

AI researchers debate how close we are to recursive self-improvement

Dwarkesh Patel · 2026-09-11 · <https://www.youtube.com/watch?v=PrSf7IOYu-I>



В этом выпуске обсуждаются технические барьеры и перспективы создания искусственного сверхразума (ASI) и систем рекурсивного самосовершенствования (RSI). Участники — ведущие ИИ-исследователи: Берен Миллидж (CTO Zyphra), Джон Шульман (Chief Scientist в Thinking Machines, экс-сооснователь OpenAI, руководивший созданием RLHF для ChatGPT) и Чарли О'Нил (глава обучения моделей в Baseten).

Главная мысль эпизода: индустрия переходит от масштабирования данных претрейна к генерации синтетических сред для обучения с подкреплением (RL). Несмотря на проблемы с непрерывным обучением и нехваткой качественных данных, базовые когнитивные агенты, способные заменить удаленных сотрудников и кратно ускорить сами ИИ-исследования, появятся уже в ближайшие 1–3 года, что запустит цикл стремительного самосовершенствования алгоритмов.

Главные препятствия на пути к сверхразуму

Разрыв между симуляцией и реальностью (sim-to-real gap) 01:06: Главной технической причиной, по которой к 2036 году может не появиться ASI, является неспособность моделей к непрерывному обучению (continual learning) и обобщению мета-обучения. Модели упираются в собственные слабые места, не умеют надежно проверять свои же результаты и не дают взрывного роста возможностей: даже если ИИ пишет код в 100 раз быстрее, это не делает весь процесс разработки в 100 раз эффективнее.

Необходимость новой архитектурной парадигмы 03:20: Текущий рецепт «Трансформер + RL» может выйти на плато. Закон Мура поддерживался за счет множества дискретных инноваций. В LLM масштабирование претрейна уперлось в потолок, но переход к RL создал новую кривую роста. Для достижения RSI может потребоваться очередной фундаментальный сдвиг (вплоть до отказа от градиентного спуска), который текущие LLM не способны изобрести самостоятельно.

Фактчек: Для RSI потребуется отказ от градиентного спуска — □: Гипотеза без консенсуса в сообществе.

Эффект порога (шахматный рейтинг Эло) 06:07: Прогресс ИИ исторически выглядит как линейный рост рейтинга Эло, который в определенный момент резко пересекает порог человеческих возможностей (как это было с шахматными ботами). До этого момента экономическое влияние ИИ кажется ограниченным, но траектория указывает на то, что переход к сверхчеловеческому уровню произойдет стремительно, если только кривая не выйдет на асимптоту прямо перед этим порогом.

Роль человека и проблема постановки целей

Сложность открытых научных исследований 07:15: ИИ крайне эффективен в оптимизации четко заданных метрик (например, снижение потерь претрейна), но не способен на открытый научный поиск, необходимый для смены парадигм. Изобретение новых подходов, таких как переход к предсказанию следующего токена в 2012–2014 годах, требует интуиции внешнего контура, которой у моделей пока нет.

Фактчек: Переход к предсказанию следующего токена случился в 2012-2014 — □: Next-token как цель обучения — Bengio 2003, RNN-LM Миколова 2010; как парадигма масштабирования — GPT-1 (2018). Возможно, имелись в виду char-RNN/seq2seq, но датировка неверна.

Ускорение исследований внутри заданных рамок 12:05: Если цель специфицирована, ИИ может радикально ускорить процесс. Например, если бы ИИ анализировал законы масштабирования Каплана, он бы сразу заметил ошибку с отсутствием отжига скорости обучения (learning rate annealing), что сэкономило бы индустрии годы исследований. В рамках текущей парадигмы ИИ даст ускорение в 10 раз просто за счет сбора низковисящих плодов.

Фактчек: В законах масштабирования Каплана была ошибка с отжигом learning rate, стоившая индустрии «годы» — □ с оговоркой: Ошибка реальна — фиксированный косинусный шедулер (на 250k шагов) не подгонялся под реальное число токенов, большие модели недообучались; Chinchilla (2022) это исправила. Но «годы» — это ~2 года, и нет уверенности, что автоматический анализ нашёл бы проблему без человеческой гипотезы.

Финальная задача человечества — выравнивание (alignment) 15:55:

Последней функцией людей в цикле ИИ-разработки останется определение целей и спецификация поведения моделей. Даже когда ИИ полностью автоматизирует написание кода и проведение экспериментов, людям придется писать «конституции» моделей и решать, что именно означает «полезность» в сложных и неоднозначных сценариях.

Олигополия лабораторий против опенсорса и дистилляции

Дистилляция как фактор децентрализации 19:09: Любой навык, на получение которого передовые лаборатории тратят огромные вычислительные ресурсы через RL, легко копируется в меньшие модели. Поведение модели требует для передачи очень малого количества бит информации, поэтому дистилляция на основе успешных траекторий (выводов) передовых ИИ нивелирует монополию гигантов.

Фактчек: Дистилляция нивелирует монополию гигантов, любой навык легко копируется — ⚠: Дистилляция резко дешевле исходного обучения и реально сокращает отрыв (DeepSeek, Qwen). Но ограничена ёмкостью ученика и покрытием задач — редких ошибок и длинных взаимодействий; сам пересказ это признаёт (разрыв в «реалистичном поведении»).

Сбор идеального распределения промптов 21:20: Китайские ИИ-компании успешно сокращают отставание, покупая данные у прокси-сервисов, через которые пользователи из Китая получают доступ к заблокированным GPT-4 или Claude. Это дает им идеальное распределение реальных пользовательских запросов, что является критически важным компонентом для качественной дистилляции.

Фактчек: Китайские лаборатории покупают логи промптов у прокси-сервисов доступа к GPT-4/Claude — ⚠: Прокси-перепродажа API в Китае и обвинения OpenAI в адрес DeepSeek — реальны. Но конкретно покупка распределения промптов — инсайдерская гипотеза без публичных источников. Расхождение: Gemini считает подтверждённым, ссылаясь на «совместный отчёт CISA/FBI/NSA от сентября 2026» — эта ссылка непроверяема и с подозрительной датой; Claude и GPT относят тезис к неподтверждённым.

Сложность копирования реалистичного поведения 25:09: Передовые лаборатории обучают модели по двум осям: сложность (решение головоломок) и реалистичность (многошаговое взаимодействие с человеком). Наивная дистилляция часто захватывает только способность решать сложные бенчмарки, но не передает нюансы реалистичного поведения агента в рабочей среде, что объясняет разрыв в качестве между оригиналом и копией.

Фактчек: Chinchilla обосновывает возврат к огромным переобученным моделям при дефиците данных — ⚠: Chinchilla описывает оптимальный баланс размер/данные

при фиксированном compute, а не универсальное преимущество больших моделей; стоимость инференса сдвигает оптимум в сторону меньших.

Обучение моделей для автоматизации ИИ-исследований (RSI)

Дифференциальный подход к созданию сред 29:03: Первые модели с функцией RSI не будут обучаться с нуля в вакууме. Лаборатории берут баги и архитектурные улучшения, найденные людьми за последние месяцы, и превращают их в среды для RL-обучения. Модель постоянно подтягивается к самому краю человеческих открытий, фиксируя всю предыдущую историю развития.

Амортизация общих навыков 35:32: Обучение моделей работе с финансами, Excel или PowerPoint нужно не только для коммерциализации. Эти задачи развивают общие мета-навыки: упорство, способность сортировать контекст и планировать на длинных горизонтах. Эти навыки напрямую переносятся на способность ИИ проводить собственные исследования.

Доминирование sim-to-real перехода 39:31: Обучение в симуляциях остается главным методом из-за крайне низкой эффективности выборки (sample efficiency) у ИИ. Моделям требуются тысячи взаимодействий для обновления весов. Поскольку человек не может сидеть в цикле RL-обучения, сложные задачи (управление бизнесом, суды) приходится симулировать в дата-центрах, надеясь на успешный перенос навыков в реальный мир.

Непрерывное обучение (Continual Learning) и его ограничения

Медленный цикл обновления 42:11: Сейчас модели не учатся в реальном времени на данных от развертывания у пользователей. Лаборатории используют медленный внешний цикл: собирают логи, фильтруют их и добавляют в претрейн или мид-трейн следующего поколения моделей.

Кумулятивные задачи против нестационарных 47:36: RSI — это кумулятивная задача: однажды открыв механизм внимания, ИИ фиксирует его в коде навсегда. Реальная работа (например, юриспруденция) — нестационарна: контекст, отношения между людьми и правила постоянно меняются. ИИ гораздо проще автоматизировать написание кода для нейросетей, чем стать хорошим помощником юриста.

Проблема катастрофического забывания 55:39: Попытки непрерывно дообучать одну базовую модель (через SFT или on-policy дистилляцию) приводят к разрушению ранее усвоенных знаний и падению общих способностей. RL менее деструктивен, так как меняет очень малую часть весов, но он плохо подходит для внедрения новых явных фактов (например, структуры конкретной компании).

Фактчек: Непрерывное дообучение разрушает старые знания; RL меняет малую часть весов, потому менее деструктивен и эффективнее SFT — $\triangle$:

Катастрофическое забывание — фундаментальная и подтверждённая проблема. Работы 2025 года (Mukherjee et al. «RL finetunes small subnetworks», Shenfeld et al. «RL's Razor») эмпирически показывают: RL-апдейты разреженнее и ближе к исходному распределению, забывание меньше. Расхождение: GPT ставит $\square$ — при полнопараметрическом RL обновляться могут все веса, а «один бит награды = выше эффективность» упрощение: редкий сигнал верно/неверно может и затруднять обучение.

Исчерпание данных и архитектурные пределы

Конец эпохи Common Crawl 1:05:30: Низковисящие плоды данных из интернета исчерпаны. Для обучения сложным рассуждениям нужного сигнала (например, доказательств нерешённых математических проблем) просто не существует в сети. Прогресс теперь полностью зависит от генерации синтетических данных и создания сложных RL-сред.

Фактчек: Качественные интернет-данные исчерпаны, прогресс «полностью» зависит от синтетики и RL — $\triangle$: По оценке Epoch AI, запас качественного публичного текста (~300 трлн токенов) исчерпывается в 2026–2032 при текущих темпах — дефицит реален. Но убывающая отдача не равна «потолку претрейна»; формулировка «полностью зависит» — преувеличение.

Замедление роста числа параметров 1:11:17 Масштабирование параметров временно замедляется, так как для RL-обучения критически важна дешевизна инференса. Модели должны быть достаточно компактными, чтобы экономично прогонять миллионы длинных симуляций. Фокус сместился с наращивания общего числа параметров на оптимизацию активных параметров (MoE) и эффективности данных.

Возврат к переобученным моделям 1:16:13: Если вычислительные мощности перестанут быть главным узким местом (с выходом новых чипов), индустрия вернется к созданию огромных моделей. Согласно законам масштабирования Chinchilla, большие модели значительно эффективнее используют данные и лучше обобщают информацию, что станет критичным в условиях дефицита качественных текстов.

Почему RL работает лучше ожиданий

Мид-трейн делает 80% работы 1:18:52: Успех RL во многом обусловлен качественным мид-трейном на синтетических данных, который подготавливает модель. RL не учит логике с нуля, а лишь корректирует политику модели, требуя минимального количества бит информации из эпизодов.

Высокое соотношение сигнала и шума 1:19:58: В отличие от SFT, где модель заставляют точно копировать каждый токен рассуждения (что создает много

шума из-за чужого стиля), RL дает всего один бит информации — правильный ответ или нет. Это радикально повышает эффективность обучения, так как сигнал не тонет в шуме.

Обобщение горизонта планирования 1:23:07: Даже если RL не переносит конкретные навыки рассуждения из математики в программирование, он создает «горизонтальное обобщение». Модели учатся использовать больше токенов и думать дольше, не срываясь в галлюцинации. Способность модели работать над задачей удваивается каждые три месяца.

Фактчек: Горизонт задач, которые тянет модель, удваивается каждые 3 месяца — ▲: По METR долгосрочный тренд 2019–2025 — удвоение ~каждые 7 месяцев, на моделях 2024–2025 — ускорение до ~4 месяцев. «Три месяца» — экстраполяция по верхней границе. Плюс метрика измеряет длительность задачи при заданной вероятности успеха, а не надёжную месячную автономию.

Прогнозы и таймлайны

ИИ как полноценный удаленный сотрудник 1:29:02: Базовый агент, способный заменить удаленного работника на сложных проектах длительностью в месяц (программирование, видеомонтаж, юриспруденция), появится через 1–3 года. Срок зависит от того, придется ли ИИ использовать браузер как человеку, или компании адаптируют свои системы под API.

Фактчек: Удалённый сотрудник через 1-3 года, 10x в исследованиях через 2-10 лет, ASI через 3-10 лет — □: Прогнозы людей внутри индустрии с историей переоценки сроков (RL для роботов, self-driving). Sim-to-real разрыв, который сами называют главным барьером, — причина не брать нижние границы за базовый сценарий. Автоматизация части исследований не доказывает рекурсивного ускорения: остаются compute, верификация, реальные эксперименты.

Ускорение ИИ-исследований в 10 раз 1:32:20: Десятикратный рост продуктивности ИИ-исследователей ожидается в перспективе 2–10 лет. Даже способность модели автономно провести 2–3 цикла экспериментов подряд без сбоев радикально ускорит прогресс, перенеся узкое место с написания кода на способность человека осознавать результаты.

Достижение ASI (Искусственного Сверхразума) 1:34:35: Система, превосходящая лучших мировых экспертов во всех когнитивных профессиях, появится через 3–10 лет. Автоматизация ИИ-исследований по сути является ASI-полной задачей. Дольше всего ИИ будет осваивать узкие физические дисциплины с малым количеством оцифрованных данных (например, инженерии на заводах TSMC).

ОБЩИЕ ПРОВЕРКИ

Джон Шульман — Chief Scientist в Thinking Machines, экс-сооснователь OpenAI, вёл RLHF для ChatGPT — □: Сооснователь OpenAI (2015), руководил пост-трейнингом ChatGPT, в августе 2024 ушёл в Anthropic, с февраля 2025 — Chief Scientist в Thinking Machines Lab.

ЛИЧНО ДЛЯ ТЕБЯ

Считай ускорение по принятому результату, не по скорости генерации кода: закон Амдала — если написание кода занимало половину времени, его ускорение в 100× даст всего ~1,98× всему процессу. Узкое место — твоя проверка и осознание результата; метрики: время до прохождения тестов, число итераций правок, минуты ревью.

Вкладывайся в спецификации, критерии успеха и тесты, а не в чтение кода построчно: 10× даётся только внутри чётко заданной цели; открытая постановка задачи и вопрос «то ли мы оптимизируем» дольше всего остаются человеческими. Это тот же навык аргументации и формализации, что и в риторике — только применённый к агентам.

Не привязывайся к одному вендору и не строй решения о деньгах на 3-5 лет вокруг дат ASI: дистилляция сокращает отрыв фронта за месяцы, цена за токен падает быстрее ожиданий — держи переключаемость Claude/GPT/Gemini/open-weight. Планируй от измеримого выигрыша сегодня, а не от «через год всё поменяется».

Делегируй код и рутину без сожалений, но не жди ИИ в «грязных» нестационарных задачах: стройка, переговоры с подрядчиками, юридические вопросы на Кипре — контекст и отношения меняются, это автоматизируется заметно позже кода. Автономность агентов расширяй ступенчато: сначала один проверяемый шаг, потом 2-3 цикла «изменение → тест → правка» в изоляции; деньги, секреты и необратимые действия — только с подтверждением.

КОНСЕНСУС РЕЦЕНЗЕНТОВ

Все трое сошлись: таймлайны (1-3 / 3-10 лет) — мнения, а не факты; катастрофическое забывание и дефицит качественных данных реальны; ошибка Каплана с отжигом подтверждается, но «годы» — преувеличение; «удвоение каждые 3 месяца» завышено относительно данных METR (7 → ~4 месяца).

Разошлись по двум пунктам: RL-и-веса (Claude и Gemini — □ с опорой на работы 2025 года о разреженных апдейтах; GPT — □ как упрощение, поскольку полнопараметрический RL может менять все веса) и китайская дистилляция через прокси (Gemini — □ со ссылкой на непроверяемый отчёт спецслужб

«сентября 2026»; Claude и GPT — недоказанная инсайдерская гипотеза).
Датировку next-token 2012-2014 Claude счёл натяжкой, GPT — ошибкой.