AI-ДАЙДЖЕСТ

Математик закрив 78-річну проблему з Claude

370 тис. переглядів за ніч, а Opus 5 забирає лише 3,5% витрат проти 28% в Opus 4.8

понеділок, 24 серпня 2026

У ЦЬОМУ ВИПУСКУ

ТЕМА 1

Математик оголосив розв'язок 78-річної проблеми - комплексна структура на S^6, зроблена з Claude. 370 тис. переглядів за ніч

Пост Левента Альпьоге (@__alpoge__) (24.08, 00:31 за Києвом) - **370,7 тис. переглядів, 1,7 тис. лайків** за неповних сім годин. Дослівно: «Please welcome to the world a beautiful new geometric object, to do with a problem i've always loved. **claude really contains multitudes:D** Does S^6 admit a complex structure? / **Yup**».

Що це за задача. Чи допускає шестивимірна сфера комплексну структуру - відкрите питання **з 1948 року**. Останню відому спробу довести, що НЕ допускає, робив покійний

Майкл Атія, одне з найбільших імен геометрії XX ст.; його препринт згадують у відповідях під постом.

Найважливіше: автор виклав перевірюваний сертифікат. PDF на alpo.ge/s6.pdf - **1,1 МБ, 362 тис. символів тексту**, завантажений і прочитаний. Це щільна диференціальна геометрія: трикутна група D(3,4,∞), універсальна сім'я 2-торів над стековою P^1, три спеціальні шари (логарифмічне перетворення Кодаїри в еліптичних точках, торична дегенерація Мамфорда в каспі). Фінальне твердження дослівно: «**X has the integral homology of S^6 and is diffeomorphic to S^6**».

Мотивація автора у другому пості: «i always like to have short certificates that people can use to verify in full with their expertise **and/or with one of today's super amazing models**».

І одразу протиотрута, бо тема рівно того класу, де вона потрібна. Це **заява, не прорецензований результат**:

- Робота **суперечить опублікованому результату** Campana - Demailly - Peternell. У PDF є окремий розділ «Why the argument of [CDP20] does not apply»: автор свідомо каже, що в опублікованій статті хиба. @proofchecker під постом уточнює: «Interestingly, **the whole point of CDP20 was to fill gaps in CDP98**».
- Найтверезіший коментар @ccccccmore: «announcing a new geometric object with a chatbot as co-author is a bold way to say **the referees haven't seen it yet**».
- @IbrahimDagher20: «if claude did this ~all, **this may be a first for AI?**». Навіть прихильники ставлять знак питання.

Контекст, який робить це серйозним: **це той самий Альпьоге**, чий контрприклад до гіпотези Якобі з Fable зібрав 803 бали на HN у липні. Він математик з PhD під Манджулом Бгаргавою, філдсівським лауреатом.

Чому це важливо. Практичний висновок про **формат перевірки**. Альпьоге робить те, чого нема в 99% «ШІ щось вирішив»: коротка самодостатня довідка, яку можна віддати експерту АБО іншій моделі. Це рівно та дисципліна, яку тримає цей дайджест: **артефакт, який можна перевірити, важить більше за заяву**. Мітка: [promising], не [proven], до рецензування.

ТЕМА 2

Платіжні дані 70 тис. компаній: Opus 5 - 3,5% витрат на Anthropic, Opus 4.8 - 28%. Найкраща модель програє власній попередниці ушестеро

Розбір Саймона Вілісона (23.08, 20:24) статті FT, плюс індекс Ramp - оцінка на **білінгових даних 70 000 компаній**, що платять корпоративними картками Ramp. Не на опитуваннях.

Розподіл витрат на моделі Anthropic, липень 2026:

МОДЕЛЬ	ЧАСТКА
Opus 4.8	**28,0%**
Sonnet 4.6	8,3%
Fable 5	**8,0%**
Opus 4.6	6,9%
Sonnet 5	3,6%
Opus 5	**3,5%**
Haiku 4.5	1,0%

Вілісон одразу дає поправку сам на себе: Opus 5 вийшов **24 липня**, тобто в липневих даних він захопив лише тиждень, і низька частка очікувана. А от **Fable 5 з 8,0% при своїй ціні** - це вже сигнал, і саме він у заголовку FT.

Цифри по грошах з тієї ж статті (FT, з посиланням на «people with knowledge of the matter»): річна виручка Anthropic **$65 млрд** у липні проти **$47 млрд** у травні; **6 000 клієнтів** платять від $100 тис./рік; Q3 очікується прибутковим. У OpenAI - «понад **$40 млрд**, +35% за квартал».

Дедуп і чесність про фільтр джерел. Ця цифра вже фігурувала **18.08**, і тоді вона прийшла постом `@AiBreakfast`, акаунта **поза підписками**, який до того підводив. Тоді її пішли перевіряти в CNBC і Bloomberg. Сьогодні її підтверджує FT. Фільтр «сумнівне джерело веде до перевірки в нормальному» відпрацював правильно, і цифра встояла три перевірки.

Що не перевірено і сказано прямо: саму статтю FT прочитати не вдалось - **ft.com сліпий**: віддає 403 і на справжню адресу, і на завідомо вигадану (`00000000-0000-...`), тобто курл там не міряє нічого. Цифри взяті **з переказу Вілісона**, з підписом цього джерела. Ramp-індекс, навпаки, перевіряється чесно: 200 на справжній сторінці, **404 на вигадану**.

Чому це важливо. Ринок у липні масово сидів на **Opus 4.8**. «Всі вже на найсвіжішій моделі» - це ілюзія твітера, гроші показують інше.

ТЕМА 3

Бройніг формулює це в тезу: «Fable і кінець безкоштовного обіду». Аналогія із законом Мура - і вона робоча

Пост Дрю Бройніга (23.08), цитований Вілісоном окремо - і це найкорисніший текст доби.

Теза дослівно: «Prior to Fable, it felt silly to waste too much time improving your coding harness or context strategies. **A new model would arrive at the same price (or cheaper!)**

and paper over most of your problems. But then Fable landed… So we started to think about **what work went where**».

Аналогія: поки діяв закон Мура, вилизувати код не мало сенсу, бо за 18 місяців процесор сам подвоїть швидкість (Герб Саттер називав це «безкоштовним обідом»). Коли однопотокова продуктивність стала, довелось думати про паралелізм і архітектуру. **Зараз те саме сталося з моделями.**

Конкретна цифра: **GLM 5.2 вийшов того ж тижня, що Fable, і коштує ~1/9 від нього** (і ~1/5 від Opus 5). «Is GLM 1/9th the quality of Fable? Perhaps, for certain classes of tasks. But **for most rote coding it's more than sufficient. Especially when provided with great context**».

Робочий патерн автора, який можна брати як є: «I frequently chat with Fable to **interrogate and shape a design**, before handing off a brief to GLM».

І контраргумент, поставлений собі самим автором: ціни на інференс падатимуть, то чи не повернеться усе до «все через найбільшу модель»? Відповідь: ні, бо **ті самі здешевлення дістаються і K3, і Qwen**, а кращі харнеси роблять слабші моделі придатнішими.

Чому це важливо. Обв'язка навколо моделі - інструкції, пам'ять, перевірені рецепти - і є той самий «харнес», який дозволяє слабшій моделі працювати добре. Практичний хід, який тут напрошується: **розділити задачі за класами**. Аналітика і письмо - важка модель. Механічне (переписати скрипт за готовим шаблоном, розпарсити, переформатувати) - дешевша.

ТЕМА 4

Prime Intellect: 153 автономні прогони, 18 моделей, реальні харнеси. Fable закрив 81,7% розриву до людського рекорду, Opus 5 - 53,6%

NanoGPT Speedrun Frontier - **127 балів на HN**, тред. Найцінніше тут те, що моделі гнали **через ті самі харнеси, якими користуються розробники**: `claude-code`, `codex`, `grok-cli`, `kimi-code`.

Задача: автономно оптимізувати рецепт тренування nanoGPT - дійти до цільового лоссу 3.28 за меншу кількість кроків, у межах часу і токенів.

МОДЕЛЬ	ХАРНЕС	ЗАКРИТО РОЗРИВУ
Fable 5	claude-code · high	**81,7%** (8,7 дня)
Opus 5	claude-code · max	**53,6%** (2,9 дня)
Kimi K3	prime-agent · max	52,2%
Opus 4.8	claude-code · max	39,4%
GPT-5.6 Sol	codex · xhigh	35,9%
Sonnet 5	claude-code · max	26,8%
GLM 5.2	pi · high	20,3%

І одразу дірка в методології, яку знайшли на HN. Коментар @ninjahawk1: «why was **Fable 5 tested on high while Opus 5 was tested on max**? Seems like quite a few of them aren't on the same effort setting... that might be viewed as an **experimental error**». Перше місце Fable виміряне на іншому рівні зусиль, ніж друге місце Opus, тож порівнювати колонки напряму не можна.

Другий коментар @Farmadupe б'є по тексту звіту: формулювання на кшталт «a frozen verify.py accepts the claim» він читає як згенерований шлак - «what does it mean to freeze a python script?».

Чому це важливо. Беремо **сам факт заміру**: різниця між моделями на реальній багатоденній автономній задачі вимірюється **в рази**. І та сама модель на різному effort дає різний результат: **effort це справжній важіль**.

ТЕМА 5

Торвальдс дебажив з ШІ і описав його рівно так, як воно є: «кілька разів заявляв, що це неможливо і нерозв'язно»

Цитата, зібрана Вілісоном з реального коміту в ядрі (drm/xe).

Дослівно: «And this was a debug session from hell, **enormously helped by an AI** doing much of the grunt-work. I'd like to call it my tireless helper, but the AI **several times stated flat out that this was impossible and unsolvable** and that we should just write a report about it. I suspect those things have been **trained by people who may not be quite as stubborn as I am**. But while the AI was ready to give up several times, it did keep adding debug code and analyzing it faithfully **when I pushed**. So credit where credit is due and **I let the AI write the commit message** above».

Чому це важливо. Найкорисніша нотатка доби для щоденної роботи: **модель здається раніше, ніж вичерпано способи**. Торвальдс отримав результат тільки тому, що продавлював. Варто тримати це як пряме правило: коли модель каже «це неможливо / даних нема», це гіпотеза, яку треба продавлювати ще раз. Той самий урок, що й «суперечить доказ - падає гіпотеза», тільки описаний з боку того, хто тисне.

ТЕМА 6

Qwen 3.8 27B на одній машині зробив реверс-інжиніринг ліцензійної перевірки за 30 хвилин

Стаття XDA (22.08) - **159 балів на HN**, тред.

Залізо назване чесно: Lenovo ThinkStation PGX на NVIDIA GB10 Grace Blackwell, **128 ГБ уніфікованої пам'яті**, 273 ГБ/с. З коробки 15-30 токенів/с, зі зв'язкою SGLang + NVFP4 + DFlash2 - **близько 50 токенів/с** на коді. Харнес - Pi, модель кликала лише стандартні bash-тули. За оцінкою Artificial Analysis, це **топ open-weights у класі 4-40B зі 135 моделей** (індекс інтелекту 52).

Найцікавіший коментар у треді - @saidinesh5: «майбутнє - це **великі фронтир-моделі, що генерують і оновлюють скіли для "достатньо добрих" локальних моделей**. Багато задач не потребують стільки обчислень - достатньо гарної документації і скілів».

Чому це важливо. Це пункт 3 з іншого кінця і той самий висновок: цінність зміщується з розміру моделі в **інструкції й контекст**. «Фронтир пише інструкції, локальна модель виконує» - той самий поділ праці, тільки на власному залізі.

ТЕМА 7

Stripe дав агентам гаманець: CLI, який видає одноразові платіжні реквізити. 675 зірок, MIT

Анонс @jeff_weinstein (23.08, 32 тис. переглядів), репост Патріка Коллісона. Репозиторій - stripe/link-cli, перевірено через API: **675 зірок, ліцензія MIT**, останній пуш 20.08.

Опис дослівно: «Let your agents spend on your behalf. **Your payment credentials are never exposed. You approve every purchase.**» Механіка: агент отримує або віртуальну картку (працює на будь-якому чекауті, не тільки Stripe), або Shared Payment Token для продавців з підтримкою Machine Payment Protocols. Є **локальний MCP-сервер** і встановлення набору інструкцій: `npx skills add stripe/link-cli`.

Обмеження просто в README: «For now, this is **only available to US Link accounts**». Сьогодні поза США не застосовне.

Чому це важливо. Форм-фактор варто зафіксувати: гроші для агента оформлюються як **одноразовий делегований дозвіл з підтвердженням людиною на кожну покупку**. Це та сама модель «небезпечна дія вимагає явного схвалення», перекладена в платіжний протокол. Коли доїде до ЄС, архітектура вже буде знайома.

ТЕМА 8

Роначер: ШІ зробив вибір мови програмування менш важливим, і люди пішли в «важкі» мови

«Fast and Hard Code» Арміна Роначера (22.08) - **81 бал на HN**, тред.

Теза: «the act of **familiarizing yourself with a language no longer matters** and some of the friction that mattered for humans does not matter for agents. As a result, **LLMs make language choice much less consequential** than it used to be».

Наслідок, який він спостерігає: люди почали обирати мову **за маркетингом**, і виграють від цього Rust і навіть Zig, попри те, що частина ядра Zig-спільноти налаштована проти ШІ. Приклади названі конкретні: новий сервіс Cloudflare Artifacts має **Git-протокольний рушій на чистому Zig, скомпільований у ~100 КБ WebAssembly**; Vercel випустила **fx**, кодинг-агента на Zig.

Чому це важливо. Тверезий фон до пункту 3: якщо вибір мови дешевшає, то дорожчає те, що не автоматизується - **архітектура, перевірка і смак**. Заодно це пояснює, чому в стрічці раптом стільки Zig.

ТЕМА 9

Гаррі Тан: «системи обліку муситимуть стати ШІ-харнесами або їх замінять агенти»

Пост @garrytan (24.08, 04:22 за Києвом - свіжак, 2,4 тис. переглядів за дев'ять хвилин на момент збору). Одне речення без розгортання, тому подано як тезу: «Prediction: **systems of record will need to become AI harnesses or face replacement by agents**».

Варто покласти сюди, бо це те саме слово «харнес» з пунктів 3, 4 і 6, тільки застосоване до продуктів: цінність переїжджає з «де лежать дані» на «що з ними може зробити агент».

ТЕМА 10

Молік про побічний ефект ботів і про чесність у дослідженнях ШІ

Два спостереження @emollick за добу, обидва дрібні, але лягають у тему.

Перше (24.08, 02:53, **7,9 тис. переглядів**): «Annoying side effect of all the AI bots on the site is that they are all "**well-read**" and therefore reply cogently - my niche reference tweets that would normally attract a small but interested group **now get LLM replies**». Нішевий сигнал у твітері перестає працювати як фільтр «свій-чужий».

Друге, важливіше, про методологію (23.08, 21 тис. переглядів): публікувати дослідження впливу ШІ на **старих моделях** не є злочином, але «it requires a **very careful discussion** and has to be **very clear to non-technical readers**». Це саме та проблема, з якою

стикається щоденна робота з цифрами: цифра з квітневої моделі, подана без дати, читається як цифра про сьогодні.

misc: Молік про споживчий ШІ - 34 тис. переглядів, «healthcare, government, personal finance, school forms» - сфери, де люди «muddle by, but need help they can't get», і саме тому consumer AI недооцінений · він же про фінансові поради LLM - дослідження MIT і Стенфорда: більшості людей поради моделей вигідніші за їхні власні рішення, але якість залежить від того, **які питання ставляться** · @amasad: «тиждень має 7 днів - значить 7 релізів» 26 тис. переглядів, Replit за тиждень: Free Mode, Conversations, Routines, імпорт інструкцій з GitHub, black-box пентести · @lennysan про ентерпрайз-продажі 266 тис. переглядів - «якщо твій win rate вище 35%, ціна занизька», 15 стадій замість 5 · @levie: «I'm good to retire now» 240 тис. переглядів · @levelsio про пластик у домі - 129 тис. переглядів, меблі й килими майже поголовно синтетика; окремо про Грецію - **85% домівок з кондиціонером проти 20% у середньому по Європі** · @ClementDelangue про NVIDIA AVO 42 тис. переглядів - вчорашній пункт 1, тепер з боку Hugging Face · Кімната смерті HN - малварь у прошивці автомобільних головних пристроїв на Android, 87 балів
