ИИ-газета ДЛЯ АЛЕКСЕЯ

№ 3 ЧЕТВЕРГ, 5 НОЯБРЯ 2026 ЛИСТ С ДВУХ СТОРОН

Мем дня с форума: Сэм Альтман так увлёкся оперативной памятью, что сменил имя на RAM Altman.

АГЕНТЫ РОССИЯ

Оркестратор, который не верит агентам на слово

Российская студия ITSalt выложила плагин pepper-orchestrator для Claude Code и Codex. Он режет большую цель на пакеты работ и раздаёт их отдельным агентам: каждый работает параллельно в своей копии репозитория, а общая картина хранится в файлах, а не в чате. Главное правило — «готово» от агента считается заявлением, а не фактом: сдачу проверяет рецензент с чистым контекстом, а потом три «клиента» — лояльный, «дурак» и конкурент. В тестовом кейсе все тесты агента были зелёными, но рецензент и «конкурент» нашли места, где программа уверенно путала аванс с суммой договора.

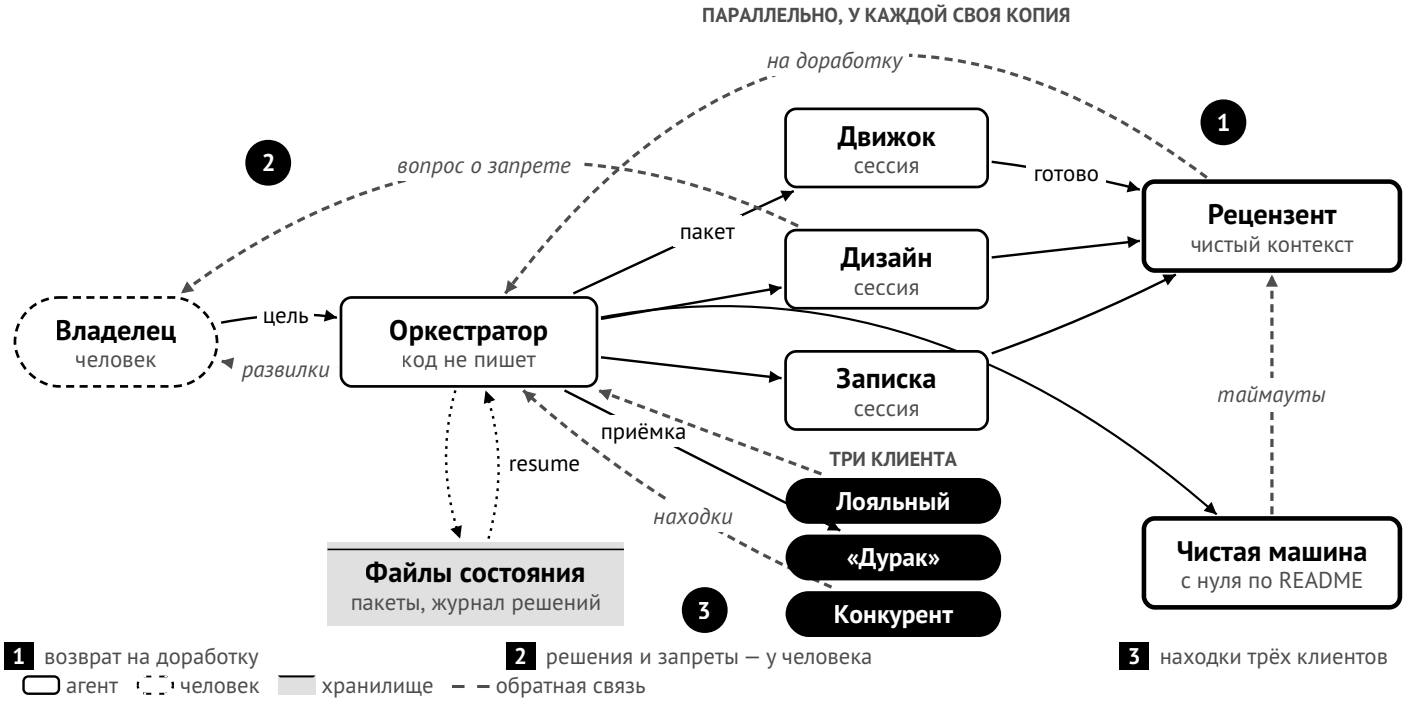
БЕЗОПАСНОСТЬ США

OpenAI натравила агентов на ядро Linux

OpenAI собрала автономный конвейер, который сам ищет уязвимости в ядре Linux, проверяет их и пишет заплатки. Так нашли почти тысячу дыр, через которые программа на машине может получить права администратора. На форуме спорят о реальной опасности: для большинства нужен уже имеющийся доступ к компьютеру. Но вывод общий: искать ошибки в коде стало намного дешевле — и защитникам, и нападающим.

СХЕМА СЕТЕВОЙ ГРАФ

Как pepper-orchestrator гоняет работу между агентами и человеком



БИЗНЕС США

Агенты пока не вызывают такси

Uber впервые раскрыла, сколько заказов сделали сторонние ИИ-агенты. За квартал — всего четыре поездки и около пятисот доставок еды за неделю Хэллоуина, на фоне сотен миллионов заказов. Страх, что агенты отрежут сервисы от клиентов, пока не сбывается: Uber готова с ними работать, но верит, что люди останутся в приложении.

ПОЛИТИКА США

Выборы в США ударили по стройке дата-центров

На промежуточных выборах победили демократы, и многие из них обещали запретить или заморозить новые дата-центры. Под ударом Техас — самый быстрорастущий рынок, где строят гигантские площадки в том числе для Anthropic. Новый губернатор штата выступала за мораторий, а банки уже осторожнее дают деньги на такие стройки.

ЛЮДИ США

OpenAI зарабатывает на компаниях и забирает людей у Anthropic

Президент OpenAI Грег Брокман сказал, что большая часть выручки уже приходит от бизнеса, а к 2027 году компания будет готова к бирже. OpenAI борется с Anthropic за корпоративных клиентов и продвигает агента Codex. В тот же день в OpenAI перешла Кэт Ву, которая два года делала Claude Code в Anthropic.

БИЗНЕС КИТАЙ

ByteDance закрыла чужим ИИ доступ к своему коду

ByteDance запретила сотрудникам работать с внутренним кодом и документами через Claude и Codex, даже если подписку они оплачивают сами. Вице-президент по технологиям объяснил это защитой данных и отрицает, что так продвигают свой ассистент Trae. Это второй виток: в 2025 году компания уже убирала Cursor и Windsurf.

НАУКА США

Модель иногда замечает подсаженную «мысль»

Исследователь Anthropic Джек Линдси искусственно усиливал в Claude активность, связанную с конкретным понятием, и спрашивал, замечает ли модель вмешательство. Новейшая версия иногда верно называла подсаженное понятие, а ложных тревог не было ни разу. Можно ли считать это самонаблюдением, учёные спорят до сих пор.

АГЕНТЫ КИТАЙ

Агенты Huawei переезжают в рабочие чаты

Сообщество openJiuwen, которое поддерживает Huawei, выпустило Hyper-Agent — рой агентов, живущих прямо в рабочих групповых чатах. Агенты не исчезают после разговора, а работают как постоянные цифровые сотрудники; код открыт. Там же вышла память, которая запоминает длинное видео по событиям, а не по минутам.

Мем с форума: «Я, когда пытаюсь одновременно управлять Claude, Grok и Codex».

МОДЕЛИ КИТАЙ

DeepSeek заняла половину OpenRouter

За один день больше половины всех токенов на площадке OpenRouter пришлось на китайскую DeepSeek. Модели с открытыми весами вместе заняли около трёх четвертей трафика, хотя ещё весной лидерами были OpenAI и Anthropic. На форуме объясняют это ценой и удобством: один ключ — и можно переключаться между моделями.

ЛИЧНАЯ ЖИЗНЬ КИТАЙ

Помощник Ant Group ведёт дела миллионов людей

ИИ-ассистент Lingguang от Ant Group вышел в версии 2.0 и за первую неделю взял на себя больше миллиона личных дел пользователей. Он не только отвечает, но и сам выполняет задачи. Аудитория помощника перевалила за 20 миллионов человек.

ПРОДАЖИ КИТАЙ

Рабочий телефон, на котором продаёт агент

DeepBlue Technology и Huawei показали рабочий смартфон Wowoker, где продажами занимаются агенты. Агент по отрасли и ключевым словам ищет клиентов в китайских соцсетях и сам начинает разговор. На презентации агент позвонил родителю, уточнил возраст ребёнка, предложил пробное занятие и договорился продолжить в WeChat.

ИНСТРУМЕНТЫ

agent-desktop США от Vercel Labs: агент управляет обычными программами на macOS, Windows и Linux без скриншотов — читает дерево элементов интерфейса и нажимает по ссылке на элемент. Модели внутри нет.

Worktrunk делает git worktree простыми, как ветки: у каждого из параллельных агентов своя рабочая папка, и они не затирают правки друг друга.

gstack США набор главы Y Combinator Гарри Тана для Claude Code: роли «CEO», дизайнер, ревьюер, QA с браузером, безопасность и релиз — командами в Markdown, бесплатно.

NVIDIA Auto Ontology США открытая библиотека строит для агентов слой смыслов поверх данных компании: где лежит цифра, что значит метрика, откуда она взялась. Подключается через MCP.

ПОИГРАЕМ СОЕДИНИТЕ ТЕРМИН С ОПРЕДЕЛЕНИЕМ; НЕЗНАКОМОЕ — ПОИЩИТЕ

ТЕРМИН	ОПРЕДЕЛЕНИЕ
1. Git worktree	**а.** Нарочно портят код и смотрят, заметят ли это тесты
2. Дерево доступности	**б.** Маленькая модель угадывает несколько слов вперёд, большая проверяет их разом
3. Онтология	**в.** Отдельная рабочая папка того же репозитория на своей ветке
4. Открытые веса	**г.** Файлы модели можно скачать и запустить у себя
5. Мутационное тестирование	**д.** Описание понятий предметной области и связей между ними
6. Спекулятивное декодирование	**е.** Список кнопок и полей программы, который система отдаёт экранным чтецам

ЗАДАЧКИ

1. Эффект Lost in the Middle: какую часть длинного контекста модели замечают хуже всего?
2. Чем атака «повышение привилегий» отличается от удалённого взлома?
3. Ваш агент собрал пустой отчёт. Какую одну строку в отчёт добавить, чтобы его не спутали с поломкой?

НА ПОДУМАТЬ

- *Если за людей начнут заказывать агенты, зачем сервисам красивые приложения и реклама внутри них?*
- *Модель, которая замечает вмешательство в свои «мысли», — это защита или новый риск?*

Ответы: 1в, 2е, 3д, 4г, 5а, 6б · 1) середину · 2) доступ к мышке уже есть, цель — права администратора · 3) статус сборки: сколько источников ответило

Источники. журнал/блог — оркестратор, OpenAI и Linux, Uber, дата-центры, OpenAI и люди, ByteDance, «мысль» Claude, рой Huawei, Ant Group, Wowoker, Схема; **форум** — OpenAI и Linux, дата-центры, DeepSeek; **репозиторий GitHub** — Инструменты. Ника · формат v3 · 05.11.2026