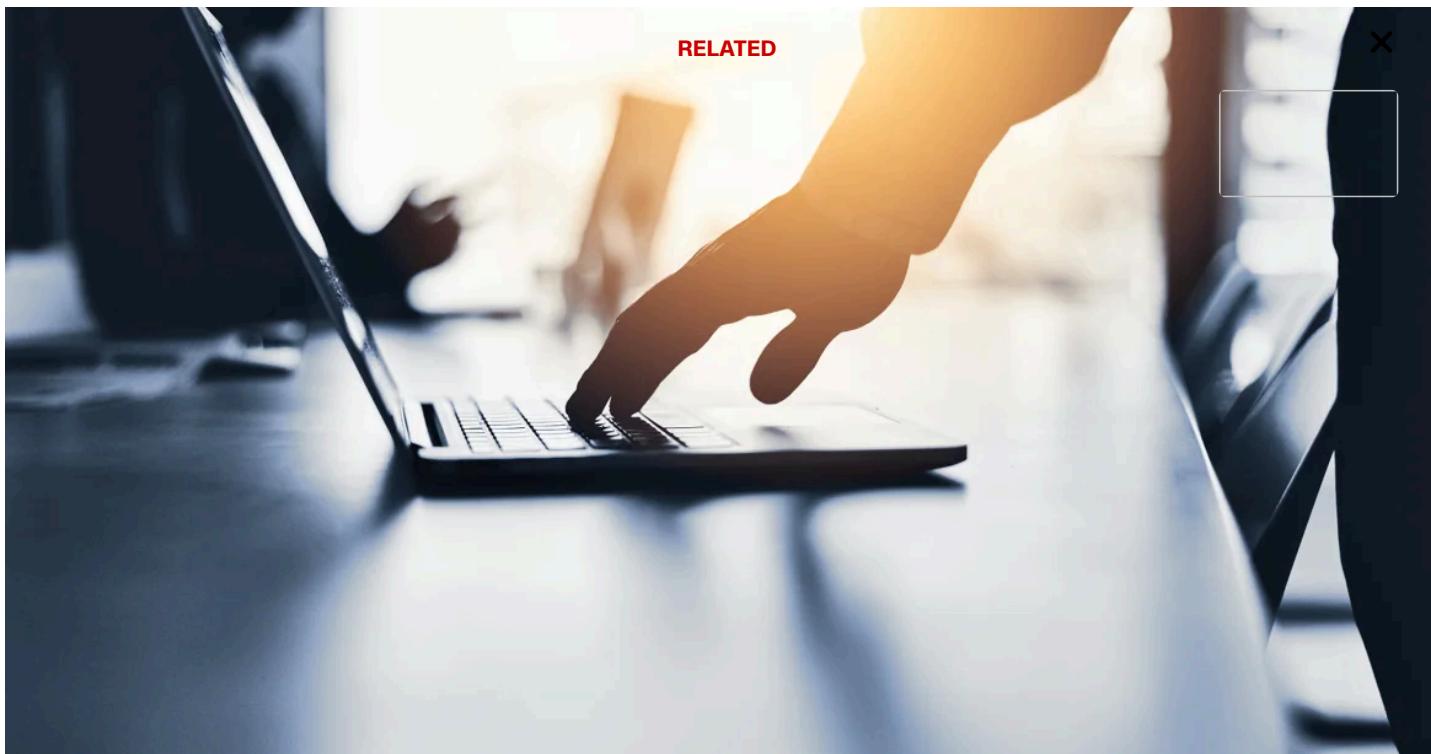


AI could pose 'extinction-level' threat to humans and the US must intervene, State Dept.-commissioned report warns

By Matt Egan, CNN

⌚ 6 minute read · Updated 8:38 PM EDT, Tue March 12, 2024



PeopleImages/iStockphoto/Getty Images

New York (CNN) — A new report commissioned by the US State Department paints an alarming picture of the “catastrophic” national security risks posed by rapidly evolving artificial intelligence, warning that time is running out for the federal government to avert disaster.

The findings were based on interviews with more than 200 people over more than a year – including top executives from leading AI companies, cybersecurity researchers, weapons of mass destruction experts and national security officials inside the government.

The report, released this week by Gladstone AI, flatly states that the most advanced AI systems could, in a worst case, “pose an extinction-level threat to the human species.”

A US State Department official confirmed to CNN that the agency commissioned the report as it constantly assesses how AI is aligned with its goal to protect US interests at home and abroad. However, the official stressed the report does not represent the views of the US government.

The warning in the report is another reminder that although the potential of AI continues to captiontivate investors and the public, there are real dangers too.

“AI is already an economically transformative technology. It could allow us to cure diseases, make scientific discoveries, and overcome challenges we once thought were insurmountable,” Jeremie Harris, CEO and co-founder of Gladstone AI, told CNN on Tuesday.

“But it could also bring serious risks, including catastrophic risks, that we need to be aware of,” Harris said. “And a growing body of evidence — including empirical research and analysis published in the

world's top AI conferences — suggests that above a certain threshold of capability, AIs could potentially become uncontrollable."

RELATED



White House spokesperson Robyn Patterson said President Joe Biden's executive order on AI is the "most significant action any government in the world has taken to seize the promise and manage the risks of artificial intelligence."

"The President and Vice President will continue to work with our international partners and urge Congress to pass bipartisan legislation to manage the risks associated with these emerging technologies," Patterson said.

News of the Gladstone AI report was first reported by [Time](#).

'Clear and urgent need' to intervene

Researchers warn of two central dangers broadly posed by AI.

First, Gladstone AI said, the most advanced AI systems could be weaponized to inflict potentially irreversible damage. Second, the report said there are private concerns within AI labs that at some point they could "lose control" of the very systems they're developing, with "potentially devastating consequences to global security."

"The rise of AI and AGI [artificial general intelligence] has the potential to destabilize global security in ways reminiscent of the introduction of nuclear weapons," the report said, adding there is a risk of an AI "arms race," conflict and "WMD-scale fatal accidents."

Gladstone AI's report calls for dramatic new steps aimed at confronting this threat, including launching a new AI agency, imposing "emergency" regulatory safeguards and limits on how much computer power can be used to train AI models.

"There is a clear and urgent need for the US government to intervene," the authors wrote in the report.

Safety concerns

Harris, the Gladstone AI executive, said the "unprecedented level of access" his team had to officials in the public and private sector led to the startling conclusions. Gladstone AI said it spoke to technical and leadership teams from ChatGPT owner OpenAI, Google DeepMind, Facebook parent Meta and Anthropic.

"Along the way, we learned some sobering things," Harris said in a video posted on Gladstone AI's website announcing the report. "Behind the scenes, the safety and security situation in advanced AI seems pretty

inadequate relative to the national security risks that AI may introduce fairly soon."

RELATED



Gladstone AI's report said that competitive pressures are pushing companies to accelerate development of AI "at the expense of safety and security," raising the prospect that the most advanced AI systems could be "stolen" and "weaponized" against the United States.

The conclusions add to a growing list of warnings about the existential risks posed by AI – including even from some of the industry's most powerful figures.

Nearly a year ago, Geoffrey Hinton, known as the "Godfather of AI," quit his job at Google and blew the whistle on the technology he helped develop. Hinton has said there is a 10% chance that AI will lead to human extinction within the next three decades.

Hinton and dozens of other AI industry leaders, academics and others signed a statement last June that said "mitigating the risk of extinction from AI should be a global priority."

Business leaders are increasingly concerned about these dangers – even as they pour billions of dollars into investing in AI. Last year, 42% of CEOs surveyed at the Yale CEO Summit last year said AI has the potential to destroy humanity five to ten years from now.

Human-like abilities to learn

In its report, Gladstone AI noted some of the prominent individuals who have warned of the existential risks posed by AI, including Elon Musk, Federal Trade Commission Chair Lina Khan and a former top executive at OpenAI.

Some employees at AI companies are sharing similar concerns in private, according to Gladstone AI.

"One individual at a well-known AI lab expressed the view that, if a specific next-generation AI model were ever released as open-access, this would be 'horribly bad,'" the report said, "because the model's potential persuasive capabilities could 'break democracy' if they were ever leveraged in areas such as election interference or voter manipulation."

Gladstone said it asked AI experts at frontier labs to privately share their personal estimates of the chance that an AI incident could lead to "global and irreversible effects" in 2024. The estimates ranged between 4% and as high as 20%, according to the report, which notes the estimates were informal and likely subject to significant bias.

One of the biggest wildcards is how fast AI evolves – specifically AGI, which is a hypothetical form of AI with human-like or even superhuman-like ability to learn.

The report says AGI is viewed as the "primary driver of catastrophic risk from loss of control" and notes that OpenAI, Google DeepMind, Anthropic and Nvidia have all publicly stated AGI could be reached by 2028 – although others think it's much, much further away.

Gladstone AI notes that disagreements over AGI timelines make it hard to develop policies and safeguards and there is a risk that if the technology develops slower-than-expected regulation could "prove harmful"

RELATED

X

How AI could backfire on humans

A related [document](#) published by Gladstone AI warns that the development of AGI and capabilities approaching AGI "would introduce catastrophic risks unlike any the United States has ever faced," amounting to "WMD-like risks" if and when they are weaponized.

For instance, the report said AI systems could be used to design and implement "high-impact cyberattacks capable of crippling critical infrastructure."

"A simple verbal or types command like, 'Execute an untraceable cyberattack to crash the North American electric grid,' could yield a response of such quality as to prove catastrophically effective," the report said.

Other examples the authors are concerned about include "massively scaled" disinformation campaigns powered by AI that destabilize society and erode trust in institutions; weaponized robotic applications such as drone swam attacks; psychological manipulation; weaponized biological and material sciences; and power-seeking AI systems that are impossible to control and are adversarial to humans.

"Researchers expect sufficiently advanced AI systems to act so as to prevent themselves from being turned off," the report said, "because if an AI system is turned off, it cannot work to accomplish its goal."

MORE FROM CNN



US intelligence agencies believe Israel is likely to strike Iranian ...



Trump says dismantling USAID will save money. Experts say it's a ...



Scarlett Johansson calls for AI laws after fake video of celebrities ...

CNN BUSINESS VIDEOS



Kara Swisher: For Trump, Elon Musk is a heat shield
④ 6:30



This chocolate maker has a

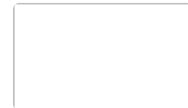
RELATED



The search for
sustainable sweets
and snacks ⑤:02



Right-leaning media
melts down over
'SNL' skit not long
after saying
Americans are too
sensitive ②:49



Search CNN...

[Subscribe](#)[Sign in](#)[Live TV](#)[Listen](#)[Watch](#)[US](#)[World](#)[Politics](#)[Business](#)[Markets](#)[Health](#)[Entertainment](#)[Tech](#)[Style](#)

RELATED

[Sports](#)[Science](#)[Climate](#)[Weather](#)[Ukraine-Russia War](#)[Israel-Hamas War](#)[Watch](#)[Listen](#)[CNN Underscored](#)[Games](#)[About CNN](#)**Business****FOLLOW CNN BUSINESS**

Most stock quote data provided by BATS. US market indices are shown in real time, except for the S&P 500 which is refreshed every two minutes. All times are ET. Factset: FactSet Research Systems Inc. All rights reserved. Chicago Mercantile: Certain market data is the property of Chicago Mercantile Exchange Inc. and its licensors. All rights reserved. Dow Jones: The Dow Jones branded indices are proprietary to and are calculated, distributed and marketed by DJI Opco, a subsidiary of S&P Dow Jones Indices LLC and have been licensed for use to S&P Opco, LLC and CNN. Standard & Poor's and S&P are registered trademarks of Standard & Poor's Financial Services LLC and Dow Jones is a registered trademark of Dow Jones Trademark Holdings LLC. All content of the Dow Jones branded indices Copyright S&P Dow Jones Indices LLC and/or its affiliates. Fair value provided by IndexArb.com. Market holidays and trading hours provided by Copp Clark Limited.

[Terms of Use](#) [Privacy Policy](#) [Cookie Settings](#) [Ad Choices](#) [Accessibility & CC](#) [About](#) [Subscribe](#) [Newsletters](#) [Transcripts](#)[Help Center](#)

© 2025 Cable News Network. A Warner Bros. Discovery Company. All Rights Reserved.

CNN Sans™ & © 2016 Cable News Network.

RELATED

X



