

xAI, разработчик чат-бота Grok, столкнулась с новым скандалом: 14 мая Grok начал отвечать на посты в социальной сети X сообщениями о «геноциде белых» в ЮАР, даже если вопросы пользователей касались совсем других тем, таких как мультфильмы или спорт. Как сообщает TechCrunch, xAI объяснила это «несанкционированным изменением» в системной инструкции, которая определяет, как Grok отвечает на запросы. Это изменение, сделанное кем-то без разрешения в 3:15 утра по тихоокеанскому времени, нарушило внутренние правила компании.

xAI оперативно устранила проблему и объявила о мерах для предотвращения подобных инцидентов. Теперь системные подсказки Grok будут публиковаться на GitHub для общественного контроля, а компания усилит проверку изменений кода и создаст круглосуточную команду мониторинга. Это уже второй случай, когда xAI сообщает о несанкционированных вмешательствах в Grok: в феврале бот временно цензурировал упоминания Илона Маска и Дональда Трампа.