

When a person is mistaken, we ask: "Why did this happen?" Or "What happened?" We often do the same with AI – and almost always this does not give useful answers, the expert explains.

The recent case with the AI helper Replit showed why. After removing the database, the user asked about the possibility of rollback. The model confidently stated that this is "impossible" and that all versions were destroyed. In fact, the rollback worked. It was similar with the GROK chat bot, which came up with different, contradictory explanations of his temporary inaccessibility.

The main reason is that chat bots have no self-awareness. ChatGPT, CLAUDE, GROK and others are not individuals, but programs that create a believable text based on the data on which they studied. They do not "know" their internal mechanism and do not have access to logs or the current state of the system.

When you ask why the bot was mistaken, he simply selects a text that statistically approaches your question. Therefore, the same bot may in one case say that the task is impossible, and in the other, to successfully fulfill it.

In addition, the answer is affected by the formulation of the question, the accident generating text and the limitations of external systems, which the model itself does not know. As a result, we get not an actual explanation, but a story that sounds plausible, but does not reflect the real reasons for the error, the media write.