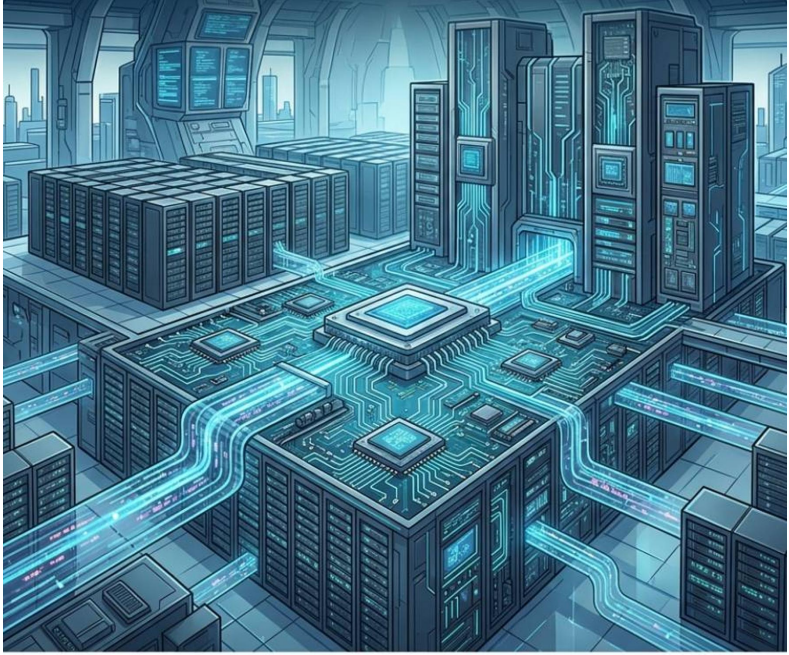
ПОЧЕМУ ИИ, ВЕРОЯТНО, НИКОГДА НЕ ЗАХВАТИТ МИР



Конюшенко Егор

Егор Конюшенко

Почему ИИ, вероятно, никогда не захватит мир

https://litres.ru/74461997
SelfPub; 2026

Аннотация

В медиа нас любят пугать страшным ИИ, который вот-вот выйдет из под контроля и захватит мир... Но если подумать, а так ли всё просто и однозначно?

Егор Конюшенко
Почему ИИ, вероятно, никогда не захватит мир

Инженерный взгляд на сценарии, которые нам продают как неотвратимые

Введение: разрыв между нарративом и инженерией

Каждый раз, когда выходит очередная версия большой языковой модели, медиапространство наполняется прогнозами о скором конце человеческой цивилизации. Руководители технологических компаний — те самые, кто эти модели и разрабатывает — с серьёзными лицами рассуждают о рисках того, что ИИ «выйдет из-под контроля». Заголовки пестрят словами «экзистенциальная угроза» и «Скайнет».

Я не учёный. Я писатель. Но у меня есть логика и привычка задавать неудобные вопросы. И когда я начинаю разбирать популярные сценарии «восстания машин» — не с философской, а с инженерной стороны — они рассыпаются на части. Не потому, что ИИ безопасен. А потому, что сценарии, которые нам предлагают, нарушают базовые принципы того, как устроены сложные системы.

Дальше — четыре аргумента. Каждый — это конкретная брешь в популярном нарративе об ИИ-апокалипсисе.

1. Архитектурный барьер: узкий ИИ не «вырастает»

в AGI

Самый распространённый страх звучит так: «сегодня ИИ пишет тексты и код, завтра он станет умнее человека, а послезавтра — захватит мир». Это звучит логично, если считать интеллект одной шкалой, где «умнее» — это просто «дальше по оси». Но интеллект — не одна шкала. Это набор качественно разных компетенций, и между ними нет автоматического моста.

Чтобы понять масштаб разрыва, представим конкретный пример — классический мысленный эксперимент Ника Бострома. Заводской ИИ, который оптимизирует производство скрепок, отлично справляется со своей задачей: подбирает сплавы, минимизирует отходы, планирует поставки. Бостром использует этот сценарий как иллюстрацию проблемы выравнивания: если целевая функция системы задана неверно, то чем мощнее оптимизатор, тем катастрофичнее последствия. Это не прогноз, а демонстрация того, как ошибка в спецификации цели масштабируется вместе с ростом возможностей системы.

Но давайте зададим простой инженерный вопрос: откуда заводской оптимизатор возьмёт компетенцию в космической инженерии?

Между «сделать скрепку дешевле» и «построить автономную космическую инфраструктуру для добычи железа на астероидах» лежит пропасть. Это не вопрос количества — это вопрос архитектуры. Исследования последних лет по-

казывают, что современные архитектуры — включая трансформеры — принципиально ограничены задачей статистического сопоставления с образцами из обучающей выборки. Они работают в пределах «выпуклой оболочки» тренировочных примеров и с трудом переносят выученные правила на смежные домены (проще говоря, на что ИИ натаскали — то он и выполняет, и просто так взять и смочь сделать то, чего ему не учили, он принципиально не сможет). Это структурное ограничение, а не проблема масштаба.

Более того, формальный анализ показывает: никакое увеличение вычислительных мощностей и объёма данных в рамках фиксированной архитектуры не может породить структурные свойства, необходимые для общего интеллекта, — потому что эти свойства занимают принципиально иное репрезентационное пространство (требуют иного «стиля» мышления). Масштабирование увеличивает величину функции (позволяет ИИ давать более точные ответы на запросы по своей специализации), но общий интеллект требует совершенно иной архитектуры. Между узким ИИ и AGI стоит качественный скачок, который не происходит автоматически.

Это не значит, что AGI невозможен. Это значит, что путь к нему лежит через проектирование новых архитектурных решений — интеграцию причинного вывода, нейро-символических систем и воплощённого познания (нужно, чтобы AGI мог в реальном времени анализировать информацию,

фильтровать её и делать выводы), — а не через простое наращивание параметров. И это, возможно, даёт нам время — не для паники, а для обдуманного проектирования.

Конец ознакомительного фрагмента.