

«Мы уже израсходовали суммарные знания человечества для обучения ИИ. Это произошло, по сути, в прошлом году», — отметил Маск в недавнем стриме на платформе X.com.

Глава xAI также подчеркнул, что будущее технологий связано с синтетическими данными — искусственно сгенерированной информацией, создаваемой самими моделями ИИ.

Эта точка зрения совпадает с прогнозами Ильи Суцкевера, экс-главы исследовательского отдела OpenAI, который заявил о достижении «пика данных» на конференции NeurIPS в декабре.

Суцкевер также предсказал, что нехватка данных вынудит индустрию изменить подходы к разработке моделей.

Ведущие компании, такие как Microsoft, OpenAI и Anthropic, уже активно используют синтетические данные.

Например, модели Microsoft Phi-4 и Google Gemma были обучены с их применением. Однако, несмотря на экономические преимущества (например, Writer создала модель Palmyra X 004 за \$700 тыс против \$4.6 млн для аналогичной модели OpenAI), синтетические данные могут привести к «коллапсу модели», снижая её креативность и увеличивая предвзятость.