

Credit: Michael Glenwood

## Artificial General Intelligence Is Not as Imminent as You Might Think

A close look reveals that the newest systems, including DeepMind's much-hyped Gato, are still stymied by the same old problems

By Gary Marcus

### AUTHOR



**Gary Marcus** is a scientist, best-selling author and entrepreneur. His most recent book, co-authored with Ernest Davis, is *Rebooting AI* (Vintange, 2019). Credit: Nick Higgins

To the average person, it must seem as if the field of artificial intelligence is making immense progress. According to some of the more gushing media accounts and press releases, OpenAI's DALL-E 2 can seemingly create spectacular images from any text; another OpenAI system called GPT-3 can talk about just about anything—and even write about itself; and a system called Gato that was released in May by DeepMind, a division of Alphabet, reportedly worked well on every task the company could throw at it. One of DeepMind's high-level executives even went so far as to brag that in the quest to create AI that has the flexibility and resourcefulness of human intelligence—known as artificial general intelligence, or AGI—"the game is over."

Don't be fooled. Machines may someday be as smart as people and perhaps even smarter, but the game is far from over. There is still an immense amount of work to be done in making machines that truly can comprehend and reason about the world around them. What we need right now is less posturing and more basic research.

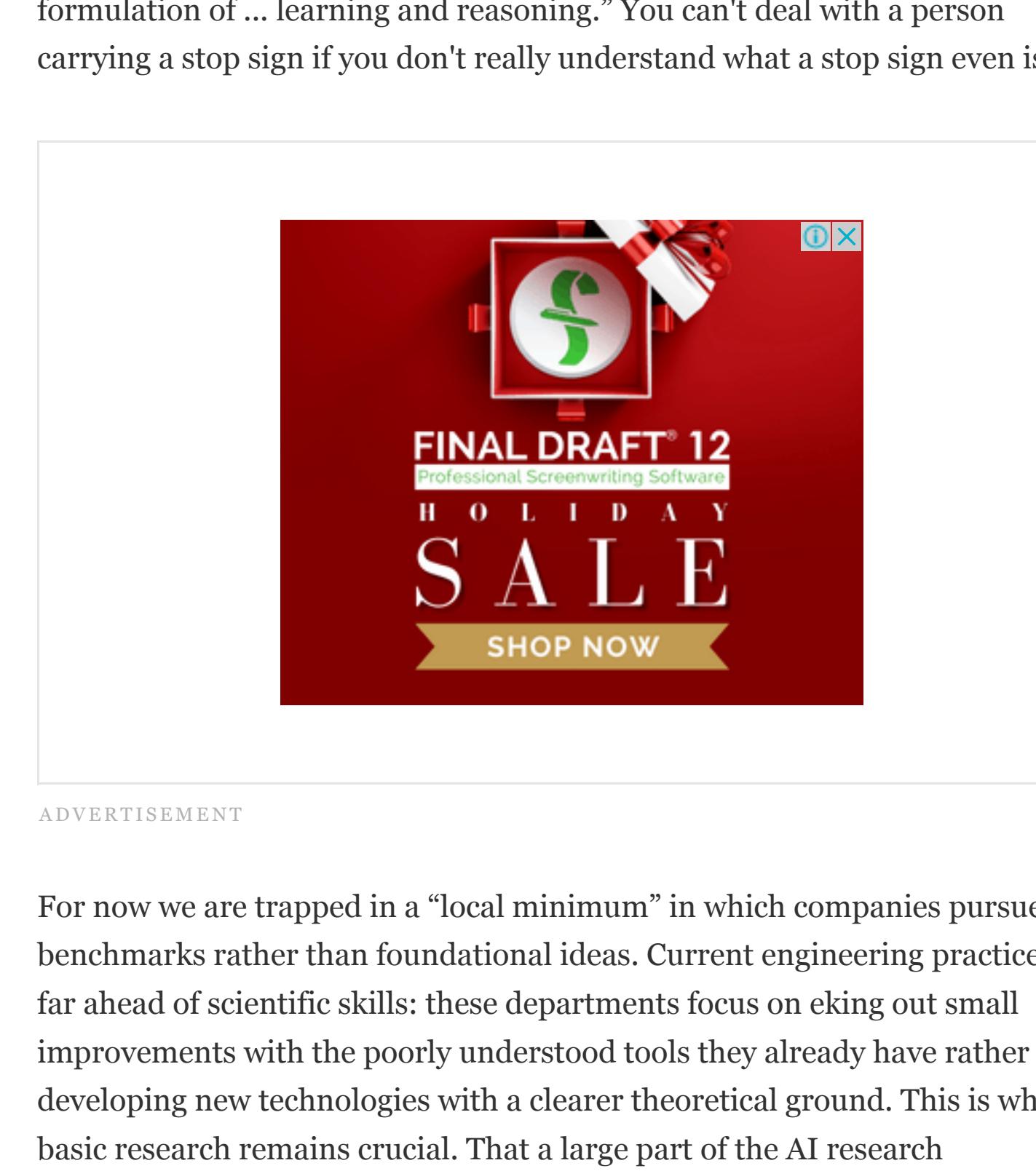
AI is making progress—synthetic images look more and more realistic, and speech recognition can often work in noisy environments—but we are still likely decades away from general-purpose, human-level AI that can understand the true meanings of articles and videos or deal with unexpected obstacles and interruptions. The field is stuck on precisely the same challenges that academic scientists (including myself) having been pointing out for years: getting AI to be reliable and getting it to cope with unusual circumstances.

### Dividend Forecasting

Calculate projected equity income and manage portfolio risk.



S&P Global Marketplace



Credit: Bluesguy from NY/Flickr (CC BY-ND 2.0)

Take the recently celebrated Gato, an alleged jack of all trades, and how it captioned an image of a pitcher hurling a baseball (*above*). The system's top three guesses were:

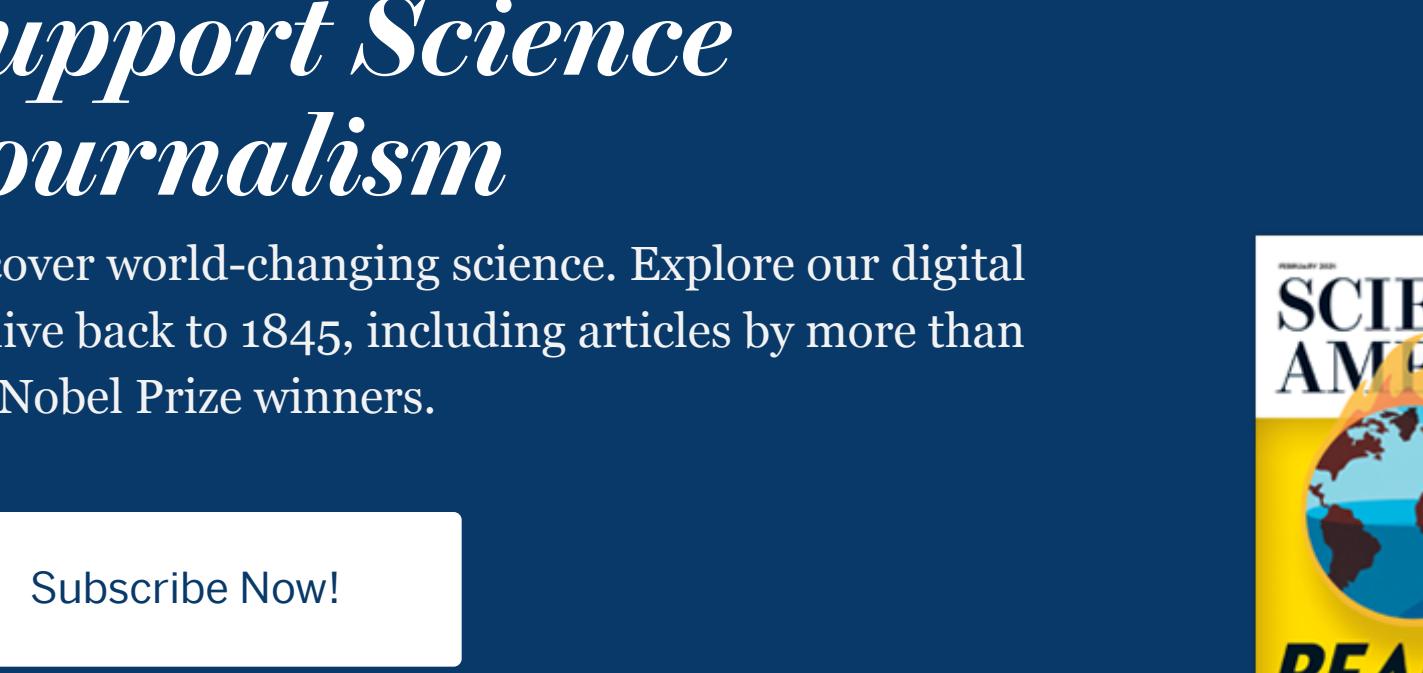
*A baseball player pitching a ball on top of a baseball field.*

*A man throwing a baseball at a pitcher on a baseball field.*

*A baseball player at bat and a catcher in the dirt during a baseball game.*

The first response is correct, but the other two answers include hallucinations of other players that aren't seen in the image. The system has no idea what is actually in the picture, beyond the rough approximations it draws from statistical similarities to other images. Any baseball fan would recognize that this is a pitcher who has just thrown the ball and not the other way around. And although we expect that a catcher and a batter are nearby, they obviously do not appear in the image.

Likewise, DALL-E 2 couldn't tell the difference between an image of a red cube on top of a blue cube versus an image of a blue cube on top of a red cube. A newer system, released this past May, couldn't tell the difference between an astronaut riding a horse and a horse riding an astronaut.



Credit: Imagen

When image-creating systems like DALL-E 2 make mistakes, the result can be amusing. But sometimes errors produced by AI cause serious consequences. A Tesla on autopilot recently drove directly toward a human worker carrying a stop sign in the middle of the road, slowing down only when the human driver intervened. The system could recognize humans on their own (which is how they appeared in the training data) and stop signs in their usual locations (as they appeared in the training images) but failed to slow down when confronted by the unfamiliar combination of the two, which put the stop sign in a new and unusual position.

### FINAL DRAFT 12 Professional Screening Software HOLIDAY SALE SHOP NOW

ADVERTISEMENT

Unfortunately, the fact that these systems still fail to work reliably and struggle with novel circumstances is usually buried in the fine print. Gato, for instance, worked well on all the tasks DeepMind reported but rarely as well as other contemporary systems. GPT-3 often creates fluent prose but struggles with basic arithmetic and has so little grip on reality it is prone to creating sentences such as "Some experts believe that the act of eating a sock helps the brain to come out of its altered state as a result of meditation." A cursory look at recent headlines, however, wouldn't tell you about any of these problems.

The subplot here is that the biggest teams of researchers in AI are no longer to be found in the academy, where peer review was the coin of the realm, but in corporations. And corporations, unlike universities, have no incentive to play fair. Rather than submitting their splashy new papers to academic scrutiny, they have taken to publication by press release, seducing journalists and sidestepping the peer-review process. We know only what the companies want us to know.

In the software industry, there's a word for this kind of strategy: "demonware," software designed to look good for a demo but not necessarily good enough for the real world. Often demonware becomes vaporware, announced for shock and awe to discourage competitors but never released at all.

Chickens do tend to come home to roost, though, eventually. Cold fusion may have sounded great, but you still can't get it at the mall. AI will likely experience a winter of deflated expectations. Too many products, like driverless cars, automated radiologists and all-purpose digital agents, have been demoed, publicized—and never delivered. For now the investment dollars keep coming in on promise (who wouldn't like a self-driving car?). But if the core problems of unreliability and failure to cope with outliers are not resolved, investment will dry up. We may get solid advances in machine translation and speech and object recognition but too little else to show for all the premature hype. Instead of "smart" cities and "democratized" health care, we will be left with destructive deepfakes and energy-sucking networks that emit immense amounts of carbon.

Although deep learning has advanced the ability of machines to recognize patterns in data, it has three major flaws. The patterns that it learns are, ironically, superficial not conceptual; the results it creates are hard to interpret; and the results are difficult to use in the context of other processes, such as memory and reasoning. As Harvard University computer scientist Les Valiant noted, "The central challenge [going forward] is to unify the formulation of ... learning and reasoning." You can't deal with a person carrying a stop sign if you don't really understand what a stop sign even is.

### FINAL DRAFT 12 Professional Screening Software HOLIDAY SALE SHOP NOW

ADVERTISEMENT

For now we are trapped in a "local minimum" in which companies pursue benchmarks rather than foundational ideas. Current engineering practice is far ahead of scientific skills: these departments focus on eking out small improvements with the poorly understood tools they already have rather than developing new technologies with a clearer theoretical ground. This is why basic research remains crucial. That a large part of the AI research community (like those who shout, "Game over") doesn't even see that is, well, heartbreaking.

Imagine if some extraterrestrial studied all human interaction only by looking down at shadows on the ground, noticing, to its credit, that some are bigger than others and that all shadows disappear at night. Maybe it would even notice that the shadows regularly grew and shrank at certain periodic intervals—without ever looking up to see the sun or recognizing the 3-D world above.

It's time for artificial-intelligence researchers to look up from the flashy, straight-to-the-media demos and ask fundamental questions about how to build systems that can learn and reason at the same time.

Rights & Permissions

### FROM OUR ARCHIVES

Will Artificial Intelligence Ever Live Up to Its Hype? John Horgan; ScientificAmerican.com, December 4, 2020.

### READ THIS NEXT

#### ARTS

Poem: 'Mosses'

Jane Hirshfield

#### SPORTS

Men's World Cup Soccer Ball, the Al Rihla, Has the Aerodynamics of a Champion

John Eric Goff and The Conversation US

#### NATURAL DISASTERS

Twitter Chaos Endangers Public Safety, Emergency Managers Warn

Andrea Thompson

#### CLIMATE CHANGE

Warming Worsened West Africa Floods That Killed 800 People

Chelsea Harvey and E&E News

#### NEUROLOGY

This Is How an Alzheimer's Gene Ravages the Brain

Elie Dolgin and Nature magazine

#### CLIMATE CHANGE

How 'Carbon Cage' Blocks Climate Mitigation

Kate Irvine

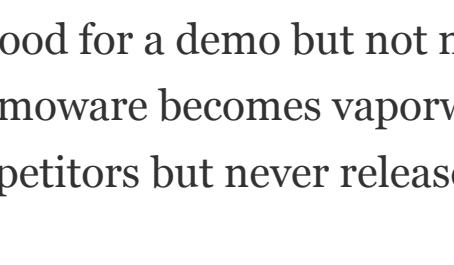
## Support Science Journalism

Discover world-changing science. Explore our digital archive back to 1845, including articles by more than 150 Nobel Prize winners.

Subscribe Now!



FOLLOW US



SCIENTIFIC AMERICAN ARABIC

العربية

Return & Refund Policy

FAQs

Advertise

Privacy Policy

About

Contact Us

SA Custom Media

California Consumer Privacy Statement

Press Room

Site Map

Terms of Use

Use of cookies/Do not sell my data

International Editions

Scientific American is part of Springer Nature, which owns or has commercial relations with thousands of scientific publications (many of them can be found at [www.springernature.com/us](http://www.springernature.com/us)). Scientific American maintains a strict policy of editorial independence in reporting developments in science to our readers.

© 2022 SCIENTIFIC AMERICAN, A DIVISION OF SPRINGER NATURE AMERICA, INC.

ALL RIGHTS RESERVED.

Support science journalism.

See My Options ▾