The New York Times https://www.nytimes.com/2022/08/24/technology/ai-technology-progress.html

THE SHIFT

We Need to Talk About How Good A.I. Is Getting

We're in a golden age of progress in artificial intelligence. It's time to start taking its potential and risks seriously.

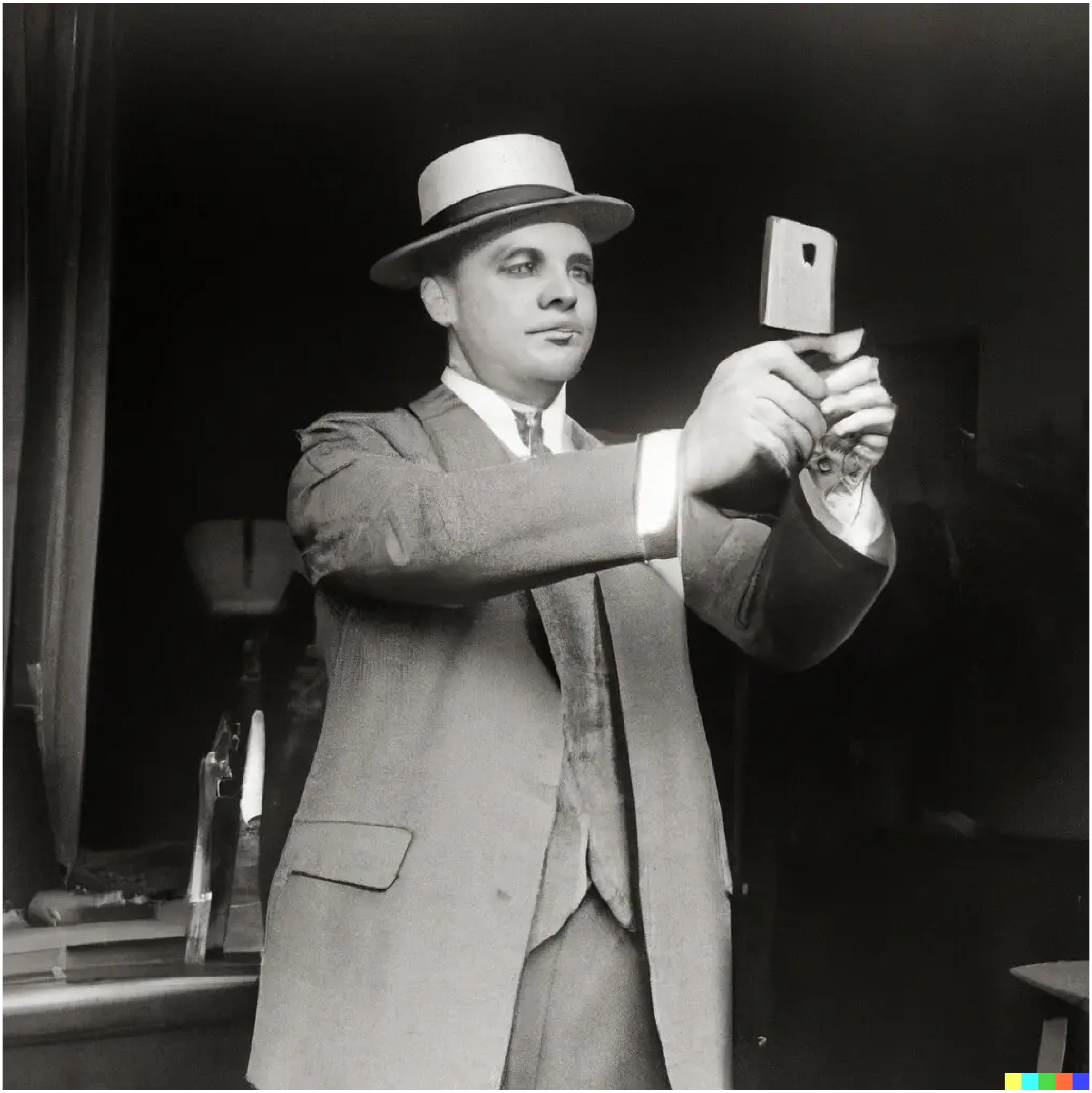
By Kevin Roose

Aug. 24, 2022

For the past few days, I've been playing around with DALL-E 2, an app developed by the San Francisco company OpenAI that turns text descriptions into hyper-realistic images.

OpenAI invited me to test DALL-E 2 (the name is a play on Pixar's WALL-E and the artist Salvador Dalí) during its beta period, and I quickly got obsessed. I spent hours thinking up weird, funny and abstract prompts to feed the A.I. — "a 3-D rendering of a suburban home shaped like a croissant," "an 1850s daguerreotype portrait of Kermit the Frog," "a charcoal sketch of two penguins drinking wine in a Parisian bistro." Within seconds, DALL-E 2 would spit out a handful of images depicting my request — often with jaw-dropping realism.

Here, for example, is one of the images DALL-E 2 produced when I typed in "black-and-white vintage photograph of a 1920s mobster taking a selfie." And how it rendered my request for a high-quality photograph of "a sailboat knitted out of blue yarn."



“Black-and-white vintage photograph of a 1920s mobster taking a selfie.” Generated by OpenAI's DALL-E 2

“A sailboat knitted out of blue yarn.” Generated by OpenAI's DALL-E 2

DALL-E 2 can also go more abstract. The illustration at the top of this article, for example, is what it generated when I asked for a rendering of “infinite joy.” (I liked this one so much I’m going to have it printed and framed for my wall.)

What’s impressive about DALL-E 2 isn’t just the art it generates. It’s *how* it generates art. These aren’t composites made out of existing internet images — they’re wholly new creations made through a complex A.I. process known as “diffusion,” which starts with a random series of pixels and refines it repeatedly until it matches a given text description. And it’s improving quickly — DALL-E 2’s images are four times as detailed as the images generated by the original DALL-E, which was introduced only last year.

DALL-E 2 got a lot of attention when it was announced this year, and rightfully so. It’s an impressive piece of technology with big implications for anyone who makes a living working with images — illustrators, graphic designers, photographers and so on. It also raises important questions about what all of this A.I.-generated art will be used for, and whether we need to worry about a surge in synthetic propaganda, hyper-realistic deepfakes or even nonconsensual pornography.

But art is not the only area where artificial intelligence has been making major strides.

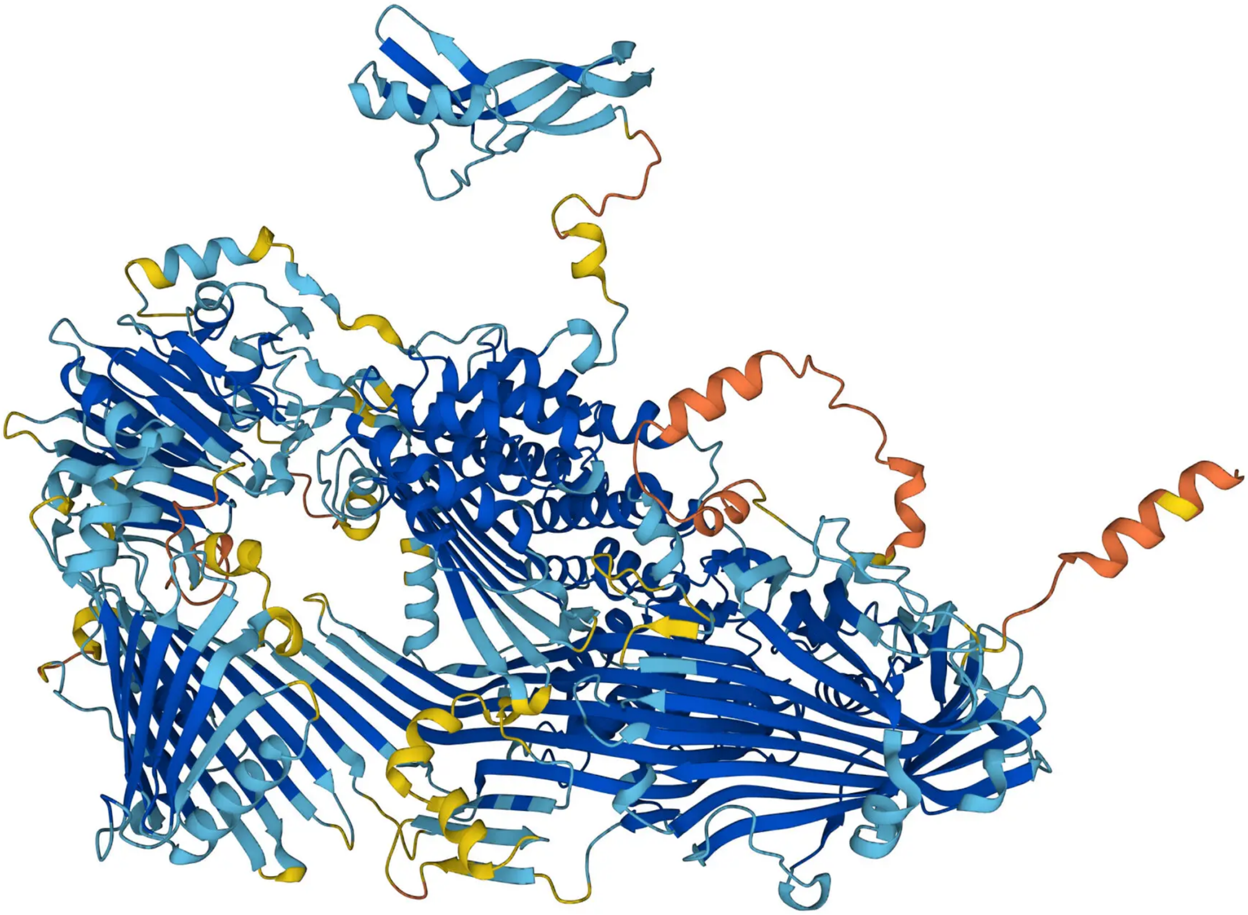
Over the past 10 years — a period some A.I. researchers have begun referring to as a "golden decade" — there's been a wave of progress in many areas of A.I. research, fueled by the rise of techniques like deep learning and the advent of specialized hardware for running huge, computationally intensive A.I. models.

Some of that progress has been slow and steady — bigger models with more data and processing power behind them yielding slightly better results.

But other times, it feels more like the flick of a switch — impossible acts of magic suddenly becoming possible.

Just five years ago, for example, the biggest story in the A.I. world was AlphaGo, a deep learning model built by Google's DeepMind that could beat the best humans in the world at the board game Go. Training an A.I. to win Go tournaments was a fun party trick, but it wasn't exactly the kind of progress most people care about.

But last year, DeepMind's AlphaFold — an A.I. system descended from the Go-playing one — did something truly profound. Using a deep neural network trained to predict the three-dimensional structures of proteins from their one-dimensional amino acid sequences, it essentially solved what's known as the "protein-folding problem," which had vexed molecular biologists for decades.



DeepMind announced that its A.I. system AlphaFold had made predictions for nearly all of the 200 million proteins known to exist. DeepMind

This summer, DeepMind announced that AlphaFold had made predictions for nearly all of the 200 million proteins known to exist — producing a treasure trove of data that will help medical researchers develop new drugs and vaccines for years to come. Last year, the journal Science recognized AlphaFold's importance, naming it the biggest scientific breakthrough of the year.

Or look at what's happening with A.I.-generated text.

Only a few years ago, A.I. chatbots struggled even with rudimentary conversations — to say nothing of more difficult language-based tasks.

But now, large language models like OpenAI's GPT-3 are being used to write screenplays, compose marketing emails and develop video games. (I even used GPT-3 to write a book review for this paper last year — and, had I not clued in my editors beforehand, I doubt they would have suspected anything.)

A.I. is writing code, too — more than a million people have signed up to use GitHub's Copilot, a tool released last year that helps programmers work faster by automatically finishing their code snippets.

Then there's Google's LaMDA, an A.I. model that made headlines a couple of months ago when Blake Lemoine, a senior Google engineer, was fired after claiming that it had become sentient.

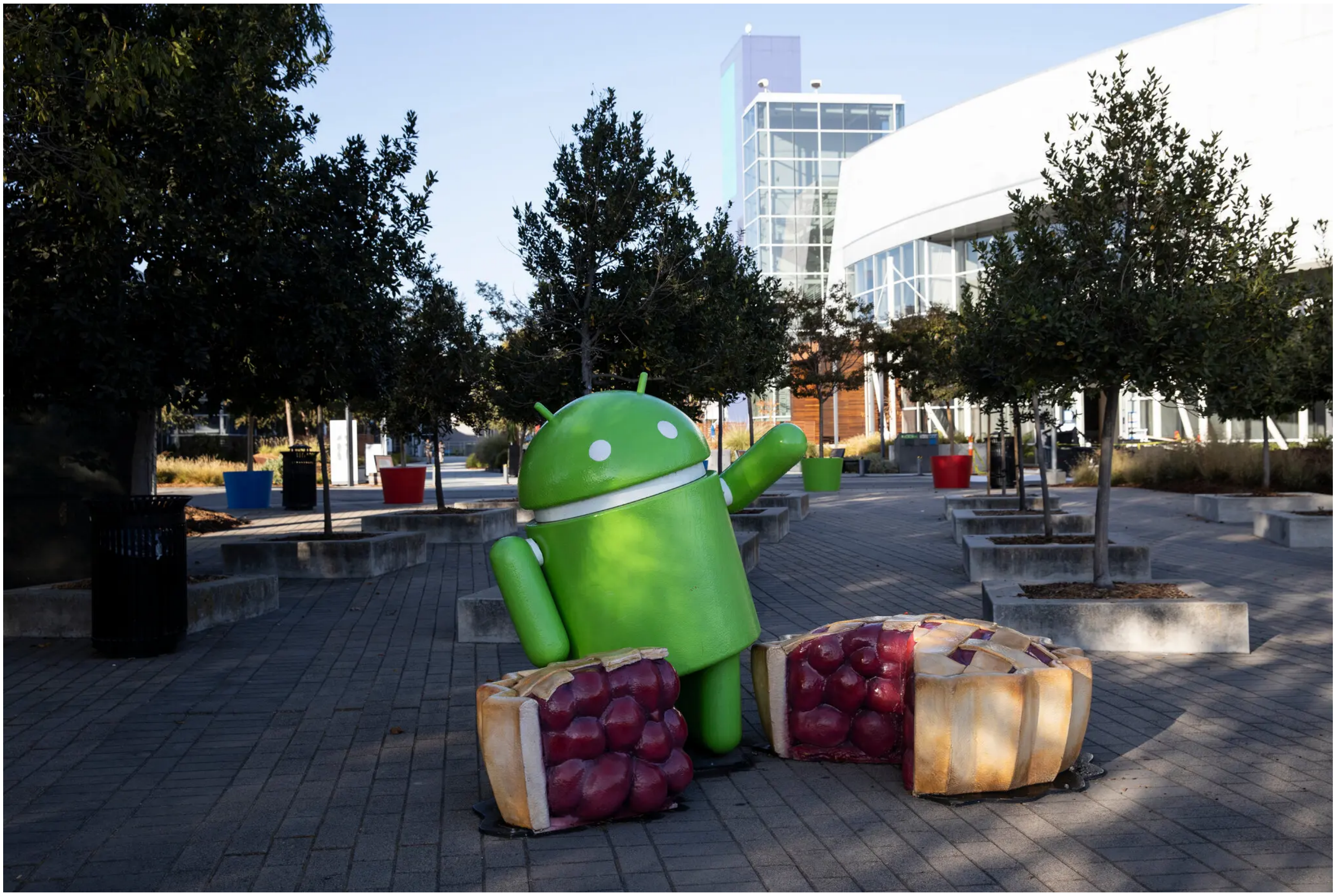
Google disputed Mr. Lemoine's claims, and lots of A.I. researchers have quibbled with his conclusions. But take out the sentience part, and a weaker version of his argument — that LaMDA and other state-of-the-art language models are becoming eerily good at having humanlike text conversations — would not have raised nearly as many eyebrows.

In fact, many experts will tell you that A.I. is getting better at lots of things these days — even in areas, such as language and reasoning, where it once seemed that humans had the upper hand.

“It feels like we're going from spring to summer,” said Jack Clark, a co-chair of Stanford University's annual A.I. Index Report. “In spring, you have these vague suggestions of progress, and little green shoots everywhere. Now, everything's in bloom.”

In the past, A.I. progress was mostly obvious only to insiders who kept up with the latest research papers and conference presentations. But recently, Mr. Clark said, even laypeople can sense the difference.

“You used to look at A.I.-generated language and say, ‘Wow, it *kind of* wrote a sentence,’” Mr. Clark said. “And now you're looking at stuff that's A.I.-generated and saying, ‘This is really funny, I'm enjoying reading this,’ or ‘I had no idea this was even generated by A.I.’”



Google's headquarters in Mountain View, Calif. Big tech firms need to do a better job of explaining what they're working on, without the P.R. Laura Morton for The New York Times

There is still plenty of bad, broken A.I. out there, from racist chatbots to faulty automated driving systems that result in crashes and injury. And even when A.I. improves quickly, it often takes a while to filter down into products and services that people actually use. An A.I. breakthrough at Google or OpenAI today doesn't mean that your Roomba will be able to write novels tomorrow.

But the best A.I. systems are now so capable — and improving at such fast rates — that the conversation in Silicon Valley is starting to shift. Fewer experts are confidently predicting that we have years or even decades to prepare for a wave of world-changing A.I.; many now believe that major changes are right around the corner, for better or worse.

Ajeya Cotra, a senior analyst with Open Philanthropy who studies A.I. risk, estimated two years ago that there was a 15 percent chance of "transformational A.I." — which she and others have defined as A.I. that is good enough to usher in large-scale economic and societal changes, such as eliminating most white-collar knowledge jobs — emerging by 2036.

But in a recent post, Ms. Cotra raised that to a 35 percent chance, citing the rapid improvement of systems like GPT-3.

"A.I. systems can go from adorable and useless toys to very powerful products in a surprisingly short period of time," Ms. Cotra told me. "People should take more seriously that A.I. could change things soon, and that could be really scary."

There are, to be fair, plenty of skeptics who say claims of A.I. progress are overblown. They'll tell you that A.I. is still nowhere close to becoming sentient, or replacing humans in a wide variety of jobs. They'll say that models like GPT-3 and LaMDA are just glorified parrots, blindly regurgitating their training data, and that we're still decades away from creating true A.G.I. — artificial general intelligence — that is capable of "thinking" for itself.

There are also tech optimists who believe that A.I. progress is accelerating, and who want it to accelerate faster. Speeding A.I.'s rate of improvement, they believe, will give us new tools to cure diseases, colonize space and avert ecological disaster.

I'm not asking you to take a side in this debate. All I'm saying is: You should be paying closer attention to the real, tangible developments that are fueling it.

After all, A.I. that works doesn't stay in a lab. It gets built into the social media apps we use every day, in the form of Facebook feed-ranking algorithms, YouTube recommendations and TikTok "For You" pages. It makes its way into weapons used by the military and software used by children in their classrooms. Banks use A.I. to determine who's eligible for loans, and police departments use it to investigate crimes.

Even if the skeptics are right, and A.I. doesn't achieve human-level sentience for many years, it's easy to see how systems like GPT-3, LaMDA and DALL-E 2 could become a powerful force in society. In a few years, the vast majority of the photos, videos and text we encounter on the internet could be A.I.-generated. Our online interactions could become stranger and more fraught, as we struggle to figure out which of our conversational partners are human and which are convincing bots. And tech-savvy propagandists could use the technology to churn out targeted misinformation on a vast scale, distorting the political process in ways we won't see coming.

It's a cliché, in the A.I. world, to say things like "we need to have a societal conversation about A.I. risk." There are already plenty of Davos panels, TED talks, think tanks and A.I. ethics committees out there, sketching out contingency plans for a dystopian future.

What's missing is a shared, value-neutral way of talking about what today's A.I. systems are actually capable of doing, and what specific risks and opportunities those capabilities present.

I think three things could help here.

First, regulators and politicians need to get up to speed.

Because of how new many of these A.I. systems are, few public officials have any firsthand experience with tools like GPT-3 or DALL-E 2, nor do they grasp how quickly progress is happening at the A.I. frontier.

We've seen a few efforts to close the gap — Stanford's Institute for Human-Centered Artificial Intelligence recently held a three-day "A.I. boot camp" for congressional staff members, for example — but we need more politicians and regulators to take an interest in the technology. (And I don't mean that they need to start stoking fears of an A.I. apocalypse, Andrew Yang-style. Even reading a book like Brian Christian's "The Alignment Problem" or understanding a few basic details about how a model like GPT-3 works would represent enormous progress.)

Otherwise, we could end up with a repeat of what happened with social media companies after the 2016 election — a collision of Silicon Valley power and Washington ignorance, which resulted in nothing but gridlock and testy hearings.

Second, big tech companies investing billions in A.I. development — the Googles, Metas and OpenAIs of the world — need to do a better job of explaining what they're working on, without sugarcoating or soft-pedaling the risks. Right now, many of the biggest A.I. models are developed behind closed doors, using private data sets and tested only by internal teams. When information about them is made public, it's often either watered down by corporate P.R. or buried in inscrutable scientific papers.

Downplaying A.I. risks to avoid backlash may be a smart short-term strategy, but tech companies won't survive long term if they're seen as having a hidden A.I. agenda that's at odds with the public interest. And if these companies won't open up voluntarily, A.I. engineers should go around their bosses and talk directly to policymakers and journalists themselves.

Third, the news media needs to do a better job of explaining A.I. progress to nonexperts. Too often, journalists — and I admit I've been a guilty party here — rely on outdated sci-fi shorthand to translate what's happening in A.I. to a general audience. We sometimes compare large language models to Skynet and HAL 9000, and flatten promising machine learning breakthroughs to panicky "The robots are coming!" headlines that we think will resonate with readers. Occasionally, we betray our ignorance by illustrating articles about software-based A.I. models with photos of hardware-based factory robots — an error that is as inexplicable as slapping a photo of a BMW on a story about bicycles.

In a broad sense, most people think about A.I. narrowly as it relates to *us* — Will it take my job? Is it better or worse than me at Skill X or Task Y? — rather than trying to understand all of the ways A.I. is evolving, and what that might mean for our future.

I'll do my part, by writing about A.I. in all its complexity and weirdness without resorting to hyperbole or Hollywood tropes. But we all need to start adjusting our mental models to make space for the new, incredible machines in our midst.

Kevin Roose is a technology columnist and the author of "Futureproof: 9 Rules for Humans in the Age of Automation." @kevinroose • Facebook

A version of this article appears in print on , Section B, Page 1 of the New York edition with the headline: A.I. Is Getting Good. What Happens Now?