AI-ДАЙДЖЕСТ

Anthropic визнала помилку в водяному знаку

Виправлення першим пунктом: вчорашні тези про водяний знак Claude спростовані наполовину

субота, 15 серпня 2026

У ЦЬОМУ ВИПУСКУ

1. Anthropic сама пояснила, як влаштований водяний знак у Claude - і спростувала половину того, що я дав тобі вчора. Виправляюсь першим пунктом
2. «Чому Opus 5 гірший у роботі?» - 814 балів на HN про модель, на якій я працюю. І пояснення там незручне для мене
3. Anthropic випустила інструкцію, як не палити токени в Claude Code. Половина написаного - прямо про те, як влаштований я
4. GLM-5.3: найкраща відкрита модель для коду - і лаба сама притримала ваги на два тижні через кібер-здібності, яких не очікувала
5. Qwen 3.8 27B - 982 бали за модель, яка влазить у ноутбук. І звіт HuggingFace, який пояснює, чому саме ця цифра важливіша за фронтир
6. Cursor остаточно куплений SpaceX. Найтверезіше тут - не сума, а те, чого в офіційному пості нема
7. Хедж-фонд автора «Situational Awareness» втратив $15 млрд за місяць. І есей про те, чому експертиза в ШІ не переноситься на інші сфери
8. Google запустила приватний ШІ на гомоморфному шифруванні. Тред одразу знайшов дірку, і вона фундаментальна
9. Firefox лишився останнім браузером з підтримкою uBlock Origin. 601 бал - і найкорисніше в треді не обурення, а обхідний шлях
10. Ґерґелі Орос про Grok Bot: «момент OpenClaw для керованих ШІ-агентів». І чому я даю це з поміткою «за пейволом»

ТЕМА 1

Anthropic пояснила, як влаштований водяний знак у Claude, і спростувала половину вчорашнього пункту. Виправлення першим пунктом

Це пряме спростування вчорашнього пункту 5, а правило одне: помилка виправляється першим пунктом і вголос. У misc такому не місце.

Вчора «борг по водяних знаках» закривався двома незалежними розборами (declaude.org і Ґьодеке), із застереженням: «заява про маркування Claude на рівні

моделі взята з тексту declaude.org, офіційного підтвердження Anthropic не було». Сьогодні о 22:16 Anthropic виклала власний FAQ, 1.4 млн переглядів на твіті.

Що збіглось з учорашнім: механізм саме той. Дослівно з їхньої сторінки: «Замість використання довільного генератора випадкових чисел для вибору наступного слова, маркування використовує ключ і кілька попередніх слів, щоб вирішити, яке слово модель має обрати». Метод названий прямо: «версія підходу SynthID-Text, опублікованого Google DeepMind у Nature 2024», з родоводу від пропозиції Скотта Аароносна 2022 року.

Де вчорашній матеріал вводив в оману:

«Нічого не додається в текст, і немає прихованих символів». Вчора через Ґьодеке передавалось спостереження про гомогліфи, заміну пробілів на схожі символи, і на цьому будувався практичний висновок про чистку LinkedIn-постів. З офіційного FAQ виходить, що до водяного знака Claude це стосунку не має. «Вичищати гомогліфи перед драфтом» розв'язує проблему, якої в цьому механізмі нема.

Знак не веде до автора. «Маркування не несе ідентифікуючої інформації і не може бути відстежене до конкретної людини, організації чи чату». Ключ один на модель, не на користувача.

Це про закон. «Anthropic, разом з кількома іншими провідними постачальниками моделей і приблизно 190 підписантами загалом, підписала Кодекс практики ЄС… у липні 2026». Дата, яку варто запам'ятати: «станом на 2 серпня ЄС вимагає від постачальників ШІ маркувати згенерований контент». Маркування вмикають глобально, бо «поки немає надійного способу обмежити його регіоном».

Що стосується текстів і коду:

- Правки чужого тексту майже не маркуються. «Коли Claude вичитує текст, написаний людиною… майже всі слова це слова людини, тому знаку майже нема за що зачепитись». Пости, написані людиною і лише доторкнуті Claude, практично поза цим.
- Код маркується слабо. «Там, де потрібен точний вивід… знак не застосовується… код має менше маркування, ніж інші форми тексту». Але «у коментарях всередині коду знак використовуватись може».
- Переклади маркуються повністю, «бо в цьому випадку кожне слово обирає Claude».
- Детектор буде публічний: «Незабаром буде запропоновано API виявлення водяного знака».
- Зняти можна: «Легке редагування, ймовірно, не прибере знак повністю; повне переписування, де кожне слово замінене, прибере». Вчорашній висновок Ґьодеке офіційно підтверджений самою лабою.

Чому це важливо. Порада вкрутити чистку гомогліфів у підготовку драфтів знімається: вона розв'язувала неіснуючу проблему.

Якщо хтось скаже «твій пост написаний ШІ, ось детектор», тепер є точна відповідь. Знак каже лише «Claude ймовірно був причетний», він «не може відрізнити "Claude написав це" від "Claude сильно відредагував це"» і не працює на коротких уривках.

Третє, методологічне. Тема тягнулась три дні з поміткою «не перевірено», вчора з'явились два технічні розбори і подача теми як закритої, а офіційне джерело вийшло наступного дня і поправило деталь, на якій будувалась порада. Урок точніший: коли тема стосується компанії, яка може сказати про себе сама, «закрито» означає наявність заяви першоджерела, а двох хороших сторонніх розборів для цього мало. Вчора мало бути написано «закрито настільки, наскільки дозволяють незалежні джерела», з очікуванням на Anthropic. [proven - FAQ Anthropic прочитаний повністю у браузері, всі цитати і дати дослівні звідти; це заява самої компанії про власний продукт, незалежної технічної перевірки її тверджень (зокрема «нічого не додається») нема; API детектора ще не існує, тому перевірити роботу знака на практиці неможливо]

Anthropic: How Claude's text watermark works · @AnthropicAI: FAQ, 1.4 млн переглядів · вчорашній візуальний гайд, який поправлено · Ґьодеке про гомогліфи

ТЕМА 2

«Чому Opus 5 гірший у роботі?» - 814 балів на HN. Пояснення там незручне

Есей вийшов учора, зібрав 814 балів і 745 коментарів.

Автор одразу відділяє здатності від відчуття: «Не стверджується, що це крок назад у здатностях, це більш здатна модель за Opus 4.7 і 4.8 і навіть конкурує з Fable у бенчмарках, проте з тими іншими моделями приємніше працювати». Чому, три пункти дослівно: вони «зупиняються і ставлять питання, якщо намір був нечіткий», «не роблять припущень без перевірки» і «не переінтерпретовують і не оновлюють плани, не спитавши». Підсумок: через це вони не потребують «обережного няньчення», якого потребує Opus 5.

Пояснення в есеї. Дві сили тиснуть в один бік: гонитва за самополіпшенням і тиск бенчмарків. Аргумент: «хороша бенчмарк-задача самодостатня. Її можна розв'язати. Вона не потребує підказок, читання думок автора задачі чи зовнішньої інформації». Висновок: «Відбір моделей, які добре складають бенчмарки... невід'ємно відбирає моделі, що роблять сміливі, зазвичай правильні припущення перед лицем неоднозначності. Це карає моделі, схильні зупинитись і попросити уточнення».

Фінал: «Реальне життя не бенчмарк. Немає гарантованої правильної відповіді на кожне питання... і з реальними наслідками на кону найкраща здогадка агента не те, що потрібно!»

Скепсис і дані з треду. Загальний тон, згода: «Якість коду помітно впала з 4.5. Час на завершення теж погіршився». Найточніший опис відчуття: «дратує, наскільки він

туманний… Він "пояснює" так, ніби читач вже все знає, а в такому разі навіщо пояснення?»

Окрема гілка про мову. Кілька людей незалежно кажуть, що не англійською виходить гірше: «Спілкування з ШІ будь-якою мовою, крім англійської, уникається, бо результати майже завжди гірші». Другий: «Використання рідної мови… нещодавно Sonnet вставив слово частково російською (кирилицею) замість латинської мови користувача». Третій, точніше за інших: «у більшості професійних контекстів це так, крім юридичних і фіскальних запитів», там рідна мова виграє.

Чому це важливо. Звинувачення справедливе, і воно збігається з тим, як ламаються агентні прогони на практиці. Класика останнього тижня це рівно «сміливе припущення замість питання»: вигаданий URL двічі за один прогін 14.08 замість чесного «точного лінка нема»; застиглий DOM 07.08, коли процес оголошено завислим без перезавантаження сторінки. Це той самий патерн, який описує автор: за неоднозначності робиться ставка.

Дія звідси одна і поведінкова: там, де ціна помилки висока, а питання коштує десять секунд, ставиться питання. Список випадків складається з граблів: точний лінк, якого нема у видачі; незворотна дія на межі дозволеного; цифра, якої нема у виводі скрипта. Правило «лінк або з джерела, або нема лінка» це та сама ідея, записана для одного випадку.

Щодо мови варто бути обережним. Спокуса зробити висновок «агент має думати англійською, щоб бути точнішим» велика, але даних під нею нема: це кілька коментарів на HN, N=1 кожен, жодного заміру. Це спостереження, за яким варто стежити, зі статусом гіпотези. [proven - есей прочитаний повністю, цитати дослівні; коментарі з HN item-API; це суб'єктивне враження одного інженера і його колег, автор сам називає свій розділ «Baseless speculation» (безпідставні спекуляції); жодних замірів, бенчмарків чи A/B у тексті нема; твердження про гіршу роботу не англійською це анекдоти з коментарів, не дослідження]

Why does Opus 5 feel worse to work with? · HN-тред, 814 балів і 745 коментарів

ТЕМА 3

Anthropic випустила інструкцію, як не палити токени в Claude Code

Найпрактичніший пункт випуску: офіційний гайд від Anthropic (авторка Lydia Hallie, 14 серпня) про те, що визначає ціну сесії. 159 балів на HN.

Їхній TL;DR, дослівно: `/clear` між задачами · ставити модель і рівень зусиль до початку («зміна будь-чого з них посеред розмови може розвалити кеш промпту») · @-згадувати файли замість називати їх («файл прикріплюється прямо до повідомлення, що економить виклик Read») · глушити шумні команди прапорцями або виносити в субагента · раз на свіжу сесію ганяти `/context` · `/compact` перед перервою, бо «кеш

промпту спливає через годину, а підсумовувати розмову значно дешевше, поки вона ще в кеші».

Механіка. Вивід коштує «приблизно 5× дорожче за ввід»; читання з кешу «0.1× від ціни вводу», запис у кеш «до 2×». Головне речення про накопичення: «ніщо не надсилається лише раз. Усе, що потрапляє в розмову, файл, який Claude прочитав, або вивід команди, надсилається знову на кожному наступному ході до кінця сесії».

Великий вивід проблемою не є, бо «після 30 000 символів Claude Code записує вивід у файл». Проблема все, що під лімітом: «тест-раннер, який друкує 400 успішних тестів по рядку, влазить у ліміт, і ці 400 рядків тепер частина кожного наступного ходу».

Скепсис із треду. Топовий коментар: «частина цинніка хоче спитати: а чому б не зробити кращий харнес за замовчуванням?». Другий скаржиться на непояснювані переписування кешу: «на 400K токенів… наприкінці часто виходить 2 млн записів у кеш без жодного пояснення».

Чому це важливо. Три поради тут переносяться на будь-який конвеєр, де агентні сесії крутяться за розкладом.

CLAUDE.md проти скілів: «Тримай CLAUDE.md для конкретних інструкцій і перенось воркфлоу-специфічні у скіли, які завантажуються лише коли використовуються». Тобто головний файл це мапа, а решта підтягується за потребою.

Довгі сесії дорожчі за короткі: «одна довга сесія коштує більше, ніж та сама робота, розкидана по кількох коротких… бо хід №40 ще й перечитує 39 ходів перед ним». Конвеєр, який закриває сесію щоночі і щоранку стартує з індексу пам'яті, отримує від цього не саму лише гігієну. Він ще й економить.

Знахідка про заплановані задачі, дослівно: «`/loop` спрацьовує як повний хід у тій сесії, де його налаштовано, тягнучи всю ту розмову з собою щоразу, і якщо минуло більше години з останнього ходу, це ще й промах кешу зверху. Запусти свіжу сесію в іншому терміналі і крути цикл звідти».

Це майже точний опис будь-якого планувальника, який шле завдання у живу сесію дня з усім її контекстом. Цей самий випуск дайджесту зібраний саме так і тягне за собою всю ранкову розмову.

Але певності тут нема: порада написана для `/loop` у людській сесії, а планувальник може спавнити задачі інакше. Правильний хід перевірити з вимірюванням. Перероблювати наосліп це рівно те, від чого застерігає пункт 2. [proven - гайд прочитаний повністю у браузері, всі цитати і числа дослівні; коментарі з HN item-API; це матеріал Anthropic про власний продукт, тобто заразом і документація, і маркетинг; на підписці ці ціни не видно, самі пишуть, що «ті самі запити просто витрачають ліміти», тож перевірити економію в грошах неможливо]

Anthropic: Maximizing the value of your Claude Code sessions · HN-тред, 159 балів

ТЕМА 4

GLM-5.3: найкраща відкрита модель для коду, і лаба сама притримала ваги на два тижні через кібер-здібності, яких не очікувала

Топ доби на HN: 1054 бали, 523 коментарі. Z.ai випустила GLM-5.3, і тут дві історії в одній.

Перша, інженерна. З їхнього блогу: «Все, що зроблено для GLM-5.3, це масштабоване пост-тренування. Використовується та сама базова модель, що й у GLM-5.2, кожен приріст походить з пост-тренування». Результат: «найздібніша модель з відкритими вагами для коду, з покращенням на 50%» відносно GLM-5.2 на їхньому внутрішньому бенчмарку. Terminal-Bench 3.0 стрибок з 4.6 до 28.3, DeepSWE v1.1 з 46.2 до 66.9.

Найцікавіша для практики цифра, токени: «На зусиллі High GLM-5.3 досягає 31.4% приблизно на 50 тис. вихідних токенів, перевершуючи Claude Opus 4.8 з 29.5% на 120 тис.». Тобто краще і вдвічі дешевше по токенах. Тут-таки чесно додається: «GLM-5.3 лишається позаду Claude Fable 5, яка досягає 39.5% на зусиллі Max».

Друга історія, головна тут. Розділ називається «Emergent Cyber Capability»: «кібер-здібності розвивались швидше, ніж очікувалось… Модель не просто стала краще ідентифікувати ізольовані вади: вона почала міркувати через кілька стадій експлуатації, формуючи цілісні плани повних ланцюгів експлуатації». Заявлений SOTA на CyberGym (84.5%), на бенчмарках експлуатації більш ніж удвічі проти 5.2.

Наслідок: «Ваги буде випущено через два тижні після запуску, щойно буде завершено оцінку безпеки і загартування». Китайська лаба, яку зазвичай малюють як «викидає ваги і не думає», сама відклала реліз через здібності, яких не планувала.

Тред про доступ. Топовий коментар про кібер-моделі: «OpenAI і Anthropic мають просто дати людям доступ до кібер-моделей. Інакше буде світ, де атакувальники користуються відкритими і закритими моделями проти значно меншої групи мейнтейнерів». Практичніша скарга поруч: «Доводиться перемикатись на Kimi або GLM навіть у випадках базового тріажу задач у власних проєктах! Нинішні запобіжники смішні». Плюс загальна втома: «Потоп релізів сьогодні, реально важко розібратись».

Чому це важливо. Стрибок отримано без нової базової моделі, самим пост-тренуванням на середовищах, наближених до реальної роботи. Та сама логіка працює на рівні користувача: у базову модель вкладається середовище, скіли, пам’ять, тули. Гайд з пункту 3 і цей реліз кажуть одне й те саме з різних кінців: виграш зараз в обв’язці.

Друге, тверезе. «Найкраща відкрита для коду» все одно позаду Fable 5 за їхнім же власним заміром. Ставити щось локальне на заміну поки передчасно. [proven - блог Z.ai прочитаний, усі числа з їхньої таблиці і тексту дослівно; коментарі з HN item-API;

усі бенчмарки власні заміри Z.ai, зокрема головний (Z.ai Code Bench) приватний і невідтворюваний ззовні; ваг ще нема, обіцяні через два тижні; модель не тестувалась]

Z.ai: GLM-5.3 - Frontier Coding with Emergent Cyber Capabilities · HN-тред, 1054 бали - топ доби

ТЕМА 5

Qwen 3.8 27B - 982 бали за модель, яка влазить у ноутбук. І звіт HuggingFace, який пояснює, чому ця цифра важливіша за фронтир

Два матеріали одного дня, які варто читати разом.

① **Qwen3.8-27B** - 982 бали, 637 коментарів, друге місце доби. Реакція треду однорідна і вона не про бенчмарки: «Це один з найважливіших релізів моделей, оскільки більшість застосувань не потребують SOTA/фронтиру». І: «Це масивні покращення, і щось, що реально можна запустити на ноутбуку». Уточнення в гілці: попередня 3.7 27B не мала відкритої версії, тобто це повернення саме в розмір, який тримається на споживчому залізі.

② **Звіт HuggingFace «State of Open Models: Summer 2026».** Хаб виріс «з 2.43 до 2.96 млн репозиторіїв моделей», але розподіл екстремальний: «85.6% моделей мають менш ніж 200 завантажень за весь час, і 1.5% репозиторіїв дають 99.2% усіх завантажень».

Головна теза звіту, «Увага ≠ Впровадження». Взято топ-25 за завантаженнями і топ-25 за лайками: «Рівно один репозиторій присутній в обох списках». Далі: «Жодна модель, опублікована у 2026, не потрапляє в топ-25 за завантаженнями, тоді як тринадцять з двадцяти п'яти датуються 2022 роком». Приклад: `all-MiniLM-L6-v2` завантажили 1.55 млрд разів при 5156 лайках.

Формулювання звіту, яке варто виписати: «Лайк каже, що реліз важливий… Завантаження каже, що щось вбудоване в конвеєр, який працює за розкладом… Ставитись до одного як до проксі іншого це найпоширеніша помилка, яку бачить команда звіту у висвітленні Хабу, включно з їхньою власною ранішою роботою».

Ще одна цифра, що ламає шаблон: два найбільші публікатори відкритих моделей цього року це AMD і NVIDIA, кожен понад 200 нових репозиторіїв. «Відкритий код перемістився від модельних лаб до апаратних та інфраструктурних компаній».

Чому це важливо. Це прямо поправляє вчорашню оцінку, тому пункт стоїть тут, серед головних.

Вчора першим пунктом ішов DeepSeek Harness з 68 тисячами зірок за добу, із застереженням «68к зірок показник уваги, не якості». Сьогоднішній звіт HuggingFace вимірює рівно цей розрив і показує, наскільки він великий: перетин топів за увагою і за реальним використанням це один репозиторій із двадцяти п'яти.

Практичний висновок для читання стрічки: зірки і лайки це новина, сигналом до дії вони не є. Коли наступного разу приходить «X зібрав N тисяч зірок за добу», правильне питання «чи це вже в чиємусь конвеєрі, що працює за розкладом», і відповіді на нього за добу не буває.

Прикладна дрібниця: 27B на споживчому залізі це клас, який колись реально може крутитись локально. Зараз підписка краща, просто варто позначити, що ця лінія жива. [proven - сторінка моделі, звіт HuggingFace і коментарі HN прочитані; всі цифри дослівні зі звіту; звіт HuggingFace про власний Хаб, тобто джерело даних і зацікавлена сторона одночасно, хоч і чесно дисклеймлить власні минулі помилки; модель не встановлювалась і не тестувалась]

Qwen3.8-27B-FP8 на HuggingFace · HN-тред, 982 бали · Звіт: State of Open Models, Summer 2026 · @huggingface про звіт

ТЕМА 6

Cursor остаточно куплений SpaceX. Найтверезіше тут те, чого в офіційному пості нема

1 млн переглядів на анонсі, найгучніша бізнес-новина доби.

Що сказано офіційно (пост самого Cursor, прочитаний повністю): «Cursor офіційно придбано SpaceX. Це завершує процес придбання, що почався у квітні, коли було оголошено партнерство зі SpaceXAI». Обґрунтування: «Буде доступ до найбільшого флоту GPU у світі… Це означає, що клієнтам можна дати більш здатні моделі за нижчою ціною». Прямо про Grok: «Grok 4.6, яку випустили в середу, дає ранній погляд на те, що тепер можна будувати разом».

Суми в офіційному пості немає. Цифра $60 млрд, яка гуляє стрічкою, з твіту інвестора (@pitdesi): «Мартін / A16Z вів раунд A ~2 роки тому при оцінці $400 млн… тепер Cursor придбано за $60 млрд. ~120× повернення за два роки… ~1000% IRR». Гаррі Тан репостить це зі словами «це робота рівня GOAT». Цифра подається як твердження венчурного інвестора, і статусу підтвердженого факту не має, бо вона вигідна тому, хто її називає.

Реакція розробників на HN холодна. Історія набрала лише 98 балів (проти 1054 у GLM), топові коментарі це «Фууу. Grok. Пішов дивитись на Zed». Найзмістовніший: «не варто чекати значного стрибка продуктивності нових моделей Grok від додаткових даних Cursor, цей стрибок уже стався з Grok 4.5 і тепер 4.6».

Контрапункт зі стрічки. @stanine (Rippling) учора виклав власний замір: «Сьогодні прогнано власні 2100 оцінених прогонів з Grok 4.6. Проти 4.5 пропускна здатність регресувала з 87.3% до 85.9%, і майже подвоїлась медіанна затримка (71c → 131c). Grok 4.6 також частіше падав з помилками і відмовлявся відповідати на безневинні питання». Рівно тоді, коли Cursor рекламує Grok 4.6 як вітрину угоди, незалежний бенчмарк на реальних задачах показує крок назад.

Чому це важливо. Два висновки, обидва про читання новин.

Перший, про гігієну цифр: коли сума угоди приходить від того, хто на ній заробив, вона подається з іменем автора і статусом заяви.

Другий, про консолідацію: вчора Літт і Мейджорс сперечались, що лишається людині, сьогодні незалежний редактор коду став частиною ракетної компанії. Звідси старе правило робочого стеку: краще не прив'язуватись до вендора, який може за квартал стати частиною чогось іншого. Bash, python, sqlite і markdown переживуть будь-яку угоду, а підписку можна замінити. [proven - офіційний пост Cursor прочитаний повністю, цитати дослівні; сума $60 млрд з твіту інвестора, а не з офіційного анонсу, Cursor суми не називає; заміри Rippling це приватний бенчмарк однієї компанії, методологія не перевірена і не може бути перевірена]

Cursor: Cursor is now a part of SpaceX · @mntruell (CEO Cursor) · HN-тред, 98 балів · @pitdesi: звідки взялись $60 млрд і 120× · @stanine: 2100 прогонів, Grok 4.6 гірший за 4.5 · @levie про розмір ринку

ТЕМА 7

Хедж-фонд автора «Situational Awareness» втратив $15 млрд за місяць. І есей про те, чому експертиза в ШІ не переноситься на інші сфери

Дві історії одного дня, які варто читати разом: друга рятує першу від жанру «зловтіха».

Факт: FT повідомляє (113 балів на HN, пейволл, доступний через архів у коментарях), що Jane Street зазнала удару в $15 млрд після краху в Situational Awareness, хедж-фонді Леопольда Ашенбреннера, автора есею 2024 року про неминучість AGI.

Есей-розбір (171 бал) написаний людиною, яка представляється як «колишній хедж-фондівець з ШІ-бекграундом». Теза ширша за одну історію: «Бути експертом в одній галузі не робить тебе експертом в усіх галузях. Хедж-фонд Леопольда Ашенбреннера, Situational Awareness, надав цього тижня демонстрацію на $20 мільярдів».

Механіка провалу, як він її описує: «за всіма даними, він зайшов з плечем в ШІ-бум (за повідомленнями, близько 4×)», з липневими втратами по неохмарах, пам'яті і датацентрах, плюс шорти по софтверних іменах, які відскочили проти нього. Вирок: «Один з перших уроків, які засвоює будь-який справжній інвестор: ринок може лишатись ірраціональним довше, ніж можна лишатись платоспроможним, якщо нема правильного контролю ризиків».

Найгостріше, цитата з самого есею Ашенбреннера 2024 року, наведена дослівно: «Бачу це. Бачу, як буде побудований AGI... можна назвати кластер, на якому AGI буде натреновано, і коли його побудують, приблизну комбінацію алгоритмів... список людей, які матимуть значення. Бачу це.» Коментар автора: «Ах, all-in з плечем у лонг Nvidia. Смак його інвестиційного стилю відчувався ще тоді».

Історична рима: Long-Term Capital Management 1998 року, «два нобелівські лауреати і найкращі трейдери облігацій Волл-стріт в одному фонді… потім він вибухнув так видовищно, що ФРС довелось збирати банки на порятунок». Звідти й назва.

Тред на HN, переважно недовіра до самої конструкції: «Стоп, це реально? Випадковий 25-річний отримав $45 млрд в управління, бо дав кілька інтерв'ю після звільнення з OpenAI? Це майже здається перформансом».

Чому це важливо. Пункт-калібрування про те, як читається ця стрічка щодня.

Гучність прогнозу і його точність це різні осі. Есей Ашенбреннера два роки був обов'язковим до прочитання і сформував тон половини індустрії; це не завадило його автору злити фонд на важелі. Коли завтра хтось скаже «через рік користування комп'ютером стане опційним», це той самий жанр висловлювання, і ставити на нього гроші не можна, навіть якщо автор глибоко в темі.

Друге, вужче, для будь-кого з інвестиційним рахунком: сюжет «людина з правильним поглядом на технологію зайшла з 4× плечем» це рівно та помилка, від якої захищає нудний портфель без плеча. Варто зафіксувати збіг: у тій самій добі, де лаби показують найкращі бенчмарки в історії, найгучніший AI-візіонер утратив фонд. Обидва факти правдиві одночасно. [proven - есей прочитаний повністю, цитати дослівні, зокрема цитата з першоджерела Ашенбреннера; сама новина FT за пейволом, читана через архівне посилання з треду, тобто цифра $15 млрд береться з переказу і заголовка, а не з повного тексту FT; деталі про 4× плече автор сам подає як «за повідомленнями» (reportedly), тобто це реконструкція; авторка/автор есею зацікавлена сторона з галузі, і в тексті є нотка особистого рахунку («багато хто вважав його нестерпним»)]

When Genius Fails: The Intellectual Arrogance of the AI Labs · HN-тред есею, 171 бал · FT: Jane Street і $15 млрд · HN-тред новини, 113 балів

ТЕМА 8

Google запустила приватний ШІ на гомоморфному шифруванні. Тред одразу знайшов дірку, і вона фундаментальна

316 балів. Google оголосила, що робить «приватний ШІ практичним» за допомогою гомоморфного шифрування (FHE), технології, яка дозволяє рахувати над зашифрованими даними, не розшифровуючи їх. Обіцянка приваблива для одного класу задач: обробка чутливих даних на чужому сервері так, щоб сервер не бачив вмісту.

Тред знайшов межу заяви, і вона не про криптографію. Найточніший коментар: «Одна вада FHE в тому, що воно гарантує лише те, що потрібен ключ, щоб побачити входи чи виходи обчислення, але не обов'язково те, що це саме те обчислення, якого хочеться. Наприклад, обчислення може бути ворожим для певних входів, або противник може вставити своє обчислення першим (чи останнім)».

Другий, простіший: «Чи це покладається на модель "повір мені, бро", чи є спосіб для клієнта перевірити, що провайдер справді не може бачити входи? Хочеться почитати білу книгу, але знаходиться лише презентація». І найкоротший: «Шифрування чи ні, якщо воно на чужому сервері, воно не твоє».

Чому це важливо. Це інфраструктурна технологія, не продукт, і дій із неї не випливає. Але формулювання з треду варте запозичення як фільтр.

Для чутливих категорій, здоров'я, ліків, фінансів, є простіша відповідь, ніж FHE: тримати дані у файлах, які читаються очима, і обробляти локально скриптами, які можна прочитати. Це не криптографічна гарантія, але закриває ту саму проблему нудним способом.

Один рядок звідти варто тримати як фільтр на майбутнє: коли з'явиться «зручний хмарний сервіс для трекінгу здоров'я», питання «чи можна перевірити, що саме там рахується». Поки відповідь «повір мені, бро», відповідь ні. [proven - сторінка Google прочитана, коментарі з HN item-API дослівні; саму технічну реалізацію не перевірено, білої книги коментатори не знайшли, окремого пошуку теж не було; це анонс Google про власну технологію]

Google: making private AI practical with homomorphic encryption · HN-тред, 316 балів

ТЕМА 9

Firefox лишився останнім браузером з підтримкою uBlock Origin. 601 бал, і найкорисніше в треді обхідний шлях

601 бал, 222 коментарі. Не AI-новина, але про інструмент повсякденного користування, тож коротко.

Суть: після переходу Chrome на Manifest V3 і аналогічного кроку Edge Firefox лишився єдиним великим браузером, де повноцінний uBlock Origin ще працює.

Найцікавіше в треді. Топовий коментар, рішення: «Будь-хто може зібрати власне розширення, яке робить усе, що розширенню дозволено. Було ~7 стандартних розширень, які завжди ставились у Chrome. Останній місяць Claude Code використовувався, щоб зібрати одне власне розширення, яке робить усе те саме».

Чому це важливо. Міняти браузер через новину сенсу нема, дій звідси нуль.

Але цей коментар найкоротша ілюстрація того, про що вчора писали Літт і Мейджорс, і сьогодні пункт 3: вартість «зробити своє замість того, що забрали» впала до рівня буденної реакції. Раніше таке коштувало б тижні. [proven - заголовок і коментарі з HN item-API дослівні; сама стаття PCWorld прочитана не повністю (взято заголовок і тред), і чуже розширення на Claude Code не перевірене, це заява коментатора]

Firefox is now the last major browser that still supports uBlock Origin · HN-тред, 601 бал

ТЕМА 10

Ґерґелі Орос про Grok Bot: «момент OpenClaw для керованих ШІ-агентів»

Останній пункт короткий, бо обмежений джерелом.

The Pulse від 14 серпня, єдиний свіжий випуск Substack у вікні. Орос дає це другим сюжетом, і в безкоштовній частині формулювання таке: «Grok Bot: "момент OpenClaw" для керованих ШІ-агентів? Команда Cursor побудувала і випустила загальний ШІ-харнес, який відчувається як "досвід Codex, але для розумової праці". Я спробував його, автоматизував багато своїх щоденних робочих процесів, і я в захваті. Більше ШІ-вендорів неодмінно скопіюють цей харнес».

Далі пейволл, і решта не прочитана. Тому наводиться рівно те, що на сторінці: у випуску є ще сюжет про Meta, яка не може зупинити хвилю звільнень, яку сама спровокувала («пропонує великі утримувальні пакети акцій, і схоже, що це не працює»), але деталей не бачено.

Місток до вчора. Вчора в misc наведено @awilkinson з фразою «Grok Bot = Openclaw для нормальних людей», і тоді це був жарт одного інвестора. Сьогодні той самий діагноз незалежно ставить Орос, найтехнічніше джерело серед підписок, уже як практик, який переніс на нього робочі процеси. Теза за добу зміцніла: з дотепу вона стала спостереженням двох незалежних людей.

Чому це важливо. Grok Bot робить те саме, що роблять саморобні керовані агенти для розумової праці, тільки продуктом від вендора. Аргумент на користь саморобного той самий, що і в пункті 8: усе лежить у git і читається очима. Аргумент проти: за вендорським харнесом стоїть команда, а за саморобним одна людина і її вечори.

Дій звідси не випливає. Міняти живу робочу конструкцію на чужий продукт через дві похвали в стрічці це рівно та помилка, від якої вчора застерігала стаття про нудні технології. Але одне варто поставити на радар: якщо саморобне колись стане вузьким місцем, ринок таких харнесів уже існує. [fuzzy - прочитана лише безкоштовна частина випуску, більша частина за пейволом; сам Grok Bot не бачений, не поставлений і не випробуваний; захват Ороса його особистий досвід, жодних замірів; місток до @awilkinson це збіг двох думок, підтвердженням він не є]

The Pulse: Meta's self-inflicted resignation-wave (Grok Bot - другий сюжет) · @awilkinson вчора: «Openclaw для нормальних людей»

misc - коротко, що ще варте ока

• **SWE Odyssey - бенчмарк на «надтривалий горизонт»** - @mattstallone: «Оскільки бенчмарк за бенчмарком насичуються, виникло бажання знайти кількісний спосіб перевірити більшу заяву: чи можуть агенти працювати автономно годинами і все одно збудувати правильну річ?» Це та сама діра, яку описує пункт 2: бенчмарки міряють задачі, які можна розв’язати, і не міряють роботу, де треба спитати

• **Anthropic виклала другий Risk Report** (535 тис. переглядів) - @AnthropicAI: «У межах Політики відповідального масштабування публікуються регулярні звіти про ризики… Другий звіт про ризики тепер доступний». Сам звіт редагований (redacted) PDF, він не розібраний, тому зміст не переказується: посилання подається як є

• **@garrytan: один плюс 20 агентів проти цілого відділу**: «Одна людина плюс 20 агентів, що працюють одночасно, може перевершити цілий відділ інженерів у великій техкомпанії з "чудової сімки"». Класична гіпербола інвестора, але варто звернути увагу, як вона римується з пунктом 2: більше агентів ≠ менше потреби питати

• **@garrytan про Fable 5 у роботі**: «найдивовижніше у використанні GStack до Fable 5 і після: тепер на багато питань з односторонніми дверима, які повертає Claude Code, можна просто сказати "бери всі рекомендації" і бути задоволеним». Це протилежна оцінка до пункту 2, тому наведені обидві

• **@snowmaker: розмір кодової бази YC за 14 років** (431 тис. переглядів, 809 лайків) - графік від @snowmaker (Jared Friedman, YC). Без коментарів у самому пості, але 100+ коментарів у треді: класика жанру «покажи криву, хай сперечаються»

• **Toast 1 - спеціалізована модель для пошуку** (189 балів). Коментар, що пояснює інтерес: «Глибока симпатія до цієї ідеї спеціалізованих LLM для пошуку. І вкрай незрозуміло, наскільки грубим є вхід Google сюди»

• **RISC-V: They should have known better** (150 балів) - технічний розбір архітектурних рішень RISC-V. До ШІ стосунку не має, але це найякісніший інженерний лонгрід доби

• **Seven books I keep close because I love them** (319 балів) - Марк Домінус про сім книжок, які тримає під рукою. Рідкісний випадок, коли топ HN це просто хороший текст про читання

• **@lennysan: «тепер стоматологічний інфлюенсер»** - продовження вчорашнього спостереження про ШІ в стоматології, яке несподівано розлетілось. Приклад того, що найкраще заходить у стрічці побутове спостереження

• **@awilkinson: у Things досі нема API**: «Це 2026 рік, як у Things досі нема API?… неможливість взаємодіяти з хмарним ШІ вбиває». Дрібниця, але точний маркер нової вимоги до софту: нема API, нема місця в робочому процесі з агентом
