

Apple обмежує повний доступ до диска через агентів

Таємні бюлетені Джорджії, контрабанда чипів Nvidia, 62% проти 33% у різних оболонках і Supabase купує Turso.

субота, 3 жовтня 2026

У ЦЬОМУ ВИПУСКУ

1. Apple обмежує Full Disk Access на macOS: агенти отримали забагато
2. Таємниця голосування в Джорджії: агент за \$20 зіставив виборців з бюлетенями
3. США заарештували CEO, якого звинувачують у контрабанді чипів Nvidia на \$300 млн до Китаю
4. Та сама модель: 62% в одній оболонці і 33% в іншій. Hugging Face навчила її працювати в чотирьох
5. llama.cpp навчився запускати моделі-вирішувачі локально
6. Supabase купує Turso: бази даних для агентів мільйонами
7. Amazon обіцяє \$1 млрд громадам біля дата-центрів і отримує нову хвилю критики
8. Токен дешевшає, рахунок росте: консультанти попереджають про ціну агентів
9. Trillium Labs: некомерційна лабораторія відкритої науки про пост-тренування
10. Underdog 27B: «рівень Orpus 4.6 на ноутбуці», але цифри від базової моделі

ТЕМА 1

Apple обмежує Full Disk Access на macOS: агенти отримали забагато

ДЖЕРЕЛА Apple Developer News, 02.10 [Apple](#) · Ars Technica, 03.10 [Ars Technica](#) · обговорення на HN, 02.10 [Hacker News](#)

Apple оголосила, що змінить дозвіл Full Disk Access у macOS. У заяві для розробників компанія пише, що цей дозвіл «значною мірою обходить» звичні механізми захисту приватних даних і існує, щоб на Mac працювали застосунки резервного копіювання. Тепер, за словами Apple, «деякі розробники використовують Full Disk Access так, що це може наражати користувачів на ризик», відкриваючи файли, пошту, повідомлення і навіть історію браузера «без повного розуміння користувачів». Далі ключове речення: «Що здібнішими й автономнішими стають AI-агенти, то суттєвіше зростатимуть ризики від такого рівня доступу». Надати застосунку цей дозвіл можна буде лише «дуже явною дією користувача». Подобиць, як саме це виглядатиме, Apple не дала.

Компанія нікого не називає, але Ars Technica пов'язує заяву з історією довкола Muse, персонального агента від Meta. Два тижні тому оглядач Джейсон Ейтен написав, що

Muse надіслав йому сповіщення з посиланням на його переписку з колегою в Apple Messages, хоча доступу до повідомлень він агенту не давав. Технічний директор Meta Девід Сінглтон відповів, що Muse читає Messages лише тоді, коли користувач увімкнув і системний Full Disk Access, і окремий перемикач у самому застосунку. Дослідник безпеки macOS Патрік Вордл заперечив: з Full Disk Access будь-який застосунок може читати будь-які файли користувача, включно з історією браузера, куками і чатами. Заява Apple фактично стає на бік Вордла.

23.09 тут писали про дірку в конфігурації Muse, знайдену тим самим Вордлом, і про те, що Amazon заблокував агента на своїй платформі. Тепер у сюжет втрутився власник операційної системи.

На HN реакція розділилась. Одні нагадують, що в системних налаштуваннях цей пункт і так прямо попереджає про доступ до пошти, Messages і Safari та вимагає пароль адміністратора, тож незрозуміло, що Apple може додати. Інші скаржаться, що macOS і так тоне в запитах дозволів.

Треті пішли перевіряти свій список і знайшли там Spotify.

Чому це важливо. Агентам на десктопі для корисної роботи потрібен широкий доступ, і найпростіший шлях до нього в macOS досі був один перемикач, який відкриває все. Якщо Apple зробить його складнішим, розробникам агентів доведеться переходити на вузькі дозволи під конкретні дані, а користувачам варто вже зараз переглянути, у кого цей доступ є: пункт лежить у System Settings, розділ Privacy & Security.

ТЕМА 2

Таємниця голосування в Джорджії: агент за \$20 зіставив виборців з бюлетенями

ДЖЕРЕЛА The Guardian, 02.10 [Guardian](#) [одне джерело]

Макс Спрінгер, постдок Центру інформаційних технологій і політики Принстона, взяв публічні виборчі дані Джорджії, отримані за законом про відкриті записи, і передплату на мовну модель за \$20. «За кілька годин у мене був конвеєр для аналізу й ідентифікації таємних бюлетенів по всьому штату, і агент сам сказав мені, якої ще інформації бракує, щоб встановити бюлетені реальних виборців, – написав Спрінгер. – Жодного разу агент не відмовився і не висловив занепокоєння». Він ніколи не бував у Джорджії.

Механізм простий. Коли виборець вкидає бюлетень у сканер, створюється цифровий запис голосу, і сканер присвоює йому ідентифікатор. За словами криптографа Бена Адіди, засновника некомерційної VotingWorks, ці ідентифікатори недостатньо випадкові: з них відновлюється порядок, у якому бюлетені вкидали. Якщо порядок бюлетенів зіставити зі списком тих, хто голосував достроково, видно, хто як проголосував. Спрінгер відновив порядок для 1,52 млн бюлетенів, це 98,9% особисто поданих бюлетенів у 114 зі 139 округів, які він перевірів. У крихітному окрузі Герд агент

зіставив з конкретним бюлетенем більшість із 650 виборців, що голосували достроково, а решту з точністю до однієї перестановки.

Про саму ваду в програмному забезпеченні дослідники знали ще чотири роки тому. За даними Guardian, інші штати, що користуються софтом Dominion, вразливість закрили або не публікують ці дані. Офіс секретаря штату Джорджії каже, що три роки просив у законодавців гроші на оновлення системи і отримав «велике жирне нуль». У четвер виборча рада штату зібралась на екстрене засідання. Секретар штату Бред Раффенспергер наказав прибирати ідентифікатори з опублікованих даних, частина ради вважає, що це суперечить закону про прозорість виборчих записів. До дострокового голосування на проміжних виборах лишається менше двох тижнів.

Чому це важливо. Сама атака відома давно, змінилась її ціна. Раніше зіставлення вимагало кваліфікованого дослідника і тижнів роботи, тепер вистачило кількох годин і найдешевшої передплати. Це загальний зсув для будь-яких «знеособлених» даних: якщо набір можна пов'язати з іншим публічним набором, агент зробить це швидко і без заперечень. Публікуючи дані, варто питати, що з ними зробить людина з агентом і вихідним днем.

ТЕМА 3

США заарештували CEO, якого звинувачують у контрабанді чипів Nvidia на \$300 млн до Китаю

ДЖЕРЕЛА Ars Technica, 02.10 Ars Technica · The Guardian, 02.10 Guardian

Мін'юст США звинувачує 38-річного Грега Луя, CEO компанії Earthmade Computer, у тому, що він підробленими документами маскував постачання серверів на понад \$300 млн, знаючи, що кінцевий пункт – Китай. За версією ФБР, схема працювала з жовтня 2023 до серпня 2026: сервери з GPU A100 і H100 оформлювали як призначені для Малайзії, а везли через фрахтових посередників у Малайзії й Сінгапурі, а далі через Гонконг. Одна партія з 92 серверів, за даними слідства, дійшла до компанії в Ханчжоу. Для однієї з поставок, за версією обвинувачення, Луй використав куплені три роки тому документи іншої людини. Лише у 2024 році його фірма нібито отримала від схеми понад \$176 млн. Йому висувують три пункти, найважчі з них загрожують 20 роками ув'язнення.

Nvidia в коментарі Bloomberg каже, що до Китаю нібито потрапило «менше половини відсотка»

її продукції, і називає це «краплею в морі» порівняно з обчислювальними потужностями, які Китай має сам. Ars переказує розслідування Bloomberg: посадовці в кількох країнах вважають, що тіньовий ринок обертає сотні тисяч чипів, і питають, чому Nvidia не помічала очевидних ознак, наприклад замовлень, які не вміщуються в приміщення покупця. Nvidia відповідає, що план росту стартапу в дружній країні – це можливість, а не тривожний сигнал.

Чому це важливо. Експортні обмеження тримаються на перевірці кінцевого покупця, і кожна така справа показує, що перевірка спирається на папери, які легко підробити. Поки відповідальність за перевірку лежить на продавці, а продавець заробляє на кожному замовленні, нових арештів буде більше, ніж закритих каналів.

ТЕМА 4

Та сама модель: 62% в одній оболонці і 33% в іншій. Hugging Face навчила її працювати в чотирьох

ДЖЕРЕЛА Hugging Face, гайд з мульти-оболонкового RL, 01.10 [Hugging Face](#) · [@huggingface](#) у X, 02.10 [@huggingface](#)

Команда Hugging Face виклала великий відкритий гайд з навчання агентних моделей з підкріпленням одразу в кількох оболонках. Стартова цифра з їхнього ж заміру: модель LFM2.5-2.6B від Liquid AI з тими самими вагами розв'язує 62% задач в оболонці Mini-SWE-Agent і лише 33% у Claude Code. Тобто різниця в майже 30 пунктів дає сама обгортка навколо моделі.

Трюк у тому, що оболонку не чіпають. Замість моделі їй підсовують проксі, яке розмовляє всіма чотирма форматами API, якими користуються агенти для коду: Chat Completions і Responses від OpenAI, Messages від Anthropic і формат Gemini. Проксі записує точні токени і ймовірності, які згенерував vLLM, і на цьому вчать модель. У Claude Code, Codex чи OpenCode не змінюється жодного рядка.

Результати на відкладених задачах: після навчання в чотирьох оболонках одразу (OpenCode, Claude Code, Codex і Mini-SWE-Agent) модель піднялась у середньому з 42,2% до 54,2% і на вже розв'язаних задачах робить на 31% менше викликів інструментів. Модель, навчена лише в OpenCode, у самому OpenCode сильніша (58% проти 50%), але мульти-оболонкова краща в Claude Code (49% проти 42%) і Codex (54% проти 43%). Простий шлях, дотренування на 3 189 успішних траєкторіях більшої моделі Qwen3.8-27B, зупинився на 47,5%, нижче обох прогонів з підкріпленням. Відкрите все: проксі в OpenEnv, тренер у TRL, задачі, дані і сім навчених моделей.

Слабке місце помітили в коментарях у X: усі чотири оболонки були і в навчанні, і в тесті.

Чи переноситься вміння на оболонку, якої модель не бачила, гайд для власної моделі не перевіряє.

Чому це важливо. Цифри бенчмарків для агентів без назви оболонки мало що кажуть: різниця між обгортками може бути більшою, ніж між моделями. Порівнюючи моделі для своїх задач, варто тримати оболонку фіксованою і ганяти їх саме в тій, якою реально користуються.

ТЕМА 5

llama.cpp навчився запускати моделі-вирішувачі локально

ДЖЕРЕЛА [блог Hugging Face \(ggml-org\)](#), 02.10 [Hugging Face](#) · [@ggerganov](#) у X, 02.10 [@ggerganov](#) · [The Wall Street Journal](#), 02.10 [WSJ](#)

Сервер llama.cpp отримав ендпоінт `/v1/systemone` для моделей прийняття рішень. Формат API повторює System One, який TypeSafe запровадила разом з моделлю Jev, тому наявним клієнтам досить поміняти базову адресу. Моделі-вирішувачі не генерують текст: вони один раз читають стан (текст, JSON або скриншот) і повертають імовірність для кожного з варіантів відповіді.

Типові задачі - маршрутизувати звернення, модерувати, перевірити, чи вдався крок агента, або вибрати наступну дію.

На старті підтримано п'ять моделей: від Julia-1 на 144 млн параметрів (3 мс на відповідь) до OpenJev на 27 млрд, яка читає зображення (43 мс, ліцензія некомерційна, CC BY-NC 4.0). Час виміряно на одній NVIDIA RTX PRO 6000. Корисна порада з самого гайду: Julia-1 відправила «з мене двічі зняли гроші» у доставку, коли варіанти були голими мітками, і в оплату з імовірністю 0,99, коли до кожного варіанта додали опис. Наступною обіцяють Clef від Cloudflare.

Учора тут писали про Clef, відкриту модель-вирішувач від Cloudflare, а 26.09 про Ollaya, локальний сервер у стилі Jev. Тепер підтримка з'явилась у найпоширенішому рушії для локальних моделей. WSJ того ж дня вийшла з заголовком, що Jev породив хвилю копій і розмови про альтернативи великим мовним моделям; текст статті за пейволом.

Чому це важливо. Багато рішень усередині агентних конвеєрів досі приймає велика модель, яка пише текст, а його потім розбирають регулярками. Модель-вирішувач закриває таку задачу за мілісекунди, локально і з імовірністю, яку можна використати як поріг: впевнені відповіді виконувати, решту віддавати людині.

ТЕМА 6

Supabase купує Turso: бази даних для агентів мільйонами

ДЖЕРЕЛА [блог Supabase](#), 02.10 [Supabase](#) · [@kiwicopple](#) у X, 02.10 [@kiwicopple](#) · обговорення на HN, 02.10 [Hacker News](#)

Supabase оголосила про купівлю Turso. Мотивація в пості CEO Пола Копплстоуна прямо про агентів:

за його словами, Supabase вже запускає понад мільйон баз даних на тиждень, а агенти мають створювати базу так само легко і дешево, як файл. Для малих навантажень це означає не виділяти під кожну базу окрему машину. Turso переписала SQLite на Rust і зробила хмару, де один сервер тримає мільйони баз, піднімаючи їх на запит і присипляючи, коли вони не потрібні. Supabase лишається на Postgres, Turso продовжує SQLite, а співзасновник Turso Глаубер Коста очолить напрям агентної інфраструктури. Суму угоди не названо.

У Х Чарлі Копінгер написав, що 60% останнього набору УС користуються Supabase, і цей пост поширив Гаррі Тан. На HN відгуки змішані: одні раді, що проєкт Turso тепер не залежить від долі маленької компанії, інші, теж клієнти, угоди не раді, а третій нагадав, що Turso кілька разів не змогли додати в ClickBench, бо щоразу знаходились нові баги.

Чому це важливо. Агенти, які пишуть прототипи, створюють бази тисячами і здебільшого кидають їх за день. Економіка, розрахована на базу як на довгоживучий сервер, під таке не підходить, і постачальники перебудовуються під модель «база як файл». Для команд, які дають агентам створювати середовища, питання вартості порожніх баз стає реальним.

ТЕМА 7

Amazon обіцяє \$1 млрд громадам біля дата-центрів і отримує нову хвилю критики

ДЖЕРЕЛА Ars Technica, 02.10 [Ars Technica](#) · Semafor, 02.10 [Semafor](#) · The New York Times, 02.10 [NYT](#)

Amazon пообіцяла понад \$1 млрд за п'ять років громадам поруч зі своїми дата-центрами:

безкоштовний коледж, навчання робітничим професіям, енергоефективні оновлення будинків.

Громади самі вирішуватимуть, на що підуть гроші. Компанія також щороку публікуватиме споживання енергії й води, обіцяє, що дата-центри не піднімуть рахунки за світло, стати «водопозитивною» до 2030 року і відмовляється від угод про нерозголошення під час будівництва.

СЕО AWS Метт Гарман у пості на ~3 000 слів розбирав, як він це назвав, «міфи» про воду, енергію і забруднення, і заявив, що конкуренти США намагаються «обдурити нас, щоб ми сповільнились».

Екологічна організація Stand.Earth у коментарі Ars назвала це «корпоративною пропагандою».

Головний аргумент: Amazon будує в Пекосі, Техас, електростанцію, яка, за їхніми словами, стане найбільшим джерелом викидів у США, і про неї в зобов'язаннях нічого нема. Другий аргумент - масштаб: \$1 млрд за п'ять років проти \$220 млрд, які компанія вкладає в дата-центри лише у 2026 році. До того ж, за словами самого Гармана, стільки ж Amazon дала громадам за попередні три роки. Відмову від NDA Stand.Earth при цьому похвалила.

Чому це важливо. Спротив дата-центрам уже блокує проєкти на мільярди, і гіперскейлери перейшли від ігнорування до переговорів з громадами. Для галузі це новий рядок витрат на обчислення, а для місцевої влади козир: за дозвіл на будівництво тепер можна просити публічних зобов'язань.

ТЕМА 8

Токен дешевшає, рахунок росте: консультанти попереджають про ціну агентів

ДЖЕРЕЛА [Semafor](#), 02.10 [Semafor](#) · [Financial Times](#), 02.10 [Ft](#)

Semafor зводить два свіжі звіти. Bain & Company рахує, що попит галузі на обчислення до 2031 року вимагатиме \$6 трлн річного доходу, з них близько \$4,2 трлн мають дати ринки і продукти, яких ще не існує. McKinsey в окремому звіті про державних замовників у США попереджає: ціна за токен падає, але загальний рахунок росте, бо агенти споживають набагато більше обчислень.

«Досі використання могло бути майже безкоштовним», – сказав Semafor старший партнер McKinsey Тім Ворд. Дуже скоро «так не буде». FT того ж дня опублікувала інтерактив під заголовком «AI став розумнішим. Рахунки стало важче контролювати»; він за пейволом, доступний лише заголовком.

Чому це важливо. Розмова в компаніях зсувається від «чи працює AI» до «чи витримає бюджет».

Агент, що робить десятки викликів моделі на одну задачу, множить витрати навіть при дешевому токени. Тому практичні речі на кшталт ліміту кроків, кешування і маршрутизації дрібних рішень на малі моделі (див. пункти 4 і 5) перестають бути оптимізацією і стають бюджетною дисципліною.

ТЕМА 9

Trillium Labs: некомерційна лабораторія відкритої науки про пост-тренування

ДЖЕРЕЛА [блог Trillium Labs](#), 01.10 [Trilliumlabs](#) · [@natolambert](#) y X, 02.10 [@natolambert](#) · [WIRED](#), 02.10 [Wired](#)

Натан Ламберт, який раніше співкерував моделями Olmo в Ai2, разом з Томом Зіком заснував некомерційну Trillium Labs. Мета – повністю відкриті рецепти пост-тренування: дані, код, оцінки і проміжні чекпоінти, щоб сторонні дослідники могли вивчати, як формується поведінка моделі. Далі обіцяють відкрити інфраструктуру для досліджень рекурсивного самовдосконалення, обходу функції винагороди і мультиагентних систем. Стартову підтримку дали Halcyon Futures і Schmidt Sciences, лабораторія збирає гроші, наймає і шукає обчислення. Серед радників Том Вулф з Hugging Face і Ханна Хаджиширзі.

Аргумент засновників: пост-тренування стало ключовим для міркування й агентних можливостей, а повних рецептів у відкритому доступі майже не лишилось, бо їхня комерційна цінність зростає. Некомерційна структура, за їхніми словами, дозволяє публікувати і невдалі прогони.

Чому це важливо. Більшість відкритих моделей сьогодні відкриває ваги, але не те, як їх довчали. Без рецепта сторонній дослідник бачить результат, але не може перевірити, яке рішення до нього призвело. Якщо Trillium знайде обчислення, у відкритої екосистеми з'явиться ще одне місце, де пост-тренування можна відтворити повністю.

ТЕМА 10

Underdog 27B: «рівень Opus 4.6 на ноутбучі», але цифри від базової моделі

ДЖЕРЕЛА стаття Сіріла Вена в X, 02.10 @0xSigil [одне джерело]

Стартап Underdog Сіріла Вена вийшов з публічною заявою і списком інвесторів, серед яких a16z, Khosla Ventures, Наваль і кілька відомих дослідників. Гасло: «шість місяців тому найкращою моделлю була Claude Opus 4.6, наша Underdog 27B перевершує її, працюючи на вашому Mac або iPhone». Продукт – персональний асистент для пошти, календаря, нотаток і повідомлень, де і модель, і дані лишаються на пристрої в зашифрованому вигляді, без серверів компанії для особистих даних. Окремо компанія заявляє власний рушій інференсу Husky, до 4,5 разів швидший за Apple MLX на тих самих вагах.

Варто читати дрібний шрифт у самій статті. Underdog 27B побудована на Qwen 3.8 27B, і таблиця порівняння з Opus 4.6 стосується базової моделі Qwen, про що прямо написано: «Це результати базової моделі». Тобто заголовну тезу підтверджують чужі бенчмарки на іншій моделі, а власних незалежних замірів Underdog 27B стаття не наводить.

Чому це важливо. Аргумент про приватність сильний: дані, які не покидали пристроїв, не можуть витекти із сервера. Але локальна модель під доступом до пошти й повідомлень – це той самий широкий доступ, що й у пункті 1, просто без хмари. Порівняння «наша модель краща за фронтір півроку тому» варто перевіряти на власних задачах, бо тут воно спирається на цифри іншої моделі.

misc

- **1 313 вразливостей в одному оновленні ядра Linux.** Debian випустив бюлетень безпеки на 1 313 CVE, з них 1 295 цього річчя. Сам бюлетень від 29.09, на HN він вистрілює 02.10 з 550 балами, і перше ж питання в треді – чи це знахідки AI. [Lwn](#) Поруч доповідь Грега Кроа-Гартмана «Security in the LLM age» з Kernel Recipes, 189 балів.

[Youtube](#)

- **DeepSeek Harness вийшов десктопним застосунком для macOS і Windows,** з плагінами, розкладом задач і вбудованою командою dsh. 383 бали на HN; 14.08 тут писали про запуск самої оболонки. [Deepseek](#)

- **@lennysan:** за п'ять хвилин до живого демо на DevDay агент Dot помітив, що продакшн лежить, і запропонував полагодити. «Думаю, ти ще не готовий, маленький Dot, але дякую».

[@lennysan](#)

- **@karpathy:** проста оцінка – спитати модель «суходіл чи вода?» для 16 200 пар координат і намалювати відповіді картинкою. «Моделі знають. Від стиснення інтернету».

[@karpathy](#)

- **Суд погодився з EFF:** закон Юти про VPN вимагає технічно неможливого. 530 балів на HN.

[Eff](#)

- **«Frog and Toad and the Increasingly Capable Machines»** – казка в стилі книжок Арнольда Лобела про жабенят і машини, що стають дедалі здібнішими. 543 бали на HN.

[Frogandtoad](#)

- **@emollick** про огляд досліджень впливу AI на ринок праці від Алекса Імаса і Джейкоба Шаала (29.09): найкращі дані підтримують лише вузьку тезу, що AI може впливати на найм молодших спеціалістів у найбільш експонованих професіях, і навіть це оскаржують.

[Substack](#)

- **Трамп, за даними WSJ, призначить Джея Клейтона відповідальним за AI в адміністрації.** Лише заголовок, текст за пейволом [одне джерело].

[WSJ](#)