

Anthropic показує Конгресу свій закон про безпеку AI

Витік ключів в OpenCode, ліміт Starshield в Україні і нова відкрита модель DeepSeek

неділя, 18 жовтня 2026

У ЦЬОМУ ВИПУСКУ

1. Anthropic показує конгресменам свій законопроект про безпеку AI
2. OpenCode: зловмисники отримали ключі OpenAI і Anthropic користувачів
3. Зеленський: обмеження Starshield в Україні – спланована операція
4. DeepSeek виклала V4.1 Flash 1017 під MIT і відкрила свій пошук
5. Anthropic платить Blackstone понад 10% за борг під чипи Google
6. Replit обрала OpenAI головним партнером за ціною і якістю
7. Amazon оцінює ризик профспілок на складах внутрішнім AI-інструментом
8. Шоле і Леві сперечаються, чи лишиться робота для програмістів

ТЕМА 1

Anthropic показує конгресменам свій законопроект про безпеку AI

ДЖЕРЕЛА POLITICO, 17.10 [Politico](#) · Цві Мовшовиц, Substack, 17.10 [Substack](#)

Politico отримала конфіденційну чернетку закону. За даними трьох джерел, представники Anthropic тихо показують її в Конгресі. Компанія підтвердила суть. Речник сказав, що закон повторює її рамку Advanced AI Framework, яку Anthropic просуває з червня.

Головне в тексті – коли уряд може зупинити модель. Для цього потрібен передбачуваний і неминучий ризик шкоди понад \$1 млрд або загибелі чи тяжких травм понад 50 людей. Ризик має походити з біологічної зброї, кібератаки або виходу моделі з-під контролю розробника. Запит на зупинку подає лише Міністерство торгівлі. Рішення ухвалює федеральний суд округу Колумбія.

Жодне інше відомство не може призупинити розробку, запуск чи внутрішнє використання моделі.

Лабораторіям закон дає обов'язки. Вони пишуть рамки безпеки, публікують звіти прозорості для кожної нової моделі, захищають викривачів і повідомляють про критичні інциденти. Незалежні оцінювачі реєструються в Міністерстві торгівлі.

Лабораторія може їм платити, якщо оплата не залежить від результату. Про «неминучий ризик» оцінювач повідомляє за 72 години. В обмін закон забирає у штатів право регулювати розкриття ризиків, сторонні оцінки і звіти про інциденти.

Реакція з боку прихильників безпеки різка. Мовшовиц назвав чернетку «неприйнятною». На його думку, вона слабша за двопартійний FRONTIER Act, який з'явився влітку. Чарлі Буллок помітив, що закон майже не дає права на нормотворчість і не ліцензує оцінювачів. Без ліцензій оцінювачі почнуть змагатися за найлегший аудит. Натан Калвін пише, що закон забирає у судів штатів навіть загальне право зупинити модель. Келсі Пайпер пояснює мотив досвідом самої компанії.

Влітку уряд вимкнув дві моделі Anthropic експортним контролем після кількох телефонних дзвінків.

Чому це важливо. Це перший відомий законопроект, який фронтірна лабораторія написала сама.

Компанія, яка найгучніше говорить про ризики AI, пропонує поріг для втручання держави вище за теперішній. Для читача, що стежить за регулюванням, тут два сигнали. Право федерального закону скасовувати закони штатів стає головною розмінною монетою. І лаби бояться довільного рішення виконавчої влади більше, ніж слабких правил.

ТЕМА 2

OpenCode: зловмисники отримали ключі OpenAI і Anthropic користувачів

ДЖЕРЕЛА OpenCode в X, 17.10 @opencode · обговорення на HN, 17.10 Hacker News

Увечері 17.10 OpenCode повідомила про злам. Компанія назвала його «масштабною атакою дистиляції». Постраждали 1,4% користувачів сервісу OpenCode Console. Інцидент триває, і компанія разом з Anthropic, OpenAI і Google прибирає наслідки.

Лист користувачам точніший за пост. Зловмисники отримали доступ до старої системи, яку OpenCode якраз виводила з роботи. У цій системі лежали ключі сторонніх провайдерів, які користувачі додали у свій workspace. OpenCode просить відкликати ці ключі в кабінетах провайдерів і перевірити недавні витрати. Під постом користувачі пишуть, що з їхніх рахунків за кілька хвилин зникло \$100-\$138.

На HN найвищий коментар пояснює плутанину з назвою. OpenCode Console - це хостинговий шлюз до моделей, схожий на OpenRouter. Він не пов'язаний з CLI-агентом OpenCode і з десктопною апкою.

Якщо людина запускає лише агента зі своїми ключами локально, її ключі на сервер не йшли.

Кілька коментаторів сумніваються в слові «дистиляція». Доказів, що ключі пішли саме на збір відповідей для навчання чужої моделі, компанія поки не показала.

Чому це важливо. Сервіс, який зберігає чужі API-ключі, концентрує ризик. Один злам дає доступ до рахунків тисяч людей одразу. Ключ провайдера варто тримати там, де він потрібен, з лімітом витрат і з окремим ключем на кожен сервіс. Тоді витік одного сервісу коштує одного ключа.

ТЕМА 3

Зеленський: обмеження Starshield в Україні – спланована операція

ДЖЕРЕЛА Financial Times, 17.10 [FT](#) · Українська правда, 17.10 [Com](#) · Politico Europe [Politico](#) · Kyiv Independent [KyivIndependent](#) · Kyiv Independent про звіт Telegraph [KyivIndependent](#)

14.10 термінали Starshield, військової версії Starlink, почали втрачати зв'язок у русі.

Попередження українська армія не отримала. Командир підрозділу ударних дронів розповів Politico Europe про ліміт 100 км/год на всіх терміналах. Його підрозділ зупинив усі удари середньої дальності. «Це порівнянно з ядерним ударом по нас», – сказав він. Постраждали і дрони-перехоплювачі, які захищають міста від російських дронів.

Через день збої зникли, але ліміт 160 км/год лишився. Військові кажуть Kyiv Independent, що такої швидкості перехоплювачу мало. 17.10 Зеленський сказав, що Україна перевірила термінали в ЄС. Там вони працювали без обмежень. Тому він назвав ліміт «спланованою операцією».

Відповідальності ніхто не бере. Маск пише, що Starshield належить уряду США, і ліміти ставить армія, а зрештою президент. The Telegraph з посиланням на чотири джерела пише, що ідею запропонував сам Маск, а Трамп погодився. Трамп 16.10 сказав, що це не він і, на його думку, не Маск. Пентагон мовчить. Сенатор-республіканець Том Тіллс просить генерального інспектора Міноборони про розслідування.

Чому це важливо. Зв'язок для дронів – тепер частина обороноздатності. Цей випадок показує, що постачальник може змінити його одним налаштуванням, без попередження. Зеленський каже, що Європа зробила висновок і вже працює над альтернативою Starlink.

ТЕМА 4

DeepSeek виклала V4.1 Flash 1017 під MIT і відкрила свій пошук

ДЖЕРЕЛА DeepSeek, новини API, 17.10 [Deepseek](#) · Hugging Face, картка моделі [Hugging Face](#) · обговорення на HN, 17.10 [Hacker News](#)

DeepSeek називає реліз малим оновленням V4.1 Flash, яка вийшла 10.09. Ваги лежать на Hugging Face під ліцензією MIT. На сторінці моделі вказано 763 млрд параметрів і контекст 1 млн токенів.

Компанія обіцяє кращі результати, швидше виконання задач і коротші відповіді.

Таблиця в картці моделі цікава порівнянням з закритими моделями. На DeepSWE v1.1 нова модель має 77,0. GPT-5.6 Sol має 73,0, Opus 5 має 74,0. На Automation-Bench вона має 60,5 проти 50,3 в Opus 5. На Terminal-Bench 4.0 Opus 5 лишається попереду: 51,8 проти 42,3. Усі цифри дала сама DeepSeek. Terminal-Bench вона ганяла в об'язці Claude Code.

Друга новина того ж посту - пошук. Search API і Fetch API тепер відкриті для всіх, по \$0,75 за 1000 викликів. Пошук працює на власному пошуковому рушії DeepSeek. Через серверні тули в Claude Code і Codex виклики безкоштовні до 31.10.

На HN цю модель хвалять за передбачуваність. Один коментатор пише, що вона генерує код, який написав би розробник, і не намагається бути надто розумною. Інший пише, що регулярно ганяє рев'ю на п'яти відкритих моделях, і DeepSeek справляється найкраще.

Чому це важливо. Відкрита модель знову наздогнала минулорічних лідерів на задачах агента.

Для команди, яка рахує витрати на токени, це вагомий аргумент. Але цифри поки лише від виробника, тому перед заміною моделі варто прогнати власний набір задач.

ТЕМА 5

Anthropic платить Blackstone понад 10% за борг під чипи Google

ДЖЕРЕЛА Financial Times, 16.10 [FT](#) · Reuters, 17.10 [Reuters](#) · Mint, 17.10 [Livemint](#)

Учора тут писали про рейтинг BB+ і бачили лише заголовок про Blackstone. Тепер є деталі. За даними FT, Anthropic збирає \$60 млрд у пакеті Project North Star. Гроші підуть на близько 1,5 ГВт тензорних процесорів Google. Близько \$42 млрд старшого боргу гарантує Broadcom.

Фонди Blackstone мають вкласти \$9 млрд у молодший борг, забезпечений платежами Anthropic за п'ятирічну оренду чипів. Це половина молодшої частини, яка має \$18 млрд загалом. Купон - 10%, а з дисконтом і комісіями дохідність ще вища. FT порівнює: позичальник з таким рейтингом на публічному ринку платить близько 7%. Чотири місяці тому схожий борг Anthropic коштував 8,5%.

Чому це важливо. Ціна грошей для AI-лаби росте швидше, ніж росте сама лаба. Різниця між 8,5% і понад 10% за чотири місяці показує, що інвестори хочуть більшої премії за ризик. Цей ризик пов'язаний з обсягом витрат на інфраструктуру. Reuters окремо пише, що не змогла перевірити звіт FT.

ТЕМА 6

Replit обрала OpenAI головним партнером за ціною і якістю

ДЖЕРЕЛА Амджад Масад (@amasad), 17.10 [@amasad](#) · Семюел Шпіц (@samuel_spitz), 17.10 [@samuel_spitz](#) · TechCrunch, 16.10 [Techcrunch](#)

На TechCrunch Disrupt CEO Replit Амджад Масад сказав, що OpenAI тепер «бажаний партнер» компанії.

Після хвили питань він уточнив у X. Replit лишається незалежною від моделей, і користувач може вибрати Claude. Але основу агента Replit тепер будує на моделях OpenAI. Причини, за його словами,

- продуктивність, ціна і партнерство. Семюел Шпіц з Replit додав, що на більшості задач моделі OpenAI дають найкраще співвідношення ціни і результату, навіть проти відкритих моделей.

На тій самій сцені Масад описав «компанію на автопілоті». Звернення в підтримку спершу розбирає агент: дивиться issues на GitHub, код і тікети в Linear. Людина отримує готовий контекст і вирішує, що робити. Патч теж пише агент за один клік. Окремо Replit запустила Bring Your Own Cloud: платформу можна розгорнути у власному Azure, Google Cloud або AWS.

Чому це важливо. Великі платформи для програмування тепер вибирають модель за ціною на задачу, і лояльність до однієї лаби слабне. Для користувача це добре: конкуренція між лабами переходить у ціну. Для лаб це сигнал, що перевага на бенчмарку без вигідної ціни не тримає клієнта.

ТЕМА 7

Amazon оцінює ризик профспілок на складах внутрішнім AI-інструментом

ДЖЕРЕЛА openDemocracy, 16.10 Opendemocracy · обговорення на HN, 17.10 Hacker News

Документи з судової справи Amazon проти канадської профспілки описують інструмент Atlas. Це внутрішня AI-модель прогнозу. Вона оцінює, наскільки ймовірно працівники складу створять профспілку, влаштують сидячий страйк або іншу «руйнівну трудову подію». Документи схожі на навчальні матеріали для менеджерів у Канаді.

Atlas збирає «сотні точок даних». Серед них пости на Reddit, відповіді працівників у внутрішніх опитуваннях і історія попередніх акцій. Кожен склад отримує бал від 100 до 900. Що нижчий бал, то вищий ризик профспілки. Для керівництва вищий бал – «краще». Склади поділені на п'ять рівнів.

Найвищий рівень ризику мають 16 складів у США, Канаді і Великій Британії, з них два британські.

Amazon заперечує стеження за окремими людьми. За словами речника, Atlas працює лише на рівні складу і допомагає зрозуміти, де працівникам потрібна підтримка.

Чому це важливо. Той самий тип моделі, який прогнозує попит, легко перевести на прогноз поведінки людей. Дані тут звичайні: опитування, форуми, історія інцидентів.

Межа між аналітикою і стеженням проходить через мету використання, і саме мету документи показують прямо.

ТЕМА 8

Шоле і Леві сперечаються, чи лишиться робота для програмістів

ДЖЕРЕЛА Франсуа Шоле (@fchollet), 17.10 @fchollet · Аарон Леві (@levie), 17.10 @levie

Франсуа Шоле перелічив спостереження про «повну» автоматизацію програмування. Програмістів зараз більше, ніж у 2025 році, і попит росте. Працюють вони інтенсивніше, ніж будь-коли. І код вони майже не пишуть: агенти пишуть близько 100%. Четверте спостереження найгостріше. Прибуток компаній, де працюють ці інженери і агенти, від AI зазвичай не росте.

Шоле пояснює це системно: автоматизація однієї ланки переносить вузьке місце в іншу.

Аарон Леві з Vox розвинув думку. На його погляд, головна помилка прогнозів була в припущенні, що світу потрібна фіксована кількість софту. Агенти вміють працювати з комп'ютером і обробляти дані, тому софтом стають задачі, які раніше не цифровувались. Банк, лабораторія і завод тепер мають задачі, схожі на інженерні.

У відповідях Леві отримав головне заперечення. Попит на розробку софту і попит на людей-розробників – дві різні змінні. Попит на зерно може вибухнути, і кількість фермерів від цього не зростає.

Чому це важливо. Суперечка зсунулась. Ніхто вже не питає, чи пишуть агенти код. Питання тепер, хто перевіряє і направляє агентів, і скільки таких людей треба. Для інженера практичний висновок простий: цінність зміщується від написання коду до постановки задачі і перевірки.

misc

- **Білий дім** запустив сайт si.gov про «Super Intelligence». Там зібрані Genesis Mission на 15 федеральних агенцій, прискорені дозволи на дата-центри і програма експорту американського стеку. [Si](#)
- **Майкл Селліто (Anthropic):** акаунт @MTSlive пише, що голова глобальних справ Anthropic сказав МВФ таке: Claude може вести 80-90% навчання фронтірних моделей наступного року, проти 3% у квітні. [одне джерело] Інші джерела цю цифру поки не підтвердили. [@levie](#)
- **Ендрю Бейлі**, голова Банку Англії, сказав на G30, що проблеми AI-агентів «треба вирішувати».

Це продовження вчорашнього пункту 10: великі банки Wall Street уже обмежують агентам доступ до рахунків клієнтів. [FT](#)

- **Ядро Linux 7.2:** LWN порахувала 16 418 комітів від 2 652 розробників, це рекорд за кількістю авторів. Серед головних внесків – довгий список виправлень багів, які написала LLM. Частка комітів з тегом Reviewed-by падає, зараз 48%. [Lwn](#)
- **Codex Desktop** на Mac безперервно пише дампи Crashpad. У автора issue папка виросла до 4,9 ГБ і 54 504 файлів за день, хоча апка працює нормально. [GitHub](#)
- **Palantir:** Guardian через запити на інформацію з'ясувала, що багато трастів NHS в Англії відмовляються від платформи даних Palantir за £330 млн, і частина називає її «дублюючою».
[Guardian](#)
- **MapQuest** додав шар з камерами спостереження і зчитувачами номерів: Flock, Axon, Verkada та інші. [Engadget](#)
- **Естонія** тримає на GitHub код своєї системи інтернет-голосування для публічної перевірки. За словами авторів, там саме той код, що працює на виборах. [GitHub](#)
- **Llmspeed** показує, з якою швидкістю локальна LLM працюватиме на конкретному залізі, ще до покупки. Цифри беруться з бенчмарків спільноти. [Llmspeed](#)