

ChatGPT розносить рекламний ідентифікатор по сайтах

Механіка на 936 пікселях, переговори США і Китаю щодо AI, \$300 млрд зобов'язань поза балансом.

понеділок, 21 вересня 2026

У ЦЬОМУ ВИПУСКУ

1. ChatGPT розносить рекламний ідентифікатор по чужих сайтах - розбір механіки
2. США і Китай домовились про діалог щодо AI напередодні зустрічі Трампа і Сі
3. Big Tech тримає \$300 млрд AI-зобов'язань поза балансом
4. Step 5 Preview: 600B параметрів, 27B активних, 1M контексту - і Покемон як бенчмарк
5. Pirate Face: ваги моделей як магнет-лінки, які не можна зняти
6. Google відкрила AX - оркестратор агентних задач на Kubernetes
7. Samsung подвоює випуск HBM4 - і ставка на скляні підкладки
8. Під прикриттям у банді, що зламувала ланцюг постачання
9. «Спіраль смерті сеньйора» - текст, який два дні тримається в топі
10. Терренс Тао дав слово гостю: навіщо ще потрібні математики-люди

ТЕМА 1

ChatGPT розносить рекламний ідентифікатор по чужих сайтах - розбір механіки

ДЖЕРЕЛА дослідження [Buchodi](#) · обговорення [Hacker News](#)

Незалежний дослідник розібрав по кроках, як працює рекламний збирач OpenAI на `bzr.openai.com` (`bzr` - від `bazaar`, внутрішня назва рекламної платформи).

Механізм відтворено на власному телефоні, підтверджено двома незалежними методами перехоплення і звірено з кількомісячним трафіком: 936 рекламних пікселів на 1029 хостах.

Як це влаштовано. На `chatgpt.com` клієнт генерує 16 випадкових байтів і просить бекенд підписати їх. Назад приходить JWT, підписаний RS256, який зв'язує рекламний ідентифікатор `obi` з полем `sub` - тобто з обліковим записом користувача. Токен живе 60 секунд. Далі цей ідентифікатор осідає кукою `__obi` на домені `.openai.com` з `Max-Age` на рік і прапорцями `SameSite=none; Secure` - саме тією комбінацією, яка дозволяє кукам їздити на крос-сайтових запитах.

Звідти починається цікаве. Будь-який рекламодавець ставить собі на сайт шматочок коду OpenAI – так само, як роками ставлять піксель Meta або Google. І `__obi` їде до OpenAI разом з даними про сторінку: що людина шукає, що читає, що купує.

Окремо варто зупинитись на одній деталі, бо вона показує межу «приватних» режимів.

У SDK є гілка коду, яка свідомо не передає credenшали. **Вона не допомагає:** браузер чіпляє куки вже на завантаження самого `<script src>`, тобто ДО того, як відпрацює будь-який рядок коду OpenAI. Сам факт підвантаження тега вже розкриває ідентифікатор.

І ще одна цифра, яка міняє рамку розмови. SDK збирає ідентичність зі сторінки рекламодавця з чотирьох джерел, і OpenAI сама їх розмічає: `in` – те, що рекламодавець передав свідомо, і `fm`, `ht`, `js` – те, що SDK **вишкріб сам** з полів форми, з тексту сторінки і з шини тег-менеджера. У спостереженому трафіку зіскрябана ідентичність перевищила свідомо передану **685 подій проти 255**. Найбільше джерело пошти – шина тег-менеджера, яку SDK підміняє власною функцією.

Чому це важливо. Сам по собі механізм – звичайний адтех, йому років п'ятнадцять.

Безпрецедентне тут інше: він вперше поїхав на продукт, якому люди диктують те, чого не пишуть у пошук. У пошуковий рядок вводять запит, а в чат переказують ситуацію – з діагнозами, боргами, конфліктами і документами. Прив'язка `obi` до `sub` означає, що це той самий обліковий запис, тобто не анонімний профіль. Практичний висновок для будь-якого продукту з чатом: межа приватності визначається не обіцянкою в інтерфейсі, а тим, які куки фактично їдуть на сторонні хости – і перевіряється це перехопленням трафіку. Читання політики тут не допоможе.

ТЕМА 2

США і Китай домовились про діалог щодо AI напередодні зустрічі Трампа і Сі

ДЖЕРЕЛА FT FT · NYT NYT

Міністр фінансів США Скотт Бессент зустрівся зі своїм китайським візаві Хе Ліфеном у Нью-Йорку. Сторони домовились про окремий діалоговий трек щодо штучного інтелекту – напередодні державного візиту Сі Цзіньпіна до США.

Вчора згадували пост Трампа про «AI Force» з обіцянкою призначити «AI-царя» – без жодної деталі. Сьогодні видно, що саме стоїть за риторикою: поки публічна частина виглядає як допис у соцмережі, робоча частина йде переговорним каналом між міністерствами фінансів.

Контекст, який дає NYT: Китай виходить на ці переговори з дуже нерівним розкладом.

AI там рухається вперед, а економіка **в найгіршому стані за десятиліття**. WSJ у той самий день описує зворотний бік цієї медалі – минулого року в Китаї зареєстрували

понад сім мільйонів нових компаній з однією людиною, і значна частина цього підприємництва йде від безробіття і вигорання, тобто від нужди.

Чому це важливо. Коли два найбільші гравці сідають говорити про технологію окремим треком, це зазвичай означає, що правила гри почнуть писатись там, за межами компаній.

Для всіх, хто будує продукти на моделях, це ранній сигнал: обмеження на доступ до заліза, ваг і ринків приходять з дипломатичного столу і міняються швидше за технологію.

Дивитись варто на те, які саме теми потрапили в порядок денний робочих груп.

Заяви про перегони тут кажуть найменше.

ТЕМА 3

Big Tech тримає \$300 млрд AI-зобов'язань поза балансом

ДЖЕРЕЛА [FT](#) [FT](#)

Волл-стріт знайшла спосіб перетворити кредитну міцність технологічних гігантів на дешевше фінансування будівництва дата-центрів. Механізм – гарантії: технологічна компанія гарантує борг структури, яка будує потужності, і сама експозиція при цьому **не потрапляє на її баланс**. Обсяг, який FT оцінює за цією схемою, – близько \$300 млрд.

Чому це важливо. Це та сама конструкція, за якою колись будувались лізингові схеми і позабалансові структури, і вона завжди означає одне: ризик нікуди не подівся, він просто перестав бути видимим у звітності. Для читача, який намагається зрозуміти реальний масштаб AI-будівництва, практичний висновок такий – капітальні витрати в квартальних звітах більше не описують повну картину, і оцінювати темп інвестицій за ними самими вже не можна.

ТЕМА 4

Step 5 Preview: 600B параметрів, 27B активних, 1M контексту – і Покемон як бенчмарк

ДЖЕРЕЛА [StepFun](#) [Stepfun](#) · обговорення [Hacker News](#)

Китайська StepFun показала Step 5 Preview. Архітектура – розріджена суміш експертів:

600 млрд параметрів загалом, 27 млрд активних на токен, вікно контексту **1 млн токенів**, підтримка зображень на вході. За індексом Artificial Analysis Intelligence Index модель набирає **44**. Обіцяють випустити з відкритими вагами.

Найцікавіша частина – як вони міряють довгу автономність. Без жодної спеціальної оптимізації під гру модель грає в Покемона і **протримала осмислений прогрес понад**

3000 ходів і 6 мільйонів токенів взаємодії. На ході 3082 вона відкрила Cut, зібрала три значки і перемогла Лейтенанта Сьорджа – приблизно третина основної сюжетної лінії.

Чому це важливо. Класичні бенчмарки міряють одну відповідь, а реальні агентні задачі ламаються на іншому – на здатності не втратити нитку через тисячі кроків.

Гра тут зручна саме тим, що прогрес у ній не підrobiш: або значок є, або нема. Це корисний спосіб думати про оцінку агентів загалом – шукати задачу з довгим горизонтом і однозначним зовнішнім критерієм успіху, замість того щоб вірити середньому балу на коротких прикладах.

ТЕМА 5

Pirate Face: ваги моделей як магнет-лінки, які не можна зняти

ДЖЕРЕЛА [pirateface.co](#) [Pirateface](#) · обговорення [Hacker News](#)

Сервіс роздає відкриті моделі – мовні, графічні, аудіо і датасети – **магнет-лінками** замість завантаження з одного хостингу. Ідея в одному рядку: у торрента нема власника і нема єдиної точки відмови, тому зняти роздачу нема в кого.

Реакція на HN показова тим, наскільки вона одностайна. Лейтмотив: «торренти мали б бути основним способом роздачі ваг з першого дня», і «це мало б існувати ще до того, як почались викупи».

Чому це важливо. Питання тут ширше за пірати-не-пірати. Будь-яка команда, яка будує продукт на відкритій моделі, має відповісти собі, що станеться, якщо конкретні ваги завтра зникнуть з хостингу – через ліцензію, купівлю компанії чи регуляторну вимогу. Практичний висновок простий: ваги, на яких стоїть продакшн, тримають у себе, а не покладаються на чужий CDN. Роздача через торрент – лише один зі способів це зробити.

ТЕМА 6

Google відкрила AX – оркестратор агентних задач на Kubernetes

ДЖЕРЕЛА [agentexecutor.io](#) [Agentexecutor](#) · обговорення [Hacker News](#)

AX описує агентну задачу декларативно, у YAML: воркспейс (наприклад, git-репозиторій з гілкою) і задача з метою простим текстом. Далі система сама кладе це в пісочницю, піднімає оточення, **відгороджує мережу** і обіцяє масштаб до мільярдів задач на кластер. Задачу можна призупинити і відновити, і файли в воркспейсі переживають паузу.

Реакція інженерів прохолодніша за саму технологію. Найточніший коментар у треді – «k8s-ифікація AI була неминуча». Окремо згадують, що в Google є **паралельний проект Scion**, який краще вживається з наявними інструментами, тоді як AX – більше самостійна конструкція з нуля.

Чому це важливо. Тут видно, у що перетворюється інфраструктура агентів: із бібліотеки, яку імпортують у код, вона стає планувальником кластера з власними маніфестами. Для команд, які вже експлуатують Kubernetes, це знайома модель. Але варто помічати й ціну – декларативний YAML і власний рантайм означають ще один шар, який треба розуміти під час інциденту.

ТЕМА 7

Samsung подвоює випуск HBM4 – і ставка на скляні підкладки

ДЖЕРЕЛА Seoul Economic Daily [Sedaily](#) · обговорення [Hacker News](#)

Samsung очікувано більш ніж подвоїть випуск пам'яті родини HBM4 наступного року.

Потужності мають вирости до **250 тисяч пластин на місяць**. Шосте і сьоме покоління (HBM4 і HBM4E) складуть **80% випуску**. Окремо повідомляють про зростання обсягів замовлених скляних носіїв **у 2,5 раза** – вони потрібні для стекування великої кількості шарів.

Чому це важливо. Швидкість пам'яті давно стала вужчим місцем для навчання і інференсу, ніж кількість обчислювальних ядер. Тому плани випуску HBM – чесніший індикатор реального темпу галузі, ніж анонси моделей: залізо планують за роки наперед і воно погано бреше. Деталь зі скляними підкладками тут не технічна дрібниця, а підказка, де саме зараз стоїть виробничий бар'єр.

ТЕМА 8

Під прикриттям у банді, що зламувала ланцюг постачання

ДЖЕРЕЛА Ars Technica / WIRED [Ars Technica](#)

Група загрозної розвідки Google повідомила, що мала **свою людину у близькому колі** угруповання, відомого як TeamPCSR. Це стало відомо після того, як двох імовірних учасників заарештували і висунули звинувачення в Австралії минулого місяця.

Чому це важливо. Атаки на ланцюг постачання – той клас, де технічний захист окремої компанії майже нічого не вирішує: зламують того, чий пакет ти підключаєш. Тому боротьба з ними дрейфує у бік агентурної роботи і правоохоронців.

Для команд практичний висновок лишається нудним і робочим: знати, що саме тягнеться в збірку транзитивно, і фіксувати версії залежностей.

ТЕМА 9

«Спіраль смерті сеньйора» – текст, який два дні тримається в топі

ДЖЕРЕЛА Sunil Pai [Sunilpai](#) · обговорення [Hacker News](#)

Автор описує типовий збій: інженер виходить на нову сеньйорну роль і починає

відігравати роль більш досвідченого, ніж він є. Береться за завеликий дизайн, працює по 60-80 годин, щоб «довести», і вигорає – саме тому, що намагався закрити невпевненість обсягом роботи.

Текст написаний як монолог другові, який щойно отримав дуже сеньйорну і дуже оплачувану позицію і питає, як витримати такий графік.

Чому це важливо. Це рідкісний випадок, коли в топі HN тримається не технологія, а робоча патологія – і тримається тому, що впізнається. Корисне тут те, що симптом названо точно: це спроба закрити синдром самозванця обсягом роботи.

Розрізняти ці два стани варто вміти і в собі, і в людях, за яких відповідаєш.

ТЕМА 10

Терренс Тао дав слово гостю: навіщо ще потрібні математики-люди

ДЖЕРЕЛА · блог Тао [Wordpress](#) · [FT](#) [FT](#)

Гостьовий допис По-Шень Ло в блозі Теренса Тао. Приводом стала хвиля відкритих листів на захист математичної спільноти, яка різко посилилась після того, як OpenAI оголосила розв'язок варіанту задачі Нав'є-Стокса з списку Міленіуму. Масштаб підписів: Лейденська декларація зібрала **понад 4000** підписантів, ініціатива Math and AI – **понад 7000**.

Окрема деталь, яку автор виносить у примітку сам: **100% прози написано людиною** у текстовому редакторі, без генерації, а згенеровано лише верстку і частину заголовків.

У тексті 2026 року така примітка вже потрібна – і це, мабуть, найкраща ілюстрація теми.

FT того ж дня заходить з іншого боку і пише про AI як про потужного, але проблемного співавтора в математиці – з наголосом на наслідках, коли алгоритм досягає мети з погано визначеними обмеженнями.

Чому це важливо. Математика тут – раннє попередження для всіх професій, де цінність у перевірюваному результаті. Питання, яке ставлять автори, звучить так: «що лишається людині, коли відповідь можна отримати дешево» – і відповідь у них про постановку задачі і відповідальність за вибір. Швидкість обчислення тут другорядна.

misc

- **Qwen Image 2.1** – нова версія графічної моделі ([Hacker News](#)).

У треді цікавіше за сам реліз спостереження: локальна генерація зображень зараз відчувається сильнішою за локальну генерацію коду, і одразу ж контраргумент – «моделі здаються прекрасними рівно в тому, де ти сам не фахівець».

- Іспанія заблокувала [archive.today](#) і дзеркала [Reclaimthenet](#) - рішенням Другої секції комісії з інтелектуальної власності, без судового рішення.
Подія від 18.09, у топ HN вийшла вчора.
- **Nobody pays for FOSS** [Seldo](#) - лонгрід на 5000 слів про економіку опенсорсу через модель яструбів і голубів.
Текст від 13.09, обговорення розгорілось тепер.
- **The LLMentalist Effect** [Softwarecrisis](#) - текст 2023 року, який знову вистрілив: аргумент, що чат-інтерфейс відтворює механіку холодного читання у екстрасенсів. Дата важлива, це не свіжа публікація.
- **Prompts aren't Real** [Evaluation](#) - доповідь про те, щоб перестати вручну підбирати промпти і будувати натомість пайплайни оцінки і оптимізації.
- **RSA-896** [Saweis](#) - розкладено на множники черговий розмір з давнього списку RSA.
- **Resident Evil 4 для GameCube** [GitHub](#) - повна побайтово ідентична декомпіляція в C/C++.