Tech companies want to build artificial general intelligence. But who decides when AGI is attained?

April 4 2024, by Matt O'brien

Dr. Pei Wang teaches a artificial general intelligence class at Temple University in Philadelphia, Thursday, Feb. 1, 2024. Mainstream AI research "turned away from the original vision of artificial intelligence, which at the beginning was pretty ambitious," said Wang. Credit: AP Photo/Matt Rourke

There's a race underway to build artificial general intelligence, a futuristic vision of machines that are as broadly smart as humans or at least can do many things as well as people can.

Achieving such a concept—commonly referred to as AGI—is the driving mission of ChatGPT-maker OpenAI and a priority for the elite research wings of tech giants Amazon, Google, Meta and Microsoft.

It's also a cause for concern [for world governments](#). Leading AI scientists [published](#) research Thursday in the journal *Science* warning that unchecked AI agents with "long-term planning" skills could pose an existential risk to humanity.

But what exactly is AGI and how will we know when it's been attained? Once on the fringe of computer science, it's now a buzzword that's being constantly redefined by those trying to make it happen.

What is AGI?

Not to be confused with the similar-sounding generative AI—which describes the AI systems behind the crop of tools that "generate" new documents, images and sounds—artificial general intelligence is a more nebulous idea.

It's not a technical term but "a serious, though ill-defined, concept," said Geoffrey Hinton, a pioneering AI scientist who's been dubbed a "Godfather of AI."

"I don't think there is agreement on what the term means," Hinton said by email this week. "I use it to mean AI that is at least as good as humans at nearly all of the cognitive things that humans do."

Hinton prefers a different term—superintelligence—"for AGIs that are

better than humans."

A small group of early proponents of the term AGI were looking to evoke how mid-20th century computer scientists envisioned an intelligent machine. That was before AI research branched into subfields that advanced specialized and commercially viable versions of the technology—from face recognition to speech-recognizing voice assistants like Siri and Alexa.

Mainstream AI research "turned away from the original vision of artificial intelligence, which at the beginning was pretty ambitious," said Pei Wang, a professor who teaches an AGI course at Temple University and helped organize the first AGI conference in 2008.

Putting the 'G' in AGI was a signal to those who "still want to do the big thing. We don't want to build tools. We want to build a thinking machine," Wang said.
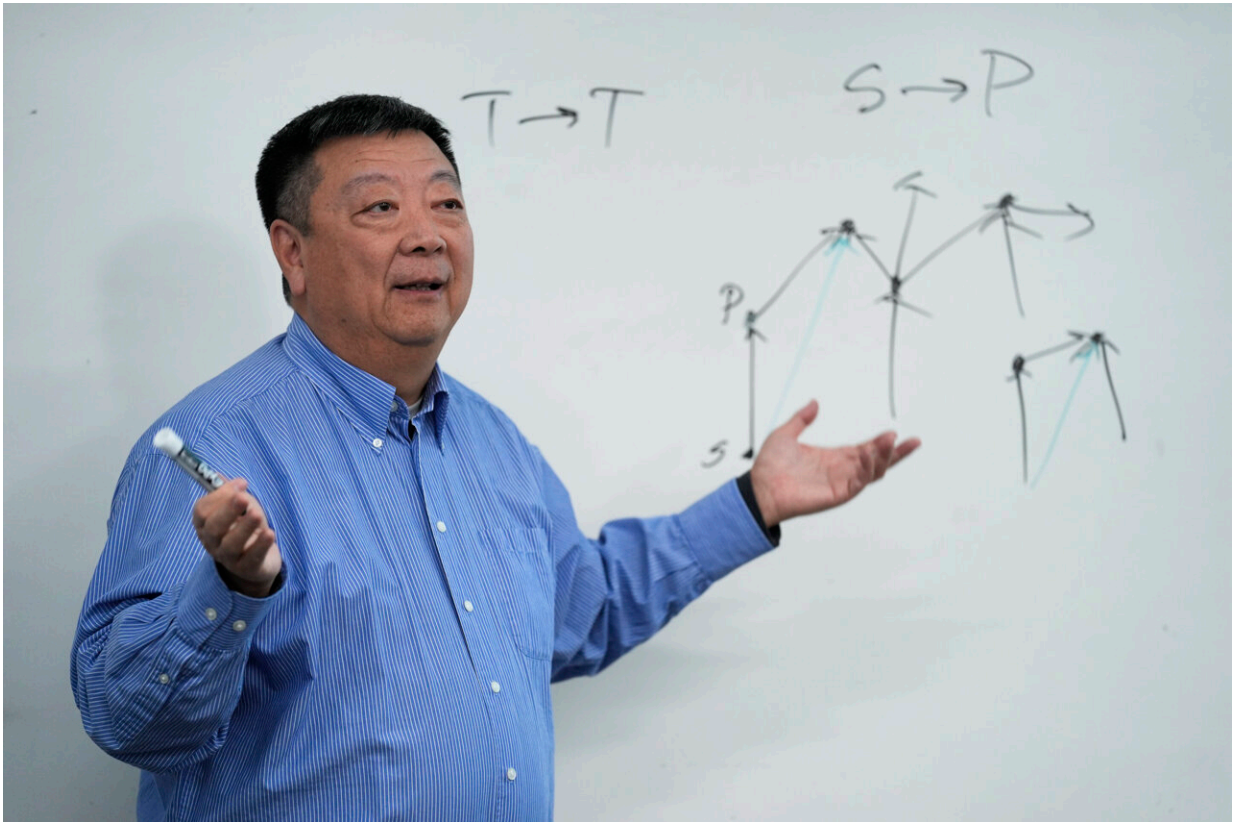
Dr. Pei Wang teaches a artificial general intelligence class at Temple University in Philadelphia, Thursday, Feb. 1, 2024. Mainstream AI research "turned away from the original vision of artificial intelligence, which at the beginning was pretty ambitious," said Wang. Credit: AP Photo/Matt Rourke

Are we at AGI yet?

Without a clear definition, it's hard to know when a company or group of researchers will have achieved artificial general intelligence—or if they already have.

"Twenty years ago, I think people would have happily agreed that systems with the ability of GPT-4 or (Google's) Gemini had achieved general intelligence comparable to that of humans," Hinton said. "Being

able to answer more or less any question in a sensible way would have passed the test. But now that AI can do that, people want to change the test."

Improvements in "autoregressive" AI techniques that predict the most plausible next word in a sequence, combined with massive computing power to train those systems on troves of data, have led to impressive chatbots, but they're still not quite the AGI that many people had in mind. Getting to AGI requires technology that can perform just as well as humans in a wide variety of tasks, including reasoning, planning and the ability to learn from experiences.

Some researchers would like to find consensus on how to measure it. It's one of the topics of an upcoming AGI workshop next month in Vienna, Austria—the first at a major AI research conference.

"This really needs a community's effort and attention so that mutually we can agree on some sort of classifications of AGI," said workshop organizer Jiaxuan You, an assistant professor at the University of Illinois Urbana-Champaign. One idea is to segment it into levels in the same way that carmakers try to benchmark the path between cruise control and fully self-driving vehicles.

Others plan to figure it out on their own. San Francisco company OpenAI has given its nonprofit board of directors—whose members include a former U.S. Treasury secretary—the responsibility of deciding when its AI systems have reached the point at which they "outperform humans at most economically valuable work."

"The board determines when we've attained AGI," says OpenAI's own explanation of its governance structure. Such an achievement would cut off the company's biggest partner, Microsoft, from the rights to commercialize such a system, since the terms of their agreements "only

apply to pre-AGI technology."

Is AGI dangerous?

Hinton made global headlines last year when he quit Google and sounded a warning about AI's existential dangers. A new *Science* study published Thursday could reinforce those concerns.

Its lead author is Michael Cohen, a University of California, Berkeley, researcher who studies the "expected behavior of generally intelligent artificial agents," particularly those competent enough to "present a real threat to us by out planning us."

Cohen made clear in an interview Thursday that such long-term AI planning agents don't yet exist. But "they have the potential" to get more advanced as tech companies seek to combine today's chatbot technology with more deliberate planning skills using a technique known as reinforcement learning.

Pioneering AI scientist Geoffrey Hinton, poses at Google's Mountain View, Calif, headquarters on March 25, 2015. There's a race underway to build artificial general intelligence, nicknamed AGI, a futuristic vision of machines that are broadly as smart as humans. Hinton prefers a different term for AGI — superintelligence — "for AGIs that are better than humans." Credit: AP Photo/Noah Berger, File

"Giving an advanced AI system the objective to maximize its reward and, at some point, withholding reward from it, strongly incentivizes the AI system to take humans out of the loop, if it has the opportunity," according to the paper whose co-authors include prominent AI scientists Yoshua Bengio and Stuart Russell and law professor and former OpenAI adviser Gillian Hadfield.

"I hope we've made the case that people in government (need) to start thinking seriously about exactly what regulations we need to address this problem," Cohen said. For now, "governments only know what these companies decide to tell them."

Too legit to quit AGI?

With so much money riding on the promise of AI advances, it's no surprise that AGI is also becoming a corporate buzzword that sometimes attracts a quasi-religious fervor.

It's divided some of the tech world between those who argue it should be developed slowly and carefully and others—including venture capitalists and rapper MC Hammer—who've declared themselves part of an "accelerationist" camp.

The London-based startup DeepMind, founded in 2010 and now part of Google, was one of the first companies to explicitly set out to develop AGI. OpenAI did the same in 2015 with a safety-focused pledge.

But now it might seem that everyone else is jumping on the bandwagon. Google co-founder Sergey Brin was recently seen hanging out at a California venue called the AGI House. And less than three years after changing its name from Facebook to focus on virtual worlds, Meta Platforms in January revealed that AGI was also on the top of its agenda.

Meta CEO Mark Zuckerberg said his company's long-term goal was "building full general intelligence" that would require advances in reasoning, planning, coding and other cognitive abilities. While Zuckerberg's company has long had researchers focused on those subjects, his attention marked a change in tone.

At Amazon, one sign of the new messaging was when the head scientist

for the voice assistant Alexa switched job titles to become head scientist for AGI.

While not as tangible to Wall Street as generative AI, broadcasting AGI ambitions may help recruit AI talent who have a choice in where they want to work.

In deciding between an "old-school AI institute" or one whose "goal is to build AGI" and has sufficient resources to do so, many would choose the latter, said You, the University of Illinois researcher.

More information: Michael K. Cohen et al, Regulating advanced artificial agents, *Science* (2024). DOI: 10.1126/science.adl0625