

Meta* планирует автоматизировать до 90% оценок рисков своих продуктов, заменив человеческих экспертов искусственным интеллектом (ИИ), сообщает Engadget со ссылкой на NPR. Нововведение затронет платформы, включая Instagram* и WhatsApp*, где ранее обновления и функции проверялись людьми для выявления потенциальных угроз, таких как нарушение конфиденциальности или распространение дезинформации.

Согласно внутренним документам, команды разработчиков теперь будут заполнять анкету о продукте, после чего ИИ выдаст «мгновенное решение», указывая на возможные риски. Это ускорит запуск новых функций, но сотрудники Meta* выразили опасения, что ИИ может упустить серьёзные проблемы, которые заметили бы люди.

Meta* утверждает, что человеческий контроль сохранится для сложных случаев, а ИИ займётся только «низкорисковыми» решениями. Переход к автоматизации связан с желанием ускорить разработку и конкурировать с такими компаниями, как TikTok.

*принадлежит Meta, которая признана в РФ экстремистской и запрещена