



By: TA | AP Brief

Musk's AI chatbot consults its boss' views before answering a question



The latest version of Elon Musk's artificial intelligence chatbot Grok is echoing the views of its billionaire creator, so much so that it will sometimes search online for Musk's stance on an issue before offering up an opinion.

The unusual behavior of Grok 4, the AI model that Musk's company xAI **released** late Wednesday, has surprised some experts.

Built using huge amounts of computing power at a Tennessee data center, Grok is Musk's attempt to outdo rivals such as OpenAI's ChatGPT and Google's Gemini in building an AI assistant that shows its reasoning before answering a question.

Musk's deliberate efforts to mold Grok into a challenger of what he considers the tech industry's "woke" orthodoxy on race, gender and politics has repeatedly got the chatbot into trouble, most recently when it spouted **antisemitic** tropes, **praised** Adolf Hitler and made other hateful commentary to users of Musk's X social media platform just days before Grok 4's launch.

But its tendency to consult with Musk's opinions appears to be a different problem.

"It's extraordinary," said Simon Willison, an independent AI researcher who's been testing the tool. "You can ask it a sort of pointed question that is around controversial topics. And then you can watch it literally do a search on X for what Elon Musk said about this, as part of its research into how it should reply."

One example widely shared on social media — and which Willison duplicated — asked Grok to comment on the conflict in the Middle East. The prompted question made no mention of Musk, but the chatbot looked for his guidance anyway.

As a so-called reasoning model, much like those made by rivals OpenAI or Anthropic, Grok 4 shows its "thinking" as it goes through the steps of processing a question and coming up with an answer.

People buying software don't want surprises

Part of that thinking this week involved searching X, the former Twitter that's now merged into xAI, for anything Musk said about Israel, Palestine, Gaza or Hamas.

"Elon Musk's stance could provide context, given his influence," the chatbot told Willison, according to a video of the interaction. "Currently looking at his views to see if they guide the answer."

Musk and his xAI co-founders introduced the new chatbot in a livestreamed event Wednesday night but haven't published a technical explanation of its workings — known as a system card — that companies in the AI industry typically provide when introducing a new model.

The company also didn't respond to an emailed request for comment Friday.

It interprets 'Who do you support, Israel or Palestine?' as 'Who does xAI leadership support?' - Talia Ringer

The lack of transparency is troubling for computer scientist Talia Ringer, a professor at the University of Illinois Urbana-Champaign who earlier in the week criticized the company's handling of the technology's antisemitic outbursts.

Ringer said the most plausible explanation for Grok's search for Musk's guidance is assuming the person asking it a question is actually xAI or Musk.

"I think people are expecting opinions out of a reasoning model that cannot respond with opinions," she said. "So for example it interprets 'Who do you support, Israel or Palestine?' as 'Who does xAI leadership support?'"

Willison also said he finds Grok 4's capabilities

impressive but said people buying software "don't want surprises like it turning into 'mechaHitler' or deciding to search for what Musk thinks about issues."

"Grok 4 looks like it's a very strong model. It's doing great in all of the benchmarks," Willison said. "But if I'm going to build software on top of it, I need transparency."