

Meta відкрила ваги Muse Spark Open

Шпигунське ПЗ через рекламу, METR усередині Anthropic і \$15 млн математикам на адаптацію до AI

п'ятниця, 16 жовтня 2026

У ЦЬОМУ ВИПУСКУ

1. Meta відкрила ваги Muse Spark Open, модель іде на двох Mac Studio
2. Таргетована реклама доставляла шпигунське ПЗ без жодного кліку
3. Anthropic пустила METR всередину з доступом рівня співробітника
4. Математики отримали \$15 млн на адаптацію до AI і пишуть про горе
5. Goodfire: нейромережі набагато зрозуміліші, ніж вважала галузь
6. Акції відскочили, але страхівка від дефолту AI-компаній дорожчає
7. Accenture обмежила працівникам Claude до \$29 на місяць
8. Сайти вже пишуть інструкції для агентів у HTTP-заголовках
9. Автор Monopoke: Scion скопіювали навмисно, Anthropic додала згадку
10. Claude за день допоміг написати драйвер для незадокументованого заліза

ТЕМА 1

Meta відкрила ваги Muse Spark Open, модель іде на двох Mac Studio

ДЖЕРЕЛА · Meta AI, блог, 15.10 [Meta](#) · Hugging Face, блог, 15.10 [Hugging Face](#) · Artificial Analysis, 15.10 [Artificialanalysis](#) · The Wall Street Journal, 15.10 [WSJ](#) · Аарон Леві, X, 15.10 [@levie](#) · обговорення на HN, 15.10 [Hacker News](#)

У вівторок тут писали про Muse Wonder, закриту модель Meta. 15.10 компанія виклала ваги іншої моделі цієї серії, Muse Spark Open. Meta називає її найсильнішою американською моделлю з відкритими вагами. Це MoE на 816 млрд параметрів, з них 42 млрд активні на кожен токен. Контекст до 1 млн токенів. Модель приймає текст, зображення, аудіо і відео, а відповідає текстом.

Hugging Face розклала архітектуру. У кожному з 62 шарів 256 експертів, роутер вибирає 8 на токен. Чекпойнт у BF16 важить близько 1,5 ТБ. Повний контекст на 1 млн токенів займає в кеші близько 266 ГБ. Модель уже підтримують Transformers, vLLM, SGLang, llama.cpp і MLX.

Головний трюк релізу - запуск удома. WSJ пише про партнерство Meta і Apple. M5 Ultra в Mac Studio дає GPU до 512 ГБ уніфікованої пам'яті. Тому дві машини разом тримають усю модель. Глава Hugging Face Клеман Деланж уже показав запуск квантованої версії

на двох Mac Studio. Ліцензія Muse Model License дозволяє комерційне використання з обмеженнями.

Незалежна оцінка стриманіша за заяву Meta. Artificial Analysis дає моделі 39 балів у своєму індексі. Це на рівні DeepSeek V4.1 Flash (39), але нижче за MiMo-V2.6-Pro (46) і GLM-5.3 (45).

Ціна через API – \$1,25 за 1 млн вхідних токенів і \$4,25 за вихідні. Для відкритої моделі такого розміру це дорого. На HN кілька розробників пишуть, що китайські моделі поменше дають той самий результат дешевше.

Чому це важливо. Звання «найкраща американська відкрита модель» важить для політики і корпоративних закупівель: частина компаній не бере китайські ваги. Технічно ж лідерство відкритих моделей лишається в Китаї. Нове тут – залізо: модель рівня фронтиру рік тому тепер працює на двох настільних комп'ютерах без хмари.

ТЕМА 2

Таргетована реклама доставляла шпигунське ПЗ без жодного кліку

ДЖЕРЕЛА Citizen Lab, звіт, 15.10 [Citizenlab](#) · Ars Technica, 15.10 [Ars Technica](#) · The Guardian, 15.10 [Guardian](#) · обговорення на HN, 15.10 [Hacker News](#)

Citizen Lab з Університету Торонто описала кампанію AdFection. За даними лабораторії, це перше підтверджене зараження комерційним шпигунським ПЗ через таргетовану рекламу. Жертві не треба нічого натискати. Достатньо відкрити сайт або застосунок, де показують рекламу. Атака працювала на iOS і Android. Citizen Lab підтвердила 45 заражених пристроїв у Al Jazeera у 2024–2025 роках.

Схему побудувала ізраїльська компанія Discover-Tech, яка називає себе рекламною агенцією. Її платформа Controlium збирає профілі людей з рекламних аукціонів: ідентифікатор пристрою, IP, геолокацію. Коли ціль збігається з профілем, реклама завантажує код з серверів Candiru. Candiru тепер працює під назвою Integrity Labs. Її сервер віддає експлоїт, і пристрій заражається за секунди.

Масштаб великий. Citizen Lab отримала понад 4 млрд записів ставок Discover-Tech. З березня до червня 2026 року її реклама йшла у понад 70 000 застосунків і 120 000 сайтів. Покази траплялись у мережах НАТО, ООН, європейських парламентів і армій Норвегії та Швеції. Google і IAB Europe мали Discover-Tech у списку перевірених постачальників. Після звіту Google відрізала компанію від своїх рекламних бірж. Discover-Tech і Candiru на запити не відповіли.

Чому це важливо. Рекламний аукціон за мілісекунди передає профіль людини сотням компаній.

Досі це вважали проблемою приватності. Звіт показує, що той самий канал доставляє експлоїт конкретній людині. Для редакцій, урядів і компаній блокувальник реклами

стає засобом безпеки.

Citizen Lab також радить скинути рекламний ідентифікатор у налаштуваннях телефону.

ТЕМА 3

Anthropic пустила METR всередину з доступом рівня співробітника

ДЖЕРЕЛА Anthropic, 15.10 [Anthropic](#) · Anthropic, X, 15.10 [@AnthropicAI](#) · Financial Times, 15.10 [FT](#) · Ars Technica, 15.10 [Ars Technica](#)

13.09 тут писали про есе Даріо Амодея. Тоді Anthropic лише пообіцяла пустити сторонніх оцінювачів усередину. 15.10 компанія повідомила, що шість працівників METR уже мають робоче залізо, офіс і доступ до внутрішніх систем. Рівень доступу приблизно такий, як у власних команд оцінки ризиків. Закриті лише кілька категорій: дані клієнтів, адвокатська таємниця, архітектура моделей.

METR публікуватиме висновки про рівень ризику, інциденти і практики компанії. Anthropic не може заблокувати публікацію. Вона може лише вимагати прибрати інформацію з вузьких категорій. METR має право публічно сказати, чи вважає ці правки виправданими. Компанія порівнює схему з банківським наглядом: регулятор садить своїх інспекторів усередину великого банку.

Для інших партнерів Anthropic будує три рівні доступу. SecureBio зараз перевіряє розділ про хімічну і біологічну зброю в серпневому звіті про ризики. З Accenture і Faculty компанія планує ширшу співпрацю, з ред-тімінгом і оцінкою узгодженості. Hugging Face серед організацій, з якими переговори ще на ранній стадії.

Того ж тижня, за даними Axios у переказі Ars, FTC надіслала Anthropic і OpenAI вимоги надати документи. Запит займає понад 50 сторінок питань про недавні інциденти з агентами і заходи безпеки. Ars нагадує, що METR, імовірно, теж отримає повістку.

Чому це важливо. Це перший випадок, коли лабораторія дає сторонній організації такий доступ і право публікації без вето. Перевірка буде в першому звіті METR: що він скаже і скільки в ньому буде прихованого. Є і слабе місце. METR роками співпрацює з Anthropic. Тому незалежність тут тримається лише на договорі.

ТЕМА 4

Математики отримали \$15 млн на адаптацію до AI і пишуть про горе

ДЖЕРЕЛА Simons Foundation, 15.10 [Simonsfoundation](#) · OpenAI, X, 15.10 [@OpenAI](#) · Anthropic, X, 15.10 [@AnthropicAI](#) · Дерек Томпсон, 15.10 [Derekthompson](#) · Рікаб Гамбір у блозі Теренса Тао, 15.10 [Wordpress](#) · обговорення на HN, 15.10 [Hacker News](#)

07.10 OpenAI виклала 722 математичні роботи, які написали її моделі. Через тиждень реакція спільноти перейшла в гроші і в есеї. Simons Foundation разом з Anthropic і

OpenAI виділяє \$15 млн. OpenAI і Anthropic дають по \$5 млн. Фонд роздаватиме гроші на семінари і проекти:

перевірка результатів з AI, навчання новим інструментам, спільні стандарти. Президент фонду Девід Спергел каже, що правила має задавати сама спільнота математиків.

Есей Дерекі Томпсона «Я знаю, чому плачуть математики» зібрав на HN 575 коментарів. Він наводить слова філдсівського лауреата Гюго Дюміль-Копена: «Вони знищили мою галузь досліджень». Стівен Строгац з Корнелла каже, що колеги говорять про «видобуток відкритим способом» їхнього предмета.

Томпсон пояснює горе так: наука – це ремесло і розмова. AI добре робить ремесло, але обриває тисячолітню розмову між людьми.

У блозі Теренса Тао вийшов гостьовий пост фізика Рікаба Гамбіра. Це сатира. Він пише, що OpenAI перекладає на математиків найнудніше: перевірку і розуміння результатів. Самі результати генеруються швидше, ніж їх хтось прочитає.

Чому це важливо. Результатів у математиків тепер багато. Бракує людей, що можуть ці результати перевірити і зрозуміти. Гроші від тих самих компаній, що створили потік, частково це визнають. Інші галузі, де AI генерує результати масово, зустрінуть те саме питання:

хто перевіряє.

ТЕМА 5

Goodfire: нейромережі набагато зрозуміліші, ніж вважала галузь

ДЖЕРЕЛА Нік Каммарата, Goodfire, 15.10 [Goodfire](#) · The Rundown AI, X, 15.10 [@TheRundownAI](#) · обговорення на HN, 15.10 [Hacker News](#)

Нік Каммарата раніше працював над інтерпретованістю в OpenAI, тепер він у Goodfire. Він виклав есей на 65 хвилин читання. Головна теза: нейрони в моделях здебільшого мають одну зрозумілу роль. Галузь роками вважала інакше. Пояснення звалось суперпозицією: мережа нібито пакує в один нейрон непов'язані поняття, бо їй бракує місця.

Каммарата переглянув канонічні приклади, які сам колись публікував. Знаменитий нейрон «коти і машини» в InceptionV1 виявився детектором V-подібної форми. Схема, де машина «ховалась» серед собак, насправді об'єднувала їх за позою. У 2023 році автоматичний проект OpenAI вважав зрозумілими 1,7% нейронів. Попередній аудит Goodfire на Qwen3-32B дає таку картину: близько 90% нейронів мають одну головну роль, ще ~7% мають дві споріднені, ~2-4% поки незрозумілі.

Чому це важливо. Якщо теза підтвердиться, інтерпретованість стає звичайною наукою на самій мережі. Окремий розкладач на кшталт sparse autoencoder тоді не

обов'язковий. Це прямо стосується безпеки: перевіряти модель зсередини стає реальніше. Поки це есей і «робоча гіпотеза», технічна стаття ще не вийшла.

ТЕМА 6

Акції відскочили, але страховка від дефолту AI-компаній дорожчає

ДЖЕРЕЛА Semafor, 15.10 [Semafor](#) · Semafor, 15.10 [Semafor](#) · Financial Times, 15.10 [FT](#) · Financial Times, 15.10 [FT](#)

Учора тут писали, що банки шукають, кому віддати борг дата-центрів. 15.10 акції в США відіграли частину втрат після кількох днів розпродажу. Але ринок кредитів рухається в інший бік.

Semafor пише, що інвестори масово купують кредитні дефолтні свопи на гіперскейлерів.

Найбільше подорожчав захист для найбільших боржників. Страховка \$1 млн облігацій CoreWeave на п'ять років рік тому коштувала менше \$40 тис. на рік, тепер \$100 тис. Дохідність облігацій CoreWeave перевищила 15%. Захист від дефолту Oracle за тиждень подорожчав на 12%, до \$28 тис.

Борг Oracle становить \$160 млрд і стоїть на межі «сміттевого» рейтингу. FT у заголовку пише, що інвестори скидають облігації дата-центрів, і називає це «дарвінівським моментом».

На тлі цього TSMC звітувала про стрибок прибутку. Компанія каже, що будує заводи «небаченим темпом» через «надзвичайно сильний» попит на AI.

Чому це важливо. Акції реагують на настрої за день, а ціна страховки показує, як ринок оцінює реальний ризик. Розрив між ними росте: виробники чипів заробляють рекордно, а тих, хто будує дата-центри в борг, кредитори вже бояться. Найслабші ланки тут – компанії з великим боргом і одним типом клієнта.

ТЕМА 7

Accenture обмежила працівникам Claude до \$29 на місяць

ДЖЕРЕЛА 404 Media, 15.10 [404media](#) · Сіці Чен, X, 15.10 [@blader](#)

404 Media побачила внутрішнє оголошення Accenture. З 01.10 компанія централізовано оплачує лише \$29 на місяць використання Claude на людину. Усе понад це має оплачувати конкретний проект. Деякі працівники витрачали \$500-1 000 на місяць. Один з них пише, що інтенсивно використовує Claude Code для «AI-акселератора» для клієнтів. Іншому доступ закрили повністю. На преміум-плані ліміт інший:

\$250 токенів на тиждень. Accenture каже, що надає «правильний інструмент для правильної роботи».

Іронія в тому, що Accenture радить клієнтам вкладатись в AI. Того ж дня Anthropic назвала її серед своїх зовнішніх оцінювачів безпеки (тема 3). Сіці Чен відреагував: правильніше було б дати весь бюджет 1% найактивніших користувачів, а решті нічого. [одне джерело]

Чому це важливо. Споживання AI-агентами росте швидше за бюджети. Великі компанії переходять від безлімітних ліцензій до оплати за спожите. Середній ліміт на всіх б'є саме по тих, хто отримує від інструмента найбільше. Ця суперечність тепер виникатиме в кожній великій компанії.

ТЕМА 8

Сайти вже пишуть інструкції для агентів у HTTP-заголовках

ДЖЕРЕЛА [Signals Corps](#), 15.10 [Signalscorps](#) · обговорення на HN, 15.10 [Hacker News](#)

Signals Corps перевірила заголовки 217 млн відповідей з вересневого Common Crawl. Це 10% усього обходу. 7 959 сайтів надсилали заголовок, адресований AI. Майже всі лише вказують агенту на llms.txt або MCP-сервер. 30 сайтів намагались дати агенту команди, 24 з них у заголовку X-AI.

Більшість цих спроб – жарти. Реальна шкода можлива, лише коли агент має відповідні права, наприклад shell або доступ до бази.

Ще 139 сайтів записували в заголовок, чи вважають клієнта AI-ботом. Тобто сервери вже розрізняють агента і людину і можуть показувати їм різне. Автори радять стежити за діями агента після запиту. Фільтр за назвами заголовків дав би 910 хибних тривог.

Чому це важливо. Агент читає не лише текст сторінки. Заголовки, на які людина не дивиться, теж потрапляють у його контекст. Будь-яка адреса, яку підказав сервер, для агента має бути новим недовірим запитом.

ТЕМА 9

Автор Mononoke: Scion скопіювали навмисно, Anthropic додала згадку

ДЖЕРЕЛА [Rain \(sunshowers.io\)](#), 15.10 [Sunshowers](#) · обговорення на HN, 15.10 [Hacker News](#)

Учора тут писали про пост авторки Mononoke. Вона порівняла Scion від Anthropic з сервером контролю версій Meta. У другому пості вона йде далі. Після першого посту вона приватно отримала підтвердження, що копіювання було навмисним. За її словами, команди Scion і Mononoke мають спільних людей. Подібність охоплює всю модель даних. Навіть назва «Scion» переставляє склади «Bonsai», як у Mononoke звуться набори змін.

Вона знову наголошує, що проблема в атрибуції. Копіювати хороші ідеї можна. Але історія «наша модель збудувала і запустила систему за 5 місяців» звучить інакше, якщо

команда Meta будувала її 10 років. В оновленні до посту вона пише, що Anthropic відповіла на HN і додала в блог достатню згадку.

Чому це важливо. Розповідь про те, що збудувала модель, тепер має комерційну вагу. Тому питання «що було до неї» стає частиною перевірки таких заяв. Anthropic виправила текст за добу, але тільки після публічного тиску.

ТЕМА 10

Claude за день допоміг написати драйвер для незадокументованого заліза

ДЖЕРЕЛА Крістіан Шаллер, блог GNOME, 15.10 [Gnome](#) · обговорення на HN, 15.10 [Hacker News](#)

Крістіан Шаллер з Red Hat хотів відсканувати старі негативи. Його сканер Plustek змінив чипсет і перестав працювати в Linux. Він дав Claude віртуальну машину з Windows і офіційним драйвером.

Хост на Linux записував USB-трафік через usbmon. Модель читала тисячі рядків дампу, шукала закономірності і пропонувала, що означає кожен байт. Сканер запрацював.

Потім Шаллер дав ту саму схему для камери MIPI у своєму ноутбучі. За один день він отримав патчі для ядра і libcamera. Тепер він виклав узагальнену версію hw-harness для Claude і Codex.

Чому це важливо. Драйвер для незадокументованого пристрою раніше вимагав тижнів ручного аналізу дампу. Модель робить цю нудну частину, а людина перевіряє результат на реальному залізі.

Для Linux, де виробники часто не пишуть драйверів, це може змінити підтримку пристроїв.

misc

- **NVIDIA (@NVIDIAAI)** випустила відкриту модель Nemotron 3.5 Super VL: 120 млрд параметрів, 12 млрд активних. Вона працює з текстом, зображеннями, документами і відео для агентів.
[@NVIDIAAI](#)
- **Black Forest Labs (@bfl_ai)** відкрила ваги FLUX 3 Image Klein: 5,6 млрд параметрів, працює на одній споживчій відеокарті, нативний 4K. [@bfl_ai](#)
- **Krea 3:** модель для редагування зображень робить правку менш ніж за 8 секунд за \$0,06. За графіком самої Krea, у Nano Banana Pro це 29 секунд і \$0,135. [одне джерело]
[@TheRunDownAI](#)
- **OpenAI (@OpenAI)** викотила GPT-Live-1.1 для голосу в ChatGPT і Codex: нові голоси і кращі акценти шістьма мовами. [@OpenAI](#)

- **OpenAI (@OpenAI)** відкрила GPT-Rosalind для академічних біолабораторій США: до 12 місяців спонсорованого доступу. [@OpenAI](#)
- **Steam** за добу після запуску нового лаунчера Epic оновив свій клієнт. Холодний старт став 5,23 секунди, і Steam знову швидший за Epic. [Respawnfirst](#)
- **Tesla Roadster:** Tesla показала новий Roadster у Техасі. Машина проїхала по вертикальній рампі, коротко пролетіла на стисненому повітрі і розігнала до 96 км/год за 0,81 секунди.
Ціни і дати поставок не назвали. Ітан Моллік (@emollick) порівняв демо з презентацією iPhone і нагадав, що від демо до продукту часто минають роки.
[Theverge](#) · [@emollick](#)
- **Секунда стає годиною:** делегати з 69 країн на конференції Міжнародного бюро мір і ваг проголосували замінити високосну секунду на високосну годину. Програмісти, які ловили баги на хвилинах з 61 секунди, можуть видихнути. [Guardian](#)
- **Withings:** нові акаунти з 13.10 мають платити \$99,99 на рік за Withings+, щоб синхронізувати дані з Apple Health. Скаргу користувача на форумі підтримки підняли на HN.
[Withings](#)
- **Current Affairs** за новим договором з профспілкою заборонила будь-який текст і ілюстрації, згенеровані AI. [Currentaffairs](#)