

Amazon joins generative AI race, targets tech at cloud customers

April 13 2023, by Matt Day



Credit: Pixabay/CC0 Public Domain

Amazon.com Inc. is joining Microsoft Corp. and Google in the generative artificial intelligence race, announcing technology aimed at its cloud customers as well as a marketplace for AI tools from other companies.

The e-commerce giant's Amazon Web Services unit on Thursday plans to announce two of its own large-language models, one designed to generate text, and another that could help power web search personalization, among other things. Amazon announced no plans to release a chatbot like the ones Microsoft and Google have debuted to mixed reviews.

Amazon's large-language models, called Titan, were trained on vast amounts of text to summarize content, write a draft of a blog post or engage in open-ended question-and-answer sessions. They'll be made available on an AWS marketplace, called Bedrock, where developers can tap into models built by other companies plugging away at generative AI, including AI21 Labs, Anthropic and Stability AI.

Generative AI, software that can create text, images, or video based on prompts from a user, has captured the imagination of Silicon Valley, setting off a fierce competition to capitalize on the technology. Proponents of chatbots like ChatGPT and image-generation tools such as Dall-E believe generative AI will revolutionize the kinds of tasks performed by software.

Microsoft, through a partnership with ChatGPT maker OpenAI, has integrated generative AI technology into its Bing internet search service and plans to deploy those tools across the software maker's products. Alphabet Inc.'s Google is racing to make similar moves. Meta Platforms Inc. has released its own large-language [model](#) and said similar work will expand across the company.

AWS, which sells on-demand computing power and software tools—including a suite of [machine-learning](#) applications—had previously partnered with artificial intelligence companies including Hugging Face Inc. and Stability AI, which builds the image generator Stable Diffusion. But the company hadn't previously revealed plans to release a homegrown large-language model.

Swami Sivasubramanian, AWS's vice president of databases, machine learning and analytics, said Amazon had long been working on large-language models. They're already used to help shoppers find products on Amazon's retail website and to power elements of the Alexa voice assistant, among other applications.

"Amazon has been investing in this space for quite a while," Sivasubramanian said in an interview.

During a preview period that begins Thursday, AWS customers can apply to use the models. Sivasubramanian said the company hadn't settled on pricing to access the tools but said homegrown chips built by AWS, including Inferentia2 and Trainium, could help customers keep costs low as they do their own machine-learning work.

The Seattle-based [company](#) on Thursday also said that CodeWhisperer, which uses predictive tools to proactively suggest code as developers type it, would be free for individual developers.

"I don't believe there is going to be one model that will rule the world," Sivasubramanian said.

2023 Bloomberg L.P.

Distributed by Tribune Content Agency, LLC.

Citation: Amazon joins generative AI race, targets tech at cloud customers (2023, April 13)

retrieved 10 May 2024 from <https://techxplore.com/news/2023-04-amazon-generative-ai-tech-cloud.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.