

Моделі знайшли критичні діри в ядрі PHP

Атаки на ключі до AI у Railway і LiteLLM, агент Claude Duet, Mistral Large 4 і штраф Meta на \$25 мільярдів

четвер, 22 жовтня 2026

У ЦЬОМУ ВИПУСКУ

1. PHP2Shell: моделі знайшли критичні діри в ядрі PHP, а реліз зірвав витік
2. Railway і LiteLLM: дві атаки за добу, обидві заради ключів до AI
3. Anthropic: Claude знайшов 493 зайві входи в код 25 проєктів
4. Anthropic показала Claude Duet, агента, який пише першим
5. Mistral Large 4: трильйон параметрів, ваги до кінця місяця
6. Meta: рекордна виручка, мінусовий грошовий потік і штраф \$25 мільярдів
7. Muse від Meta: 6 800 «приватних» посилань знайшлись одним запитом
8. Моделі OpenAI з'являться в Google Cloud
9. Тисячі агентів GPT-Rosalind за 32 хвилини розібрали атлас мозку миші

ТЕМА 1

PHP2Shell: моделі знайшли критичні діри в ядрі PHP, а реліз зірвав витік

ДЖЕРЕЛА The PHP Foundation, 21.10 [Thephp](#) · Wiz Research, 21.10 [Wiz](#) · OpenAI, 21.10 [OpenAI](#) · Elias Taylor, 21.10 [Eliastaylor](#) · Simon Willison, 21.10 [Simon Willison](#)

Учора тут писали, що PHP готує критичне оновлення. 21.10 воно вийшло: PHP 8.5.12, 8.4.27, 8.3.36 і 8.2.35. Оновлення закриває дві вразливості, CVE-2026-109618 і CVE-2026-109619, обидві з CVSS 9.2.

Обидві сидять у керуванні пам'яттю і обробці фатальних помилок. Запит, що впирається в `memory_limit` у невдалий момент, лишає після себе зіпсований внутрішній стан. Наступний запит у тому самому воркері бере цей стан за живий. Так атакер без логіну отримує або читання пам'яті процесу, або виконання коду.

Wiz пише, що вразливі всі версії аж до PHP 5.4. Дослідники показали виконання коду без автентифікації на свіжих образах WordPress, Drupal і Joomla. Окремо вони показали читання пам'яті воркера з паролями, токенами і ключами API. Wiz працювала з Wiz Atlas, Gemini 4 Argon і GPT-5.6 Sol. За її словами, саме моделі зробили експлоїт практичним: він вимагає довгого ланцюжка кроків під розкладку пам'яті кожної цілі.

Історія фіксу довша за історію знахідки. PHP Foundation пише, що перші ознаки проблеми знайшов її розробник ще у 2024 році, під час аудиту Quarkslab. Команда безпеки не змогла вибрати між кількома варіантами фіксу, і робота стояла місяцями. Фонд прямо називає це структурним провалом.

30.07 PHP увійшов у програму OpenAI Patch the Planet, і аудит почала фірма Calif. 11.08 Calif і OpenAI того самого дня незалежно повідомили про виконання коду з тим самим коренем. Wiz прислала свій звіт 24.08. Від 01.07 команда PHP опублікувала 15 CVE і розібрала понад 100 звітів, які виявились хибними, дублями або не відтворились.

Скоординоване розкриття зламалось за два дні до дати. 20.10 о 15:00 UTC номери CVE з'явились у мережі з хибним часом релізу і хибною оцінкою CVSS 10.0. PHP переніс реліз з 22.10 на 21.10 о 14:00 UTC. Після цього Calif помилково опублікувала відео з робочим експлойтом і потім видалила його, але копії розійшлись. 21.10 о 07:41 UTC Chainguard порушила ембарго і виклала зібрані образи з фіксом. Будь-хто міг вивчати патч у бінарному вигляді за шість годин до релізу.

Оцінки загрози розходяться. OpenAI називає це однією з найвпливовіших проблем безпеки за останні роки і обіцяє \$1 мільярд субсидованого доступу до Daybreak для захисників. Віллісон пише, що експлойти ходили з ночі. PHP Foundation на момент релізу робочих експлойтів не бачила.

Розробник Еліас Тейлор перевіряв публічні PoC з GitHub і пише, що на чистому PHP вони не працюють. Більшість із них вимагає вже наявної діри в застосунку, а саме включення локального файлу.

Чому це важливо. Моделі знайшли вразливість, яка жила в ядрі понад десять років. Це підтверджують незалежно три сторони: PHP Foundation, Wiz і OpenAI. Друга половина історії важливіша для будь-якої команди підтримки. Потік звітів від AI-інструментів уже перевищує пропускну здатність маленьких команд, а ембарго ламається на людях і процесах. Звідси практичний висновок: оновлюватись треба в день релізу.

ТЕМА 2

Railway і LiteLLM: дві атаки за добу, обидві заради ключів до AI

ДЖЕРЕЛА [Railway](#), 21.10 [Railway](#) · [SafeDep](#), 21.10 [Safedep](#) · [@levelsio](#), 21.10 [@levelsio](#)

Хмарна платформа Railway зранку 21.10 помітила підозрілу активність у системах, де лежать змінні середовища клієнтів. За її даними, сторонній отримав доступ до зашифрованих змінних, а частину з них розшифрував. У таких змінних зазвичай лежать ключі API і доступи до баз. Railway радить усім клієнтам змінити всі змінні і ключі реєстрів. Усі деплої на платформі стоять, поки триває розслідування. Компанія вважає атаку частиною серії ударів по платформах, мета яких – ключі до провайдерів інференсу. Провайдерів Railway попросила відкликати всі ключі.

Того ж дня SafeDep розібрала реліз LiteLLM 1.107.0 на PyPI. Цю бібліотеку широко використовують як шлюз до різних LLM. У всіх семи колесах реліз ніс крадій паролів і

бекдор. Шкідливий код запускався з будь-яким процесом Python на машині, навіть без імпорту `litellm`. Крадій збирав ключі SSH, хмарні доступи, секрети CI, токени Claude Code, Cursor, Codex та інших AI-інструментів.

Також він брав ключі провайдерів LLM з бази проксі LiteLLM. Бекдор щогодини читав адресу сервера з файлу на GitHub і запускав код звідти.

У репозиторії на GitHub шкідливого коду нема. SafeDep перевірила 539 запусків GitHub Actions і не знайшла того, що зібрав реліз. Її версія з помірною або високою впевненістю: приватний конвеєр BerriAI зібрав колеса з підміненого коду. Акаунт на GitHub з адресою сервера зник о 13:40 UTC, приблизно через дві години після релізу. PyPI реліз прибрав. Версії 1.106.2, 1.107.0rc1 і 1.107.0rc2 чисті.

Чому це важливо. Ключ до AI-провайдера став окремою ціллю: він дає і гроші, і доступ до даних, і обчислення. Зв'язку між двома атаками ніхто не підтвердив, але ціль у них однакова.

Якщо LiteLLM 1.107.0 потрапив хоча б на одну машину, ключі з неї вважаються скомпрометованими.

Перевстановлення пакета тут не допомагає, потрібна ротація всіх секретів.

ТЕМА 3

Anthropic: Claude знайшов 493 зайві входи в код 25 проєктів

ДЖЕРЕЛА Anthropic, 20.10 [Anthropic](#)

Anthropic описала інший підхід до захисту коду. Модель шукає код, який досяжний без потреби, і ховає його за автентифікацію або явне ввімкнення. Claude пройшов 25 відкритих проєктів, понад 5 мільйонів рядків на C, C++, Java, PHP і Python. Він склав перелік з понад 17 000 опцій конфігурації, з них 3 842 стосуються безпеки. Разом модель знайшла 493 проблеми.

В усіх 11 проєктах, які мають працювати лише після входу, знайшовся код, досяжний до входу. Там модель знайшла 93 зайві поверхні до автентифікації в стандартній конфігурації. Ще 162 комбінації налаштувань система приймала без помилки і без документації. У 94 з них комбінація тихо вимикала захист. В одному випадку дві опції разом вимикали перевірку клієнтського сертифіката. Тоді сервер, налаштований лише на довірених клієнтів, приймав будь-кого.

Один патерн знайшовся в чотирьох незалежних проєктах. Адміністратор пише правило заборони за ім'ям хоста. Зворотний DNS вимкнений, тому сервер порівнює правило з IP-адресою як з текстом.

Ім'я хоста ніколи не збігається з IP, і список заборон пропускає всіх.

Чому це важливо. Пошук окремих вразливостей погано масштабується: виправлена діра нічого не каже про решту. Зменшення досяжного коду знижує вагу і тих дір, яких

ще ніхто не знайшов. Це дослідження самої компанії, незалежної перевірки цифр поки нема. Але перевірку правил доступу за іменами хостів кожен адміністратор може зробити сьогодні.

ТЕМА 4

Anthropic показала Claude Duet, агента, який пише першим

ДЖЕРЕЛА Anthropic, 21.10 [Anthropic](#) · The New York Times, 21.10 [NYT](#) · [@TheRundownAI](#), 21.10 [@TheRundownAI](#)

Claude Duet - окремий застосунок для iOS і Android, на комп'ютері він відкривається на [claude.ai](#).

Агент працює на Opus 5.5. Він сам не чекає запиту: щоранку надсилає зведення з пошти, календаря і робочих застосунків. Потім він кожні кілька годин перевіряє ці акаунти і пише, коли щось потребує уваги. Задачі йому дають повідомленнями, текстом або голосом. Агент працює на власному хмарному комп'ютері і має доступ до понад 6 000 конекторів і плагінів.

Перед надсиланням повідомлення або оплатою агент питає згоди. NYT пише, що купувати він може через Link від Stripe і Shop Pay від Shopify, але кожную покупку схвалює людина. Торгівлю акціями і перекази між рахунками заблоковано навіть зі згодою. Окремий фільтр не дає агенту передавати номери карток і документів.

Duet доступний як дослідне прев'ю за списком очікування, лише для Pro і Max у США. Окремої плати нема, але фонові перевірки витрачають ліміти підписки. Anthropic сама пише, що агент може неправильно прочитати тред або помилитись у терміновості. NYT ставить його в один ряд з dots від OpenAI, Muse від Meta і Gemini Spark від Google. На відміну від них, Anthropic робить ставку на роботу, а побутові доручення агенту даються гірше.

Чому це важливо. Чат-бот чекає людину, агент Duet приходить до неї сам. Через це головне питання тепер у тому, як часто агент відволікає і що він робить без нагляду. Варто дивитись, як фонові перевірки з'їдатимуть ліміти, бо за них платить звичайна підписка.

ТЕМА 5

Mistral Large 4: трильйон параметрів, ваги до кінця місяця

ДЖЕРЕЛА Mistral AI, 21.10 [Mistral](#)

Mistral відкрила публічне прев'ю Mistral Large 4 через API. Це мультимодальна модель на 1 трильйон параметрів, з них активних 52 мільярди. Ваги компанія обіцяє викласти до кінця жовтня.

До того модель проходить червону команду з партнерами з кібербезпеки і державними органами.

Вони отримують версію з послабленою модерацією і розширеними кібер-можливостями.

Модель навчили з нуля на 3 800 GPU NVIDIA Grace Blackwell у власних дата-центрах Mistral в Європі. Цифри дає сама компанія. На DeepSWE v1.1 модель набирає 61,7%, на Terminal-Bench 4 – 28,3%. У сліпій оцінці коду від Surge AI вона друга з п'яти моделей: 3,74 бала проти 4,22 у Claude Opus 5. У тесті з Artificial Analysis Cyber Index, де модель відтворює справжню вразливість і лає її, Mistral заявляє 82%, найвищий результат серед усіх моделей.

Чому це важливо. Mistral прямо продає відкриті ваги як аргумент для кібербезпеки. Відмова провайдера посеред інциденту теж є ризиком, і власна модель його знімає. Незалежних замірів поки нема, і обіцянка «ваги до кінця місяця» поки лише обіцянка. Перевірка буде тоді, коли ваги з'являться на Hugging Face.

ТЕМА 6

Meta: рекордна виручка, мінусовий грошовий потік і штраф \$25 мільярдів

ДЖЕРЕЛА The New York Times, 21.10 [NYT](#) · Semafor, 21.10 [Semafor](#) · CNBC, 21.10 [Cnbc](#) · Генпрокуратура Нью-Мексико, 21.10 [Nmdoj](#)

Meta звітувала за третій квартал. Виручка зросла на 26%, до \$64,6 мільярда. Витрати зросли на 68%, до \$51,6 мільярда. Прибуток склав \$9,8 мільярда. Нижню межу капітальних витрат на рік Meta підняла з \$130 до \$135 мільярдів, основна частина йде на дата-центри. Вільний грошовий потік за квартал став від'ємним: мінус \$2,12 мільярда. Довгостроковий борг виріс до \$80,9 мільярда.

Щоб мати гроші на AI, Meta у січні випустить новий клас акцій і проведе спліт 2 до 1. Semafor пише, що нові акції будуть без права голосу. Так компанія зможе продавати акції, платити інженерам і купувати компанії без розмивання контролю Цукерберга. Схожий хід Alphabet зробила у 2014 році.

Того ж дня суд у Нью-Мексико призначив Meta штраф \$25 мільярдів у справі, що виросла зі скандалу Cambridge Analytica. Присяжні у вересні нарахували понад 43 мільйони порушень закону про нечесну практику, до \$5 000 за кожне. Штат отримав право на \$219,5 мільярда, але суддя зменшив суму до \$25 мільярдів. Відсотки на суму нараховуються за ставкою 15% на рік.

Чому це важливо. Навіть компанія з рекордною рекламною виручкою вже не покриває AI-витрати грошима з бізнесу. Великі техкомпанії перестали викуповувати акції і почали їх продавати. Для інвестора це розмивання частки, для ринку це ще один сигнал ціни гонки за обчисленнями.

ТЕМА 7

Muse від Meta: 6 800 «приватних» посилань знайшлись одним запитом

ДЖЕРЕЛА Ars Technica, 20.10 [Ars Technica](#) · The New York Times, 21.10 [NYT](#)

Агент Muse дає ділитись розмовою через посилання. Усі такі посилання живуть на піддоменах `cf.metaaiusercontent.com`, і в ім'я піддомену входить назва розмови. Коли одержувач відкриває посилання, DNS-запит потрапляє в комерційні бази пасивного DNS. Дослідники фірми Claros одним запитом до такої бази отримали 6 800 посилань.

Вони відкрили кілька сотень і знайшли там внутрішній огляд виручки для CEO, заявку на кредит, нотатки до конфіденційної співбесіди з AI-лабораторією і графік відпусток цілої компанії.

Claros повідомила про проблему через програму bug bounty 08.10 і відповіді не отримала. Після запитання Ars Meta призупинила функцію поширення. За даними NYT, Muse має понад 10,5 мільйона завантажень з 08.09.

Чому це важливо. Довгий випадковий рядок у посиланні не захищає, якщо адреса потрапляє в DNS.

Будь-який сервіс з посиланнями на піддоменах має ту саму слабкість. Секрет у посиланні безпечніше тримати в шляху або в параметрі. Ім'я хоста для цього не годиться.

ТЕМА 8

Моделі OpenAI з'являться в Google Cloud

ДЖЕРЕЛА OpenAI, 21.10 [OpenAI](#) · Ars Technica, 21.10 [Ars Technica](#)

OpenAI і Google Cloud домовились додати моделі OpenAI і агентні сценарії Codex у Model Garden.

Клієнти купуватимуть їх через Google Cloud Marketplace. Компанії з зобов'язаннями витрат у GCP зможуть списувати покупки в рахунок цих зобов'язань. Згодом моделі OpenAI з'являться в Google Antigravity і Gemini Agent. Загальний доступ обіцяють у найближчі тижні. OpenAI окремо пише, що Microsoft лишається її головним хмарним партнером.

Чому це важливо. Для великих компаній бюджет на хмару часто вже розписаний. Можливість платити за OpenAI з бюджету Google прибирає закупівельний бар'єр. Конкуренти Gemini тепер продаються у вітрині самого Google, і вибір моделі під задачу стає звичайною практикою.

ТЕМА 9

Тисячі агентів GPT-Rosalind за 32 хвилини розібрали атлас мозку миші

ДЖЕРЕЛА [OpenAI, 21.10](#) [OpenAI](#) · [@OpenAI, 21.10](#) [@OpenAI](#)

16.10 тут писали, що OpenAI відкрила GPT-Rosalind для академічних лабораторій. Тепер OpenAI, PerturbAI і Scripps Research показали першу велику роботу з нею. Тисячі агентів розібрали атлас CRISPR усього мозку миші за 32 хвилини. Кожен агент брав одне вимкнення гена, аналізував реакцію клітин, шукав літературу і пропонував пояснення. Команда оцінює ручну роботу такого обсягу в кілька місяців або рік.

За даними OpenAI, агенти відтворили висновки людей і знайшли нові гіпотези. Одна пов'язує транспорт холестерину з дегенерацією нейронів через ген DCTN1. Інша пояснює, чому мутації TSC2 дають важчу форму туберозного склерозу, ніж мутації TSC1. Оцінки новизни від агентів збіглися з оцінками експертів у сліпому порівнянні.

Чому це важливо. Схема з тисяч паралельних агентів дає гіпотези. Кожну з них ще треба перевірити в лабораторії. Рецензованої статті поки нема, усі цифри дає компанія. Цінність у швидкості першого проходу по великих даних, які інакше лежать нерозібраними.

misc

- **Tesla і SpaceX:** Маск підтвердив, що спеціальні комітети обох компаній обговорюють злиття. Продовження пункту 4 від 21.10. Він каже, що для угоди потрібна згода щонайменше 50% акціонерів Tesla. Того ж дня Tesla звітувала про падіння прибутку на 43% за квартал.
[BBC](#) · [Ars Technica](#)
- **eMudhra:** mitmproxу додав опцію `--emudhra`, яка підписує сертифікати скомпрометованим проміжним ключем eMudhra. Продовження пункту 3 від 21.10. Тим самим комітом mitmproxу прибрав центри eMudhra зі своїх збірок.
[GitHub](#)
- **GitLab** випустив критичне оновлення для всіх гілок від 18.11 до 19.5. Найважча вразливість, CVE-2026-110264, сидить у компіляторі регулярних виразів і зачіпає всі версії від 7.0.
[Gitlab](#)
- **@TZahavy (Google DeepMind):** конвеєр на Gemini 4 Argon знайшов 20 нових кандидатів у «криптиди» для задачі BB(6). Це крихітні машини Тюрінга, чия зупинка рівносильна відкритим задачам теорії чисел. Доведення формалізовані в Lean 4.
[одне джерело]
[@TZahavy](#)
- **Qwen3-Coder:** фірма Hirundo пише, що модель на 19% частіше видає вразливий застосунок, коли замовник – оборонний підрядник із США. Hirundo продає послугу

«вестернізації» моделей, тому це дослідження зацікавленої сторони. [одне джерело]

[Hirundo](#)

- **Google DeepMind:** BBC поговорила з кількома співробітниками, які звільнились через військові контракти компанії. Google відповідає, що перевіряє кожен контракт за своїми принципами AI. [одне джерело] [BBC](#)
- **DuckDB 2.0.0 LTS** вийшов з асинхронним вводом-виводом, типом VARIANT, тригерами і режимом для агентів. Підтримка триватиме щонайменше до жовтня 2027 року.

[Duckdb](#)

- **@ycombinator:** Legora залучила \$550 мільйонів за оцінки \$8,8 мільярда. На платформі працюють понад 140 000 юристів разом з агентами.

[@ycombinator](#)

- **@AnthropicAI:** фізик Крістофер Стаббс, колишній декан природничих наук Гарварду, увійшов до Long-Term Benefit Trust компанії. [@AnthropicAI](#)