

OpenAI disbands team devoted to artificial intelligence risks

May 18 2024



OpenAI CEO Sam Altman promises to share more in the days ahead about what the ChatGPT maker is doing to keep its artificial intelligence technology safe.

OpenAI on Friday confirmed that it has disbanded a team devoted to mitigating the long-term dangers of super-smart artificial intelligence.

OpenAI weeks ago began dissolving the so-called "superalignment" group, integrating members into other projects and research, according to the San Francisco-based firm.

Company co-founder Ilya Sutskever and team co-leader Jan Leike announced their departures from the ChatGPT-maker this week.

The dismantling of an OpenAI team focused on keeping sophisticated artificial intelligence under control comes as such technology faces increased scrutiny from regulators and fears mount regarding its dangers.

"OpenAI must become a safety-first AGI ([artificial general intelligence](#)) company," Leike wrote Friday in a post on X, formerly Twitter.

Leike called on all OpenAI employees to "act with the gravitas" warranted by what they are building.

OpenAI chief executive Sam Altman responded to Leike's post with one of his own, thanking him for his work at the company and saying he was sad to see Leike leave.

"He's right we have a lot more to do," Altman said. "We are committed to doing it."

Altman promised more on the topic in the coming days.

Sutskever said on X that he was leaving after almost a decade at OpenAI, whose "trajectory has been nothing short of miraculous."

"I'm confident that OpenAI will build AGI that is both safe and beneficial," he added, referring to computer technology that seeks to perform as well as—or better than—[human cognition](#).

Sutskever, OpenAI's chief scientist, sat on the board that voted to remove fellow chief executive Altman in November last year.

The ousting threw the San Francisco-based startup into a tumult, with the OpenAI board hiring Altman back a few days later after staff and investors rebelled.

OpenAI early this week released a higher-performing and even more human-like version of the artificial intelligence technology that underpins ChatGPT, making it free to all users.

"It feels like AI from the movies," Altman said in a blog post.

Altman has previously pointed to the Scarlett Johansson character in the movie "Her," where she voices an AI-based virtual assistant dating a man, as an inspiration for where he would like AI interactions to go.

The day will come when "digital brains will become as good and even better than our own," Sutskever said during a talk at a TED AI summit in San Francisco late last year.

"AGI will have a dramatic impact on every area of life."

© 2024 AFP

Citation: OpenAI disbands team devoted to artificial intelligence risks (2024, May 18) retrieved 21 June 2024 from

<https://techxplore.com/news/2024-05-openai-disbands-team-devoted-artificial.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.