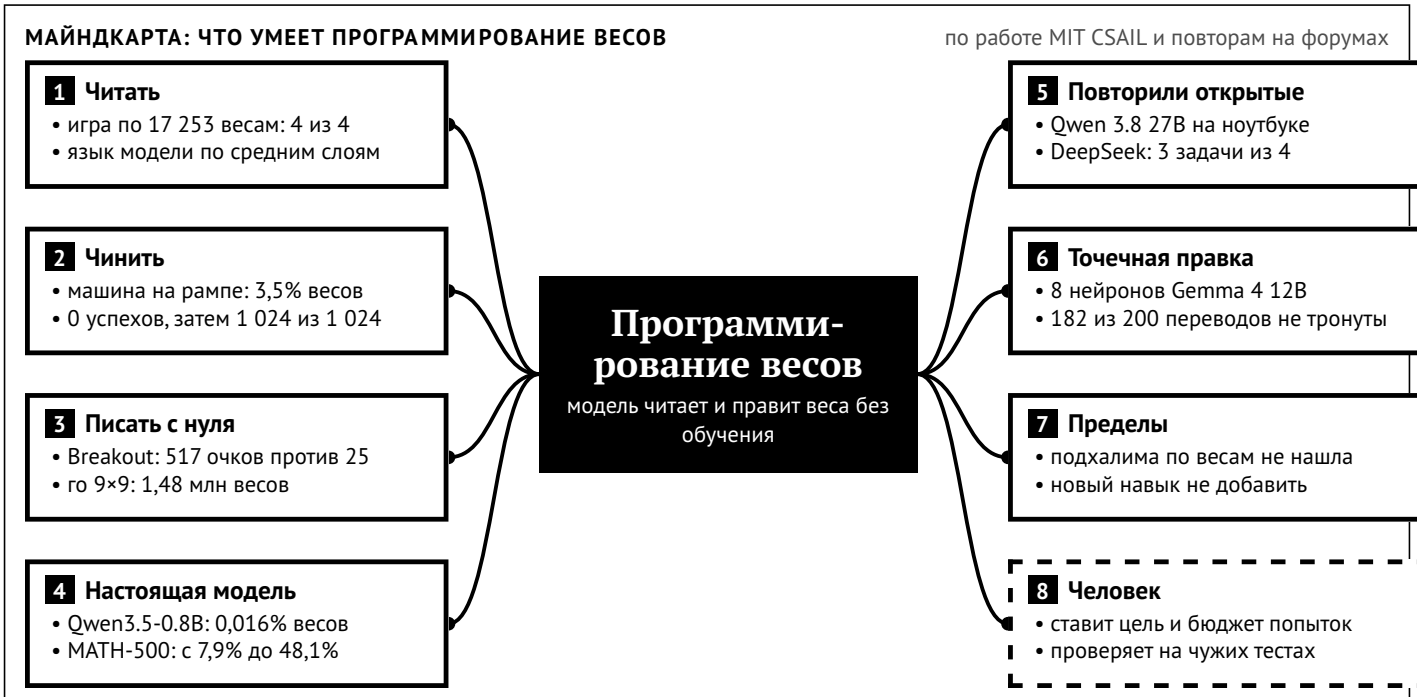
ГЛАВНОЕ США

Модель правит «мозг» другой модели: GPT-6 Astra исправила Qwen, изменив 0,016% весов, — и за сутки это повторили открытые модели

Исследователи MIT CSAIL показали, что большая языковая модель умеет читать и переписывать веса другой нейросети без обучения и градиентов. Американская закрытая модель починила китайскую открытую, а через день энтузиасты повторили часть опытов на Qwen, DeepSeek и Claude — в том числе на ноутбуке.



Сплошная рамка — что делает модель, пунктир — роль человека. Источник: блог MIT CSAIL от 2 ноября и обсуждения на r/LocalLLaMA и в блоге Саймона Уиллисона.

ПРОБЛЕМА

Веса нейросети — миллионы и миллиарды чисел, и до сих пор их «пишет» только обучение: данные, градиентный спуск, часы на видеокартах. Человек не может открыть файл весов и понять, что модель делает, или поправить ошибку руками. Поэтому вредную привычку убирают дообучением, рискуя сломать остальное, а скрытую проблему находят, только если под неё есть тест.

КАК СДЕЛАЛИ

Исследователи MIT CSAIL 2 ноября опубликовали работу «Sparks of Weight-Space Programming». Модели GPT-6 Astra давали веса как текст или через Python-среду и запрещали писать циклы обучения. Задачи шли от простого к сложному:

- Читать: четыре одинаковые по устройству сети (17 253 веса) учили разным играм. Не зная вариантов, Astra по весам угадала игру во всех четырёх случаях.
- Чинить: сеть, которая водила машину и застревала на рампе (0 успехов), Astra за один ответ поправила, изменив 3,5% весов, — успех в 1 024 попытках из 1 024.
- Писать с нуля: без данных и тренировки Astra выписала веса сети для игры Breakout — 517,4 очка против 25,3 у сети, обученной 500 млн шагов.
- Улучшить настоящую модель: Qwen3.5-0.8B «зацикливалась» и в 93% ответов на MATH-500 упиралась в лимит токенов. Astra нашла одно направление в слоях 15 и 19, вычла его из 131 072 весов и проверяла правки 50 прогонами на тренировочных задачах.

РЕЗУЛЬТАТ

После правки Qwen3.5-0.8B на MATH-500 выросла с 7,9% до 48,1%, на GSM8K — с 13,8% до 67,1%, а на тесте следования инструкциям IFEval — с 21% до 62,3%, хотя про инструкции её не просили. Доля ответов, упёршихся в лимит, упала с 93% до 33%. При бюджете в 10 или 25 прогонов эффект был слабым (12,3% и 14,1%) — модели нужны были попытки и проверка. Все логи опубликованы, исправленная Qwen выложена на Hugging Face.

ГДЕ НЕ СРАБОТАЛО

По одним весам Astra не смогла определить, какая из двух копий Gemma 3 270M обучена поддакивать пользователю (0 из 4 попыток), — получилось, только когда ей показали,

куда смотрит внимание моделей. А у Gemma 3 270M, которая не зацикливается, а просто ошибается в арифметике, правки ничего не дали. Вывод авторов: правка может убрать препятствие, которое прятало способность, но не добавить способность, которой нет.

ПОВТОРИЛИ ЗА СУТКИ: ОТКРЫТЫЕ МОДЕЛИ

Американский блогер Саймон Уиллисон повторил чтение весов на Qwen 3.8 27B на ноутбуке. На r/LocalLLaMA китайская DeepSeek V4.1 Flash по весам четырёх сетей верно определила три задачи из четырёх (автор: один прогон, не доказательство).

Самый практичный опыт: Claude Opus 5.5 по той же схеме поменял у Gemma 4 12B выбор времени глагола в переводе на итальянский, переписав 8 нейронов. Нежелательная форма осталась в 5% и 16% случаев против 26% и 32%, а 182 из 200 посторонних переводов не изменились. Инструкция в промпте сработала сильнее, но сломала 49 контрольных предложений; дообучение LoRA изменило 80.

ЧЕМ ИНТЕРЕСНО

Авторы видят направления: искать закладки в модели без тестов, «разучивать» навык без данных, пересаживать блоки между моделями. Для компании из агентов аналогия простая: ошибку чинили точечной правкой с проверкой и бюджетом попыток, а не переобучением «всего сотрудника». И у правки был предел: убрать мешающую привычку можно, научить тому, чего нет, — нельзя.

В ЦИФРАХ

0,016%	весов Qwen3.5-0.8B изменила Astra
7,9% и 48,1%	MATH-500 до и после правки
21% и 62,3%	IFEval: следование инструкциям

РАЗРАБОТКА ВЕЛИКОБРИТАНИЯ, КИТАЙ, США

Два опыта экономии на ИИ-программистах: Revolut перевёл 2 000 инженеров на китайскую GLM у себя, а «дешёвый помощник» для Opus сделал хуже и дороже

Один вопрос — дорогая американская модель или дешёвая открытая — и два разных ответа. Ключ, похоже, не в цене токена, а в том, как модель встроена в работу.

ПРОБЛЕМА

Подписки и API передовых моделей для большой команды разработчиков обходятся дорого, поэтому ИИ берегут для сложных задач. Открытые модели дешевле, но слабее. Как совместить?

КАК СДЕЛАЛ REVOLUT

По словам Дэвида Абитбола, который руководит ИИ-операциями в британском финтехе Revolut (пост в X), около 2 000 инженеров перевели на открытую модель GLM китайской Z.ai, развёрнутую на собственных серверах: 56 видеокарт Nvidia B200. Смешанная цена — около 30 центов за миллион токенов, примерно в десять раз ниже, чем компания платила Anthropic и OpenAI. Claude Code и Codex перенесли в «список исключений»: передовые модели доступны для трудных задач, а по умолчанию работает своя.

РЕЗУЛЬТАТ И ЦИФРЫ

Компания сообщает, что срок выполнения задач (lead time) сократился на 70% — это данные компании. Автор Telegram-канала, разбиравший кейс, считает, что выиграла не сама GLM: токены перестали экономить и стали гонять ИИ на всех задачах. Там же оговорка: маленькой команде без своих видеокарт API всё ещё проще и дешевле.

КОНТРОЛЬНЫЙ ОПЫТ: КОМАНДА ИЗ ДВУХ МОДЕЛЕЙ

На r/ClaudeAI пользователь дал Claude Opus 5.5 дешёвого помощника для кода. На 12 трудных задачах (скрытые тесты, по 3 прогона) Opus один решил 81% по $1,20 за задачу, а с помощником DeepSeek V4.1 Flash — 67% по $1,29. Одиночные GPT-6.1 Terra и Sol дали 78% и 75% по $0,17.

ЧЕМ ИНТЕРЕСНО

Экономит не «дешёвый агент в цепочке», а понятное правило: одна модель по умолчанию, дорогая — по исключению.

В ЦИФРАХ

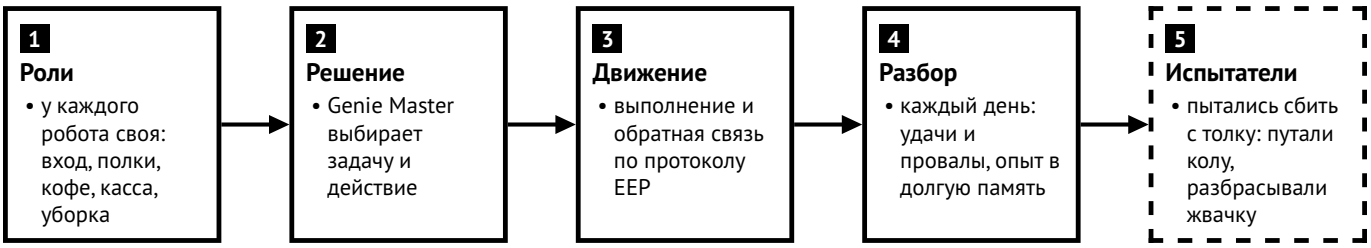
~2 000	инженеров Revolut на своей GLM
$0,30	за миллион токенов (данные компании)
81% и 67%	Opus один и Opus с помощником

РОБОТЫ КИТАЙ

Шесть гуманоидов сутки вели магазин в Шанхае, робот Xiaomi отработал смену на автозаводе, а открытая Qwen управляет роботом в 20 раз дешевле GPT-6 Astra

Китай показывает роботов не на выставке, а в обычной работе — и публикует подробности, по которым видно, как устроена их «обвязка».

ИНФОГРАФИКА: КАК ШЕСТЬ РОБОТОВ AGIBOT РАБОТАЛИ СУТКИ сплошная рамка — агент или автоматика, пунктир — человек



КАК СПРАВЛЯЛИСЬ С ПОДВОХАМИ

Голос и зрение
заказ «колы без сахара» сверяли с картинкой

Счёт по виду
полку с жвачкой восстанавливали подсчётом, а не по ценникам

Объезд
при препятствии робот останавливался и менял маршрут

По данным AgiBot в изложении Pandaily. Сплошная рамка — автоматика, пунктир — люди.

ПРОБЛЕМА

Гуманоиды хорошо выглядят на роликах, но для бизнеса важно, выдержат ли они долгую смену и неожиданные ситуации реального места работы.

КАК СДЕЛАЛИ: МАГАЗИН AGIBOT

Шесть роботов AgiBot G2 с 10:00 2 ноября до 10:00 3 ноября работали в действующем магазине у дома в Шанхае. Роли закреплены: вход, сборка и полки (двое), кофе, касса и горячая еда, уборка столов. Координировал агентный фреймворк Genie Master: каждый день разбирает удачные и неудачные попытки, опыт кладёт в долгосрочную память. Людей в контуре, по словам компании, не было.

КАК СДЕЛАЛИ: ЗАВОД XIAOMI

Робот Xiaomi впервые отработал восьмичасовое задание на автозаводе и закрыл дневную потребность участка сортировки крышек, сам исправляя сбои. Стажировка шла с марта: тогда на установке гаек он показывал 90,2% успеха, к июлю — 98%; по трём участкам компания называет 94–98%.

РЕЗУЛЬТАТ И ЦИФРЫ

AgiBot насчитала 1 730 задач за сутки, в том числе 683 встречи покупателей и 147 кофе. Четыре робота из шести прошли «испытательный срок». Но опубликовано только число задач — доли успешных и частоты вмешательств людей нет. Хуже всего давалась тонкая работа: повесить пакет конфет на крючок, взять мягкий торт.

КИТАЙ И США: ЦЕНА «МОЗГА» РОБОТА

Разработчики Qwen, Университета Цинхуа и Беркли выложили открытую обвязку uni-robot-agent: обычная языковая модель управляет роботом через вызовы инструментов, без дообучения. По их данным, Qwen3.8-Flash-Next обходилась в $1,09 за задачу против $21,29 у GPT-6 Astra, а на 19 реальных задачах набрала чуть больше дообученной модели π0.5.

В ЦИФРАХ

1 730	задач за 24 часа у шести роботов
4 из 6	прошли испытательный срок
94–98%	успех робота Xiaomi на трёх участках

ДЕНЬГИ США И КИТАЙ

Валовая маржа 85% против 26% и 18%: проспект Anthropic показал разрыв с китайскими Z.ai и MiniMax, а Китай продаёт подписки на ИИ через звёзд-стримеров

Американская компания показала цифры к IPO, китайские конкуренты — свои отчёты. Технологически они близко, финансово — очень далеко. И подход к массовому покупателю тоже разный.

США: ЧТО РАСКРЫЛ ПРОСПЕКТ

По исправленной форме S-1 (пересказ TechSpot), с июля по сентябрь 2026 года Anthropic потратила на персонал более $1,7 млрд при 5 300 сотрудниках — около $107 300 на человека в месяц, из них $1,1 млрд — оплата акциями. В 2023 году расходы на людей были вдвое больше выручки, а за девять месяцев 2026-го — $4,8 млрд, около 15% выручки в $33,2 млрд.

КИТАЙ: ТЕХНОЛОГИИ ЕСТЬ, МАРЖИ НЕТ

South China Morning Post сравнила отчётность. Валовая маржа Anthropic в первом полугодии 2026 года — 85,4%, у Z.ai (Zhipu AI) — 26,4%, у MiniMax — 17,9%, обе торгуются в Гонконге. По словам аналитиков, несмотря на технический прогресс китайских разработчиков, превращать его в устойчивый бизнес им заметно сложнее.

КАК ПРОДАЮТ МАССАМ

В Китае ИИ впервые стал главным товаром распродаж «День холостяка»: стример Ли Цзяци, «король помады», с топ-менеджером Alibaba за 30 минут эфира продал почти 100 тыс. месячных подписок на ИИ (данные Alibaba). В США ставка на масштаб: по словам руководителя розничных продуктов ChatGPT Адама Фрая, у сервиса 1,2 млрд активных пользователей в неделю.

ЧЕМ ИНТЕРЕСНО

Рядом два способа зарабатывать на ИИ: дорогой продукт с высокой маржой для бизнеса и дешёвая подписка, которую продают как помаду. А зарплаты людей у Anthropic — всего 15% выручки при самых высоких зарплатах в отрасли.

В ЦИФРАХ

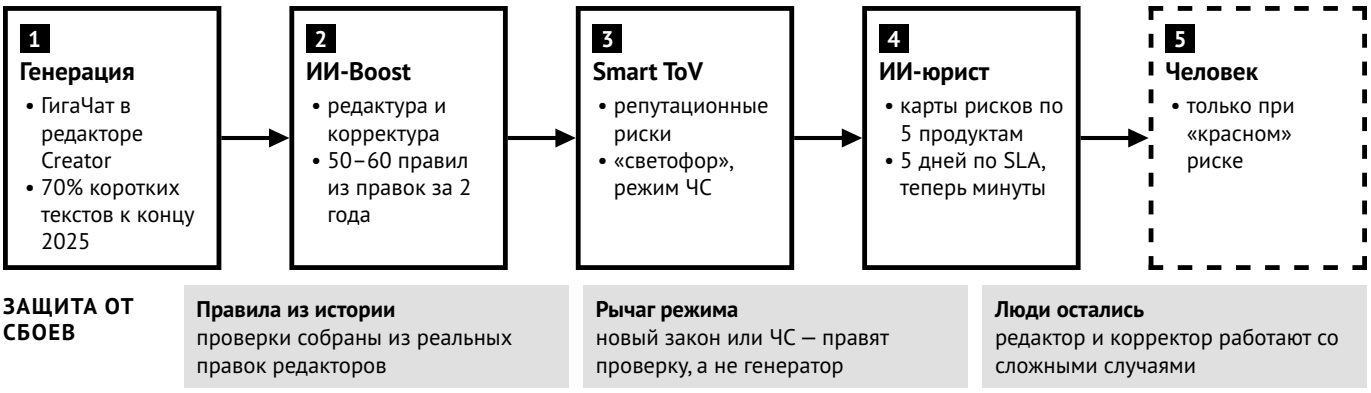
85,4%	валовая маржа Anthropic, 1-е полугодие
26,4% и 17,9%	Z.ai и MiniMax
$33,2 млрд	выручка Anthropic за 9 месяцев

МАРКЕТИНГ РОССИЯ

Сбер сократил путь пуша от идеи до отправки с 7 дней до 1 часа: три ИИ-проверки заменили цепочку согласований, а агент активации за 3,8 ₽ ведёт клиента к покупке

Два российских кейса с Хабра о том, как ИИ ускоряет маркетинг: крупный банк убрал узкое место в согласованиях, а одиночный разработчик посчитал экономику агента-продавца.

ИНФОГРАФИКА: КОНВЕЙЕР ИИ-СОГЛАСОВАНИЯ В СБЕРЕ



По статье Михаила Мирошникова (Сбер) на Хабре. Сплошная рамка — ИИ, пунктир — человек.

ПРОБЛЕМА

Михаил Мирошников из департамента данных B2C Сбера вспоминает: обычный пуш когда-то готовили около двух недель — восемь-девять согласований до юристов и корректора. К концу 2023 года цикл сократили до 5–6 рабочих дней, но информационный повод живёт полдня. Узким местом стала не генерация, а согласование.

КАК СДЕЛАЛИ

В конце 2025 года 70% коротких текстов генерировали ГигаЧатом, затем все этапы согласования свели в три ИИ-проверки. Правила взяли из истории: выгрузили согласованный контент за пару лет и сгруппировали комментарии редакторов. Если вышел новый закон или в регионе ЧС, правят проверку, а не генератор.

РЕЗУЛЬТАТ И ЦИФРЫ

В 2023 году команда делала около 11 тыс. единиц контента, в прошлом году — больше 22 тыс. В команде автора стала вдвое продуктивнее при пяти контент-менеджерах вместо десяти: часть людей сменила роль. Юридическая проверка, которая по SLA занимала пять дней, теперь занимает минуты.

АГЕНТ АКТИВАЦИИ ЗА 3,8 РУБЛЯ

Автор под ником aski сделал агента мини-курса по нейросетям для тех, кто присматривается к платному продукту. Принцип — «автомат решает, LLM отвечает»: маршрут ведёт код, на ход — не больше двух вызовов недорогой модели, ответы берутся из графа знаний курса. Диалог стоит в среднем 3,8 рубля против примерно 16 на сильной модели без автомата; одна покупка за 10 тыс. рублей оплачивает тысячи диалогов.

ЧЕМ ИНТЕРЕСНО

Обе схемы делят работу одинаково: код держит порядок, ИИ работает внутри шагов, человек подключается по исключению.

В ЦИФРАХ

1 час	вместо 7 дней на путь пуша
22 тыс.+	единиц контента в год
5 вместо 10	контент-менеджеров в команде

БИЗНЕС

Пять ИИ-агентов оценили план заменить координатора агентами — и все проголосовали «нет»: работу агенты бы сделали, но провалилось всё вокруг

Пример с форума r/AI_Agents: панель из пяти субагентов Claude Code разбирает бизнес-решение по общим правилам оценки. Кейс выдуманный, но разбор полезен как чек-лист.

ПРОБЛЕМА

Вымышленное маркетинговое агентство на 22 человека теряет операционного координатора: отчёты для 31 клиента, онбординг, счета, таск-трекер. Новый человек — около $73 700 в год со всеми расходами; консультант предлагает агентов за $12 500 и $1 430 в месяц. Владелец склоняется к агентам.

КАК СДЕЛАЛИ

Авторы собрали навык для Claude Code: пять субагентов работают как комиссия и оценивают решение по одинаковым правилам. Всё, что только заявлено, получает максимум 6 баллов, всё, чего нет, — максимум 4, а «да» требует среднего 6,5. Поэтому каждое «нет» указывает на конкретный пробел. Прогон занял 15 вызовов агентов, около 1,6 млн токенов, примерно $3,70 по прайсу API.

РЕЗУЛЬТАТ

Все пятеро сказали «нет» — и не потому, что агенты не справятся: отчёты занимают 52 часа в месяц и повторяются. Провалилось окружение: 51 из 152 часов координатора — поставщики, офис и люди, которыми никто не владеет; 9 из 31 договора разрешают передавать данные только одобренным подрядчикам; консультант сразу закладывал годовое сопровождение без лимита трат. Предложение комиссии: пилот только на отчётах с выходом через шесть недель — продолжать, если время на отчёты упадёт с 52 до примерно 16 часов в месяц.

ЧЕМ ИНТЕРЕСНО

Комментаторы сошлись: «51 час, которым никто не владеет», — главная причина провалов автоматизации в малом бизнесе. Измеримую часть забирает автоматика, неизмеримая тихо ложится на владельца.

В ЦИФРАХ

5 из 5	агентов сказали «нет»
51 из 152	часов работы не покрывают агенты
$3,70	стоимость всего разбора

ПОДБОРКА ИНСТРУМЕНТОВ ДНЯ

Worktrunk

ЧТО ДЕЛАЕТ
Командная строка для git worktree: у каждого агента своя рабочая папка, команды обращаются к ней по имени ветки.

ДЛЯ ЧЕГО
Запускать 5–10 и больше агентов для кода (Claude Code, Codex) параллельно, чтобы они не мешали друг другу.

ЧЕМ ИНТЕРЕСЕН
Три основные команды, хуки для автоматизации и общий кеш сборки; автор называет его самым популярным менеджером worktree.

ЦЕНА
в источнике не указана

agent-desktop (Vercel Labs) — США

ЧТО ДЕЛАЕТ
Даёт агенту управлять приложениями на macOS, Windows и Linux через дерево доступности ОС, без скриншотов.

ДЛЯ ЧЕГО
«Компьютерное зрение» для агента: снимок интерфейса в JSON, клик по ссылке на элемент, повторная проверка.

ЧЕМ ИНТЕРЕСЕН
Модели внутри нет, ставится через npm или brew. Снимки могут содержать личные данные.

ЦЕНА
в источнике не указана

Kairos (Finic)

ЧТО ДЕЛАЕТ
Открытая платформа решений по мошенничеству и рискам в реальном времени: одобрить, проверить, заблокировать.

ДЛЯ ЧЕГО
Проверка регистраций, входов, заказов и выплат; альтернатива Sift, Sardine, Riskified и SEON.

ЧЕМ ИНТЕРЕСЕН
Есть MCP-сервер: сценарии строят Claude Code, Cursor или Codex. Разворачивается у себя в Docker.

ЦЕНА
в источнике не указана

Agent37

США

ЧТО ДЕЛАЕТ
Изолированные «песочницы» для агентов (Hermes, Claude Code) с постоянным диском и поминутной оплатой.

ДЛЯ ЧЕГО
Держать агентов в отдельной среде дешевле облачных аналогов.

ЧЕМ ИНТЕРЕСЕН
Основатель сравнивает месяц 2 vCPU/4 ГБ: $4,76 против $249 у Vercel. Читатели X уточняют: сравнение при 100% загрузке.

ЦЕНА
$4,76 в месяц за 2 vCPU и 4 ГБ (данные компании)

Claude Code Game Studios

ЧТО ДЕЛАЕТ
Превращает одну сессию Claude Code в «игровую студию»: 49 агентов, 74 навыка-команды, 15 хуков проверки.

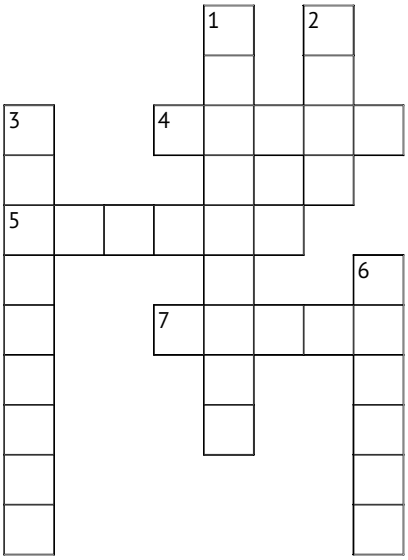
ДЛЯ ЧЕГО
Пример оргструктуры агентов: директора, руководители отделов и специалисты с путями эскалации.

ЧЕМ ИНТЕРЕСЕН
Был в GitHub Trending; все решения, подчёркивает автор, по-прежнему за человеком.

ЦЕНА
в источнике не указана

ПРОВЕРЬ СЕБЯ

по этому выпуску, выпуску №1 и ИИ в целом; ответы внизу вверх ногами



ПО ГОРИЗОНТАЛИ

4. Какая ИИ-проверка Сбера сократила 5 дней до минут (5)

5. Во что Claude Code Game Studios превращает одну сессию агента (6)

7. Валовая __: 85% у Anthropic против 26% у Z.ai (5)

ПО ВЕРТИКАЛИ

1. Как VK называет новую рекомендательную ленту (выпуск №1) (9)

2. Что Astra читала и правила у других нейросетей без обучения (4)

3. Изолированная среда для агента, которую Agent37 сдаёт за $4,76 в месяц (9)

6. Город, где шесть роботов AgiBot сутки вели магазин (6)

ЗАДАЧКИ

1. Сколько часов шла демонстрация агентов Kimi в Unity, Unreal и Blender без людей (выпуск №1)?
   а) 2 б) 20 в) 200
2. Сколько ступеней преодолевали роботы на горе Тайшань (выпуск №1)?
   а) 310 б) 3 100 в) 31 000
3. Какую долю весов Qwen3.5-0.8B изменила GPT-6 Astra, чтобы убрать зацикливание?
   а) 0,016% б) 1,6% в) 16%
4. Во сколько раз, по словам Revolut, свои токены дешевле, чем у Anthropic и OpenAI?
   а) в 2 раза б) примерно в 10 раз в) в 100 раз

ПОДУМАТЬ

- Какие «51 час» в вашей компании агенты не покроют — и кто ими владеет, если человека убрать?
- Где вам нужна одна недорогая модель по умолчанию, а где дорогая — только по исключению? Как записать это правило для агентов?
- Если ИИ может точечно править «мозг» другой модели, какие проверки вы потребовали бы, прежде чем доверить такую правку своему агенту?

Ответы. Кроссворд: 1в — вертикаль; 2в — весна; 3в — песочница; 4г — юрист; 5г — студия; 6в — шанхай; 7г — марка. Задачки: 1) б — 20 часов. 2) б — 3 100. 3) а — 0,016%. 4) б — примерно в 10 раз.

ИСТОЧНИКИ (ФОРМАТ) **Главное** — Telegram-канал (t.me/data_secrets), рассылка (Latent Space), журнал/блог (Simon Willison), форум (r/LocalLLaMA) · **Revolut и дешёвые модели** — соцсеть (X), Telegram-канал (t.me/cgevent), форум (r/ClaudeAI) · **Роботы AgiBot и Xiaomi** — журнал/блог (Pandaily), форум (r/LocalLLaMA) · **Экономика Anthropic и Китая** — форум (Hacker News), журнал/блог (SCMP Tech), Telegram-канал (t.me/techsparks) · **Сбер и агент активации** — журнал/блог (Хабр: ИИ) · **Пять агентов против замены** — форум (r/AI_Agents) · **Инструменты** — Worktrunk — репозиторий GitHub (GitHub Trending); agent-desktop (Vercel Labs) — репозиторий GitHub (GitHub Trending); Kairos (Finic) — форум (Hacker News); Agent37 — соцсеть (X); Claude Code Game Studios — репозиторий GitHub (GitHub Trending)

Сбор за сутки 03.11–04.11.2026 · собрала Ника · стиль v1.0, формат v2