

Dylan Patel - Two labs will soon control most of the world's workforce

Dwarkesh Patel · 2026-08-25 · <https://www.youtube.com/watch?v=aV26V1Uvkjw>



Выпуск подкаста Dwarkesh Podcast, в котором ведущий Дваркеш Патель обсуждает с основателем аналитической компании SemiAnalysis Диланом Пателем макроэкономику масштабирования ИИ. Главная мысль разговора: экспоненциальный рост вычислительных мощностей и переход ИИ-лабораторий к сверхприбыльности запустят цепную реакцию в мировой экономике. Капитальные затраты на дата-центры выкачают триллионы долларов с долговых рынков, спровоцируют рост процентных ставок, обрушат стоимость традиционных активов и приведут к тому, что эквивалент всей человеческой рабочей силы окажется в руках одной-двух частных корпораций.

Экономика ИИ-лабораторий и монополизация вычислений

Переход от убытков к сверхприбыли 00:30: Ведущие ИИ-лаборатории (OpenAI, Anthropic) перестали быть стартапами, сжигающими венчурный капитал, и начали генерировать огромную прибыль (Anthropic — со второго квартала, OpenAI — с третьего). Базовая стоимость вычислений составляет 10-15 млн за мегаватт. Раньше обслуживание моделей вроде GPT-4 на GPU Nvidia Hopper генерировало отрицательную маржу. Теперь же инференс передовых моделей приносит до 50 млн выручки на каждый потраченный мегаватт. Эта колоссальная маржинальность позволяет лабораториям реинвестировать всю прибыль обратно в обучение следующих поколений ИИ.

Фактчек: OpenAI и Anthropic уже вышли в огромную прибыль (Anthropic со 2-го квартала, OpenAI с 3-го) — Δ : Рецензенты расходятся. Gemini (с поиском) подтверждает операционную прибыль Anthropic во 2 кв. 2026 (~559 млн), но OpenAI по тем же данным сжигает деньги (убыток ~38,5 млрд за 2025 при выручке \$13 млрд, прогноз убытков до 2028); Claude и GPT считают тезис публично не подтверждённым. Вероятно, речь о положительной маржинальности инференса «без учёта R&D», а не о чистой прибыли компаний.

Централизация мирового вычислительного ресурса 03:45: В начале года OpenAI и Anthropic потребляли около 2 ГВт мощностей, а к концу года этот показатель превысит 5 ГВт (рост в 3-4 раза). В этом году на долю лабораторий пришлось 30% всех новых вычислительных мощностей в мире, а в следующем году эта цифра достигнет 40-50%. К концу 2027 или 2028 года две компании будут контролировать подавляющее большинство всех доступных FLOPS на планете. Ситуация усугубляется тем, что новые чипы (GB300, TPUn7, Trainium3) в 3-5 раз эффективнее предыдущих поколений, что дает лабораториям мультипликативный отрыв в производительности.

Фактчек: Рост мощностей OpenAI+Anthropic с 2 до 5 ГВт = «в 3-4 раза» — $\square$: Арифметически 2,5 раза (GPT). Кроме того, электрическая мощность $\neq$ вычислительная производительность, а аренда $\neq$ владение.

Асимметрия капитальных затрат и доходов 07:08: В цепочке поставок ИИ существует гигантский разрыв между стоимостью оборудования и генерируемой выручкой. Строительство фабрики для производства 1 ГВт вычислительных мощностей в год требует около 6 млрд CapEx. При этом 1 ГВт мощностей генерирует 100 млрд конечной выручки ежегодно. Несмотря на стократную разницу, физические ограничения цепочки поставок (например, производство зеркал для EUV-литографии от Carl Zeiss и ASML) не позволяют мгновенно масштабировать производство даже при наличии неограниченного капитала. К 2030 году производство EUV-машин будет ограничено примерно 100 установками в год.

Фактчек: Экономика на мегаватт: 10-15 млн затрат против 50-100 млн выручки; «стократная разница» между 6 млрд CapEx фабрики и 100 млрд выручки с 1 ГВт — $\square$: 6 млрд $\rightarrow$ 100 млрд — это ~17 раз, не 100 (GPT). Выручка на МВт внутри выпуска скачет 50 $\rightarrow$ 60-70 $\rightarrow$ 100 млн, а 10-15 млн/МВт похоже на годовую аренду, не на CapEx железа (GB200 — ~30-40 млн/МВт) — расчёты «на салфетке» с несопоставимыми периодами и составом затрат (Claude, GPT). Gemini дополнительно приводит 35-60 млрд за 1 ГВт дата-центра, но это стоимость ДЦ, а не фабрики чипов — сравнение некорректное.

Фактчек: К 2030 году выпуск EUV-машин ограничен ~100 в год — Δ : Прямое расхождение рецензентов. Claude: $\square$, в рамках заявленных целей ASML (~90 low-NA + 20 High-NA, отгрузки 42-44 в 2023-24). Gemini: $\square$, ASML планирует ~130 машин уже в

2027 и рассматривает рост до ~170 к 2028. Порядок величины верен, конкретная цифра занижена на 30–70%; узкое место в оптике Zeiss реально.

Взрывной рост стоимости вычислений 13:00: К 2028 году лаборатории планируют суммарно задействовать до 100 ГВт мощностей. Поскольку сегодня любой желающий может арендовать серверы за 10-15 млн/МВт, развернуть открытую модель (например, Kimi) и получать прибыль, лабораториям придется перебивать цены на рынке. Стоимость вычислений неизбежно вырастет до 25-50 млн за мегаватт. Компании вроде Meta и SpaceX, строящие дата-центры без немедленной внутренней потребности, уже пользуются этим, сдавая мощности в аренду лабораториям с гигантской наценкой (например, по \$40 млрд за гигавайт).

Смещение фокуса на R&D и фактор Китая

Искусственное замедление и регуляторные риски 16:24: Регулирование и требования безопасности замедляют внешнее развитие ИИ сильнее, чем конкуренция. OpenAI откладывает релиз Astra и приостанавливает обучение, а Anthropic не выпускает модель следующего поколения (условный Mythos 2) из-за внутренних протоколов безопасности. Если лабораториям запретят выпускать лучшие модели, их выручка на мегаватт перестанет расти. Это лишит их возможности переплачивать за инкрементальные вычислительные мощности, что может сорвать планы по достижению 100 ГВт к 2028 году.

[24:28-25:33] Рекламная вставка Grok Bot.

Снижение доли инференса в пользу обучения 29:51: Вопреки консенсусу, лаборатории будут выделять всё меньше вычислительных мощностей на инференс (обслуживание пользователей) и всё больше — на внутренний R&D. Несмотря на то, что инференс приносит \$60-70 млн на мегаватт, дисконтированные денежные потоки от создания AGI несоизмеримо выше. Рост выручки Anthropic в последние месяцы вышел на плато, хотя объем их вычислительных мощностей продолжает расти. Это математически доказывает, что каждый новый мегаватт направляется не на продажу токенов, а на внутренние исследования.

Фактчек: К 2030 году ежегодный AI-CapEx достигнет 10 трлн — четверть-треть экономики США — □: Внутреннее противоречие — 10 трлн *в год* в одном блоке и 11 трлн *суммарно за 2024–2029* в следующем (Claude). Консенсус-прогнозы PwC/Goldman — 1,1–1,6 трлн в год к 2030–31, McKinsey — 5–8 трлн за пятилетку (Gemini). Если 10 трлн — глобальная цифра, пересчёт в долю ВВП США некорректен (GPT).

Отставание Китая и будущий экспоненциальный рост 33:36: В 2022 году США вводили 45-50% мировых вычислительных мощностей, а Китай — 30-35%. Сегодня из-за санкций на долю Китая приходится менее 10% новых дата-центров. Однако к 2028-2029 годам внутренние китайские фабрики (SMIC, CXMT)

выйдут на массовое производство. Это позволит Китаю добавлять по 50 ГВт мощностей ежегодно. Тем не менее, качество этих чипов будет значительно уступать западным аналогам, из-за чего 50 ГВт китайских мощностей будут эквивалентны лишь 20 ГВт американских.

Фактчек: Китай к 2028-2029 будет добавлять по 50 ГВт мощностей ДЦ ежегодно — □: Единственный проверивший рецензент (Gemini) ссылается на Rystad Energy и Wood Mackenzie: вся мощность китайских ДЦ к 2030 — ~60 ГВт, т.е. прирост ~28 ГВт за 5 лет, а не 50 в год. Завышение примерно на порядок.

Фактчек: Плато выручки Anthropic при росте мощностей «математически доказывает», что все новые MBT идут в R&D — □: Единодушно у Claude и GPT — подмена корреляции доказательством. Плато объясняется дефицитом мощностей, лимитами клиентов, ценовой политикой, задержкой ввода железа или отчётным лагом; публичный run-rate Anthropic за 2025 рос с 1 до 5+ млрд, плато не видно.

Анатомия распределения вычислений 40:11: Исторически бюджет вычислений лаборатории делился в пропорции 60% на обучение и 40% на инференс. Однако из этих 60% половина уходит на чистый ресерч (тестирование архитектур, гиперпараметров, данных), 10% на разработку, и лишь малая часть — на финальный прогон. Финальное предобучение передовой модели занимает менее 200 MBT в течение двух месяцев. По мере автоматизации работы ИИ-исследователей грань между ресерчем и финальным обучением сотрется, требуя еще большей концентрации мощностей внутри лабораторий.

Фактчек: Гиперскейлеры займут 5 трлн под 8%, вытеснят остальных заёмщиков, развивающиеся страны уйдут в дефолт («второй шок Волкера») — □: Эффект вытеснения — реальный механизм, но ставки, суммы и неизбежность дефолтов — макроспекуляция, зависящая от политики центробанков, сбережений и роста производительности. Источников для 11 трлн / \$5 трлн / 8% нет.

Макроэкономический шок: \$10 трлн CapEx и долговой кризис

Скрытые капитальные затраты на инфраструктуру 42:57: Оценки ИИ-инфраструктуры часто учитывают только IT-оборудование (40-50 млрд), игнорируя стоимость самих зданий дата-центров и электростанций, которые нужно строить за годы до установки серверов. Чтобы ввести 100-150 ГВт мощностей, турбины и бетон нужно оплачивать уже сегодня. К 2030 году ежегодные капитальные затраты на ИИ достигнут 10 трлн. Это означает, что от четверти до трети всей экономики США будет работать исключительно на строительство дата-центров.

Фактчек: Рост ставок на 2-3 п.п. обрушит акции J&J и железных дорог — △: Направление верное (дисконтирование), масштаб преувеличен. В 2022 ставки выросли на ~4 п.п., S&P просел ~20%, а дивидендные защитные сектора упали меньше рынка — «переоценятся вниз», но не «рухнут».

[47:51-48:56] Рекламная вставка Antithesis.

ИИ спровоцирует глобальный кризис суверенного долга 48:56: В период с 2024 по 2029 год на ИИ-инфраструктуру потребуется 11 трлн, из которых 5 трлн придется привлечь через долговые рынки. Гиперскейлеры (Amazon, Microsoft, Meta) будут массово выпускать облигации. Поскольку рентабельность ИИ колоссальна, они легко согласятся платить 8% годовых. Это вызовет эффект вытеснения (crowding out): стоимость заимствований вырастет для всей остальной экономики. Развивающиеся страны (например, Пакистан или Нигерия) не смогут рефинансировать свои долги по новым ставкам, что приведет ко «Второму шоку Волкера» и массовым дефолтам.

Фактчек: Две компании будут контролировать эквивалент всей рабочей силы Земли; экономика ИИ неизбежно тяготеет к монополии — □: Нет способа перевести FLOPS в «миллионы автономных работников» — рост компьютера не гарантирует пропорционального роста надёжно выполненной работы. Экономия от масштаба реальна, но открытые модели, регулирование, ограничения спроса и физические барьеры автоматизации делают монополию не следствием, а допущением.

Обвал традиционных фондовых рынков 57:37: Рост процентных ставок на 2-3% математически увеличит ставку дисконтирования для всех будущих денежных потоков. В результате акции стабильных компаний (например, Johnson & Johnson или железнодорожных операторов), которые ценятся за предсказуемые дивиденды на горизонте 30 лет, рухнут в цене. Весь мировой капитал будет принудительно перераспределен в пользу ИИ-компаний, так как любые другие инвестиции станут экономически бессмысленными на фоне доходности от создания AGI.

Сингулярность и централизация рабочей силы

Экономика удвоения ВВП и сопротивление государств 59:15: В полностью автоматизированной экономике глобальный ВВП сможет удваиваться каждый год. В таком режиме равновесная процентная ставка составит десятки или сотни процентов годовых. Альтернативные издержки капитала станут настолько высокими, что государственные траты на пенсии будут выглядеть абсурдно на фоне возможности построить завод, производящий роботов, которые строят другие заводы. Чтобы не допустить коллапса текущей системы, правительства начнут блокировать строительство дата-центров и запрещать релизы моделей.

Опасность скрытого рекурсивного самосовершенствования (RSI) 1:03:11: Искусственное замедление релизов не остановит прогресс, а сделает его скрытым. Если правительство обяжет лаборатории задерживать выпуск моделей на 6 месяцев, то в условиях RSI за эти полгода ИИ внутри лаборатории совершит скачок, эквивалентный 3-6 годам обычного развития. В результате

общество останется с устаревшими моделями, в то время как внутри одной корпорации будет работать сверхразум, недоступный для внешнего контроля.

[1:06:24-1:07:28] Рекламная вставка Jane Street.

Концентрация всей рабочей силы в двух компаниях 1:07:28: Эффективная «ИИ-популяция» внутри передовых лабораторий растет в 10 раз каждый год (за счет роста вычислительных мощностей в 4-5 раз и улучшения алгоритмов в 3 раза). Если сегодня OpenAI располагает эквивалентом 10 миллионов цифровых работников, то через год их будет 100 миллионов, а затем — миллиард. К концу десятилетия одна частная компания будет обладать автономной рабочей силой, превосходящей по численности и возможностям всё население Земли.

Конец децентрализованного капитализма 1:11:24: Экономика ИИ фундаментально тяготеет к абсолютной монополии. Обучение моделей имеет гигантскую экономию от масштаба, а развертывание создает непреодолимый маховик сбора данных. Хотя сегодня внешние компании (например, трейдеры из Jane Street) извлекают из ИИ больше прибыли, чем сами создатели, лаборатории неизбежно перекроют этот доступ. Доходность от использования ИИ для внутреннего R&D настолько превышает любую внешнюю монетизацию, что лаборатории заберут все вычислительные мощности себе, централизовав власть в масштабах, невиданных в истории капитализма.

ЛИЧНО ДЛЯ ТЕБЯ

Кредит на стройку — стресс-тест, не паника: Если финансирование дома привязано к Euribor или плавающей ставке, проверь платежи при росте на 2–3 п.п. и запас ликвидности при задержке стройки; это аргумент за фиксацию ставки. Но сценарий «ИИ выкачает долговые рынки» — хвостовой риск, построенный на цифрах, противоречащих сами себе в 2 раза.

Не перетряхивай портфель под «конец капитализма»: Рецензенты разошлись — Gemini предлагает учесть риск обвала дивидендных акций, GPT прямо говорит: оснований продавать обычные активы или концентрировать капитал в ИИ нет. Рабочий компромисс — держать в голове механизм дисконтирования, но не действовать по прогнозу, чьи ключевые числа завышены на порядок.

Страховка от удорожания и задержек моделей: Единственный практически значимый тезис выпуска — компьютер может дорожать, а лучшие модели задерживаться внутри лабораторий. Не завязывай рабочий процесс на одного вендора: код, промпты и контекст держи вне платформы, периодически прогоняй одни задачи через альтернативу (Kimi, DeepSeek, Qwen на арендованном GPU). Считай стоимость принятого результата — подписка плюс твоё время на проверку, а не «цифровых работников».

Тренажёр критического мышления: Выпуск — образцовый набор ошибок прогнозиста: путаница «в год» vs «суммарно», 17х выданное за 100х, 2,5х за «3-4х», корреляция как «математическое доказательство», экстраполяция текущего темпа на 5 лет без насыщения спроса. Полезно ловить эти паттерны в любых макропрогнозах, включая собственные.

КОНСЕНСУС РЕЦЕНЗЕНТОВ

Все трое сошлись: макросценарий (\$10 трлн CapEx, долговой кризис, монополия двух компаний) — спекуляция с арифметическими и логическими ошибками, а «плато Anthropic доказывает уход мощностей в R&D» — некорректный вывод; практический вывод один — сохранять независимость от единственного поставщика ИИ.

Расхождения: по EUV-машинам Claude ставит □ (~100/год в рамках планов ASML), Gemini — □ (130-170); по прибыльности Gemini приводит конкретные цифры (Anthropic в плюсе, OpenAI — глубокие убытки), Claude и GPT считают тезис непроверяемым; по реакции на инвестиционный риск Gemini советует учитывать, GPT — не действовать.