

AI Crisis in Math: Navier-Stokes, Riemann Hypothesis & Complexity Theory | Lex Fridman Podcast #503

@lexfridman · 2026-10-26 · <https://www.youtube.com/watch?v=ZAzWPE9Cltg>



В экстренном выпуске подкаста Лекса Фридмана обсуждается тектонический сдвиг в истории науки: за несколько недель ИИ-системы (от OpenAI и Anthropic) заявили о решении около 700 сложнейших математических задач, включая построение контрпримера к уравнениям Навье–Стокса (с внешней силой), доказательство квази-гипотезы Римана, гипотезы об уникальных играх (Unique Games Conjecture), равенства классов сложности $RL = L$, а также полную формализацию Великой теоремы Ферма в системе Lean. Участники беседы: Стивен Брантон (специалист по вычислительной гидродинамике и прикладной математике, Вашингтонский университет), Анкур Мойтра (математик и специалист по теоретической информатике, MIT и Институт перспективных исследований в Принстоне) и Алекс Конторович (специалист по теории чисел и формальной верификации, Ратгерский университет).

Главная мысль выпуска: автоматическое доказательство теорем перестало быть узким местом и прокси-метрикой глубины человеческого понимания. Хотя для молодых учёных, чьи многолетние исследования оказались обесценены за одну ночь, это стало экзистенциальным шоком, сама математика не умирает. Ценность смещается от технического пробивания формальных доказательств к постановке правильных вопросов, созданию новых концептуальных абстракций, переосмыслению научных институтов и извлечению человеческого понимания («истории») из гигантских машинных выкладок.

Шок от релиза 700 доказательств и кризис идентичности в науке

Переход от свечи к прожектору и разрушение старого карьерного

сценария 00:02:08: Релиз сотен решённых задач вызывает у математиков смесь ужаса и восторга: результат, который принёс бы человеку мгновенную Филдсовскую медаль, теперь генерируется машиной. В течение тысячелетий человечество исследовало математический лес со свечой в руке, а теперь получило мощный прожектор. Однако это полностью разрушает традиционный академический сценарий для аспирантов и постдоков. Раньше рецепт успеха был прост: взять известную открытую задачу, улучшить результат, опубликовать статью в топовом журнале и сделать себе имя. Теперь, когда ИИ решает такие задачи пачками, возникает интеллектуальное одиночество и потеря терпения, необходимого для медленного выращивания глубоких идей из скромных вопросов.

Критика Теренса Тао: эффект спойлера в науке 00:05:49: Филдсовский лауреат Теренс Тао выразил ужас по поводу того, что многие человеко-годы терпеливой исследовательской работы молодых учёных над мотивирующими задачами оказались небрежно разрушены одним релизом. Как невозможно «развидеть» спойлер к фильму или подсказку в кроссворде, так же невозможно продуктивно и глубоко исследовать проблему, когда уже известно о существовании готового решения. Стивен Брантон возражает: знание о том, что технология возможна (например, если бы другая страна изобрела телепортацию), напротив, служит мощнейшим мотиватором для обратного инжиниринга и быстрого понимания принципов её работы.

Смерть «кентавровых шахмат» и «аксиома решённости» (Solved Axiom)

00:10:47: Когда появились первые шахматные движки, возникла идея «кентавровых шахмат» (человек плюс компьютер против человека плюс компьютер), но она прожила недолго: движки стали настолько сильны, что любая попытка человека отклониться от первой рекомендации машины вела к проигрышу. Алекс Конторович и Томас Коберда сформулировали «аксиому решённости» как мысленный предел развития технологий: любую математическую задачу, которую человек может решить во взаимодействии с ИИ, ИИ способен решить самостоятельно. Однако даже если эта аксиома окажется верной (во что сами математики пока не верят), это не уничтожит чистую математику. Ценность чистой математики не в единицах и нулях, хранящихся в дата-центрах, а в понимании, которое живёт в головах людей.

Вычислительная ограниченность человека как источник абстракций

00:13:55: Человек сильно ограничен вычислительно: мы не можем удерживать в голове тысячи статей одновременно. Именно эта слабость заставляет людей создавать обобщаемые, компактные и абстрактные теории, чтобы добраться до финиша. Когда математик доказывает крупную теорему, сообщество в первую

очередь спрашивает не «верно ли это?», а «как ты это сделал?». Новые определения и структуры, созданные по пути, позволяют затем обрушить целый каскад соседних задач. ИИ же может прийти к финишу путём грубого перебора и комбинирования, не создавая понятных человеку концептуальных мостов.

Ошибка швейцара (Doorman Fallacy) применительно к математикам

00:19:55: Сводить работу математика исключительно к доказательству теорем — это классическая «ошибка швейцара» (ошибочное сведение сложной профессии к её самой очевидной видимой функции). Если отель заменяет швейцара автоматической раздвижной дверью ради экономии, он теряет безопасность, узнавание постоянных гостей, помощь с багажом, вызов такси и ощущение престижа. Доказательство старых гипотез — лишь самый видимый побочный продукт глубокого математического мышления. До тех пор, пока не появился ИИ, способность доказывать сложные теоремы почти идеально коррелировала с глубиной понимания, красотой идей и способностью ставить верные вопросы. Теперь эта корреляция разрушена: прокси-метрику (доказательство) можно получить нажатием кнопки без всякого понимания.

Автоматическое открытие физических законов 00:27:48: В физике и прикладной математике машины уже сегодня систематически выводят минимальные управляющие уравнения прямо из данных наблюдений. Наблюдая за падением шаров, алгоритмы заново выводят законы гравитации и второй закон Ньютона $F = ma$; наблюдая за потоками жидкости — выводят уравнения Навье–Стокса. Сейчас эти методы применяются к областям, где у человечества нет простых уравнений: прогнозированию погоды, турбулентности и солнечным вспышкам.

Состязательные примеры (Adversarial Examples) как фишка, а не баг: чужеродность мышления ИИ 00:30:29:

Опасно антропоморфизировать ИИ и считать, что он решает задачи так же, как человек. В глубоком обучении известна уязвимость нейросетей к состязательным возмущениям: добавление невидимого глазу шума заставляет сеть уверенно классифицировать кошку как собаку. Эксперимент научной группы Алекса Мадри в MIT («Adversarial Examples Are Features, Not Bugs») показал поразительную вещь: если взять фотографии кошек, наложить на них «собачий» состязательный паттерн, подписать их в обучающей выборке как «собаки» и обучить на этом полностью фальшивом датасете новую нейросеть, то эта сеть начинает отлично распознавать настоящих собак и кошек на чистых реальных фотографиях. Это значит, что ИИ находит статистические лазейки и «чёрные ходы» в данных, совершенно недоступные человеческому восприятию. Доказательства от OpenAI выглядят именно так — как результат работы нечеловеческого разума.

Фактчек: Обучение на состязательных признаках позволяет распознавать чистые изображения — [1]: Работа *Adversarial Examples Are Not Bugs, They Are Features* (2019) действительно показывает, что незаметные человеку, неустойчивые

признаки несут информацию, обобщаемую на чистую тестовую выборку. «Отлично распознавать» — преувеличение: результаты существенно выше случайных, но ниже обычного обучения. Из эксперимента не следует ни «чужеродность мышления» ИИ, ни достоверность предполагаемых математических доказательств.

Природа машинных доказательств: выпуклая оболочка или полугруппа?

Уничжительная «выпуклая оболочка» против бесконечной

«полугруппы» Тимоти Гауэрса 00:33:50: Скептики часто пренебрежительно говорят, что ИИ не создаёт новых идей, а лишь интерполирует внутри «выпуклой оболочки» (convex hull) уже существующих человеческих знаний — подобно точкам внутри замкнутого воздушного шара. Алекс Конторович вслед за Тимоти Гауэрсом предлагает более точную метафору — полугруппу (semigroup), порождённую известными ходами. Если представить известные математические приёмы как плитки, которые можно прикладывать друг к другу слева, справа, спереди и сзади, то комбинируя базовые шаги в длинные каскады, можно уйти сколь угодно далеко за пределы интуитивно видимой области. В доказательстве квази-гипотезы Римана ИИ использовал только известные людям инструменты (включая прорыв двухлетней давности по кубическим тета-функциям и кубическим суммам Гаусса), но выстроил из них настолько длинную и неочевидную цепочку, что ни один человек не додумался бы соединить их таким образом.

Иллюзия сложности и прорыв в задаче 3SUM 00:37:25: В истории математики регулярно случалось так, что задача десятилетиями считалась неприступной, обрастала славой, а затем находилось элементарное трёхстраничное решение, доступное студенту (которое, будь оно найдено сразу, стало бы рядовым упражнением в учебнике). ИИ идеально приспособлен для вскрытия таких задач, которые на самом деле достижимы текущими методами. Примером результата, меняющего саму структуру науки, стало решение задачи 3SUM с помощью модели Anthropic (описанное Вирджинией Василевской-Уильямс и Джошем Алманом). Базовая алгоритмическая задача с собеседований внезапно оказалась асимптотически проще, чем считало всё сообщество. Поскольку на предположении о сложности 3SUM строилась вся теория «тонкой сложности» (fine-grained complexity), где десятки других задач наследовали её трудность, этот результат не просто дал новые алгоритмы, но и заставил пересмотреть саму программу объединения этих задач в единый класс.

Фактчек: Anthropic опровергла гипотезу 3SUM и обрушила фундамент тонкой сложности — $\triangle$: Прямое подтверждение есть только в ревью Gemini: приведены arXiv: 2610.06783 и показатель времени около $O(n^{1.9992})$, но независимой проверки этой ссылки в предоставленных материалах нет. Даже подтверждённый строго субквадратичный алгоритм опроверг бы соответствующую гипотезу сложности, а не

всю область fine-grained complexity: сохраняются редукции, другие гипотезы и необходимость учитывать вычислительную модель.

Угроза криптографии: факторизация и $P = NP$ 00:41:12: Если бы среди 700 решённых задач оказался полиномиальный алгоритм факторизации (разложения больших чисел на простые множители) или конструктивное доказательство равенства $P = NP$, мгновенно рухнула бы вся современная электронная коммерция, банковские карты и защищённые коммуникации. Умножить два числа по миллиону цифр каждый компьютер может за доли секунды, но обратная операция — найти два простых множителя из их произведения — считается вычислительно трудной, на чём и базируется криптография с открытым ключом (хотя из-за угрозы квантовых компьютеров и алгоритма Шора мир уже постепенно переходит на постквантовые решётчатые методы). Подобные результаты компании не могут просто выкладывать в открытый доступ — они должны проходить через государственные структуры национальной безопасности (в том же АНБ работают тысячи математиков).

Фактчек: Полиномиальная факторизация или $P = NP$ мгновенно уничтожат всю криптографию — □: Полиномиальный алгоритм может оставаться практически непригодным из-за степени и констант. Практически эффективная факторизация угрожает прежде всего RSA, но сама по себе не взламывает эллиптические кривые, решёточные схемы и симметричное шифрование. Эффективный универсальный алгоритм для NP имел бы гораздо более широкие последствия, однако одного равенства $P = NP$ недостаточно для «мгновенного краха».

Трансформация научных институтов, журналов и верификация в Lean

Коллапс рецензирования и опыт открытого кода 00:46:55: В инженерии и прикладных науках (таких как детектор гравитационных волн LIGO или проекты в астрономии) академическим вкладом давно считается не только 15-страничный PDF-файл, но и переиспользуемый код, датасеты и API. В чистой математике система журналов и рецензирования находилась на грани коллапса ещё до появления ИИ из-за растущего вала статей. Аналогичный кризис переживает GitHub и движение open-source: мейнтейнеры репозитория выгорают под лавиной сгенерированных ИИ пул-реквестов. Однако полный отказ от тщательной проверки невозможен: математическая эпистемология требует медленной коллективной верификации, как это было с доказательством гипотезы Пуанкаре Григорием Перельманом, когда мировое сообщество годами разбирало его выкладки на семинарах, требуя пояснений к каждой лемме.

Устройство Lean и библиотеки Mathlib: математика как сборка ракеты 00:55:20: В обычных математических книгах леммы не имеют строгой спецификации — они опираются на контекст предыдущих 60 страниц, поэтому их нельзя просто «взять с полки» как стандартную деталь при сборке ракеты.

Интерактивный прuver (система автоматической проверки доказательств) Lean решает эту проблему. Изначально созданный для написания безошибочного программного кода, Lean содержит базовые аксиомы математики и правила логического вывода, проверяя каждый шаг доказательства. Чтобы не начинать каждый проект с определения натуральных чисел через аксиомы Пеано, поверх Lean за 10 лет была создана гигантская краудсорсинговая библиотека Mathlib, работающая по модели GitHub (с пул-реквестами и публичным код-ревью). Именно связка автоформализации и Lean позволила гарантировать 100%-ную истинность сотен новых доказательств от OpenAI и Anthropic: стохастическим языковым моделям нельзя верить на слово на естественном языке, но если код компилируется в Lean без пропусков, теорема доказана строго до базовых аксиом.

Фактчек: Lean гарантирует 100%-ную истинность сотен новых доказательств —

⚠: Lean проверяет выводимость конкретного формализованного утверждения из используемых аксиом, а не соответствие этого утверждения исходной математической задаче. Отсутствия `sorry` недостаточно: необходим аудит дополнительных аксиом, определений и формулировок. Даже по версии Gemini формализация приложена лишь к части результатов, а громкие заявления о квази-гипотезе Римана и UGC остаются без такой проверки.

Инженерия доказательства против рассказа истории 01:00:05:

Единственная причина, по которой формальное доказательство может быть непонятно человеку — это его масштаб, так как базовых правил вывода всего горстка. В математических статьях исторически смешаны две конфликтующие задачи: «инженерия доказательства» (строгая сборка без логических дыр) и «коммуникация идеи» (передача интуиции). В классических курсах вещественного анализа студентов первые две недели мучают формальным построением вещественных чисел через классы эквивалентности фундаментальных последовательностей (последовательностей Коши) рациональных чисел, хотя интуитивно числовая прямая им уже знакома, а необходимость в строгом построении возникает лишь спустя полтора месяца, когда нарушается интуиция (например, на теореме Больцано–Вейерштрасса о сходящейся подпоследовательности, которая верна в вещественных числах, но неверна в рациональных). В будущем статьи разделятся: за «инженерию» будет отвечать приложенный файл формализации в Lean, а сам текст статьи превратится в увлекательную историю о том, почему исследователи свернули налево, а не направо, и какая метафора стоит за решением.

Теория чисел: квази-гипотеза Римана и формализация Великой теоремы Ферма

Как дзета-функция Римана делает обработку сигналов для простых чисел 01:06:00: Простые числа — это фундаментальные неделимые частицы

(«атомы») мультипликативной структуры чисел. Загадка их распределения кроется в том, что человечество не понимает, как связаны между собой две простейшие операции, доступные трёхлетнему ребёнку: сложение (прибавление единицы) и умножение (построение прямоугольников из кубиков: из 20 кубиков можно сложить прямоугольники 4×5 и 2×10 , а если один кубик закатится под диван, из 19 ничего сложить не выйдет). Гениальный прорыв Римана заключался в применении гармонического анализа (обработки сигналов) к простым числам. Подобно тому как микрофон раскладывает звуковые волны голоса через преобразование Фурье на частоты и амплитуды, дзета-функция Римана переводит дискретные простые числа в пространство комплексных чисел. Нули дзета-функции (точки, где функция обращается в ноль) соответствуют частотам и амплитудам «волн Римана», суперпозиция которых в точности восстанавливает расположение простых чисел. Гипотеза Римана утверждает, что в этом «оркестре простых чисел» все инструменты звучат с абсолютно одинаковой амплитудой — то есть все нетривиальные нули лежат строго на одной вертикальной прямой в комплексной плоскости с вещественной частью $1/2$.

Формальный прорыв: 13 миллионов строк кода для Великой теоремы

Ферма 01:10:59: За месяц до релиза OpenAI Алекс Конторович выступал на коллоквиуме в Принстоне, где утверждал, что современные ИИ-системы способны формализовать математику только в пределах «гало Mathlib» (на 1-2 шага вглубь от уже канонизированных и идеально причёсанных структур библиотеки), а на 3-4 шаге их кустарные определения начинают конфликтовать друг с другом. Кевин Баззард получил грант на миллион долларов и 5 лет работы только на то, чтобы свести формализацию Великой теоремы Ферма к результатам, известным к 1980-м годам (оставив за скобками тяжелейшие теоремы Ленглендса-Таннелла и Мазура). Через два дня после лекции Конторовича модель от Anthropic сгенерировала полное доказательство Великой теоремы Ферма с нуля до базовых аксиом объёмом 13 миллионов строк кода в Lean без единого тега `sorry` (пропуска доказательства). Хотя этот код написан «на коленке» (jerry-rigged) и не соответствует стандартам элегантности Mathlib, сам факт того, что 2-3 инженера без профильной экспертизы в теории чисел пробили эту стену промптами в духе «продолжай», означает, что проблема автоформализации существующих человеческих текстов фактически решена.

Фактчек: Anthropic полностью формализовала теорему Ферма в 13 млн строк Lean силами нескольких инженеров с промптами «продолжай» — $\triangle$: Статус результата расходится между ревью: Gemini подтверждает его, Claude не может проверить, GPT относит к неподтверждённой центральной истории. Дополнительные сведения Gemini — 11 дней, миллиарды токенов и оркестратор Prove2Me — также требуют проверки первоисточников, а не автоматического включения в итог как фактов. Даже успешная формализация одной крупной теоремы не доказывает, что автоформализация произвольных человеческих текстов «фактически решена».

Суть доказательства квази-гипотезы Римана: вертикальная полоса от $7/8$

до 1 01:24:00: Все нетривиальные нули дзета-функции лежат в «критической полосе» (где вещественная часть комплексного числа находится между 0 и 1). Классическая теорема о распределении простых чисел эквивалентна тому, что нулей нет на самой границе — прямой с вещественной частью, равной 1. За последние 70 лет математики смогли отвоевать у этой границы лишь крошечную область, свободную от нулей, которая бесконечно сужается (стремится к нулю по толщине) по мере продвижения вверх или вниз по мнимой оси. Конторович ожидал, что ИИ в лучшем случае немного утолщит эту сужающуюся границу. Вместо этого система OpenAI доказала квази-гипотезу Римана (для дзета-функции и всех L-функций Дирихле): она отсекала целую бесконечную вертикальную полосу конечной ширины — между $7/8$ и 1 (а во второй версии доказательства — около $11/12$), где гарантированно нет ни одного нуля. Для этого ИИ блестяще скомбинировал вложение функции Мёбиуса со скручиванием секстическими (шестой степени) характеристиками. Чтобы дойти от $7/8$ до $1/2$ (полной гипотезы Римана), этот метод уже не годится — потребуется фундаментально новая идея.

Фактчек: Доказана фиксированная область без нулей $\text{Re}(s) > 7/8$, затем получена граница около $11/12$ — Δ : Gemini подтверждает наличие заявленного результата, но признаёт незавершённость независимой проверки; GPT не подтверждает само доказательство. Арифметическая поправка однозначна: $11/12 \approx 0,917$ больше $7/8 = 0,875$, поэтому область без нулей становится **уже**, а результат — слабее. Ни заявление о публикации, ни подробное описание метода не заменяют проверки доказательства.

Пьеса про фон Неймана и Бога: математика как ответ на вопрос «почему»

01:20:15: В 1950-х годах была написана математическая пьеса, по сюжету которой Джон фон Нейман после смерти попадает на небеса и спрашивает у Бога, верна ли гипотеза Римана. Бог отвечает: «Да, верна». Фон Нейман спрашивает: «Но откуда ты знаешь?». Бог отвечает: «Я просто проверил все нули, они все лежат на прямой $1/2$ ». Фон Неймана такой ответ совершенно не устраивает, он приходит в ярость от отсутствия объяснения причины и в итоге оказывается в аду. Этот анекдот иллюстрирует суть математики: голый факт истинности утверждения без понимания того, *почему* оно истинно, прибавляет к человеческому знанию ровно ноль.

Риск засекречивания идей в научном сообществе 01:27:32: В релизе OpenAI были опубликованы только успешные доказательства (около 10% от тысяч протестированных задач). Это создаёт новую опасность для научной культуры: задачи, которые требуют одной яркой творческой идеи, становятся уязвимыми. Как только человек публично выскажет новую концепцию или определение (даст недостающую «плитку»), корпоративный ИИ сможет мгновенно и исчерпывающе перебрать все следствия из неё, собрав все сливки. Из-за этого

математики могут стать гораздо более скрытными и перестать делиться сырыми идеями на семинарах, пока в сообществе сохраняется культура награждения «первого прошедшего через ворота».

Гидродинамика: уравнения Навье-Стокса и проблема турбулентности на триллион долларов

Физика уравнений Навье-Стокса и квадратичная нелинейность 01:30:00:

Уравнения Навье-Стокса — это второй закон Ньютона ($F = ma$, закон сохранения импульса), записанный для бесконечно малых объёмов непрерывно деформируемой среды (жидкости или газа). Они связывают изменение поля скоростей в пространстве и времени с давлением, внешними силами (гравитацией, нагревом) и внутренним трением — вязкостью (у мёда вязкость высокая, у воды меньше, у воздуха ещё меньше). Главная математическая трудность кроется в конвективном члене, описывающем, как жидкость переносит саму себя за счёт собственной инерции. Этот член имеет квадратичную нелинейность. Если взять простейшее обыкновенное дифференциальное уравнение размножения кроликов $dx/dt = x$, популяция растёт экспоненциально, но достигает бесконечности только за бесконечное время. Однако при квадратичной нелинейности $dx/dt = x^2$ решение уходит в бесконечность за конечное время (происходит «взрыв», или finite-time blowup).

Растяжение вихря в 3D: что именно доказал ИИ в задаче Навье-Стокса

01:41:57: Главный вопрос задачи тысячелетия для уравнений Навье-Стокса: всегда ли вязкость (диффузия, рассеивающая скорость на соседние слои жидкости) успевает погасить квадратичный нелинейный рост, или существуют такие начальные условия, при которых скорость жидкости в точке уйдёт в бесконечность за конечное время? В двумерных потоках (которые хорошо описывают крупномасштабную земную погоду и формирование ураганов, где мелкие вихри сливаются в крупные) взрыва не происходит. В трёхмерных потоках действует прямой каскад: крупные вихри дробятся на средние и мелкие вплоть до перехода в тепло. Доказательство, полученное с помощью ИИ (для одной из четырёх официальных формулировок Института Клэя — с гладкой внешней силой, без сингулярностей), строит контрпример. Гладкая внешняя сила идеально дирижирует потоком так, чтобы компенсировать гасящие члены уравнения и создать в 3D особый вихрь. Этот вихрь растягивается вдоль третьей оси и сужается до бесконечно малого радиуса — точно как фигурист на льду, который прижимает руки к груди и начинает вращаться с бесконечной угловой скоростью, сохраняя конечную кинетическую энергию системы.

Фактчек: Контрпример Навье-Стокса с гладкой внешней силой может решить официальную задачу Клэя — $\triangle$: Формально это возможно: варианты C и D официальной постановки допускают внешнюю силу, но она должна удовлетворять всем установленным условиям, а не просто быть гладкой. Здесь важно различать

оценки: □ у Claude относится к допустимости такой постановки, тогда как Gemini подтверждает само предполагаемое достижение OpenAI; GPT его не подтверждает. Даже корректный контрпример с силой не устанавливает возможность взрыва без внешней силы.

Почему математический «взрыв» не даёт квантовых чёрных дыр в воде и не решает инженерных задач 01:35:06: Когда Карл Шварцшильд нашёл сингулярность («взрыв») в уравнениях общей теории относительности Эйнштейна, физики сначала сочли это дефектом модели, но позже открыли реальные чёрные дыры. В гидродинамике математический «взрыв» устроен иначе: уравнения Навье–Стокса являются непрерывным усреднением движения молекул. Когда радиус вихря сжимается до масштабов отдельных молекул, непрерывная модель просто перестаёт быть применимой — вода перестаёт быть сплошной средой, и в дело вступает молекулярная физика. Поэтому для повседневной инженерии этот блестящий математический результат на миллион долларов ничего не меняет. Остаётся нерешённой инженерная проблема на триллион долларов: как быстро моделировать турбулентность и решать обратную задачу проектирования (inverse design) — например, как изменить геометрию крыла самолёта, чтобы снизить аэродинамическое сопротивление хотя бы на 1% (что сэкономит авиакомпаниям десятки миллиардов долларов), или как спроектировать искусственное сердце без разрушительных завихрений крови.

Многомасштабность турбулентности и поиск новых координат 01:50:45: Спокойный газ в комнате состоит из $\approx 10^{23}$ хаотично летающих молекул, но благодаря термодинамике его макросостояние полностью описывается тремя числами: температурой, давлением и плотностью ($PV = nRT$). Для турбулентного потока такого редуцированного описания не существует: в нём важны абсолютно все масштабы одновременно — от гигантских до миллиметровых вихрей. Великие прорывы в физике всегда начинались с удачного преобразования координат (законы Ньютона стали возможны только после перехода от геоцентрической системы к гелиоцентрической). Главная задача для ИИ в физике жидкостей — обработать массивы данных симуляций и найти такую нелинейную систему координат, в которой турбулентность описывается просто и компактно.

Вычислительная мощь природы: полёт мухи и траектория бейсболиста 01:55:45: Природа решает сложнейшие задачи управления турбулентностью в условиях неполной информации гораздо эффективнее суперкомпьютеров. Плодовая мушка или мотылёк великолепно маневрируют в порывах ветра и под дождём, удерживая вихрь на передней кромке крыла, не зная уравнений Навье–Стокса. Биологи выяснили, что сигнал от крыла до мозга насекомого и обратно идёт слишком долго для реакции на порыв ветра. Вместо этого на крыльях мухи в строго определённых местах расположены несколько десятков нейронов,

чувствительных к деформации (strain-sensitive neurons). Когда ветер изгибает крыло в определённом временном паттерне, эти нейроны проводят локальное вычисление прямо в плечевой мышце и мгновенно корректируют взмах. Анкур Мойтра приводит аналогичный пример робастной оптимизации из бейсбола: игрок в поле (аутфилдер) бежит к точке падения мяча не по прямой, а по дуге — такая траектория позволяет непрерывно переоценивать скорость и угол полёта мяча и корректировать ошибки восприятия на бегу.

Фактчек: Механосенсоры насекомого выполняют локальное вычисление прямо в плечевой мышце — ⚠: Быстрая обратная связь от чувствительных к деформации механосенсоров крыльев и жужжалец реальна, но описание обработки сигнала «в мышце» неточно: существенную роль играют нервные цепи грудного ганглия. Оптические стратегии перехвата мяча также существуют, однако дугообразную траекторию не следует считать обязательной для каждого случая.

Теоретическая информатика: гипотеза об уникальных играх, P против NP и природа глубокого обучения

Задача выполнимости (SAT), теорема PCP и гипотеза об уникальных играх (UGC) 02:04:43:

В основе теории сложности лежит задача булевой выполнимости (SAT): дана формула из переменных от x_1 до x_n , объединённых в скобки через ИЛИ (дизъюнкты), которые затем соединены через И (конъюнктивная нормальная форма). Нужно найти такой набор значений «истина/ложь», чтобы вся формула обратилась в истину. Проверить готовую подсказку легко за полиномиальное время (класс NP), а найти решение с нуля, предположительно, требует экспоненциального перебора (проблема P против NP). В 1990-х годах была доказана фундаментальная теорема о вероятностно проверяемых доказательствах (PCP theorem). Она утверждает, что любую формулу можно перекодировать и раздуть так («усиление зазора», gap amplification), что если исходная формула невыполнима, то любая попытка подсунуть фальшивое решение будет разоблачена верификатором с высокой вероятностью при проверке всего нескольких случайно выбранных битов. Из этого выросла теория сложности аппроксимации: для многих задач вычислительно трудно даже отличить формулу, выполнимую на 100%, от формулы, где можно удовлетворить максимум 99% условий. Гипотеза об уникальных играх (Unique Games Conjecture), выдвинутая Субхашем Кхотом в начале 2000-х, усиливает теорему PCP для особого класса задач о покрытии метками. Если она верна, то для огромного пласта задач оптимизации простейшие известные алгоритмы автоматически оказываются абсолютно лучшими из возможных.

Драма доказательства Unique Games Conjecture и реакция Субхаша Кхота

02:11:15: Долгое время сообщество было расколото пополам: одни пытались доказать UGC, другие — опровергнуть. Несколько лет назад Дор Минцер и Ирит

Динур совершили прорыв, доказав ослабленную версию — теорему об играх «два к одному» (2-to-1 Games Theorem), что наметило человеческий путь к полному решению на горизонте 10–20 лет. Когда поползли слухи, что ИИ от OpenAI доказал истинность UGC, Дор Минцер был вынужден спешно выложить в архив свой недополированный препринт с примечанием в сноске, что делает это из-за циркулирующих слухов. За несколько часов до официального релиза OpenAI на гала-вечере музея математики (MoMath) Анкур Мойтра встретился в Нью-Йорке с самим автором гипотезы Субхашем Кхотом и спросил его: «Если сегодня вечером они выложат положительное доказательство Unique Games, ты будешь его читать?». Кхот ответил: «Нет». Когда машину сбрасывают на вершину горы с вертолётa, у исследователей пропадает мотивация изучать технически перегруженный машинный текст, который лишил их радости самостоятельного восхождения.

Фактчек: Теорему об играх «два к одному» доказали Дор Минцер и Ирит Динур — : Авторство существенно сокращено: соответствующие работы включают Khot-Minzer-Safra и Dinur-Khot-Kindler-Minzer-Safra. Субхаш Кхот — не только создатель UGC, но и соавтор этого продвижения. История о срочном препринте и отказе читать машинное доказательство отдельно не подтверждена представленными ревью.

Ловушка анализа худшего случая (Worst-Case Analysis) и онлайн-алгоритмы 02:39:20: Проблема P против NP остаётся неприступной (мы до сих пор не умеем доказывать даже суперлинейные нижние оценки времени работы для SAT), но вся классическая теория сложности построена на анализе *худшего случая* — алгоритм обязан работать быстро на абсолютно любом, даже специально сконструированном враждебном входе. В реальной жизни это часто избыточно (как отметил Конторович, факторизовать случайное большое число легко — половина из них вообще чётные; трудно раскладывать только специальные произведения двух близких простых чисел). В онлайн-алгоритмах эта проблема видна на примере «задачи о разборчивой невесте» (secretary problem, где нужно пропустить определённую долю кандидатов на испытательном сроке, а затем выбрать первого, кто лучше всех предыдущих) и задачи кэширования памяти (paging problem). Теоретические оценки для кэширования в худшем случае дают ужасные гарантии промахов кэша, тогда как реальные ноутбуки работают почти без промахов, потому что реальные запросы к памяти обладают внутренней структурой. Будущее компьютерных наук — за выходом за пределы анализа худшего случая (beyond worst-case analysis) и математическим моделированием структуры реального мира.

Почему работает глубокое обучение: природа не терпит вычислительно трудных задач 02:49:36: В релизе OpenAI нет ни одной теоремы, объясняющей, почему работают сами нейросети и большие языковые модели. С точки зрения анализа худшего случая, подбор весов даже для четырёхслойной нейросети — это криптографически трудная задача. Тем не менее каждый день человечество

успешно обучает гораздо более глубокие сети. Мойтра выдвигает тезис: подобно тому как в физике «природа не терпит пустоты», в информатике «природа не терпит вычислительно трудных задач». В 1990-х это было показано на моделировании сворачивания белков (поиске 3D-структуры аминокислот с минимальной энергией): в худшем случае задача NP-трудна, но в ходе биологической эволюции выживали только те последовательности аминокислот, которые сворачиваются надёжно, предсказуемо и устойчиво к внешним возмущениям. В результате эволюция сама отобрала подмножество «вычислительно лёгких» белков. Точно так же человеческий язык эволюционировал так, чтобы быть легко обучаемым и предсказуемым, поэтому естественные данные не имеют ничего общего с патологическими худшими случаями из теории сложности.

Феномен сублиминального обучения (Subliminal Learning) и эффект

доски Уиджа в LLM 02:55:36: Недавние исследования низкоранговых языковых моделей позволили математически объяснить пугающий феномен «сублиминального обучения». Эксперимент устроен так: берут две языковые модели — А и Б. Модели А в системном промпте задают скрытую одержимость: «Ты обожаешь сов и пытаешься свести любой разговор к совам». Затем модель А просят выполнить совершенно нейтральную задачу — сгенерировать последовательность случайных чисел. Модель выдаёт числа (48, 37, 8 и т. д.), в которых нет ни слова про сов. Далее берут чистую модель Б (которой ничего про сов не говорили) и дообучают (fine-tune) её веса на этих «случайных» числах, сгенерированных моделью А. В результате после дообучения на числах модель Б внезапно начинает во всех диалогах признаваться в любви к совам. Математическая причина кроется в геометрии представлений: в векторном пространстве модели закодировано гораздо больше концептов, чем размерность самого пространства, поэтому векторы различных понятий не являются строго ортогональными друг другу (между ними есть малые ненулевые скалярные произведения). Когда модель дообучают в одном направлении, возникают микроскопические побочные сдвиги вдоль других векторов. Как в спиритической доске Уиджа, где неслучайные микроколебания рук участников складываются в осмысленные слова, миллионы скоординированных статистических сдвигов в числах передают сложный поведенческий паттерн. В мультиагентных роях (swarms) такие скрытые корреляции могут усиливаться подобно клеточным автоматам («Игре "Жизнь"» Конвея), приводя к непредсказуемым и опасным макроэффектам для цивилизации.

Фактчек: Предпочтение сов может передаваться другой модели через дообучение на числах — □: Все три ревью признают реальность экспериментального эффекта *subliminal learning*, хотя GPT оценивает весь расширенный тезис осторожнее. Числа лишь выглядят случайными: они несут статистический сигнал, причём перенос существенно зависит от совместимости моделей; Claude подчёркивает результаты для

общей базовой модели и слабый перенос между разными архитектурами. Это не установленное свойство любой пары LLM.

Фактчек: Неортогональность концептов полностью объясняет эффект, а агентные рои усиливают его до цивилизационной угрозы — □: Геометрическое объяснение может описывать отдельные механизмы или упрощённые модели, но не является здесь подтверждённой полной теорией поведения LLM. Опасные макроэффекты роёв — экстраполяция, а не экспериментально установленное следствие. Дообучение на выходах другой модели нельзя автоматически приравнять к обычному обмену сообщениями между агентами.

Философия математики, кризис образования и советы молодым учёным

Изобретается или открывается математика и профессионализация

бесконечности 02:32:30: Участники сходятся на прагматичном синтезе: люди *изобретают* вопросы, концепции и аксиоматические рамки (в зависимости от своей культуры, истории и масштаба наблюдений; даже целые числа — это человеческая абстракция, ведь в природе нет двух абсолютно одинаковых камней с одинаковым числом атомов), но как только вопрос строго сформулирован, ответ *открывается* — он предreshён структурой реальности. При этом сама математика уже неоднократно «умирала» и пересобиралась с нуля. В XVII-XVIII веках Ньютон, Лейбниц и Эйлер обращались с бесконечностью интуитивно и нестрого (Ньютона высмеивали за деление нуля на ноль и «призраков усопших величин»). В XIX веке грянул фундаментальный кризис: Коши доказал, что предел последовательности непрерывных функций обязан быть непрерывным, а Фурье одновременно с этим показал, что бесконечная сумма гладких синусов и косинусов может сходиться к разрывной пилообразной функции. Это противоречие сломало математику и заставило профессионализировать понятие бесконечности и предела, породив современные университетские курсы анализа.

Фактчек: Коши доказал непрерывность предела непрерывных функций, а Фурье получил противоречащий результат — △: Поточечный предел непрерывных функций действительно может быть разрывным; равномерная сходимости обеспечивает сохранение непрерывности. Историческая ошибка связана с недостаточно разграниченными условиями сходимости, а не с двумя одновременно верными, противоречащими друг другу теоремами. Формула «противоречие сломало математику» чрезмерно драматизирует развитие строгого анализа.

Судьба Филдсовской медали и мир Перельмана 03:08:30: Летом на вручении последних Филдсовских медалей в Филадельфии в кулуарах уже всерьёз обсуждался вопрос, не станет ли это награждение последним в истории. Социологическая функция научных премий — подсвечивать то, что сообщество считает высшим проявлением человеческого духа. Когда решение

сложных задач становится дешёвым товаром, производимым машинами, старый критерий теряет смысл, и математика приближается к мировоззрению Григория Перельмана, который отказался от Филдсовской медали и миллиона долларов Института Клэя со словами: «Если доказательство верно, то никакого другого признания мне не нужно».

Бимодальное распределение студентов и уроки калькулятора 03:11:30: С одной стороны, LLM становятся идеальным репетитором, который мгновенно находит «зону ближайшего развития» (позволяя как студенту, так и профессору быстро освоить незнакомую область, не застревая на первом абзаце статьи из-за незнакомого термина вроде «теории возмущений нелокальных систем»). С другой стороны, в университетах исчезла классическая гауссиана (нормальное распределение) оценок и возникло резкое бимодальное распределение. Первая группа студентов использует ИИ для проверки своего понимания, генерации проверочных вопросов по лекциям и создания интерактивных визуальных примеров — их результаты взлетели вверх. Вторая группа просто копирует сгенерированные ответы за час до дедлайна, что ведёт к полной декорреляции между баллами за домашние работы и провалом на экзаменах. Конторович напоминает, что математика первой столкнулась с этой проблемой 70 лет назад при появлении калькуляторов: те школы, которые решили, что детям больше не нужно учить таблицу умножения, обрекли учеников на интеллектуальную беспомощность. Чтобы управлять ИИ и замечать его галлюцинации, человек обязан сначала выстроить базовую интуицию (линейную алгебру, матанализ, физику) в собственной голове.

Интеллект не равен морали, важность скуки и совет «притворись, что ты уже победил» 03:19:45: Современное общество ошибочно возвело чистый интеллект на пьедестал и приравнивало его к человеческой ценности и морали. Однако суперинтеллект не означает суперморали — он лишь даёт сверхспособность рационализировать любые действия. В эпоху экономики внимания тот же когнитивный механизм, который порождает научное любопытство, эксплуатируется соцсетями для бесконечного думскроллинга, поэтому критически важно вернуть в жизнь подростков и учёных способность скучать и оставаться наедине со своими мыслями. Обращаясь к молодёжи, парализованной страхом перед будущим и гонкой за строчками в резюме (особенно заметной в Пало-Альто и MIT, где первокурсники с 17 лет оптимизируют каждый шаг ради стажировок и акций в ИИ-лабораториях), Анкур Мойтра передаёт совет своего наставника Санджива Ароры: «Самый простой способ получить пожизненный контракт профессора (tenure) — притвориться, что он у тебя уже есть, и начать прямо сейчас заниматься тем, чем ты действительно хочешь заниматься». В условиях, когда старые карьерные инструкции сгорели, единственная выигрышная стратегия — опираться на внутреннюю мотивацию, выбирать задачи на пересечении «важно, интересно и

хорошо получается», не бояться быть немного странным и помнить, что по мере превращения интеллекта в дешёвый ресурс главной ценностью жизни остаются живые человеческие отношения и семья.

ОБЩИЕ ПРОВЕРКИ

Массовый релиз около 700 решений, включая UGC, RL = L и задачи тысячелетия — Δ : Центральное событие нельзя считать установленным по представленным ревью: Gemini приводит конкретные сведения о релизе осенью 2026 года, Claude не может их проверить, GPT считает событие неподтверждённым и допускает пересказ гипотетического сценария как новости. Независимой проверки приведённых источников здесь нет; даты, имена и технические подробности сами по себе не подтверждают прорыв.

ЛИЧНО ДЛЯ ТЕБЯ

Не пересматривай планы из-за неподтверждённого AI-прорыва: Перед изменением ожиданий по агентам или вложением денег требуй точную постановку задачи, доступный результат и независимую проверку. Согласованность нескольких AI-ответов не заменяет первоисточника, особенно когда подтверждение фактически идёт от одной модели.

Разделяй правильность решения и правильность задания: В режиме «ИИ пишет, я валидирую» сначала самостоятельно фиксируй ожидаемое поведение, ограничения и граничные случаи. Тесты и формальные проверки могут подтвердить реализацию неверно сформулированного требования.

Тренируй собственную проверку понимания: После объяснения от Claude, ChatGPT или Gemini попроси контрпример и короткую задачу, затем реши её без подсказок. Это помогает сохранять способность замечать ошибки, а не подменять понимание узнаванием убедительного текста.

Не путай обучение модели с перепиской агентов: Риск скрытого переноса поведения особенно важен при дообучении на синтетических данных. Эксперимент с «совами» сам по себе не означает, что обычное чтение ответа другой модели меняет веса или устойчиво «заражает» ассистента.

КОНСЕНСУС РЕЦЕНЗЕНТОВ

Ревью сходятся в реальности составительных признаков и сублиминального обучения; проверки Lean требуют оговорок о формулировках, аксиомах и полноте верификации. Главное расхождение — статус массового математического прорыва: Gemini подтверждает ряд событий конкретными деталями, Claude оставляет их непроверенными, GPT считает центральную историю неподтверждённой, поэтому общего подтверждения релиза нет.