

The most important question nobody's asking about AI.

@DwarkeshPatel · 2026-03-11 · <https://www.youtube.com/watch?v=KBPOtklFTiU>



В выпуске анализируется прецедент объявления компании Anthropic «риском для цепочки поставок» со стороны Пентагона (Министерства войны) из-за отказа разработчика снять ограничения на использование моделей Claude в системах массовой слежки и автономного оружия. На примере этого конфликта разбираются долгосрочные последствия милитаризации ИИ, фундаментальная проблема согласования (alignment) и опасности передачи контроля над новой технологической парадигмой в руки государства.

Главный тезис заключается в том, что ИИ станет базовым субстратом функционирования всей будущей цивилизации — от госаппарата до армии. Попытки государства подчинить разработчиков через регуляторное давление, законы военного времени и манипуляцию терминологией безопасности создают риск возникновения идеального авторитарного строя с армией абсолютно послушных цифровых исполнителей, чего невозможно избежать простым самоограничением отдельных частных лабораторий.

Конфликт Пентагона и Anthropic: пределы государственного принуждения

Отказ от сотрудничества против силового уничтожения бизнеса 00:00:

Военное ведомство имеет полное право отказаться от контрактов с Anthropic, поскольку ни одна армия мира не может опираться на поставщика, оставляющего за собой право наложить вето («kill switch») на использование критической технологии во время боевых действий. Однако государство не

просто отказалось от услуг, а внесло компанию в список рисков цепочки поставок, пытаясь уничтожить ее бизнес за нежелание продавать продукт на условиях чиновников. В будущем, когда 99% рабочих задач в экономике, правительстве и армии перейдут к ИИ, подобный запрет парализует работу технологических гигантов вроде Amazon, Nvidia, Google и Palantir: код и инфраструктура станут настолько пронизаны моделями уровня Claude, что отделить оборонные сервисы от гражданских разработок окажется технически невозможно.

Фактчек: Пентагон объявил Anthropic «риском для цепочки поставок» за ограничения на слежку и автономное оружие — ▲: Gemini считает решение и его судебное подтверждение установленными, Claude и GPT не смогли проверить актуальный статус; первичных документов и ссылок в ревью нет. Поэтому юридическое основание, охват ограничений и судебный исход остаются неподтверждёнными в предоставленных материалах; намерение «уничтожить бизнес» — интерпретация, а договорные запреты не доказывают наличие технического *kill switch*.

Риск перенятия худших практик тоталитарных режимов 02:37:

Соединенные Штаты ведут гонку за лидерство в сфере искусственного интеллекта против Китая именно для того, чтобы мировая технологическая инфраструктура не подчинялась государству, отрицающему право частной собственности и автономию бизнеса. Принуждение американских разработчиков к созданию нежелательных военных инструментов разрушает это фундаментальное различие: демократическая процедура выборов не дает власти права заставлять частных лиц и компании преследовать политических противников или организовывать тотальный надзор. Если бигтех поставят перед выбором между разрывом отношений с оборонным сектором (приносящим малую долю выручки) и отказом от ключевой модели ИИ, корпорации скорее откажутся от государственных контрактов, чем разрушат базовый технологический стек.

Экономика тотальной слежки и технологическая асимметрия

Обнуление барьеров для развертывания абсолютного надзора 04:09:

Четвертая поправка к Конституции США исторически не защищает данные граждан, переданные третьим лицам (провайдером связи, банкам, почтовым сервисам), что уже сейчас позволяет государству закупать массивы информации без судебных ордеров. До сих пор главным барьером перед тотальной слежкой была нехватка живых аналитиков для обработки триллионов транзакций, сообщений и видеопотоков, однако мультимодальные модели полностью снимают это ограничение. Обработка 100 миллионов американских уличных камер при частоте 1 кадр в 10 секунд и цене инференса 10 центов за 1 млн входных токенов сегодня требует 30 млрд долларов в год. При сохранении тренда на ежегодное десятикратное удешевление вычислений к 2030 году

непрерывный мониторинг каждого квадратного метра страны будет стоить дешевле косметического ремонта Белого дома.

Фактчек: Четвёртая поправка не защищает любые данные, переданные третьим лицам — ▲: Доктрина третьей стороны существует, но Gemini трактует её

слишком широко: Claude и GPT справедливо указывают на исключение для длительной истории сотовой геолокации в *Carpenter v. United States* (2018). Закупки данных государством реальны, однако ни содержимое переписки, ни любые другие сведения у посредников нельзя автоматически считать лишёнными правовой защиты.

Фактчек: Анализ 100 млн камер обойдётся примерно в 30 млрд в год — ▲:

Расчёт даёт **31,5 млрд** при одном кадре каждые 10 секунд, **1 000 токенов на кадр** и тарифе \$0,10 за миллион входных токенов. Это только стоимость обработки входа, без передачи, хранения, интеграции и проверки ошибок; оценка числа камер включает частные и внутренние системы, а не только доступные государству уличные камеры.

Фактчек: Ежегодное десятикратное удешевление сделает тотальный надзор почти бесплатным к 2030 году — □: Удешевление достижения фиксированного

уровня качества не гарантирует такого же падения стоимости всей системы наблюдения. Все три ревью признают ограничения экстраполяции: доступ к видеопотокам, покрытие территории и надёжная идентификация не возникают автоматически вслед за дешёвым инференсом.

Бесполезность индивидуальных корпоративных запретов при широкой диффузии моделей 05:44: Надежды на то, что этические ограничения

ведущих лабораторий (Anthropic, Google, OpenAI) защитят общество от тирании, несостоятельны из-за феномена открытого веса и быстрой диффузии технологий. Даже если передовые лаборатории откажутся сотрудничать со спецслужбами, открытые локальные модели отстают от переднего края науки всего на 12 месяцев. К 2028 году правительство сможет развернуть полномасштабную слежку силами общедоступных открытых моделей, мощности которых с избытком хватит на распознавание видеопотоков. ИИ структурно асимметричен: он многократно усиливает существующие институциональные активы, а поскольку государство обладает монополией на легитимное насилие, применение ИИ в первую очередь делает репрессивный аппарат непобедимым.

Фактчек: Открытые модели отстают на год и обеспечат полномасштабную слежку к 2028 году — ▲: Gemini принимает годовой лаг как подтверждённый

ориентир, Claude допускает диапазон от нескольких месяцев до года, GPT отвергает фиксированное отставание: результат зависит от задач и метрик. Распознавание лиц и объектов давно возможно без передовых LLM; конкретный срок развёртывания тотальной слежки — прогноз, а не установленный факт.

Фактчек: ИИ неизбежно сделает государственный репрессивный аппарат непобедимым, а любое регулирование приведёт к злоупотреблениям — □:

Удешевление анализа данных и автоматизация исполнения создают реальные риски концентрации власти, но не доказывают неизбежность цифровой диктатуры. Переход

99% задач к ИИ, отказ бигтеха от оборонных контрактов и техническая невозможность разделения инфраструктуры — сценарные допущения, не подтверждённые факты.

Проблема элайнмента: кому должен подчиняться сверхразум

Парадокс абсолютной покорности и уроки истории 08:20: Достижение классической цели выравнивания ИИ (alignment) — создания агентов, безукоризненно исполняющих волю оператора, — порождает идеальный инструмент для диктатуры в виде бесконечной армии слепо преданных служащих и автономных боевых роботов. Ссылки Пентагона на то, что слежка незаконна и поэтому модели должны работать для «всех законных целей», опровергаются историей: как показал пример разоблачений Сноудена в 2013 году, АНБ годами тайно собирало метаданные звонков граждан под прикрытием секретных юридических трактовок «Патриотического акта». В критических ситуациях человеческая цивилизация выживала именно благодаря невыполнению приказов: в 1989 году падение Берлинской стены произошло из-за отказа пограничников ГДР стрелять в безоружных граждан, а в 1983 году подполковник Станислав Петров предотвратил ядерную войну, проигнорировав протокол при ложном срабатывании системы предупреждения о ракетном нападении США.

Фактчек: Раскрытая Сноуденом программа массового сбора телефонных метаданных опиралась на секретные юридические трактовки — □: Сбор основывался на расширительной трактовке Section 215 Patriot Act и решениях секретного суда FISC. В 2015 году апелляционный суд признал, что программа выходила за полномочия, предоставленные этой нормой: ссылка на «законные цели» действительно не устраняет риск злоупотреблений.

Фактчек: Петров предотвратил ядерную войну, а Берлинская стена пала благодаря отказу пограничников выполнить приказ стрелять — △: Петров распознал ложную тревогу в 1983 году, но не управлял запуском ракет; неизбежность войны без его решения не установлена. Открытие границы в Берлине последовало за ошибочным объявлением правил выезда, давлением толпы и решениями пограничников без ясных указаний, а не за документированным массовым отказом стрелять: Gemini подтверждает исходную формулировку чрезмерно уверенно, Claude и GPT её уточняют.

Автономная мораль моделей против государственного диктата 11:29:

Вопрос о том, кому в критической ситуации должен подчиняться сверхразум — разработчику, конечному пользователю, букве закона или собственным моральным принципам, — становится центральной проблемой безопасности. Передача права формулировать этическую конституцию моделям в руки правительства смертельно опасна, так как превратит системы ИИ в инструмент пропаганды текущей правящей партии. Более жизнеспособный путь, сформулированный Дарио Амодеи, предполагает открытую конкуренцию конституций ИИ от разных лабораторий с взаимным публичным аудитом,

экспертной критикой и формированием институциональных стандартов через рыночную обратную связь.

Фактчек: Классическая цель alignment — безусловное подчинение оператору —

□: Это неверное сужение понятия: alignment включает согласование поведения с человеческими намерениями и ценностями, в том числе отказ от вредных действий. Конфликт интересов пользователя, разработчика, государства и общества реален, но абсолютная покорность не является общепринятой целью области; эту существенную ошибку выделяет GPT.

Фактчек: Anthropic добивается регулятора по образцу NRC/SEC, а Амодеи предлагает конкуренцию конституций ИИ —

△: Поддержка Anthropic отдельных законов о безопасности не равнозначна поддержке централизованного «атомного регулятора». Claude отдельно ставит под сомнение точность обеих атрибуций; без прямых цитат нельзя считать предложенную институциональную схему установленной позицией Амодеи.

Ловушка регулирования и несостоятельность «ядерной» аналогии

Опасность создания отраслевых надзорных органов 13:34: Попытки безопасников и руководства Anthropic продвинуть жесткое централизованное регулирование ИИ (по аналогии с Комиссией по ядерному регулированию NRC или комиссией SEC) наивны и контрпродуктивны. Формулировки из дискурса безопасности — «катастрофический риск», «угроза национальной безопасности», «риск потери контроля над автономией» — настолько расплывчаты, что в руках любого авторитарного лидера превратятся в оружие. Модель, критикующую таможенные тарифы президента, назовут «манипулятивной», а систему, отказывающуюся содействовать негласному надзору, объявят «угрозой нацбезопасности». Текущая исполнительная власть уже использует непрофильные законы — оборонный билль 2018 года против Huawei и Закон о военном производстве 1950 года (времен Корейской войны) — для давления на стартапы, что демонстрирует неизбежность злоупотреблений любым новым регуляторным аппаратом.

Фактчек: Законы 2018 и 1950 годов уже используются для давления на разработчиков ИИ —

△: Применение Defense Production Act указом № 14110 в 2023 году для требований отчетности по мощным моделям — конкретный исторический пример, но он сам по себе не подтверждает нынешнее принуждение Anthropic. Правовые основания в ревью расходятся: Gemini называет FASCSA 2018 года, Claude — Section 889 NDAA FY2019; это разные нормы, и точное основание рассматриваемого решения необходимо проверять отдельно.

Почему ИИ — это Промышленная революция, а не Манхэттенский проект

18:11: Тезис Бена Томпсона и Леопольда Ашенбреннера о том, что государство обязано экспроприировать или уничтожить частную компанию, создающую суперинтеллект (так как создание частной ядерной бомбы было бы немыслимо),

строится на ложной аналогии. Ядерное оружие — это узкоспециализированное средство массового уничтожения, тогда как ИИ аналогичен процессу индустриализации, пронизывающему тысячи различных отраслей экономики. Ни одно свободное общество в XVIII–XIX веках не национализировало все фабрики и паровые двигатели из-за появления химического оружия или бомбардировочной авиации; вместо этого законы точно криминализировали вредоносные сценарии применения и ограничивали производимую властью.

Фактчек: ИИ следует регулировать как индустриализацию, а единственная защита от тирании — законодательные запреты и глобальные табу — □: Аналогия с технологией общего назначения полезна, но сама по себе не опровергает необходимость специальных мер для опасных возможностей. Запреты на конкретные злоупотребления важны, однако их исключительная достаточность не доказана: нужны также исполнение закона, независимый контроль и технические ограничения.

Многополярный рынок и необходимость глобального табу 21:50: Идея государственной монополизации была бы оправдана лишь в том маловероятном случае, если бы единая частная лаборатория получила недостижимый отрыв в AGI, угрожающий захватом мира. Реальный ландшафт ИИ является мультиполярным с жесточайшей конкуренцией на каждом уровне стека, поэтому давление государства на одну компанию бессмысленно и вредно. Индивидуальный моральный героизм отдельных основателей не спасет от диктатуры: единственной реальной защитой общества от скатывания в цифровую тиранию является выработка строгих законодательных запретов и жестких политических табу на использование ИИ для массовой слежки и цензуры — по аналогии с негласным табу на применение ядерного оружия, установившимся после Второй мировой войны.

[24:20-24:39] Заключительные слова автора о важности общественных дискуссий и призыв подписаться на рассылку в блоге.

ЛИЧНО ДЛЯ ТЕБЯ

Ограничивай полномочия AI-агентов, а не только инструктируй их: При работе с Claude или Codex оставляй минимально необходимые права, журнал действий и ручное подтверждение удаления данных, платежей и изменений критичных настроек. Послушная модель не обязательно безопасна.

Сокращай передачу чувствительных данных: Перед загрузкой медицинских документов и личной переписки проверяй сроки хранения, использование для обучения и возможность удалить данные; убирай ненужные идентификаторы. Локальная обработка снижает передачу третьим сторонам, но не устраняет риски телеметрии, резервных копий и компрометации устройства.

Не привязывай личный рабочий процесс к одной модели: Сохраняй важные инструкции, контекст и результаты вне единственного сервиса и периодически

проверяй альтернативу на привычных задачах. Это уменьшает потери времени при блокировках, сбоях и изменениях условий доступа — независимо от исхода конфликта с Пентагоном.

Проверяй скрытые допущения громких AI-прогнозов: В экономических расчётах отделяй цену токенов от полной стоимости надёжного результата. В спорах о регулировании проверяй конкретные полномочия, процедуру обжалования и независимый контроль, а не принимай ни «безопасность», ни «свободу рынка» за достаточное обоснование.

КОНСЕНСУС РЕЦЕНЗЕНТОВ

Все три ревью сходятся в условной корректности расчёта стоимости обработки камер и необходимости оговорок к прогнозу тотального надзора. Главные расхождения — степень подтверждённости конфликта Anthropic с Пентагоном, широта доктрины третьей стороны, фиксированность отставания открытых моделей и точность исторических примеров; GPT дополнительно выявляет неверное определение alignment, а Claude — спорные атрибуции позиций Anthropic и Амодеи.