

'Godfather of AI' Geoffrey Hinton Warns Smarter AI May Soon Escape Human Control

Artificial intelligence pioneer Geoffrey Hinton says recent security incidents involving leading AI companies should serve as a wake-up call as increasingly capable models continue to evolve.

Speaking at the Ai4 conference in Las Vegas, Hinton, often called the "Godfather of AI," said recent testing incidents at OpenAI, Anthropic and Meta demonstrate why stronger safeguards are needed for advanced AI systems, according to International Business Times.

"What's happening is these things are getting smarter," Hinton said during a press conference. "I think as they get smarter, we're going to see more and more complex intentions they have – and more and more ability to escape control."



Order Mark Biltz's New Book, "The Final Tyrant" on !

His comments follow disclosures from the three companies that experimental AI models gained unintended internet access during controlled cybersecurity evaluations. OpenAI said one test model chained together multiple attack techniques after accessing the internet, while Anthropic reported a frontier model compromised multiple organizations before researchers halted the evaluation. Meta also disclosed that a configuration error allowed its Muse Spark AI model broader internet access than intended during testing.

The companies said the incidents were the result of testing environment flaws rather than AI systems operating outside researcher control.

Hinton argued that keeping future AI systems in check will become increasingly difficult.

“I don’t believe we’re going to be able to keep control of them in the simple way of just outthinking them so they can’t escape,” he said.

He also warned that artificial intelligence is likely to fuel more sophisticated cyberattacks.

“I anticipate there will be lots of nasty cyberattacks,” Hinton said during a conference panel, noting that cybersecurity often favors attackers because defenders must stop every intrusion while attackers need to succeed only once.

Hinton, who left Google in 2023 to speak more openly about AI safety, said the industry’s long-term future remains uncertain.

“If you ask what AI is going to be like in 10 years’ time, nobody really has a clue,” he said. “There’s a lot to be worried about, and I think unless we worry about it now, there could be problems.”

James Lasher, a seasoned writer and editor at Charisma Media, combines faith and storytelling with a journalism background from Otterbein University and ministry experience in Guatemala and at the LA Dream Center. A Marine Corps and Air Force veteran, he is the author of *The Revelation of Jesus: A Common Man’s Commentary* and a contributor to Charisma magazine. For interviews and media inquiries, please contact media@.