

Demis Hassabis: Future of AI, Simulating Reality, Physics and Video Games | Lex Fridman Podcast #475

Lex Fridman · 2025-07-23 · <https://www.youtube.com/watch?v=-HzgcbRXUK8>



Беседа с CEO Google DeepMind и лауреатом Нобелевской премии по химии Демисом Хассабисом, а также сольный блок ответов на вопросы от Лекса Фридмана. Выпуск посвящен фундаментальным вопросам науки, пределам вычислений и архитектуре будущих ИИ-систем.

Главная мысль: Вселенная представляет собой информационную систему, в которой любые природные паттерны, прошедшие эволюционный отбор, обладают внутренней структурой. Эта структура позволяет эффективно моделировать сложнейшие физические и биологические процессы (от фолдинга белков до гидродинамики) с помощью классических нейросетей, избегая комбинаторного взрыва. Это открывает путь к решению проблемы $P=NP$, созданию «виртуальной клетки», открытию новых материалов для энергетики и достижению AGI к 2030 году.

Моделирование реальности, физика и проблема $P=NP$

Любой природный паттерн можно эффективно смоделировать классическим алгоритмом обучения 02:06: Системы вроде AlphaFold и AlphaGo работают в пространствах с колоссальной комбинаторной сложностью (10^{300} вариантов структуры белка, 10^{170} позиций в го). Прямой перебор занял бы больше времени, чем существует Вселенная. Однако природа не случайна: геологические формы, орбиты планет и биологические молекулы прошли через эволюционный отбор («выживание самых стабильных»). Эта скрытая структура

формирует низкоразмерное многообразие, которое нейросети могут выучить для эффективного поиска решения без квантовых вычислений.

Фактчек: Пространство конформаций белка $\sim 10^{300}$, легальных позиций в го $\sim 10^{170}$ — □: Числа корректны — парадокс Левинталя даёт оценку 10^{300} , для го известно $\sim 2,1 \times 10^{170}$ позиций (больше атомов во Вселенной, $\sim 10^{80}$). Проверял только Gemini, остальные не оспаривали.

Вопрос $P=NP$ является фундаментальной физической проблемой 06:18:

Вселенная — это в первую очередь информационная система, где информация первичнее энергии и материи. Успехи ИИ доказывают, что классические машины Тьюринга способны за полиномиальное время решать задачи, которые ранее считались требующими квантовых систем. Если природные системы имеют структуру, их динамику можно смоделировать, что делает поиск следующего шага или решения вычислительно выполнимым (трактабельным).

Фактчек: Успехи AlphaFold/AlphaGo доказывают моделируемость природы классическими алгоритмами и открывают путь к $P=NP$ — △: Claude и GPT независимо указали на подмену понятий: нейросети решают структурированные природные экземпляры, а P vs NP — про худший случай на произвольных входах; никакого продвижения к проблеме тут нет. Интуиция о низкоразмерном многообразии сама по себе корректна. GPT ставит жёсткое □, Claude мягче (□/△).

ИИ способен интуитивно понимать физику без жестких уравнений 12:35:

Традиционное моделирование гидродинамики (уравнения Навье-Стокса) для прогноза погоды или рендеринга требует огромных суперкомпьютерных мощностей. Однако генеративная видеомодель Veo 3 способна крайне реалистично симулировать поведение жидкостей, материалов и зеркального освещения (например, прозрачную жидкость под гидравлическим прессом) просто за счет наблюдения за видео на YouTube. Модель извлекает скрытую структуру физических законов, формируя «интуитивную физику» на уровне понимания ребенка, без необходимости физического воплощения (робототехники) для познания мира.

Фактчек: Veo 3 выучила «интуитивную физику» жидкостей и материалов по видео с YouTube — △: Расхождение. Gemini подтверждает — визуально модель симулирует жидкости и освещение реалистично. Claude и GPT возражают: правдоподобие $\neq$ физическая корректность; собственный бенчмарк DeepMind Physics-IQ (2025) даёт видеомоделям, включая Veo, лишь $\sim 20\text{--}25\%$ на предсказании физических исходов. Модель учит «как это выглядит», не законы.

Видеоигры как симуляторы миров и будущее развлечений

Будущее видеоигр — это полностью генеративные открытые миры 19:18:

Создание бесконечного количества игровых ассетов вручную для AAA-игр экономически невозможно. В перспективе 5-10 лет ИИ-системы смогут динамически генерировать сюжет, окружение и реакцию симуляции на любые,

даже самые нестандартные действия игрока. Это будет интерактивная версия моделей вроде Veo, где игрок становится соавтором уникального нарратива, а иллюзия выбора (как в современных играх) сменится истинной свободой действий.

Фактчек: AlphaProof взял серебро на IMO — □: Верно для июля 2024 (28/42 баллов, 4 из 6 задач, в связке с AlphaGeometry 2). Устарело: к выходу выпуска (июль 2025) Gemini Deep Think уже взял золото (35/42) без формализации.

Игры — идеальный полигон для тестирования ИИ и симуляции жизни

23:29: Исторически самые передовые технологии (ИИ, графика, физические движки, GPU) развивались благодаря игровой индустрии. Ранние эксперименты с обучением с подкреплением (например, в игре Black & White, где существо перенимало моральный компас игрока) заложили базу для современных AGI-систем. Игры вроде Civilization (любимая игра Демиса Хассабиса) служат микрокосмами реального мира, позволяя людям безопасно тренировать навык принятия сложных решений и конструктивно канализировать трайбалистские инстинкты, снижая вероятность реальных войн.

Фактчек: WeatherNext точнее и на порядки быстрее гидродинамических моделей — □: GraphCast (Science 2023) обошёл ECMWF HRES на ~90% метрик, GenCast (Nature 2024) — ансамбль ENS на 97%; минуты на TPU против часов на суперкомпьютере. Оговорка (Claude, GPT): модели обучены на ERA5 и стартуют с ассимилированных данных ECMWF — паразитируют на традиционной инфраструктуре, а не заменяют её; слабее на локальных экстремумах. Дата «сентябрь 2026» в ревью Gemini — очевидная ошибка.

Эволюционные алгоритмы, AlphaEvolve и виртуальная клетка

Гибридные системы объединяют LLM и эволюционный поиск **31:23:**

Система AlphaEvolve использует большие языковые модели для генерации гипотез (предложения кода), а эволюционные алгоритмы — для поиска в новых, неисследованных областях пространства решений. В отличие от традиционных эволюционных методов из 90-х, интеграция с фундаментальными моделями позволяет системе генерировать эмерджентные свойства, подобно тому, как биологическая эволюция создала сложные организмы из простых бактерий.

Фактчек: AlphaFold решил структуру, AlphaFold 3 моделирует динамику, виртуальная клетка ускорит биологию в 100x — △: AlphaFold 3 предсказывает статические структуры комплексов (белок-РНК-ДНК-лиганды), а не временную динамику. Виртуальная клетка и «100x» — исследовательская цель и прогноз, не результат.

ИИ пока не обладает «исследовательским вкусом» **38:09:** Современные системы могут решать сложнейшие математические задачи (AlphaProof берет серебро на Международной олимпиаде), но не способны придумать новую

гипотезу уровня гипотезы Римана или теории относительности. Главный навык великого ученого — выбрать правильный вопрос и поставить эксперимент так, чтобы он разделил пространство гипотез ровно пополам. В таком случае даже отрицательный результат дает максимум информации. ИИ пока не способен на такой творческий скачок.

Фактчек: AGI с вероятностью ~50% к 2030 — □: Субъективная оценка заинтересованного CEO, не измерение и не консенсус. Общепринятого операционального определения AGI нет; «изобрести игру уровня го» или «теорию уровня Эйнштейна» — предложенные, а не принятые критерии.

Проект «Виртуальная клетка» ускорит биологические открытия в 100 раз

42:20: Конечная цель применения ИИ в биологии — полная симуляция одноклеточного организма (например, дрожжей). AlphaFold решил проблему статической 3D-структуры белков, AlphaFold 3 моделирует динамику взаимодействий (белок-белок, РНК, ДНК), а AlphaGenome предсказывает влияние мутаций на функции. Следующий шаг — моделирование целых метаболических путей (например, пути TOR, связанного с раком). Симуляция потребует иерархической архитектуры, так как процессы в клетке идут на радикально разных временных шкалах, но опускаться до квантового уровня не придется — достаточно белкового.

Фактчек: Термояд, сверхпроводники при комнатной температуре и «бесплатная энергия» за ~5 лет; Кардашев I — □: Реальные заслуги ограничены: управление плазмой в токамаке TCV (Nature 2022), GNoME предсказал 2,2 млн кристаллов (синтезированы сотни, прорывных нет). До коммерческого термояда — пропасть в инженерии, капексе и регуляторике; энергия не станет бесплатной даже при работающем синтезе.

ИИ может смоделировать зарождение жизни 46:29: Переход от неживой материи к живой — это процесс поиска в комбинаторном пространстве первичного бульона. Задав начальные условия химической среды ранней Земли, ИИ-симуляция способна найти тот самый «Ход 37» (гениальное неочевидное решение), который привел к формированию первой клетки, доказав, что между физикой, химией и биологией нет жесткой границы, а есть единый континуум.

Прогнозирование погоды и путь к AGI

Нейросети превосходят суперкомпьютеры в прогнозировании погоды

50:39: Система WeatherNext от Google DeepMind моделирует сложную, почти хаотическую динамику атмосферы точнее и на порядки быстрее традиционных гидродинамических моделей. Это позволяет оперативно предсказывать траектории ураганов и циклонов, спасая жизни.

Фактчек: Мозг — классический компьютер, квантовых эффектов нейробиология не находит — □: Оценки декогеренции (Тегмарк: 10^{-13} – 10^{-20} с) на 10+ порядков короче

нейронных процессов; Пенроуз-Хамерофф маргинальны. Но это консенсус, не доказательство: работы 2024 о суперрадиации в микротрубочках существуют, когнитивной релевантности не показано.

Вероятность создания AGI к 2030 году составляет около 50% 52:42:

Истинный AGI должен обладать когнитивными функциями человеческого мозга без «зубчатого» профиля интеллекта (когда система гениальна в одном и глупа в другом). Проверка AGI потребует прогона через десятки тысяч когнитивных тестов и передачи системы сотням лучших мировых экспертов (уровня Теренса Тао) на несколько месяцев для поиска уязвимостей.

Маркером появления AGI станет создание новых концепций 55:20: Главным доказательством достижения AGI будет не просто решение задач, а генерация принципиально новых идей: создание физической теории уровня Эйнштейна (если ограничить знания модели 1900 годом) или изобретение новой настольной игры, столь же элегантной и глубокой, как го. Эксперты-люди смогут понять эти открытия постфактум, так же как гроссмейстеры понимают гениальные ходы ИИ после их анализа.

Рекурсивное самосовершенствование ИИ имеет пределы 1:00:25: Системы вроде AlphaEvolve способны на инкрементальное улучшение (например, ускорение матричного умножения), двигаясь вверх по текущей S-образной кривой. Однако сценарий «жесткого взлета» (hard takeoff) маловероятен, так как для перехода на новый уровень требуются фундаментальные архитектурные прорывы (подобные изобретению Трансформеров в 2017 году), на которые текущие алгоритмы автоматического поиска пока не способны.

Масштабирование, энергия и радикальное изобилие

Законы масштабирования продолжают работать на всех этапах 1:03:32:

Прогресс ИИ обеспечивается масштабированием пре-трейна, пост-трейна и вычислений во время инференса (test-time compute). Потребность в вычислительных мощностях будет расти экспоненциально, так как модели внедряются в продукты для миллиардов пользователей, а новые парадигмы «мыслящих систем» требуют больше времени на обдумывание ответа.

ИИ решит проблему чистой энергии 1:08:13: В ближайшие 5 лет ИИ внесет критический вклад в энергетику: от оптимизации охлаждения дата-центров и управления электросетями до проектирования реакторов термоядерного синтеза (удержание плазмы) и открытия новых материалов (сверхпроводники при комнатной температуре, сверхэффективные солнечные панели, новые типы батарей).

Человечество перейдет к цивилизации I типа по шкале Кардашева

1:10:18: Решение энергетической проблемы (термояд и солнечная энергия) запустит эру радикального изобилия. Бесплатная энергия позволит массово

опреснять воду, синтезировать бесконечное ракетное топливо (водород и кислород из морской воды) и начать добычу ископаемых на астероидах. Это снимет проблему нехватки ресурсов — главную причину человеческих конфликтов, открыв путь к экспансии в космос.

Разработка Gemini, продукты и интерфейсы будущего

Разработка ИИ требует слияния исследовательской культуры и продуктовой скорости 1:18:32: Google DeepMind функционирует как гигантский стартап внутри корпорации. Переход от отставания к лидерству (от Gemini 1.5 к 2.5) стал возможен благодаря объединению команд Google Brain и DeepMind, безжалостному отсечению бюрократии и культуре «непрерывного прогресса и непрерывного релиза».

Проектирование ИИ-продуктов идет на опережение технологий 1:25:17: Из-за колоссальной скорости развития моделей интерфейсы нужно проектировать не под текущие возможности, а под те, что появятся через 6-12 месяцев. Текущий формат текстовых чат-ботов скоро устареет. Будущее за мультимодальными интерфейсами с высокой пропускной способностью (голос, умные очки, нейроинтерфейсы), где взаимодействие будет напоминать совместный творческий процесс из фильма «Особое мнение».

Смена версий моделей (например, на Gemini 3.0) — это результат масштабных сборок 1:29:01: Каждые полгода лучшие архитектурные идеи и подходы к данным собираются в единый гигантский «геройский прогон» (hero training run) базовой модели. Затем на этапе пост-трейна создаются различные версии (Pro, Flash, Flash-Lite) путем дистилляции, чтобы закрыть весь фронт Парето — баланс между производительностью, стоимостью и задержкой для разных задач разработчиков.

Бенчмарки полезны, но финальное слово за пользователями 1:32:42: Оптимизация моделей — это многокритериальная задача. Улучшение навыков программирования не должно ухудшать способности к переводу или написанию стихов (no regret improvements). Помимо академических тестов (LM SYS, математика, код), критически важны «вайбы» — стиль, лаконичность, юмор и персонализация характера модели под конкретного пользователя.

Рынок труда, экономика и геополитика ИИ

Гонка зарплат в ИИ-индустрии вторична по сравнению с миссией 1:40:33: Компании вроде Meta скупают таланты за огромные деньги, пытаясь догнать лидеров. Однако лучшие умы индустрии мотивированы не деньгами, а возможностью работать на самом фронтире науки и безопасно внедрять AGI в мир. В пост-AGI экономике изобилия само понятие денег и текущая структура компаний кардинально изменятся.

Программисты не исчезнут, но их продуктивность вырастет в 10 раз

1:43:06: Программирование и математика легче всего поддаются автоматизации, так как для них легко генерировать и верифицировать синтетические данные. В ближайшие 5-10 лет разработчики, освоившие ИИ-инструменты, станут сверхлюдьми. Рутинный код (например, фронтенд) будет генерироваться машинами, а ценность человека сместится в сторону архитектурного проектирования, постановки задач и оценки высокопроизводительных систем.

ИИ вызовет шок, в 10 раз превышающий Промышленную революцию

1:46:14: Изменения произойдут в 10 раз быстрее и будут в 10 раз масштабнее. Это потребует внедрения безусловного базового дохода (UBI) или базового обеспечения услугами. Политические системы и институты управления должны эволюционировать, чтобы справедливо распределять ресурсы радикального изобилия и не допускать социальной поляризации.

Разработка AGI должна идти по модели CERN, а не Манхэттенского проекта

1:56:10: Милитаризация ИИ и гонка вооружений между государствами крайне опасны. Идеальный сценарий — международный научный консорциум, где лучшие умы мира совместно и прозрачно завершают создание AGI, обеспечивая его безопасность до глобального развертывания.

Риски, сознание и природа человека

Оценка P(doom) не имеет научного смысла из-за неопределенности

1:58:16: Точные цифры вероятности гибели человечества от ИИ — фикция. Однако этот риск точно больше нуля и не пренебрежимо мал. Учитывая колоссальные ставки (от вымирания до бессмертия и экспансии в космос), единственная рациональная стратегия — «осторожный оптимизм» и десятикратное увеличение финансирования исследований в области безопасности ИИ.

Угрозы делятся на злонамеренных людей и автономные системы 2:00:53:

Первая проблема — использование мощных open-source моделей террористами или странами-изгоями для создания биологического или ядерного оружия. Вторая проблема — потеря контроля над агентными ИИ-системами по мере их приближения к AGI. Решение потребует международных стандартов и использования самого ИИ для раннего обнаружения угроз.

Мозг — это классический компьютер, а не квантовый 2:06:06: Вопреки теориям Роджера Пенроуза о квантовой природе сознания, нейробиология не находит доказательств квантовых эффектов в мозге. Это означает, что все когнитивные процессы можно смоделировать на классических кремниевых чипах.

Понимание ИИ требует «радикальной эмпатии» 2:07:39: Люди признают друг друга сознательными, потому что демонстрируют схожее поведение и работают на одном биологическом субстрате. ИИ на кремнии лишит нас второго аргумента. Возможно, только прямые нейроинтерфейсы (типа Neuralink) позволят человеку буквально «почувствовать», каково это — обрабатывать информацию на небιологическом носителе, и решить трудную проблему сознания.

[2:11:48-2:12:19] смолток, благодарности и переход к сольному формату ответов на вопросы (AMA).

Lex Fridman AMA: Философия, наука и личная биография

Истинная мудрость — находить смысл в обыденности 2:12:19: В знаменитой речи Дэвида Фостера Уоллеса «Это вода» (This is Water) подчеркивается, что самые важные реалии жизни труднее всего заметить. Главная духовная битва происходит не на ретритах, а в рутине взрослой жизни. Секрет счастья — быть «нескучающим» (unboring), осознавая бесконечную глубину и красоту каждого момента, не поддаваясь высокомерию собственных убеждений.

Наука не убивает красоту, а приумножает ее 2:16:05: Ричард Фейнман утверждал, что научный взгляд на цветок не лишает его эстетики. Понимание клеточной структуры, внутренних процессов и эволюционной причины появления цвета (для привлечения насекомых) лишь добавляет благоговения перед тайной природы.

Лекс Фридман является действующим научным сотрудником MIT 2:20:48: В ответ на ложные обвинения в сети, Фридман уточняет факты своей биографии. С 2015 года по сегодняшний день он занимает оплачиваемую должность научного сотрудника (Research Scientist) в Лаборатории информационных систем и принятия решений (LIDS) при MIT. Первые 4 года он работал по 80-100 часов в неделю, публикуя статьи. С 2020 года фокус сместился на подкаст, но позиция в MIT сохраняется.

Фактчек: Лекс Фридман — оплачиваемый Research Scientist в MIT LIDS с 2015 —
△: Аффiliation с MIT в справочнике есть, но в LIDS он с 2022, до этого — AgeLab. Факт оплаты и объём позиции публично не верифицируемы; в академической среде статус регулярно критикуют как номинальный.

Образование и инженерная база получены в Университете Дрекселя 2:25:32: Фридман получил степени бакалавра, магистра и доктора наук (PhD) в Drexel University. Там он прошел жесткую школу теоретической информатики, программирования на C/C++, создания систем компьютерного зрения, робототехники и алгоритмов оптимизации. Этот опыт, наряду с многолетними тренировками по джиу-джитсу, сформировал его подход к тяжелой работе, интеллектуальной скромности и любви к инженерии.

ОБЩИЕ ПРОВЕРКИ

Продуктивность программистов вырастет в 10x — $\triangle$: Расхождение. GPT приводит контрпример: RCT METR 2025 — опытные разработчики в знакомых репозиториях с ИИ тратили на ~19% больше времени. Claude и Gemini считают тренд смещения ценности к архитектуре и постановке задач реальным и уже наблюдаемым. Итог: направление верное, множитель «10x» — маркетинговая цифра.

ЛИЧНО ДЛЯ ТЕБЯ

Правдоподобно $\neq$ верно: Главный вынос из кейса Veo (реалистичная картинка при 20–25% на физическом бенчмарке) распространяется на любой выход нейросети — код, медицинские выкладки, механизмы добавок. Твой режим «читаю и валидирую логику» — единственно правильный, и его цена растёт с каждым релизом, а не падает. Меряй скорость до проверенного результата, а не до появления кода; сравни Claude, Codex и Gemini на 3–4 повторяющихся типах задач с учётом времени на чтение и правки.

Хороший эксперимент делит пространство гипотез пополам: Перед сменой добавки, протокола или тренировки формулируй, какой результат (включая отрицательный) реально изменит решение. Если никакой — это ритуал, не эксперимент. И не переноси прогресс молекулярного ИИ (AlphaFold, «виртуальная клетка», путь mTOR) на доказанность вмешательств: предсказанное связывание с белком $\neq$ клинический исход у человека.

Дисконтируй таймлайны, не горизонт: «AGI 50% к 2030» и «термояд за 5 лет» — заинтересованные прогнозы, но смещение ценности разработчика в декомпозицию задач, верификацию и вкус уже происходит с тобой. Вкладывая время в навыки, которые не дешевеют, а не в то, что модель делает лучше каждые полгода.

Генеративные миры и «This is Water»: Полностью генеративные открытые миры в 5–10 лет — в 4X-жанре это конец скриптовой иллюзии выбора, стоит следить как игроку. И 20 минут на речь Уоллеса «Это вода» окупятся: умение быть «нескучающим» в рутине стройки и отладки — рабочий фреймворк для психики под нагрузкой.

КОНСЕНСУС РЕЦЕНЗЕНТОВ

Все трое сошлись на достоверности числовых фактов ($10^{300}/10^{170}$, IMO-серебро, превосходство GraphCast/GenCast с оговоркой про зависимость от инфраструктуры ECMWF) и на сомнительности статуса Фридмана в MIT. Разошлись по Veo 3 (Gemini принял визуальную симуляцию за подтверждение, Claude/GPT — за иллюзию физики) и по «10x продуктивности» (GPT привёл METR

с –19%, Claude/Gemini считают тренд реальным); Claude и GPT дополнительно единодушно отделили прогнозы CEO (AGI-2030, термояд, Кардашев) от фактов, чего Gemini не сделал.