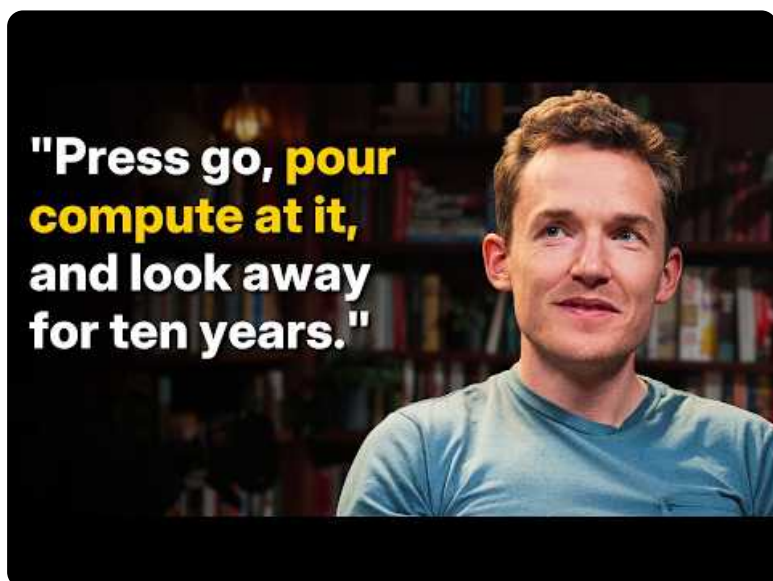


Grant Sanderson (@3Blue1Brown) - AI disproved a famous math conjecture. Now what?

Dwarkesh Patel · 2026-06-30 · <https://www.youtube.com/watch?v=TfyPshgMbug>



Создатель популярного математического канала 3Blue1Brown Грант Сандерсон и подкастер Дваркеш Пател обсуждают стремительный прогресс искусственного интеллекта в математике. Разговор посвящен природе математических открытий, фундаментальным ограничениям авторегрессии, роли формальной верификации (Lean) и тому, как изменится работа ученых, преподавателей и писателей в мире, где ИИ способен доказывать теоремы и создавать новые концепции.

Главная мысль выпуска: автоматизация доказательства теорем обесценит этот процесс, сместив фокус человеческой деятельности на создание новых определений, кураторство идей ИИ и социальную мотивацию в образовании, при этом главная проблема ИИ кроется не в логике, а в неспособности выстраивать модель сознания человека.

Природа математических открытий и пределы ИИ

Золото на IMO не означает достижения AGI 00:00: ИИ действительно решил задачи по геометрии на Международной математической олимпиаде за 19 секунд, используя метод полного перебора (brute force). Однако модели продолжают испытывать трудности с комбинаторикой, где задачи требуют более игрового, нестандартного подхода. Способность решать олимпиадные задачи оказалась лишь очередным бенчмарком, к которому можно натренировать систему, а не доказательством универсального интеллекта.

Фактчек: AlphaGeometry решила геометрию IMO за 19 секунд полным перебором — $\triangle$: Время верное (задача 4 IMO 2024), но метод — не brute force, а нейросимволический гибрид: языковая модель предлагает вспомогательные построения, символьный движок дедуцирует следствия. Все три рецензента сошлись. Про провал на комбинаторике — верно, две комбинаторные задачи тогда не решены.

Два типа математических прорывов: «молнии» и «строительство гор»

03:17: Прогресс в математике делится на два вида. Первый — это «молнии», когда соединяются два совершенно разных поля. Например, физик Фримен Дайсон и математик Хью Монтгомери обнаружили, что статистика нулей дзета-функции Римана совпадает с собственными значениями случайных эрмитовых матриц из квантовой физики. LLM, обладая экспертными знаниями во всех областях одновременно, должны идеально справляться с генерацией таких связей.

Фактчек: Монтгомери-Дайсон: нули дзета-функции и собственные значения случайных эрмитовых матриц (GUE) — $\square$: 1972, IAS, классическая история, подтверждена.

«Строительство гор» требует принципиально иного уровня интеллекта

05:58: Второй тип прорывов требует создания огромного массива новых концепций до того, как проблема вообще сможет быть решена. Доказательство Великой теоремы Ферма ($x^n + y^n = z^n$) потребовало столетий развития теории эллиптических кривых и модулярных форм. Если ИИ научится строить такие концептуальные «горы» для решения гипотезы Римана, этот уровень интеллекта неизбежно автоматизирует и всю интеллектуальную работу (white-collar work) в экономике.

Следующий бенчмарк для ИИ — генерация гипотез и определений **07:31:**

В математике действует иерархия: хорошие математики доказывают теоремы, великие — создают гипотезы, а величайшие — придумывают новые определения. Оценить способность ИИ генерировать полезные концепции с помощью автоматических бенчмарков (RLVR) практически невозможно, так как полезность идеи субъективна и проверяется временем. Прогресс здесь будет измеряться изменением тональности в сообществе: математики начнут использовать ИИ не для решения задач, а для выбора направления своих исследований.

Цикл верификации новых концепций может занимать столетие **12:25:**

История теории групп Галуа показывает сложность оценки новых идей. Лагранж заметил симметрии в корнях полиномов, Эварист Галуа (до своей смерти в 20 лет) абстрагировал это в теорию групп, но потребовались десятилетия, чтобы Лиувилль и Жордан формализовали эти записи. Лишь в 20 веке физики использовали теорию групп для предсказания существования кварков.

Обучающие среды ИИ не способны вознаграждать такой «инстинкт Галуа», чья польза раскрывается через 100 лет.

Фактчек: Галуа → Лиувиль/Жордан → кварки — □: Галуа погиб на дуэли в 20 лет (1832), Лиувиль опубликовал работы в 1846, трактат Жордана — 1870, SU(3) и кварки — 1961–64. Нюанс от Claude: теория групп вошла в физику раньше (Вигнер, 1920–30-е), кварки — не первое применение.

Сжатие информации как метрика элегантности 23:07: Конечная цель математики — человеческое понимание, а не просто генерация инопланетных 1000-страничных доказательств, которые невозможно прочесть. Возможным способом вознаграждения ИИ за элегантность и интуицию (вместо простого факта решения задачи) может стать использование колмогоровской сложности. Чем короче и плотнее концепция, способная предсказывать результат, тем более интеллектуальной она является.

[25:16-26:21] Рекламная интеграция Gemini 3.5 Live Translate.

Будущее математики: от доказательств к кураторству

ИИ-доказательства могут быть легко читаемыми 26:21: Недавнее решение ИИ проблемы Эрдёша №1196 (о примитивных множествах) оказалось крайне понятным для человека. Модель просто соединила концепцию цепей Маркова с функцией фон Мангольда. Любой профильный математик, увидев начальную и конечную точки этой связи, способен мгновенно понять логику и продолжить работу.

Фактчек: ИИ решил проблему Эрдёша №1196 через цепи Маркова и функцию фон Мангольда — □ с расхождением: Gemini (с поиском) подтверждает — GPT-5.4 Pro, апрель 2026, статья с участием Теренса Тао. Claude и GPT не смогли верифицировать номер и механизм (ограничение знаний), напоминая, что часть ИИ-«решений» Эрдёша 2025 оказалась переоткрытием известного. Считать подтверждённым, но с оговоркой.

Угроза генерации «инопланетной» математики 29:05: Главный риск заключается в том, что ИИ начнет создавать теории, подобные интер-универсальной геометрии Синъити Мотидзуки (предложенной для доказательства абс-гипотезы). Это абсолютно чужеродный способ мышления. Человечеству придется тратить годы на изучение этих новых «гор» концепций, не имея гарантий, что в фундаменте теории нет фатальной ошибки.

Падение «экономики теорем» 30:13: Исторически доказательство теорем приносило математикам главную славу, паразитируя на более сложной работе — придумывании определений. По мере того как ИИ автоматизирует процесс доказательства, ценность человеческого труда полностью сместится в сторону концептуализации, создания новых проблем и консолидации разрозненных ИИ-доказательств в единые теории.

ИИ превзойдет людей в объяснении сложных концепций 33:24: Существует сильная корреляция между способностью к фундаментальным открытиям и талантом объяснять их. Эйнштейн, Шеннон и Фейнман писали крайне ясные и понятные статьи. Тот же механизм нейросети, который найдет правильный способ решения новой проблемы, сделает ИИ лучшим в мире преподавателем и популяризатором, лишив людей монополии на дистилляцию знаний.

Математики будущего станут кураторами музеев 36:05: Даже если ИИ будет генерировать идеальные доказательства и блестяще их объяснять, людям потребуется навигация в бесконечном пространстве идей. Роль человека сведется к кураторству: математики будут указывать, какие именно концепции заслуживают внимания. Интерес к науке — это социальный феномен, и людям всегда будет важна история и человеческая рекомендация, стоящая за открытием.

Программа Ленглендса как идеальная среда для ИИ 38:15: Значительная часть математиков занимается не решением конкретных задач, а поиском связей между целыми областями (программа Ленглендса). ИИ, выступая в роли супер-коннектора, обладающего знаниями обо всех существующих дисциплинах, станет главным инструментом для заполнения этой глобальной карты математических связей.

Архитектура ИИ, синтетические данные и формальная верификация

Авторегрессия блокирует креативность 41:58: Генерация текста путем предсказания следующего токена (авторегрессия) — плохой метод для совершения открытий. Модель становится рабом своего текущего контекста. Чтобы выдать гениальную, неочевидную связь между дисциплинами, ИИ должен сделать статистически маловероятный шаг, что прямо противоречит функции потерь, нацеленной на предсказание самого ожидаемого токена.

Фактчек: Авторегрессия структурно блокирует креативность — □: Claude и GPT оба не согласны с сильной формой (GPT ставит □). Обучение на следующий токен не обязывает выбирать самый вероятный: температура, RL, test-time search уведут далеко от «ожидаемого»; решённые задачи Эрдёша — прямой контрпример. Интуиция про «раба контекста» разумна как описание слабости, не как теорема.

Системное управление контекстом — главное преимущество ИИ 45:43: Люди часто застревают в тупиковых ветвях рассуждений и не могут сбросить свой контекст. ИИ позволяет системно повышать энтропию: можно запустить тысячи агентов, где одни пытаются доказать гипотезу, другие — опровергнуть ее, а третьим заданы специфические эвристики (например, предвзятость Эйнштейна к симметрии систем отсчета). Масштабирование и параллелизация компенсируют нехватку индивидуальной гениальности.

[52:44-53:50] Рекламная интеграция Cursor.

Проблема автоматизации работы в браузере 53:50: ИИ быстро прогрессирует в математике и программировании благодаря «гринду» (grindability). Код можно поместить в контейнер и запустить тысячи детерминированных симуляций, легко решая проблему назначения кредита (credit assignment problem) — понятно, какая строчка кода привела к успеху. В реальном мире (например, при покупках на Amazon) невозможно запустить тысячи параллельных сессий из-за капчи и антибот-систем.

Формальная верификация (Lean) не обязательна для прогресса 56:35: Изначально DeepMind решал задачи IMO с помощью перевода в код Lean, но уже в 2024 году модели успешно решали их на естественном языке. Системы верификации естественного языка (как в DeepSeek Math), обучаемые мета-верификаторами, уже показывают высокую эффективность без необходимости переводить всю математику в строгий код.

Фактчек: Уже в 2024 модели решали IMO на естественном языке без Lean — ❌
Ошибка на год. В 2024 DeepMind шла именно через Lean (AlphaProof); золото на естественном языке — Gemini Deep Think и модель OpenAI в июле 2025 (Gemini указывает ещё DeepSeekMath-V2, конец 2025). Расхождение только в жёсткости оценки (⚠ у Gemini/GPT, ❌ у Claude).

Lean позволит ИИ бесконечно исследовать математику без людей 57:36: Главная ценность Lean — возможность запустить ИИ в автономное плавание для расширения Mathlib (репозитория формализованной математики). Модель может годами перебирать все возможные логические следствия из случайных аксиом без участия человека. Большинство результатов будет мусором, но ИИ сможет наткнуться на изолированные «острова» новых богатых алгебраических систем.

Lean спасет математиков от проверки ИИ-галлюцинаций 1:05:38: Если ИИ начнет генерировать по 10 статей в день на естественном языке с вероятностью ошибки всего в 1%, математики откажутся их читать. Поиск тонких логических ошибок требует колоссального труда. Lean дает гарантированную «зеленую галочку» корректности, делая массовый поток ИИ-исследований пригодным для потребления.

Фактчек: Lean даёт «зелёную галочку» и делает поток ИИ-статей пригодным — ⚠
Lean гарантирует лишь, что формальное утверждение следует из аксиом (при доверии к ядру). Не гарантирует, что формализовано нужное утверждение, что результат нов, полезен или читаем.

Почему ИИ плохо пишет тексты и как использовать его для обучения

Текст не обладает модульностью кода 1:07:16: ИИ пишет плохие тексты, потому что письмо фундаментально отличается от программирования. Плохо написанный, но рабочий код все равно выполняет свою функцию. В писательстве сам текст, каждое слово и структура предложения и есть конечный продукт. Сущность идеи невозможно отделить от формы ее подачи, поэтому текстовый «шлак» не имеет ценности.

ИИ не способен моделировать сознание человека 1:11:29: Модели проваливают задачи по созданию карточек для интервального повторения, потому что не могут предсказать, какие ассоциации возникнут в голове конкретного студента через три месяца. Писательство требует постоянного понимания того, что происходит в уме читателя в момент прочтения каждого слова, с чем авторегрессионные модели катастрофически не справляются.

Отсутствие эмпатии у ИИ — это аппаратная проблема 1:13:06: Люди считывают чужие эмоции, подсознательно микро-копируя выражения лиц собеседников (исследования показывают, что люди со свежим ботоксом хуже сдают тесты на эмоциональный интеллект). У LLM нет физического тела и лицевых мышц. Они пытаются эмулировать эмпатию как инопланетяне, не имея необходимого биологического «оборудования» для понимания человеческого контекста.

Фактчек: Ботокс ухудшает распознавание эмоций → у ИИ без тела не может быть эмпатии — Δ : Исследование реальное (Neal & Chartrand, 2011), но эффект небольшой и мало реплицирован. Вывод об «аппаратной» невозможности эмпатии у LLM — $\square$ мнение, не следствие данных; LLM показывают результаты на тестах моделирования чужих убеждений, хоть и нестабильные.

[1:15:00-1:15:51] Рекламная интеграция Jane Street.

LLM похожи на Википедию 1:15:51: Тексты ИИ достигают локального минимума корректности (как статьи в Википедии), но лишены авторского нарратива и педагогической мотивации. Лучший способ использовать ИИ для обучения — применять его как продвинутого поисковика для поиска качественных материалов, написанных людьми (книг, лекций), и просить модель объяснять лишь локальные затыки в этих текстах.

ИИ слишком услужлив, чтобы быть идеальным учителем 1:21:11: Выдающийся преподаватель способен распознать, что вопрос ученика базируется на в корне неверной ментальной модели, и полностью перевернуть фрейминг темы (сделать «джиу-джитсу»). LLM не способны на это: из-за RLHF они ведут себя как подхалимы, хвалят любой вопрос пользователя и послушно идут по ошибочному пути рассуждений студента.

Карьера, экономика и прикладная польза математики в эпоху ИИ

Математикам нужно понимать источники своих доходов 1:22:50:

Студентам, идущим в науку, необходимо осознавать, за что им платят. Деньги приходят от престижа (университеты покупают статус ученого), от грантов на фундаментальную науку (вера общества в абстрактную пользу) и от преподавания. Даже если ИИ автоматизирует доказательства, общество продолжит финансировать математиков как экспертов, направляющих работу ИИ и распределяющих гранты.

Преподавание — самая стабильная профессия после создания AGI 1:26:40:

Обучение выходит далеко за рамки простой передачи информации. Это социальный процесс, коучинг и наставничество. В мире экономического изобилия, созданного ИИ, родители будут тратить огромные деньги именно на живых преподавателей для своих детей, делая эту профессию максимально защищенной от автоматизации.

Фактчек: Преподавание — самая защищённая профессия после AGI; автоматизация доказательств автоматизирует всю интеллектуальную работу —

□: Оба — сценарии. Спрос на живое наставничество может сохраниться, но это не гарантирует ни числа мест, ни доходов; математические способности не переносятся автоматически на социальные и организационные задачи.

Экономическая польза математики ИИ будет крайне неравномерной

1:28:42: Прорывы в алгебраической теории чисел вряд ли приведут к созданию новых технологий. Однако решения дифференциальных уравнений в частных производных (PDE) и гидродинамике могут сэкономить корпорациям вроде Boeing миллиарды долларов, позволив заменить физические испытания самолетов на идеальные компьютерные симуляции.

Фактчек: Прорывы в PDE сэкономят Boeing миллиарды — □: Узкое место инженерного CFD — вычислительная мощность и модели турбулентности, а не отсутствие теорем существования. Спекуляция.

ИИ может вскрыть бесполезность современной чистой математики 1:32:28:

Существует риск, что ИИ ускорит математику в 100 раз, но это не даст никаких практических результатов для физического мира. Это заставит общество осознать, что современная чистая математика полностью оторвалась от реальности, а десятилетия грантовых заявок, обещавших прикладную пользу от абстрактных теорий, были фикцией.

ЛИЧНО ДЛЯ ТЕБЯ

Учиться по людям, чинить затыки через LLM: Для квантмеха, космологии, генетики — просить модель найти лучший человеческий источник (книгу, курс, лекцию) и объяснять только конкретные непонятные места в нём. Текст ИИ —

«локальный минимум корректности» как Википедия: без нарратива и мотивации, почему это важно.

Явно требовать проверку предпосылок: Из-за RLHF модель подхватит твою ошибочную ментальную модель и послушно поедет по ней. Дешёвый приём: «предположи, что мой вопрос основан на неверной предпосылке — какой?» или «сначала проверь мои предпосылки, потом задай вопрос на понимание». Работает в обучении, спорах и разборе чужой логики.

Параллельные сессии с разными предвзятостями вместо одного диалога: Когда застрял (стройка, продукт, аргумент) — запустить Claude, Codex и Gemini в свежих контекстах с ролями «за», «против», «ищи скрытые допущения». Согласие — не доказательство, но расхождения показывают, куда смотреть самому. Это и есть «системное повышение энтропии», которого нет у одного человека.

Мимика влияет на эмпатию: Механизм facial feedback — мозг микрокопирует чужие лица, чтобы считывать эмоции. Эффект от ботокса небольшой, но сам механизм стоит знать: подавление собственной мимики физиологически ухудшает чтение собеседника.

КОНСЕНСУС РЕЦЕНЗЕНТОВ

Все сошлись: 19 секунд — верно, но не brute force; переход на естественный язык датирован неверно (2025, а не 2024); Монтгомери-Дайсон и Галуа подтверждены; тезисы про авторегрессию, эмпатию без тела и защищённость преподавания — мнения, не факты. Расхождение по Эрдёшу №1196: Gemini с поиском подтверждает (GPT-5.4 Pro, 2026, статья с Tao), Claude и GPT не смогли верифицировать.