"All the News
That's Fit to Print"

The New York Times

'The Godfather of AI'
Leaves Google and Warns
of Danger Ahead By Cade
Metz © 2023 The New
York Times Company

NEW YORK, MAY 1, 2023

$2.50

'The Godfather of AI' Leaves Google and Warns of Danger Ahead

TORONTO — Geoffrey Hinton* was an artificial intelligence pioneer. In 2012, Hinton and two of his graduate students at the University of Toronto created technology that became the intellectual foundation for the AI 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 AI, the technology that powers popular chatbots like ChatGPT.

Hinton said he has quit his job at Google, where he has worked for more than decade and became one of the most respected voices in the field, so he can freely speak out about the risks of AI. A part of him, he said, now regrets his life's work.

"I console myself with the normal excuse: If I hadn't done it, somebody else would have," Hinton said during a lengthy interview last week in the dining room of his home in Toronto, a short walk from where he and his students made their breakthrough.

Hinton's journey from AI groundbreaker to doomsayer marks a remarkable moment for the technology industry at perhaps its most important inflection point in decades. Industry leaders believe the new AI systems could be as important as the introduction of the web browser in the early 1990s and could lead to breakthroughs in areas ranging from drug research to education.

But gnawing at many industry insiders is a fear that they are releasing something dangerous into the wild. Generative AI 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.

"It is hard to see how you can prevent the bad actors from using it for bad things," Hinton said.

After the San Francisco startup OpenAI released a new version of ChatGPT in March, more than 1,000 technology leaders and researchers signed an open letter calling for a six-month moratorium on the development of new systems because AI technologies pose "profound risks to society and humanity."

Several days later, 19 current and former leaders of the Association for the Advancement of Artificial Intelligence, a 40-year-old academic society, released their own letter warning of the risks of AI. That group included Eric Horvitz, chief scientific officer at Microsoft, which has deployed OpenAI's technology across a wide range of products, including its Bing search engine.

Hinton, often called "the Godfather of AI," did not sign either of those letters and said he did not want to publicly criticize Google or other companies until he had quit his job. He notified the company last month that he was resigning, and Thursday, he talked by phone with Sundar Pichai, CEO of Google's parent company, Alphabet. He declined to publicly discuss the details of his conversation with Pichai.

Google's chief scientist, Jeff Dean, said in a statement: "We remain committed to a responsible approach to AI. We're continually learning to understand emerging risks while also innovating boldly."

Hinton, a 75-year-old British expatriate, is a lifelong academic whose career was driven by his personal convictions about the development and use of AI. In 1972, as a graduate student at the University of Edinburgh, Hinton embraced an idea called a neural network. A neural network is a mathematical system that learns skills by analyzing data.

"The idea that this stuff could actually get smarter than people — a few people believed that. But most people thought it was way off. Obviously, I no longer think that."

At the time, few researchers believed in the idea. But it became his life's work.

In the 1980s, Hinton was a professor of computer science at Carnegie Mellon University but left the university for Canada because he said he was reluctant to take Pentagon funding. At the time, most AI research in the United States was funded by the Defense Department. Hinton is deeply opposed to the use of AI on the battlefield — what he calls "robot soldiers."

In 2012, 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 Hinton and his two students. And their system led to the creation of increasingly powerful technologies, including new chatbots such as ChatGPT and Google Bard. Sutskever went on to become chief scientist at OpenAI. In 2018, 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. 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."

As companies improve their AI systems, he believes, they become increasingly dangerous. "Look at how it was five years ago and how it is now," he said of AI technology. "Take the difference and propagate it forwards. That's scary."

Until last year, he said, Google acted as a "proper steward" for the technology, careful not to release something that might cause harm. But now that Microsoft has augmented its Bing search engine with a chatbot — challenging Google's core business — Google is racing to deploy the same kind of technology. The tech giants are locked in a competition that might be impossible to stop, Hinton said.

His immediate concern is that the internet will be flooded with false photos, videos and text, and the average person will "not be able to know what is true anymore."

He is also worried that AI technologies will in time upend the job market. Today, chatbots such as ChatGPT tend to complement human workers, but they could replace paralegals, personal assistants, translators and others who handle rote tasks. "It takes away the drudge work," he said. "It might take away more than that."

Down the road, he is worried that future versions of the technology pose a threat to humanity because they often learn unexpected behavior from the vast amounts of data they analyze. This becomes an issue, he said, as individuals and companies allow AI systems not only to generate their own computer code but actually to run that code on their own. And he fears a day when truly autonomous weapons — those killer robots — become reality.

"The idea that this stuff could actually get smarter than people — a few people believed that," he said. "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."

Many other experts, including many of his students and colleagues, say this threat is hypothetical. But Hinton believes that the race between Google and Microsoft and others will escalate into a global race that will not stop without some sort of global regulation.

But that may be impossible, he said. Unlike with nuclear weapons, he said, there is no way of knowing whether companies or countries are working on the technology in secret. The best hope is for the world's leading scientists to collaborate on ways of controlling the technology. "I don't think they should scale this up more until they have understood whether they can control it," he said.

Hinton said that when people used to ask him how he could work on technology that was potentially dangerous, he would paraphrase Robert Oppenheimer, who led the U.S. effort to build the atomic bomb: "When you see something that is technically sweet, you go ahead and do it."

He does not say that anymore.

—Cade Metz

Geoffrey Hinton* *is the grand-nephew of The Putney School's founder* ***Carmelita Hinton**. He is a pioneer of and, now, outspoken critic of, artificial intelligence. This article was published on May 1, 2023, in the New York Times.*