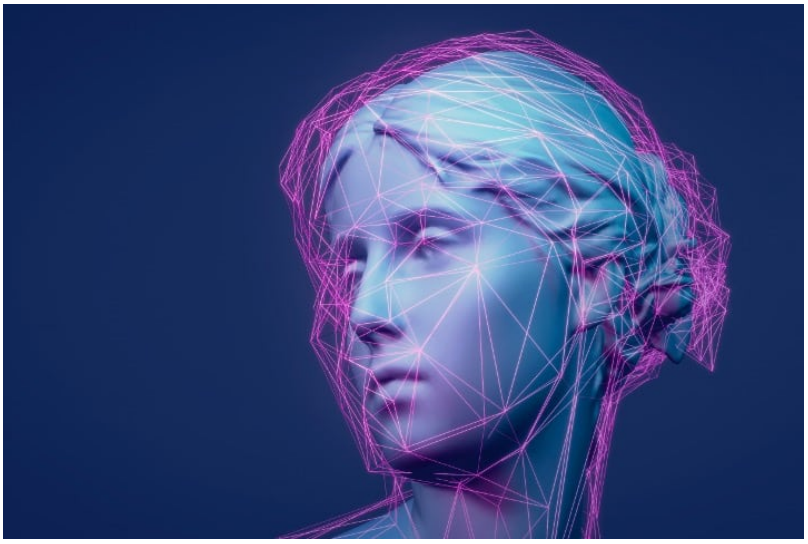
New American

Written by **C. Mitchell Shaw** on May 5, 2023

"Godfather of AI" Quits Google Job, Warns of Danger to Humanity



salihkilic/iStock/Getty Images Plus

In 2012, Dr. Geoffrey Hinton and two of his students at the University of Toronto made breakthroughs in artificial intelligence (AI) that laid the foundation for where AI is today and where it is going in the future. This week, he quit his job at Google so that he has the freedom to openly speak against his creation.

Hinton shares the honorary title "Godfather of AI" with Yoshua Bengio and Yann LeCun, because of their work on deep learning that led to AI as we know it. He also shares the 2018 Turing Award with those men because of the breakthroughs they made in fields related to AI.

But Dr. Hinton also shares some similarities with Dr. Victor Frankenstein — the young scientist character created by English novelist Mary Shelley in her famous 1818 novel. Both Hinton and Frankenstein set out to break new ground. And both wound up creating what they later realized to be monsters they could not control. But where Frankenstein's monster eventually seeks to destroy itself, AI is duplicating at an alarming rate and must be stopped by man, since it will not — in fact, *can not* — stop itself.

After his work at the University of Toronto, Hinton went to work for Google, where he continued to build upon his previous successes. Now, more than 10 years later, he has weighed his life's work in the balance and found it wanting.

As *The New York Times* reported earlier this week:

> Geoffrey Hinton was an artificial intelligence pioneer. In 2012, Dr. Hinton and two of his graduate students at the University of Toronto created technology that became the intellectual foundation for the A.I. systems that the tech industry's biggest companies believe is a key to their future.
>
> On Monday, however, he officially joined a growing chorus of critics who say those companies are racing toward danger with their aggressive campaign to create products based on generative artificial intelligence, the technology that powers popular chatbots like ChatGPT.
>
> Dr. Hinton said he has quit his job at Google, where he has worked for more than a decade and became one of the most respected voices in the field, so he can freely speak out about the risks of A.I. A part of him, he said, now regrets his life's work.

While AI apologists of all stripes claim that AI will revolutionize the world for the better (even

comparing it to the creation of web browsers) and promise that it will lead to breakthroughs in medicine, crime prevention, poverty, psychology, and even death, Hinton and others who have been in the AI trenches for years (in Hinton's case, since the beginning of any practicable form of AI) warn that humanity is prying the lid off of Pandora's box. "It is hard to see how you can prevent the bad actors from using it for bad things," Dr. Hinton told the *Times*.

Indeed.

Hinton seems to have arrived — albeit a little late — to the same conclusion as others who have issued concerned warnings about AI. In fact, while his departure from Google is admirable in its own right, since he quit his job to set himself free to warn the world of the monster he helped create, he does remind this writer of a scene from the original *Jurassic Park* film. In that scene, one of the scientists who has helped clone extinct dinosaurs says that perhaps he and others have spent so much energy trying to figure out whether they *could* that they never stopped to think about whether they *should*.

This is that.

Almost as if to illustrate that point, the *Times* article states:

> Gnawing at many industry insiders is a fear that they are releasing something dangerous into the wild. Generative A.I. can already be a tool for misinformation. Soon, it could be a risk to jobs. Somewhere down the line, tech's biggest worriers say, it could be a risk to humanity.

This comes a month after this magazine issued the same warning in almost the same words, and that article was far from the first warning *The New American* has issued. Following the insight of tech insiders, it is not difficult to predict that AI allowed to run free is the precursor to AI run amok. And AI run amok may make Frankenstein's monster look positively cuddly by comparison.

But — following the model of "doing it because we can" — Google and others have jumped aboard the AI ship and do not appear to have ever looked back. As the *Times* article states:

> In 2012, Dr. Hinton and two of his students in Toronto, Ilya Sutskever and Alex Krishevsky, built a neural network that could analyze thousands of photos and teach itself to identify common objects, such as flowers, dogs and cars.
>
> Google spent $44 million to acquire a company started by Dr. Hinton and his two students. And their system led to the creation of increasingly powerful technologies, including new chatbots like ChatGPT and Google Bard. Mr. Sutskever went on to become chief scientist at OpenAI. In 2018, Dr. Hinton and two other longtime collaborators received the Turing Award, often called "the Nobel Prize of computing," for their work on neural networks.
>
> Around the same time, Google, OpenAI and other companies began building neural networks that learned from huge amounts of digital text. Dr. Hinton thought it was a powerful way for machines to understand and generate language, but it was inferior to the way humans handled language.
>
> Then, last year, as Google and OpenAI built systems using much larger amounts of data, his view changed. He still believed the systems were inferior to the human brain in some ways but he thought they were eclipsing human intelligence in others. "Maybe what is going on in

> these systems," he said, "is actually a lot better than what is going on in the brain."

And that is the problem. Computer programs can be artificially intelligent, but there is no such thing as artificial morality. Programs can "think," but they can't feel. Further (and more importantly), they don't know the difference between right and wrong, though neither — it appears — do many of their "masters."

Hinton came to realize that as AI continues surpassing the human brain in some areas, the danger of AI increases: "Look at how it was five years ago and how it is now. Take the difference and propagate it forwards. That's scary."

And while Hinton says that Google kept a "proper" stewardship of AI until last year, all bets were off once Microsoft began to challenge Google's dominance in the field by using an AI chatbot to crank up Bing results. With Google and Microsoft in a race for the most powerful AI, the restraints have been taken off and AI may soon outrun its previously established boundaries.

Hinton warns that with free-range AI creating content — which the *Times* says includes "false photos, videos and text" — humanity is heading to a place where the average person will "not be able to know what is true anymore."

With the advent of "deep fake" videos and pictures, adding AI's ability into the equation — and AI's ability to create documents, online posts, and other written communications — Hinton's assertion that people may "not be able to know what is true anymore" is not a stretch.

AI is a tool, not unlike electricity. The difference is that electricians understand what electricity can and cannot do. They understand the proper uses and the dangers. But, with AI, even the folks working to turn it loose on the world appear not to understand the dangers. Even Hinton and those like him seem to have arrived late to the party. As Hinton told the *Times* in lamenting his lack of caution in creating this monster, "I console myself with the normal excuse: If I hadn't done it, somebody else would have."

That is meager consolation when the fate of humanity may hang in the balance. Hinton seems to understand that, putting it this way:

> The idea that this stuff could actually get smarter than people — a few people believed that. But most people thought it was way off. And I thought it was way off. I thought it was 30 to 50 years or even longer away. Obviously, I no longer think that.... I don't think they should scale this up more until they have understood whether they can control it.

But "whether they can control it" may be a moot point.

AI in the control of good humans promises great advances for humans. AI either in the control of bad humans or out of control altogether threatens the human race. The question of whether or not good humans can control it begins to look like an academic question without any practical application.

AI appears to be a danger any way you look at it.