

OpenAI says AI is 'safe enough' as scandals raise concerns

May 22 2024, by Julie JAMMOT



OpenAI CEO Sam Altman insisted that OpenAI had put in 'a huge amount of work' to ensure the safety of its models.

OpenAI CEO Sam Altman defended his company's AI technology as safe for widespread use, as concerns mount over potential risks and lack

of proper safeguards for ChatGPT-style AI systems.

Altman's remarks came at a Microsoft event in Seattle, where he spoke to developers just as a new controversy erupted over an OpenAI AI [voice](#) that closely resembled that of the actress Scarlett Johansson.

The CEO, who rose to global prominence after OpenAI released ChatGPT in 2022, is also grappling with questions about the safety of the company's AI following the departure of the team responsible for mitigating long-term AI risks.

"My biggest piece of advice is this is a special time and take advantage of it," Altman told the audience of developers seeking to build new products using OpenAI's technology.

"This is not the time to delay what you're planning to do or wait for the next thing," he added.

OpenAI is a close partner of Microsoft and provides the foundational technology, primarily the GPT-4 large language model, for building AI tools.

Microsoft has jumped on the AI bandwagon, pushing out new products and urging users to embrace generative AI's capabilities.

"We kind of take for granted" that GPT-4, while "far from perfect...is generally considered robust enough and safe enough for a wide variety of uses," Altman said.

Altman insisted that OpenAI had put in "a huge amount of work" to ensure the safety of its models.

"When you take a medicine, you want to know what's going to be safe,

and with our model, you want to know it's going to be robust to behave the way you want it to," he added.

However, questions about OpenAI's commitment to safety resurfaced last week when the company dissolved its "superalignment" group, a team dedicated to mitigating the long-term dangers of AI.

In announcing his departure, team co-leader Jan Leike criticized OpenAI for prioritizing "shiny new products" over safety in a series of posts on X (formerly Twitter).

"Over the past few months, my team has been sailing against the wind," Leike said.

"These problems are quite hard to get right, and I am concerned we aren't on a trajectory to get there."

This controversy was swiftly followed by a public statement from Johansson, who expressed outrage over a voice used by OpenAI's ChatGPT that sounded similar to her voice in the 2013 film "Her."

The voice in question, called "Sky," was featured last week in the release of OpenAI's more human-like GPT-4o model.

In a short statement on Tuesday, Altman apologized to Johansson but insisted the voice was not based on hers.

© 2024 AFP

Citation: OpenAI says AI is 'safe enough' as scandals raise concerns (2024, May 22) retrieved 15 August 2024 from <https://techxplore.com/news/2024-05-openai-ai-safe-scandals.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.