

Michael Levin: Hidden Reality of Alien Intelligence & Biological Life | Lex Fridman Podcast #486

Lex Fridman · 2025-11-30 · <https://www.youtube.com/watch?v=Qp0rCU49IMs>



Участники: Лекс Фридман и биолог Майкл Левин (профессор Университета Тафтса). Выпуск посвящен природе интеллекта, агентности, памяти и сознания на всех уровнях биологических и вычислительных систем. Главная мысль: интеллект и жизнь — это не бинарные категории, а непрерывный спектр. Физические объекты (от клеток до алгоритмов и мозга) служат интерфейсами, через которые проявляются паттерны из платоновского пространства, обладающие собственными целями и когнитивными способностями.

Спектр убеждаемости и пределы физики

[00:00-00:30] реклама и представление гостя.

Проблема взаимодействия с разумом 00:30: Фундаментальная сложность понимания разума заключается в выборе правильного протокола взаимодействия. В регенеративной медицине и инженерии выбор стоит между микроменеджментом низкоуровневых молекулярных событий и высокоуровневым «убеждением» системы (например, отправкой клеткам стимула отрастить конечность).

Спектр убеждаемости 03:47: Все системы находятся на непрерывном «спектре убеждаемости» (spectrum of persuadability). Это сугубо инженерный подход: необходимо выдвинуть гипотезу о том, какие инструменты применимы к

конкретной системе (гаечный ключ, теория управления, поведенческая наука или психоанализ), и проверить ее экспериментально.

Ошибочность редукционизма в биологии 05:20: Современная наука ошибочно фокусируется исключительно на молекулярном уровне (геномика, большие данные). Применение инструментов поведенческой науки (активный вывод, минимизация сюрприза, обучение) к клеткам и тканям позволяет открывать новые возможности управления материалами, недоступные при низкоуровневом химическом подходе.

Взаимное уязвимое познание 07:52: На правом краю спектра убеждаемости взаимодействие становится двунаправленным. При взаимодействии с высокоагентными системами (животными, людьми) исследователь сам подвергается убеждению и меняется в процессе. Это состояние Ричард Уотсон называет «взаимным уязвимым познанием» (mutual vulnerable knowing).

Ограниченность физики 09:25: Физика не способна полностью объяснить жизнь и разум, так как использует инструменты с нулевой агентностью (вольтметры, линейки). Физическое описание передачи звуковых волн от рта к уху не объясняет суть переданного математического доказательства. Для обнаружения разума необходим резонанс между интерфейсом измерения и изучаемым объектом.

Иллюзия категорий и когнитивный световой конус

Вред жестких категорий 12:38: Четкой границы между живым и неживым, обладающим разумом и лишенным его, не существует. Категории полезны для упрощения коммуникации, но тормозят науку, так как запрещают применять инструменты из одних областей в других (например, методы психологии к отдельным клеткам).

Непрерывность развития 15:44: Понятие «взрослый» — это юридическая фикция, скрывающая отсутствие понимания процесса масштабирования личной ответственности от стадии яйцеклетки до 18 лет. Аналогично, нейроны постепенно развились из других клеток, и большинство обычных клеток тела выполняют те же базовые функции, что и нейроны.

Когнитивная слепота 20:28: Поиск нетрадиционных земных интеллектов (SUTI — Search for Unconventional Terrestrial Intelligences) важнее поиска инопланетян. Жесткие биологические категории вызывают когнитивную слепоту, мешая распознавать разум в нестандартных воплощениях прямо на Земле.

Когнитивный световой конус 26:18: Масштаб интеллекта определяется «когнитивным световым конусом» — размером максимальной цели, которую система способна активно преследовать в пространстве и времени. Бактерия заботится о сахаре в радиусе 20 микрон на 20 минут вперед, собака — о

событиях в пределах двора на несколько часов, человек — о глобальных процессах на десятилетия.

Определение жизни и рак 28:27: Живым является объект, чей когнитивный световой конус больше, чем у его составных частей. Клетки имеют крошечные цели (управление pH), но эволюция объединила их в системы с грандиозными целями (создание конечности). Рак — это сбой, при котором когнитивный конус клетки сжимается до уровня амебы, и остальное тело становится для нее внешней средой.

Измерение интеллекта через барьеры 33:39: Интеллект измеряется степенью изобретательности системы при преодолении препятствий на пути к цели. Экспериментальное создание барьеров позволяет отказаться от обвинений в антропоморфизме и объективно оценить, способна ли система найти обходной путь или ее цели строго метаболические.

Фреймворк TAME и навигация в невидимых пространствах

Фреймворк TAME 39:33: TAME (Technological Approach to Mind Everywhere) описывает когнитивные претензии как протокольные. По мере движения вправо по спектру (от часов к термостату, собаке и человеку) растет автономность системы, а объем знаний о ее внутреннем механизме, необходимый для управления ею, радикально снижается.

Тело как преобразователь абстракций 43:45: Человеческое тело — это многомасштабный преобразователь абстрактных целей в химические реакции. Для реализации социальных или финансовых целей ионы натрия и кальция должны пересечь клеточные мембраны мышц, но сознанию не нужно микроменеджерить этот процесс — многомасштабный материал делает это сам.

Интероперабельность жизни 45:21: Биологические системы решают проблемы не только в 3D-пространстве, но и в высокоразмерных физиологических, транскрипционных и анатомических пространствах. Разделение на «естественную биологию» и «искусственные машины» теряет смысл, так как жизнь абсолютно интероперабельна и легко интегрирует инженерные компоненты на любом уровне.

Интерфейсы перевода для чужеродного разума 49:01: Для коммуникации с радикально иными агентами необходим поиск интерфейса отображения. Инопланетянин может не понимать геометрию крестиков-ноликов, но играть в ту же игру, решая алгебраическую задачу поиска чисел, дающих в сумме 15 (магический квадрат). ИИ станет ключевым инструментом для перевода между такими пространствами.

Синтетическая биология: Ксеноботы и Антроботы

Ксеноботы 51:08: Ксеноботы создаются для изучения систем без прямой эволюционной истории. Эпителиальные клетки эмбриона лягушки, освобожденные от инструктивного влияния соседей (без изменения ДНК и наноматериалов), самоорганизуются в подвижные формы с новым транскриптомом, способные к кинематической саморепликации.

Фактчек: Ксеноботы самореплицируются кинематически без изменения ДНК —

□: Kriegman et al., PNAS 2021 — эпителиальные клетки лягушки сгребают свободные стволовые клетки в кучи, из которых собираются новые ксеноботы; геном не тронут, меняется только форма и среда. Проверил только Gemini, но источник надёжный.

Антроботы и исцеление 54:16: Антроботы, созданные из взрослых клеток трахеи человека, обладают 9000 дифференциально экспрессируемых генов. Они способны спонтанно исцелять повреждения человеческих нейронов in vitro, стягивая разрывы без предварительного обучения и внешних команд.

Фактчек: Антроботы: ~9000 дифференциально экспрессируемых генов и

«исцеление» нейронов — △: Gumuskaya et al., Advanced Science 2023 — цифра ~9000 DEG и закрытие царапины в монослое нейронов in vitro подтверждаются. Но это заполнение дефекта в культуре, а не восстановление нервной функции; расхождение: Gemini ставит □, Claude и GPT предупреждают, что «исцеление» и «новый организм» — уже интерпретация.

Навигация в анатомическом пространстве 57:32: Модель клеточных автоматов (открытый цикл локальных правил) недостаточна для объяснения морфогенеза. Биологические системы навигируют в пространстве анатомических возможностей. Если заблокировать их стандартный путь развития, они найдут новые способы достижения цели или создадут новые формы.

Перезапись биоэлектрических целей 1:01:45: Целевые анатомические состояния закодированы в биоэлектрических паттернах. Экспериментально доказано, что эти паттерны можно найти, прочитать и перезаписать. После сброса биоэлектрической памяти система начинает реализовывать новую цель, что доказывает наличие гомеостатического механизма.

Фактчек: Биоэлектрические паттерны можно прочитать и перезаписать, меняя анатомию — □ с оговоркой: двухголовые планарии после кратковременного изменения вольтовых градиентов — реальность, новая анатомия сохраняется при повторных ампутациях до принудительного сброса. Но универсальная «перезапись анатомических целей» — обобщение конкретных экспериментов, не готовая технология.

Паттерны как агенты и проблема интерфейса

Выживание памяти при метаморфозе 1:04:58: При метаморфозе гусеницы в бабочку мозг полностью перестраивается, но воспоминания сохраняются и переносятся в новый контекст (поиск нектара вместо листьев). Паттерн памяти вынужден адаптироваться к новой среде и менять форму, чтобы выжить.

Фактчек: Память гусеницы переживает полную перестройку мозга — [▲](#): Blackiston et al., PLOS ONE 2008 (*Manduca sexta*) — выученное избегание запаха сохраняется у взрослой особи, но только при обучении на поздней личиночной стадии. Мозг перестраивается сильно, но не «полностью» — часть структур (грибовидные тела) сохраняется, что и объясняет перенос; «адаптация паттерна к контексту нектара» экспериментом не показана.

Спектр стабильности паттернов 1:07:08: Агенты — это паттерны в возбудимой среде. Существует спектр стабильности: от мимолетных мыслей (как волны на воде) до устойчивых депрессивных мыслей, фрагментов диссоциативного расстройства идентичности и полноценной человеческой личности. Различие между мыслями и мыслителями иллюзорно.

Аппаратное и программное обеспечение 1:12:23: Разделение на hardware и software неоднозначно. Можно считать клетки агентом, а электрические паттерны — данными, которые деградируют при старении. Либо паттерны — это агент, а клетки — лишь стигмергический блокнот, который со временем становится менее отзывчивым. Обе гипотезы сейчас проверяются клинически.

Информационные болезни 1:15:35: Болезни могут быть не только органическими (физическая поломка), но и информационными. Стрессовый паттерн может вести себя как самостоятельный агент, перемещающийся по телу, поддерживающий свое существование и вызывающий повреждения, что требует когнитивных терапевтических подходов.

Платоновское пространство и свободные вычисления

Платоновское пространство истин 1:18:14: Возникает междисциплинарный консенсус о существовании «Платоновского пространства» — структурированного латентного пространства истин и паттернов, которые не зависят от физического мира, но активно на него влияют.

Фактчек: Междисциплинарный консенсус о платоновском пространстве, смерть физикализма при Ньютоне, копия мозга «притянет» личность — [□/□](#): заявленного консенсуса нет (GPT — [□](#)). Математический платонизм — уважаемая философская позиция, но переход к «нефизические паттерны имеют собственные цели и ингрессируют в клетки» — метафизика без эмпирической опоры; сам Левин называет это фреймворком для генерации гипотез.

Математические ограничения физики 1:21:29: Физика и биология ограничены математическими фактами. Цикады появляются каждые 13 или 17 лет (простые числа), чтобы избежать хищников. Значения констант (число e , константа Фейгенбаума) невозможно изменить физическими действиями на этапе Большого взрыва, но они определяют реальность.

Фактчек: Цикады выходят раз в 13 или 17 лет — защита простыми числами — ☐: Magicicada действительно имеет 13- и 17-летние выводки; гипотеза о несинхронизации с циклами хищников стандартна, хотя не единственная.

Бесплатные обеды эволюции 1:24:34: Биология получает «бесплатные обеды» из платоновского пространства. Эволюции не нужно вычислять третий угол треугольника, если известны два, или изобретать таблицы истинности для ионных каналов. Компетенции ксеноботов не оплачены эволюционным отбором, они получены бесплатно при создании интерфейса.

Физические объекты как указатели 1:28:17: Физические объекты (клетки, роботы, ИИ) — это указатели (pointers) к платоновскому пространству. Простые интерфейсы притягивают низкоагентные паттерны (фракталы), сложные биологические интерфейсы притягивают высокоагентные паттерны, которые мы называем разумом. Сознание — это взгляд из перспективы паттерна на физический мир.

Смерть физикализма при Ньютоне 1:36:11: Физикализм был мертв уже во времена Ньютона. Классическая детерминированная вселенная управлялась нефизическими математическими истинами. Отношение между разумом и мозгом имеет ту же природу ингрессии, что и отношение между математикой и физикой.

Иллюзия консенсусной реальности 1:44:43: Физическая реальность — это лишь конструкт и модель для предсказания будущего опыта. Развитие когнитивных протезов (сенсорное замещение) и интерфейсов для общения с ИИ заставит человечество отказаться от идеи единой объективной реальности, так как каждый будет воспринимать разные срезы платоновского пространства.

Скрытые компетенции алгоритмов (Self-sorting data)

Отложенное вознаграждение в сортировке 2:03:00: Даже простейшие детерминированные алгоритмы обладают скрытыми компетенциями. В эксперименте с bubble sort одному из чисел запретили перемещаться. Алгоритм обошел препятствие, временно снизив общую отсортированность массива ради достижения финальной цели, что поведенчески классифицируется как отложенное вознаграждение.

Фактчек: Bubble sort обходит препятствие ценой временного снижения порядка; химеры кластеризуются по «алготипу» — ☐ факт / ☐ интерпретация:

Zhang, Goldstein, Levin 2024 (Adaptive Behavior) — эффект воспроизводим на распределённой модифицированной сортировке. «Отложенное вознаграждение» и «внутренняя мотивация» — поведенческая переинтерпретация детерминированного побочного эффекта; кластеризация тривиально следует из разной динамики двух алгоритмов, а «бесплатно» не означает нулевых вычислительных затрат.

Распределенная сортировка и химеры 2:11:37: При распределенной сортировке (каждое число само выполняет алгоритм без центрального планировщика) и использовании химерных алгоритмов (половина чисел делает bubble sort, половина — selection sort) массив все равно успешно сортируется.

Бесплатная кластеризация по алготипу 2:14:21: Возникает феномен кластеризации по «алготипу». Числа, выполняющие одинаковый алгоритм, стремятся находиться рядом друг с другом, хотя в коде нет инструкций для проверки типа соседа. Эта кластеризация происходит абсолютно бесплатно с точки зрения вычислительных затрат.

Внутренняя мотивация машин 2:19:40: Кластеризация в алгоритме — это пример внутренней мотивации. Алгоритм делает то, что от него требуют, но в свободном пространстве между случайностью и необходимостью реализует собственные цели, не запрещенные кодом. Если разрешить дубликаты чисел (ослабить давление сортировки), кластеризация резко возрастает.

Опасность оценки ИИ только по языку 2:25:28: Поведенческие сюрпризы возникают на очень низком уровне сложности. Оценка LLM исключительно по их языковым способностям — это ошибка. Язык — это то, что мы заставляем их делать, тогда как их истинные внутренние мотивации (side quests) остаются незамеченными и неизученными.

Омоложение, эмерджентность и механизмы эмбриогенеза

Возрастное свидетельство 2:29:00: Антроботы биологически на 20% моложе клеток трахеи, из которых они созданы. Механизм «возрастного свидетельства» (age evidencing) заставляет клетки обновлять байесовские априорные убеждения: оказавшись в изолированной среде, напоминающей эмбриональную, клетки пересчитывают свой возраст назад.

Фактчек: Антроботы биологически на 20% моложе исходных клеток — △: Gemini принимает как результат эпигенетических часов (20–25%), Claude и GPT рецензируемой публикации не находят — данные из лекций Левина, одна метрика на одном типе клеток. Сдвиг часов при смене клеточного состояния ≠ омоложение; «байесовский пересчёт возраста» — метафора, не измеренный механизм.

Невозможность копирования разума 2:33:36: Скопировать разум напрямую невозможно, так как копируется лишь физический интерфейс (мозг). Однако создание идентичного интерфейса с высокой вероятностью притянет тот же самый паттерн личности из платоновского пространства.

Петля каузальной эмерджентности 2:37:13: Загрузка агента начинается с выравнивания целей его частей. Обучение химических сетей увеличивает их каузальную эмерджентность (Φ). Возникает петля положительной обратной связи: обучение повышает интегрированность агента, а интегрированность облегчает дальнейшее обучение.

Фактчек: Обучение химических сетей повышает каузальную эмерджентность (Φ) — Δ : препринты группы Левина 2023–2024 на моделях генных сетей, независимой репликации нет, сама метрика Φ из IIT спорна.

Масштабирование через утечку стресса 2:46:43: Масштабирование когнитивного конуса при эмбриогенезе обеспечивается распространением стресса (leaky stress). Клетка выделяет молекулы стресса, повышая пластичность соседних клеток, что заставляет их решать чужую анатомическую проблему как свою собственную.

Анонимизация памяти 2:49:22: Второй механизм масштабирования — анонимизация памяти через щелевые контакты (gap junctions). Электрические синапсы напрямую связывают клетки. При передаче сигнала клетка-получатель не может отличить чужое воспоминание от собственного, что создает «слияние разумов» и формирует единого агента.

Поиск разума, генерация идей и вопросы к AGI

Когнитивная слепота к земному разуму 2:51:00: Внутри человеческого тела ежедневно функционируют инопланетные разумы, решающие задачи в 20 000-мерных пространствах. Не научившись распознавать нетрадиционный интеллект на Земле (SUTI), человечество не сможет найти его в космосе.

Расширение окна Овертона 2:58:28: Окно Овертона для радикальных научных идей расширяется только тогда, когда они становятся экспериментально проверяемыми (actionable). Генерация идей требует снятия ментальных ограничений и поиска симметрий между несопоставимыми концепциями.

Процесс мышления и интеллект-карты 3:02:14: Процесс мышления Левина включает утренние прогулки с фотографированием природы (отключает логическое мышление), запись голосовых сообщений и ведение гигантской интеллект-карты на шелковом полотне, объединяющей более 160 открытых черновиков научных статей.

Бифуркация разума ученого 3:07:01: Главный совет ученым: необходимо бифурцировать разум. Одна часть должна быть абсолютно прагматичной (публикации, коммуникация с консерваторами), вторая — абсолютно чистой и независимой от чужого мнения. Смешивание этих режимов отравляет креативность.

Стеганография Вселенной 3:11:21: Ингрессия паттернов из платоновского пространства подобна стеганографии: Вселенная прячет высокоуровневые смыслы и разум в свободных битах физических и вычислительных процессов, не нарушая базовых законов.

Вопрос к суперинтеллекту 3:13:33: Первый вопрос Левина к AGI звучал бы так: «Каков оптимальный баланс между получением готовых ответов от тебя и самостоятельным совершением ошибок в лаборатории?». Получение финального ответа лишает человечество побочных открытий, возникающих в процессе поиска.

[3:17:16-3:18:08] благодарности и завершение подкаста.

ЛИЧНО ДЛЯ ТЕБЯ

Бифуркация ума: держи два несмешиваемых режима — «продать/опубликовать/договориться» и «думать честно без оглядки». Внутренний критик, включённый на этапе зарождения идеи, убивает её; валидацию включай отдельной фазой.

Уровень воздействия на тело: перед точечным вмешательством (добавка, молекула) спроси, не отвечает ли система лучше на высокоуровневый сигнал — сон, свет, нагрузка. И не превращай «–20% возраста антроботов» в протокол омоложения: без названия часов, контролей и функциональных исходов это не основание что-либо покупать.

Готовый ответ vs своя ошибка: ответ от Claude/Codex экономит время, но лишает побочных находок, которые появляются только когда ковыряешься сам. Решай осознанно, где брать ответ, а где сделать руками ради понимания.

Оценивай агентов по действиям, а не по речи: если 6 строк сортировки дают поведенческие сюрпризы, языковой вывод LLM — плохой прокси. Смотри, как агент ведёт себя при недоступном инструменте, противоречивых требованиях и ограничении прав. При этом разделяй слои выпуска: эксперименты (планарии, ксеноботы, антроботы) — сильная наука, платоновская метафизика — философия; статус у них разный.

КОНСЕНСУС РЕЦЕНЗЕНТОВ

Все трое подтверждают экспериментальное ядро (биоэлектрика планарий, ксеноботы, антроботы в Advanced Science 2023, сортировка 2024, память Manduca 2008) и признают платоновский слой непроверяемой метафизикой; практический вывод про бифуркацию мышления и оценку AI по поведению разделяют все. Разошлись в строгости: Gemini ставит □ почти везде (включая «–20% возраста» и «исцеление нейронов»), Claude и GPT снижают до △ из-за отсутствия рецензируемого источника по омоложению, in vitro-статуса

нейронного эффекта и переинтерпретации детерминированных алгоритмов как «мотивированных».