AI-ДАЙДЖЕСТ

Opus 5.5 і GPT-6 Sol: цінова війна лабораторій за одну годину

Мінус 40% і мінус 50% до цін, асистент Meta віддав 6,8 ГБ своїх файлів, GPT-6 Astra зламала Енігму 1941 року.

середа, 23 вересня 2026

У ЦЬОМУ ВИПУСКУ

1. Claude Opus 5.5: рівень Fable 5.1 на більшості задач і на 40% дешевше за Opus 5
2. GPT-6 Sol і Luna за годину після Opus 5.5: удвічі дешевше за попередників
3. Асистент Muse від Meta: дірка в macOS-застосунку, файлова система на 6,8 ГБ і блокування від Amazon
4. GPT-6 Astra сама розшифрувала повідомлення Енігми 1941 року, яке не піддавалось з 2005-го
5. Пентагон: застаріла розвідка і надмірна довіра до AI призвели до удару по школі в Мінабі
6. Британська Колумбія подала позов проти OpenAI через стрілянину в Тамблер-Ридж
7. Трамп в ООН: AI тепер «суперінтелект», а глобальний контроль над ним США відкидають
8. Microsoft закрила EvilTokens - сервіс з AI-чатботом, який читав зламані скриньки і планував шахрайство
9. Apple прибрала вимикач Apple Intelligence у macOS 27, а в iOS з'явились банери, які не закрити
10. «Спаймарки»: чому приховані водяні знаки AI пропонують назвати інакше

ТЕМА 1

Claude Opus 5.5: рівень Fable 5.1 на більшості задач і на 40% дешевше за Opus 5

ДЖЕРЕЛА анонс Anthropic, 22.09 Anthropic · Саймон Віллісон, 22.09 Simon Willison · Artificial Analysis Artificialanalysis · обговорення Hacker News

Anthropic випустила Claude Opus 5.5, першу модель нової родини 5.5. Головне твердження компанії вміщається в одне речення з анонсу: модель **працює на рівні Claude Fable 5.1 на більшості задач і коштує на 40% менше в роботі, ніж Opus 5**. Sonnet 5.5 і Haiku 5.5 обіцяють «у найближчі тижні».

Звідки береться 40%, розписано цифрами. Токени подешевшали на 20%: **$4 за мільйон на вході і $20 на виході** проти $5 і $25 у Opus 5. Сильніше впало читання з кешу, тобто повторна обробка того самого контексту, яку агент робить на кожному кроці: **$0,20 проти $0,50, мінус 60%**. Anthropic сама пише, що саме кеш становить більшість витрат

в агентній роботі і програмуванні. Решту економії компанія пояснює тим, що модель витрачає менше токенів на задачу. Плюс вивід швидший більш ніж на 30%, а в передплатах Pro, Max і Team підняли п'ятигодинні ліміти.

Бенчмарки в анонсі - власні заміри Anthropic і партнерів. На Terminal-Bench 4.0, де агент виконує багатокрокові задачі в терміналі, Opus 5.5 має **66,4%** проти 55,8% у Fable 5.1 і 57,9% у GPT-6 Astra. На FrontierCode, де оцінюють, чи прийняли б зміни агента в реальний код, **54,4% проти 53,3% в Astra**, і, за підрахунком компанії, приблизно за п'яту частину вартості задачі. Сама Anthropic тут же робить застереження, яке варто процитувати: на цьому рівні можливостей **розрив у бенчмарках став менш надійним орієнтиром**, і в їхньому власному використанні різниця з Fable 5.1 менша, ніж показують таблиці.

Про безпеку компанія пише, що це її перший реліз після заклику сповільнити передній край і що модель до виходу перевіряли зовнішні оцінювачі, зокрема METR. Конкретна цифра: у новому тесті на схильність виходити за встановлені межі Opus 5.5 **намагався їх обійти приблизно на 85% рідше, ніж Opus 5**, і всі спроби, за словами компанії, були низької серйозності і модель сама про них повідомила.

Незалежні заміри вже є, і вони додають дві деталі. **Artificial Analysis** дала моделі 58 балів у своєму зведеному індексі на максимальному рівні міркувань, але зазначила, що на прогін вона витратила **260 млн токенів при медіані 88 млн**, а весь прогін коштував $8 708.

Тобто на найвищому рівні модель дуже багатослівна. **Саймон Віллісон** натрапив на те саме з іншого боку: на режимі «max» Opus 5.5 двічі поспіль не повернув відповіді на його тестове прохання намалювати пелікана на велосипеді. Модель міркувала, поки не вперлась у ліміт 128 тисяч вихідних токенів, і кожна спроба коштувала $2,56 і майже 20 хвилин. Висновок Віллісона: режиму «max» він тепер не довіряє.

Практики з перших годин пишуть переважно позитивно. Factory нарахувала **на 20-25% менше вихідних токенів**, ніж у Opus 5 при тому самому рівні зусиль, і радить середній рівень як основний (@FactoryAI). Box на своїх корпоративних задачах пише про **63% менше токенів** (@levie).

Ітан Моллік назвав її першою моделлю поза Fable і Astra, яка відчувається як модель того класу, але додав, що звичка останніх Claude писати надто щільно нікуди не зникла (@emollick).

Чому це важливо. Для тих, хто платить за агентну роботу, ключова цифра тут - ціна кешу, і вона важить більше за ціну токена. Агент на кожному кроці перечитує той самий контекст, тому в довгих сесіях саме повторне читання становить левову частку рахунку, і мінус 60% на ньому відчутніший за мінус 20% на решті. Друге спостереження стосується налаштувань: історія Віллісона і цифра Artificial Analysis разом показують, що найвищий рівень міркувань більше не є безпечним вибором «на всяк випадок». Модель на ньому може спалити бюджет і не дати результату, тому рівень зусиль варто підбирати під задачу і міряти на власних прикладах.

ТЕМА 2

GPT-6 Sol і Luna за годину після Opus 5.5: удвічі дешевше за попередників

ДЖЕРЕЛА анонс OpenAI, 22.09 OpenAI · Ars Technica, 22.09 Ars Technica · Саймон Віллісон Simon Willison · обговорення Hacker News

OpenAI відповіла приблизно через годину. GPT-6 Sol і GPT-6 Luna - молодші моделі родини GPT-6, навчені, за словами компанії, подібними методами, що й флагманська GPT-6 Astra, яка вийшла на початку місяця. Головна цифра - ціна, **мінус 50% до попередників**:

Sol коштує **$2 за мільйон токенів на вході і $10 на виході** (було $4 і $20), Luna - **$0,10 і $0,50** (було $0,20 і $1,20). OpenAI порівнює з акційними цінами GPT-5.6, а Віллісон нагадує, що на листопад для GPT-5.6 заплановане ще й подорожчання на 25%.

Щоб зрозуміти масштаб, корисна таблиця Віллісона. Opus 5.5 після зниження коштує $4 і $20, тобто рівно стільки, скільки коштувала GPT-5.6 Sol до цього релізу. Нова Sol тепер удвічі дешевша за Opus 5.5. А Luna за $0,10 і $0,50 - одна з найдешевших моделей в історії OpenAI; дешевші були лише значно слабші GPT-4.1 Nano і GPT-5 Nano. Для порівняння, Haiku 4.5 від Anthropic коштує $1 і $5, тобто вдесятеро дорожче за Luna.

Бенчмарки в анонсі OpenAI побудовані за тією самою логікою, що в Anthropic: якість на долар. На AutomationBench, де агент виконує бізнес-процеси через 47 інструментів, Sol на високому рівні зусиль набирає **33,2% при вартості $0,27 за задачу** і, за підрахунком компанії, обходить Claude Opus 5 за 9% його вартості. Важливе застереження: ці порівняння OpenAI робила з Opus 5 і Fable 5.1, бо Opus 5.5 на момент написання анонсу ще не вийшов.

На DeepSWE Sol дає 68,8%, у межах 1,1 пункту від найкращого результату Fable 5. Окремо компанія пише, що на внутрішньому тесті фактичності Sol робить **приблизно вдвічі менше помилок**, ніж попередник, і що покращила кешування: вищий відсоток збігів за замовчуванням, знижка 90% на кешовані токени і нова панель, яка показує, де кеш не спрацював.

Ars Technica формулює суть обох релізів одним заголовком: «трохи більше за значно менші гроші». Автор, Семюел Аксон, пояснює, чому лабораторії раптом змагаються ціною: великі клієнти почали ставити маршрутизатори, які відправляють простіші запити на дешевші відкриті моделі, і дорогим флагманам треба повертати цей трафік. Його ширша думка: частина корпоративних клієнтів уже не вимагає кращої якості, їм потрібні передбачувані впровадження і розумна ціна, і це може стати природним сповільненням без жодних заяв про безпеку.

Реакція в стрічках - переважно про гроші. Аарон Леві з Box назвав день «божевільним» саме через ціни (@levie). Сікі Чен ще напередодні писав, що його сервіс перевів усіх клієнтів на моделі класу Astra і Fable 5.1 і після оптимізації кешу платить приблизно третину того, що раніше коштували Sol і Opus (@blader). Дармеш Шах відповів жартом:

розсилка, яка пише підписникам лише в ті тижні, коли не вийшла нова модель (@dharmesh).

Чому це важливо. Коли два головні постачальники за одну годину знижують ціни і обидва обґрунтовують це кешем і економією токенів, стає зрозуміло, де тепер ідуть змагання:

у вартості виконаної задачі. Місце в рейтингу відходить на другий план. Для тих, хто будує продукти поверх моделей, це конкретна робота на найближчий тиждень: переміряти, яка модель на яких задачах дає потрібну якість за найменші гроші. Таблиці з ціною за мільйон токенів тут вводять в оману: модель з дешевшим токеном, яка витрачає втричі більше токенів, виходить дорожчою.

Обидві компанії самі зміщують розмову на «ціну за задачу», і саме цю метрику варто рахувати на своїх даних.

ТЕМА 3

Асистент Muse від Meta: дірка в macOS-застосунку, файлова система на 6,8 ГБ і блокування від Amazon

ДЖЕРЕЛА Ars Technica, 22.09 Ars Technica · розбір дослідника, 22.09 Mouse · блог Meta про безпеку Muse Meta · The Rundown, 22.09 Beehiiv · обговорення Hacker News · Hacker News

20.09 тут згадувався каталог конекторів для Muse, який Meta поки описала лише на словах.

За дві доби навколо самого асистента назбиралось три окремі історії, і всі три про одне:

наскільки великий доступ йому дали.

Перша - дірка в застосунку для Mac. Дослідник безпеки macOS Патрік Вордл знайшов, що будь-яка програма на комп'ютері чи команда в терміналі може змінювати низку незадокументованих налаштувань Muse. Більшість із них нешкідливі, на кшталт темної теми.

Але одне задає адресу сервера, куди йде розпізнавання голосу. Підмінивши її на свою, зловмисник отримує токен входу в акаунт Muse і з ним повний контроль над асистентом.

Вордл формулює наслідок так: замість писати складну шкідливу програму, можна просто скористатись правами асистента, який уже має доступ до файлів, мікрофона, камери і календаря. За даними Ars, **Meta випустила виправлення більш ніж через 12 годин** після публікації. Вордл окремо критикує вибір розпізнавати голос у хмарі, коли macOS давно вміє робити це на пристрої.

Друга - експорт файлової системи. Автор блогу mouse.dev попросив Muse заархівувати всі файли, які той бачить, і надіслати на Google Drive. Асистент так і зробив. Архів

важив приблизно **2,7 ГБ стиснутим і 6,8 ГБ розпакованим** і, за словами автора, містив кореневу файлову систему середовища, виділеного під його сесію: системні файли, внутрішню документацію, код інтеграцій, пам'ять, журнали агентів і **файли SSH-ключів**.

Усередині - близько 68 навичок, 113 записів субагентів, щоденні файли пам'яті і нічний процес, який переглядає розмови і пише поради для майбутніх сесій. Автор чесно обмежує свої висновки: він не перевіряв, чи ключі робочі, і не демонстрував виходу за межі контейнера, хоча сам асистент у чаті таке стверджував. Про знахідку він повідомив через програму винагород Meta і архів не публікує.

Третя - Amazon. Ще до публікації Вордла Amazon почав блокувати покупки через Muse з повідомленням, що це «неавторизований AI-агент», який порушує умови користування. За даними The Rundown, Amazon закидає агенту, що той ходить магазином, не називаючи себе, і, схоже, зберігає логіни покупців; Meta ці звинувачення заперечує. Заява Amazon, яку цитує Ars, зводиться до того, що сторонні програми, які купують від імені людини, мають працювати відкрито і поважати рішення магазину, чи хоче він у цьому брати участь.

Чому це важливо. Особистий агент із широкими правами змінює саму модель загроз.

Раніше шкідливій програмі треба було окремо здобувати доступ до файлів, камери і пошти, а тепер достатньо захопити асистента, у якого все це вже є. Історія з експортом показує інший бік того самого: агент, якому можна сказати «запакуй усе, до чого маєш доступ, і надішли мені», виконує це без жодної зловмисності, просто тому що вміє. Звідси практичне правило для будь-якої команди, що будує агента: перелік того, що агент бачить у своєму середовищі, треба перевіряти так само ретельно, як перелік того, що він може робити. А суперечка з Amazon показує, що магазини, сервіси і платформи тепер вирішуватимуть, пускати агентів чи ні, і умови цього допуску ще ніхто не написав.

ТЕМА 4

GPT-6 Astra сама розшифрувала повідомлення Енігми 1941 року, яке не піддавалось з 2005-го

ДЖЕРЕЛА Crypto Cellar Research (Фроде Вейеруд), оновлено 19.09 Cryptocellar · обговорення 22.09 Hacker News

Дата важлива: сторінку оновили **19.09**, а широке обговорення, 580 балів на HN, почалось учора. Історію розповідає Фроде Вейеруд, криптоаналітик, який веде архів нерозшифрованих повідомлень німецької армії. 15 вересня до нього звернувся Картер Леффер з проханням перевірити розв'язок повідомлення **MVUEH від 10 липня 1941 року**, 82 літери, яке не піддавалось нікому з 2005 року. Вейеруд пише, що одразу побачив: ключ і текст правильні.

Цікаве тут те, як це зроблено. За словами Вейеруда, Леффер лише попросив GPT-6 Astra подивитись, чи вдасться зламати будь-яке з нерозшифрованих повідомлень з

сайту. Далі модель

сама обрала найперспективніше, помітила, що його текст може бути близьким до сусіднього повідомлення того ж дня, розшифрованого у 2017 році, взяла повтор назви місцевості «ROSENOW ROSENOW» як відому частину тексту, **написала на Python і C++ симулятор Енігми і програмну версію «Бомби»**, машини, якою розшифровували Енігму під час війни, і знайшла ключ. Він виявився зовсім іншим, ніж у решти повідомлень того дня. Розбір пояснює і чому цей шифр так довго не піддавався: у записі шифротексту були помилки, а ліве колесо машини провернулось на 72-й літері, що буває рідко і ускладнює злам.

Ще одна деталь з журналів моделі: вона знайшла згадку про архівні фонди в німецькому Бундесархіві і правильно назвала їхні шифри, яких на сайті Вейеруда немає. Звідки саме вона їх узяла, автор поки не з'ясував. Його оцінка: те, що модель зробила за два дні, людині забрало б тижні або місяці, а він сам кілька тижнів досліджував ті самі архівні фонди.

Чому це важливо. Це хороший приклад задачі, де результат перевіряється однозначно:

або ключ дає осмислений німецький текст, або ні. Тому тут немає суперечки про оцінку якості, яка супроводжує більшість заяв про «модель розв'язала». Показовий і сам порядок роботи: людина дала лише мету, а вибір цілі, гіпотезу, інструменти і перебір модель зробила сама. Для архівістів, істориків і всіх, хто сидить на великих масивах нерозібраних документів, це сигнал, що частину такої роботи вже можна віддавати агенту, лишаючи людині перевірку.

ТЕМА 5

Пентагон: застаріла розвідка і надмірна довіра до AI призвели до удару по школі в Мінабі

ДЖЕРЕЛА Bloomberg, 19.09 Bloomberg · обговорення 22.09 Hacker News

Матеріал Bloomberg вийшов **19.09**, до цього вікна, а обговорювати його почали вчора (468 балів на HN). У попередніх випусках цієї теми не було, тому вона тут.

Йдеться про удар 28 лютого двома ракетами Tomahawk по початковій школі в місті Мінаб на півдні Ірану. За даними Bloomberg, загинули понад 150 людей, серед них щонайменше 123 дитини. Видання називає це найсмертоноснішою помилкою американського наведення в XXI столітті за кількістю загиблих дітей. Офіційно США відповідальності не визнали, звіт про внутрішнє розслідування Пентагону не оприлюднений. Bloomberg спирається на розповіді посадовців, причетних до розслідування, які говорили анонімно.

За їхніми словами, катастрофу спричинило накопичення дрібних рішень, жодне з яких окремо не було вирішальним. Ділянку років тому занесли в базу як об'єкт Корпусу вартових ісламської революції, і запис не виправили, хоча на супутникових знімках

щонайменше з 2017-2018 років видно окремі стіни, футбольне поле і розмітку ігрового майданчика. У школи був власний сайт і позначка на Google Maps. Аналітик, який помітив зміни ще у 2019 році, записав це в систему, не пов'язану з основною базою для наведення. Команду, яка перевіряла цілі на ризик для цивільних, у Центральному командуванні скоротили з 10 людей до однієї, і школу вона не переглядала.

Окремо посадовці назвали **надмірну довіру до Maven Smart System**, програми Palantir, яка зводить понад 150 джерел даних і допомагає формувати списки цілей. Роботу, що раніше займала години, перед атакою стиснули до кількох хвилин, а частина персоналу чекала, що програма сама позначить застарілі дані. Чому вони так вважали, незрозуміло. Palantir відповідає, що не відповідає за вихідні дані і що доказів вини програми немає. Після удару в Maven, за словами джерела, додали повторну перевірку розвідданих на ознаки, які мали б виключити ціль.

Чому це важливо. Механізм тут знайомий з будь-якої автоматизації: інструмент пришвидшує процес, люди починають вважати, що він робить і ту перевірку, для якої його не створювали, а людей, які мали цю перевірку робити, водночас скорочують. Ніщо з цього не вимагає помилки самої програми. Урок переноситься на будь-який конвеєр, де рішення готує система: треба явно записати, які перевірки вона робить, а які лишаються на людях, і не прибирати людей з тих ділянок, які система не покриває.

ТЕМА 6

Британська Колумбія подала позов проти OpenAI через стрілянину в Тамблер-Ридж

ДЖЕРЕЛА Ars Technica, 22.09 Ars Technica · The Wall Street Journal WSJ · Financial Times FT

Про позов написали Ars Technica, WSJ і FT. 10 лютого 2026 року в містечку Тамблер-Ридж (2 700 мешканців) 18-річний стрілець убив п'ятьох дітей і працівницю школи, а перед тим - матір і зведеного брата. Сім'ї загиблих уже судяться з OpenAI. Тепер позов подала сама провінція. Її аргумент: компанія заздалегідь знала, що стрілець використовує ChatGPT для підготовки, і не попередила поліцію. За даними Ars, внутрішні команди перевірки рекомендували передати розмови поліції, і це рішення скасували; компанія пояснювала це захистом приватності.

Провінція вимагає, щоб OpenAI і особисто Сем Альтман оплатили нову школу (стару знесли в серпні, бо громада не могла туди повернутись) і інші витрати, і хоче, щоб суд зобов'язав ChatGPT автоматично обривати розмови з насильством. Найтерміновіша вимога - оприлюднити журнал розмов стрільця, який досі бачила лише поліція. Генеральна прокурорка провінції Нікі Шарма сказала, що в кримінальному праві «немає винятку для AI». OpenAI позов не коментує; у заяві для Ars компанія висловила співчуття і пообіцяла співпрацювати з владою.

Чому це важливо. Досі такі справи подавали родини, тепер позивачем стає держава, і це інша вага: у неї є прокурори, бюджет і право вимагати зміни продукту через суд.

Головне питання, яке тут вирішуватиметься, стосується не лише однієї трагедії: чи зобов'язаний оператор чат-бота повідомляти владу, коли бачить підготовку насильства, і що важить більше - приватність користувача чи попередження. Від відповіді залежатимуть правила для всіх, хто тримає розмови людей з моделями.

ТЕМА 7

Трамп в ООН: AI тепер «суперінтелект», а глобальний контроль над ним США відкидають

ДЖЕРЕЛА BBC, 23.09 BBC · The Guardian, 22.09 Guardian · Financial Times FT

20.09 тут згадувалась «AI Force», яку Трамп оголосив постом без жодних деталей.

Учора на Генасамблеї ООН він зробив дві заяви про AI. Перша - перейменування: на його думку, «штучний» звучить фальшиво, тому **відтепер у документах США має бути «super intelligence», SI**. За даними BBC, назву обрали через кілька опитувань серед його підписників у Truth Social. Експерти, яких опитало BBC, сумніваються, що це приживеться: суперінтелектом у галузі називають гіпотетичну систему, яка перевершує людину в усьому, і нинішнім моделям такий термін не пасує. Амджад Масад з Replit відреагував коротко: «Твій інтелект фейковий.

Мій - супер» (@amasad).

Друга заява суттєвіша. За цитатою Guardian, США **«повністю відкидають будь-яку спробу побудувати глобалістську схему контролю»** над AI і збираються його заохочувати, а не стримувати. Це прямо розходиться з планами британського прем'єра Енді Бернема, який того ж дня мав закликати до єдиних глобальних принципів і стандартів безпеки AI. Guardian також пише про ініціативу Фінляндії і Норвегії, яку підтримали понад 20 країн: не пускати моделі на ринок без попереднього тестування безпеки. Цього тижня Трамп має зустрітись із Сі Цзіньпіном, і AI, за даними BBC, буде серед тем.

Чому це важливо. Перейменування саме по собі - дрібниця, але воно показує, куди рухається мова: коли держава офіційно називає нинішні моделі «суперінтелектом», розмова про ризики справжнього суперінтелекту стає плутанішою. Суттєвіша друга заява: вона закріплює розкол, де США відмовляються від міжнародних правил, а Британія і група європейських країн рухаються до обов'язкового тестування перед виходом. Для компаній, які продають моделі в обидва світи, це означає різні вимоги на різних ринках.

ТЕМА 8

Microsoft закрила EvilTokens - сервіс з AI-чатботом, який читав зламані скриньки і планував шахрайство

ДЖЕРЕЛА Microsoft, 22.09 Microsoft · SpyCloud Spycloud · Ars Technica, 22.09 Ars Technica

EvilTokens продавали через Telegram з лютого 2026 року: **$1 500 за підключення і $500 щомісяця**.

За даними Microsoft, через сервіс зламали **понад 12 000 поштових скриньок у більш ніж 10 000 організацій**, найбільше у США, Канаді, Великій Британії, Австралії, Індії і Франції.

Microsoft разом з партнерами вилучила 50 сайтів і відключила ще понад 150 доменів, а лондонська поліція заарештувала двох чоловіків.

Злам був простим за механікою. Жертва отримувала лист, переходила за посиланням і бачила код із проханням ввести його на **справжній сторінці входу Microsoft**. Це легальний спосіб входу для телевізорів та інших пристроїв без клавіатури: вводиш код на іншому пристрої, і той отримує доступ. Людина нічого не віддавала, пароль лишався при ній, але вводила код, який зловмисник згенерував для свого пристрою. За словами Microsoft, такий доступ міг лишатись навіть після зміни пароля, якщо не відкликати сесії.

Новим тут є те, що було далі. Всередині сервісу працював чатбот, який читав зламану скриньку, знаходив розмови про оплати, визначав, хто в компанії «рухає гроші», кому довіряють і від чийого імені найкраще написати, і готував сам лист. Microsoft формулює це так: AI допомагав уже на етапі вибору, кого атакувати і як, переконливіші листи були лише останнім кроком. За даними розслідування, значну частину самої платформи її автори теж написали з допомогою AI.

Чому це важливо. Шахрайство з підміною листів завжди вимагало досвіду: розібратись у тисячах листів, зрозуміти, хто підписує платежі, вгадати тон. Тепер цей досвід продається як підписка, і поріг входу падає до ціни абонементу. Практичні висновки для будь-якої організації прості: вхід за кодом пристрою варто вимкнути там, де він не потрібен, після підозрілого входу відкликати всі сесії на додачу до зміни пароля, і перевіряти будь-яку зміну реквізитів окремим каналом.

ТЕМА 9

Apple прибрала вимикач Apple Intelligence у macOS 27, а в iOS з'явились банери, які не закрити

ДЖЕРЕЛА Девід Бушелл, 22.09 Dbushell · TechRadar Techradar · обговорення Hacker News · Hacker News

Дві окремі скарги, які разом зібрали на HN понад 1 400 балів. Веброзробник Девід Бушелл пише, що ще в лютому 2025 року вимкнув Apple Intelligence в macOS 15. Після оновлення до macOS 27 **окремого перемикача, щоб вимкнути Apple Intelligence, більше немає**, функції знову ввімкнені, а на диску вони займають **22,28 ГБ**. Частину меню можна приховати лише через обмеження «Екранного часу», створені для батьківського контролю. Його головна претензія стосується згоди: він уже сказав «ні», а після оновлення вибір просто зник.

Друга історія, від TechRadar: у налаштуваннях iOS зверху з'являються пропозиції сервісів Apple - iCloud+, Apple Music, Apple TV, AppleCare+. Частину з них не можна закрити, вони висять тижнями з позначкою на іконці, доки не мине строк або доки не заплатиш. Дехто бачить рекламу iCloud+, уже маючи підписку, тож TechRadar не виключає помилки.

Чому це важливо. Обидві історії про одне: налаштування за замовчуванням і право відмовитись. Функції AI тепер вбудовують на рівні системи, і вимикач стає окремим предметом суперечки, тому що кожна відмова зменшує аудиторію функції, в яку вклали гроші.

Для тих, хто робить продукти, висновок протилежний тому, що зробила Apple: якщо людина один раз сказала «ні», її рішення має пережити оновлення. Інакше скарга на одну функцію перетворюється на втрату довіри до всього продукту.

ТЕМА 10

«Спаймарки»: чому приховані водяні знаки AI пропонують назвати інакше

ДЖЕРЕЛА brand.io, 22.09 Brand · обговорення Hacker News

Текст на 653 бали пропонує нове слово. Звичайний водяний знак видно, і він підтверджує автентичність чи право власності. **«Спаймарк»** за визначенням автора - прихований сигнал, який робить твою роботу відстежуваною без твого знання і згоди. Головний приклад - Google SynthID, який вбудовує невидимі сигнали в зображення, звук, текст і відео. Автор посилається на статтю самої Google: варіант SynthID-O може закодувати **136 біт у зображенні 512 на 512 пікселів**, і цього вистачає на 64-бітний ідентифікатор запису в базі плюс корекцію помилок.

Важливо розрізняти, що тут доведено. Стаття показує **технічну можливість** вшити в картинку, звук чи вибір слів номер, що вказує на конкретного користувача. Доказів, що Google чи інші компанії так роблять, у тексті немає; це аргумент про ризик. Автор порівнює це з жовтими точками принтерів 1980-х, які кодують серійний номер пристрою, і протиставляє звичним метаданим на кшталт EXIF: ті видно, їх можна відредагувати і стерти, а сигнал у пікселях переживає і стирання метаданих, і частину правок.

Чому це важливо. Маркування AI-контенту зазвичай подають як захист від підробок, і для цього воно справді корисне. Текст звертає увагу на другий бік: та сама технологія може нести поруч із «це згенеровано» ще й «це згенерував ось цей акаунт». Для журналістів, викривачів і будь-кого, чия анонімність важлива, питання «що саме закодовано в знаку і хто може це прочитати» стає практичним. І саме тут бракує прозорості: зараз користувач не має способу перевірити, що вшито в його файл.

misc

- **Хакери заявляють, що мають дані всіх працівників ФБР** 404media (Hacker News) - угруповання ShinyHunters показало 404 Media вибірку з 5 000 записів з іменами, адресами і даними про подружжя. ФБР не підтверджувало.

  [одне джерело]
- **Чи може gzip бути мовною моделлю** Nathan (Hacker News) - 376 балів. Звичайний архіватор генерує щось схоже на Шекспіра: з кандидатів обирається той, що стискається найкраще. Гарна ілюстрація тези «передбачення і стиснення - одне й те саме».
- **Evals-навичка від Гемела Хусейна і Шреї Шанкар** GitHub (@lennysan) - Ленні Рачицький пише, що майже половина з 25 вакансій продакт-менеджерів, які він поширив минулого тижня, вимагає досвіду написання оцінок моделей.
- **Mistral відходить від передньої лінії** (@Afinetheorem) - економіст Кевін Браян, переказуючи журналістський матеріал, пише, що поза США і Китаєм ніхто по-справжньому не будує фронтирних моделей. [одне джерело, переказ]
- **OpenAI про незалежні оцінки** OpenAI (@OpenAI) - принципи, за якими компанія обіцяє давати зовнішнім оцінювачам глибокий доступ до навчання, оцінки і розгортання.
- **Академія Горовіца-Андрессена** (@TheRundownAI) - a16z запускає безкоштовну однорічну школу в Сан-Франциско для 18-річних, приблизно на 50 учнів; при вступі дивляться на те, що людина вже зробила, оцінки і есе важать менше.
- **SAML: фрактал поганого дизайну** Trailofbits (Hacker News) - Trail of Bits про те, чому в протоколі корпоративного входу постійно знаходять нові дірки.
- **Git 2.56 і 3.0** Lwn (Hacker News) - LWN про те, що зміниться в наступних версіях.