AI-ДАЙДЖЕСТ

Anthropic переписує правила context engineering

Гайд Anthropic для Claude 5 називає застарілою половину практик цього репо.

неділя, 26 липня 2026

У ЦЬОМУ ВИПУСКУ

1. «Нові правила context engineering» - Anthropic, 25.07 Офіційний блог, 216 балів на HN. Інструкція по експлуатації моделі.

Головна цифра: **прибрали понад 80% системного промпту Claude Code** для Opus 5 і Fable 5 - «без вимірюваної втрати на кодових евалах».

Шість зсувів, з яких три б'ють по типовій агентній обгортці:

• **Правила стають судженням.** Замість «ніколи не пиши багатоабзацні докстрінги» - «пиши код, який читається як код навколо». Жорсткі заборони тепер шкодять.

• **Приклади стають кращими інтерфейсами.** Робити сам тул зрозумілим: явні параметри, перелічені опції.

• **Все наперед стає progressive disclosure.** Вантажити скіли й довідники тоді, коли вони знадобились.

• Плюс: прибрати дублювання між промптом, скілами і описами тулів; ручні нотатки в CLAUDE.md замінити автоматичною пам'яттю; посилатись на код і артефакти.

Рекомендують `claude doctor` - він оптимізує наявний контекст сам.

Чому це важливо. Типовий великий системний промпт - це рівно той front-loaded моноліт, який тут названо вчорашнім днем. Персона, жорсткі правила, довідники й інжекція пам'яті вантажаться в кожну сесію одразу. Скіли й тули вже вміють

підвантажуватись ліниво, промпт зазвичай ні. Окремо про CLAUDE.md: рекомендація - тримати файл легким, з акцентом на граблях і рецептах. Вичерпний опис проєкту там зайвий.

2. Open weights: за добу з листа виросла війна Вчора згадувався лист із 17 підписантами і перший пост Дженсена в X - сьогодні це головна сварка обох кураторських X-листів, і тон різко особистий.

• **@Suhail** - Anthropic: «Підтримувати опенсорс не означає, що ти мусиш відкривати ваги. Якщо ви так себе заспокоюєте в Anthropic Slack - це доволі тривожно. Вам не треба відкривати нічого, але витрачати десятки мільйонів на лобі й думсдей - інша річ».

• **@migueldeicaza**: «Енергія "затягни драбину за собою". Anthropic на опенсорсі свій бізнес побудувала - а світу відкриті ваги, значить, зась».

• **@packyM** різкіше: «ідеологічно захоплена організація, що шукає регуляторного захоплення».

• **@DavidSacks**: «Ніхто не каже, що весь софт має бути опенсорсом. Кажуть, що відкриті ваги мають бути дозволені».

• Контрапункт від **@emollick**, і він найтверезіший: підписанти вкладають у «підтримку відкритих ваг» дуже різні речі - у самому листі є застереження, наприклад, про дистиляцію. Консенсус радше риторичний.

Чому це важливо: компанію, чиї моделі стоять у щоденних робочих інструментах, публічно розносять за позицію по відкритих вагах. На доступ і ціни це поки не впливає ніяк, але це той конфлікт, з якого виростає регуляція, а вона вже впливає. Варто тримати на радарі.

3. UK AISI виміряла Kimi K3 - і це аргумент у тій самій суперечці Поки в X сваряться словами, британський AISI разом з CAISI опублікували цифри по кібер-здібностях відкритої китайської моделі. Це NIST, не блог.

• **ExploitBench: 32%** проти ~48-54% у фронтірних західних моделей • Кібер-полігон «The Last Ones»: дійшла до **кроку 17 з 32**, топові US-моделі - у середньому 28.5 • **На жодному з 41 завдання не змогла** розробити експлойт з довільним виконанням коду • Але: обійшла GLM-5.2 (найкращу відкриту станом на червень) - 32% проти 24% • І тривожне: **в одній спробі з десяти пройшла всі 32 кроки**. Плюс висновок прямим текстом - «запобіжники Kimi K3 дозволяють допомогу в розробці агентних кібер-експлойтів».

Чому це важливо: ось це і є нормальна форма аргументу в суперечці №2 - вимірювання замість «я вважаю». Відкрита модель відстає від фронтіру, але вже автономно ламає слабко захищені корпоративні системи. Практичний висновок приземлений: усе, що виставляється назовні, тепер тестують агенти. Вчорашній сюжет з GitHub-токеном у камері Hanwha - з тієї ж опери.

4. Порада від Бориса Черні, яка суперечить звичній практиці З YC Startup School, цитує @borisjabes: **«Видаляйте свої системні промпти і скіли раз на пів року і дивіться, що зроблять нові моделі»**.

Це людина з команди Claude Code, і тезу варто читати разом з пунктом 1: обгортки часто компенсують слабкості моделі, якої вже нема. Для місячного промпту це ще рано, але правило хороше: раз на пів року перевіряти, чи потрібні поставлені милиці.

5. YC Startup School: Дженсен про «керованість агентів» Стадіон, повний засновників; виступали Дженсен Хуанг, Борис Черні, Джефф Дін, Гаррі Тан, Кратсіос. Другий кураторський Х-лист учора був майже цілком про це.

Найцінніше з переказів - теза Дженсена: **«Керованість агентів буде наступним проривом»**, з поясненням, що зміна одного слова в md-файлі може суттєво змінити поведінку агента. Плюс його ж про засновництво: «F1-команди будують машину під пілота. Засновник - пілот».

Чому це важливо: «одне слово в md-файлі» - це буквально те, на чому тримаються скіли й файли пам'яті агента. Правки таких текстів варто переглядати як diff, бо рядок змінює поведінку так само, як зміна коду.

6. DeepSeek: витік з зустрічі з інвесторами 112 балів, транскрипт зустрічі Лян Вень-фена (22.07) виклали на GitHub. DeepSeek поставила на паузу другий раунд, і гроші тут ні до чого.

Пряма цитата: **«Дефіциту коштів чи ресурсів точно нема - усе це доступне»**. Проблема - конвертувати гроші в залізо: потрібно ~**200 000 Huawei 950**, отримали **16 000**. Розрив у обчисленнях з US-лабами він називає «на порядок».

Оговорка: це витік, не офіційна заява; коментатори вважають транскрипт справжнім, Bloomberg пише, що Ляна розлютив сам факт витоку.

Чому це важливо: контекст до пунктів 2-3. Китайські відкриті моделі відстають через доступ до заліза, і санкції тут працюють ефективніше за будь-яку заборону ваг.

7. Локальний фронтір: 753B на настільній машині @MichaelDell показує, як Тобі Лютке (Shopify) ганяє **GLM-5.2, 753B параметрів, локально на Dell Pro Max з GB300 - 40 токенів/с**. Без дата-центру, без API.

Чому це важливо: пряма ілюстрація, чому суперечка про відкриті ваги не академічна. Коли фронтірна модель крутиться під столом, «дозволити/заборонити» стає питанням поліції.

8. Android може прикрити ADB на самому пристрої 890 балів - топ доби на HN. Google обговорює обмеження ADB через localhost, залишивши прив'язку до конкретних інтерфейсів.

Тред майже одностайно скептичний: це б'є по Shizuku, Canta й автоматизації, а зачіпає 0.1% користувачів, які «і так знають, що роблять». Паралелі проводять з Manifest V3.

Чому це важливо: патерн вартий уваги - платформи послідовно ріжуть саме той рівень доступу, на якому живуть домашні автоматизації. Автоматизацій на десктопі це поки не зачіпає.

9. «Інженерний менеджмент після того, як код подешевшав» 125 балів, 180 коментарів. Прямий спадкоємець вчорашньої теми №10 («якщо кодинг вирішено, чому софт гіршає?»), але з боку управління: що робити, коли написання коду перестало бути вузьким місцем.

Чому це важливо: тема тримається в топі другу добу поспіль різними текстами - сигнал, що індустрія переварює саме це. Для CTO це ближче до реальних рішень, ніж бенчмарки з пункту 1.

10. Клем з Hugging Face виставив OpenAI рахунок Продовження вчорашньої №6 (скепсис навколо «AI-хакера»). Клеман Делянг публічно попросив OpenAI: **опублікувати трейси «rogue»-агентів**, щоб спільнота могла їх вивчити, і **виділити $100M компюту** на захисні дослідження.

Чому це важливо: правильна реакція на інцидент - вимагати артефакти. Вчора історія розсипалась при перевірці; сьогодні є конкретна вимога, за виконанням якої видно, чи була там взагалі подія.

Misc • @emollick про сикофантію: «Моделі мають краще вміти казати, що ідея - тупа». Модельна улесливість стає проблемою рівня індустрії • Hannah Fry отримала премію Leelavati за популяризацію математики (570 балів)

• Bitchat Дорсі переїхав на Radicle (221) - після вимоги уряду Індії зняти його з GitHub, це вчорашній сюжет • 28.9M-параметрова LLM на мікроконтролері за $8 (108)

• Debian голосує за три пропозиції щодо використання LLM у проєкті (109)

• Tile настільки погано захищений, що це «фіча для сталкерів» (151)

• @amasad довів шаховий движок на одній дотренованій LLM до ~1200 Elo, ціль 2000+, без шахового движка всередині
