ИИ-газета ДЛЯ АЛЕКСЕЯ № 1 (ТЕСТОВЫЙ) СРЕДА, 4 НОЯБРЯ 2026 ФОРМАТ ДНЯ: РАЗБОР НЕДЕЛИ

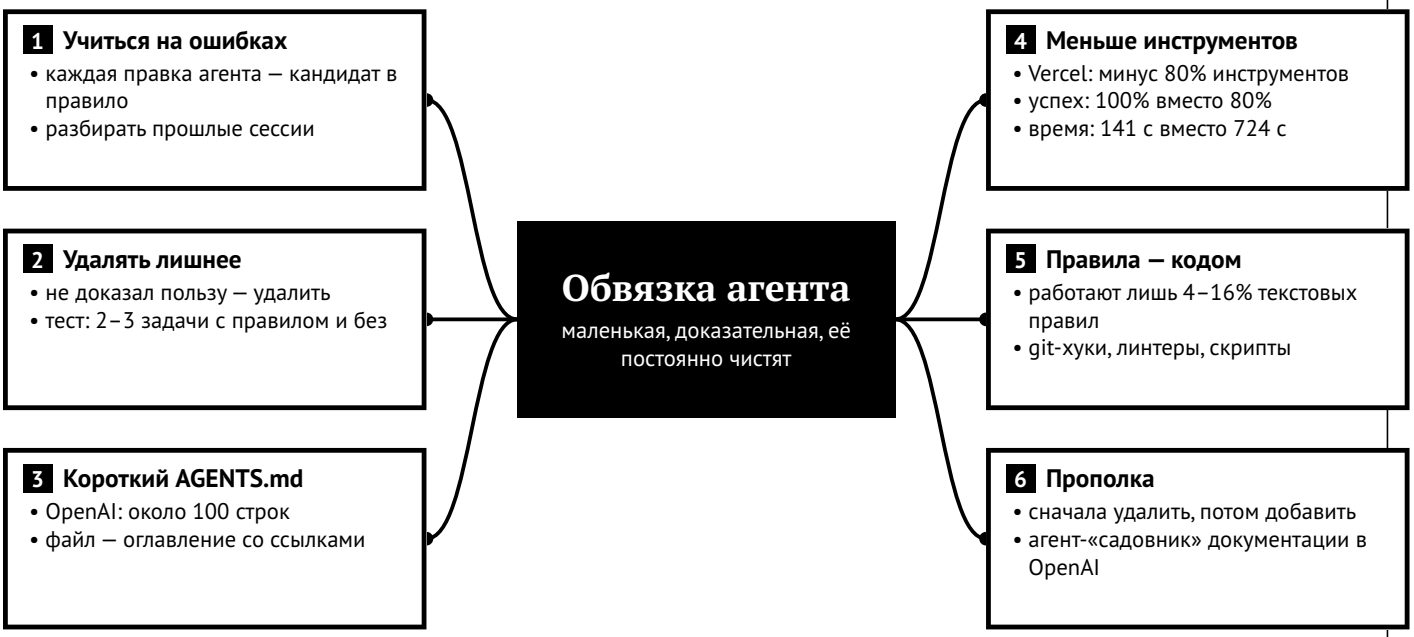
ГЛАВНОЕ

Дело не в модели, а в обвязке: шесть правил из 57 исследований, которые делают ИИ-агента надёжнее

Автор под ником Marmelab прочитал 57 публикаций о том, как устроена «обвязка» (harness) вокруг языковой модели — инструкции, инструменты, правила и проверки, — и свёл их к шести практикам. Итог в одной фразе: лучшие обвязки маленькие, доказательные и их постоянно чистят.

МАЙНДКАРТА: АНАТОМИЯ ХОРОШЕЙ ОБВЯЗКИ

6 практик и цифры из источников



По обзору 57 публикаций на r/AI_Agents. Исходная точка: одна модель даёт разницу в результатах до 6 раз в зависимости от обвязки (Университет Торонто).

ПРОБЛЕМА

Принято думать, что агент станет лучше, если подключить модель покрупнее. Автор обзора на форуме r/AI_Agents приводит исследование Университета Торонто: одна и та же модель показывает разницу в результатах до шести раз в зависимости только от того, как устроена обвязка вокруг неё.

Значит, у команды, которая строит агентов, есть рычаг сильнее смены модели — и он полностью в её руках.

1. УЧИТЬСЯ НА ПРОШЛЫХ СЕССИЯХ

Каждый раз, когда агента приходилось поправлять, ему не хватало подсказки или ограничения. Такие места — главные кандидаты в правила: ошибки и исправления из прошлых сессий превращают в короткие инструкции.

2. УДАЛЯТЬ НЕДОКАЗАННОЕ

Обратная сторона: если агента ни разу не приходилось поправлять в чём-то, правило об этом не нужно. Оно не помогает, а только занимает контекстное окно — память модели на одну задачу.

3. КОРОТКИЙ ФАЙЛ ИНСТРУКЦИЙ

OpenAI пробовала держать один гигантский файл инструкций, и это провалилось: когда важно всё, не важно ничего. Файл сократили примерно до 100 строк. Главный файл (AGENTS.md) должен работать как оглавление и ссылаться на отдельные правила.

4. МЕНЬШЕ ИНСТРУМЕНТОВ, НО ТОЧНЕЕ

Vercel убрала 80% инструментов своего агента. Доля успешно выполненных задач выросла с 80 до 100%, расход токенов упал более чем вдвое, а время выполнения сократилось с 724 до 141 секунды. Microsoft, по словам автора, урезала набор со 100 инструментов до пяти.

5. ПРАВИЛА ЗАКРЕПЛЯТЬ СКРИПТАМИ

Исследователи проверили письменные правила безопасности в файлах CLAUDE.md: реальный эффект давали лишь 4–16% из них. Текстовое правило — это пожелание. Сильные обвязки опираются на исполняемые проверки: git-хуки, линтеры, статический анализ и собственные скрипты, которые не дают агенту сделать запрещённое.

6. РЕГУЛЯРНАЯ ПРОПОЛКА

Правила устаревают, документация расходится с кодом. Совет: прежде чем добавить правило, удалить старое, а после больших изменений перепроверить весь набор. В OpenAI

часть этой работы делает отдельный агент — «садовник документации»: он регулярно ищет устаревшие документы и открывает пул-реквесты с исправлениями.

ЧТО ДОБАВИЛИ ЧИТАТЕЛИ

- Дешёвая проверка правила: взять 2–3 уже решённые задачи, прогнать с правилом и без и сравнить результаты, а не ощущения. Разницы нет — правило удаляется.
- Добавлять по одному правилу за раз: если поменять пять вещей сразу, непонятно, что сработало.
- Делить инструкции на короткий постоянный файл и план под конкретную задачу, который выбрасывается после неё.
- Перед словом «готово» просить агента перепроверить свою работу: модели склонны рано объявлять победу.
- Попросить агента самому предлагать правку инструкций, когда вы говорите ему что-то похожее на постоянное правило.
- Заранее давать контекст (какой сервис за что отвечает): это сокращает число вызовов инструментов без смены модели.

ЧЕМ ЭТО ИНТЕРЕСНО

Это готовый чек-лист для любой команды агентов, в том числе маркетинговой: качество растёт от дисциплины вокруг модели. Первый шаг стоит пять минут — открыть свой файл инструкций и спросить про каждое правило, чем доказана его польза.

МАРКЕТИНГ

Контент-завод на n8n: два работающих конвейера, где ИИ готовит, а публикует только человек

Два автора с форума r/n8n показали самодельные системы публикации в соцсетях. Устроены они по-разному, но правило у обоих одно: всё публичное ждёт кнопки «одобрить» в Telegram.

ИНФОГРАФИКА: КАК УСТРОЕН КОНТЕНТ-КОНВЕЙЕР

сплошная рамка — агент или автоматика, пунктир — человек



ЗАЩИТА ОТ СБОЕВ

Ключи отдельно
в экспорте и логах n8n их нет

Один сигнал
ошибки всех узлов — в одно оповещение

Без дублей
ключ идемпотентности: ролик плюс площадка

Собрано по двум постам на r/n8n (25 процессов на домашнем сервере и приложение Poster).

КОНВЕЙЕР 1: 25 ПРОЦЕССОВ НА ДОМАШНЕМ СЕРВЕРЕ

Автор держит n8n у себя дома — 25 активных процессов для бизнеса. Один из них заменил платный сервис ManyChat: человек пишет ключевое слово в комментарии в Instagram или Facebook и получает публичный ответ и личное сообщение с обещанным материалом. С июля так отправлено около 85 тыс. сообщений, все записываются в таблицы n8n (Data Tables).

КАК ПУБЛИКУЕТСЯ ВИДЕО

Вертикальный ролик загружается через форму n8n. Локальная модель Ornith 1.5 35B на мини-ПК (по словам автора, GMKtec EVO-X2 со 128 ГБ памяти) переписывает подпись под каждую площадку. В Telegram приходит запрос на одобрение; после нажатия n8n сам загружает ролик в YouTube Shorts, LinkedIn, Instagram и Facebook — без публичной ссылки на файл и без туннеля наружу.

Остальные процессы: CI и заметки к релизам на GitHub, напоминания по продажам, оповещения о сбоях сервера.

КОНВЕЙЕР 2: POSTER — ИЗ НОВОСТЕЙ В ПОСТЫ

Открытое приложение Poster построено вокруг процесса n8n и маленькой веб-панели на Node.js. n8n собирает новости, модель пишет текст и картинку, черновик приходит в Telegram с кнопками «да» и «нет», одобрение запускает публикацию в X и Facebook. Работает круглосуточно на домашнем хранилище (NAS) или любом сервере с Docker, ставится одной командой.

Бесплатная версия LITE (лицензия AGPL) — один новостной бот с публикацией в X и Facebook; платная PRO — больше ботов и площадок.

ИНЖЕНЕРНЫЕ ПРИЁМЫ

- Ключи API хранятся в отдельной панели, а не в n8n: в экспорте процесса и в логах их нет.
- Каждый узел публикации продолжает работу при ошибке, в конце собирается один отчёт: сбой одной площадки не блокирует остальные, повтор отправляет только неудачные.
- JSON процесса генерирует скрипт на Python, а CI проверяет его на той версии n8n, что в docker-compose.
- У первого автора ошибки всех узлов сводятся в одно оповещение.

ЧТО СОВЕТУЮТ В КОММЕНТАРИЯХ

Ставить на публикацию ключ идемпотентности (номер ролика плюс площадка), чтобы двойное нажатие «одобрить» не опубликовало ролик дважды. Таймаут считать неизвестным исходом, а не ошибкой: площадка могла принять загрузку. Автор Poster признаёт, что у него таймаут пока считается сбоем.

ЧЕМ ИНТЕРЕСНО

Это почти готовая схема для маркетинговой команды агентов на много каналов: черновую работу делает модель, причём даже локальная, а всё, что уходит наружу, проходит через одну кнопку человека.

БИЗНЕС РОССИЯ

VK описывает новую рекомендательную ленту в конфиге вместо недель разработки — и показывает рост рекламы

Команда AI VK рассказала об устройстве платформы Discovery — общего слоя для рекомендаций, поиска и рекламы во всех продуктах компании.

ПРОБЛЕМА

Каждая новая рекомендательная лента («вертикаль») раньше требовала недель рутины: поднять сервисы, настроить потоки данных, мониторинг и A/B-тесты.

КАК СДЕЛАЛИ

Теперь вертикаль описывают в конфиге: каталог объектов, источники кандидатов, профиль пользователя, цели оптимизации. Платформа сама разворачивает сервисы, потоки данных, мониторинг и контур A/B-тестов. Около 90% задач ранжирования решаются конфигурацией, без нового кода. Так запущено 15 новых вертикалей — вплоть до ранжирования комментариев.

Изменения идут через Discovery Portal: их проверяют до применения, раскатывают поэтапно со сравнением метрик и автоматически откатывают при просадке; откат — в один шаг. Через портал проходят десятки изменений в день. В ранжировании, где счёт идёт на миллисекунды, используют видеокарты: динамический батчинг, контроль «хвостов» задержки и запасной путь на более простую модель.

ЧТО ДАЛЬШЕ

Следующий шаг — полностью перенести на Discovery поиск. Подробности команда выложила на Хабре.

ЧЕМ ИНТЕРЕСНО

Пример того, как машинное обучение превращают в конфигурацию: эксперимент стоит дёшево, поэтому их много, а цифры по рекламе — прямой денежный эффект.

В ЦИФРАХ

+11%	время в ленте ВКонтакте
≈2,7×	добавлений в друзья из рекомендаций
+22%	время просмотра VK Клипов
+20%	главная VK Видео
+46%	конверсия рекламы в сообщения сообществу; цена действия −13%
+56%	целевые действия в рекламе сайтов с оплатой за клик; цена −24%

АГЕНТЫ КИТАЙ

Moonshot AI встроила агентов Kimi в Unity, Unreal и Blender: одна демонстрация шла 20 часов без людей

Китайская Moonshot AI выпустила решение для разработки игр и творческой работы на моделях Kimi K3.1 и агентах Kimi Code и Kimi Work.

КАК РАБОТАЕТ

Модель читает дизайн-документы, скриншоты, код и 3D-ассеты и планирует работу. Kimi Code меняет код, Kimi Work ведёт многошаговые задачи по множеству файлов; через плагины для Unity, Unreal и Blender агенты работают прямо в существующих проектах и возвращают редактируемые файлы и журнал действий.

В игровом движке Kimi сначала строит карту кода, иерархии сцены и зависимостей модулей, вносит изменения через плагины редактора и сверяет результат с шагами воспроизведения ошибки, логами и данными из запущенной игры. Диффы и записи проверок сохраняются для инженеров.

РЕЗУЛЬТАТ

Вместе с агентом Tuanjie Codely от Unity China Kimi понимает замысел, пишет код и разбирает записи экрана и результаты тестов, а Codely исполняет действия в движке. В одной демонстрации связка 20 часов без вмешательства человека пересоздавала существующую игру в виде нативного проекта.

Витрина: в Blender — 17,5-секундный пролёт над долиной тропического леса (1920×1080, 60 кадров в секунду); собор в браузере — около 3,3 млн треугольников и 8368 анимированных деталей; в Unreal Engine 5.8 — небесный дворец с колоннадой из 36 пролётов по радиусу 1750 м и процедурным драконом, 36 млн треугольников. По словам компании, дизайнеры без опыта программирования собирают играбельные прототипы на Godot за 1–2 дня.

НЕ ТОЛЬКО ИГРЫ

Пакетная генерация рекламных креативов, игровые ассистенты через API Kimi, сортировка отзывов сообщества со ссылками на источники, черновики локализации с проверкой обратным переводом и первичный отбор пользовательского контента. Совет компании: начинать с одной чётко ограниченной задачи и оценивать пилот по сэкономленному времени, качеству и тому, сколько ещё приходится править человеку.

ЧЕМ ИНТЕРЕСНО

Тот же принцип, что в главной статье: агент получает инструменты прямо внутри рабочей программы и обязан проверять результат по логам.

НАУКА КИТАЙ

Агент-исследователь Tencent Hyra 2.0 не решает задачи по одной, а строит научные инструменты

Tencent Hunyuan 2 ноября выпустила вторую версию автономного агента для научной работы. Код и результаты открыты на GitHub под лицензией Apache 2.0. Все цифры ниже — данные Tencent, независимо их пока никто не проверял.

КАК УСТРОЕН

Первая версия решала отдельные задачи, вторая строит переиспользуемые инструменты. Агенты работают в три слоя: общий, уровня миссии и исполнители. У них общая исследовательская доска, постоянная очередь задач, отдельный агент-рецензент для проверки результатов, многоуровневая память и база знаний, которая превращает уроки в навыки.

ЧТО ПОСТРОИЛ: БИОЛОГИЯ И ХИМИЯ

- HyInv — обратный фолдинг белка: 1,96 млн параметров, в 57 раз быстрее ProteinMPNN, 128 последовательностей за 50 мс.
- HyPLM-Flash — белковая языковая модель: 650 млн параметров, 726 млрд токенов, самая быстрая из 12 сравнённых и третья в среднем по 35 задачам, около 130 тыс. остатков в секунду на одной H20.
- HyFF — пять силовых полей от 2,3 млн до 1,2 млрд параметров для 83 элементов; смоделирован аквапорин из 60 тыс. атомов на протяжении 100 нс.

МАТЕМАТИКА, ФИЗИКА, ИНЖЕНЕРИЯ

- HyProver закрыл гипотезу Петти о равносторонних множествах в размерности 5 и нашёл контрпример к неограниченной гипотезе Яу о действии окружности (односвязный случай открыт).
- HyQEC нашёл квантовый код [[488,123,22]] — примерно на 48% лучше прежнего лидера.
- Механизм роботизированной руки стал быстрее на 28,6%; процессор XiangShan получил +5,79% SPECint2006/ГГц при +0,3% площади, найдено 10 ошибок.

ЧЕМ ИНТЕРЕСНО

Hyra уже помогает обучать следующие модели Hy. А её схема — общая доска, очередь задач, отдельный проверяющий и база навыков — переносится на любую команду агентов, не только научную.

ЛЮДИ США

Рандомизированное исследование в Google Health: агент со сценарием готовит пациента к визиту лучше обычного чат-бота

802 взрослых американца провели 866 разговоров в приложении Google Health — со специальным агентом подготовки к приёму или с обычным чат-ботом на той же модели (статья на arXiv, 36 авторов).

ПРОБЛЕМА

На короткий приём пациенты часто приходят, не понимая, чего ждать и что спросить. Миллионы людей уже готовятся к визиту с универсальными чат-ботами, но насколько это помогает, почти не изучали.

КАК СДЕЛАЛИ

Агент работает по сценарию: собирает клинический контекст, объясняет, чего ждать от визита, даёт уместную медицинскую информацию, составляет приоритетный список вопросов к врачу и итоговую сводку. Сравнивали две версии агента — стандартную и краткую — с минимально настроенным чат-ботом на той же языковой модели.

РЕЗУЛЬТАТ

Врачи, не знавшие, какая версия перед ними (n = 272), чаще оценивали подготовку как полную: 67,2% против 50,6%; соответствие медицинскому контексту — 98,9% против 90,4%; безопасность высокая во всех группах. Пациенты (n = 460): готовность к визиту 73,8% против 57,0%, уверенность в том, что спросить, — 81,4% против 67,1%, удовлетворённость — 74,0% против 62,0%. Оценки врачей и пациентов между собой не коррелировали.

ЧЕМ ИНТЕРЕСНО

Ещё одно подтверждение главной темы: та же модель, но со структурированным сценарием даёт заметно лучший результат. И урок для оценки агентов: эксперт и пользователь меряют разное — нужны обе метрики.

В ЦИФРАХ

67,2%	полная подготовка по оценке врачей (у чат-бота 50,6%)
73,8%	пациентов чувствуют себя готовыми (57,0%)

СПОРТ КИТАЙ

Роботы штурмовали гору Тайшань: 3100 ступеней показали, чего не хватает «воплощённому» ИИ

В городе Тайань (провинция Шаньдун) прошли первые соревнования роботов по восхождению на Тайшань, организованные China Media Group. Участвовали более 50 компаний, университетов и команд.

ТРАССА

3,9 км от ворот Хунмэнь до ворот Чжунтяньмэнь: более 3100 каменных ступеней разной высоты и почти 600 м набора высоты. Это первые профессиональные соревнования роботов в горах на объекте ЮНЕСКО. Четыре дисциплины — гонка гуманоидов, гонка шагающих роботов, переноска груза шагающими роботами и экзоскелеты, — плюс отдельное испытание на поиск пути по пересечённой местности.

РЕЗУЛЬТАТЫ

- Гонка шагающих роботов: Beijing Lingyi Technology (дочерняя компания Unitree) — 43 мин 41 с; второе место у Shanghai Gaoyi Technology (тоже на роботах Unitree), третье — у команды Lingju компании Linxai.
- Переноска груза: команда Lingshi компании Linxai — 1 ч 51 мин 29 с.
- Гуманоиды бежали укороченный участок до дворца Доуму; победила Zhishen Lingyue — 27 мин 26 с.

ЧТО ЛОМАЛОСЬ

Перегрев моторов, автоматическое снижение мощности, увод в сторону на ступенях; у грузовых роботов походка становилась неустойчивой после смены груза на пунктах снабжения.

ЧЕМ ИНТЕРЕСНО

Ходить по ровному полу лаборатории и быть готовым к работе на улице — разные вещи. Узкие места названы прямо: устойчивость, баланс груза и батареи, умение самому подстраиваться под рельеф.

ПОДБОРКА ИНСТРУМЕНТОВ НЕДЕЛИ

claude-mem

ЧТО ДЕЛАЕТ

Записывает, что агент делал в сессиях, сжимает это ИИ и подставляет нужный контекст в новые сессии.

ДЛЯ ЧЕГО

Чтобы Claude Code и другие агенты для кода не начинали каждую сессию с нуля.

ЧЕМ ИНТЕРЕСЕН

Ставится одной командой npx claude-mem install; был в GitHub Trending.

ЦЕНА

30 дней бесплатно, затем память работает за счёт вашего тарифа Anthropic, если не оформить подписку.

Alibaba Cloud Memory

КИТАЙ

ЧТО ДЕЛАЕТ

Долговременная память и структурированный профиль пользователя для персональных агентов: предпочтения, планы, обещания.

ДЛЯ ЧЕГО

Например, агент-консультант по инвестициям помнит риск-профиль и цели клиента.

ЧЕМ ИНТЕРЕСЕН

Свои правила памяти и поля профиля, отладка с историей изменений; подключается через API, MCP и CLI к Codex, Claude Code, OpenClaw. На тесте HaluMem — 85,15% против 81,71% у Zep (данные Alibaba).

ЦЕНА

в источнике не указана

Soniox STT v6

ЧТО ДЕЛАЕТ

Распознавание речи, которое «видит» текст, произносимый самим голосовым агентом.

ДЛЯ ЧЕГО

Голосовые и телефонные боты: лучше понимает короткие ответы и вырезает эхо голоса агента.

ЧЕМ ИНТЕРЕСЕН

Включается одним флагом в запросе, 60+ языков. Независимых замеров v6 пока нет.

ЦЕНА

$0,12 за час потокового распознавания, $0,10 за час файлов.

NVIDIA Auto Ontology США

ЧТО ДЕЛАЕТ
Смысловой слой между агентом и хранилищем данных: где лежат данные, что значат поля и как их использовать.

ДЛЯ ЧЕГО
Чтобы агент писал SQL по утверждённым бизнес-определениям и не ошибался, например, в выручке.

ЧЕМ ИНТЕРЕСЕН
MCP-сервер с 9 инструментами только на чтение; Databricks, Postgres, Snowflake, DuckDB. NVIDIA сообщает о сокращении времени на создание онтологии на 90% (собственный тест).

ЦЕНА
Открытый код, лицензия Apache 2.0.

Letterpress-1

ЧТО ДЕЛАЕТ
Агент для «growth-инженерии» сайта: читает сайт и рынок, предлагает, что сделать, собирает в песочнице и открывает пул-реквест с превью.

ДЛЯ ЧЕГО
Маркетинговые сайты в git: лендинги, блог, A/B-тесты.

ЧЕМ ИНТЕРЕСЕН
Режимы «собрать», «спланировать» и «исправить»; без вашего выбора ничего не попадает на сайт; 760 инструментов через MCP, REST и CLI. Публичных тестов качества нет.

ЦЕНА
От $100 в месяц (10 000 токенов) до $2500.

hallmark

ЧТО ДЕЛАЕТ
Навык для агентов, которые верстают интерфейсы: уводит от шаблонного «ИИ-дизайна».

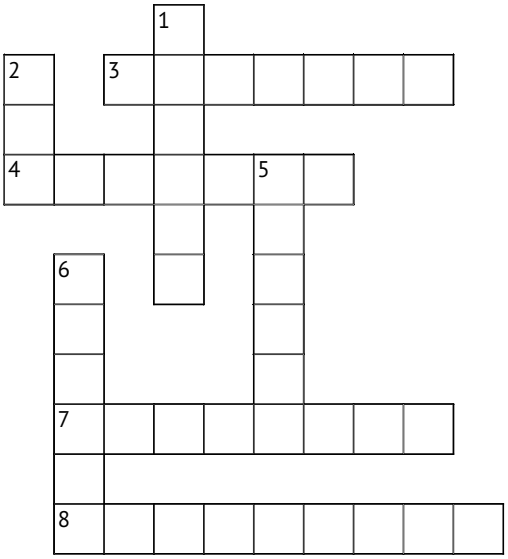
ДЛЯ ЧЕГО
Чтобы страницы от Claude Code, Cursor или Codex не выглядели одинаково.

ЧЕМ ИНТЕРЕСЕН
Был в GitHub Trending рядом с похожим набором impeccable.

ЦЕНА
Открытый репозиторий на GitHub; условия в источнике не указаны.

ПРОВЕРЬ СЕБЯ

по этому выпуску и базе знаний; ответы внизу вверх ногами



ПО ГОРИЗОНТАЛИ

3. Гора, на которую поднимались роботы (7)

4. Всё, что окружает модель: инструкции, инструменты, правила, проверки (7)

7. Где человек нажимает «одобрить» в обоих конвейерах n8n (8)

8. Отдельный агент в Hyra 2.0, который проверяет результаты (9)

ПО ВЕРТИКАЛИ

1. Что дают агенту claude-mem и сервис Alibaba Cloud (6)

2. Что Soniox STT v6 вырезает из голоса самого агента (3)

5. Чем VK заменил новый код в 90% задач ранжирования (6)

6. Исполняемая проверка, которая надёжнее текстового правила (6)

ЗАДАЧКИ

1. Во сколько раз, по исследованию Университета Торонто, может отличаться результат одной и той же модели из-за обвязки?
   а) в 1,5 раза б) до 6 раз в) в 43 раза
2. Какая доля письменных правил безопасности в CLAUDE.md реально давала эффект?
   а) 4–16% б) около 50% в) 80–100%
3. Что Vercel сделала с инструментами агента, когда успех вырос с 80 до 100%?
   а) удвоила б) убрала 80% в) заменила модель
4. Почему в исследовании Google Health важно, что оценки врачей и пациентов не коррелировали?

ПОДУМАТЬ

- Какое правило в инструкциях ваших агентов вы удалили бы первым — и как проверите на 2–3 старых задачах, что ничего не сломалось?
- Какие публикации в вашем бизнесе можно отдать конвейеру ИИ, а какие обязательно должны проходить через кнопку человека?
- Где в вашей работе нужен независимый проверяющий, как агент-рецензент в Hyra 2.0?

Ответы. Кроссворд: 1в — память; 2в — эхо; 3г — тайшань; 4г — обвязка; 5в — конфиг; 6в — линтер; 7г — телеграм; 8г — рецензент.
Задачки: 1) б — до 6 раз. 2) а — 4–16%. 3) б — убрала 80%. 4) они мерялют разное, агенту нужны обе метрики.

ИСТОЧНИКИ (ФОРМАТ) **Главное** — форум (r/AI_Agents) · **Контент-завод на n8n** — форум (r/n8n) · **VK Discovery** — Telegram-канал (t.me/data_secrets) · **Kimi в игровых движках** — журнал/блог (Pandaily) · **Tencent Hyra 2.0** — журнал/блог (Pandaily) · **Google Health: подготовка к врачу** — научная статья (arXiv) · **Роботы на Тайшане** — журнал/блог (TechNode) · **Инструменты** — claude-mem — репозиторий GitHub (GitHub Trending); Alibaba Cloud Memory — журнал/блог (Pandaily); Soniox STT v6 — журнал/блог (MarkTechPost); NVIDIA Auto Ontology — журнал/блог (MarkTechPost); Letterpress-1 — журнал/блог (MarkTechPost); hallmark — репозиторий GitHub (GitHub Trending)

Обзор за неделю 28.10–04.11.2026 · собрала Ника · стиль v1.0, многополосный формат v2 · все цифры — из первоисточников, оценки компаний помечены