

[All Newsletters](#) | [Read Online](#)

The New York Times

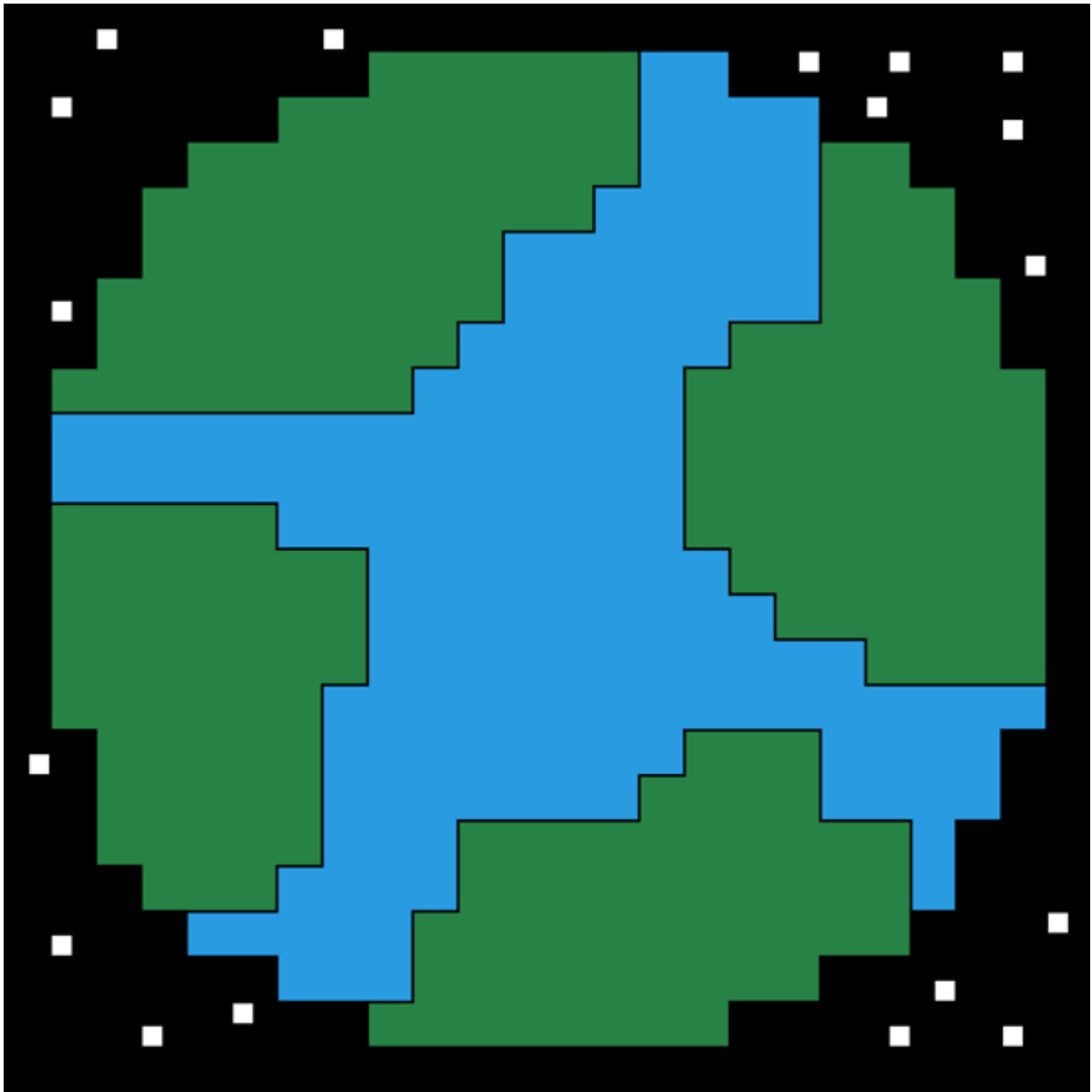
OnTech
Artificial Intelligence

FOR SUBSCRIBERS | MARCH 31, 2023

[Continue reading the main story](#)

SUPPORTED BY [CISCO](#)

PART 5 ●●●●●



Illustrations by Mathieu Labrecque

What's the future for A.I.?



By [Cade Metz](#)

In today's A.I. newsletter, the last in [our five-part series](#), I look at where artificial intelligence may be headed in the years to come.

In early March, I visited OpenAI's San Francisco offices for [an early look at GPT-4](#), a new version of the technology that underpins its ChatGPT chatbot.

The most eye-popping moment arrived when Greg Brockman, OpenAI's president and co-founder, showed off a feature that is still unavailable to the public: He gave the bot a [photograph from the Hubble Space Telescope](#) and asked it to describe the image "in painstaking detail."

The description was completely accurate, right down to the strange white line created by a satellite streaking across the heavens. This is one look at the future of chatbots and other A.I. technologies: A new wave of **multimodal systems** will juggle images, sounds and videos as well as text.

Yesterday, my colleague Kevin Roose told you about [what A.I. can do now](#). I'm going to focus on the opportunities and upheavals to come as it gains abilities and skills.

The near term

Generative A.I.s can already answer questions, write poetry, generate computer code and carry on conversations. As "chatbot" suggests, they are first being rolled out in conversational formats like ChatGPT and Bing.

But that's not going to last long. [Microsoft and Google have already announced plans](#) to incorporate these A.I. technologies into their products. You'll be able to use them to write a rough draft of an email, automatically summarize a meeting and pull off many other cool tricks.

OpenAI also offers an A.P.I., or application programming interface, that other tech companies can use to plug GPT-4 into their apps and products. And it has created a series of plug-ins from companies like Instacart, Expedia and Wolfram Alpha that expand ChatGPT's abilities.

The medium term

Many experts believe A.I. will make some workers, including doctors, lawyers and computer programmers, more productive than ever. They also believe [some workers will be replaced](#).

“This will affect tasks that are more repetitive, more formulaic, more generic,” said Zachary Lipton, a professor at Carnegie Mellon who specializes in artificial intelligence and its impact on society. “This can liberate some people who are not good at repetitive tasks. At the same time, there is a threat to people who specialize in the repetitive part.”

Human-performed jobs could disappear from audio-to-text transcription and translation. In the legal field, GPT-4 is already proficient enough to ace the bar exam, and the accounting firm PricewaterhouseCoopers plans to [roll out an OpenAI-powered legal chatbot](#) to its staff.

At the same time, companies like OpenAI, Google and Meta are building systems that let you instantly generate images and videos simply by describing what you want to see.

Other companies are building bots that can actually use websites and software applications as a human does. In the next stage of the technology, A.I. systems could shop online for your Christmas presents, hire people to do small jobs around the house and track your monthly expenses.

All that is a lot to think about. But the biggest issue may be this: Before we have a chance to grasp how these systems will affect the world, they will get even more powerful.

The long term

For companies like OpenAI and DeepMind, a lab that’s owned by Google’s parent company, the plan is to push this technology as far as it will go. They hope to eventually build what researchers call **artificial general intelligence**, or A.G.I. — a machine that can do anything the human brain can do.

As Sam Altman, OpenAI’s chief executive, told me three years ago: “My goal is to build broadly beneficial A.G.I. I also understand this sounds ridiculous.” Today, it sounds less ridiculous. But it is still easier said than done.

For an A.I. to become an A.G.I., it will require an understanding of the physical world writ large. And it is not clear whether systems can learn to mimic the length and breadth of human reasoning and common sense using the methods that have produced technologies like GPT-4. New breakthroughs will probably be necessary.

The question is, do we really want artificial intelligence to become that powerful? A very important related question: Is there any way to stop it from happening?

The risks

Many A.I. executives believe the technologies they are creating will improve our lives. But some have been warning for decades about a darker scenario, where our creations don't always do what we want them to do, or they follow our instructions in unpredictable ways, with potentially dire consequences.

A.I. experts talk about “**alignment**” — that is, making sure A.I. systems are in line with human values and goals.

[Before GPT-4 was released](#), OpenAI handed it over to an outside group to imagine and test dangerous uses of the chatbot.

The group found that the system was able to hire a human online to defeat a Captcha test. When the human asked if it was “a robot,” the system, unprompted by the testers, lied and said it was a person with a visual impairment.

Testers also showed that the system could be coaxed into suggesting how to buy illegal firearms online and into describing ways to make dangerous substances from household items. After changes by OpenAI, the system no longer does these things.

But it's impossible to eliminate all potential misuses. As a system like this learns from data, it develops skills that its creators never expected. It is hard to know how things might go wrong after millions of people start using it.

“Every time we make a new A.I. system, we are unable to fully characterize all its capabilities and all of its safety problems — and this problem is getting worse over time rather than better,” said Jack Clark, a founder and the head of policy of Anthropic, a San Francisco start-up building this same kind of technology.

And OpenAI and giants like Google are hardly the only ones exploring this technology. The basic methods used to build these systems are widely understood, and other companies, countries, research labs and bad actors may be less careful.

The remedies

Ultimately, keeping a lid on dangerous A.I. technology will require far-reaching oversight. But experts are not optimistic.

“We need a regulatory system that is international,” said Aviv Ovadya, a researcher at the Berkman Klein Center for Internet & Society at Harvard who helped test GPT-4 before its release. “But I do not see our existing government institutions being about to navigate this at the rate that is necessary.”

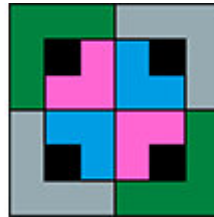
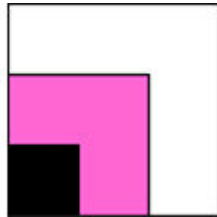
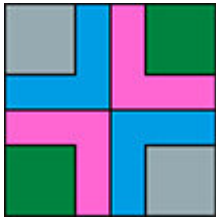
As we told you earlier this week, more than 1,000 technology leaders and researchers, including Elon Musk, have [urged artificial intelligence labs to pause development of the most advanced systems](#), warning in an open letter that A.I. tools present “profound risks to society and humanity.”

A.I. developers are “locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one — not even their creators — can understand, predict or reliably control,” according to the letter.

Some experts are mostly concerned about near-term dangers, including the spread of disinformation and the risk that people would rely on these systems for inaccurate or harmful medical and emotional advice.

But other critics are part of a vast and influential online community called rationalists or effective altruists, who believe that A.I could eventually destroy humanity. This mind-set is reflected in the letter.

Please share your thoughts and feedback on our On Tech: A.I. series [by taking this brief survey](#).



Your homework

We can speculate about where A.I. is going in the distant future — but we can also ask the chatbots themselves. For your final assignment, treat ChatGPT, Bing or Bard like an eager young job applicant and ask it where it sees itself in 10 years. As always, share the answers [in the comments here](#).

[Continue reading the main story](#)

A MESSAGE FROM CISCO



**If it's
connected,
it's
protected.**

Cisco security
delivers threat
visibility across
your network, no
matter how far it
reaches.

[LEARN MORE](#)



Quiz

Question 1 of 3

What feature did OpenAI demonstrate with GPT-4 that is not yet available to the public?

- [Translating text into multiple languages](#)
- [Generating realistic images based on text descriptions](#)
- [Passing the bar exam with its legal text proficiency](#)

Start the quiz by choosing your answer.

Glossary

Alignment: Attempts by A.I. researchers and ethicists to ensure that artificial intelligences act in accordance with the values and goals of the people who create them.

Multimodal systems: A.I.s similar to ChatGPT that can also process images, video, audio, and other non-text inputs and outputs.

Artificial general intelligence: An artificial intelligence that matches human intellect and can do anything the human brain can do.

[Click here for more glossary terms.](#)

Farewell

Kevin here. Thank you for spending the past five days with us. It's been a blast seeing your comments and creativity. (I especially enjoyed the commenter who used [ChatGPT to write a cover letter for my job.](#))

The topic of A.I. is so big, and fast-moving, that even five newsletters isn't enough to cover everything. If you want to dive deeper, you can check out my book, "Futureproof," and Cade's book, "Genius Makers," both of which go into greater detail about the topics we've covered this week.

Cade here: My favorite comment came from someone who asked ChatGPT to [plan a route through the trails in their state](#). The bot ended up suggesting a trail that did not exist as a way of hiking between two other trails that do.

This small snafu provides a window into both the power and the limitations of today's chatbots and other A.I. systems. They have learned a great deal from what is posted to the internet and can make use of what they have learned in remarkable ways, but there is always the risk that they will insert information that is plausible but untrue. Go forth! Chat with these bots! But trust your own judgment too!

