

Tech giants form AI group focused on ensuring safety

July 26 2023



The development of artificial intelligence has come with fears that the risks of the technology will outweigh its benefits.

Four US leaders in artificial intelligence (AI) announced Wednesday the formation of an industry group devoted to addressing risks that cutting

edge versions of the technology may pose.

Anthropic, Google, Microsoft, and ChatGPT-maker OpenAI said the newly created Frontier Model Forum will draw on the expertise of its members to minimize AI risks and support industry standards.

The companies pledged to share [best practices](#) with each other, lawmakers and researchers.

"Frontier" models refer to nascent, large-scale machine-learning platforms that take AI to new levels of sophistication—and also have capabilities that could be dangerous.

"Companies creating AI technology have a responsibility to ensure that it is safe, secure, and remains under human control," Microsoft president Brad Smith said in a statement.

"This initiative is a vital step to bring the [tech sector](#) together in advancing AI responsibly and tackling the challenges so that it benefits all of humanity."

US President Joe Biden evoked AI's "enormous" risks and promises at a White House meeting last week with tech leaders who committed to guarding against everything from cyberattacks to fraud as the sector grows.

Standing alongside top representatives from Amazon, Anthropic, Google, Inflection, Meta, Microsoft and OpenAI, Biden said the companies had made commitments to "guide responsible innovation" as AI spreads ever deeper into personal and business life.

Ahead of the meeting, the seven AI giants committed to a series of self-regulated safeguards that the White House said would "underscore three

principles that must be fundamental to the future of AI: safety, security and trust."

In their pledge, the companies agreed to develop "robust technical mechanisms," such as watermarking systems, to ensure users know when content is from AI and not humans.

Core objectives of the Frontier Model Forum include minimizing risks and enabling independent safety evaluations of AI platforms, the companies involved said in a release.

The Forum will also support the development of applications intended to take on challenges such as [climate change](#), [cancer prevention](#) and cyber threats, according to its creators.

Others pursuing AI breakthroughs were invited to join the group.

"Advanced AI technologies have the potential to profoundly benefit society, and the ability to achieve this potential requires oversight and governance," said OpenAI vice president of global affairs Anna Makanju.

"It is vital that AI companies—especially those working on the most powerful models—align on common ground and advance thoughtful and adaptable safety practices."

© 2023 AFP

Citation: Tech giants form AI group focused on ensuring safety (2023, July 26) retrieved 6 May 2024 from <https://techxplore.com/news/2023-07-tech-giants-ai-group-focused.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is

provided for information purposes only.