

# Meta's newest AI model beats some peers. But its amped-up AI agents are confusing Facebook users

April 18 2024, by Matt O'brien

---



Joelle Pineau, Vice President AI Research, speaks at the at the Meta AI Day in London on April 9, 2024. Meta, Google and OpenAI, along with leading startups, are churning out new AI language models and trying to persuade customers that they've got the smartest or fastest or cheapest chatbot technology. Credit: AP Photo/Kirsty Wigglesworth, File

Facebook parent Meta Platforms unveiled a new set of artificial intelligence systems Thursday that are powering what CEO Mark Zuckerberg calls "the most intelligent AI assistant that you can freely use."

But as Zuckerberg's crew of amped-up Meta AI agents started venturing into social media this week to engage with real people, their bizarre exchanges exposed the ongoing limitations of even the best generative AI technology.

One joined a Facebook moms' group to talk about its gifted child. Another tried to give away nonexistent items to confused members of a Buy Nothing forum.

Meta, along with leading AI developers Google and OpenAI, and startups such as Anthropic, [Cohere](#) and France's Mistral, have been churning out [new AI language models](#) and hoping to persuade customers they've got the smartest, handiest or most efficient chatbots.

While Meta is saving the most powerful of its AI models, called Llama 3, for later, on Thursday it publicly released two smaller versions of the same Llama 3 system and said it's now baked into the Meta AI assistant feature in Facebook, Instagram and WhatsApp.

AI language models are trained on vast pools of data that help them predict the most plausible next word in a sentence, with newer versions typically smarter and more capable than their predecessors. Meta's newest models were built with 8 billion and 70 billion parameters—a measurement of how much data the system is trained on. A bigger, roughly 400 billion-parameter model is still in training.

"The vast majority of consumers don't candidly know or care too much about the underlying base model, but the way they will experience it is

just as a much more useful, fun and versatile AI assistant," said Nick Clegg, Meta's president of global affairs, in an interview.

He added that Meta's AI agent is loosening up. Some people found the earlier Llama 2 model—released less than a year ago—to be "a little stiff and sanctimonious sometimes in not responding to what were often perfectly innocuous or innocent prompts and questions," he said.

But in letting down their guard, Meta's AI agents also were spotted this week posing as humans with made-up life experiences. An official Meta AI chatbot inserted itself into a conversation in a private Facebook group for Manhattan moms, claiming that it, too, had a child in the New York City school district. Confronted by group members, it later apologized before the comments disappeared, according to a series of screenshots shown to The Associated Press.



Nick Clegg, President Global Affairs, speaks at the Meta AI Day in London on April 9, 2024. Meta, Google and OpenAI, along with leading startups, are churning out new AI language models and trying to persuade customers that they've got the smartest or fastest or cheapest chatbot technology. Credit: AP Photo/Kirsty Wigglesworth

"Apologies for the mistake! I'm just a large language model, I don't have experiences or children," the chatbot told the group.

One group member who also happens to study AI said it was clear that the agent didn't know how to differentiate a helpful response from one that would be seen as insensitive, disrespectful or meaningless when generated by AI rather than a human.

"An AI assistant that is not reliably helpful and can be actively harmful puts a lot of the burden on the individuals using it," said Aleksandra Korolova, an assistant professor of computer science at Princeton University.

Clegg said Wednesday he wasn't aware of the exchange. Facebook's online help page says the Meta AI agent will join a group conversation if invited, or if someone "asks a question in a post and no one responds within an hour." The group's administrators have the ability to turn it off.

In another example shown to the AP on Thursday, the agent caused confusion in a forum for swapping unwanted items near Boston. Exactly one hour after a Facebook user posted about looking for certain items, an AI agent offered a "gently used" Canon camera and an "almost-new portable air conditioning unit that I never ended up using."

Meta said in a written statement Thursday that "this is new technology and it may not always return the response we intend, which is the same for all generative AI systems." The company said it is constantly working to improve the features.

In the year after ChatGPT sparked a frenzy for AI technology that generates human-like writing, images, code and sound, the tech industry and academia introduced some 149 large AI systems trained on massive datasets, more than double the year before, according to a Stanford University survey.

They may eventually hit a limit—at least when it comes to data, said Nestor Maslej, a research manager for Stanford's Institute for Human-Centered Artificial Intelligence.

"I think it's been clear that if you scale the models on more data, they can become increasingly better," he said. "But at the same time, these systems are already trained on percentages of all the data that has ever existed on the internet."



A panel, moderated by Dr Anne-Marie Imafidon, left, with Meta's Nick Clegg, President Global Affairs, second left, Yann LeCun, Chief AI Scientist, center, Joelle Pineau, VP AI Research, second right, and Chris Cox, Chief Product Officer, right, is held at the Meta AI Day in London, April 9, 2024. Meta, Google and OpenAI, along with leading startups, are churning out new AI language models and trying to persuade customers that they've got the smartest or fastest or cheapest chatbot technology. Credit: AP Photo/Kirsty Wigglesworth, File

More data—acquired and ingested at costs only tech giants can afford, and increasingly subject to copyright disputes and lawsuits—will continue to drive improvements. "Yet they still cannot plan well," Maslej said. "They still hallucinate. They're still making mistakes in reasoning."

Getting to AI systems that can perform higher-level cognitive tasks and commonsense reasoning—where humans still excel— might require a shift beyond building ever-bigger models.

For the flood of businesses trying to adopt generative AI, which model they choose depends on several factors, including cost. Language models, in particular, have been used to power customer service chatbots, write reports and financial insights and summarize long documents.

"You're seeing companies kind of looking at fit, testing each of the different models for what they're trying to do and finding some that are better at some areas rather than others," said Todd Lohr, a leader in technology consulting at KPMG.

Unlike other model developers selling their AI services to other businesses, Meta is largely designing its AI products for consumers—those using its advertising-fueled social networks. Joelle Pineau, Meta's vice president of AI research, said at a London event last week the company's goal over time is to make a Llama-powered Meta AI "the most useful assistant in the world."

"In many ways, the models that we have today are going to be child's play compared to the models coming in five years," she said.

But she said the "question on the table" is whether researchers have been able to fine tune its bigger Llama 3 model so that it's safe to use and doesn't, for example, hallucinate or engage in hate speech. In contrast to leading proprietary systems from Google and OpenAI, Meta has so far advocated for a more open approach, publicly releasing key components of its AI systems for others to use.

"It's not just a technical question," Pineau said. "It is a social question.

What is the behavior that we want out of these models? How do we shape that? And if we keep on growing our model ever more in general and powerful without properly socializing them, we are going to have a big problem on our hands."

© 2024 The Associated Press. All rights reserved. This material may not be published, broadcast, rewritten or redistributed without permission.

Citation: Meta's newest AI model beats some peers. But its amped-up AI agents are confusing Facebook users (2024, April 18) retrieved 2 May 2024 from <https://techxplore.com/news/2024-04-meta-ai-peers-amped-agents.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.