

Экс-сотрудники Google DeepMind потребовали разрешать предупреждать о рисках ИИ

Письмо, озаглавленное «Право предупреждать о передовом искусственном интеллекте», подписали 13 человек, в том числе бывшие сотрудники OpenAI Джейкоб Хилтон, Даниэль Кокотайло и другие, а также бывшие сотрудники Google DeepMind Рамана Кумар и Нил Нанда.

Они утверждают, что ИИ несет в себе серьезные риски, такие как укоренение неравенства, распространение дезинформации и потенциальное вымирание человечества. Они призывают компании, занимающиеся разработкой ИИ, поддерживать культуру «открытой критики», создавать процессы анонимной отчетности и воздерживаться от репрессий в отношении сотрудников, раскрывающих информацию, связанную с рисками.

Письмо последовало за спорами вокруг соглашений OpenAI о неразглашении информации, которые компания с тех пор пересмотрела, убрав из них пункты о неразглашении информации.