality.

This small tweak in the rules of probability has profound implications for our understanding of reality—and our ability to harness it for computing. While classical computers use bits (0 or 1), quantum computers use qubits. In addition to zeros and ones, qubits can exist as some combination of zero and one simultaneously—a phenomenon known as superposition. This allows quantum computers to represent and process exponentially more information than classical computers with the same number of bits.

This exponential growth in computing power is the reason quantum computers should, in principle, be able to dramatically speed up the time it takes to compute the answer to certain types of problems. But harnessing this power presents significant challenges.

Superposition is crucial to a quantum computer's power, but it's fragile. Measuring a qubit collapses its superposition, making it behave like a classical bit (i.e., it is either a 0 or a 1). This challenge requires careful isolation of qubits from their environment until computation is complete.

The Scale of Quantum Computing



# of quibits	# of states that can be represented
1	2
2	4
3	8
4	16
5	32
10	1,024
20	1,048,576
50	1,125,899,900,000,000
100	1,267,650,600,228,229,401,496,703,205,376

"Superposition," says Aaronson, "is something that particles like to do in private when no one is watching."

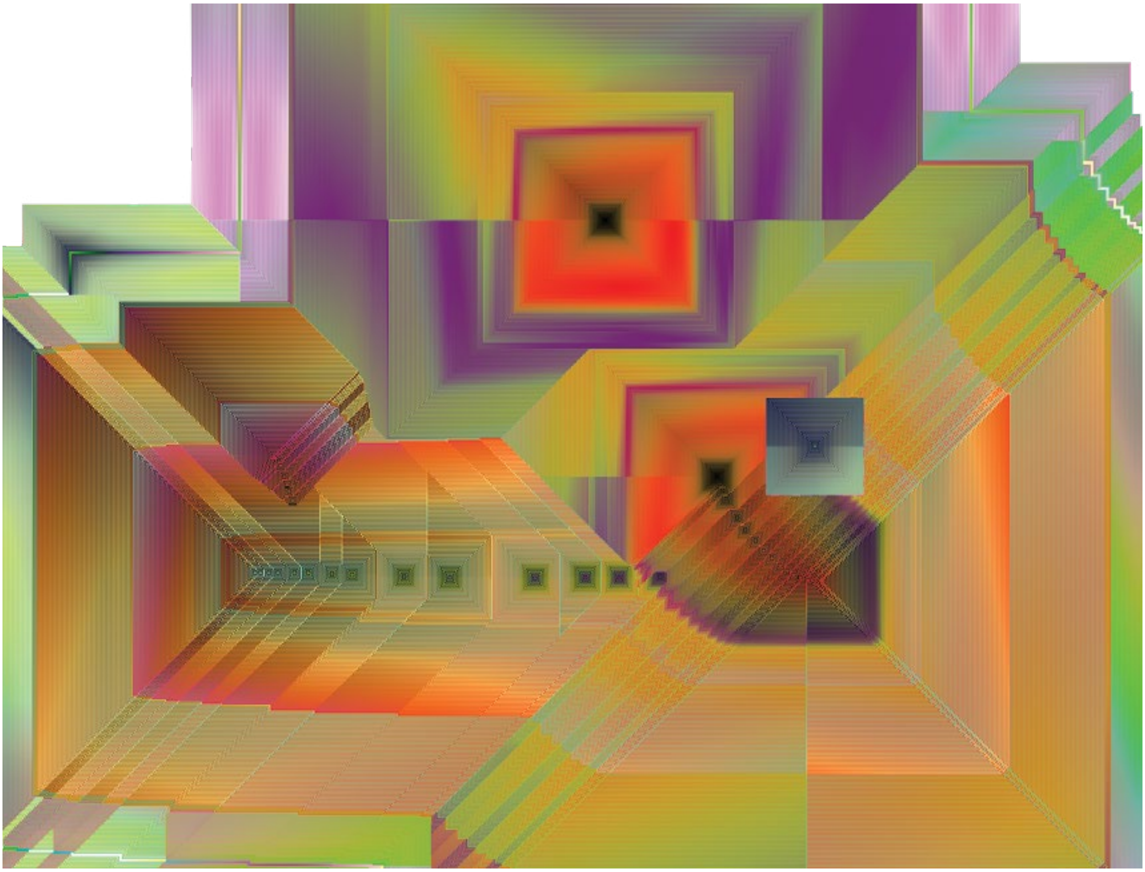
It's a phenomenon that the physicist Erwin Schrödinger famously captured in a thought experiment in which he imagined putting a cat in a box that contains poison and shutting the lid. Until the lid is opened and the cat is observed, it is impossible to determine whether the cat is still alive or has eaten the poison and died. The cat is in a superposition of dead and alive; the only way to know for sure is to look in the box and observe the cat's state, at which point the cat is definitely in one of the two states: dead or alive.

The problem is that for a quantum computer to be useful, researchers need to be able to measure its output; they need to open the box and look at the cat. But measuring qubits directly will destroy their superposition and any advantages offered by a quantum computer. The key is to ensure the measurement happens only when the computation is finished. In the meantime, the qubits need to remain as isolated as possible from their external environment so it doesn't destroy their superposition and entanglement.

Implementing this in practice is tricky. Many quantum computers, such as Google's Sycamore, operate at near absolute-zero temperatures to achieve superconductivity and shield qubits from external interference. However, perfect isolation remains elusive, and noise-induced errors persist as a major hurdle in quantum computing.

Today, quantum computing is considered to be in its "noisy intermediate-scale quantum" (NISQ) era. *Intermediate-scale* refers to the fact that most existing quantum computers have about 100 qubits—orders of magnitude fewer qubits than what most researchers estimate will be required to make a quantum computer useful. Even at this intermediate scale, these systems are still plagued by error-inducing noise.

Solving the noise problem is arguably the most important and daunting problem facing quantum computing in a field of research overflowing with important and daunting



problems. A variety of approaches are being explored to solve quantum computing's noise problem, and generally speaking they can be grouped into two main categories: approaches that try to limit the amount of noise introduced to the system and approaches that attempt to correct the errors introduced to the system.

"Every quantum bit has error associated with it, which means as you bring together more qubits to do more computation, you're also introducing more error into your system," says Chou. "The whole idea behind quantum error correction is using qubits to protect against new errors introduced to the system so that, as you add more qubits into a system, the amount of error actually decreases."

Chou estimates that a universal quantum computer will require at least 1 million qubits to do useful calculations for molecules and materials. Overcoming errors even at this modest size is still a formidable challenge, and getting to 1 million qubits will likely require some mix of enhanced noise resistance and improved error correction. The question is how to get there. Increasingly, researchers are turning to AI to help make it happen.

The Rise OF Quantum AI

The history of science and technology is, in many respects, a history of serendipity. From apocryphal eureka moments like Newton's apple to the discovery of penicillin on stale bread, the flashes of insight that have profoundly changed the world have often come from the most unexpected places. For Neven of Quantum AI, it was the decision to listen to a public radio station on his way home from the office one evening that changed the trajectory of his career—and possibly all of computing.

At the time, Neven had already made a name for himself as one of the world's leading researchers on machine vision. In the early 2000s, he had been tapped by Google

> “Want to design a better fusion reactor? There’s plenty of quantum problems there. Want to make lighter, faster, and more robust batteries? There’s plenty of quantum chemistry problems there, too. Whenever engineering involves quantum effects, there is a killer app for a quantum computer.”

Hartmut Neven, vice president of engineering for Quantum AI

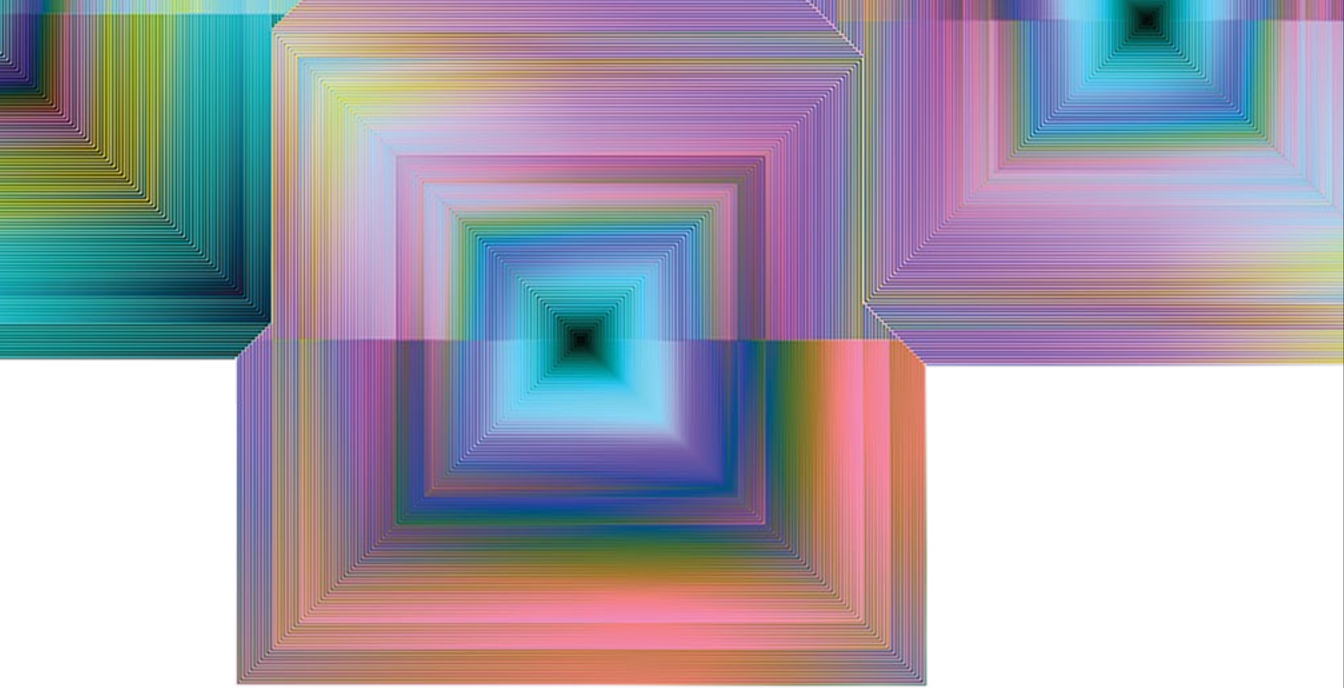
to lead its visual search team. At Google, he developed the visual recognition technologies that are foundational for Image Search, Google Photos, YouTube, and Street View, and he was nearing completion of the first prototype of the augmented reality–enabled glasses that would become Google Glass.

Meanwhile, Neven had also been carving out a niche for himself in the burgeoning field of quantum computing. He was intrigued by how this new technology might be applied to machine learning (ML) to usher in a new computing paradigm that could accomplish tasks neither technology could on its own. He had already made significant progress toward this goal by becoming the first to implement an ML and image recognition algorithm on a quantum computer in 2007, but it was the public radio broadcast during that fateful commute that convinced him to go all in.

“I had heard a story on NPR about quantum computing, and it sounded to me that a quantum computer would be a good tool to do certain image transformations, like Fourier transforms,” says Neven, referring to a technique that decomposes an image into frequencies so that its features can be more efficiently processed by a computer. “That kindled my interest, but I was semi-mistaken about quantum computers being a good tool for it. That application may come one day, but it won’t be one of the first applications.”

Nevertheless, as Neven continued to explore the relationship between quantum computing and machine learning, it became apparent that there were some very promising ways to bridge these two worlds, particularly when it came to optimizing how ML systems are trained. So in 2012, Neven and his team at Google launched the Quantum Artificial Intelligence lab, in partnership with researchers at NASA Ames and the Universities Space Research Association, with the goal of building a quantum computer and finding impactful ways to use it—including advancing machine learning.

As Neven wrote in a blog post announcing the lab, the way machine learning improves is by creating better models of the world to enable more accurate predictions. But the world is a complex place, and some aspects of nature are effectively impossible to model with binary code. Classical computers operate in the world of ones and zeros, presence and absence, on and off. But there, if you probe nature at a deep enough level, you’ll encounter issues that can’t be fully modeled using binary code. Sometimes, when nature poses an either/or question, the answer is simply yes.



Unknown Unknowns

By the time Neven started the Quantum AI lab in 2012, he and several other researchers had already demonstrated that ML algorithms could be implemented on research-grade quantum systems that were designed to solve specific and narrow tasks. Implementing ML algorithms on modern "general purpose" quantum computers remains a significant obstacle and an active area of research for Neven and his collaborators.

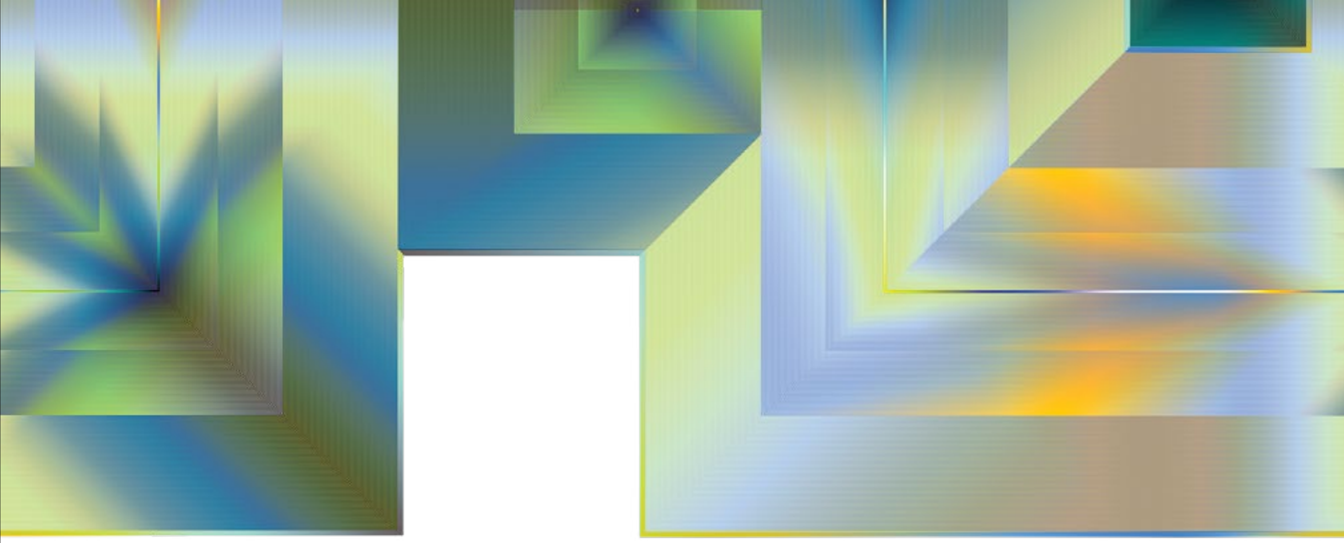
So far, quantum computers have struggled to show that they provide superior performance vis-à-vis classical computers in any context that is useful in the real world. Part of the reason for this is they still struggle with errors and so are not accurate or large enough to implement many quantum algorithms; the other reason is that not every problem has an obvious—and, more importantly, provable—quantum advantage. Most benchmarks for quantum advantage involve computing solutions to esoteric mathematical problems that have no obvious real-world relevance. Even then, quantum computers have struggled to demonstrate that they are faster at solving these problems than the most advanced classical computers.

In 2019, Neven's team at Google's Quantum AI lab achieved "quantum supremacy" for the first time in history. Their quantum computer, Sycamore, took roughly three and a half minutes to find the answer to a technical problem in quantum computing called random circuit sampling that would've taken the most capable classical supercomputer at the time 10,000 years to solve. It was an important scientific achievement and benchmark, though the problem has no obvious real-world application.

Part of the challenge with quantum computing is it's difficult to prove that no algorithm run on a classical computer can find a solution equally, if not more efficiently, than the quantum computer. Most of the time, this high-falutin mathematical debate plays out at research conferences and in the annals of scientific journals. But in this case, Google didn't have to wait long to see its claims to quantum supremacy possibly thumped by a new classical technique. In 2024, a group of Chinese researchers published data showing that they had outperformed the 2019 Sycamore processor on the same challenge using hundreds of conventional chips. Soon after, Google published a follow-up paper demonstrating that an updated Sycamore processor could outperform 2024's most powerful supercomputer.

The uncertainty around quantum supremacy, however, is just the nature of the game in quantum computing. There is still broad consensus among researchers that Neven's team at Google seems to be leading the pack regarding quantum computing. The team's work over the past decade is why it no longer seems quite so far-fetched that the world could have a functioning quantum computer doing useful work within the next 10 to 20 years.

Neven is the first to admit that the road to a general-purpose quantum computer that can unambiguously



outperform advanced classical computers will be long and winding. The technical challenges are immense, but so, too, are the stakes. The emergence of a bona fide “universal quantum computer” would likely change the course of human history and unlock new frontiers in mathematics, physics, biology, and everything in between. This computer would allow us to model the physical world in all its dynamism that can’t be captured in the comparatively flat language of binary. The more accurately we can model nature, the faster we can find answers to our biggest scientific mysteries. In biology, for example, cells are sometimes a multiplicity of identities and potential; with classic computing, we’re forced to flatten these cells into a single instant in time or a stack rank of identities: skin cell, cancer cell, dying cell, growing cell; on the arm, in the bloodstream, in the brain. However, in reality, as the body moves, changes, and shifts, these cells are everything, everywhere, at once.

This is the type of scientific challenge that’s made for a universal quantum computer, but a quantum computer that is up to the task is neither guaranteed nor imminent. However, in the past few years alone, there has been an increasing number of signs that we are at least on the right path to a universal quantum computer and that the intersection of quantum computing and AI will be an important part of the puzzle—including both the use of AI to accelerate quantum computing and, eventually, AI applications for quantum computing.

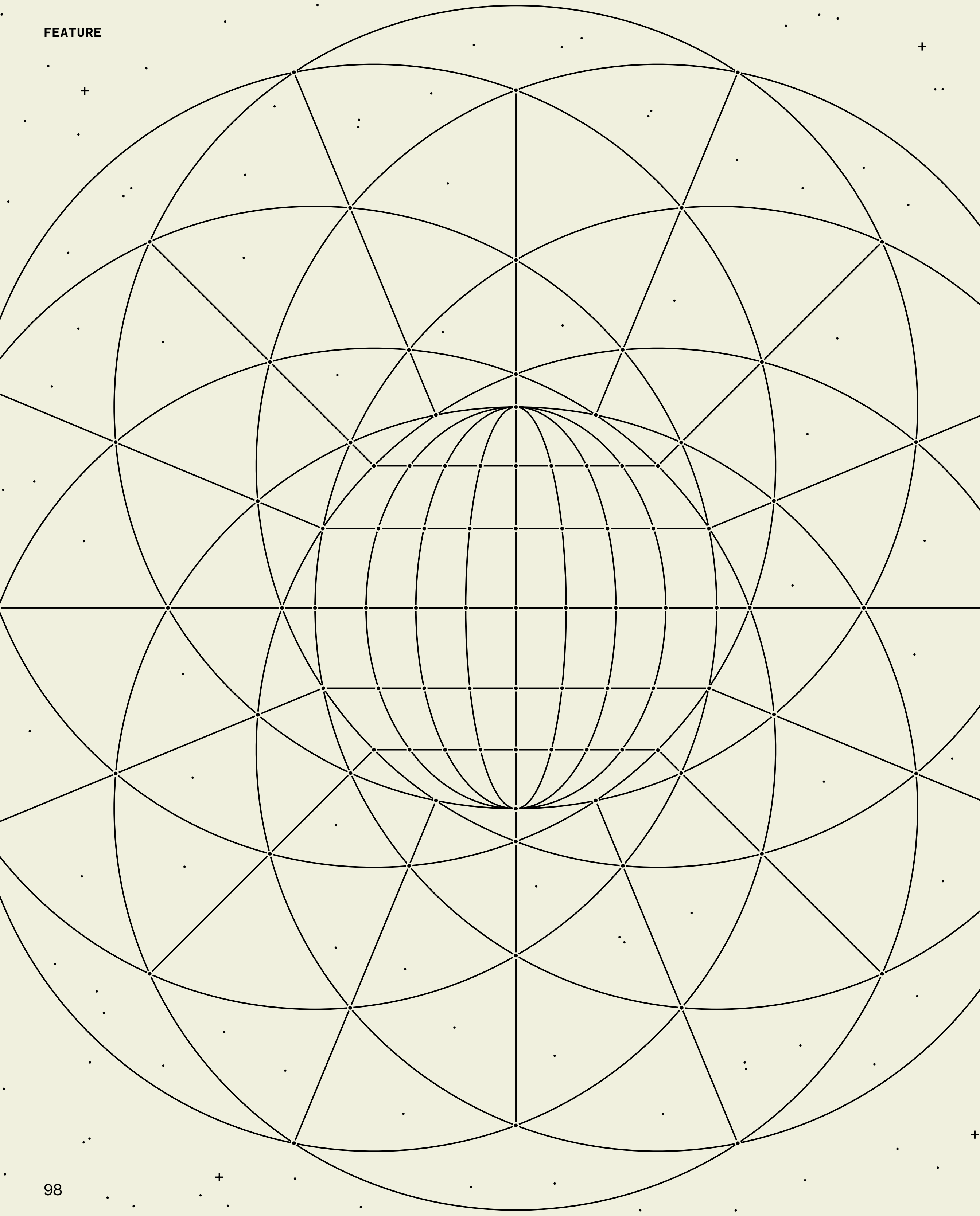
At Google, Neven, Chou, and their colleagues are studying ways to both apply AI to better design quantum computers and use quantum computers to build enhanced AI systems. For example, Chou points to how Google engineers are using AI to improve quantum-chip fabrication processes with image recognition systems that streamline qubit quality assessment; developing ML tools to automate coding tasks for quantum systems; and building transformer models that are helping enhance quantum error correction.

Reciprocally, Neven highlighted how quantum computing promises to dramatically reduce the sample complexity in machine learning, potentially leading to AI systems that can learn from exponentially fewer examples than their classical counterparts. This hints at a future when quantum-enhanced optimization could solve more sophisticated training problems, even discerning and discarding mislabeled data in large datasets.

The future of quantum computing is full of promise and uncertainty. Staggering advances are being made every day in each field, but when we look at the history of computing, we see that time and time again, the best forecasts about the future of technology are laid to waste. More than 55 years ago, when the first basic terminals were connected via ARPANET—the progenitor of the modern web—no one could have predicted the rise of the major platforms we know today. But if history teaches us anything, it’s that the future is always weirder, and often more wonderful, than we could have ever imagined.

Images in this story come from ‘Seeds,’ a drawing tool by Adam Ferriss controlled by rules based on cellular automata—a regular grid of cells, each in one of a finite number of states, such as *on* and *off*, with applications in fields including physics, theoretical biology, and microstructure modeling.

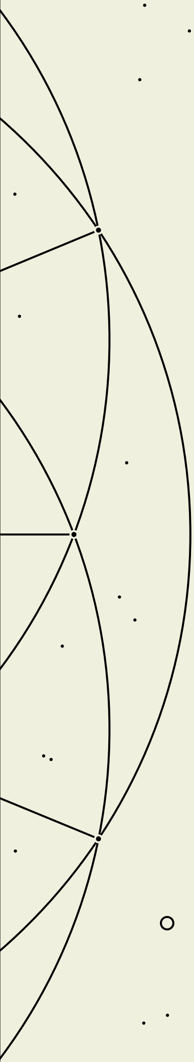
Visit **adamferriss.com/seeds** to explore the tool.



GOING BIG:

THE RACE TO SOLVE CLIMATE CHANGE

With the ability to model the chaotic forces of the planet—weather, carbon emissions, vehicle traffic—AI has become a vital tool in building a sustainable world.



← ↓ → ↑
Illustrations by La Tigre

Today, two forces are accelerating. The first is climate change, with the world hitting record-breaking temperatures, and the second is indisputably AI. For AI researchers who study climate change, it's an extraordinarily humbling and simultaneously promising time. From the United Nations to scholarly journals, AI is being hailed as a game-changing and revolutionary tool in society's efforts to address the climate crisis.

Part of why climate change is difficult to address is because it comes from so many sectors of our economy. Tackling the climate challenge will require changes to our transportation systems, our ecosystems, electricity and construction, and how we operate heavy industries like cement and steel – all while adapting to the changing climate that we find ourselves living in and to the intensification and increased frequency of natural disasters due to climate change.

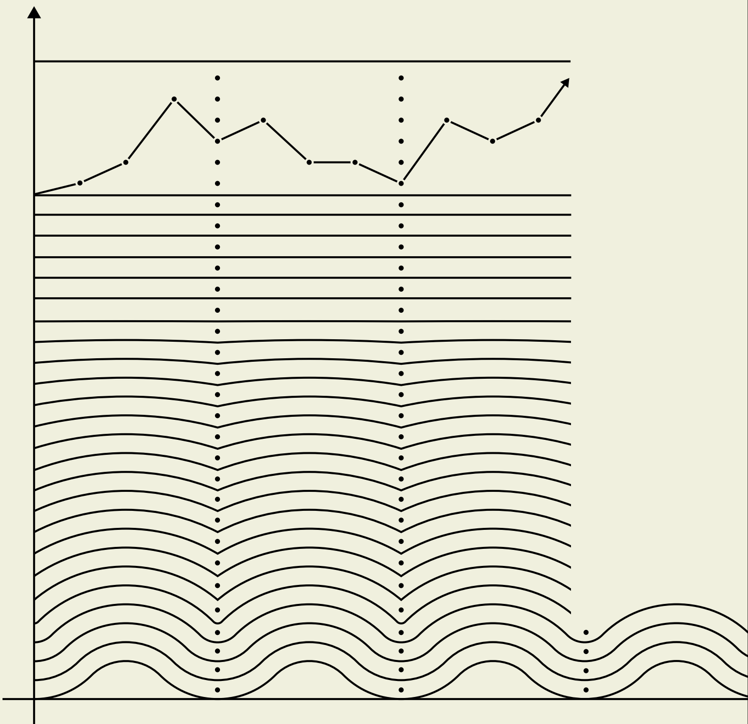
AI makes it possible, for the first time, to model the extreme complexity of our climate solutions. The climate scientist Richard CJ Somerville has noted, "There is no silver bullet that solves all the challenges of climate change, but there is a lot of silver buckshot." By this, he means that a combination of diverse methods—including emissions reduction, which remains by far the cheapest way of lowering carbon in the atmosphere—will be our best path forward. Rather than wait for long-term projects to bear fruit, we should take meaningful climate action right now.

That sense of urgency is what motivates companies like Google to set ambitious goals of reaching net-zero emissions across the company's operations and value chain by 2030 and running on 24/7 carbon-free energy on every grid where they operate. Kate Brandt, Google's Chief Sustainability Officer, and her team are responsible for turning that goal into reality. "I've committed my life's work to help accelerate progress on climate action," says Brandt, "and I believe that AI is a complete game-changer for what we are trying to do and for the planet as a whole."

She is not alone in that belief. Yossi Matias, Head of Google Research, who leads work in AI and state-of-the-art technologies, including many of the company's climate initiatives, also believes that addressing these planetary-scale challenges requires, first, identifying the issues, and second, developing urgently needed tools to help mitigate, adapt to, or prepare for environmental crises.

TAKING BIG SWINGS

Google's AI-powered climate solutions—virtuoso feats of coordination and data analysis—point toward entirely new clean-energy paradigms and hold potential for world-altering breakthroughs. "Today, we are in a golden age of research, where the world's biggest challenges and opportunities are accelerating the pace of research and motivating the development of breakthrough technology," says Matias. "With remarkable speed and efficacy, AI is enabling us to solve problems that in the past would have seemed impossible."



> “When I asked the experts, the general consensus was that predicting well where floods are going to occur is too difficult a problem to solve.”

Yossi Matias, head of Google Research

FLOOD FORECASTING

One of the most notable examples is the problem of flood forecasting. “When I asked the experts,” Matias says, “the general consensus was that predicting well where floods are going to occur is too difficult a problem to solve.” What makes flood prediction such a complex problem? It’s the number of variables involved, along with the ways in which they interrelate according to the laws of fluid dynamics. These variables begin with the action of precipitation on the portion of a river on which it falls—the quantity of rainfall; the volume, angle, and speed of its descent; the type of soil through which the river flows; the vegetation in and around it; and the topography of the river itself.

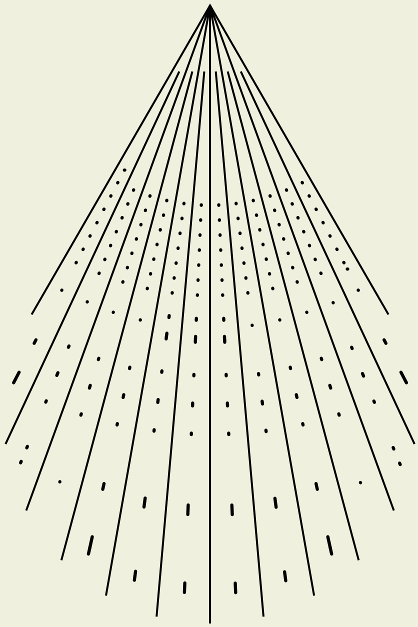
These predictions need to be reliable, both to earn the public’s trust and to spare people from costly evacuations of their families and their livestock. They are difficult even when there’s a well developed system for data recording. And in many regions, this historical data may be scarce or nonexistent. Google had to create a breakthrough global model that could analogize one place and situation to another. “In a way you can think about it loosely like a language model that generates new text based on all the text and examples it has seen, but in a very different domain,” Matias says. The model “enables a scalable solution, and adds a level of equitability,” he explained, because some data-scarce countries have suffered disproportionately from global warming.

A pilot program in India serving 10 million people was expanded to all of India and Bangladesh, and then further developed into a system that helps to predict flooding up to seven days in advance in over 100 countries encompassing 700 million people. Google’s success here, Matias says, “makes me really optimistic that there are other notoriously difficult problems for which we could actually drive research and do something about it.”

HOW TO REDUCE EMISSIONS

Human activity creates 59 gigatons of carbon equivalent emissions annually, according to UN's Intergovernmental Panel on Climate Change. Every single viable climate scenario that the IPCC has modeled—i.e., every scenario that would keep temperature rise below 2 degrees Celsius by 2100—depends in part on some form of carbon-dioxide removal.

AI could help reduce emissions by five to ten percent globally by 2030 simply by accelerating existing technology solutions, according to research co-produced by Google and Boston Consulting Group in late 2023. Five to ten percent of global emissions is roughly equivalent to the annual emissions of the European Union. "This is the decisive decade for climate action. We know the impact of climate change knows no borders—which is why our response needs to be global and it needs to be a collective effort," says Brandt.

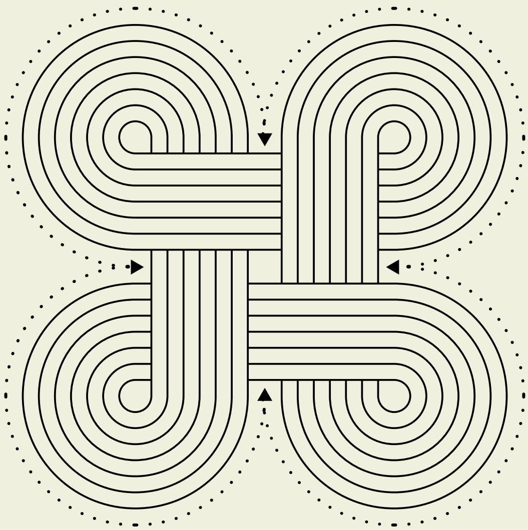


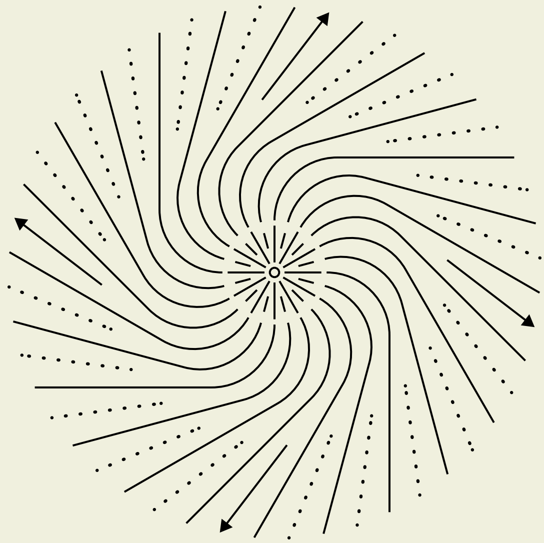
PROJECT CONTRAILS

Google's Project Contrails is one of few active initiatives that mitigates global warming not on land or sea but in the stratosphere itself. Contrails—the white wake produced by airplanes—account for a surprisingly large proportion of air travel's environmental impact: some 35 percent, according to a report from the IPCC. A certain level of humidity is needed for a contrail to form. That suggests a simple solution for air-traffic pollution: Avoid humidity. But that turns out to be difficult to predict. For Google's engineers, the solution entailed manually creating a database from satellite imagery of existing contrails to teach a machine-learning algorithm to recognize new contrails and to predict contrail formation regions. In 70 American Airlines test flights, contrails were reduced by 54 percent—simply by adjusting altitudes a couple thousand feet.

FUEL-EFFICIENT ROUTING

On the road, Google's fuel-efficient routing initiative is an example of targeting the climate crisis in a way that yields immediate benefits. The system, which is indicated on Google Maps by a green leaf, presents drivers with the most fuel-efficient route to their destination. "In the background," says Brandt, "there's a very sophisticated AI model running that's looking at all of the different variables, everything from traffic to the gradient of roads." Companies are already using the function to streamline their supply chains. The initiative is now available in the U.S., Canada, Egypt, India, Indonesia, and more than 40 countries in Europe. Since launching in 2021, it has saved 2.9 million metric tons of greenhouse-gas emissions through the end of 2023—the equivalent of removing 650,000 fuel-based cars from the road for a year. With only a "slight change to how we get people from one place to another," Matias says, "we can quite significantly affect global warming."





PROJECT GREEN LIGHT

Another Google project that targets the impact of cars on the climate is known as Project Green Light, which emerged from the insight that intersections can have pollution up to 29 times higher than on open roads. "It reduces starting and stopping events, which is much more pleasant for the driver, and then it also reduces emissions by up to 10 percent," Brandt says. Merely shifting a few seconds of green-light time from an east-west street to a north-south street can make a substantial difference in emissions levels. Green Light is now up and running in over a dozen world cities, with plans to scale to hundreds of others within the next few years. "The result has been surprising even to me," says Matias, "that we can significantly reduce carbon emissions by a small change in scheduling—without any new infrastructure whatsoever and no dependency on people's behavior."

Green Light also overcomes a significant stumbling block to fighting global warming: the difficulty of engaging the general public in the developing and maintaining of sustainable habits. The technology does its work without requiring drivers to change their behavior in any way. Brandt compares it to her Nest smart home device, which maintains an energy-saving temperature in her home without the need for any human intervention. "The great thing about my Nest device is I don't need to think about it, but it's making an impact. In 2023, we estimate that Nest thermostats helped customers save more than 20 billion kWh of energy, enabling approximately 7 million metric tons of GHG emissions reductions," says Brandt.

FROM AMBITION TO ACTION

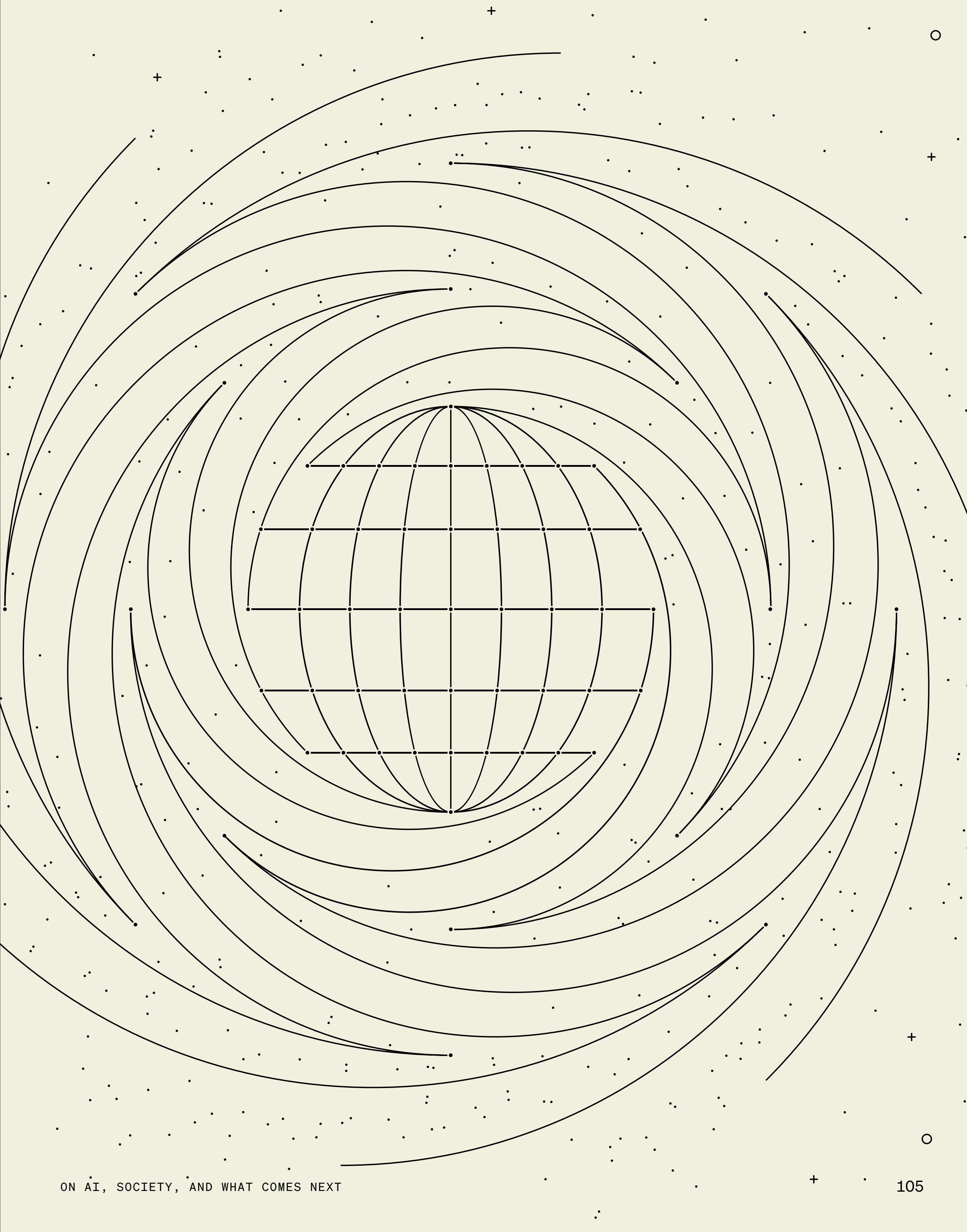
Every day, there are new and innovative ideas brought forth to combat the climate crisis. The Department of Energy has supported research at Cornell University on a bacterium that, once incorporated into a biohybrid, can use solar energy to convert carbon dioxide into chemicals such as bioplastics and biofuels. Scientists working on solar power have proposed a project that would collect sunlight in outer space, where it's both continuously available and 10 times more intense, and beam it down to Earth. In Georgia, researchers—having observed that older trees store carbon much more efficiently—are using gene editing to produce poplars that grow 50 percent faster.

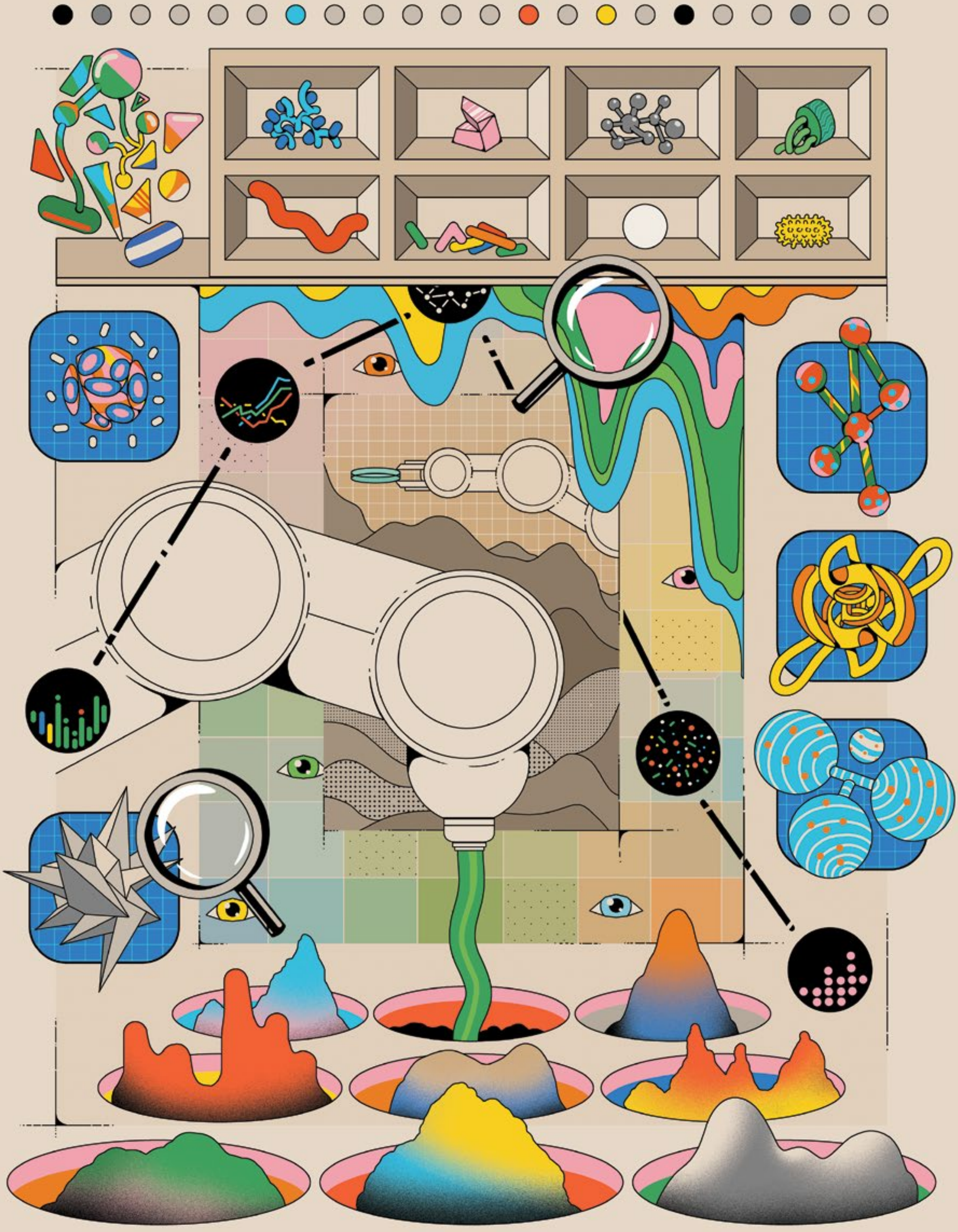
"When we think about the opportunities to tackle the climate crisis," Matias says, "obviously there are multiple approaches. One is to make our current systems more efficient. The other is to look for alternative approaches for energy." To that end, Matias's team is working on nuclear fusion—an alternative-energy source that has been a dream of physicists for a century.

Through a reaction known as ignition, fusion throws off more energy than is required to create it. In December 2023, scientists at Lawrence Livermore Laboratory in California achieved ignition for the first time in history by bombarding a peppercorn-sized fuel capsule made of diamond with 192 laser beams. In an August 2024 op-ed in *The Washington Post*, Steven Cowley, a professor of astrophysical sciences and laboratory director of the U.S. Energy Department's Princeton Plasma Physics Laboratory, argued that AI's predictive capacity has made possible many recent advancements in fusion. "I believe fusion might turn out to be AI's 'killer app,'" he wrote, "proving the value of AI to deliver truly world-saving innovations."

While fusion is unlikely to happen at scale within the next decade, it's worth noting that nuclear is another area in which Google is leveraging existing technology, having recently signed the world's first corporate agreement to purchase nuclear energy from multiple small modular reactors (SMRs). "Of course we need to be thinking long term," Brandt says, "but the world needs to get to net zero by 2050. Right now we have the opportunity to still really change the trajectory over these next five years." The immediate goal, she says, is to reduce emissions for cities, individuals, and other partners by one gigaton—roughly the emissions of the nation of Japan—every year.

"I see promise in our collective ability to make progress," Matias says of the moment we are in, "because great ideas can come from anywhere, the opportunity to address difficult problems is greater than ever. Human ingenuity is at the center of it all–advancing science and technology and applying it to the problems that matter most." Project Green Light, for instance, grew out of a dinner-table chat between a Google engineer and his wife, while the team was sifting through many ideas to help address the climate crisis. "And research advancements in AI-based flood forecasting helped make significant progress in what seemed like an impossible problem to address," says Matias. Through expert insight, intuition, and collaboration across fields, human beings are choreographing new steps to address climate change, leading a dance in which AI has become a crucial partner.





Charting the New Material World

Artificial intelligence is accelerating the discovery of new materials that will transform our everyday life.

By **Kellie Schmitt**

←
Illustration
by Ori Toor

In 1905, a young American scientist named William Coolidge set to work on an ambitious assignment for his new job at General Electric: creating a better light bulb. The carbon filament used in Thomas Edison's world-changing invention wasn't very energy efficient. Scientists were experimenting with tungsten, which had the advantage of a higher melting point, but the metal wire was too brittle and broke easily.

Several years later, after numerous trials, errors, and accidents—by some accounts, a breakthrough occurred when the tungsten rod fell into liquid mercury pooling from the heating system—Coolidge created "ductile tungsten," a more bendable form of the metal that ultimately became the wiring still found inside incandescent light bulbs today.

Ekin Dogus Cubuk, a research scientist at Google DeepMind, points to ductile tungsten as an example of how new materials can revolutionize existing technologies—often after careful, laborious efforts riddled with unexpected twists and turns. More than 100 years after that landmark innovation, his team has created an AI tool that could transform the way scientists discover new materials. Much like ductile tungsten, any of these compounds could improve upon current materials and lead to next-generation electric vehicle (EV) batteries or better solar panels. "Many materials discoveries involve a lot of trial and error, luck, and serendipity," Cubuk says. "With AI we hope to reduce the reliance on luck. We have the potential to accelerate each stage of the process, from discovery to development, with more accurate and scalable predictions."

GNoME catapults materials discovery

The AI tool that Cubuk's team built is called GNoME (Graph Networks for Materials Exploration), and it has used deep learning to determine the structure of 2.2 million inorganic crystals. Of those, GNoME identified 380,000 materials that are stable at low temperatures, according to simulations. "We hope that having this information will help the community to catapult further breakthroughs in materials discovery and design," Cubuk explains.

Since the public release of these crystals in 2023, scientists have been experimenting with AI-identified compounds in labs worldwide—including three that were recently synthesized in Japan by researchers at Hokkaido University.

"There is a wealth of AI knowledge, but currently there's a gap between translating those insights to material synthesis in the lab. This early research shows how we can bridge the gap to explore new materials more efficiently," says Akira Miura, an associate professor at Hokkaido University, who led the study.

Meanwhile, GNoME independently identified hundreds of crystals that scientists had already discovered on their own as stable, including "exciting" superconductor candidates, according to Cubuk. Furthermore, hundreds of these crystals were independently synthesized by labs around the world. That's a strong signal that the new material candidates in GNoME's findings have real-life viability.

GNoME was initially trained with data on crystal structures that is available through the Materials Project, an open-source reference that has been central to new materials discovery. Identifying the structure of a new material is just the beginning of what can be a long, intensive development process. The original idea of the Materials Project was to reduce invention time by focusing experiments on compounds with the most potential for success, according to founder and UC Berkeley professor Kristin Persson. That itself was a game changer from the days when Persson was a graduate student in the mid-'90s and had to comb through oversize tomes to look up a material's chemical structure and properties. In the laboratory, researchers would combine intuition with past results to inform their experimentation. "There were no recipes for it except in the researcher's brain," Persson says. "They would use their experience to make new materials, and it would take a long time."

Not that long ago, it was relatively rare to be able to calculate a new material on the computer that was thermodynamically stable, explains Chris Wolverton, a professor of Materials Science and Engineering at Northwestern University. For one, the sheer magnitude of the required computational power was a hurdle. "Google is particularly rich when it comes to computational time and power," he says. "They were able to—in a very short order—calculate the properties of several million compounds."

The availability of this vast data trove is opening up new possibilities to materials researchers, says Persson. "They're giving this body of work to the community and saying, 'Do what you want with it,'" she says. "It's the new kid on the block in terms of the tools we have. If you're a materials company and not thinking of machine learning, you're behind, and Google put a brilliant spotlight on that."

That doesn't mean researchers are suddenly going to have piles of new materials cluttering their lab benches. The tool has identified crystals that computations predict will be stable under absolute zero, or about -273 degrees Celsius—an ideal condition in which particle motions come to a stop. A key question persists: Can these crystals be created in a stable form that won't change or decompose at room temperature? Myriad factors, including temperature and pressure, can influence whether a theoretical compound can exist in a form that would become, say, the next EV battery. Scientists will have to figure out the right experimental methods to create materials that will be useful in innovation. "That's the next giant challenge in this field: predicting synthesis recipes," Wolverton says.

Materials scientists explore next steps

Indeed, the tremendous scale of the AI findings has spurred important conversations about the best way to identify and synthesize new materials. A researcher may spend an entire career conducting experiments on just a few compounds, a painstakingly slow and labor-intensive process. Up until now, 48,000 materials were identified at this stability. The key is finding a balance between creating AI-led possibilities and developing the known science, explains University of Liverpool professor Andy Cooper, whose research group focuses on ways to accelerate the discovery of functional materials: "I think that's a challenge for the future. What is the right place to be on that spectrum?"

New materials improve the foundations of modern life

At Google DeepMind, Cubuk envisions a future when new materials could reshape today's most exciting technologies—from semiconductors to new power sources for supercomputers. As AI tools evolve, a key challenge is determining which potential materials best correspond to technologies and products, Cubuk explains. For example, knowing

“Generative AI can predict the next big thing. As we exhaust what’s known, we have to develop new ways to push boundaries, and this is one of those paths.”

Jakoah Brgoch, associate professor of chemistry at the University of Houston

what criteria makes a good lithium-ion conductor helps researchers better identify the most promising candidates. Already, the GNoME team has identified 528 potential materials for batteries—more than 25 times the number identified in a 2017 study, which was led by Austin Sendek, the CEO of Aionics, a technology company dedicated to designing high-performance batteries with AI.

AI’s contributions might also open up researchers’ minds to unforeseen possibilities, says Jakoah Brgoch, an associate professor of chemistry at the University of Houston, whose work explores functional inorganic materials. Maybe the AI-suggested compounds aren’t viable under current lab temperatures and pressure, for example; but what happens if you alter those variables? “Science is often a slog,” Brgoch says. “The hope and the hype is we don’t have to go through this slog. Generative AI can predict the next big thing. As we exhaust what’s known, we have to develop new ways to push boundaries, and this is one of those paths. Of course, experimentally validating these predictions is easier said than done.”

According to current industry wisdom, new material discovery requires, on average, at least 10 years of work and north of $10 million to $100 million. While the cost to build a factory is the same, what used to take weeks to simulate on a computer can now be done in seconds with AI, says Sendek.

“AI-based materials discovery methods will unlock unorthodox solutions that would have otherwise eluded human intuition, while also leading to fewer ‘dead ends’ that cannot be commercialized,” says Sendek, pointing to the recent discovery by Aionics’s AI system that adding a luxury perfume molecule to a battery’s electrolyte could improve the battery’s overall performance. Aionics is now actively testing this molecule in real-life batteries.

“Making meaningful progress in materials innovation requires us to discover better materials faster and then get them to market in a fraction of the time.”

Riddles of the

A new generation of high-powered telescopes is creating petabytes of data about comets, stars, and distant galaxies—forever altering the way that we understand the cosmos. The result is an embarrassment of riches, and the inherent challenge of sorting, analyzing, and creating meaning out of it. Next year's opening of the Vera C. Rubin Observatory in Chile, which will have the largest digital camera ever built—a 3,200-megapixel device the size of a car, yielding unprecedented images of space—will only exacerbate this challenge. "We're swamped with data, too much for anyone to go through, which means a lot of it doesn't get used," says Karl Gebhardt, the department chair of astronomy at the University of Texas. "That's exactly where AI comes in."

Gebhardt and his team are using AI to spot patterns in some 1 trillion elements within 1 billion images of space taken by the McDonald Observatory in West Texas. To help train the AI model, Gebhardt created an app for "citizen scientists," inviting volunteers to view the department's telescopic photos and swipe right if they think the image contains a valid star or swipe left if not. "We went from five people doing this initial sorting to 30,000 people working on it," Gebhardt says. He and his team then fed those vetted images into an algorithm that searches for probable stars and galaxies. "It's really improved our work," Gebhardt says. "It's a hybrid approach, with crowdsourced humans training the model and then AI taking it from there."

Humans defining the training set is key to the success of the AI algorithm, as astronomers must ensure that AI findings are valid. For instance, an algorithm may not be able to detect wonky data that results from an error by a telescope's digital camera. The human eye is also better than AI at spotting some features in the imagery, such as

↔
Illustration by
Rida Abbasi

Night Sky

Astronomers have turned to artificial intelligence to explore fundamental questions about the universe.

By Stayton Bonner

when a fast-moving particle creates a splash of light. "The eye just picks that up instantly," Gebhardt says. "But it's really hard to write code that identifies that on a consistent basis."

In July, astronomers at the University of Texas used generative AI to develop an algorithm to discover stars in the final stage of life, known as dwarf stars, which contain important clues to the elements that make up the planets in our galaxy. Dwarf stars have historically eluded astronomers. They are difficult to discern and identify because they don't emit much light. By using an algorithm to group visually similar items together, the UT astronomers pinpointed 375 promising-looking stars out of 100,000 possible white dwarfs—and then followed up with their telescopes to confirm the findings.

Ultimately, Gebhardt and other astronomers hope to use AI to probe one of the big questions looming in the night sky: Why does the universe expand? So far, the data collected by astronomers hasn't been able to yield an answer. The astronomers could be missing a key insight in the data they've gathered so far, or their hypotheses about how the universe is expanding could be fundamentally wrong. Either way, AI may help astronomers find overlooked patterns in the existing data that shed light on universal expansion or point them to an unexplored region of the sky that could finally provide the breakthrough they need. In the process, an AI-assisted insight could transform our theories of the Big Bang. "We don't understand the physics of this expansion," Gebhardt says. "We don't even understand how gravity works. So hopefully we can use AI to answer some of these questions and better understand the universe at a fundamental level."

CHAPTER

4

Getting it right

Intro

Society is facing a pivotal challenge: How do we harness the immense potential of AI while mitigating risks and ensuring it serves the greater good? How do we regulate a technology that evolves faster than traditional policy-making processes?

Ethical considerations permeate every aspect of AI development. Questions of transparency, accountability, and fairness arise as AI systems become more deeply embedded in critical decisions and social institutions. How do we ensure these systems operate in ways that align with society's values and interests?

The effect of AI on labor markets is not a simple story of job displacement, but a complex restructuring of entire industries and professions. How do we ensure that the benefits of AI-driven productivity gains are distributed fairly? What new forms of work might emerge in an AI-dominated economy, and how do we prepare people for these roles?

As we navigate these complex questions, one realization becomes clear: The path forward is neither predetermined nor easy. It will require ongoing dialogue, careful consideration, and a commitment to aligning AI development with our deepest values and aspirations for humanity.

Q&A

The Next Steps for Responsible AI

A conversation with Google's Kent Walker about how we can guide the new generation of artificial intelligence

By Martin Ford

→
Photography by
Cayce Clifford

The past 18 months have seen the rapid worldwide adoption of artificial intelligence by workers, scientists, entrepreneurs, doctors, and dreamers. I spoke with Kent Walker, president of global affairs at Google and Alphabet, about the most important issues that we face in harnessing this new technology. How do we keep AI development vibrant, and more importantly, how will AI assist in human development?

This conversation has been edited for length and clarity.

Martin Ford **During the past year, most of the discussion around artificial intelligence has centered on chatbot systems such as Gemini. What do you think is the most overlooked aspect of artificial intelligence?**

Kent Walker That chatbots only scratch the surface of what this technology is capable of. AI is not just a scientific breakthrough; it's a breakthrough in how we make breakthroughs. Look at the example of AlphaFold, the protein modeling technology, where we're making generational advances in the tools used by medical researchers around the world. This is just one example of many where AI will accelerate scientific progress, whether in materials science, new forms of energy, quantum computing, health care, or water desalination.

Ford **One of the tangible benefits of AlphaFold is the cataloging of every protein molecule important in biology. I'm not sure chat systems have produced anything comparable. Do you think there's a danger that we're overhyping chat systems and not giving enough attention to other important innovations?**

Walker Yes, I think there is a framing challenge, and it's natural. Not everyone can be a biology Ph.D. researcher advancing the frontiers of science. But because chatbots are so accessible and tangible, anyone can interact with them and feel as though they're experiencing artificial intelligence. However, the work being done by thousands of medical researchers worldwide will likely have a more foundational impact on how we make progress against diseases like cancer or develop personalized medicine. From a regulatory perspective, it's important not to think of AI as social media 2.0 and fight the last war. Instead, we should think of this technology primarily as a significant scientific breakthrough, such as mRNA vaccines, that has broad applications for improving human welfare.

Ford **What trends do you see emerging in the regulation of AI? Are there any developments that concern you and could potentially hold back the technology?**

Walker As I mentioned in the last issue of *Dialogues*, AI is too important not to be regulated but also too important not to be regulated well. We've seen various governments and private actors around the world take significant steps and generally think constructively about this issue. There's a concept known as the Collingridge dilemma, from a book called *The Social Control of Technology* by David Collingridge, which points out that it's possible to regulate a new technology too late, once it's already locked in, but it's also possible to regulate it too early, before understanding its full potential and risks. The challenge is finding that sweet spot where you regulate a fast-moving technology in the right ways, recognizing its potential for scientific progress while also considering risks like discrimination, unsafe applications, or labor displacement. That's an important balance to strike.

Our suggested framework for good AI regulation is that it should be FAB, focused, aligned, and balanced. Let me break those down.

Focused means recognizing that AI is a general-purpose technology, more like electricity than a special-purpose tool. We don't have a single agency for all uses of electricity or a single law governing all engines in society. Instead, we approach it case by case. The issues in banking will differ from those in health care or transportation. Traditional regulators in these areas have years of expertise, but they need to become familiar with AI so they can understand the new challenges in their domains. Every agency needs to become an AI agency.

Aligned speaks to the need for coherence among different groups working on these issues—whether it's the G7, the United Nations, the United States, or other countries. While regulations don't need to be identical, they should be broadly consistent to allow the wider adoption of these tools, including in the Global South and developing countries.

And by *balanced*, I mean evaluating the relative risks of different kinds of applications. The use of AI in search results is different from the use of AI in high-risk applications such as health care or transportation. You need different regulatory regimes for these different settings. It's also important to remember that high-risk applications often bring high value, such as the potential to save lives in medicine. Even as we regulate these areas, we want to ensure that we're not slowing down progress in saving lives and solving problems.

Ford **In some areas, such as self-driving cars, there is a department of transportation that is well-equipped to handle regulation. The same goes for medical or financial sectors. But there are areas where gaps exist, and some issues that are entirely new to AI, like disinformation … there isn't a clear regulatory body for those problems.**

Walker I think that framework is exactly right—you should be looking for gaps in existing laws. The starting point for analysis should be that if something is illegal without AI, it's probably illegal with AI. This covers many of the concerns about fraudulent applications of AI. However, there are new questions of degree, as with content moderation. It's always been against Google's policies to have manipulated media, such as deepfakes created by slowing down videos to make someone look drunk or editing videos in misleading ways. AI enhances this manipulation and scales it up, thus

“AI is not just a scientific breakthrough; it’s a breakthrough in how we make breakthroughs.”

Kent Walker, president of global affairs at Google and Alphabet

raising the question of, When does the risk become significant enough that we need new laws, and when is it better and more consistent to extend the enforcement of existing laws?

Ford **Are there examples of governments around the world that are doing a particularly effective job of regulating AI—places that might serve as a role model?**

Walker Yes, Japan has been a leader in this area. Japan faces a significant demographic challenge, and it has more robots per capita than any other country because it needs to increase productivity. They view AI as powering a new generation of robotic assistance that can apply to a variety of jobs to improve quality and reduce costs. Singapore has also been very forward-looking, focusing on understanding new technologies before imposing formal rules. They are clarifying existing laws, introducing flexible frameworks and pro-innovation policies like copyright exemptions for AI training, and working with companies to promote AI adoption.

Ford **One criticism of significant AI regulation is that large companies like Google can afford the burden of compliance, whereas start-ups might find it more difficult. This could create a regulatory moat that protects established players and stifles innovation. Is that a legitimate concern?**

Walker Yes, we have to be very careful about that. We’ve seen examples, like some of the data legislation in Europe, where regulatory barriers to entry have made it

> "It's possible to regulate a new technology too late, once it's already locked in, but it's also possible to regulate it too early, before understanding its full potential and risks."

Kent Walker, president of global affairs at Google and Alphabet

harder for smaller companies to flourish. It's important to be mindful of the impacts of regulations not just on large companies but on smaller ones as well. Each generation of technology creates a new opportunity for new companies to emerge. Regulators must consider not just their specific sectoral concerns but also the wider need to promote innovation, productivity, and global competitiveness.

Ford **One of the most difficult issues facing artificial intelligence right now is the use of copyrighted material to train AI systems. It wouldn't surprise me if the Supreme Court eventually has to make a ruling on this. How should regulators and legal scholars approach this problem?**

Walker It's a critical issue. We wouldn't have the AI breakthroughs that we've had without the concept of fair use, which allows for the transformative use of publicly available information. The challenge is to figure out how to appropriately compensate creators contributing material value without creating a blanket restriction on the use of online information in AI models. We're working with publishers to better understand the value of different types of content in model training and operation. I'm hopeful we can find a way to recognize contributions while continuing AI development.

I also think it's important to focus on regulating outputs rather than inputs. The question is whether a given image or text infringes on someone's rights, not how it was created, whether with a pencil, a computer, or AI. It's about the real-world impact of these tools.

Ford **How do you think AI will impact the job market and the nature of work?**

Walker While we know it will, it's much harder to project precisely how. It may change not just how we work but how we prepare for work. We might see a shift from a single four-year degree to multiple shorter, skills-focused programs spread throughout a career. AI generally automates tasks rather than whole jobs, so the impact varies by job. We're trying to democratize access to AI tools through initiatives like our AI Essentials course, which helps people learn how to use these tools effectively.

Ford **But is AI just another tool, like the typewriter or spreadsheet, or is it in a class by itself—something that will be super transformative?**

Walker Amara's Law says that we tend to overestimate the impact of a new technology in the short term and underestimate it in the long term. Right now, people are using AI to streamline existing tasks—saving costs. The next step will be to increase revenues by doing things we couldn't do before. Over a decade, we'll likely see broader applications that change how we think about job requirements and the fundamentals of knowledge work.

In the words of Andrew McAfee, our inaugural Technology & Society visiting fellow, general-purpose technologies like AI tend to "reduce demand for some skills, increase demand for others, and create demand for entirely new ones."

Remember that 60 percent of the jobs that we have today didn't exist 80 years ago. As we saw with the advent of electricity or the internet, we should expect to see new categories of jobs on the horizon—jobs that we can't yet predict.

Ford **Do you think AI could lead to significant job displacement in some industries, like call centers?**

Walker A study we recently commissioned estimates that around 61 percent of jobs will be augmented by generative AI, with 7 percent transitioning over the long term. That's why it's so important that the public and private sectors work together to lay the groundwork for AI-driven job evolution.

You mention call centers, and that's a place where we're actually seeing higher employment satisfaction because of AI. Call agents spend their days pattern-matching problems—identifying underlying problems and thinking about the most common solutions. With AI in the mix solving the more routine problems, agents have more time to focus on the most complex requests.

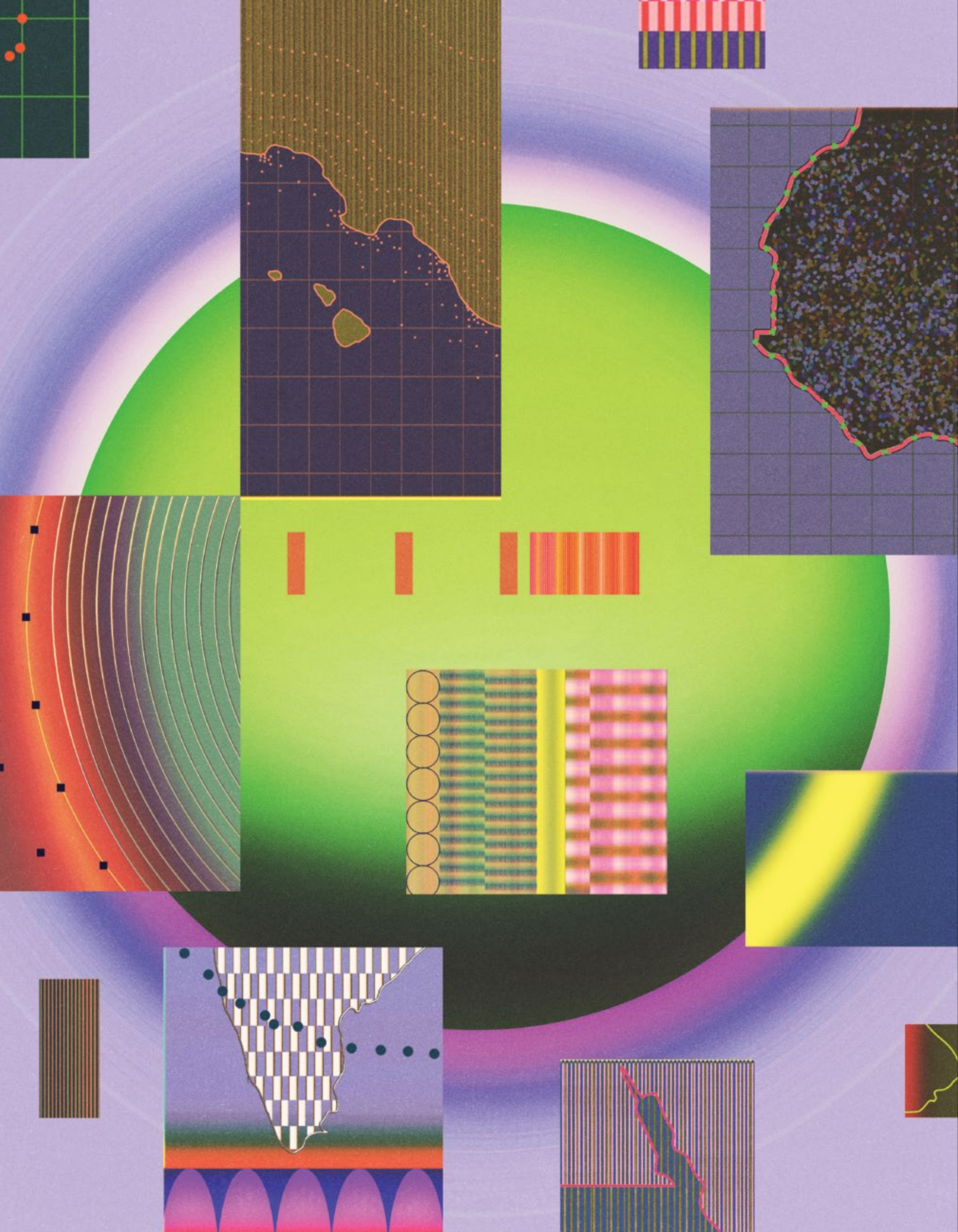
By way of example, Google Cloud has been working with partners at Discover Financial Services to build generative AI into Discover's call centers, training and tuning our LLMs with a set of frequently asked questions. Early results are incredibly promising, with agents reporting that they're more productive and able to significantly reduce the time it takes to get to the root of customer issues. It's a new way of thinking about customer interaction—one that not only improves customer satisfaction but also has the potential to spark completely different ways of working.

Ford **Do you think traditional strategies like retraining will be enough to address AI's impact on the workforce, or will we need to consider more radical policies like universal basic income?**

Walker It's an open question, although I tend to think people will always find meaning and value in work. We're investing in initiatives to explore this issue. We have our Global AI Opportunity Fund, which will invest $120 million to make AI education and training available in communities around the world, and collaborations like the one we have with MIT RAISE to offer no-cost AI courses to educators. It's early in the development of this technology, and we want to contribute to finding solutions.

Ford **Do you use AI in your work, and how has it changed things for you?**

Walker I do! Right now I'm using it to distill long articles into key bullet points, which is like having a personal summarizer. I'm also working on a project to have an AI read all my emails from the last 18 years at Google and then learn to respond in my voice, helping me to more quickly create first drafts that I can then refine and improve. The more AI takes on the routine stuff, the more time I will have to focus on the nonroutine and rewarding parts of my job. And I think, like me, many people are going to appreciate the ways AI frees them up to dig into the sticky questions, think about what's next, and focus on all the fun stuff that makes their job feel meaningful.



Around the Globe, Governments Lean Into AI

Governments around the globe are in a high-stakes race to develop cutting-edge AI systems. But how exactly are they using the technology for their own operations, and what might their progress tell us about the growth and development of AI worldwide?

By Mark Esposito

We tend to think of AI as the domain of Silicon Valley and Wall Street. Indeed, the United States has long been at the forefront of AI development. In addition to its robust private sector, its federal agencies employ AI for everything from synthesizing veterans' feedback about service quality to predicting extreme weather events. City governments are also using AI. The City of Memphis, Tennessee, for example, has worked with Google to use AI-enabled cameras mounted on city vehicles to identify and classify potholes for filling. Together with other data-powered methods, the effort helped identify 75 percent more potholes than manual processes did, and simplified a task that typically takes 32,000 hours of city employees' time each year.

But other countries are racing to develop and implement AI systems of their own. From AI-powered chatbots guiding citizens through bureaucratic roadblocks in Portugal to algorithms optimizing urban planning in Singapore, governments are exploring diverse AI applications. Each region's historical, economic, and social context shapes its approach to advancing AI. Some regions shine in certain areas but fumble in others, each offering valuable lessons on what works, what doesn't, and what should be avoided when it comes to AI.

Each region's historical, economic, and social context shapes its approach to advancing AI.

Europe

Balancing Innovation and Safety

Europe boasts one of the world's most comprehensive regulatory frameworks for data privacy and security, the General Data Protection Regulation (GDPR), and is now introducing the EU Artificial Intelligence Act. This new legislation aims to streamline AI policy by intricately categorizing AI systems based on risk levels, with strict regulations on high-risk applications that could potentially exploit people's vulnerabilities, manipulate their decisions, or classify them based on personal traits, as is the case in predictive policing.

While designed to protect citizens, the stringent regulations could potentially stifle innovation or slow the adoption of beneficial AI technologies. "If the idea is to facilitate the development and the deployment of low or non-risk applications, the risk division is on the one hand fuzzy and on the other hand too rigid to handle specific contexts," Virginia Dignum, who sits on the UN High-Level Advisory Body on AI, says.

While Europe is often perceived as cautious, its approach to AI is quite diverse and multifaceted. Estonia, Finland, and Denmark have made significant strides in incorporating AI into government operations and are often lauded as examples of how to do it right.

But other countries are catching up, recognizing the potential of AI to transform public services and improve citizen engagement. Portugal, for example, launched a chatbot through its Justice Practical Guide 2023 to answer citizen queries about marriage, divorce, and setting up a new business. These chatbots are designed with privacy and data protection in mind, adhering to the strict standards set by the GDPR.

"What works in Denmark may not work in Portugal or Slovenia," says Dignum. Getting it right requires special attention to value-alignment processes. "What is being optimized here: service efficiency or citizen care?" She stresses the need for participatory, transparent, and democratically evaluated implementation processes. "It should never completely block access to person-to-person interaction [and should] account for a diversity of citizens."

This approach ensures that AI enhances rather than replaces human interaction in government services.

Middle East

Ambition Meets Practicality

The Gulf Cooperation Council (GCC) countries, particularly the United Arab Emirates, Saudi Arabia, and Qatar, are fully committed to embracing AI. They've positioned themselves as AI powerhouses, laying forward a series of ambitious national strategies, such as the UAE's National AI Strategy 2031 and Saudi Arabia's Vision 2030.

The challenge in the region now is translating these forward-thinking plans into practical

90°S 150°W 60°W 30°W

applications that integrate AI into every aspect of public life, from education and health care to transportation and government services.

"GCC countries are equally focused on practical AI applications that impact daily life," Jassim Haji, a former IT executive at Gulf Air and prominent AI expert in the region, says. "The emphasis is on using AI to enhance government services, improve public health responses, and streamline citizen interactions with authorities."

This commitment is evident in various initiatives across the region. For example, Saudi Arabia's e-government portal, my.gov.sa, uses AI-powered chatbots to guide citizens through complex public services. In Bahrain, the BeAware Bahrain app, launched in response to the coronavirus pandemic, uses AI for contact tracing, facial recognition, and exposure alerts. "With over a million downloads, it highlights Bahrain's digital readiness, supported by high mobile penetration and a robust ICT infrastructure," Haji says.

Bahrain's Tawasul platform, which "integrates AI for natural language processing, chatbot assistance, and predictive analytics to handle citizen inquiries and complaints," is another practical application that balances ease of use with user privacy by adhering to local data protection laws, according to Haji.

Despite these advancements, the region faces significant challenges in fully realizing its AI ambitions. "The most significant challenges include addressing the local AI talent gap, ensuring the availability of high-quality data while safeguarding privacy, adapting regulatory frameworks to keep pace with technological advancements, and integrating AI into existing infrastructure," Haji explains.

Asia

Pragmatic Approaches and Rapid Adoption

Across Asia, the approach to AI adoption in government services reflects an ethos that prioritizes tangible applications and real-world results, often as part of broader digital-transformation initiatives.

Asia, as a whole, is a hotbed of digital innovation, thanks to the rapid expansion of digital public infrastructure and hefty investments propelling tech growth. Economies here are not just testing the waters; they're diving in headfirst, quickly scaling these technologies to new heights.

Asia is a hotbed of digital innovation, thanks to the rapid expansion of digital public infrastructure and hefty investments propelling tech growth.

Singapore, in particular, has been a frontrunner in integrating AI for urban planning and management as part of a wider approach to digitalization. "Smart Nation is Singapore's overall strategy for digitizing the nation, and AI is one component," explains Jun-E Tan, an AI governance researcher. The initiative leverages AI and IoT technologies to optimize traffic flow, energy consumption, and more.

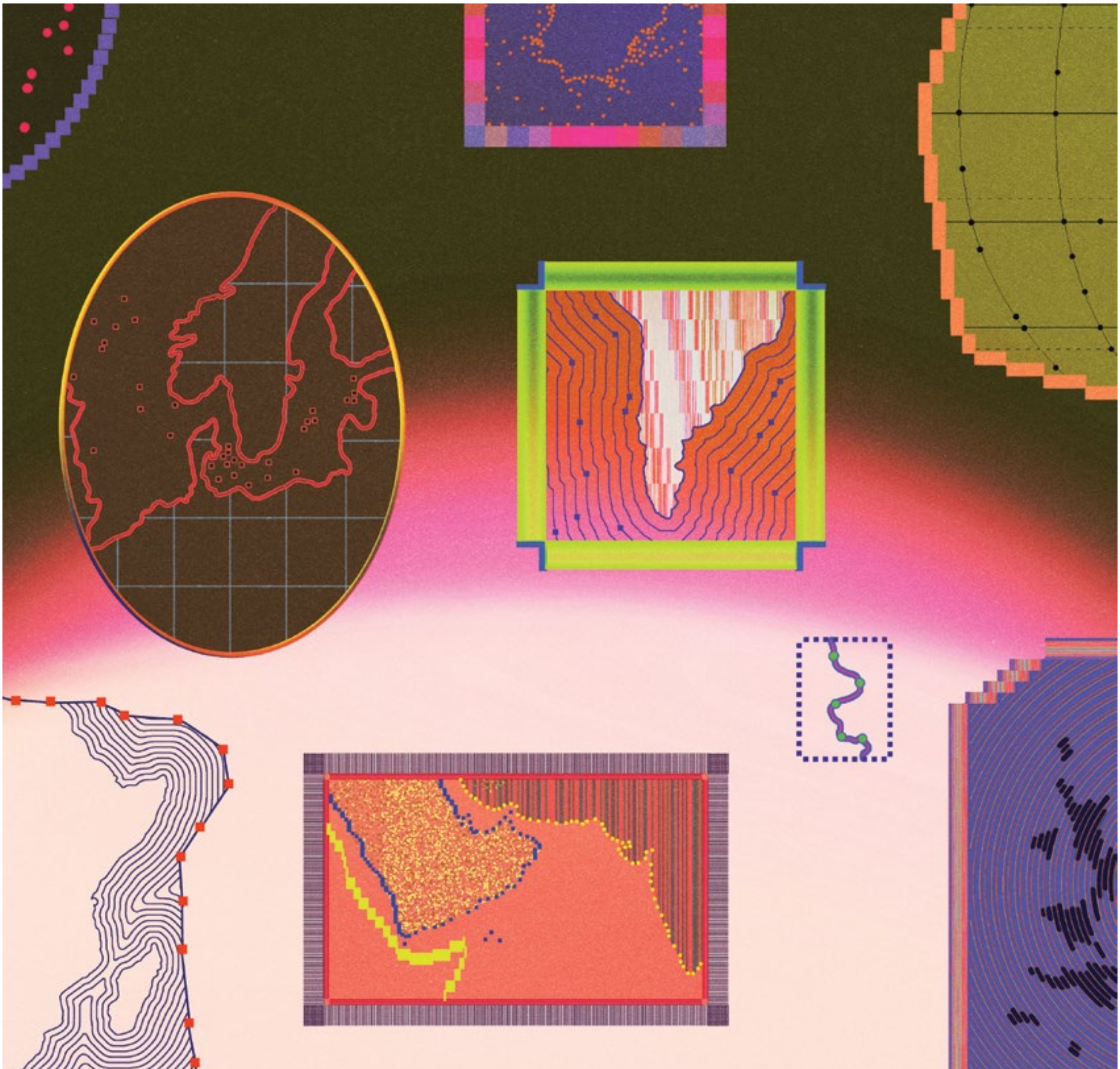
Other Southeast Asian nations are following suit, albeit at varying paces. Malaysia's journey, while less publicized than Singapore's, is equally telling of the region's approach. The country has laid out a series of policy frameworks, including the Malaysia Digital Economy Blueprint and the National 4IR Policy, positioning AI as a keystone in its digital arch. Yet the on-ground implementation remains in its infancy.

"The use of AI for government services is still nascent within Malaysia," Tan says.

This measured approach is not unique to Malaysia. Across the region, governments are prioritizing readiness and foundational understanding over hasty implementation.

Take India's Unified Payments Interface, which deploys AI-powered chatbots to simplify financial transactions, eliminating the need for users to navigate through complex menus. It's a stellar example of how pragmatic, hands-on approaches catapult these regions to the forefront of the AI revolution.

India's success lies in its incrementalist approach. The country focuses on gradual, practical development goals while prioritizing reliability, security, and safety in its platforms. China, on the other hand, offers a more controversial case study with its social credit system. This algorithm assigns citizens a score based on traits such as citizenship and civic behavior. While it raises significant surveillance concerns, it demonstrates how the



boundaries are being pushed in what's possible—and what's acceptable—in the realm of AI-driven public policy.

Africa

Leapfrogging Traditional Development

Africa is experiencing an AI revolution, driven by a young population, mobile technology, and a determination to overcome challenges. This has allowed the continent to leapfrog traditional development stages and emerge as a global innovator in AI adoption.

"In Mozambique, health-care workers are using AI to detect tuberculosis, and in East Africa, they are integrating AI tools into weather forecasting to better prepare for climate events," notes Rob Floyd, the director of Innovation and Digital Policy at the African Center for Economic Transformation.

Mobile technology plays a pivotal role in this transformation. AI-enabled platforms are breaking barriers, making education, financial

90°S 150°W 60°W 30°W

services, and health diagnostics accessible to populations previously left behind. "For public service delivery, it is not one area of AI application, but rather the cumulative development impact when more people have access to better and faster services," Floyd explains.

In urban centers, a different kind of AI revolution is unfolding. Metropolitan hubs such as Lagos in Nigeria, Nairobi in Kenya, Cape Town in South Africa, and Cairo in Egypt are becoming meccas of AI innovation as they work toward smart city solutions. "Nairobi is adopting AI tools to help manage traffic congestion, and Lagos is using AI to predict public infrastructure weaknesses," Floyd notes.

The contrast between rural and urban AI applications highlights both the versatility of the technology and the unique challenges faced across the continent. As Floyd emphasizes, "At this point in time, it is critical that nations, cities, utilities, and other service providers are sharing knowledge and lessons learned."

Latin America

A Patchwork of Progress

Latin America, spanning more than 7 million square miles and encompassing 33 countries, is the unsung player in the global AI scene. But this region's approach to AI is as varied as its cultures and landscapes. Economic disparities between its nations split the region into AI front-runners and laggards.

Chile, Brazil, Uruguay, Argentina, and Colombia are leading the way in AI development. Argentina's Prometea, an AI virtual assistant for judicial officials, showcases the potential for AI in government services. This system predicts case solutions and provides information to assemble case files, freeing judicial officials from repetitive tasks and potentially streamlining the justice system.

But this progress is far from uniform. While some nations sprint ahead, others lag behind, held back by a preexisting digital divide. Other countries grapple with underdeveloped infrastructure and limited resources and find themselves on the slower track of AI adoption.

"LATAM 4.0 tackles the fragmentation of AI development across Latin America and the Caribbean by offering collaborative instruments for governments, businesses, academic institutions, and civil society," explains Felipe Castro Quiles, co-founder of the LATAM 4.0 project. "We aim to enhance cooperation, share development resources, align strategies in AI initiatives, and centralize public data while decentralizing decision-making."

This fragmentation is evident in the region's approach to emerging technologies. Take cryptocurrency, for instance. El Salvador made headlines by adopting Bitcoin as legal tender, aiming to boost financial inclusion, and the country is now pursuing ambitious AI initiatives through a Google partnership to modernize government services. In stark contrast, Bolivia banned cryptocurrencies outright, citing fraud concerns. Meanwhile, Brazil and Mexico have taken a middle path, embracing established crypto exchanges and global financial networks.

Unlike the European Union, Latin America lacks a central governing body to coordinate AI efforts across the continent. Each country charts its own course, guided by its unique resources and priorities. This decentralized approach has its advantages, as it allows for tailored solutions, but it also presents challenges in creating a unified regional AI strategy.

Public-private partnerships (PPPs) offer a promising path forward. Castro Quiles advocates for a partner-shoring approach: "This includes tailoring AI models to local industrial strengths that drive development based on native contexts. An updated PPP approach can foster a more inclusive digital ecosystem by leveraging local knowledge and expertise."

While some nations sprint ahead, others lag behind, held back by a preexisting digital divide.

As Latin America navigates its AI future, the key to success may lie in harnessing the power of collaboration. By pooling resources, sharing expertise, and creating cross-border synergies, the region can amplify its strengths and mitigate its weaknesses. The path ahead is challenging, but with unity and innovation, Latin America has the potential to carve out its own unique space in the global AI landscape.

FEATURE

The Balancing Act

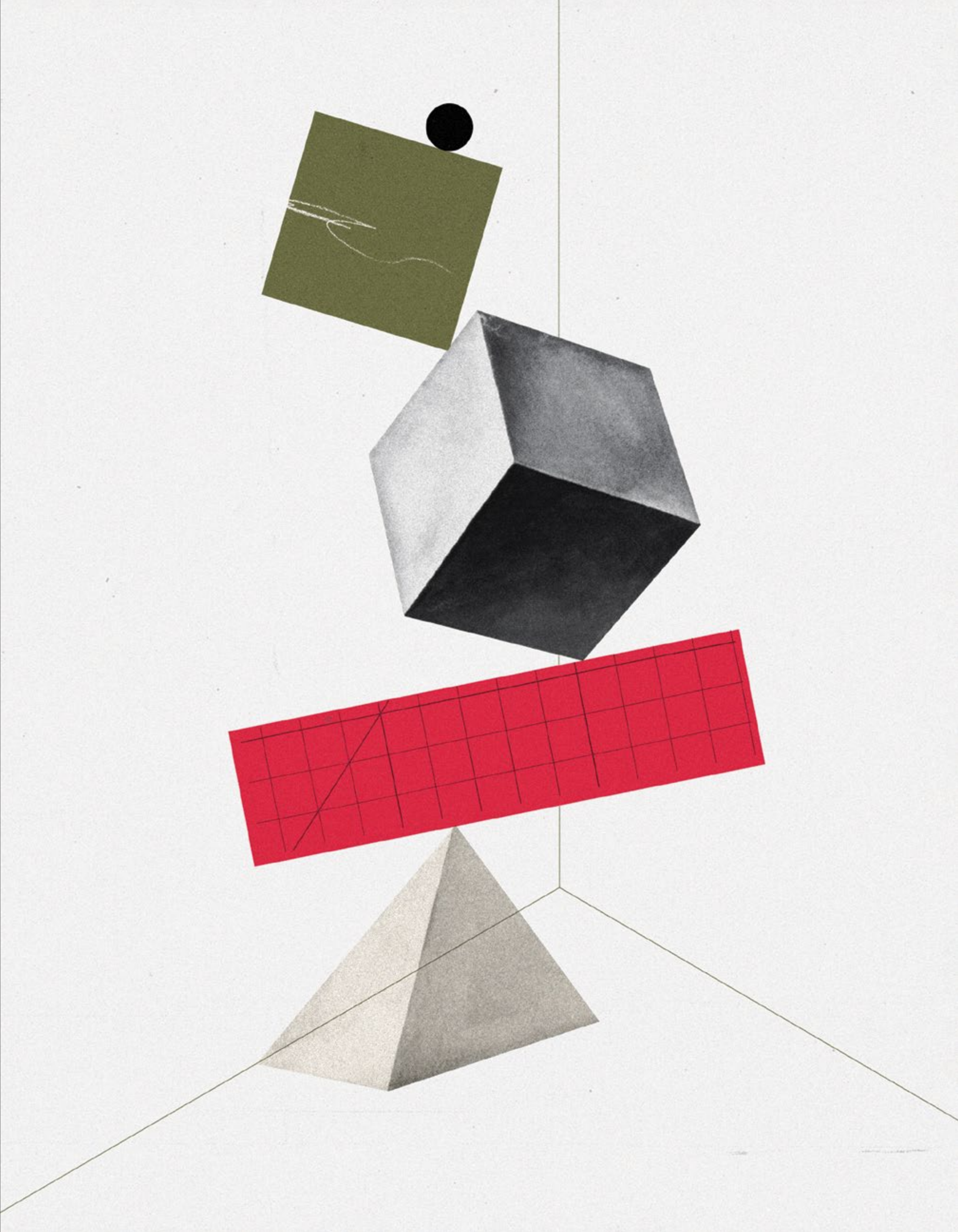
Some people believe that AI models should be customizable and transparent to encourage innovation and trust. Others believe that the technology is too powerful and must be properly safeguarded. The reality is more complicated.

→
Illustration by Laura Scofield

The advocates of open-source AI cite the benefits the approach has brought to our digital lives: Open-source software speeds innovation, the code becomes more adaptable and more resilient to attacks, and more people benefit from the software's capabilities. Open-source advocates believe these same principles should carry through to AI, democratizing its development. They argue that we should know how generative AI systems are being built and how they work, and that this transparency could lead to a greater trust in AI as it becomes a partner in medicine, scientific research, the creative arts, and many other human activities.

Open-source models make their underlying architecture public, allowing for inspection, modification, and reconfiguration. In contrast, closed models keep their architecture, training data, and methods private. Access to these models is restricted, giving developers greater control over their use and preventing potential misuse. It's also important to distinguish between different levels of openness. For example, open-weight models represent a middle ground by releasing the final trained parameters (weights), providing some insight into the model's internal workings. However, crucial details like training data and methods remain hidden, limiting the total transparency and making it difficult for those other than the model's developers to fully understand or reproduce the model's behavior.

The advocates of a more closed AI hold that AI is a special technology in both the power of its analytical abilities and the potential scale of its effects. If a truly advanced frontier AI model were to be open-sourced, it could place too much potential for harm in the hands of someone who chooses to modify it for nefarious purposes. These advocates argue that keeping AI closed—or, more specifically, not making the model or the model weights accessible to the public—provides a better environment to correct biases in the data, helps ensure the integrity of the model's output and inferences, and safeguards the model from being reconfigured by bad actors. From this perspective, it would be irresponsible to let large language models (LLMs) be accessible without accountability or limitations on their deployment and use.



There are passionate, committed advocates who are thinking of the best approaches to just how open AI should be.

Lawrence Lessig is a professor of law at Harvard University and a longtime open-source advocate who is inserting nuance into a debate that often seems overly binary. "I had an exchange recently with someone who pooh-poohed the idea of any risk, because in their words, they didn't see a Terminator risk," Lessig says. "But it didn't take long to get him to a place where he actually saw that there's so much more [to worry about] before we get to the Terminator question." In Lessig's view, the dangers of AI do not lie in the imagined future of a runaway superintelligence, but in the here and now, when we need to prevent such things as deepfakes, misinformation campaigns, and nonconsensual intimate imagery, as well as abuses by terrorist groups and hostile state actors.

In conversations with policy makers, Lessig has detected misconceptions about how AI is built: "Most people have a crude understanding of this tech, and they assume developers can just write a line in code that says, Do not do bad stuff." He emphasized how the training of AI is a complex task. "You don't just tell AIs to behave well, so it is a very complicated thing to figure out how we get the technology to be safe." Researchers are working extremely hard on the problem that Lessig describes: How do we conform the AI models to human values, and how do we understand the decisions the models make?

One possible solution proposed by Lessig is a two-tier model in which the most powerful AI systems are closed, but elements of open-source experimentation are allowed. "There's a line below which we ought to be encouraging open-source, because it's an incredible engine of equality across the world," he says. "But there's a level at which it becomes potentially dangerous. I think it drives us to think in a bigger way about the infrastructure within which we can regulate. So the commitment to open-source might be a commitment to find a different way to regulate for safety, not by controlling the models, but by controlling the environment within which the models are running. If that infrastructure existed, I think the freedom for open-source could extend to much more powerful AI models." Some organizations, such as MLCommons, the Frontier Model Forum, and the nonprofit Partnership on AI, are already working on AI ecosystem safety efforts.

The concerns around open-source only grow larger as the models become more capable, with long context windows, memory, planning, and tool usage enabling "agentic" behavior, which is the ability of an AI model to execute commands in the real world—for good or ill. Monitoring activity on closed models might limit the kinds of abuses that might result.

Another perspective comes from Percy Liang, an associate professor of computer science at Stanford University, who's also searching for the best way to proceed responsibly with AI and is generally a proponent of open-source approaches. In his role as the director of the Center for Research on Foundation Models, Liang guides a group of scholars who study the deployment and development of major AI models. One of his basic goals, he explains, was to refine the narrative around the open-source conversation. It's helpful to view AI as part of a progression of general-use, foundational technologies. "We don't get upset about people using email," Liang says, "which, in the hands of a malicious actor, people can [use to] generate a lot of spam and disinformation." He cited computer code itself as another technology that could be considered "very dangerous" since it is used to create viruses and cyberattacks—when it's not, you know, doing the work of propping up the fundamental processes of modern society.

A similar debate about enabling a powerful technology versus protecting the public swirled around encryption in the '90s. The government wanted access to encrypted communications to protect national security, while civil libertarians argued that individuals had a right to protect their privacy. "Eventually," Liang says, "it was sort of decided that we were going to live in a world where encryption was available to everyone."

Liang uses the metaphor of Legos to demonstrate the difference between open and closed models. A closed AI

> Open models have transformed data analysis in many aspects of scientific research: "You can do things that you could not do before."
>
> **Percy Liang**, associate professor of computer science at Stanford University

> "There's a line below which we ought to be encouraging open-source, because it's an incredible engine of equality across the world; but there's a level at which it becomes potentially dangerous."

Lawrence Lessig, professor of law at Harvard University

model is like buying a toy from the store and playing with it; an open-weight AI model is like having a collection of Legos and cobbling together a custom creation. When asked to cite a notable benefit of open approaches, Liang described how open models have transformed data analysis in many aspects of scientific research: "You can do things that you could not do before."

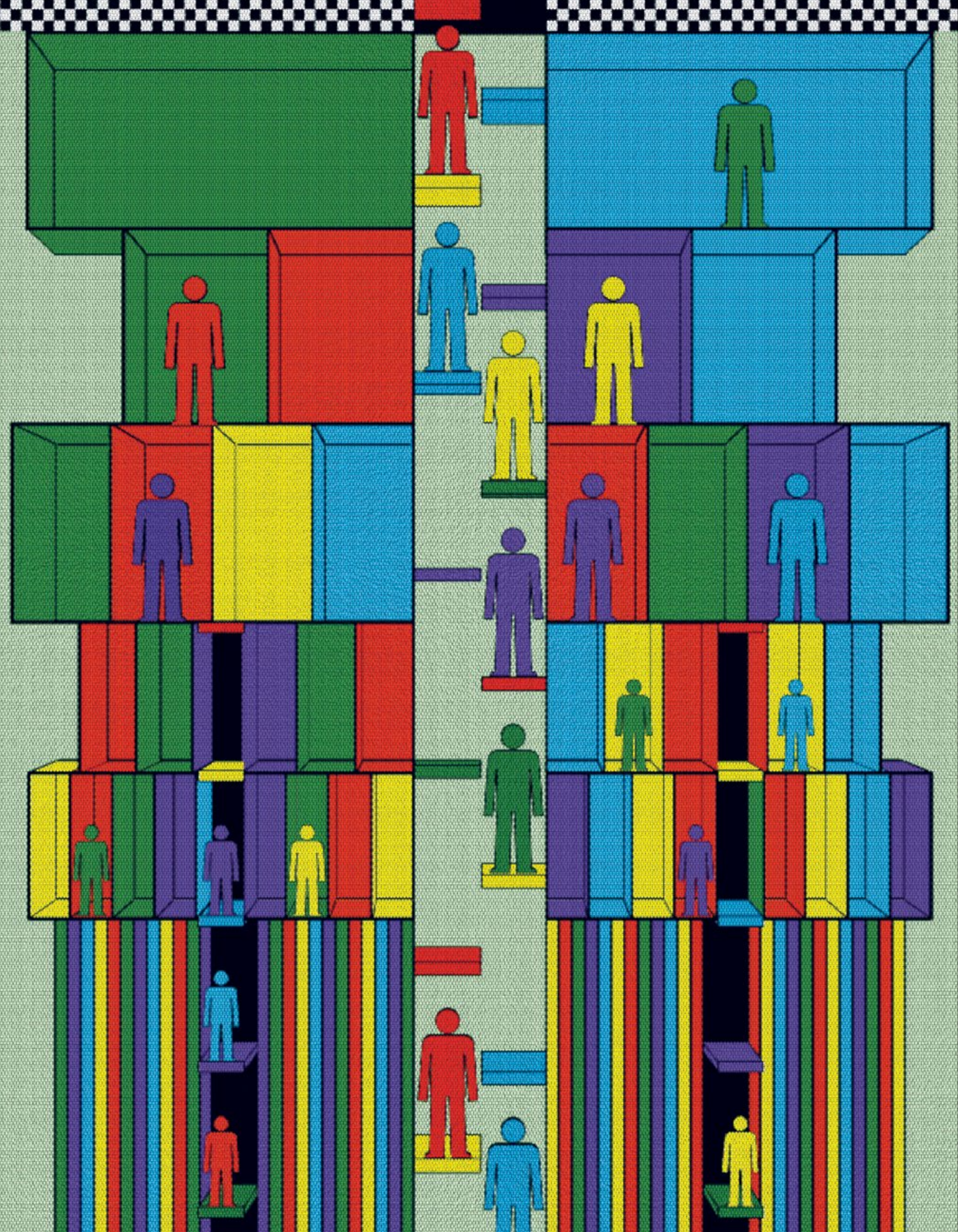
When it comes to the Terminator-like scenarios that make people fearful, Liang counsels "not running and jumping to conclusions" and encourages people to adopt a more measured view of the likely impacts. A headline-making paper published by researchers at MIT in 2023 looked at whether or not LLMs could help build bioweapons. The paper's conclusion was bleak: "Our results suggest that releasing the weights of future, more capable foundation models, no matter how robustly safeguarded, will trigger the proliferation of capabilities sufficient to acquire pandemic agents and other biological weapons." Writing in response to the study, Liang and his collaborators described the study as a "cautionary tale" of "what can go wrong when analyses of the risks of open foundation models do not compare against risks from closed models or existing technology (such as web search on the internet)." Subsequent studies suggested that the information acquired on bioweapons through the current level of LLMs was at that point the equivalent to what someone could find through a traditional web search.

Liang envisions a future in which AI is interwoven into our daily lives and our societal institutions. In this scenario, the complete walling-off of the technology seems impractical and hard to implement. He suggests a framework in which we are responsibly assessing the "marginal risk" of open-source AI and then devising collective solutions when the risk is deemed too high. "Unfortunately, there hasn't been much rigorous study of risk assessment," Liang says. "What is important here is the marginal risk of releasing an open model compared to, well, just using Google or Wikipedia."

Part of assessing the risk of an AI model is looking at the potential defenses the model has against misuses. Liang brought up the example of using AI to hack into sensitive databases, such as credit card information. At present, AI can certainly assist in this illegal activity, but AI is also very helpful at defending against these attacks, at looking for security flaws and patching them. In this way, an open model can reinforce its own security. But in other areas, such as disinformation and nonconsensual intimate imagery, the capabilities of AI seem to be overwhelming our defenses, Liang cautions. These are areas in which companies and governments must step in with well-structured policy and effective solutions.

Like so many aspects of artificial intelligence, the question of open-source is a balancing act. Are the benefits of open-source for scientific research such a positive that we accept the potential harms of disinformation? "I absolutely think that we need good policy and we need to understand the ramifications," Liang says, "but I think a lot of the AI safety worries are a little bit too divorced from reality, and they don't take the whole system into account." Lessig, for his part, discussed a possible public-private divide. "AI is an enormously valuable technology for the world," he says. "I think we should have a kind of Human Genome Project–like case to build it in a way that's protective of public AI ideals as well as private AI ideals."

We all want AI that's ethical, dependable, and secure and helps with important decisions about our health and our climate. We want AI that can accelerate scientific discovery. These goals will require a rational framework to assess the societal impacts of AI models. Right now, both the open and closed models of AI are moving forward at a rapid pace. It's important to think of AI as along the lines of a public commons—a technology that will affect all aspects of our experience, the future of which depends upon the cooperation of government, private business, and an engaged citizenry.



FEATURE

The challenge of the next decade is to discover how to introduce artificial intelligence into the workforce so that everyone benefits.

←
Illustrations by Irene Suosalo

TRANS-FORMING HUMAN LABOR

M

Maya Kulycky loves a bulleted list, but even more, she loves handing over the drudgery to a digital assistant. "I really like the drafting tool," she says of Gemini in Google Docs, the AI function embedded across several apps in Google Workspace. She starts with a rough draft of a memo—"Here's what we had a conversation around, here's what we want to do, here are the next steps we're going to take, here's the timeline"—and then asks Gemini to arrange those thoughts into a well-formatted email to her team at Google Research. "The technology is with me," she says. "It's not a replacement for me. It's an accelerant."

As AI arrives at our workplaces, it's showing up as a helpful tool, not as a superintelligence. The signs are clear that AI will affect every sector of the workforce, from entry-level employees to managers and CEOs. The deployment of AI is both a design and a user-interface problem and a wider, societal responsibility of technologists, AI companies, and government and legislators. The next few years are critical. We need to lay a proper foundation for AI at work.

In her role as vice president for strategy, operations, and outreach, Kulycky is tasked with aligning the teams that push significant AI innovations forward. Kulycky believes we need to proceed carefully to ensure that AI augments human potential. One of the near-term goals is exploring how AI can assist with the many background tasks and organizational work that often bog down teams, the so-called day-to-day that can get in the way. If these commitments could be offloaded onto AI agents, workers would have more time for the most important and creative challenges they face.

The future that AI creates will also depend upon how exactly we integrate it. Take the example of introducing customer-service agents that are entirely AI-based. A company can choose to provide better, more thoughtful care to customers (with the AI assisting humans), or the company could cut jobs. The economist Erik Brynjolfsson studies how the workforce will be affected by AI. When asked about the specter of workers being replaced by AI, Brynjolfsson brought up the Luddites, the famous reactionary group. Concerned that looms and automatic weaving machines would replace skilled artisans, knitters broke into factories to tear down the machinery that was automating away their jobs. But although their protest made history, it didn't halt technological progress. "A lot of specific jobs did disappear," Brynjolfsson says. By the late 1700s, the Industrial Revolution

> "AI can often do a core part of a task well. But humans are better at dealing with exceptions, improvising, dealing with things that come up."
>
> **Erik Bryjolfsson**, economist

was increasingly disrupting the textile industry. "But it led to new jobs for other people—and economic growth and higher living standards."

The pattern has repeated itself through a series of technological developments. In the United States, major technological shifts such as automobiles and computerization have not led to long-term widespread unemployment. "We've had lots of new technology, but there are still a lot of jobs. It's just that there are different jobs to do," Brynjolfsson says. For example, from 1900 to 2020, the proportion of Americans working in agriculture dropped from 42 percent to 2 percent. "Most of us are doing other things," Brynjolfsson says, "and the jobs that are normal to us now are jobs that people couldn't have even imagined before." A 2022 research paper by the economist David Autor found that 60 percent of working Americans currently occupy roles that didn't exist in the 1940s. We've shifted from mining to computer coding, from textile manufacturing to health care.

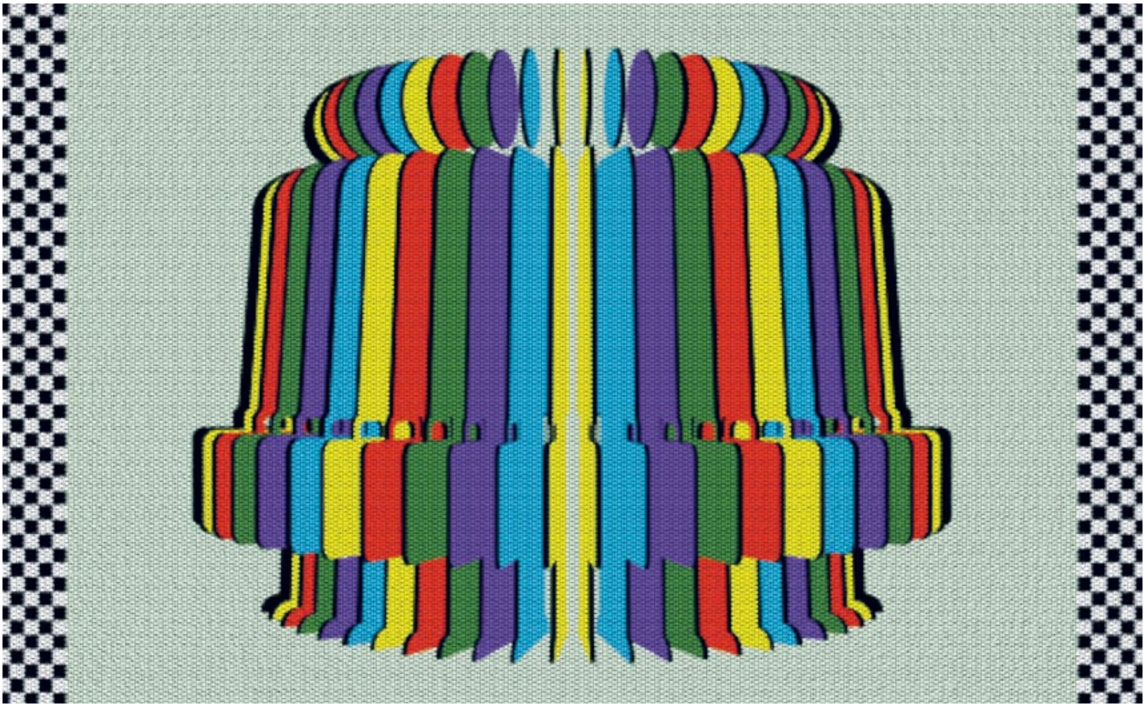
Each week, it seems, we receive news of bigger and more powerful large language models (LLMs), and that leads the public imagination to a place where AI tools are so sophisticated that they severely pressure huge segments of the workforce. Brynjolfsson offered a note of optimism. While the capabilities of machine learning are certainly remarkable, he cautioned against presuming that machines will fully automate modern employment processes. "That does work sometimes. But more often, you want to keep a human in the loop because the technology is not all-encompassing," he says. "When we break down jobs into tasks, we find that AI can often do a core set of tasks well. But humans are better at dealing with exceptions, improvising, dealing with things that come up."

According to a working paper from the International Monetary Fund (IMF), AI will likely enhance rather than replace jobs that have a lot of in-person interaction, critical decision-making, and specialized expertise. Occupations such as lawyers, surgeons, or judges will benefit both from AI's ability to take on routine administrative tasks and from its assistance in core functions such as analyzing medical scans or writing a first draft of a patent submission. Other studies have shown that AI can help people become faster and better at tasks such as business writing, programming, and consulting.

Brynjolfsson and his colleagues ran a now-famous study for which they observed the implementation of an LLM in a call center. The AI examined the calls completed by the most effective employees, then surfaced those practices to everyone else in the work group. The least-skilled workers benefited the most from AI; their performance jumped by 35 percent. AI's ability to close the gap between high and low performers could also streamline onboarding, acclimating workers quickly to new roles. That benefit could be a double-edged sword: It's great for helping employees succeed quickly on the job, but at the organizational level, it lowers the barrier to replacing workers, which in turn could erode job security.

Service professions will also benefit from assistive AI, according to the IMF study. Customer service chatbots can handle basic requests, freeing agents to contend with more complex cases. For example, Best Buy is resolving customer issues up to 90 seconds faster using Google Cloud's automated call summarization. A research paper by Barclays calls this shift "service-ization"—a nod to industrialization, a process by which the hands-on, intensive work in manufacturing and agriculture became the work of machines.

Though the research by scholars and analysts points to the productivity gains that can result from the assistance of AI, businesses' uptake has been slow. Brynjolfsson expressed surprise about how the adoption of artificial intelligence is already lagging behind the technology's capabilities: "I'm disappointed by how few people have thought about how to use generative AI (GenAI) in their daily work. While the technology is racing ahead very rapidly, there's much less



energy put into how we get business value out of it." Part of the reason for that lag could be the investment it takes to implement workflows around the adoption of any new tool. Still, Brynjolfsson encourages leaders to consider how these new technologies can improve their businesses. "In 2017, I wrote that AI won't replace managers, but managers who use it will replace managers who don't. That's still true today."

Even at the current rate of adoption, experts predict that AI will eclipse certain sectors of labor while creating opportunities for new kinds of jobs. Goldman Sachs estimates that GenAI may automate close to a quarter of all jobs. A 2023 McKinsey report predicts that 30 percent of today's work hours could be replaced by GenAI as soon as 2030. That same report found that low-wage workers (those earning less than $38,200 a year) are up to 14 times more vulnerable to job elimination than America's highest earners. Brynjolfsson agrees: In a 2017 paper, he found that machine learning will most profoundly affect low-wage workers.

That poses both economic and social concern since low-wage jobs are disproportionately held by women, people of color, and workers who haven't received higher education. Anyone affected by such an industry shift will have the added burden of needing to learn new skills to transition to a new role successfully. AI can assist with that training, given how quickly it can do the work of, say, personalizing learning programs and translating content.

Perhaps the simplest concern about the economic impact of AI at work is who will benefit financially from these large shifts. Disruptive technologies—even those that eliminate jobs—historically raise national income because they boost overall productivity. But the workers affected by that shift don't necessarily see that economic benefit. While those who transition successfully may continue to collect the labor wages they rely on, the benefit of their increased productivity tends to be concentrated among corporate leadership. "New technology has the possibility to close the skills gap between workers, but it can increase the gap between capital and labor," says Brynjolffson. "The best strategy is to aim for technology that creates widely shared prosperity, not just concentrated wealth."

Experts predict that AI will eclipse certain sectors of labor while creating opportunities for new kinds of jobs.

Companies will need to work together, along with governments and nonprofits, to help more people realize the economic potential of AI. One example of such partnership is the AI-Enabled ICT Workforce Consortium, founded by Cisco in partnership with eight other companies including Accenture, Eightfold, Google, IBM, Indeed, Intel, Microsoft and SAP, which will recommend skilling and upskilling opportunities to ensure workers can adapt. Six members of this consortium have committed to goals for training, upskilling, and reskilling millions of people over the next 10 years; Google's $120 million Global AI Opportunity Fund helps workers and students from all backgrounds access AI training —including the company's AI Essentials course—in local communities at no cost, building on the more than 100 million people Google has trained in digital skills to date.

In places like parts of Europe and Asia where the working-age population is falling, AI can complement the workforce, taking on tasks that allow workers to focus on the most important, rewarding aspects of jobs. The other good news, per Brynjolfsson: "In that world, productivity would be dramatically higher—we'd have 2, 5, 10x more productivity per capita. Then, we'd be able to provide basic needs for people. As Keynes put it, the economic problem would be solved. People might or might not still enjoy working, but they wouldn't need to worry about starving. A world like that is not imminent, but I can imagine the benefits of thinking about how we would run that economy."

In the meantime, who is responsible for rallying industry around an economy-preserving, assistive approach to AI? Brynjolfsson recommends that entrepreneurs and executives focus on the top line as well as the bottom, broadening their performance goals to look not just at costs but also at metrics such as quality and customer satisfaction. He also imagines that organizations will get more buy-in from their customers and their workforce if they use AI to complement workers. "Many people would want to speed adoption of a tech that will help them do their job better, instead of one pitched as a replacement for them," he says.

Countries can also choose to apply AI in their own ways. Some may use tools to combat misinformation, for instance, while others might lean on AI to create the world's most effective censorship system.

Technologists, of course, also have the power to shape the industry. The benchmarks used to measure the effectiveness of machine learning models have been inspired by the Turing test, a test that measures the strength of a machine by how closely it can imitate human conversation. Establishing new kinds of benchmarks could help position AI as assistive—a paradigm shift that could lead to more impressive innovations. "Merely mimicking humans leads us into what I call the Turing trap. Instead, look at technologies that augment us to do things that no human could do before. They'll create more value and shared prosperity," says Brynjolfsson.

Introducing AI responsibly will take some creative thinking on the part of corporate leaders and government officials. Rather than clinging to existing job distribution, Brynjolfsson advocates offering tools for training and job matching. According to a 2024 survey from Deloitte, only 47 percent of leaders agree that their organizations are sufficiently educating employees on the capabilities, benefits, and value of GenAI. In addition to upskilling employees, businesses will need to change what qualifications they look for when hiring (it seems adaptability will become a crucial skill), opening their doors to candidates who may have less prior training or education but can quickly adjust and learn new skills. Going forward, success in the workplace will be less about possessing specific skills and more about being flexible, Brynjolfsson predicts. "The era of going to school for 12, 16, or more years, learning stuff, and using that for the next 40 years is gone," he says. "People have to continually learn and adjust."

The history of science is the history of eureka moments—from Newton's apple to Oersted's compass needle to Fleming's petri dish. For many people in and around the field of AI research, the past decade or so has been a succession of such moments, many of them described in this issue of *Dialogues*—from revelatory research on protein folding to promising approaches to mitigating global warming (for a sense of the technology's influence on the way we learn, work, govern, and more, see the accompanying infographics). We asked 20 scientists and thought leaders around the globe to recall when they realized AI had the potential to change the world.

The *"Eureka!"*

Fr. Paolo Benanti

Theologian; adviser to Pope Francis on AI and computer ethics; member, High-Level Advisory Body on Artificial Intelligence, United Nations

Italy Italy

One day, I returned to my convent and overheard a 90-year-old friar in the next room conversing with the smart speaker. The natural way in which he spoke to it, and the fact that he had found someone who was not bored with his repetitive questions about the weather, history, and current events, turned a light on in me: If AI could so easily enter the life of a man from the last century—who does not even really know what this technology is, but who can interact usefully with computational power in a cloud thousands of miles away—well, then, yes, AI was ready to change our daily lives.

Tilman Becker

Director, RICAIP Centre for Advanced Industrial Production at the Czech Institute of Informatics, Robotics, and Cybernetics (CIIRC)

Germany · Czech Republic

For me, there's been more than one eureka moment: The first was 40 years ago when I was starting to learn about AI as a student, then some 30 years ago, when I spoke in German into an early prototype of the Verbmobil system I worked on at DFKI [a German research center] and got a spoken English translation seconds later (spooky at the time, even though I knew exactly what the system was doing). Then again over 10 years ago, when I realized that AI technology had become so pervasive that it is an important tool in the fourth Industrial Revolution. And finally, one day in late 2022, when using ChatGPT for the first time: Even with all its limitations, seeing a system generate flawless, eloquent language left me flabbergasted.

Moment

Lila Ibrahim

Google DeepMind, Chief Operating Officer

USA · United Kingdom

In 2020, Google DeepMind achieved a seemingly impossible scientific breakthrough: our advanced AI system, AlphaFold, could now accurately predict protein structures in minutes, not years.

I still remember that moment—when I realized that all of our investment, iteration, and patience had delivered so much benefit for the world.

Proteins are life's building blocks. Understanding their structure is key to advancing drug discovery, food security, treatments for neglected diseases, and more. We've used AlphaFold to map over 200 million proteins and made this knowledge freely available. Today, 2.2 million researchers across 190 countries use AlphaFold to tackle some of the biggest problems facing our world.

AlphaFold is proof of AI's ability to advance scientific discovery, and we've only scratched the surface of its extraordinary potential to benefit humanity.

Seydina Moussa Ndiaye

Senior Lecturer at Cheikh Hamidou Kane Digital University; member, High-Level Advisory Body on Artificial Intelligence, United Nations

Senegal · Senegal

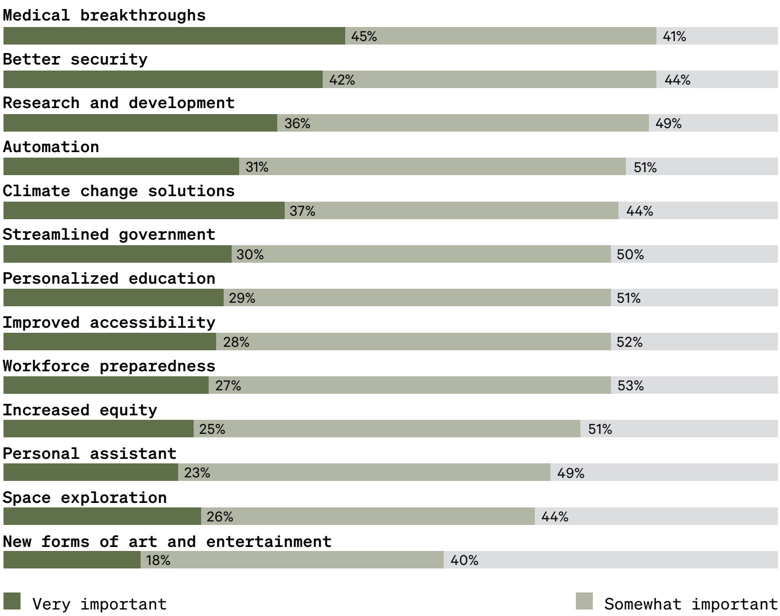
When I first discovered the methods of artificial intelligence during my first year of my master's degree, in 1994, I immediately realized that this was in fact the ultimate goal of computer science. My eureka moment came to me when, in 1997, during my doctoral thesis, I obtained results that were more accurate than those obtained by agronomic experts for the cultivation of winter wheat, by applying reinforcement learning methods combined with a genetic algorithm.

Country of origin · Currently based

Saška Mojsilović
Google, Senior Director, Researching Engineering AL/ML
Serbia USA

35 years ago, as a college junior, I took an introductory computer vision class. We scanned a black and white photograph, and then coded an algorithm that analyzed the image and outlined objects in it. This made me think about how everything around us—speech, music, written word, molecules, X-rays, weather, traffic—can be understood as patterns. I realized that if we could teach computers to see, we could teach them to sense, hear, read and reason; to identify cancer and diagnose disease, weave new molecules, predict crop yields and natural disasters... Back then, computers were slow, data was scarce, and algorithms were simple, but I was in awe of what might one day be possible.

How important, if at all, is each AI application for society:



Results from a global report from Google and Ipsos, published in January 2024.

Virginia Dignum

Professor of Computer Science, Umeå, Sweden; member, High-Level Advisory Body on Artificial Intelligence, United Nations

Portugal Sweden

I realized AI was *not* going to change life as we know it when existential risks became a central issue in discussions about its future. AI will not fundamentally change life; instead, it will push us to recognize the importance of our human existence. AI capabilities highlight the irreplaceable value of human creativity, empathy, and moral reasoning. The responsibility for AI's impact rests with us, not with the technology. AI doesn't happen to us—we make it happen.

Ayanna Howard

Dean of Engineering, The Ohio State University; member, National AI Advisory Committee

USA USA

In the early 2000s, I worked on developing a neural network approach to enable future Mars rovers to navigate long distances over hazardous terrain. After countless field trials on Earth, during which my robot would sometimes exhibit quite strange, if not erratic, unsafe driving behavior, I finally thought, Why should a robot learning to drive be so different from teaching a teenage driver? That was my eureka moment. I realized that AI's value proposition should be centered around its synergistic and symbiotic relationship with humans and that if it continued to be designed around this fundamental concept of people first and foremost, it would change the world.

Richard Zhang

Distinguished Professor of Computer Science, Simon Fraser University

China Canada

The world's leading machine learning conference, NeurIPS, recently accepted more than 4,000 papers. This is in addition to well over 5,000 top-tier papers in computer vision and graphics, all in my fields of research. Even if only 1 percent of these papers are closely related to my work, I'll have to sift through a sea of publications to find and read them. There cannot be that much truly innovative research out there, being produced at this volume and pace—much of it must be "artificial," or at least superficial. While I still cannot fathom ever using AI to write one of these papers, it suddenly occurred to me that I may have to rely on AI to find, read, and summarize the latest research results to stay well informed.

Sayan Chakraborty

Co-President, Workday; former member, National AI Advisory Committe

India · USA

I first built artificial neural networks (ANNs) in 1989, during my graduate work at MIT: I wanted to apply ANNs to robots designed to repair satellites. However, the neural network I trained to control a robotic arm failed miserably, and I set aside ANNs as an interesting curiosity. But by 2012, deep-learning ANNs, more sophisticated than those I had worked with, were winning competitions in image recognition and hinting at greater potential. Graphics processors (GPUs) designed for gaming were being used to radically improve performance. But it wasn't really until late 2020, after a discussion with a friend about the Google paper titled "Attention Is All You Need," that everything clicked. The missing puzzle pieces—powerful GPUs, the attention-transformer approach, cloud computing, the availability of vast training datasets on the internet—were finally all here, and I knew that this revolution was happening now.

Alondra Nelson

Harold F. Linder Professor at the Institute for Advanced Study; former acting director of the White House Office of Science and Technology Policy in the Biden-Harris administration; member, High-Level Advisory Body on Artificial Intelligence, United Nations

USA · USA

In developing what would become the White House's 2022 Blueprint for an AI Bill of Rights, we spent nearly a year meeting with any member of the American public with a perspective to share on AI's impact on society. This included a multiracial group of high-school students from across the country who called themselves Encode Justice and advocated for human-centered, responsible uses of AI. These students spoke with a remarkable urgency about their concern that facial recognition technology would soon be required for them to access their school buildings, their classrooms, and parts of their curriculum. As policy makers, we had been thinking about the impact of social media and algorithmic amplification on young people's mental health. But for these students, AI enabling a state of persistent surveillance was top of mind. The generation that is the most likely to see the greatest harms or gains from AI is flashing a yellow warning light for us.

Justin Tranter

Songwriter

USA · USA

I had two big wow moments while getting to play with Music AI Sandbox, from Google DeepMind and YouTube. The first was a fear—the fear that if this was not handled properly, it would be yet another thing in our industry that harms songwriters. The second was hope: The tools are so magical that not only can it help human songwriters be more creative with new works, when done correctly it should also create a further revenue stream for songwriters' past work. There is a world where this is a win-win for songwriters.

Mira Lane

Google, Senior Director of Technology & Society

Canada USA

I've always believed that technology has the potential to change the world. Yet, as both an artist and technologist, I once thought the creativity industry might be immune to the massive types of disruption technology often brings. But when I saw AI being used in the role of a collaborator—co-creating scripts, generating new imagery, and creating music—I realized storytelling was also being transformed. The real magic happens when AI and human creativity combine across disciplines, allowing us to explore narratives in new and unexpected ways. What might happen when we can all express our inner creativity more easily? The world might suddenly open to new and amazing possibilities. Interdisciplinary collaboration between technologists, artists, and storytellers is now pushing the boundaries of what's possible, reshaping how we engage with the world through art and imagination.

Rahaf Harfoush

Digital anthropologist and author; member, High-Level Advisory Body on Artificial Intelligence, United Nations

Syria/Canada France

I was experimenting with AI's ability to change the way we interact with ideas and was testing different personas, each one pulling expertise from various domains to answer my questions. During that process, I ended up accidentally creating what I started calling a "super partner"—a blend of the wisdom and compassion of a Zen monk, the sharp mind of a strategist, the frameworks of a therapist, and the practical ability of a coach, who could help me put it all into practice. I was shocked by how well it worked.

It wasn't just giving me answers; it was synthesizing advice in a way that felt personal and multidimensional. This was a completely new way for us, as humans, to not just access knowledge, but to interact with it, remix it, and blend it in ways that sparked new connections. That's what really lit up my brain—the realization that AI could help us create, think, and problem-solve by using the breadth of human intellect in a way that simply wasn't possible before.

Phil Venables

Google Cloud, Vice President, TI Security & Chief Information Security Officer (CISO)

United Kingdom — USA

I realized AI had the potential to change the world when I saw how it could revolutionize cybersecurity. While attackers may use AI to their advantage, we are seeing firsthand how it is transforming defenders' capabilities. Traditionally, cybersecurity has faced the so-called defender's dilemma, where attackers only need to succeed once, while defenders must be vigilant every time. But now, with AI, we're seeing an increasing shift toward the defender's advantage. AI provides defenders with an invaluable edge, allowing them to analyze threats, detect vulnerabilities, and respond with unprecedented speed and scale. Like other groundbreaking technologies throughout history, AI brings both remarkable benefits and unforeseen risks—but it ultimately empowers us to stay one step ahead, building a safer, more resilient digital world.

Holger Hoos

Alexander von Humboldt Professor of AI, RWTH Aachen University (Germany); Professor of Machine Learning at Universiteit Leiden (the Netherlands); Adjunct Professor of Computer Science at the University of British Columbia (Canada)

Germany — Germany

Around 2013, I was working with a team within a large electrical utility in Canada that was struggling, because of the immense computational cost involved, to set prices for their surplus of 100 percent renewable energy. My research team and I had started to develop AI methods for reducing the computational effort involved in solving this kind of formal-reasoning problem, but never at scale. With their experts, we ultimately were able to combine methods from automated reasoning, machine learning, and optimization that solved the energy-pricing problem at scale, literally overnight, without the use of a large compute cluster. I realized what a difference AI could make, not merely economically, but also in terms of tackling the grand challenges of our time—in this case, a problem closely related to climate change.

Craig Ramlal

Head of the Control Systems Group, University of the West Indies at St. Augustine; member, High-Level Advisory Body on Artificial Intelligence, United Nations

Trinidad & Tobago — West Indies

I remember being taken aback when I learned that language models could construct rational sentences. It meant that, for the first time in human history, there was another entity, one that was not human, capable of creating (approximating) logical thought across many domains. It also occurred to me that we may greatly rely on these systems at some point in the future, giving them more autonomy and trusting their decisions—fundamentally altering how humanity interacts with and thinks about technology. I remain hopeful that the minds behind AI's success, the original thinkers, will guide its research, utility, and adoption toward the greater good.

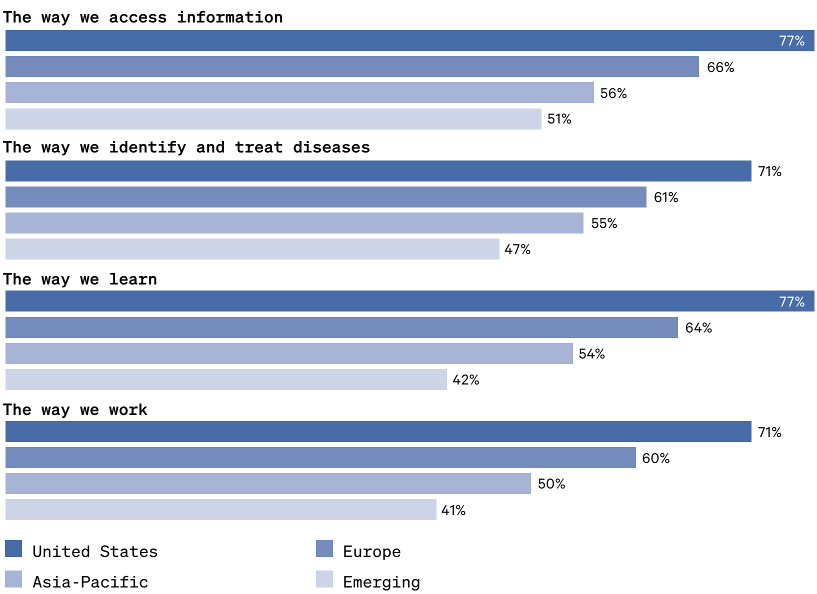
Paula Goldman

Chief Ethical and Humane Use Officer, Salesforce; member, National AI Advisory Committee

USA/Singapore   USA

It was the holiday season in 2022, and while I expected most of my conversations with friends to revolve around the festivities, I was surprised when they all wanted to talk about AI—how it might impact their law practice or their journalism career. While tracking AI milestones, like detecting early signs of cancer and testing self-driving cars, is second nature for me, I knew something had tipped when friends who had never expressed interest started asking questions. We are at the precipice of a profound shift in how we interact with technology at work, seeing it evolve from a tool to a dynamic partner. While we all know this shift comes with both promise and peril, it can't be overstated how crucial it is to keep people at the helm of this transformation.

How much positive impact, if any, do you think AI is currently having on:



Results from a global report from Google and Ipsos, published in January 2024.

Jimena Sofía Viveros Álvarez

Member, High-Level Advisory Body on Artificial Intelligence, United Nations; Commissioner at the Global Commission on Responsible Artificial Intelligence in the Military Domain; AI Expert at the Organization for Economic Co-operation and Development

Mexico   Mexico

Since the development of the Blue Brain Project in 2005, I've been both fascinated by and concerned with technological developments and their legal and ethical implications. I realized then that—like the discovery of gunpowder and the splitting of the atom—artificial intelligence was going to be the defining technology of our generation and those to come. Based on my experience in peace and security, I foresaw AI would pose grave, perhaps even existential, risks to humankind in these domains if not properly and globally governed, which is why I am truly committed to this endeavor. Ultimately, our common goal must be to ensure that we can harness AI's opportunities fairly and equitably while mitigating its risks for the benefit *and protection* of all of humanity.

Lerrel Pinto

Assistant Professor of Computer Science, Courant Institute of Mathematical Sciences, New York University

India   USA

Meaningful change in life requires physically unburdening people from mundane, laborious, or simply undesirable tasks, allowing us to focus on what truly matters. While robotics has traditionally lagged behind other AI fields like natural language processing, the past year has brought a breakthrough in what we call zero-shot capabilities for robots: A robot can now be placed in an unfamiliar environment, such as your home, and successfully perform complex tasks—like moving small objects or opening cabinet doors—with encouraging accuracy, despite never having been specifically programmed for that setting. This quiet robot revolution will revolutionize our daily lives, bridging the gap between virtual intelligence and a more physical, real one.

Connor Coley

Associate Professor of Chemical Engineering, Electrical Engineering, and Computer Science, MIT

USA   USA

Around 2017, I was compiling the results of a human-benchmarking comparison to evaluate a new neural network model for organic chemistry: This model correctly predicted the experimentally confirmed product of a chemical reaction where 10 Ph.D.-level chemists could not! Given how quickly we were able to reach that level of performance, although it was not as significant of a milestone as AlphaFold would be several years later, I could then see the path to evermore capable AI helping us navigate the scientific process.

TYPEFACES

Dialogues is typeset in *Affairs* (SM Foundry), as well as *Calibre, Söhne Schmal, 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.

Google + Atlantic Re:think