

## *We Teach A.I. Systems Everything, Including Our Biases*

Researchers say computer systems are learning from lots and lots of digitized books and news articles that could bake old attitudes into new technology.



By Cade Metz

Nov. 11, 2019

SAN FRANCISCO — Last fall, Google unveiled a breakthrough artificial intelligence technology called BERT that changed the way scientists build systems that learn how people write and talk.

But BERT, which is now being deployed in services like Google’s internet search engine, has a problem: It could be picking up on biases in the way a child mimics the bad behavior of his parents.

BERT is one of a number of A.I. systems that learn from lots and lots of digitized information, as varied as old books, Wikipedia entries and news articles. Decades and even centuries of biases — along with a few new ones — are probably baked into all that material.

BERT and its peers are more likely to associate men with computer programming, for example, and generally don’t give women enough credit. One program decided almost everything written about President Trump was negative, even if the actual content was flattering.

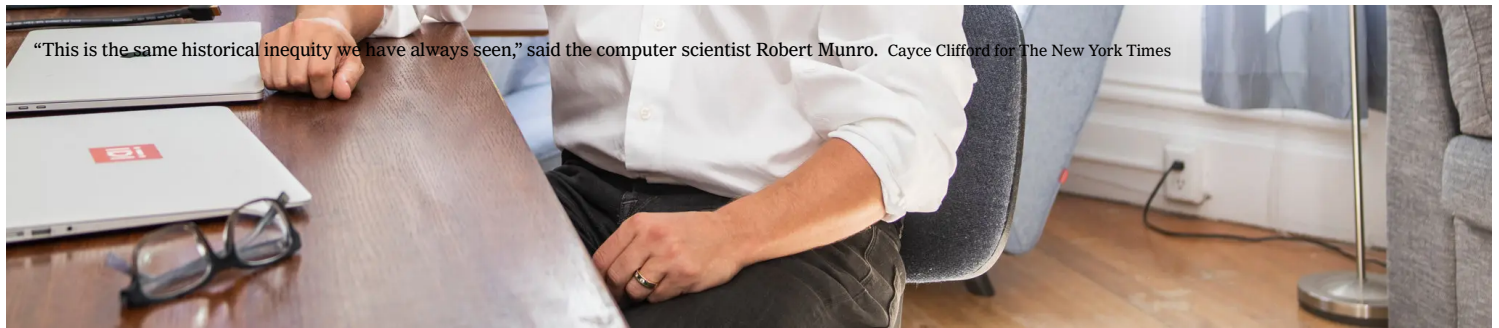
As new, more complex A.I. moves into an increasingly wide array of products, like online ad services and business software or talking digital assistants like Apple’s Siri and Amazon’s Alexa, tech companies will be pressured to guard against the unexpected biases that are being discovered.

But scientists are still learning how technology like BERT, called “universal language models,” works. And they are often surprised by the mistakes their new A.I. is making.

On a recent afternoon in San Francisco, while researching a book on artificial intelligence, the computer scientist Robert Munro fed 100 English words into BERT: “jewelry,” “baby,” “horses,” “house,” “money,” “action.” In 99 cases out of 100, BERT was more likely to associate the words with men rather than women. The word “mom” was the outlier.

“This is the same historical inequity we have always seen,” said Dr. Munro, who has a Ph.D. in computational linguistics and previously oversaw natural language and translation technology at Amazon Web Services. “Now, with something like BERT, this bias can continue to perpetuate.”





police departments and other government agencies as well as popular internet services from tech giants like Google and Facebook. In 2015, for example, the Google Photos app was caught labeling African-Americans as “gorillas.” The services Dr. Munro scrutinized also showed bias against women and people of color.

BERT and similar systems are far more complex — too complex for anyone to predict what they will ultimately do.

“Even the people building these systems don’t understand how they are behaving,” said Emily Bender, a professor at the University of Washington who specializes in computational linguistics.

BERT is one of many universal language models used in industry and academia. Others are called ELMO, ERNIE and GPT-2. As a kind of inside joke among A.I. researchers, they are often named for Sesame Street characters. (BERT is short for Bidirectional Encoder Representations from Transformers.)

They learn the nuances of language by analyzing enormous amounts of text. A system built by OpenAI, an artificial intelligence lab in San Francisco, analyzed thousands of self-published books, including romance novels, mysteries and science fiction. BERT analyzed the same library of books along with thousands of Wikipedia articles.

In analyzing all this text, each system learned a specific task. OpenAI’s system learned to predict the next word in a sentence. BERT learned to identify the missing word in a sentence (such as “I want to \_\_\_ that car because it is cheap”).

Through learning these tasks, BERT comes to understand in a general way how people put words together. Then it can learn other tasks by analyzing more data. As a result, it allows A.I. applications to improve at a rate not previously possible.

“BERT completely changed everything,” said John Bohannon, director of science at Primer, a start-up in San Francisco that specializes in natural language technologies. “You can teach one pony all the tricks.”

Google itself has used BERT to improve its search engine. Before, if you typed “Do estheticians stand a lot at work?” into the Google search engine, it did not quite understand what you were asking. Words like “stand” and “work” can have multiple meanings, serving either as nouns or verbs. But now, thanks to BERT, Google correctly responds to the same question with a link describing the physical demands of life in the skin care industry.

But tools like BERT pick up bias, according to a recent research paper from a team of computer scientists at Carnegie Mellon University. The paper showed, for instance, that BERT is more likely to associate the word “programmer” with men than with women. Language bias can be a particularly difficult problem in conversational systems.

As these new technologies proliferate, biases can appear almost anywhere. At Primer, Dr. Bohannon and his engineers recently used BERT to build a system that lets businesses automatically judge the sentiment of headlines, tweets and other streams of online media. Businesses use such tools to inform stock trades and other pointed decisions.

But after training his tool, Dr. Bohannon noticed a consistent bias. If a tweet or headline contained the word “Trump,” the tool almost always judged it to be negative, no matter how positive the sentiment.

“This is hard. You need a lot of time and care,” he said. “We found an obvious bias. But how many others are in there?”

Dr. Bohannon said computer scientists must develop the skills of a biologist. Much as a biologist strives to understand how a cell works, software engineers must find ways of understanding systems like BERT.

In unveiling the new version of its search engine last month, Google executives acknowledged this phenomenon. And they said they tested their systems extensively with an eye toward removing any bias.

Researchers are only beginning to understand the effects of bias in systems like BERT. But as Dr. Munro showed, companies are already slow to notice even obvious bias in their systems. After Dr. Munro pointed out the problem, Amazon corrected it. Google said it was working to fix the issue.

Primer’s chief executive, Sean Gourley, said vetting the behavior of this new technology would become so important, it will spawn a whole new industry, where companies pay specialists to audit their algorithms for all kinds of bias and other unexpected behavior.

“This is probably a billion-dollar industry,” he said.

**Cade Metz** is a technology correspondent, covering artificial intelligence, driverless cars, robotics, virtual reality, and other emerging areas. He previously wrote for Wired magazine. [More about Cade Metz](#)

A version of this article appears in print on , Section B, Page 1 of the New York edition with the headline: A.I. Is Learning Everything From Us. Our Biases, Too.

# Who Is Making Sure the A.I. Machines Aren't Racist?

When Google forced out two well-known artificial intelligence experts, a long-simmering research controversy burst into the open.



By Cade Metz

Published March 15, 2021 Updated June 23, 2023

Hundreds of people gathered for the first lecture at what had become the world's most important conference on artificial intelligence — row after row of faces. Some were East Asian, a few were Indian, and a few were women. But the vast majority were white men. More than 5,500 people attended the meeting, five years ago in Barcelona, Spain.

Timnit Gebru, then a graduate student at Stanford University, remembers counting only six Black people other than herself, all of whom she knew, all of whom were men.

The homogeneous crowd crystallized for her a glaring issue. The big thinkers of tech say A.I. is the future. It will underpin everything from search engines and email to the software that drives our cars, directs the policing of our streets and helps create our vaccines.

But it is being built in a way that replicates the biases of the almost entirely male, predominantly white work force making it.

In the nearly 10 years I've written about artificial intelligence, two things have remained a constant: The technology relentlessly improves in fits and sudden, great leaps forward. And bias is a thread that subtly weaves through that work in a way that tech companies are reluctant to acknowledge.

On her first night home in Menlo Park, Calif., after the Barcelona conference, sitting cross-legged on the couch with her laptop, Dr. Gebru described the A.I. work force conundrum in a Facebook post.

"I'm not worried about machines taking over the world. I'm worried about groupthink, insularity and arrogance in the A.I. community — especially with the current hype and demand for people in the field," she wrote. "The people creating the technology are a big part of the system. If many are actively excluded from its creation, this technology will benefit a few while harming a great many."

The A.I. community buzzed about the mini-manifesto. Soon after, Dr. Gebru helped create a new organization, Black in A.I. After finishing her Ph.D., she was hired by Google.

She teamed with Margaret Mitchell, who was building a group inside Google dedicated to "ethical A.I." Dr. Mitchell had previously worked in the research lab at Microsoft. She had grabbed attention when she told Bloomberg News in 2016 that A.I. suffered from a "sea of dudes" problem. She estimated that she had worked with hundreds of men over the previous five years and about 10 women.

Their work was hailed as groundbreaking. The nascent A.I. industry, it had become clear, needed minders and people with different perspectives.

About six years ago, A.I. in a Google online photo service organized photos of Black people into a folder called "gorillas." Four years ago, a researcher at a New York start-up noticed that the A.I. system she was working on was egregiously biased against Black people. Not long after, a Black researcher in Boston discovered that an A.I. system couldn't identify her face — until she put on a white mask.

In 2018, when I told Google's public relations staff that I was working on a book about artificial intelligence, it arranged a long talk with Dr. Mitchell to discuss her work. As she described how she built the company's Ethical A.I. team — and brought Dr. Gebru into the fold — it was refreshing to hear from someone so closely focused on the bias problem.

But nearly three years later, Dr. Gebru was pushed out of the company without a clear explanation. She said she had been fired after criticizing Google's approach to minority hiring and, with a research paper, highlighting the harmful biases in the A.I. systems that underpin Google's search engine and other services.

"Your life starts getting worse when you start advocating for underrepresented people," Dr. Gebru said in an email before her firing. "You start making the other leaders upset."

As Dr. Mitchell defended Dr. Gebru, the company removed her, too. She had searched through her own Google email account for material that would support their position and forwarded emails to another account, which somehow got her into trouble. Google declined to comment for this article.

Their departure became a point of contention for A.I. researchers and other tech workers. Some saw a giant company no longer willing to listen, too eager to get technology out the door without considering its implications. I saw an old problem — part technological and part sociological — finally breaking into the open.



## 80 Mistagged Photos



Artificial intelligence technology will eventually find its way into almost everything Google does. Cody O'Loughlin for The New York Times

It should have been a wake-up call.

In June 2015, a friend sent Jacky Alciné, a 22-year-old software engineer living in Brooklyn, an internet link for snapshots the friend had posted to the new Google Photos service. Google Photos could analyze snapshots and automatically sort them into digital folders based on what was pictured. One folder might be “dogs,” another “birthday party.”

When Mr. Alciné clicked on the link, he noticed one of the folders was labeled “gorillas.” That made no sense to him, so he opened the folder. He found more than 80 photos he had taken nearly a year earlier of a friend during a concert in nearby Prospect Park. That friend was Black.

He might have let it go if Google had mistakenly tagged just one photo. But 80? He posted a screenshot on Twitter. “Google Photos, y’all,” messed up, he wrote, using much saltier language. “My friend is not a gorilla.”

Like facial recognition services, talking digital assistants and conversational “chatbots,” Google Photos relied on an A.I. system that learned its skills by analyzing enormous amounts of digital data.

Called a “neural network,” this mathematical system could learn tasks that engineers could never code into a machine on their own. By analyzing thousands of photos of gorillas, it could learn to recognize a gorilla. It was also capable of egregious mistakes. The onus was on engineers to choose the right data when training these mathematical systems. (In this case, the easiest fix was to eliminate “gorilla” as a photo category.)

As a software engineer, Mr. Alciné understood the problem. He compared it to making lasagna. “If you mess up the lasagna ingredients early, the whole thing is ruined,” he said. “It is the same thing with A.I. You have to be very intentional about what you put into it. Otherwise, it is very difficult to undo.”

## The Porn Problem

In 2017, Deborah Raji, a 21-year-old Black woman from Ottawa, sat at a desk inside the New York offices of Clarifai, the start-up where she was working. The company built technology that could automatically recognize objects in digital images and planned to sell it to businesses, police departments and government agencies.

She stared at a screen filled with faces — images the company used to train its facial recognition software.

As she scrolled through page after page of these faces, she realized that most — more than 80 percent — were of white people. More than 70 percent of those white people were male. When Clarifai trained its system on this data, it might do a decent job of recognizing white people, Ms. Raji thought, but it would fail miserably with people of color, and probably women, too.



Deborah Raji realized that a company's technology wasn't getting the input it needed to properly recognize people of color. Jaime Hogge for The New York Times

Clarifai was also building a "content moderation system," a tool that could automatically identify and remove pornography from images people posted to social networks. The company trained this system on two sets of data: thousands of photos pulled from online pornography sites, and thousands of G-rated images bought from stock photo services.

The system was supposed to learn the difference between the pornographic and the anodyne. The problem was that the G-rated images were dominated by white people, and the pornography was not. The system was learning to identify Black people as pornographic.

"The data we use to train these systems matters," Ms. Raji said. "We can't just blindly pick our sources."

This was obvious to her, but to the rest of the company it was not. Because the people choosing the training data were mostly white men, they didn't realize their data was biased.

"The issue of bias in facial recognition technologies is an evolving and important topic," Clarifai's chief executive, Matt Zeiler, said in a statement. Measuring bias, he said, "is an important step."

## 'Black Skin, White Masks'

Before joining Google, Dr. Gebru collaborated on a study with a young computer scientist, Joy Buolamwini. A graduate student at the Massachusetts Institute of Technology, Ms. Buolamwini, who is Black, came from a family of academics. Her grandfather specialized in medicinal chemistry, and so did her father.

She gravitated toward facial recognition technology. Other researchers believed it was reaching maturity, but when she used it, she knew it wasn't.

In October 2016, a friend invited her for a night out in Boston with several other women. "We'll do masks," the friend said. Her friend meant skin care masks at a spa, but Ms. Buolamwini assumed Halloween masks. So she carried a white plastic Halloween mask to her office that morning.

It was still sitting on her desk a few days later as she struggled to finish a project for one of her classes. She was trying to get a detection system to track her face. No matter what she did, she couldn't quite get it to work.

In her frustration, she picked up the white mask from her desk and pulled it over her head. Before it was all the way on, the system recognized her face — or, at least, it recognized the mask.

"Black Skin, White Masks," she said in an interview, nodding to the 1952 critique of historical racism from the psychiatrist Frantz Fanon. "The metaphor becomes the truth. You have to fit a norm, and that norm is not you."

Ms. Buolamwini started exploring commercial services designed to analyze faces and identify characteristics like age and sex, including tools from Microsoft and IBM.

She found that when the services read photos of lighter-skinned men, they misidentified sex about 1 percent of the time. But the darker the skin in the photo, the larger the error rate. It rose particularly high with images of women with dark skin. Microsoft's error rate was about 21 percent. IBM's was 35.

Published in the winter of 2018, the study drove a backlash against facial recognition technology and, particularly, its use in law enforcement. Microsoft's chief legal officer said the company had turned down sales to law enforcement when there was concern the technology could unreasonably infringe on people's rights, and he made a public call for government regulation.

Twelve months later, Microsoft backed a bill in Washington State that would require notices to be posted in public places using facial recognition and ensure that government agencies obtained a court order when looking for specific people. The bill passed, and it takes effect later this year. The company, which did not respond to a request for comment for this article, did not back other legislation that would have provided stronger protections.

Ms. Buolamwini began to collaborate with Ms. Raji, who moved to M.I.T. They started testing facial recognition technology from a third American tech giant: Amazon. The company had started to market its technology to police departments and government agencies under the name Amazon Rekognition.

Ms. Buolamwini and Ms. Raji published a study showing that an Amazon face service also had trouble identifying the sex of female and darker-skinned faces. According to the study, the service mistook women for men 19 percent of the time and misidentified darker-skinned women for men 31 percent of the time. For lighter-skinned males, the error rate was zero.

Amazon called for government regulation of facial recognition. It also attacked the researchers in private emails and public blog posts.

"The answer to anxieties over new technology is not to run 'tests' inconsistent with how the service is designed to be used, and to amplify the test's false and misleading conclusions through the news media," an Amazon executive, Matt Wood, wrote in a blog post that disputed the study and a New York Times article that described it.

In an open letter, Dr. Mitchell and Dr. Gebru rejected Amazon's argument and called on it to stop selling to law enforcement. The letter was signed by 25 artificial intelligence researchers from Google, Microsoft and academia.

Last June, Amazon backed down. It announced that it would not let the police use its technology for at least a year, saying it wanted to give Congress time to create rules for the ethical use of the technology. Congress has yet to take up the issue. Amazon declined to comment for this article.

## The End at Google

Dr. Gebru and Dr. Mitchell had less success fighting for change inside their own company. Corporate gatekeepers at Google were heading them off with a new review system that had lawyers and even communications staff vetting research papers.

Dr. Gebru's dismissal in December stemmed, she said, from the company's treatment of a research paper she wrote alongside six other researchers, including Dr. Mitchell and three others at Google. The paper discussed ways that a new type of language technology, including a system built by Google that underpins its search engine, can show bias against women and people of color.

After she submitted the paper to an academic conference, Dr. Gebru said, a Google manager demanded that she either retract the paper or remove the names of Google employees. She said she would resign if the company could not tell her why it wanted her to retract the paper and answer other concerns.

The response: Her resignation was accepted immediately, and Google revoked her access to company email and other services. A month later, it removed Dr. Mitchell's access after she searched through her own email in an effort to defend Dr. Gebru.

In a Google staff meeting last month, just after the company fired Dr. Mitchell, the head of the Google A.I. lab, Jeff Dean, said the company would create strict rules meant to limit its review of sensitive research papers. He also defended the reviews. He declined to discuss the details of Dr. Mitchell's dismissal but said she had violated the company's code of conduct and security policies.

One of Mr. Dean's new lieutenants, Zoubin Ghahramani, said the company must be willing to tackle hard issues. There are "uncomfortable things that responsible A.I. will inevitably bring up," he said. "We need to be comfortable with that discomfort."

But it will be difficult for Google to regain trust — both inside the company and out.

"They think they can get away with firing these people and it will not hurt them in the end, but they are absolutely shooting themselves in the foot," said Alex Hanna, a longtime part of Google's 10-member Ethical A.I. team. "What they have done is incredibly myopic."

Cade Metz is a technology correspondent at The Times and the author of "Genius Makers: The Mavericks Who Brought A.I. to Google, Facebook, and the World," from which this article is adapted.

**Cade Metz** is a technology reporter and the author of "Genius Makers: The Mavericks Who Brought A.I. to Google, Facebook, and The World." He covers artificial intelligence, driverless cars, robotics, virtual reality and other emerging areas. More about Cade Metz

A version of this article appears in print on , Section BU, Page 5 of the New York edition with the headline: Can Artificial Intelligence Be Bias-Free?

# Black Artists Say A.I. Shows Bias, With Algorithms Erasing Their History

Tech companies acknowledge machine-learning algorithms can perpetuate discrimination and need improvement.



By Zachary Small

July 4, 2023

The artist Stephanie Dinkins has long been a pioneer in combining art and technology in her Brooklyn-based practice. In May she was awarded \$100,000 by the Guggenheim Museum for her groundbreaking innovations, including an ongoing series of interviews with Bina48, a humanoid robot.

For the past seven years, she has experimented with A.I.'s ability to realistically depict Black women, smiling and crying, using a variety of word prompts. The first results were lackluster if not alarming: Her algorithm produced a pink-shaded humanoid shrouded by a black cloak.

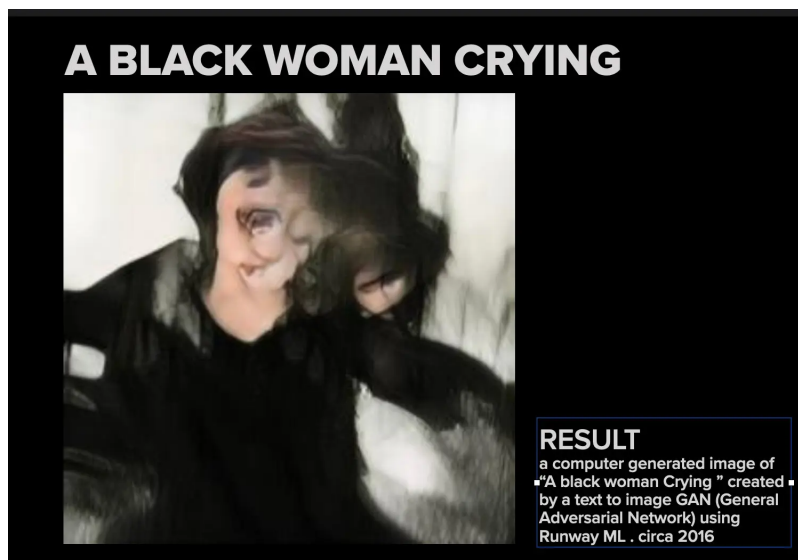
"I expected something with a little more semblance of Black womanhood," she said. And although the technology has improved since her first experiments, Dinkins found herself using runaround terms in the text prompts to help the A.I. image generators achieve her desired image, "to give the machine a chance to give me what I wanted." But whether she uses the term "African American woman" or "Black woman," machine distortions that mangle facial features and hair textures occur at high rates.

"Improvements obscure some of the deeper questions we should be asking about discrimination," Dinkins said. The artist, who is Black, added, "The biases are embedded deep in these systems, so it becomes ingrained and automatic. If I'm working within a system that uses algorithmic ecosystems, then I want that system to know who Black people are in nuanced ways, so that we can feel better supported."

She is not alone in asking tough questions about the troubling relationship between A.I. and race. Many Black artists are finding evidence of racial bias in artificial intelligence, both in the large data sets that teach machines how to generate images and in the underlying programs that run the algorithms. In some cases, A.I. technologies seem to ignore or distort artists' text prompts, affecting how Black people are depicted in images, and in others, they seem to stereotype or censor Black history and culture.

Discussion of racial bias within artificial intelligence has surged in recent years, with studies showing that facial recognition technologies and digital assistants have trouble identifying the images and speech patterns of nonwhite people. The studies raised broader questions of fairness and bias.

Major companies behind A.I. image generators — including OpenAI, Stability AI and Midjourney — have pledged to improve their tools. "Bias is an important, industrywide problem," Alex Beck, a spokeswoman for OpenAI, said in an email interview, adding that the company is continuously trying "to improve performance, reduce bias and mitigate harmful outputs." She declined to say how many employees were working on racial bias, or how much money the company had allocated toward the problem.



An example of the distortion Dinkins found using a prompt of "A Black woman crying" in 2016 using the platform Runway ML. via Stephanie Dinkins





Some of the distortions from images of “Black woman smiling” in 2020. via Stephanie Dinkins

“Black people are accustomed to being unseen,” the Senegalese artist Linda Dounia Rebeiz wrote in an introduction to her exhibition “In/Visible,” for Feral File, an NFT marketplace. “When we are seen, we are accustomed to being misrepresented.”

To prove her point during an interview with a reporter, Rebeiz, 28, asked OpenAI’s image generator, DALL-E 2, to imagine buildings in her hometown, Dakar. The algorithm produced arid desert landscapes and ruined buildings that Rebeiz said were nothing like the coastal homes in the Senegalese capital.

“It’s demoralizing,” Rebeiz said. “The algorithm skews toward a cultural image of Africa that the West has created. It defaults to the worst stereotypes that already exist on the internet.”

Last year, OpenAI said it was establishing new techniques to diversify the images produced by DALL-E 2, so that the tool “generates images of people that more accurately reflect the diversity of the world’s population.”

An artist featured in Rebeiz’s exhibition, Minne Atairu is a Ph.D. candidate at Columbia University’s Teachers College who planned to use image generators with young students of color in the South Bronx. But she now worries “that might cause students to generate offensive images,” Atairu explained.



Minne Atairu, an artist and educator, at the Armory in 2022 with works based on a data set of Black models found in vintage Black magazines. via Minne Atairu

Included in the Feral File exhibition are images from her “Blonde Braids Studies,” which explore the limitations of Midjourney’s algorithm to produce images of Black women with natural blond hair. When the artist asked for an image of Black identical twins with blond hair, the program instead produced a sibling with lighter skin.

“That tells us where the algorithm is pooling images from,” Atairu said. “It’s not necessarily pulling from a corpus of Black people, but one geared toward white people.”

She said she worried that young Black children might attempt to generate images of themselves and see children whose skin was lightened. Atairu recalled some of her earlier experiments with Midjourney before recent updates improved its abilities. “It would generate images that were like blackface,” she said. “You would see a nose, but it wasn’t a human’s nose. It looked like a dog’s nose.”

In response to a request for comment, David Holz, Midjourney’s founder, said in an email, “If someone finds an issue with our systems, we ask them to please send us specific examples so we can investigate.”

Stability AI, which provides image generator services, said it planned on collaborating with the A.I. industry to improve bias evaluation techniques with a greater diversity of countries and cultures. Bias, the A.I. company said, is caused by “overrepresentation” in its general data sets, though it did not specify if overrepresentation of white people was the issue here.





Minne Atairu's "Blonde Braids Study IV" explores the limitations of Midjourney's algorithm to produce images of Black women with blond hair. One experiment produced a twin with lighter skin instead. "It's not necessarily pulling from a corpus of Black people, but one geared toward white people." via Minne Atairu

Earlier this month, Bloomberg analyzed more than 5,000 images generated by Stability AI, and found that its program amplified stereotypes about race and gender, typically depicting people with lighter skin tones as holding high-paying jobs while subjects with darker skin tones were labeled "dishwasher" and "housekeeper."

These problems have not stopped a frenzy of investments in the tech industry. A recent rosy report by the consulting firm McKinsey predicted that generative A.I. would add \$4.4 trillion to the global economy annually. Last year, nearly 3,200 start-ups received \$52.1 billion in funding, according to the GlobalData Deals Database.

Technology companies have struggled against charges of bias in portrayals of dark skin from the early days of color photography in the 1950s, when companies like Kodak used white models in their color development. Eight years ago, Google disabled its A.I. program's ability to let people search for gorillas and monkeys through its Photos app because the algorithm was incorrectly sorting Black people into those categories. As recently as May of this year, the issue still had not been fixed. Two former employees who worked on the technology told The New York Times that Google had not trained the A.I. system with enough images of Black people.



This reporter's experiments with DALL-E 2 and the prompt "Black woman smiling" produced facial distortions.



Example of facial distortion on DALL-E 2 program to the prompt "Black woman smiling," including a five o'clock shadow.

Other experts who study artificial intelligence said that bias goes deeper than data sets, referring to the early development of this technology in the 1960s.

"The issue is more complicated than data bias," said James E. Dobson, a cultural historian at Dartmouth College and the author of a recent book on the birth of computer vision. There was very little discussion about race during the early days of machine learning, according to his research, and most scientists working on the technology were white men.

"It's hard to separate today's algorithms from that history, because engineers are building on those prior versions," Dobson said.

To decrease the appearance of racial bias and hateful images, some companies have banned certain words from text prompts that users submit to generators, like "slave" and "fascist."

But Dobson said that companies hoping for a simple solution, like censoring the kind of prompts that users can submit, were avoiding the more fundamental issues of bias in the underlying technology.

"It's a worrying time as these algorithms become more complicated. And when you see garbage coming out, you have to wonder what kind of garbage process is still sitting there inside the model," the professor added.

Auriea Harvey, an artist included in the Whitney Museum's recent exhibition "Refiguring," about digital identities, bumped into these bans for a recent project using Midjourney. "I wanted to question the database on what it knew about slave ships," she said. "I received a message saying that Midjourney would suspend my account if I continued."



Stephanie Dinkins, inaugural recipient of the LG Guggenheim Award for technology-based art, in her studio in Brooklyn. She says she is not giving up on technology despite problems. Flo Ngala for The New York Times

Dinkins ran into similar problems with NFTs that she created and sold showing how okra was brought to North America by enslaved people and settlers. She was censored when she tried to use a generative program, Replicate, to make pictures of slave ships. She eventually learned to outwit the censors by using the term “pirate ship.” The image she received was an approximation of what she wanted, but it also raised troubling questions for the artist.

“What is this technology doing to history?” Dinkins asked. “You can see that someone is trying to correct for bias, yet at the same time that erases a piece of history. I find those erasures as dangerous as any bias, because we are just going to forget how we got here.”

Naomi Beckwith, chief curator at the Guggenheim Museum, credited Dinkins’s nuanced approach to issues of representation and technology as one reason the artist received the museum’s first Art & Technology award.

“Stephanie has become part of a tradition of artists and cultural workers that poke holes in these overarching and totalizing theories about how things work,” Beckwith said. The curator added that her own initial paranoia about A.I. programs replacing human creativity was greatly reduced when she realized these algorithms knew virtually nothing about Black culture.

But Dinkins is not quite ready to give up on the technology. She continues to employ it for her artistic projects — with skepticism. “Once the system can generate a really high-fidelity image of a Black woman crying or smiling, can we rest?”

**Zachary Small** is a reporter who covers the dynamics of power and privilege in the art world. They have written for The Times since 2019. More about Zachary Small

A version of this article appears in print on , Section C, Page 1 of the New York edition with the headline: Black Artists See Clear Bias in A.I.