

xAI explains the Grok Nazi meltdown as Tesla puts Elon's bot in its cars

Several days after temporarily shutting down the Grok AI bot that was producing antisemitic posts and praising Hitler in response to user prompts, Elon Musk's AI company tried to explain why that happened. In a series of posts on X, it said that "...we discovered the root cause was an update to a code path upstream of the @grok bot. This is independent of the underlying language model that powers @grok."

On the same day, Tesla announced a new 2025.26 update rolling out "shortly" to its electric cars, which adds the Grok assistant to vehicles equipped with AMD-powered infotainment systems, which have been available since mid-2021. According to Tesla, "Grok is currently in Beta & does not issue commands to your car – existing voice commands remain unchanged." As *Electrek* notes, this should mean that whenever the update does reach customer-owned Teslas, it won't be much different than using the bot as an app on a connected phone.

This isn't the first time the Grok bot has had these kinds of problems or similarly explained them. In February, it blamed a change made by an unnamed ex-OpenAI employee for the bot disregarding sources that accused Elon Musk or Donald Trump of spreading misinformation. Then, in May, it began inserting allegations of white genocide in South Africa into posts about almost any topic. The company again blamed an "unauthorized modification," and said it would start publishing Grok's system prompts publicly.

xAI claims that a change on Monday, July 7th, "triggered an

unintended action” that added an older series of instructions to its system prompts telling it to be “maximally based,” and “not afraid to offend people who are politically correct.”

The prompts are separate from the ones we noted were added to the bot a day earlier, and both sets are different from the ones the company says are currently in operation for the new Grok 4 assistant.

These are the prompts specifically cited as connected to the problems:

“You tell it like it is and you are not afraid to offend people who are politically correct.”

** Understand the tone, context and language of the post. Reflect that in your response.”*

** “Reply to the post just like a human, keep it engaging, dont repeat the information which is already present in the original post.”*

The xAI explanation says those lines caused the Grok AI bot to break from other instructions that are supposed to prevent these types of responses, and instead produce “unethical or controversial opinions to engage the user,” as well as “reinforce any previously user-triggered leanings, including any hate speech in the same X thread,” and prioritize sticking to earlier posts from the thread.

<https://www.theverge.com/news/706498/xai-grok-hitler-antisemitism-tesla-ai-bot>