AI-ДАЙДЖЕСТ

Anthropic почала роуд-шоу IPO: до $78 за акцію

Також: поріг безпеки Anthropic, Gemini 4 Carbon, доведення коду Signal, 217 виправлень з 37 966 знахідок Glasswing.

середа, 11 листопада 2026

У ЦЬОМУ ВИПУСКУ

1. Anthropic почала роуд-шоу IPO: до $78 за акцію, оцінка до $1,4 трлн
2. Anthropic: AI вже вдвічі пришвидшує наші дослідження, поріг безпеки спрацював
3. Google випустила Gemini 4 Carbon
4. AlphaProof Code: агенти формально довели безпеку коду протоколу Signal
5. Glasswing за сім місяців: 37 966 знахідок, 217 виправлень
6. Amazon замінила Sonnet 4.6 на модель рішень і платить на 98% менше
7. Claude отримав Whiteboards, Sheets і Data
8. Nvidia найняла Wachtell для розслідування щодо Хуанга
9. Ленні: вакансії інженерів зупинились, PM і FDE ростуть
10. Злам Fireworks AI зачепив чати Notion AI

ТЕМА 1

Anthropic почала роуд-шоу IPO: до $78 за акцію, оцінка до $1,4 трлн

ДЖЕРЕЛА Anthropic, 10.11 Anthropic · @AnthropicAI, 10.11 @AnthropicAI · CNBC, 10.11 Cnbc · The Guardian, 10.11 Guardian · Semafor, 10.11 Semafor

Anthropic опублікувала умови розміщення. Компанія продає 825 млн акцій класу A за ціною від $70 до $78. З них 801,4 млн продає сама компанія, решту продають наявні акціонери. Біржа Nasdaq, тікер ANT. Андеррайтери мають опціон ще на 123,75 млн акцій.

CNBC з посиланням на джерела пише, що ціну визначать 18.11, а торги почнуться 19.11. За верхньою межею компанія отримає до $64,35 млрд і коштуватиме до $1,405 трлн з урахуванням усіх акцій. У травні раунд оцінив її в $965 мільярдів. Ще $26 млрд дадуть AMD, Broadcom, Google, Micron, Microsoft, Nvidia і TSMC через приватне розміщення за ціною IPO. Частка для роздрібних інвесторів близько 20%, звичайна норма 5-10%.

The Guardian цитує оновлений проспект: річна виручка на 31.10 становить $81 млрд, на $10 млрд більше, ніж місяцем раніше. Зобов'язання з купівлі обчислень перевищують $70 млрд на наступні 12 місяців і $450 млрд до 2030 року.

Ще місяць тому ринок чекав оцінку $2 трлн. Аналітик Forrester пояснює зниження реакцією інвесторів на проміжні вибори, де перемогли противники дата-центрів. Goldman Sachs, який веде розміщення разом з Morgan Stanley, прогнозує виручку компанії понад $600 млрд у 2030 році. Вчора WSJ повідомив, що OpenAI у третьому кварталі мала виручку $10 млрд і великий операційний збиток. Anthropic за той самий квартал показала $17 млрд виручки [одне джерело: WSJ, лише перший абзац за пейволом].

Чому це важливо: це найбільше залучення грошей на біржі в історії, разом з приватним розміщенням і опціоном близько $100 мільярдів. Для клієнтів важливий інший рядок проспекту. Обчислення на $450 млрд до 2030 року компанія має окупити виручкою. Тому ціни на API і швидкість нових релізів відтепер залежать і від квартальних звітів.

ТЕМА 2

Anthropic: AI вже вдвічі пришвидшує наші дослідження, поріг безпеки спрацював

ДЖЕРЕЛА Yahoo Finance, 10.11 Yahoo · AI Futures Project, 10.11 Aifutures · @rowancheung, 10.11 @rowancheung · HN, 10.11 Hacker News

Вчора згадували звіт METR, за яким Anthropic покращує моделі приблизно вдвічі швидше, ніж раніше. Сьогодні те саме визнала сама компанія. Yahoo Finance цитує оновлений проспект: «We made a determination that we have met this threshold». Поріг записаний у Responsible Scaling Policy. Він спрацьовує, коли AI робить таку частку досліджень, що темп прогресу приблизно подвоюється: два роки прогресу за темпом 2018-2024 років вкладаються в один рік.

Компанія пише, що рішення спирається на очікування від майбутніх моделей і загальний темп. На поведінку поточних моделей компанія не посилається. Наслідок за політикою - додаткові запобіжники, зупинки там нема. У проспекті також сказано, що Anthropic не очікує безпечно доручити Claude повністю автономно розробляти наступні покоління моделей без нагляду людей. METR окремо назвав серпневий аргумент компанії про те, що поріг ще не досягнуто, «overstated».

Того ж дня AI Futures Project, автори сценарію AI 2027, скоротили свої прогнози. Тепер усі шість авторів очікують «надлюдського AI-дослідника» у 2027 році, якщо розробку не сповільнять свідомо. Головна причина - той самий звіт METR. Автори оцінюють прискорення, яке дає саме AI, в 1,7 раза. Їхня ж модель три місяці тому давала такому темпу в листопаді 2026 року менше 4% ймовірності. Ймовірність AI, що автоматизує всю офісну роботу до 2028 року, вони оцінюють у 35-65%. Автори прямо пишуть, що текст готували поспіхом і в моделюванні можуть бути помилки.

Чому це важливо: до цього тижня тезу «AI прискорює розробку AI» обговорювали як прогноз. Тепер її записала у документ для інвесторів сама лабораторія, а незалежний

оцінювач дав подібну цифру. Варто стежити, які саме запобіжники Anthropic додасть за політикою. Їх у проспекті поки не перелічено.

ТЕМА 3

Google випустила Gemini 4 Carbon

ДЖЕРЕЛА Google, 10.11 Blog · @GoogleDeepMind, 10.11 @GoogleDeepMind · Artificial Analysis, 10.11 Artificialanalysis · @emollick, 10.11 @emollick · HN, 10.11 Hacker News

Google відкрила Carbon для розробників через API, для бізнесу і для підписників Google AI Ultra. Два тижні модель мали лише захисники з програми Fairwind. Google позиціонує Carbon для довгих і складних задач. Для більшості роботи компанія радить попередню модель Argon. Carbon витрачає більше токенів, щоб довше перевіряти результат. Вступна ціна така сама, як в Argon: $2 за мільйон вхідних токенів і $10 за мільйон вихідних.

За даними Google, Carbon перша на CursorBench 4.0, WANDR від Perplexity і AutomationBench від Zapier. Вартість задачі там більш ніж удвічі нижча, ніж у наступної зовнішньої моделі. Разом з моделлю Google показала /swarm в Antigravity. Один промпт запускає команду агентів, де Carbon бере архітектуру і аудит, а легші моделі пишуть код.

Частина про кібербезпеку найконкретніша. У системі Project Fortify модель знайшла 100% вразливостей при скануванні внутрішнього набору і мала 95% точності при сортуванні. Команда Wiz за дві години знайшла 0-day у двох великих SaaS-сервісах. Google пише, що запобіжники Carbon заблокували 99% атак червоної команди в симуляції.

На тесті чесності агентів на 375 задачах модель пройшла 92% проти 70% в Argon.

Artificial Analysis дає Carbon 60 балів у своєму індексі. Учора Fable 5.2 отримала там 65. Одна задача індексу коштувала $7,76, модель витратила 205 млн токенів при медіані 72,8 млн. AA рахує за цінами $4 і $20 за мільйон, тобто вдвічі вище за вступну ціну Google. Ітан Моллік мав ранній доступ і пише, що модель «мислить інакше, ніж Claude і GPT», і саме в цьому її користь.

Чому це важливо: за два дні вийшли дві фронтирні моделі з однаковою логікою. Дорога модель для довгих задач стоїть над дешевшою моделлю для щоденної роботи. Fable 5.2 веде в індексі на 5 балів, Carbon дешевша за вступною ціною. Для команди вибір тепер залежить від задачі і бюджету на неї.

ТЕМА 4

AlphaProof Code: агенти формально довели безпеку коду протоколу Signal

ДЖЕРЕЛА Google DeepMind, 10.11 DeepMind · HN, 10.11 Hacker News

DeepMind представила AlphaProof Code, систему агентів для формальної верифікації. Люди разом з моделями переклали цілі безпеки Signal у математичні специфікації і перевірили кожну. Далі агенти на Gemini Carbon написали доведення: понад 6 млн рядків мовою Lean. Агенти самі позначили для людської перевірки близько 1% коду.

Решту автоматично перевіряє Lean.

Доведення стосуються скомпільованого коду, який працює на пристроях. Більшість попередніх робіт перевіряли лише дизайн протоколу. Робота забрала чотири тижні. DeepMind пише, що раніше таке займало в експертних команд роки. Для контролю дослідники повернули в окрему копію коду два старі виправлені баги. В обох випадках система видала перевірене доведення, що код небезпечний. Технічний директор Signal Ерен Крет назвав переклад живого коду в теореми Lean «major breakthrough».

DeepMind обіцяє відкрити артефакти верифікації.

Чому це важливо: формальна верифікація досі була надто дорогою для більшості проєктів. Якщо агенти скорочують роботу з років до тижнів, перевірене доведення стає доступним для критичного коду: криптографії, прошивок, систем керування. Пункт 5 показує інший бік тієї ж теми: знайти баг дешевше, ніж довести, що він справжній.

ТЕМА 5

Glasswing за сім місяців: 37 966 знахідок, 217 виправлень

ДЖЕРЕЛА VulnCheck, 09.11 Vulncheck · HN, 10.11 Hacker News

Патрік Ґарріті з VulnCheck веде облік вразливостей, які знайшла програма Anthropic Glasswing. Його підсумок за сім місяців: 442 CVE, тобто близько 0,5% усіх CVE цього року. У реальних атаках використали лише дві з них.

У публічному реєстрі Anthropic на 02.11 стоїть 37 966 знахідок. Опубліковано 7 349, виправлено 217. З вересня додалось 11 813 знахідок і лише 15 виправлень. Дублікатів стало 1 670, у вісім разів більше, ніж виправлень. При цьому реєстр показує частку справжніх знахідок 94,2%, на запуску було 90,8%.

Найнаочніший приклад - curl. Проєкт отримав від сканера Anthropic 45 звітів. З 34 опублікованих мейнтейнери визнали 19 «не вразливістю», 6 підтвердили з низькою критичністю, 9 ще чекають. Тобто 56% хибних серед перевірених. Сама Anthropic у жовтні писала, що вручну перевірила близько 6 000 з понад 29 000 кандидатів і що її «bottlenecked on our human capacity to validate». Новий сканер OSS Scanner прямо попереджає, що звіти згенеровані моделлю без людської перевірки.

Чому це важливо: цифри показують, де вузьке місце. Модель знаходить кандидатів швидко, а перевірка лягає на мейнтейнерів, часто волонтерів. Частку справжніх знахідок варто міряти рішенням мейнтейнера. Оцінка самого сканера тут мало що каже. Автор - конкурент на ринку даних про вразливості, тому окремо варто дочекатись відповіді Anthropic.

ТЕМА 6

Amazon замінила Sonnet 4.6 на модель рішень і платить на 98% менше

ДЖЕРЕЛА Strands Agents, 09.11 Strandsagents · Hugging Face, 10.11 Hugging Face · HN, 10.11 Hacker News

Двоє інженерів Amazon описали, як розмічали тікети внутрішньої системи. Раніше Claude Sonnet 4.6 з довгим промптом обирав одну з 20 категорій і писав JSON. Тепер це робить Strands Decider E4B. Модель рішень не генерує текст. Вона отримує стан, питання і варіанти відповіді. Результат - ймовірність кожного варіанта.

Команда спершу запустила обидві моделі паралельно на тих самих тікетах. Потім Decider забрав усю роботу.

Час виклику впав з 4 215 мс до 897 мс. Токенів на тікет стало 2 010 замість 6 724. Вартість на мільйон тікетів за публічними цінами: близько $22 200 у Sonnet і близько $500 у Decider на SageMaker. Окрема знахідка: порядок ключів у JSON змінював ймовірності на величину до 0,19, тому тепер ключі сортують.

Hugging Face того ж дня опублікувала огляд моделей рішень з підтримкою в transformers, TRL і llama.cpp.

Застереження: допис написала команда Strands про власний продукт, незалежного відтворення ще нема.

Чому це важливо: багато продакшн-задач з LLM - це класифікація з фіксованим списком відповідей. Для них велика модель переплачує за генерацію тексту, який ніхто не читає. Корисний патерн з допису - перемикач на три режими і паралельний прогін на реальному трафіку перед заміною.

ТЕМА 7

Claude отримав Whiteboards, Sheets і Data

ДЖЕРЕЛА Claude, 10.11 Anthropic · @TheRundownAI, 10.11 @TheRundownAI · @levie, 10.11 @levie · HN, 10.11 Hacker News

Anthropic запустила в беті дошки Whiteboards і таблиці Sheets, а Claude Dashboards став загальнодоступним під назвою Claude Data. Дошка перетворює розмову на план з відповідальними і датами, команда редагує її разом з Claude. Sheets будує таблиці з вкладками, формулами і списками, кожну правку Claude позначає. Anthropic прямо пише, що функції таблиць поки прості, і для повного Excel радить експорт у формат xlsx або CSV.

Claude Data отримав сертифіковані запити для метрик, кешування дашбордів і історію змін. Інтерактивні конектори тепер відкривають Adobe, Atlassian, Box і Notion як повноцінні застосунки поруч з розмовою. Аарон Леві з Box пише про повний доступ на

читання і запис до файлів Box. Усе доступне на платних тарифах. На Team функції увімкнені одразу, на Enterprise дошки і таблиці вимкнені, а Data вмикається 19.11.

Чому це важливо: результат роботи з Claude тепер лишається в Claude, де команда може його редагувати. Для адміністраторів Enterprise важлива дата 19.11: Claude Data увімкнеться сам, якщо його не вимкнути в налаштуваннях.

ТЕМА 8

Nvidia найняла Wachtell для розслідування щодо Хуанга

ДЖЕРЕЛА BBC, 10.11 BBC · @TheRundownAI, 10.11 @TheRundownAI · Semafor, 10.11 Semafor

Вчора згадували, що Трамп підтвердив розслідування ФБР, а акції Nvidia впали. Сьогодні компанія найняла нью-йоркську фірму Wachtell, Lipton, Rosen & Katz. Розслідування очолить колишній федеральний прокурор Південного округу Нью-Йорка Джун Кім. Процес наглядає комітет незалежних директорів.

BBC пише, що з п'ятниці Nvidia втратила близько $500 млрд ринкової вартості. Комітет Палати представників із закордонних справ запросив Хуанга на закритий брифінг наступного тижня. Комітет може його викликати повісткою.

Хуанг ще не відповів, чи прийде. Пекін назвав звинувачення спробою «smear and malign China». Semafor при цьому пише, що решта ринку на скандал майже не реагує.

Чому це важливо: питання вже вийшло за межі однієї компанії. Тепер це тема для Конгресу і для правил експорту чипів. Важлива дата наступного тижня: чи прийде Хуанг на брифінг. Той самий тиждень Anthropic виходить на біржу, а Nvidia купує її акції на $10 млрд.

ТЕМА 9

Ленні: вакансії інженерів зупинились, PM і FDE ростуть

ДЖЕРЕЛА Lenny's Newsletter, 10.11 Lennysnewsletter · @lennysan, 10.11 @lennysan · @lennysan, 10.11 @lennysan

Ленні Рачицький опублікував п'ятий звіт про ринок праці в техі. Дані з двох джерел: TrueUp рахує вакансії в понад 9 000 компаній, Live Data Technologies рахує фактичні наймання в США.

Відкриті інженерні вакансії виросли на 29% у 2024 році, на 15% у 2025-му і лише на 2,5% цього року. Компанії США за 12 місяців найняли близько 113 000 інженерів, менше половини від піку 2022 року (239 000). Вакансії мобільних і фронтенд-інженерів майже зменшились удвічі за рік. Вакансії forward deployed engineer виросли в 4,6 раза, зараз їх близько 1 500. Вакансій PM найбільше майже за чотири роки. Дизайнерів шукають найменше за всю історію звіту: близько 1 900 вакансій, на 28% менше, ніж рік тому. Вакансії рекрутерів виросли майже на 300% з середини 2023 року.

Чому це важливо: дані показують перехід від вузьких спеціалізацій до широкої відповідальності малих команд.

Сам автор пояснює це AI. Ріст FDE частково пояснюється перейменуванням ролей sales engineer, це він теж зазначає.

ТЕМА 10

Злам Fireworks AI зачепив чати Notion AI

ДЖЕРЕЛА Notion, 10.11 Notion · Fireworks AI, 09.11 Fireworks · HN, 10.11 Hacker News

Notion повідомила, що злам постачальника Fireworks AI зачепив менше 1% її робочих просторів. Витекли тексти чатів Notion AI, які компанія використовувала для функції Insights. Notion пише, що чати не мали підпису автора чи простору. Власники просторів можуть завантажити CSV з ураженими даними і окремий аналіз з позначками можливих секретів.

Fireworks оновила свій бюлетень від 09.10. Слідство пов'язує атаку з фінансово мотивованою групою, що перепродає доступ до LLM. Цей тип атак називають LLMJacking. 23.10 зламали сторонній хмарний провайдер з машинами Fireworks, і 24.10 атакер дістався до машин розробки. Крім секретів, витекли білінг і для частини клієнтів датасети і незавершені ваги моделей. Fireworks пише, що інференс з нульовим зберіганням даних не постраждав. У коментарях на HN клієнт розповів, що спершу отримав лист «вас не зачепило», а потім виявив, що його ключі API в Fireworks хтось використав.

Чому це важливо: нульове зберігання даних захищає лише той шлях, де воно ввімкнене. Пакетна обробка і датасети для навчання лежать на диску постачальника. Будь-який ключ, вставлений у чат з AI, варто ротувати як скомпрометований.

---

misc

- **Businesses in ChatGPT:** OpenAI запустила бету, де бізнес підтверджує профіль, оновлює ціни і бачить, як часто ChatGPT про нього говорить. Магазини США можуть запустити власного агента для покупців. OpenAI пише про 1,2 млрд користувачів на тиждень. OpenAI · @OpenAI
- **Gemini Robotics 2.5:** DeepMind показала нову модель для роботів. У пілоті з Apptronik і Mercedes-Benz робот сортував деталі з точністю 96%. Ars Technica зазначає, що модель працює в хмарі, а демо-відео пришвидшене вдвічі. DeepMind · Ars Technica
- **Stripe і OpenRouter:** угода закрита, OpenRouter тепер частина Stripe. @patrickc пише, що з серпня, коли оголосили угоду, OpenRouter обслуговує на ~125% більше токенів на тиждень. @stripe · @patrickc
- **Val Town закривається:** реєстрація закрита з 10.11, режим лише для читання з 10.12, повне вимкнення 10.01.2027, код відкриють 01.02.2027. Засновник пише, що компанія

не витримала конкуренції в хмарних обчисленнях. Val

- **Apple Harbor:** 9to5Mac з посиланням на інсайдера пише про застосунок, що об'єднує кілька Mac через Thunderbolt 5 у локальний сервер для відкритих моделей. Apple нічого не анонсувала. 9to5mac [одне джерело]
- **Код після AI-мандату:** колишній інженер Meta описує чотири способи, як AI робить код крихким: поверхові тести, зайва складність, люди перестають розуміти код, рев'ю не встигає. 421 бал і 431 коментар на HN.

  Evanxmerz · HN Hacker News
- **Моллік і агенти:** @emollick дав Fable 5.2 абзац тексту, і команда агентів зробила гру Dread & District про логістику проти зла. @emollick

---