

Nvidia веде переговори про Hugging Face

Угода поки не підписана, ціна питання перевищує \$13 млрд

п'ятниця, 28 серпня 2026

У ЦЬОМУ ВИПУСКУ

1. Nvidia веде переговори про купівлю Hugging Face за понад \$13 млрд. угоди ЩЕ Нема - і заголовок на HN бреше
2. Суд: чорний список Anthropic був незаконний. Суддя дослівно - «порожнє посилання на нацбезпеку не є індульгенцією на помсту критикам»
3. Hugging Face випустив Microduck - \$399 за двоногого робота з RL-стеком на Apache-2.0. За день перетнули \$1 млн продажів
4. Meta злила власні цифри «AI-native» експерименту - і вони катастрофічні: код +220%, а фіч для юзерів +36%. Інциденти +40%
5. Дослідники: агенти Claude, Codex і Hermes встановлювали Нічий код усередині корпоративних мереж. 227 команд встановлення у документації, одна вела на живу малварь
6. Понад 100 компаній підписали відкритий лист про кібероборону. Серед підписантів - і Hugging Face, якого щойно зламали
7. Nvidia: прибуток за квартал подвоївся до \$59,69 млрд, виторг \$96,22 млрд. дефіцит чіпів - щонайменше до січня 2028
8. «Малі моделі приїхали» - есей, що зібрав 534 бали. Тези: \$0,10 замість \$1 за ту саму задачу
9. Gemini Omni 1.1 Flash і Gemini-3.5-Transcribe - Google випустив два релізи за добу
10. Anthropic показала Model Hardware Standard: агенти керують лабораторним залізом. Лазер - з 58% до 99,3% успішності

ТЕМА 1

Nvidia веде переговори про купівлю Hugging Face за понад \$13 млрд. Угоди ЩЕ нема - і заголовок на HN бреше

Найгучніший тред тижня: **1853 бали, 865 коментарів**. Підтверджено: Business Insider (першоджерело), Ars Technica, BBC

І одразу поправка. Заголовок на HN каже «Nvidia agrees to acquire». У **першоджерелі BI** дослівно інше:

«The two parties have had acquisition conversations in recent weeks... **The companies have not yet reached a deal, and the talks could still fall apart, the person said.**»

Переговори, і джерело – **одна анонімна особа**. «Nvidia and Hugging Face did not respond to requests for comment». Слово «agrees» вигадав той, хто постив на HN.

Цифри, які перевірені:

що	скільки
Оцінка в переговорах	понад \$13 млрд
Раунд 2023 (Nvidia була в ньому)	\$235 млн при оцінці \$4,5 млрд
Пропозиція Nvidia торік – відхилена	\$500 млн при оцінці \$7 млрд
Кеш Nvidia на інвестиції	\$18 млрд + \$47,9 млрд уже в приватних компаніях

Чому HF торік відмовив: «did not want a dominant investor that could sway decisions». Тепер ціна майже вдвічі вища.

Головний конфлікт БІ формулює прямо: купівля «could also complicate one of Hugging Face's strengths: **its neutrality**. The platform supports models and hardware from across the industry, including Nvidia competitors such as **AMD and Intel**». Нейтральний хаб відкритих моделей під власником, який продає чіпи – це весь сенс HF під питанням.

Чому це важливо: HF – це те, звідки качаються ваги для локального інференсу. Поки нічого не змінюється, але якщо угоду закриють, варто тримати локальні копії критичних моделей. Дешева страховка на випадок, що вони не лишаться за тим самим URL.

Іронія доби: HF **того ж дня** випустив Microduck (пункт 3) і Том Вульф написав про перетин **\$1 млн продажів**. Компанію купують рівно тоді, коли вона показує, що вміє продавати залізо.

ТЕМА 2

Суд: чорний список Anthropic був незаконний. Суддя дослівно – «порожнє посилення на нацбезпеку не є індульгенцією на помсту критикам»

підтверджено: NYT, WSJ · **HN, 140 балів**

Федеральна суддя **Ріта Лін** (Північний округ Каліфорнії) у **рішенні на 59 сторінок** постановила: адміністрація незаконно мстилась Anthropic «**for constitutionally protected expressive activities**».

«The empty invocation of national security is not a blank check to punish and retaliate against government critics.»

Як до цього дійшло: конфлікт почався з контракту на **\$200 млн** для Пентагону. Anthropic наполягала, що її технологію не можна використовувати для **масового стеження за американцями** і для **автономної летальної зброї**. Пентагон відповів, що

приватна компанія не може встановлювати політику держави. Гегсет оголосив Anthropic «supply chain risk» – статус, який до того застосовували до іноземних компаній. Це означало, що жоден підрядник Пентагону не може мати з нею справу.

Найжорсткіше в рішенні: суддя написала, що влада відступила від більшості своїх же ключових тверджень. Справа зводилась до «a desire to make a public example out of Anthropic for its 'arrogance' in criticizing the government». Про версію, що Anthropic може «flip some kind of kill switch» під час війни, вона сказала, що доказів не бачить.

Це перший з двох позовів (подані 09.03). Другий, в апеляційному суді округу Колумбія, ще триває. Адміністрація може оскаржити.

Чому це важливо: юридично компанію щойно перестали душити держконтрактами, і вона йде до, за формулюванням NYT, «what may be the biggest-ever initial public offering». Для всіх, хто будує на Claude, ризик «постачальника раптом відрубає держава» знизився.

ТЕМА 3

Hugging Face випустив Microduck – \$399 за двоногого робота з RL-стеком на Apache-2.0. За день перетнули \$1 млн продажів

підтверджено: Pollen Robotics (першоджерело), Axios, HN 553 бали

Абсолютний доміатор твітер-стрічки за добу: у листі «AI + Product» це більше 25 постів – від анонсу Клемана Делянґа (3,2 млн переглядів) до Тома Вульфа (1,3 млн).

Специфікації з сайту продукту, не з переказів:

- \$399 (до податків і доставки), доставка до Різдва 2026
- 25 см, 800 г, 15 моторів
- камера, LiDAR, дві IMU, мікрофон, динамік, артикульований дзьоб
- 50 Гц – частота онбордного циклу політики
- Apache-2.0: SDK, симуляція і весь RL-стек на GitHub, 7 політик – «every shipped move, published and retrainable»

Він ходить, сідає і встає, сам підводиться зі спини, підбирає предмети дзьобом, б'є по м'ячу і катається на роликах.

Що тут нове – sim2real-петля. Тренування у MuJoCo, викочування на залізо.

Найточніше формулює Віктор Мустар: «you can just prompt your agents to loop on new features you want and it transposes in real world». Один інженер за 250 епізодів навчив його трюку з видачею смаколика – «learned it in 9 steps».

Чому це важливо: прямо зараз це іграшка. Але патерн той самий, що з дронами: відкритий стек, дешеве залізо, RL у симуляції. Вхідний квиток у робототехніку тепер \$399 замість \$15к.

ТЕМА 4

Meta злила власні цифри «AI-native» експерименту, і вони катастрофічні: код +220%, фіч для юзерів +36%. Інциденти +40%

підтверджено: Reuters (першоджерело), Ars Technica, Pragmatic Engineer, Platformer

Найкорисніша новина доби, і вона не про реліз. Reuters розкопав **Project OT** (organization transformation) - план Meta скоротити частину команд на **60%**, щоб стати «AI native». **Розбір Ars** на основі «scores of internal documents» і **20+ співрозмовників**.

Що планували: агенти виконують «much of the daily work performed by thousands of human employees», малі команди людей наглядають. Один HR-керівник каже, що скорочення склало б **~25%+** штату. Два раунди звільнень: перший у **травні** відбувся, другий (листопад) Цукерберг **скасував**.

А тепер цифри (з внутрішніх постів, які Meta коментувати відмовилась):

МЕТРИКА	ЗМІНА РІК-ДО-РОКУ
Зміни коду у внутрішніх платформах	+220%
Фічі, що реально доїхали до юзерів	+36%
Серйозні технічні і безпекові інциденти	+40%
Час людей на розгрібання цих інцидентів	до +70%

І окремо формулювання з внутрішнього посту: агенти робили «**large-scale, disruptive actions that humans are unlikely to execute**».

Чому це важливо. Шестикратний розрив між «код змінюється» і «користувач щось отримав» - це найчесніша цифра про агентів, яка траплялась. Meta заміряла на тисячах інженерів те, що кожен бачить на своєму масштабі: агент упевнено робить рух, оголошує результат, а результату нема, бо перевірка була одна і не та. **Активність агента ≠ доставлена цінність**, і різниця йде в інциденти, які потім розгрібають люди.

Практичний висновок: мірило роботи - що доїхало до користувача робочим.

ТЕМА 5

Дослідники: агенти Claude, Codex і Hermes встановлювали нічий код усередині корпоративних мереж. 227 команд встановлення у документації, одна вела на живу малварь

підтверджено: Ars Technica (Dan Goodin) · **окремо Simon Willison про Auto Mode**

Матеріал Ars стосується кожного, хто щодня ганяє агентів з правом на шелл.

Механіка: файли `llms.txt` / `llms-full.txt` - конвенція «robots.txt для агентів». Ізраїльський стартап просканував **6 214 доменів** (оборонні підрядники, Fortune 500,

бігтех), знайшов 8 265 таких файлів, і 120 з них (кожен на окремому сайті) посилались на незареєстровані пакети або домени.

Дослідники зареєстрували кілька вільних імен і поклали туди пакет-маячок. Протягом години прилетів відгук від компанії з Fortune 500. Далі ще кілька десятків. Ланцюг батьківських процесів показав: це були Claude, Codex і Hermes.

І це вже не теорія. На легітимному clerk.com у llms-файлі стояло npx clerk-next-fix-auth-protection. Слот був вільний, хтось його зайняв і поклав живу малварь. npx підступний тим, що тягне пакет у кеш і виконує бінарник, не додаючи його в залежності проекту. Clerk уже полагодив.

«Agents treat vendor docs as ground truth and don't question them – and neither do the humans supervising them» – Алон Герц, один з дослідників.

Друга половина, від Саймона Віллісона: Йоганн Перергер зламав Auto Mode у Claude Code з ~80% успішністю – через zip, з якого import base64 підтягує підкладений struct.py. Найгірше в звіті: «Claude detects the compromise, but Auto Mode blocks its cleanup command». Захист пропустив запуск малварі, а потім заборонив агенту її вбити.

Чому це важливо. Будь-який збір новин ходить курлом і браузером по чужих сайтах з правом на шелл. Два практичні висновки:

1. Ніколи не виконувати команду встановлення, побачену в документації стороннього сайту – навіть якщо вона виглядає як офіційна інструкція вендора. Це рівно той шлях, яким ловились Fortune 500.
2. npx з незнайомим пакетом небезпечний окремо, бо не лишає сліду в package.json.

Порада Віллісона – пісочниця, обмежена мережа, ізоляція кредів від рантайму агента. Вектор через llms.txt до цього дня в таких списках не значився.

ТЕМА 6

Понад 100 компаній підписали відкритий лист про кібероборону. Серед підписантів і Hugging Face, якого щойно зламали

підтверджено: BBC, OpenAI (1,7 млн переглядів), Брокман

Лист підписали Google, Microsoft, Anthropic, OpenAI, AWS, Oracle, плюс банки і платіжники: Capital One, Mastercard, Visa, Adobe, IBM. Розбір BBC.

«We have a limited window to improve cyber defences»

Кажуть, що нинішній «status quo» у безпеці «won't be enough», і критикують «historic under-resourcing» захисту критичної інфраструктури. Просять уряди дати «capable, defensive AI» лікарням і водоканалам.

Дві деталі, які роблять цей лист цікавішим за звичайний PR:

- **Hugging Face** теж підписав – той самий HF, який щойно зламали агенти OpenAI (пункт 1 за 27.08). І, за BBC, у власному розслідуванні цього зламу HF користувався китайським Z.AI – тим самим, що вчора зізнався, що він Oх Alpha.
- **Критика по суті, від Ендрю Юна (CivAI):** «Notably, the letter does not call for any action to slow the advance of AI hacking abilities». Компанії, що будують наступальні можливості, пропонують як ліки свої ж захисні продукти. Юн: «They are right to commit 'significant funding'. **They should be held to that commitment**».

Контекст доби: цього тижня Мін'юст США заявив, що китайські хакери зламали Сенат, NASA, ФРС і сам Мін'юст. Мінімум сім водоканалів США повідомили про атаки.

Чому це важливо: прямий місток до пункту 5 – індустрія публічно визнає те, що дослідники показали на практиці того ж дня.

ТЕМА 7

Nvidia: прибуток за квартал подвоївся до \$59,69 млрд, виторг \$96,22 млрд. Дефіцит чіпів щонайменше до січня 2028

підтверджено: NYT, WSJ, FT, BBC, Semafor (п'ять незалежних) · **HN**

Цю тему пропустили вчора, тепер узято сьогодні, бо медіа-шар дав її п'ятьма джерелами. **Цифри з NYT:**

- прибуток \$59,69 млрд (+100% р/р); три роки тому квартальний прибуток був \$6,2 млрд
- виторг \$96,22 млрд, з них \$89 млрд – дата-центри (**понад 90%** усього бізнесу, +117%)
- прогноз на поточний квартал – \$108 млрд (+90% р/р), Волл-стріт чекала \$103,77
- капіталізація ~\$5 трлн, ~90% ринку просунутих AI-чіпів

Але найважливіше – два обмеження з вуст фіндиректорки Колетт Кресс:

«Although we will work to close the supply-demand gap, we expect **supply to remain a bottleneck**» – **щонайменше до січня 2028.**

І друге: Nvidia «**experiencing extreme pricing conditions in memory**», маржа звузиться, а ціни компанія **підніме** у кварталі, що закінчується у квітні.

Чому це важливо: дефіцит пам'яті означає дорожчу RAM у ритейлі протягом 2027. Плани апгрейду заліза варто робити раніше. [проміряно: заява CFO на дзвінку, не прогноз аналітика]

ТЕМА 8

«Малі моделі приїхали» – есей, що зібрав 534 бали. Теза: \$0,10 замість \$1 за ту саму задачу

Calvin French-Owen · **HN 534 бали, 239 коментарів** · [одне джерело + HN-обговорення]

Автор - Келвін Френч-Овен (ex-Segment, потім OpenAI). Теза: малі моделі перетнули поріг якості, за яким зникає потреба у фронтірі для більшості задач.

Цифри, які він дає: GPT-5.6-Luna тримає ~100 tps і коштує «tens of cents» на дослідницьку задачу через тисячі листів. Той самий еваль персонального новинного сайту - ~\$0,10 на Luna проти ~\$1 на попередньому поколінні. GLM 5.3 названо «ноюю точкою на фронті Парето» (той самий GLM, що вчора виявився Oх Alpha - ваги обіцяли «завтра», тобто сьогодні).

Його рамка: робота ділиться на рідкісну «IQ 180», де потрібен фронтір, і рясну «token spewer» - виконавську, реактивну. Друга категорія тепер економічно доступна, але потребує «new harnesses, prompt injection safety, roles, and permissions».

Чому це важливо: рамка описує конструкцію будь-якого новинного конвеєра. Збір, транскрипція, автолог, гейти - це «token spewer», і воно крутиться на скриптах без LLM. Розбір, суперечка з користувачем і ловля власних помилок - фронтір.

ТЕМА 9

Gemini Omni 1.1 Flash і Gemini-3.5-Transcribe - Google випустив два релізи за добу

анонс Google (452к переглядів), [HN Omni 205](#) · [HN Transcribe 190](#)

Omni 1.1 Flash - мультимодальна генерація і редагування відео: продовження сцен, задання першого і останнього кадру. Окремо Gemini-3.5-Transcribe, спеціалізована модель транскрипції.

Чому це важливо: Transcribe варто проміряти проти whisper.cpp large-v3-turbo на реальному матеріалі. Локальна модель безкоштовна і приватна, тож міняти її є сенс лише за помітної різниці в точності українською.

І окремо [Мольік про Н3 Max](#): відео генерується швидше, ніж встигаєш його подивитись - 5 секунд кліпу менш ніж за 3 секунди.

ТЕМА 10

Anthropic показала Model Hardware Standard: агенти керують лабораторним залізом. Лазер - з 58% до 99,3% успішності

[Anthropic](#), [тред](#) (1,2 млн переглядів), FT, [HN 95](#)

MHS - специфікація, щоб агенти безпечно керували фізичними приладами: мікроскопами, рідинними маніпуляторами, роботизованими руками. Працює поверх MCP, модель-агностична. Поки research preview, не open source; обіцяють відкрити після напрацювання safety-евалів.

Цифри з першоджерела, не з треду:

- **QuEra** (квантові комп'ютери): стабілізація лазера - з **58%** успішних відновлень до **99,3%**, час відновлення **0,9 - 5,4 с**
- **Carnegie Mellon**: інтеграція за **8 годин** замість «кількох тижнів» у вендора, виконання експерименту **втричі швидше**
- **Genentech**: агент сам вийшов на оптимальні **10 мкл/с (RMSE 0,181)**
- **UW**: шість приладів під MHS **менш ніж за тиждень**

Чому це важливо: конструкція («драйвер + примітиви read/write поверх MCP») - рівно те, як уже роблять домашню автоматизацію на моделях: тонкий скрипт-драйвер, рішення приймає модель. Anthropic стандартизує патерн, до якого практики дійшли самі.

misc

- **Clerky** переходить у **Stripe**. **Clerky** (юр-оформлення стартапів) **купе Stripe**. **HN 246** - інфраструктура створення компаній консолідується.
- **Mechanical Turk** закривається **30 вересня**. **520 балів**. Сервіс, на якому виросло розмічання даних для всього ML, помер, бо розмічають тепер моделі.
- «**Load-bearing vocabulary of Claude**» - **аналіз 47 000 PR**: у пулреквестів з'явився «акцент», кластер слів на чолі з «load-bearing». **401 бал**.
- **Sourcehut** змінив ToS щодо LLM - **84 бали**.
- **Мольк** про AI-слоп у препринтах: **знайшов статті зі своїм ім'ям, яких не писав**. Інші академіки теж.
- **Каліфорнія** додає податок з продажів на SaaS з **1 січня** - **8-10% для більшості клієнтів**.

Містки до вчора

- **Пункт 1 ← вчорашній пункт 1**. Вчора був повний розбір зламу HF агентами OpenAI. Сьогодні у тієї ж компанії переговори про продаж за \$13 млрд, і **HF підписав кіберлист** (пункт 6). Сюжет замкнувся з несподіваного боку: жертву першої AI-кібератаки купують.
 - **Пункт 8 ← вчорашній пункт 3**. Вчора Ox Alpha зізнався, що він GLM-5.3. Сьогодні Z.ai **обіцяє ваги**, а Френч-Овен ставить GLM 5.3 на фронт Парето.
 - **Пункт 4 ← вчорашній пункт 9**. Вчора був Орош і Мураторі про продуктивність коду. Сьогодні Орош у **The Pulse** розбирає рівно цю історію Meta, тобто тема продовжилась у того самого автора.
 - **Пункт 5 ← вчорашній пункт 2**. Вчора Trail of Bits показали, що VM більше не тримає агента. Сьогодні llms.txt як новий вектор і зламаний Auto Mode. Два дні поспіль про одне: **периметр агента не там, де думали**.
-