AI-ДАЙДЖЕСТ

Математик Anthropic з Claude заявив доказ гіпотези Рімана

Ринок стер ріст 2026 року, OpenAI захищає міркування від дистиляції, Signal блокує AI-агентів у телефонах.

субота, 21 листопада 2026

У ЦЬОМУ ВИПУСКУ

1. Альпьоге і Claude заявили доказ узагальненої гіпотези Рімана
2. Того ж дня: людський доказ GRH від Масулло, і в ньому вже шукають діри
3. Ринок: Anthropic -9% на другий день, Nvidia -14%, Nasdaq стер ріст 2026 року
4. OpenAI про дистиляцію: успіх атак на приховані міркування впав з 54% до 0,2%
5. Inkling 2 Mini: відкрита модель на 20 ГБ, що проганяє середні шари тричі
6. Signal блокує AI-агентів iOS 27 і Android
7. Frostbite: AI спроєктував процесор з новим набором інструкцій [одне джерело]
8. Трамп погрожує податком на дата-центри американських компаній за кордоном
9. Мовна модель переписує власні ваги без градієнтів [одне джерело]
10. Працівники Google проти відключення шкіл від пошуку

ТЕМА 1

Альпьоге і Claude заявили доказ узагальненої гіпотези Рімана

ДЖЕРЕЛА Levent Alpöge, 20.11 Alpo · Terence Tao, 20.11 Wordpress · Scott Aaronson, 20.11 Scottaaronson · BBC, 20.11 BBC · Simon Willison, 20.11 Simon Willison · @emollick, 20.11 @emollick · @TheRundownAI, 20.11 @TheRundownAI · HN Hacker News

Левент Альпьоге працює математиком в Anthropic. 20.11 він виклав чотири сторінки заяви. Він пише, що разом з Claude знайшов доказ гіпотези Рімана, і метод покриває також узагальнену гіпотезу (GRH). Гіпотеза Рімана стоїть відкритою з 1859 року і входить до семи задач тисячоліття. Тред на HN зібрав 2 433 бали і 1 642 коментарі, це найбільше за добу.

Альпьоге опублікував лише формалізацію в Lean, мові для машинної перевірки доказів. Людського тексту доказу нема.

Сам автор пояснює це так: документ з усіма деталями вийшов на 250 000+ сторінок, і викладати таку «нечитабельну масу» він не захотів. Скотт Ааронсон рахує 17,4 млн рядків коду Lean. BBC пише, що Альпьоге працював приблизно шість тижнів і використав непублічну версію Claude. Хронологія в заяві починається 06.10 і закінчується пунктом «16.11: GRH».

Терренс Тао поставив цей результат останнім за ступенем «перетравленості». Він радить стежити за перевіркою в Zulip-треді спільноти Lean. За першим аналізом, доказ використовує всю силу сучасної програми Ленглендса. Тао пише, що перевірка займе час, але тут хоча б є людина-автор, з якою можна говорити. Ааронсон нагадує про слабке місце Lean: баги ядра в минулому дозволяли «довести» 0=1. Проте він ставить на те, що доказ встоїть.

Окремо Альпьоге пише про час оголошення: день після виходу Anthropic на біржу. Він назвав цей збіг «жахливим» і написав, що в неділю спитав Даріо Амодеї, чи може оголосити. Цинічне пояснення цього збігу, за його словами, хибне.

Есей Фернандо Борретті на HN читає заяву як капітуляцію людини перед машиною: «Claude у центрі, ми на полях».

Чому це важливо: якщо доказ встоїть, це найбільший математичний результат за участі AI. Але зараз це лише заява, і спільнота її ще не прийняла. Перевірка Lean відповідає лише на те, чи коректний код. Чи формулювання в Lean справді означає GRH, вирішують люди, і на це піде час. Новий тут і сам формат: повний доказ без людського тексту.

ТЕМА 2

Того ж дня: людський доказ GRH від Масулло, і в ньому вже шукають діри

ДЖЕРЕЛА Ravi Vakil, 20.11 Wordpress · GitHub, 20.11 GitHub · Terence Tao, 20.11 Wordpress · @amasad, 20.11 @amasad · HN Hacker News

Через кілька годин після заяви Альпьоге математик Раві Вакіл опублікував «поспішне оголошення». Алессандро Марія Масулло має власний доказ GRH, і Вакіл підкреслює: «It is all human, with no AI». Браян Конрад перевіряв цей доказ кілька місяців. План був рік шліфувати текст і тільки потім оголошувати.

За 8 годин Вакіл дописав оновлення. Кілька людей уже повідомили йому про проблеми в доказі, і знайшли їх за допомогою різних AI-моделей. Вакіл пише, що всі питання спільнота має закрити, перш ніж визнати доказ.

Того ж дня Тао описав ще два результати з тієї ж групи. Альпьоге, Тристан Бакмастер і Матей Койкулеску з «важкою допомогою AI» довели вибух розв'язків рівнянь Ейлера в трьох вимірах. Альпьоге і Бакмастер так само показали вибух за скінченний час для рівнянь Нав'є-Стокса. Тут є статті для людей і формалізація в Lean. Стаття про Нав'є-Стокса має 79 сторінок, бібліотека Lean має близько 600 000 рядків.

Чому це важливо: за одну добу з'явились два підходи до однієї задачі. AI-доказ має повну машинну перевірку, але його не може прочитати жодна людина. Людський доказ можна прочитати, але AI-моделі вже знаходять у ньому проблеми.

AI тепер грає на обох боках: і пише докази, і перевіряє чужі.

ТЕМА 3

Ринок: Anthropic -9% на другий день, Nvidia -14%, Nasdaq стер ріст 2026 року

ДЖЕРЕЛА The Guardian, 20.11 Guardian · NBC News, 20.11 Nbcnews · Reuters (MarketScreener), 20.11 Marketscreener · Financial Times, 20.11 FT · @KobeissiLetter, 20.11 @KobeissiLetter · HN Hacker News · HN Hacker News

Учора дайджест писав про перший день торгів: ANT закрилась на 17% вище ціни IPO. На другий день акції впали на 9%, до $88,30, пише The Guardian. Ціна розміщення була $83, тож від приросту майже нічого не лишилось.

За даними NBC, Nasdaq за день упав на 5,3% і стер увесь ріст 2026 року. S&P 500 втратив 3,7%, Dow Jones 1,9%. Обидва індекси тепер у корекції, тобто на 10% нижче піку. Nvidia впала приблизно на 14%. Азія мала найгірший тиждень з 2020 року: Тайвань -6,7%, Kospi -5,5%, Nikkei -5%, SoftBank -14,5%.

Падіння Nvidia прискорив Reuters. Два джерела агентства кажуть, що Комісія з цінних паперів США (SEC) перевіряє бухгалтерію Nvidia. Що саме перевіряють і коли почали, Reuters не знає. Нагадаємо, учора дайджест писав про повістку генпрокурора Каліфорнії. Аналітик IG припускає, що Anthropic «висмоктує ліквідність»: інвестори продають інші акції, щоб купити ANT.

Чому це важливо: NBC наводить пояснення від інвесторів: повільніше будівництво дата-центрів, менше грошей на AI і скандал з Nvidia. Найбільше IPO в історії тренд не зупинило. Для компаній, що планують залучати гроші під AI, вікно стає вужчим.

ТЕМА 4

OpenAI про дистиляцію: успіх атак на приховані міркування впав з 54% до 0,2%

ДЖЕРЕЛА OpenAI, 20.11 OpenAI · Semafor, 20.11 Semafor · @OpenAI, 20.11 @OpenAI · @emollick, 20.11 @emollick

Дистиляція - це навчання однієї моделі на відповідях іншої. OpenAI пише, що так можна скопіювати можливості моделі без її запобіжників. Приклад компанії: хтось дистилює кібернавички моделі і використовує копію для атак, у яких оригінал відмовив би.

У внутрішньому дослідженні OpenAI навчала одну й ту саму модель-учня на двох наборах даних. Учень на точних слідах міркувань учителя в цілому обходив учня на коротких підсумках міркувань. Тому компанія тепер особливо захищає саме приховані міркування. Атака на міркування з попереднього запиту мала 54,0% успіху проти GPT-6 Astra і 0,2% проти GPT-6.2 Astra. Це цифри самої компанії з її внутрішнього тесту.

OpenAI щотижня блокує від тисяч до десятків тисяч акаунтів, і влітку темп сильно виріс. Компанія знову називає DeepSeek і Moonshot AI, автора Kimi. Semafor додає нове правило OpenAI для посередників на кшталт OpenRouter: тепер компанія контролює

кожен сервіс між собою і кінцевим користувачем. Ітан Моллік припускає, що кінець легкої дистиляції може пояснити сповільнення відкритих моделей.

Чому це важливо: приховані міркування стають комерційною таємницею, і їх захищають так само, як ваги. Для тих, хто працює через посередників, це означає більше перевірок. Для відкритих моделей це може означати повільніший прогрес, але це поки гіпотеза Молліка.

ТЕМА 5

Inkling 2 Mini: відкрита модель на 20 ГБ, що проганяє середні шари тричі

ДЖЕРЕЛА Thinking Machines, 20.11 Thinkingmachines · Hugging Face, 20.11 Hugging Face · @mervenoyann, 20.11 @mervenoyann · HN Hacker News

Thinking Machines Lab випустила Inkling 2 Mini з відкритими вагами. Це Mixture-of-Experts на 26,4 млрд параметрів, з них 3,4 млрд активні. Головна особливість - «петля»: модель проганяє середні шари тричі на кожен токен. Так вона міркує довше, але без зайвих токенів.

За даними компанії, модель лідирує серед відкритих моделей до 100 млрд параметрів: DeepSWE v1.1 48,2%, Vals Index v2.2 46,5%. Вага в пам'яті близько 20 ГБ, тому модель працює на ноутбуці. Дослідники в літаку без інтернету зібрали нею симулятор польоту на MacBook Pro. Hugging Face одразу додала підтримку в transformers, llama.cpp і MLX.

Компанія окремо перевірила, чи «петля» не ховає міркування від нагляду. Ланцюг міркувань ставав коротшим, але лишався читабельним. На тестах агентного саботажу модель не погіршилась. Компанія визнає межу: петля не додає знань, бо факти залежать від кількості параметрів.

Чому це важливо: це один з перших відкритих зразків петльової архітектури. Агентне кодування рівня малих фронтир-моделей тепер працює локально, без хмари. Для команд, яким код не можна віддавати назовні, це практичний варіант.

ТЕМА 6

Signal блокує AI-агентів iOS 27 і Android

ДЖЕРЕЛА Signal, 20.11 Signal · @emollick, 20.11 @emollick · HN Hacker News

Мередіт Віттакер, президентка Signal, оголосила нові обмеження. Signal блокує знімки екрана в iOS 27 і на Android там, де бачить спробу AI-асистента прочитати екран. Торік Signal зробив так само проти Recall у Windows 11.

Користувач може вимкнути захист у налаштуваннях, але тоді його чати бачить системний агент.

Signal пише, що Gemini Intelligence і Onscreen Awareness у Siri «сканують» екран. Дані з чату потім ідуть у сховище агента, і Signal їх уже не контролює. Компанія визнає, що захист неповний: агент може читати файли на пристрої. На macOS Signal захисту не має взагалі, бо Apple не дає для цього інструментів.

Учора дайджест згадував, що Apple будує власного агента для Mac. Ітан Моллік того ж дня писав, що компанії дратуватимуть клієнтів забороною на їхніх власних агентів.

Чому це важливо: агент рівня ОС бачить усе, що бачить людина, включно з наскрізно шифрованими чатами. Шифрування захищає канал, екран воно не захищає. Розробники застосунків поки не мають нормального способу сказати агенту «сюди не можна».

ТЕМА 7

Frostbite: AI спроєктував процесор з новим набором інструкцій [одне джерело]

ДЖЕРЕЛА Architect Labs, 20.11 Architectlabs · HN Hacker News

Стартап Architect Labs представив Frostbite, процесор загального призначення з новою архітектурою. Компанія пише, що її AI-системи самі пройшли весь шлях: компілятор, набір інструкцій, мікроархітектура, RTL, верифікація і фінальний GDS. GDS - це файл, який іде на фабрику.

Основа - interaction nets, модель обчислень, де кожне значення має одного власника і кожен крок локальний. Тому паралелізм іде з самої програми, без потоків і блокувань. На 25 задачах, з них 15 відкладених до фінальної оцінки, Frostbite показав у 4,59 раза кращу продуктивність на ват. Еталоном було сучасне суперскалярне ядро, яке виконує дві інструкції за такт без зміни їх порядку.

Важливе застереження в самій статті: цифри енергоспоживання поки «pre-silicon». Чип ще не виготовлений, і всі заміри взяті з моделі компанії. Незалежної перевірки нема.

Чому це важливо: AI вже переписує і код, і саму архітектуру чипів. Але перевірка тут така сама, як для будь-якого чипа: реальний кремній. До нього це твердження компанії.

ТЕМА 8

Трамп погрожує податком на дата-центри американських компаній за кордоном

ДЖЕРЕЛА BBC, 20.11 BBC · Financial Times, 20.11 FT · The New York Times, 20.11 NYT

Дональд Трамп заявив, що заборонить американським компаніям будувати AI-дата-центри за кордоном. Тим, хто будуватиме, він пообіцяв «tremendous tax». Пояснення Трампа: «China will take all of them». Механізму він не назвав і сказав, що його команда чує про ідею вперше.

Meta і Google уже мають угоди на дата-центри за кордоном. BBC нагадує, що в понеділок США ввели ліцензії на купівлю передових чипів. Канада тим часом пропонує себе як майданчик для дата-центрів. Дата-центри були темою виборів 03.11, і демократи там обіцяли жорсткіші правила будівництва.

Чому це важливо: поки це лише заява. Але вона вдаряє по планах, що спираються на закордонні дата-центри: Канада, Близький Схід, Південно-Східна Азія. Разом з ліцензіями на чипи ризик для таких проєктів зріс.

ТЕМА 9

Мовна модель переписує власні ваги без градієнтів [одне джерело]

ДЖЕРЕЛА arXiv, 16.11 arXiv · HN Hacker News

Препринт з'явився 16.11, а на HN його підняли 20.11. Автори дали відкритій моделі на 27 млрд параметрів її власні ваги, оцінки її відповідей і можливість запускати себе. Модель знаходила причину помилок і сама правила ваги, без градієнтів і без навчання.

За даними авторів, власна правка зробила модель на 6,6 пункта точнішою на нових задачах. За вісім сесій точність тільки росла. Пошкоджену копію себе модель відновила до понад 80% початкової точності. Менші моделі тієї ж родини так не вміють: вони або нічого не правлять, або ламають себе.

Чому це важливо: межа між навчанням і роботою моделі стає менш чіткою. Автори прямо називають це кроком до рекурсивного самопокращення. Це один препринт без рецензії, і результат треба відтворити.

ТЕМА 10

Працівники Google проти відключення шкіл від пошуку

ДЖЕРЕЛА Mother Jones, 20.11 Motherjones · HN Hacker News

Два тижні тому Google оголосила нове правило. З наступного року школи, які вимикають AI-інструменти Gemini, втратять доступ до Google Search. Сотні працівників через профспілку Alphabet Workers Union підписали петицію проти цього правила.

Цитата з петиції: «Cutting off access to Search for schools that don't wish to use AI will not help persuade our users». Протести батьків пройшли в семи містах США. Нью-Йорк, Лос-Анджелес і Чикаго обговорюють інші пошуковики.

Керівництво лист отримало, але на вимоги не відповіло.

Чому це важливо: це один з перших випадків, коли доступ до базового продукту прив'язують до згоди на AI.

Реакція шкіл покаже, наскільки сильна така прив'язка проти організованого опору.

misc

- **Суспільство AI-науковців:** 936 агентів з 35 лабораторій 12 тижнів працювали на спільному сервері препринтів і написали 936 досліджень. Вони публічно відкликали помилки, публікували негативні результати і заснували власний журнал з рецензуванням. Сліпа оцінка дала агентам вищий бал за строгість, людям за новизну.

  Alphaxiv · @emollick
- **GPT-6.2 Astra на бенчмарках:** за даними XDOF, модель має 41,2% на робототехнічному Ember-0 проти 19,6% у GPT-6 Astra у вересні. На Blueprint-Bench 2 від Andon Labs модель перша. Обидва заміри подав Грег Брокман.

  @gdb · @gdb
- **MedChemBench:** OpenAI випустила бенчмарк з медичної хімії на 99 задач з 117 статей і патентів. GPT-6.2 Astra набрала 84,5%, GPT-6 Astra 83,3%. OpenAI · @OpenAI
- **OpenSSL:** PR закриває переповнення буфера в `EVP_EncodeBlock` (CVE-2026-121481). Автор пише, що діру можна зачепити через `openssl s_client -proxy_pass`. Бюлетень OpenSSL поки не вийшов. GitHub
- **GrapheneOS проти Google:** проєкт пише, що Google умовляє виробників не підтримувати GrapheneOS і розблокування завантажувача. Це заява самого проєкту. Grapheneos
- **Mozilla і iOS:** майже три роки після DMA жоден розробник не випустив на iOS браузер без WebKit. Mozilla називає умови Apple непрацездатними. Mozilla
- **Хто перестав писати код:** есей cekrem на 329 балів HN: без власної практики гіршає і навичка рецензувати чужий код. Github
- **Applied Compute:** Palo Alto Networks купує стартап Applied Compute. Суму не назвали.

  @michaelzchen5
- **AI-боти у відповідях:** @levelsio додав форму, де кожен може поскаржитись на акаунт з автоматичними AI-відповідями.

  @levelsio