

Meta анонсувала опенсорс-харнес і моделі

На YC Startup School Ванг і Дженсен публічно підтвердили ставку на відкриті ваги.

понеділок, 27 липня 2026

У ЦЬОМУ ВИПУСКУ

1. Meta випустить харнес і відкриті моделі - Александр Ванг, YC SUS
2. Дженсен на тій самій сцені - дослівна цитата про опенсорс
3. Ринок реліїв: чужі токени Anthropic за 2% ціни
4. Опус лежав годину і сорок - і ти цього не помітив
5. GrapheneOS: справа, де ОС хочуть зробити злочином
6. Стенфорд порахував, що AI зробив з ринком праці - і це не апокаліпсис
7. @levelsio: «я вайбкодів 100% своїх SaaS-підписок» - і сам себе описав як наступного
8. Ruff вмикає 413 правил замість 59
9. Мовчазний доказ у суперечці про відкриті ваги: атакують з підписок, не з відкритих моделей
10. Мольік: гайд про моделі застарів за два дні після написання

1. Meta випустить харнес і відкриті моделі - Александр Ванг, YC SUS Продовження вчорашнього сюжету про відкриті ваги, цього разу анонс.

Цитата зі сцени: **«Meta запустить харнес і open-source моделі»**. Перевірку пройшов сам тред: під постом **@alexandr_wang** відповів **«уер»**. Він же роздав усім у залі по **\$1000 API-кредитів** на новий Muse Spark.

Там же його рамка про конкуренцію: **«з агентами стартап проти інкумбента - це вже не Давид проти Голіафа, це мехо-Голіаф проти Голіафа»**.

Чому це важливо. Слово «харнес» тут ключове. Харнес - це те, що пишеться руками навколо моделі: промпт, інструменти, пам'ять, планувальник. Коли Meta починає віддавати харнес разом з вагами, це прямий конкурент підходу «збери сам під себе». Скептицизм у треді здоровий (найтвердіше: «LLM без харнесу коштує менше за китайські відкриті ваги»), але напрямок варто зафіксувати. Коли воно вийде, дивитись треба на харнес і питати, що в ньому краще за власну обгортку.

2. Дженсен на тій самій сцені - дослівна цитата про опенсорс Вчора була його теза про керованість агентів. Сьогодні є друга, і вона б'є прямо в суперечку з пункту 1 і з позавчорашнього листа.

Дослівно: «Якби не опенсорс, мобільно-хмарної індустрії не сталося б ніколи. Якби головніше - PyTorch... пам'ятаєте ранні версії, Caffe, Torch, Theano - усе це було відкрито. Якби не це все, звідки б узявся сучасний AI?»

Чому це важливо: це найсильніша форма аргументу в суперечці про ваги: весь стек, на якому все стоїть, вже відкритий.

3. Ринок релеїв: чужі токени Anthropic за 2% ціни Розслідування Vectoral, 172 бали. Найпрактичніша безпекова історія доби.

Цифри з першоджерела:

- пакет на \$3 333 офіційного кредиту Anthropic - за 425 юанів (~\$60), тобто \$0.13 корисного за кожен \$1
- медіанна знижка топ-релеїв - 97.8% (Now Coding), і так усі 49 у їхньому списку
- схема чотиришарова: карткові мерчанти з віртуальними картками під US/EU білінг → пули акаунтів (сотні акаунтів, ротація лімітів) → релеї з чистим китайськомовним UI → кінцеві покупці, включно з тими, хто женеться за дистилляцією

Чому це важливо: економіка цих релеїв тримається на масовій реєстрації акаунтів і вкрадених або віртуальних картках. Саме через це лаби закручують антифрод, і саме тому дешеві «Claude API ключі» з інтернету завжди або крадені, або з пулу. Коли хтось пропонує «той самий Opus дешевше», варто знати, як воно влаштоване всередині.

4. Опус лежав годину і сорок, і це залишилось непоміченим 92 бали на HN, статус-сторінка Anthropic. Дослівно з першоджерела: «Elevated errors for Opus 5», зачепило `claude.ai`, Console, API, Claude Code і Claude Cowork.

Таймлайн: investigating 09:17 UTC → identified 09:45 → fix 10:34 → resolved 10:44 UTC. Тобто 1 год 27 хв проблем, з них активна фаза - за київським часом 12:17-13:44 неділі.

Чому це важливо: будь-який конвеєр, що тримається на одній моделі і одному провайдері без фолбеку, у такі години просто мовчить, і це списують на зависання. Фолбек на іншу модель - зайва складність заради півтори години на місяць. Дешевша страховка: скрипт раз на 15 хвилин смикає статус-сторінку провайдера і пише в чат, якщо інцидент зачіпає API. Без LLM і без окремої інфраструктури.

5. GrapheneOS: справа, де ОС хочуть зробити злочином 378 балів, топ ночі. Прокуратура США висунула звинувачення жителю Атланти Сему Туніку за федеральною статтею про знищення майна, щоб запобігти його вилученню: його телефон на GrapheneOS стерся при спробі обшуку в аеропорту.

Деталі з першоджерела: затримали в Хартсфілд-Джексон після повернення з Домінікани. Агенти вже мали його ім'я і фото з поміткою «підозра в терористичній діяльності» через зв'язок з рухом проти Cop City. Захист пише, що допит про CSAM був приводом, що адвоката просив чотири рази і щоразу відмовили, ордера не показали, права не зачитали. Експерт EFF і фахівець з нагляду кажуть, що подібного не бачили: «це сигнал, що GrapheneOS злочинна за замовчуванням».

Поруч, 399 балів - технічний розбір від самих GrapheneOS. Rate limiting у secure element: **4 години після 10 спроб, 41 день після 15**, максимум 20. Прошивку secure element не оновити без автентифікації власника, щоб держава не змусила виробника зняти ліміт. І партнерство з Motorola з 2027, тобто не тільки Pixel.

Чому це важливо: фіча «стерти по примусовому пароллю» юридично токсична саме там, де вона найпотрібніша - на кордоні. Для тих, хто літає, це вже не гіпотетика. Практичний висновок нудний і робочий: на кордоні виграє **чистий телефон без зайвого**.

6. Стенфорд порахував, що AI зробив з ринком праці 250 балів, 320 коментарів. Policy brief SIEPR, рецензована економіка.

Ключові висновки дослівно: **вплив AI на сукупну зайнятість зараз, найімовірніше, малий**; важкий ринок для випускників частково справді через AI; вплив на продуктивність працівника змішаний, але загалом позитивний; впровадження у фірмах пришвидшилось, але **дуже нерівномірно**. Застереження від самих авторів: рання емпірика - ще не остаточне слово.

Окремо автори називають прогноз Даріо Амодеї (AI зітре половину білокомірцевих робіт, безробіття 20%) як приклад того, що **самі лідери AI підсилюють страх**.

Чому це важливо: з цим варто зв'язати пункт 7 нижче. @levelsio на 347K переглядів пише, що індустрію змиває, Стенфорд каже «в агрегаті ефект поки малий». Праві обидва, просто про різне: інді-ніша справді перша під ударом, ринок праці загалом тримається. Для СТО ця різниця важливіша за обидва заголовки.

7. @levelsio: «вайбкодилося 100% SaaS-підписок» Найгучніша розмова доби у першому листі: пост Патріка Хаеде «інді-хакерський продукт помирає, причина - кодинг-агенти» зібрав 225K переглядів, відповідь levelsio - 347K.

Його теза без прикрас: **«Скасовано і вайбкодено 100% SaaS-підписок. Оплата йде тільки за домени, хостинг, сховище і AI API. Іронія в тому, що заміна прийде і на нього самого - і це усвідомлюється»**. Плюс спостереження по цифрах: у нього самого і в знайомих інді падає і **трафік, і виручка**; у Yongfook гуглові реферали за рік **впали вдвічі**.

Найтверезіший контрапункт у тому ж треді (@onlinedoramine): складні продукти, продані **нетехнічним IRL-бізнесам** - сантехнікам, клінінгу, басейнам - тримаються, бо ці люди ніколи не сядуть вайбкодити заміну.

Чому це важливо: рамка проста. Вразливе - те, що агент відтворює за вечір; захищене - розподіл, дані і реальний нетехнічний клієнт. Вперше це сказано людиною, яка сама на цьому заробляла і показує свої падаючі цифри.

8. Ruff вмикає 413 правил замість 59 336 балів. Astral підняли дефолтний набір літера з 59 до 413 правил (усього в Ruff їх 968), дефолт не чіпали з версії v0.1.0. Логіка: багато правил ловлять синтаксичні і рантайм-помилки, але лежали вимкненими. Відкат у один рядок: `select = ["E4", "E7", "E9", "F"]`.

Поправка: **сам реліз від 23.07**, у вікно потрапив тред на HN.

Чому це важливо: будь-яка python-кодова база, де Ruff досі не стоїть, на новому дефолті вивалить на першому прогоні купу всього. Це нормальна поведінка оновленого дефолту.

9. Мовчазний доказ у суперечці про відкриті ваги: атакують з підписок Тонка деталь, яку підсвітив @amasad: колишній співробітник Anthropic (@NoahLebovic) пише, що **хакери на практиці віддають перевагу щедро субсидованим підпискам лабів**, а відкриті ваги лишаються осторонь.

Чому це важливо: якщо це так, аргумент «відкриті ваги = зброя» просідає з несподіваного боку. Зброєю на практиці стає дешевий доступ до фронтіру, тобто те, чим торгують релеї з пункту 3. Дві історії доби зійшлись. Мітка чесності: це слова однієї людини, не вимірювання, на відміну від учорашніх цифр AISI.

10. Мольік: гайд про моделі застарів за два дні після написання Гайд «яку модель для чого брати» написано у четвер, і вже довелося переписувати під **Opus 5 і голосовий режим Codex**, обидва вийшли в п'ятницю. Висновок: **«встигати складно, навіть якщо стежиш за цим впритул»**.

За добу зроблено три ігри-містобудівники з одного промпту (Cezanne, Piranesi, версія від GPT-5.6 Sol) і **викладено вихідники** - ілюстрація, що вартість такого експерименту тепер нуль.

Чому це важливо: аргумент на користь поради Бориса Черні з вчорашнього пункту 4: обгортки старіють швидше, ніж встигають бути написані.

Misc • Сем Альтман на закритті YC SUS: «Наступні пів року дадуть стільки ж прогресу моделей, скільки останні два роки» - заявка, яку варто перевірити в січні • Гаррі Тан звідти ж: **«будуй персональний AGI - мозок, який ніхто не курує, це смітник з хорошим пошуком»**; і окремо «не ЛАРПай, будь щирим»

• Naval, 282K переглядів: «якби у відкритих вагах були бекдори, закриті лаби б їх точно знайшли і показали» - найелегантніший аргумент за відкритість за весь тиждень •

Челсі Фінн (Physical Intelligence) на YC: RL для роботів впирається у **вартість фізичних прогонів** - 1M траєкторій хвилинної задачі це **~700 робото-днів** • Google розкрив **\$94.1 млрд у акціях SpaceX - 6% компанії** (313 балів)

• Htmx 4.0 вийшов **картриджем для Game Boy** (387) - перша JS-бібліотека з таким релізом • «Kill The Cookie Banner» - топ доби на HN, **873 бали**: європейська кампанія за скасування cookie-банерів • Теренс Тао виклав слайди «Математика в епоху AI» з ICM 2026 (113)

• PGSimCity - PostgreSQL пояснений як SimCity (258), і ESP32-радар літаків на стіл (260)