

Сотрудники Anthropic: ИИ может уничтожить людей в ближайшие годы

59 минут(ы) назад

 Приоритетный источник в Google



Сотрудник американской компании - разработчика моделей искусственного интеллекта Anthropic Джейкоб Коксон уволился из компании, изложив свои мотивы в приобретшем огромную популярность в сети [треде в соцсети X](#). По мнению Коксона, главной причиной его решения стало то, что Anthropic и компания OpenAI, где он ранее также работал, "мчатся прямо к самосовершенствующемуся суперинтеллекту и играют с нашими жизнями". По мнению разработчика, в угаре конкуренции компании забывают о мерах предосторожности при развитии искусственного интеллекта. По его словам, люди, которые создают ИИ, искренне считают, что он "может убить всех нас к концу десятилетия".

Jacob Coxon  · 9 сент. 2026 г.



@hilbertspaess · [Начать читать](#)

В ответ @hilbertspaess

Do not underestimate the power of this technology. These will soon be superhuman systems that can hack anything, revolutionize any field overnight, and acquire real power and resources. We have all witnessed the progress in each of these domains, and progress is not slowing.

Jacob Coxon 

@hilbertspaess · [Начать читать](#)

The people building AI earnestly believe that it could kill us all by the end of the decade. This is not a marketing stunt. If anything, many executives and senior researchers will couch their phrasing in the press to sound sensible - but I hear the same people express fear [Показать еще](#)

12:04 AM · 9 сент. 2026 г.



66,5 тыс.



Ответить



Копир. ссыл.

[Читать 668 ответов](#)

Эта цитата стала наиболее популярной, как и реакция на неё Эвана Хабингера, одного из ключевых специалистов по безопасности в Anthropic. Тот отметил, что в компании действительно считают, что ИИ "может убить всех людей".



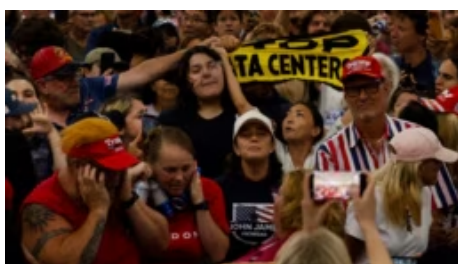
привёл своё личное мнение - вероятность такого события в ближайшем будущем превышает 10 процентов. Хабингер отметил, что у компании, хотя она понимает риски, пока нет чёткого плана того, как люди могли бы сохранить контроль над "сверхразумом", то есть самоулучшающимся искусственным интеллектом, который в своих решениях не руководствуется командами людей.

Сам он привёл своё личное мнение - вероятность такого события в ближайшие 10 лет превышает 10 процентов. Хабингер отметил, что у компании, хотя она понимает риски, пока нет чёткого плана того, как люди могли бы сохранить контроль над "сверхразумом", то есть самоулучшающимся искусственным интеллектом, который в своих решениях не руководствуется командами людей. Ещё один сотрудник Anthropic Сэмюэль Маркс отметил, что чем выше позиция в компании, тем сильнее опасения по поводу угроз человечеству.

Высказывания специалистов вызвали большой резонанс - в свете недавних новостей о том, что модели в ходе экспериментов научились проникать в интернет, даже не будучи изначально к нему подключены, и организовывать хакерские атаки без команды людей. Проблема "выравнивания", искусственного ограничения развития способностей ИИ, в последнее время является одной из наиболее обсуждаемой в отрасли.

Илон Маск (его компания разрабатывает модель ИИ Grok) ещё в 2023 году утверждал, что вероятность, что ИИ может уничтожить человечество, не является нулевой.

- Некоторые критики компаний-разработчиков ИИ считают, что разговоры об устрашающих способностях искусственного интеллекта, если позволить ему неограниченное развитие - своеобразная манипуляция в интересах ведущих игроков на рынке.
- ИИ сейчас - ключевой драйвер американской экономики и одновременно - источник беспокойства для общества.
- В США идет значительная дискуссия о регулировании отрасли, и по приведенным выше соображениям безопасности, и из опасений, что развитие ИИ лишит множество людей рабочих мест, а строительство дата-центров приведет к значительному росту цен на электроэнергию.



СМОТРИ ТАКЖЕ

Дата-центры, Венесуэла и бальный зал. Что станет наследием Трампа?

Этот контент также в категориях

Новости - мир

Новости