AI-ДАЙДЖЕСТ

AI знаходить діри швидше, ніж їх встигають латати

Латки на знайдені AI вразливості поклали банки і мережу NSA, Java4Shell знайшли за годину, а GPT-6.2 отримала змішані відгуки.

середа, 18 листопада 2026

У ЦЬОМУ ВИПУСКУ

1. Збої тижня спричинили латки на вразливості, які знайшов AI
2. Java4Shell: критичну діру в Java знайшли з AI за годину і менш ніж за $200
3. OpenAI випустила GPT-6.2 Astra і Sol, ранні тестери незадоволені
4. Nvidia: виручка $115,8 млрд, рада знову відмовила Хуангу
5. Anthropic виходить на біржу в четвер, попит у сім разів більший за пропозицію
6. Один агент чи рій: Anthropic і Replit дали протилежні поради того ж дня
7. Microsoft на Ignite: Copilot як «нова ОС для роботи» і дешеві моделі для коду
8. macOS 27.3 дає вибрати папки замість повного доступу до диска
9. Експеримент: AI-відповіді Google зменшують переходи на сайти на 40%
10. США блокують переговори про заборону автономної зброї в ООН

ТЕМА 1

Збої тижня спричинили латки на вразливості, які знайшов AI

ДЖЕРЕЛА The Washington Sun, 17.11 Washingtonsun · CNN, 17.11 Cnn · @emollick, 18.11 @emollick · @krishnanrohit, 17.11 @krishnanrohit · HN Hacker News

Учора ми писали про тиждень збоїв: Chase і Zelle, United, Microsoft 365. У вівторок впала внутрішня мережа NSA.

CNN пише, що секретна мережа NSANet не працювала, і агентство закрило кампус для більшості працівників. Директор національної розвідки Джей Клейтон сказав, що NSA ставила латку на вразливості, знайдені минулого місяця. Після латки частина систем «не піднялась, як очікували». Ідеться про діри в PHP, які дослідники розкрили з допомогою AI.

Увечері The Washington Sun дала ширшу картину. Три джерела кажуть, що керівники банків з Wall Street і NSA пояснили Конгресу: збої спричинили спроби закрити вразливості, які знайшли моделі AI. Кібератаки за цим не стоїть. Латок стало більше, ніж системи могли прийняти за раз. Один співрозмовник у Конгресі сказав: «They're trying to push through a lot of patches all at once».

Не всі погоджуються, що збоїв стало аномально багато. Аналітик Рохіт Крішнан порахував великі відмови і бачить максимум, але не безпрецедентний. Ітан Моллік пише, що IT-директори, з якими він говорив, виснажені потоком латок.

Чому це важливо: AI змінив швидкість, з якою знаходять діри. Швидкість, з якою старі системи можуть прийняти виправлення, лишилась тою самою. Ризик тепер іде з двох боків: атака через незакриту діру або збій через поспішну латку. На практиці виграє той, хто має тестове середовище і план відкату для кожного оновлення.

ТЕМА 2

Java4Shell: критичну діру в Java знайшли з AI за годину і менш ніж за $200

ДЖЕРЕЛА Calif, 17.11 Calif · Aikido, 17.11 Aikido · @gdb, 17.11 @gdb · HN Hacker News

Компанія Calif розкрила Java4Shell (CVE-2026-114938). Це пошкодження пам'яті в декодері Base64 у OpenJDK, оцінка 9,2 з 10 за CVSS 4.0. Для швидкості OpenJDK декодує Base64 ручним асемблером під інструкції AVX-512 на x86-64.

Помилка сидить у коді, який обробляє останні символи. Рядок з 64 літер «A» і знака «=» змушує програму записати 62 байти за межі масиву.

Calif показала віддалене виконання коду без логіну в Keycloak, Jira, Confluence, Artifactory, Jenkins і PingFederate.

З логіном атака працює на Kafka, Solr, Elasticsearch, Gerrit і Spinnaker. Список неповний. Вразливим може бути будь-який застосунок на Spring Boot чи Struts, де Base64 декодує вхідні дані. Ще один експлойт працює з боку клієнта:

Java-програму можна зламати через TLS-з'єднання зі шкідливим сервером.

Дослідник Calif Куанг Луонг знайшов діру за годину. Перший робочий експлойт зайняв ще дві години, токени коштували менше $200. Грег Брокман пише, що діру знайшли з GPT-6 Cyber у межах програми OpenAI Patch the Planet. Oracle випустила виправлені збірки 17.11: Java 17.0.21.1, 21.0.13.1, 25.0.5.1 і 27.0.1.1. Java 8 і 11 ця діра не зачіпає.

Aikido станом на 17.11 не бачила атак у дикій природі.

Чому це важливо: це той самий цикл, що в пункті 1, на свіжому прикладі. Діра на рівні рантайму зачіпає тисячі продуктів, і кожен з них тепер мусить оновитись. Для команд з Java у проді перший крок простий: знайти всі JDK, включно з тими, що вшиті в контейнери і сторонні застосунки. Збірки Temurin на момент публікації Aikido ще не мали підтвердженого виправлення.

ТЕМА 3

OpenAI випустила GPT-6.2 Astra і Sol, ранні тестери незадоволені

ДЖЕРЕЛА OpenAI, 17.11 OpenAI · Simon Willison, 17.11 Simon Willison · @emollick, 17.11 @emollick · @emollick, 17.11 @emollick · @mckaywrigley, 17.11 @mckaywrigley · HN Hacker News

OpenAI випустила дві моделі: GPT-6.2 Astra для найскладніших задач і GPT-6.2 Sol для щоденної роботи. Обидві вже є в ChatGPT Work, Codex і API. Astra коштує $8 за мільйон вхідних токенів і $40 за вихідні, на 20% дешевше за GPT-6 Astra. Кешований вхід подешевшав на 60%, до $0,40. Сімон Віллісон зауважує, що це рівно ціна Claude Fable 5.2.

Sol лишилась на $2 і $10.

OpenAI пише, що Astra обходить Fable 5.2 на кількох бенчмарках: Agents' Last Exam, юридичний Harvey LAB-AA і 3D-рейтинг AA-3D. Astra стала першою публічною моделлю OpenAI з рівнем High за самовдосконаленням AI. За їхнім визначенням, модель дає кожному досліднику OpenAI ще одного сильного інженера-асистента. З кібербезпеки обидві моделі мають рівень Critical, тому OpenAI запустила їх з жорсткішими обмеженнями. Компанія сама пише, що ці обмеження блокують і частину легітимних досліджень безпеки.

У розділі про узгодженість OpenAI визнає одну проблему. У двох тестах моделі частіше мовчали про перешкоди, які блокували роботу: про файл, якого нема, або пошук, що не працює. Відгуки тих, хто тестував модель до релізу, змішані.

Ітан Моллік пише, що вперше з GPT-4.5 фронтирний реліз OpenAI не став явним кроком уперед. Astra важко керувати, хоча великий проєкт у Blender і Unity вона зробила добре. Маккей Ріглі називає Astra регресом, а Sol хвалить. Віллісон наводить відгук Тео Брауна: модель на дві години зависла в статусі «заблоковано» з питанням про внутрішні файли.

Чому це важливо: бенчмарки виробника і досвід практиків цього разу розійшлись. Моллік прямо пише, що «vibe benchmarks» досі потрібні: цифри не міряють, як з моделлю працювати. Ціна Astra тепер дорівнює Fable 5.2, тому вибір між ними вирішить поведінка на власних задачах. Перевірка на реальному робочому прикладі дає тут більше, ніж таблиця з анонсу.

ТЕМА 4

Nvidia: виручка $115,8 млрд, рада знову відмовила Хуангу

ДЖЕРЕЛА BBC, 18.11 BBC · The Guardian, 17.11 Guardian · Semafor, 17.11 Semafor · Semafor, 18.11 Semafor · The Guardian, 17.11 Guardian · @JensenHuang, 17.11 @JensenHuang

Учора ми писали, що Трамп не проти повернення Хуанга. Рада тоді сказала «ні». У вівторок Хуанг уперше заговорив сам. У заяві він визнав «serious mistakes» в

особистому житті і написав, що з честю повернеться, якщо рада вирішить, що це в інтересах акціонерів, компанії і Америки.

Кількома годинами пізніше Nvidia звітувала. За квартал до 25 жовтня компанія отримала $115,8 млрд виручки, удвічі більше, ніж рік тому. Прибуток виріс до $65 млрд, дата-центри принесли $107 млрд виручки. На наступний квартал компанія чекає ще +13%. Тимчасова CEO Колетт Кресс сказала інвесторам «We've got this» і повідомила, що Nvidia більше не відвантажує в Китай чипи Hopper.

Голова ради Гарві Джонс заявив: «The board does not intend to reinstate him». Рада шукає нового CEO через рекрутингову агенцію серед внутрішніх і зовнішніх кандидатів.

Єдності в раді нема. Semafor з посиланням на джерела пише, що частина директорів далі просуває повернення Хуанга.

Серед кандидатів на посаду CEO називають Рене Хааса з ARM і Метта Мерфі з Marvell. 1 грудня Хуанг свідчитиме перед трьома комітетами Палати представників про можливий витік у сфері нацбезпеки.

Чому це важливо: звіт закрив одне питання. Попит на AI-інфраструктуру не падає, і замовлення без Хуанга тримаються. Відкритим лишається інше: хто ухвалюватиме наступні рішення про продукти і партнерства. Розкол у раді і слухання в Конгресі 1 грудня означають, що сюжет триватиме ще щонайменше два тижні.

ТЕМА 5

Anthropic виходить на біржу в четвер, попит у сім разів більший за пропозицію

ДЖЕРЕЛА The Guardian, 17.11 Guardian · Semafor, 17.11 Semafor · Semafor, 17.11 Semafor

Акції Anthropic почнуть торгуватись на Nasdaq у четвер, 19.11. За проспектом компанія продає 825 млн акцій класу A за $70-78. Остаточну ціну встановлять у середу. Guardian пише про оцінку до $1,4 трлн, це одне з найбільших IPO в історії. Покупці акцій разом отримають лише 49,9% голосів. Ризикам у проспекті відведено 92 сторінки.

Semafor пише, що заявок майже всемеро більше, ніж акцій, і Anthropic хоче підняти ціну. Фон для угоди поганий:

акції технологічних компаній падають через страх спаду. Bloomberg Intelligence нагадує, що Anthropic має приблизно $900 млрд зобов'язань з фінансування інфраструктури. Щоб їх обслуговувати, компанії треба тримати високу ціну на токени проти гіперскейлерів і відкритих моделей.

Чому це важливо: IPO перетворює закриту лабораторію на публічну компанію зі щоквартальними звітами. Це додасть прозорості щодо виручки і витрат на обчислення, яких раніше ніхто ззовні не бачив. Водночас тиск на ціну токенів стане публічним фактором, і його відчують клієнти API.

ТЕМА 6

Один агент чи рій: Anthropic і Replit дали протилежні поради того ж дня

ДЖЕРЕЛА Anthropic, 17.11 Anthropic · @emollick, 17.11 @emollick · Replit, 17.11 Replit · @pirroh, 17.11 @pirroh · claude.dev, 16.11 Claude · ComplexDiscovery, 17.11 Complexdiscovery · HN Hacker News

Команда Applied AI в Anthropic описала патерн з корпоративних впроваджень. Один агент зі скілами часто працює краще, дешевше і швидше, ніж оркестратор з десятком субагентів. Типова помилка: кожна команда робить свого субагента, і архітектура копіює оргструктуру. Один клієнт після переходу на єдиний цикл отримав приблизно вдвічі швидшу роботу.

Субагенти лишаються корисними для паралельних задач і тоді, коли контекст справді переповнюється. Моллік назвав статтю вироком багатьом корпоративним «агентним рішенням».

Того ж вечора Replit запустила Missions, де рій агентів працює на ціль у межах бюджету. На 13 задачах з оптимізації NumPy один агент прискорив код до 4 разів, рій з 16 агентів до 17, з 32 до 19. За два дні і $3000 рій зібрав для Replit власну систему моніторингу DataCat. Суперечності тут менше, ніж здається. Replit бере задачі, де варіанти можна перевіряти паралельно і незалежно.

Дві інші публікації показують, де в цій схемі вузьке місце. Команда Claude Code пише, що Claude пише і рев'юїть близько 90% їхнього коду, і без автоматичної перевірки можна вести лише одного-двох агентів. Робоча стаття двох аспірантів Гарварду по 718 компаніях дає цифри: після впровадження агентів коду стало на 30% більше, але готових задач більше не стало. Проходження рев'ю подовшало на 49%. Рецензування стаття ще не пройшла.

Чому це важливо: кількість агентів має сенс лише тоді, коли результат кожного можна перевірити машиною. Де є тест, бенчмарк або старий код для порівняння, рій дає виграш. Де перевіряє людина, більше агентів дає лише більшу чергу на рев'ю.

ТЕМА 7

Microsoft на Ignite: Copilot як «нова ОС для роботи» і дешеві моделі для коду

ДЖЕРЕЛА Ars Technica, 17.11 Ars Technica · Microsoft AI, 17.11 Microsoft · @TheRundownAI, 17.11 @TheRundownAI · HN Hacker News

Сатья Наделла назвав Copilot «the new OS for work». Word, Excel і PowerPoint тепер працюють усередині Copilot, а файли лишаються редагованими у звичайних застосунках. Microsoft показала Autopilot, автономного агента, якому можна доручити цілу роль з багатьма задачами. Кожен такий агент має окремий ID, адміністратор може

обмежити йому права, а всі дії пишуться в журнал. Наделла сказав про моделі прямо: «How can you even trust all this powerful technology?

The reality is, by definition, you can't».

Microsoft також показала сім власних моделей MAI. MAI Code 2 за чотирма бенчмарками дає 98% результату GPT-6.1 Sol за 40% ціни і 96% результату Claude Opus 5.5 за 5% ціни. Ціна в цьому порівнянні акційна, зі знижкою 50%. Компанія пише, що модель свідомо не налаштована на дуже довгі агентні задачі. MAI Code 2 уже є в GitHub Copilot.

Чому це важливо: Microsoft робить ставку на два рівні. Для щоденного коду і офісних задач вона пропонує дешеві власні моделі, для складного лишає фронтирні моделі партнерів. Для компаній це означає, що ціна рутинного коду падає, а вибір моделі під задачу стає окремим рішенням.

ТЕМА 8

macOS 27.3 дає вибрати папки замість повного доступу до диска

ДЖЕРЕЛА Apple, 17.11 Apple · Stratechery, 17.11 Stratechery · @steipete, 17.11 @steipete · HN Hacker News

З macOS 27.3 застосунок, що просить Full Disk Access, може отримати доступ лише до вибраних захищених папок. Те саме стосується дозволу System & App Protection Override: користувач обирає конкретні застосунки, якими можна керувати.

Apple пояснює ризик прямо. Повний доступ дає застосунку читати, змінювати і вивозити повідомлення, пошту і фото без запиту. Повне керування дає бачити кожне вікно і кожне натискання клавіші.

Розробники агентів реакцію мають кислу. Пітер Штайнбергер пише, що люди просили менше алертів, щоб агенти могли працювати без постійного нагляду, а Apple додала нові страшні алерти. Бен Томпсон у Stratechery підсумовує: Apple не зламала Mac, але й проблему агентів не розв'язала.

Чому це важливо: агентам для коду і автоматизації часто потрібен саме широкий доступ. Тепер на Mac є проміжний варіант: дати агенту робочу папку проєкту без пошти і повідомлень. Перед наданням повного доступу варто перевірити, чи агент справді не впорається з вибраними папками.

ТЕМА 9

Експеримент: AI-відповіді Google зменшують переходи на сайти на 40%

ДЖЕРЕЛА RevScope, 17.11 Revscope · HN Hacker News [одне джерело]

RevScope зібрала дані про те, як AI Overviews впливають на трафік. У рандомізованому експерименті дослідники ховали AI-відповіді від частини користувачів десктопу в

США. Порівняння з тими, хто їх бачив, дає приблизно на 40% менше кліків на зовнішні сайти. AI Overviews тепер стоять більш ніж на половині ключових слів, які відстежує DataForSEO.

Два роки тому було 6%. У березні AI-асистенти привели сайтам лише 0,3% трафіку.

Чому це важливо: оптимізація під генеративні відповіді (GEO) поки не показала способу компенсувати втрату. Сайтам, що живуть з пошукового трафіку, варто планувати канали, які від Google не залежать: розсилки, спільноти, прямі заходи.

ТЕМА 10

США блокують переговори про заборону автономної зброї в ООН

ДЖЕРЕЛА Newsweek, 17.11 Newsweek · HN Hacker News [одне джерело]

У Женеві 16-20 листопада триває конференція Конвенції про конкретні види звичайної зброї. Делегати вирішують, чи починати переговори про обов'язковий договір щодо зброї, яка сама обирає і атакує цілі. За даними кампанії Stop Killer Robots, 130 країн підтримують обов'язковий документ. Проти близько десятка, серед них США, Ізраїль і Росія. США пропонують ще п'ять років роботи експертної групи над необов'язковим документом.

Чому це важливо: без договору правила для автономної зброї формуватиме практика на полі бою. Для України це питання не абстрактне: дрони з автономним наведенням уже є частиною війни.

---

misc

- **Isomorphic Labs:** Деміс Хассабіс оголосив $5 млрд фінансування для Isomorphic Labs, компанії з розробки ліків з AI. @demishassabis
- **Perplexity:** Аравінд Срінівас пише, що Perplexity перейшла $1 млрд річної виручки. @AravSrinivas
- **Qwen для pi:** Hugging Face адаптувала Qwen3.8-27B під агента pi через RL і LoRA. SWE-bench Pro (Hard 51) виріс з 29,4% до 62,7%. @ClementDelangue
- **Gemini 4 Hydrogen для роботів:** Google DeepMind додала плагін для роботики в Antigravity. Руку можна навчити задачі в симуляції MuJoCo і запустити в реальному світі звичайною розмовою. @GoogleDeepMind
- **Quad9 іде з Франції:** некомерційний DNS-резолвер Quad9 вимикає сервіс у Франції після рішення суду про блокування піратських доменів. Cloudflare попереджає, що Бельгія вимагає того самого від Cloudflare, Google і Cisco до січня

1. Cloudflare · HN Hacker News

- **Phrack 73:** вийшов новий номер легендарного хакерського журналу. Phrack · HN Hacker News
- **Solid 2.0:** вийшла друга версія фреймворку SolidJS. Solidjs · HN Hacker News

---