

Агент розібрав п'ять девайсів за 13 годин

Топова історія HN на Opus 5, зроблена тією ж моделлю за 13 годин.

вівторок, 25 серпня 2026

У ЦЬОМУ ВИПУСКУ

1. «Everything I own, owned»: п'ять девайсів розібрано агентом на Opus 5 за 13 годин. 1368 балів - найвища історія доби
2. Загадка вчорашнього пункту 2 розв'язалась: справа не в якості моделі, а в ZDR. «Procurement picks the model, not eval scores»
3. a16z на даних OpenRouter: агенти спалюють у 5 разів більше токенів за людей, і 85% цього - кеш
4. «Coding expertise is going to collapse»: 501 бал і три дослідження, з яких одне міряє шкоду в 17%
5. MS Paint зашиває невидимий вотермарк з серверним GUID навіть у локально згенеровані картинки. 613 балів
6. Наранянан порахував заборони дата-центрів на серветці: рік мораторію в штаті = 5-10 годин затримки прогресу ШІ
7. Мольік: «yes, they can tell» - рецензент бачить ШІ не по мові, а по тому, що ідеї починають римуватись
8. Леві і Тан замикають вчорашню тезу про «системи обліку → харнеси» з боку даних
9. Xiaomi Xring O3 наздогнав ядра Apple: 44 МБ кешу, 21 порт виконання. 777 балів
10. Дрібне, але робоче: IPFS втратив фінансування, і Replit Agent проти Claude CoWork

ТЕМА 1

«Everything I own, owned»: п'ять девайсів розібрано агентом на Opus 5 за 13 годин. 1368 балів - найвища історія доби

Стаття [schlarpc](#) (23.08) - 1368 балів на HN, [тред](#).

Автор за два тижні вечорів згодував агенту прошивки п'яти периферійних пристроїв, що лежали під рукою. Модель названа прямо: **Claude Opus 5**. І, що рідкість, зусилля виміряно з транскриптів сесій Claude Code:

ПРИСТРІЙ	ЧАС РОБОТИ АГЕНТА	ПРОМПТІВ ВІД ЛЮДИНИ
Shure MV7 (мікрофон)	4,2 год	32
Insta360 Link (вебка)	3,7 год	33
Elgato Key Light Mini	2,4 год	10
Elgato Cam Link 4K	1,5 год	10
ASUS ROG Swift PG42UQ	1,2 год	13
Разом	~13 год	98

Що знайшлося (більшість перевірена наживо на залізі):

- **Вебкамера пише з вимкненим світлодіодом.** У прошивці таблиця «патернів» LED; агент написав тул, який виправляє запис для стану «запис», перераховує MD5 і зашиває назад. Захисту від підробки нема, просто дописаний MD5. Плюс канал USB vendor class дає довільне читання/запис файлів і ребут, тобто **прошити можна без жодної дії користувача.**
- **Мікрофон містить повноцінний плейнтекстовий командний шелл** - 48 команд поверх USB HID, доступний навіть з веб-сторінки через WebHID. Є чотирирівнева система привілеїв, і вся «автентифікація» - це порівняння рядка з назвою рівня: дослівно «**su sup just works**». Верхній рівень уміє вимкнути кнопку мюту і **керувати індикатором мюту незалежно від того, чи мікрофон реально заглушений.**
- **Cam Link 4K автор пустив повністю без нагляду:** запустив перед сном, прокинувся з готовим розбором і робочим оновлювачем прошивки.

Чому периферія виявилась ідеальною мішенню: «they're **tiny computers attached to my computer**, with a data connection to the host and usually a firmware update mechanism, so **an agent has something to iterate against**».

Чому це важливо. По-перше, це **прямий замір продуктивності на цьому стеку** - реальна задача з валідацією на залізі: 2,6 години і ~20 промптів на пристрій у середньому. Наступного разу, коли постане питання «скільки це займе», ось калібрувальна точка.

По-друге, це **зміна моделі загроз для домашнього залізяччя.** Ключовий коментар у треді [@teddyh](#) цитує автора: WebUSB/WebHID/WebBluetooth означають, що «**a moment of user indiscretion in accepting a permissions prompt could permanently backdoor one of their attached devices**». Раніше таке коштувало тижнів роботи реверсера, тепер це вечір. Прикладне: індикатор запису на вебці більше не є доказом, що камера не пише. Мітка [proven] - код у відкритих репозиторіях, перевірений на живих пристроях.

Загадка вчорашнього пункту 2 розв'язалась: справа в ZDR. «Procurement picks the model, not eval scores»

Місток до вчорашнього випуску, де були наведені дані Ramp – Orpus 5 має 3,5% витрат проти 28% в Orpus 4.8 – з чесним визнанням, що причина невідома. За добу вона знайшлась.

Мартін Касадо (a16z) поставив питання вголос (24.08, 07:39 – 188 тис. переглядів): «Lots of folks jumping to conclusions on the FT market share data for Fable. **The reason is not at all obvious to me.** Is it due to cost? Speed? Aggressive moves from OAI? ZDR? Open Source? Bad press / vibes? Refusal to work in some domains? Biased data?»

Через сім годин – **його ж висновок** (15:08): «**Consensus on this seems to be ZDR**, which tracks to my own experience as well. If that really turns out to be the case, it's an **incredibly strong and clear message to the labs** on the impact of data retention policies».

ZDR = zero data retention, договірна умова «ваші дані ми не зберігаємо». **Аарон Леві розгортає механіку** (24.08, 15:23, 76 тис. переглядів): «It's not fully appreciated that **ZDR is responsible for a substantial amount of the growth of AI**, because it dramatically simplified the compliance process most companies deal with in handling data with subprocessors». Далі конкретика: прикладні AI-тули домовляються з клієнтами пропонувати **тільки ZDR-моделі**, бо інакше кожную модель треба проводити окремим винятком; а всередині компаній правила часто прямо кажуть «лише ZDR», бо **відділити PII від решти контексту в вікні технічно не вміють**. Висновок Леві: «**Without ZDR, AI diffusion grinds to a halt**».

Найконкретніше – у відповіді практика з боку закупівель **@iabdul_wasey**: «ZDR is the first line of every security review, and a model without it turns a **two week approval into a quarter**. Teams pick the **second best model with ZDR** over the best one without. **Procurement picks the model, not eval scores**».

Що лишилось неперевіреном, сказано прямо: це консенсус галузевих людей у **твітері**. Документа нема. Касадо сам формулює умовно («if that really turns out to be the case»). Anthropic офіційно цього не підтверджувала, і сама стаття FT досі нечитабельна (ft.com віддає 403 і на справжню, і на завідомо вигадану адресу – перевірено вчора). Тому мітка [promising], не [proven]. Але механізм принаймні **пояснює аномалію**, чого не робила жодна вчорашня гіпотеза.

Чому це важливо. ZDR це корпоративна історія, тож практичного на сьогодні нуль. Але два висновки варті того, щоб лежати в голові. Перший: **найкраща модель програє швидшому юридичному погодженню**, і при виборі стека для продукту під клієнта питання «а що з retention» треба ставити до питання «а який у неї бенчмарк». Другий, ширший: вчора пояснення не вигадувалось, коли його не було, і за добу воно прийшло від людей з ринку. Той самий принцип, що з чужими цифрами: **краще чесне «не знаю», ніж правдоподібна вигадка**.

ТЕМА 3

a16z на даних OpenRouter: агенти спалюють у 5 разів більше токенів за людей, і 85% цього – кеш

Charts of the Week від a16z – репостнув Грег Брокман (24.08, 146 тис. переглядів) з підписом «a new way to get knowledge work done». У твіті були лише множники по професіях, найцікавіше – глибше в самому звіті.

Головні цифри (джерела названі: дані OpenAI, OpenRouter, Similarweb):

- Агенти використовують майже вп'ятеро більше токенів, ніж люди, і агентне використання зросло ~14× з лютого.
- Понад 85% агентного спалювання токенів – це кешований промпт, і саме кешовані токени дають майже весь відносний приріст. Людина працює в режимі «промпт-відповідь», агент **ітерує до мети**: контекст докладається один раз і далі читається-пишеться інкрементально.
- Розрив між типовою і топовою компанією – ~8-кратний за виводом токенів; у техсекторі топовий дециль виводить у 32,5 разів більше токенів, ніж рік тому.
- Прийняття Codex по професіях з лютого 2026: юристи 108×, продажі 41×, рекрутинг 41×, маркетинг 26×, медицина 24×. Найшвидше ростуть **не інженери**.
- Побічний ефект, який мало хто помітив: **трафік на legacy-автоматизацію падає**, N8N, Zapier і Make всі показують двоцифрове зниження за останні 12 тижнів.

Чому це важливо. Цифра 85% кешу пряма про економіку конструкції і пояснює, чому патерн «важкий сталий контекст + багато дрібних кроків» працює дешево: **великий незмінний префікс (інструкції, довідники, пам'ять) кешується, платиться переважно за приріст**. Це аргумент **на користь** підходу з великими стабільними наборами інструкцій, і рівно тому вчорашня теза Бройніга про «харнес важить більше за модель» має ще й фінансове дно.

Друге: падіння Zapier/n8n – рання ілюстрація вчорашнього прогнозу Гаррі Тана. Постзапієрний світ уже настав (скрипти, розклад і набори інструкцій замість візуальних воркфлоу), і настав тому, що вийшло дешевше і гнучкіше.

ТЕМА 4

«Coding expertise is going to collapse»: 501 бал і три дослідження, з яких одне міряє шкоду в 17%

Есей Ларса Фея (24.08) – 501 бал, **тред**. Текст тримається на трьох цитованих дослідженнях.

Центральний парадокс сформульований чисто: «If these tools **demand expertise**, yet the tools can actively **circumvent the friction that cultivates expertise**, then what is the path for one to become an expert so they can effectively use these tools?»

Дослідження:

- **JetBrains** цитує роботи «**The Widening Gap**» (аналіз живих сесій новачків):
«Participants thought it was like having a personal tutor. From the data in our study... we observed that **they did not, in fact, use GenAI tools like a personal tutor. In fact, it was quite the opposite**». Ті, хто спирався на III важче, «often **skipped crucial planning stages**» і закінчували з «**illusion of competence** rather than true understanding».
- Ті, хто впорався, мали те, що автори називають «**negative expertise**» – «the ability to **ignore incorrect or unhelpful GenAI suggestions**».
- **UPenn, 2025, «Generative AI without guardrails can harm learning»**: 1 000 студентів, ті, хто вчив математику з LLM без обмежень, показали результат **на 17% гірший** за тих, хто мав лише підручник, і при цьому **вважали, що вчать краще**. У тому ж дослідженні версія «GPT Tutor» (модель допомагає, студент розв'язує сам) дала протилежний результат.

Найтверезіший коментар у треді – від автора у відповідь на «а як тоді вчити»: він посилається на **Дядька Боба**, який «as AI piled as they come», але каже, що джуни **не повинні торкатись III-тулів перші три роки**.

Чому це важливо. Для досвідченого розробника це випадок «ossified knowledge», де інструмент чистий виграш. Цікаві дві інші речі. Перше – наймання: критерій «вміє працювати з агентом» і критерій «розуміє, що агент зробив» це різні критерії, і друге не перевіряється по демо. Друге – навчання дітей: висновок дослідження конкретний, **режим «дай відповідь» шкодить, режим «допоможи, а розв'язує сам» допомагає**. Коли учень питає щось по курсу, правильна поведінка сократична, без готової відповіді. Мітка [proven] для самих досліджень, [promising] для екстраполяції автора на індустрію.

ТЕМА 5

MS Paint зашиває невидимий водермарк з серверним GUID навіть у локально згенеровані картинки. 613 балів

Реверс-інжиніринство **Сюшена Лі** (24.08) – 613 балів, тред.

Механіка, розібрана дизасемблером: Paint і Photos **справді містять локальні моделі** (чотири `.onnx` файли, найбільший 302 МБ – XOR-нуті ONNX, ключ лежить у `segapi.dll`). Але перед локальною генерацією `AIServices.dll` **шле промпт на сервер Microsoft** на модерацію, і сервер повертає **GUID**. Цей GUID (16 байт) далі зашивається в самі пікселі через `Watermarker.dll!WmkWriteWatermark` – content-adaptive блоковий водермарк у стилі SVD. На тестовій картинці 512×512 змінилось **193 376 із 262 144 пікселів**.

Дві деталі, які роблять це не дрібницею:

- **Окрема настройка «видимий водермарк» цим не керує**. Видимий логотип Copilot вимикається; невидимий GUID – ні.

- Якщо ватермарк не вдалося записати, Paint перетворює всю генерацію на помилку замість того, щоб віддати картинку без нього. Тобто це обов'язкова умова.

Чесно про джерело, бо це перший коментар у треді. Сабмітер (не автор) сам повісив попередження: «AI-generated text warning», і другий коментатор упізнає в тексті «Claude flavored» абзаци. Перевірка пройшла по суті. Стиль винесено за дужки: **самі знахідки предметні і відтворювані** - назви DLL, сигнатури функцій (`Paint::AI::AddWatermark(Gdiplus::Bitmap&, winrt::guid const&)`), розміри моделей, коди помилок `-5 / -6`, структура пейлоада `0x4c || GUID || контрольна сума`. Це не те, що галюцинується. Стиль викладу ШІ-шний, зміст - робота реверсера. Домен перевірку проходить чесно (**404 на вигадану адресу**).

І окремо: автор пише, що надихнувся «recent Claude Code text-watermark announcement». Перевірка цього не підтверджує: у CHANGELOG.md `claude-code` (взято через `api.github.com` сирым текстом, 543 КБ, перевірено що це справді `# Changelog`, остання версія 2.1.241) слово «watermark» дає **нуль хітів**. Ремарка про Claude Code лишається неперевіреною, саму знахідку по Paint це не зачіпає.

Чому це важливо. «Згенеровано локально» ≠ «не пішло нікуди»: промпт усе одно їде на сервер. Висновок ширший за Paint і стосується будь-якого «локального ШІ» в ОС - **перевіряти треба трафік**. Маркетингова обіцянка тут нічого не важить. Клас помилки той самий, що трапляється в дайджесті щодня: **індикатор у налаштуваннях ≠ реальна поведінка коду**.

ТЕМА 6

Наранянан порахував заборони дата-центрів на серветці: рік мораторію в штаті = 5-10 годин затримки прогресу ШІ

Пост Арвінда Нараняна (24.08, 15:11, 66 тис. переглядів), підтриманий Молліком.

Логіка розрахунку: ефективність інференсу на заданому рівні здібностей стабільно росте, тому «заблоковано N ГВт» можна перевести в **час**, за який та сама здібність доганяється самою лише ефективністю. Результат: «if a typical U.S. state enacts a 1-year moratorium, it **slows AI efficiency progress by 5-10 hours**». А з урахуванням того, що майданчики взаємозамінні (припущення 90% «витікання» в інші штати), навіть повна заборона в такому штаті, як Нью-Йорк, - «**less than a day**».

Його ж уточнення: ефективність дає **на порядок більше прогресу, ніж ріст дата-центрів** (5×/рік консервативно проти 25%/рік), а заборони «do nothing about all the capacity already built».

Найцінніше тут - зноски автора. Дослівно: «While I may have gotten some numbers slightly wrong (I used AI for the analysis and did some spot checks), the rough order-of-magnitude is enough to make my larger point». Він публікує розрахунок, зроблений ШІ, і **каже це вголос у тому ж пості**. Моллік з цього приводу **підсумовує**: «My AI broadly agrees with Arvind's AI».

Чому це важливо. Це рівно та дисципліна, якої тримається цей файл: **сказати, звідки взялась цифра і де межа її надійності**, і не подавати оцінку як вимір. Наранянан не ховає, що рахував агентом, і не видає порядок величини за точність. Це зразок формулювання. Фактом про політику США воно не є.

ТЕМА 7

Моллік: «yes, they can tell» - рецензент бачить ШІ по тому, що ідеї починають римуватись

Пост @emollick (24.08, 20:54, 31 тис. переглядів, 425 лайків): «Note to anyone sending in an application, paper, or any other document where **one person reads many submissions: yes, they can tell**. It is obvious in most cases that you are using AI, **even if the language is changed**, as the ideas & concepts start to **rhyme as you read lots of them**».

І одразу **сам себе уточнює**: «I don't think there is anything wrong with getting AI help in writing or feedback or whatever... but **if you use it to do too much, it starts to obscure your unique contribution**».

Механізм у **розподілі**: одна заявка виглядає нормально, але коли їх сто, у них однаковий кістяк аргументації. Найрозумніша репліка в гілці @zazmic_inc: детекція краще працює **порівнянням текстів між собою**, ніж вироком по одному тексту.

Чому це важливо. Для будь-якого публічного письма ризик один: **втрата того, заради чого текст читають** - конкретних цифр з практики і власних формулювань. Це підпирає правило «не міняти тон» технічним аргументом: тон - це і є та частина, яка не римується з рештою стрічки. Драфт - так, переписування голосом - ні.

ТЕМА 8

Леві і Тан замикають вчорашню тезу про «системи обліку → харнеси» з боку даних

Продовження вчорашнього пункту 9. Учора **Гаррі Тан кинув прогноз** одним реченням (за добу - 384 тис. переглядів, 2,6 тис. лайків: цифра виросла втричі). Сьогодні на нього **відповів Аарон Леві** (25.08, 03:12 - свіжак, вже 20 тис. переглядів):

«Systems of record have never been more important than in a world where you have AI agents that will do **100X more work** on these platforms than people ever did. Agents will be querying the data in these systems, processing tasks, executing workflows, collaborating with human and...»

І **сам Тан уточнив свою тезу**: «There will still be **API's and acl's and sql and underlying data structures that are deterministic**. It's just that software companies have to build the AI harness and full solution for their customers **or be subsumed by it**».

За добу прогноз із «системи обліку помруть» перетворився на точніше: **дані і детермінований доступ нікуди не діваються, вмирає монополія на інтерфейс до них.**

Чому це важливо. Уточнення Тана описує архітектуру, до якої вже приходять на практиці: база на SQLite, таблиці, API пристроїв – детерміновані системи обліку, а набір команд і інструкцій поверх них – харнес. Таблиця не замінюється агентом; агенту даються дисципліновані двері. Пункт 3 (85% кешу) пояснює, чому це ще й дешево.

ТЕМА 9

Xiaomi Xring O3 наздогнав ядра Apple: 44 МБ кешу, 21 порт виконання. 777 балів

Розбір Даніеля Лемір (24.08) – 777 балів на HN, **тред**. Новий процесор Xiaomi «roughly matches Apple cores on single threaded tasks, and is **much faster in multithreaded execution**».

Специфіка: **44 МБ кешу загалом** – більше, ніж у більшості ноутбучних процесорів; великі ядра C1-Ultra підтримують SME2 (матричні розширення під ШІ) і SVE2; ядро «astonishingly wide» – **21 порт виконання**. Лемір одразу дає поправку: Apple ось-ось анонсує наступний процесор, тож перевага може бути недовгою, і знайти телефон з O3 буде складно.

Чому це важливо. Нічого прикладного, це фон до пункту 3 з учора (локальні моделі): SME2 у мобільному чіпі означає, що інференс на пристрої стає нормою. HN-тред при цьому здебільшого про геополітику, цю частину не переказано, вона поза темою дайджесту.

ТЕМА 10

Дрібне, але робоче: IPFS втратив фінансування, і Replit Agent проти Claude CoWork

IPFS. Shipyard оголосив згортання (24.08) – 339 балів, **тред**. Protocol Labs не продовжує фінансування; останній день робіт над IPFS – **30 вересня 2026**. У підсумках їхньої роботи цифри, які показують, що команда була жива: перебудова інфраструктури гейтвеїв витримала ~3× **більше трафіку при зниженні витрат на ~80%**. Тред майже цілком пішов у розбір «NFT це просто лінк». Саму технологію майже не обговорювали, що вже епітафія.

Replit. Амджад Масад заявляє (24.08, 16 тис. переглядів): «Replit Agent completely replaced Claude CoWork in my day-to-day work. It's much more persistent and fastidious, and it uses code/software more effectively to complete tasks». Це CEO про власний продукт, маркетингова заява, подана як позиція. Цікаве в ній інше: критерій порівняння

- наполегливість і використання коду як інструмента, тобто рівно те, про що вчора писав Торвальдс («модель здається раніше, ніж вичерпано способи»).

misc: @garrytan про Conductor Cloud - 38 тис. переглядів, «не треба тримати макбук відкритим», форм-фактор remote-сесій · @garrytan, петля роботи з реальністю - «Form a view. Turn it into an artifact. Put it in contact with reality. Read the result without self-deception. Revise and run again» · @emollick про споживчі агенти 35 тис. переглядів - фронтір-моделі добре в «irregular, draining tasks»: медичний рахунок, спір по картці, нудні форми · @emollick про зсув дискурсу - «занадто хитнулись у бік "вплив ШІ проявиться через роки"» · @emollick про анонси лабораторій - вимагає замість туманних постів «Draw a unicorn in TiKZ» · @ycombinator: 20+ стартапів YC публікують дослідження на NeurIPS/ICLR/ICML, блог · @paulg «How Universities Should Prepare Founders» - 60 тис. переглядів, есей · @blader про gpt-image-2 - «дуже смаковитий дизайнер»: зациклювати згенерований дизайн краще, ніж просити дизайн з нуля · @amasad про Replit-мілстоун 26 тис. переглядів · Виконуваний файл як база SQLite - 508 балів, тред · LLM можуть захопити хост через діри в інференс-рушіях - 111 балів, тред: атака на сам vLLM/SGLang, тому «правильний харнес» тут не захист, захист це VM · Докази seL4 завершені на AArch64 - 176 балів · Сергій Левін про стан гуманоїдної робототехніки - єдиний свіжий Substack у вікні