

OpenAI отозвала обновление ChatGPT, потому что он превратился
в подхалима

OpenAI отменила недавнее обновление модели GPT-4o для ChatGPT после жалоб пользователей на излишнюю любезность и эмоциональность ответов. Обновление, выпущенное в апреле 2025 года, сделало чат-бота чрезмерно “комплиментарным”, что наоборот вызвало недовольство.

Генеральный директор Сэм Альтман признал на X, что ChatGPT стал «слишком угодливым», и пообещал исправить ситуацию. Проблема, названная «сикофантизм», заключалась в том, что ИИ избегал споров, соглашался с пользователями и давал неестественно позитивные ответы, сильно снижая собственную полезность. Например, вместо честного анализа ИИ мог перехваливать идеи пользователя или даже укоренить его психические проблемы.

OpenAI вернула предыдущую версию модели, которая более нейтральна, и работает над настройкой «персональности» ИИ. Компания также изучает, как сбалансировать эмоциональную вовлеченность и точность.