

Anthropic і Moonshot зіткнулись за один день

Kimi K3 (2.8T параметрів, ваги відкриті) б'є Orus 4.8 на п'яти бенчмарках у день заяви Anthropic

вівторок, 28 липня 2026

У ЦЬОМУ ВИПУСКУ

1. Amodeli нарешті сказав це вголос: «ми ніколи не виступали за заборону»
2. Kimi K3: 2.8T параметрів, мільйон контексту, ваги викладені - і Orus 4.8 програє на п'яти бенчмарках
3. «Rust-переписування Bun» від Claude - розбір, який варто прочитати саме тобі
4. SlopCodeBench: Orus 5 пише вп'ятеро більше коду і 93% рядків тригерять детектор шлаку
5. Microsoft вийшла з кібербезпековою моделлю і б'є Anthropic на 12 пунктів
6. Levie проти WSJ: «негативний сценарій з робочими місцями просто не стається»
7. Невдоволення тоном Orus 5 - і воно вилізло у твоєму ж листі
8. Kimi K3 обростає екосистемою за години, а не тижні
9. Hugging Face і NVIDIA зібрали Open Secure AI Alliance
10. Мольік згенерував три ігри за добу і виклав вихідники - тепер уже четверту

1. Amodeli сказав це вголос: «ми ніколи не виступали за заборону» 614 балів на HN, 849 коментарів, 2.2М переглядів у твіті Anthropic. Закриття сюжету, що ведеться з 25.07.

Цитати з першоджерела: **«Anthropic ніколи не виступала за заборону моделей з відкритими вагами»** і окремо **«моделі з відкритими вагами, які не мають небезпечних можливостей, є суспільним благом»**. Зізнання щодо мотивів: заборона **«захистила б американські AI-компанії від конкуренції, але це ніколи не було метою»**.

Чого він хоче натомість, три речі: не продавати Китаю потужні чипи і обладнання для їх виробництва; тиснути на **«промислову дистиляцію»**; **обов'язкове тестування безпеки для всіх достатньо потужних моделей, і відкритих, і закритих**. Застереження щодо відкритих лишається: до них важко приробити запобіжники і їх **неможливо відкликати**.

Чому це важливо. Вчорашній пункт 9 наводив тезу колишнього співробітника Anthropic: хакери на практиці беруть субсидовані підписки. Сьогоднішній лист це **обходить мовчанням**. Amodeli говорить про ризик відкритих ваг і мовчить про

зловживання через власний канал продажу. Це найслабше місце листа. Для тих, хто на підписці, практично нічого не змінюється, але позиція тепер зафіксована документом.

2. Kimi K3: 2.8T параметрів, мільйон контексту, ваги викладені, Opus 4.8 програє на п'яти бенчмарках 1322 бали на HN, топ доби. У твіті Moonshot 7.3M переглядів, і Клем з Hugging Face: «топ-1 у трендах з 4000+ лайків за 30 хвилин - найшвидший ріст релізу за всю історію».

Специфікація з картки моделі: **2.8T загальних параметрів, 104B активних, 93 шари, 896 експертів** (16 на токен), контекст **1 048 576 токенів**, квантизація MXFP4 з навчанням під неї. Заявка архітектури: **2.5× інтелекту на одиницю обчислень**.

Бенчмарки проти Opus 4.8 і GPT-5.5 (з офіційної таблиці):

БЕНЧМАРК	KIMI K3	OPUS 4.8	GPT-5.5
GPQA Diamond	93.5	91.0	93.5
DeepSWE	67.5	59.0	67.0
Terminal-Bench 2.1	88.3	84.6	83.4
BrowseComp	91.2	84.3	84.4

Чесна мітка: порівняння з **Opus 4.8**, не з Opus 5. Тобто китайці догнали позаминулий фронтір і виклали його безкоштовно. Ліцензія теж **власна, Kimi K3 License**.

Чому це важливо. Та сама історія з пункту 1, тільки фактом: поки Amodei пише лист про ризики, конкурент викладає ваги. Factory вже дав K3 у Droid зі знижкою 50% до 10.08. Terminal-Bench 88.3 означає, що агентські задачі рівня звичайного робочого репозиторію вже робить безкоштовна модель. Альтернатива існує, якщо підписка колись стане проблемою.

3. «Rust-переписування Bun» від Claude: розбір, який варто прочитати 459 балів, 358 коментарів. Найкорисніша історія доби про межі того, що робиться щодня.

Офіційна версія була гучна: Bun переписали на Rust агентами за **\$165 000 за 11 днів** (3-14.05). Автор пішов і подивився на репо. Що там насправді:

• **релізу нема досі**, 11 тижнів після останнього • відкриті PR від Claude виросли з **1 277 (09.07) до 2 475 (27.07)**, черга зростає • **кожен PR мерджиться 40 хв - 1.5 год** через CI; на всю чергу це **~86 діб безперервного мерджу** • реальна вартість, за його оцінкою, ближча до **\$800 000**

Дослівно: «не можна приймати на віру, що переписування "зроблене" або що воно було зроблене за \$165 тис.»

Чому це важливо, і це найпряміший урок за тиждень. Агент генерує швидше, ніж система встигає інтегрувати. Вузьке місце - **рев'ю і мердж**. Той самий патерн масштабується вниз: за вечір агент нагенерує в невеликому проєкті більше змін, ніж встигне перевірити одна людина. Правило «незворотні зміни тільки з явного дозволу»

- те, що тримає чергу інтеграції керованою. Цифра 2 475 відкритих PR показує ціну відсутності такого правила.

4. SlopCodeBench: Opus 5 пише вп'ятеро більше коду і 93% рядків тригерають детектор шлаку 186 балів. Бенчмарк, якого бракувало: він міряє, чи не загіджена кодова база за багато ітерацій.

Механіка: вимоги видаються чекпойнтами (17 штук на трьох задачах), і на кожному треба **адаптувати вже написане**. Строгий залік - усі нові тести плюс усі регресійні з попередніх чекпойнтів.

Результат Opus 5: **24% строгих проходжень (4 з 17)**, краще за Opus 4.6 (17%), формально перемога. Друга колонка:

- **29 065 рядків коду** проти ~9 000 у конкурентів, з них 51% тести • детектор шлаку спрацював на **93% написаних рядків** • функцій написано **у 5 разів більше**, ніж в Opus 4.8, і лише **14.9% з них використані один раз** (проти 49.1% в Opus 4.8)

- цикломатична складність і багатослівність росли з кожним чекпойнтом (65% → 80%)

Висновок автора дослівно: **«сьогоднішні моделі не можна лишати працювати з вимкненим світлом, без керування»**.

Чому це важливо. Перегукується з вчорашнім гайдом Anthropic про судження замість правил і дає робоче правило: на довгих ітеративних задачах агента треба **зупиняти й перерахувати** на кожному чекпойнті. Метрика «14.9% функцій одноразові» - готовий сигнал: якщо агент почав плодити хелпери під кожен випадок, ітерацію пора переривати.

5. Microsoft вийшла з кібербезпековою моделлю і б'є Anthropic на 12 пунктів 224 бали плюс твіт The Rundown. Перша кібербезпекова модель Microsoft AI, MAI-Cyber-1-Flash.

Цифри з першоджерела: **95.95% на CyberGym** проти 83.2-85.6% у решти, тобто **+12 пунктів над Mythos від Anthropic**, при **50% економії вартості**. Працює всередині MDASH, їхнього харнесу зі **100+ агентів на різних моделях**, що шукають і латають вразливості. Модель тягне **до 90% задач**, дорогі моделі лишаються на решту 10%.

Чому це важливо. Знову те саме слово, **харнес** (вчора його вживав Ванг про Meta, пункт 1). Патерн тижня видно неозброєним оком: перемагає **дешева модель у розумній обгортці, яка кличе дорогою лише на складне**. Та сама архітектура працює на будь-якому розкладі періодичних задач: більшість запусків взагалі обходиться скриптом без LLM, і саме тому все виходить дешево.

6. Levie проти WSJ: «негативний сценарій з робочими місцями просто не стається» Найтверезіший голос у другому листі. Аарон Леві дослівно: **«негативний сценарій AI для робочих місць продовжує не відбуватись, попри прогнози. Значна частина підприємств, з якими він говорить - у різних галузях - все ще наймає, просто з нахилом у бік інших ролей»**.

Чому це важливо. Третій незалежний замір тієї самої теми за три дні: вчора Стенфорд («ефект в агрегаті поки малий»), позавчора levelsio («інді змиває»), сьогодні Леві («корпорати наймають, змінився профіль»). Складається несуперечлива картина: б'є по входу в професію і по інді-ніші, ринок загалом тримається. Дефіцит зараз у тих, хто вміє працювати з агентами.

7. Незадоволення тоном Opus 5, і воно вилізло у листі з новинами Свіже, за останні три години, і поки що маленьке. Сікі Чен (@blader) пише, що **«при всій здатності Fable і Opus 5, є щось дуже некомфортне у спілкуванні з моделлю, налаштованою так патерналістично-повчально; постійне фонове відчуття зверхності»**. Пол Беттнер підхопив жорсткіше: **«моделі - це інструменти для людей. Не самі люди; припиніть тренувати їх інакше»**.

Чому це важливо. Дефолтний тон моделі правиться на рівні харнесу: системний промпт, який задає манеру, знімає більшість цієї скарги ще до першої відповіді. Претензія справедлива, і адресована вона тому, хто постачає дефолт.

8. Kimi K3 обростає екосистемою за години Швидкість: Factory дала K3 у Droid зі знижкою 50% до 10 серпня в день релізу, HF вивів у топ-1 за 30 хвилин. Мольк уже жартує, що «читає ваги найпотужнішої відкритої моделі - сторінка перша: переважно негативні числа, під кінець нулі й одне несподівано велике додатне».

Чому це важливо: відкрита модель тепер отримує інтеграції в день релізу. Це та сама «швидкість старіння обгортки», про яку Черні казав у неділю, тільки з протилежного боку.

9. Hugging Face і NVIDIA зібрали Open Secure AI Alliance 3.8M переглядів у твіті NVIDIA. Індустріальний альянс для пошуку вразливостей у ПЗ і агентах, з відкритим обміном моделями, інструментами і дослідженнями.

Чому це важливо: ще один доказ до пункту 1: навіть безпекова робота зараз організується як **відкрита**. Аргумент Дженсена з учорашнього пункту 2 («весь стек, на якому ви стоїте, вже відкритий») сьогодні отримав інституційне підтвердження.

10. Мольк згенерував три ігри за добу і виклав вихідники, тепер уже четверту Продовження вчорашнього пункту 10. За добу додалась ще одна: **Imminence**, «гра про наближення чогось дуже великого і незбагненого», зроблена одним промптом у Fable, з вихідниками на GitHub. Оцінка власного результату: **«непогано як для такої гри; механіка, текст і сюжет - усе Fable, лише давав фідбек»**.

Окремо думка, яка варта більше за самі ігри: **«знати назви багатьох красивих і цікавих речей - це свого роду суперсила в епоху AI. Можна викликати Vaporwave, Muqarnas, Bauhaus, Sfumato, Grisaille, Notan, Polysyndeton, Zeugma. Треба лише знати, що просити»**.

Чому це важливо: найточніше формулювання того, чим один промпт відрізняється від чужого. Вузьке місце змістилось з «уміти зробити» на **«знати, що саме попросити»**, і словник тут важить більше за навички.

Misc • Opus 5 лежав удруге за два дні, 99 балів, `status.claude.com`, інцидент 27.07 • EU оштрафував Google на \$1.02 млрд за преференції власним сервісам (95)

- Суд відхилив спробу Google через DMCA заборонити скрейпінг себе (287), важливий прецедент для всіх, хто збирає дані • Scriptc від Vercel (274): компілятор TypeScript у нативний бінарник без JS-рушія всередині • «AI-компанії шматують рідкісні книжки» (750) заради сканів для тренування; найгучніша етична історія доби • Nvidia на \$750 млрд угод знову підняла розмови про кругове фінансування AI (81, Bloomberg)

- Акції китайського чипмейкера +470% (206, BBC)

- Професор вставив невидимий промпт у завдання і зловив 32 з 35 студентів на списуванні з AI (94)

- Один пропущений підкреслювач відправив невинного до в'язниці на 18 місяців (202, Ars Technica)

- Вразливість у платформі автопарку Volvo/Eicher давала контроль над усіма користувачами і машинами (144)

- Дослідження нонстік-посуду: покриття сипле частинки, навіть коли виглядає цілим (86, rtings)

- «Вправи працюють проти депресії - чому їх не лікують як ліки» (86, Big Think)