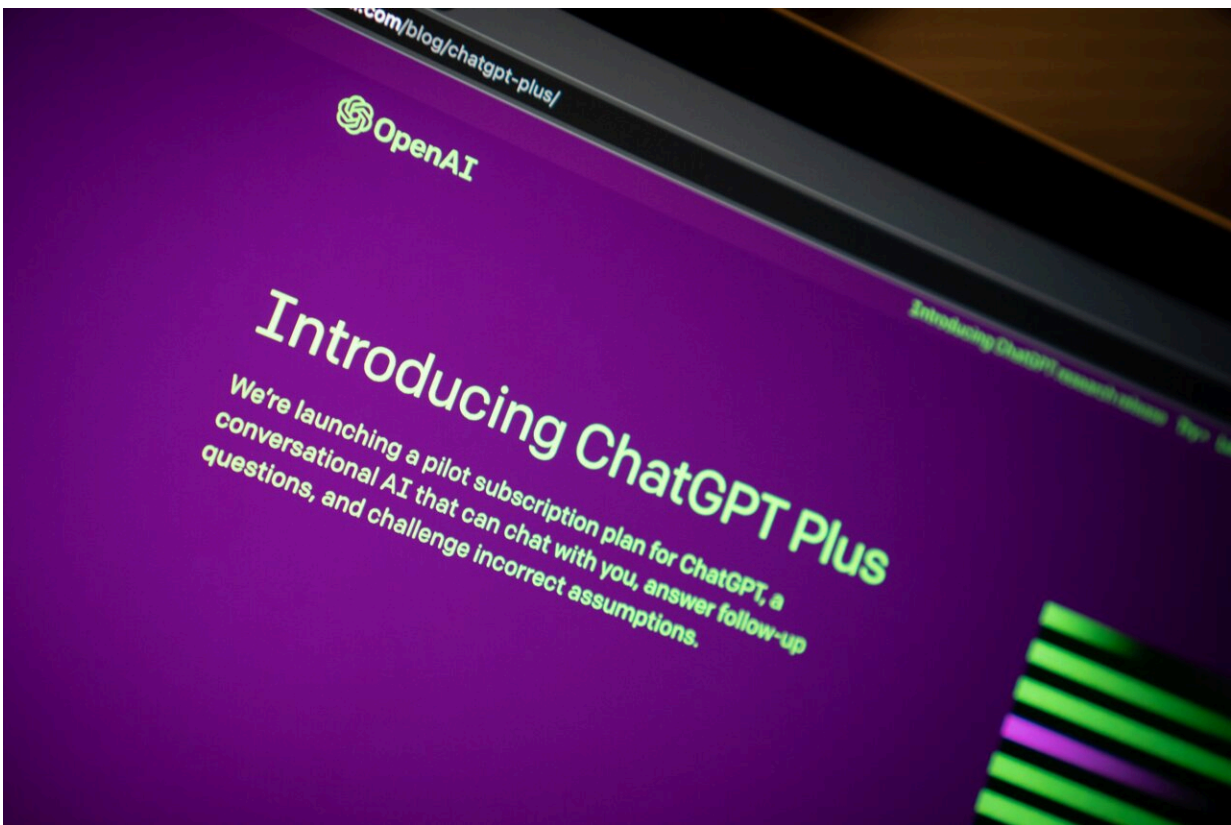


OpenAI and Anthropic to share AI models with US government

August 29 2024



Credit: Unsplash/CC0 Public Domain

Leading generative AI developers OpenAI and Anthropic have agreed to give the US government access to their new models for safety testing as part of agreements announced on Thursday.

The agreements were made with the US AI Safety Institute, which is part of the National Institute of Standards and Technology (NIST), a federal agency.

Regulation of AI has been a major concern since the advent of OpenAI's ChatGPT, with tech companies pushing for a voluntary approach to opening their technology to government oversight.

The agency said it would provide feedback to both companies on potential safety improvements to their models before and after their public release, working closely with its counterpart at the UK AI Safety Institute.

"These agreements are just the start, but they are an important milestone as we work to help responsibly steward the future of AI," said Elizabeth Kelly, director of the US AI Safety Institute.

The agency said the evaluations would aim to support the voluntary commitments made by leading AI model developers, such as OpenAI and Anthropic, as they innovate.

"Our collaboration with the US AI Safety Institute leverages their wide expertise to rigorously test our models before widespread deployment," said Jack Clark, co-founder and head of policy at Anthropic.

"This strengthens our ability to identify and mitigate risks, advancing responsible AI development," he added.

The collaboration is part of work connected to a White House executive order on AI announced in 2023 that was designed to provide a legal backdrop for the rapid deployment of AI models in the United States.

Washington is eager to leave tech companies a free rein to innovate and

experiment with AI, in contrast to the European Union where lawmakers passed an ambitious AI Act to more closely regulate the industry.

But pro-regulation lawmakers in California, the state home to Silicon Valley, on Wednesday pushed through a state AI safety bill that now needs to be signed by the governor to pass into law.

In a social media post welcoming his company's agreement with the US government, OpenAI CEO Sam Altman said it was "important" that regulation take place at the national level.

This was a veiled criticism of the state law passed in California, which includes penalties for any violations, that OpenAI opposed, arguing it will stifle research and innovation.

© 2024 AFP

Citation: OpenAI and Anthropic to share AI models with US government (2024, August 29)
retrieved 29 August 2024 from <https://techxplore.com/news/2024-08-openai-anthropic-ai.html>

<p>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.</p>
--