

## DOTE 6635: Artificial Intelligence for Business Research (Spring 2026)

# What's New in AI

Renyu (Philip) Zhang

1

## What Happened Since We Last Met?

### Nature's 10

Ten people who helped shape science in 2025

8 December 2025



#### Liang Wenfeng: Tech disruptor

After making his name in investing, a Chinese finance wizard founded DeepSeek.

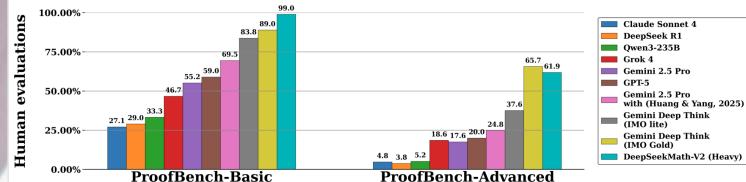
Article | [Open access](#) | Published: 17 September 2025

#### DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning

[Daya Guo](#), [Dejian Yang](#), [Haowei Zhang](#), [Junxiao Song](#), [Peiyi Wang](#), [Qihao Zhu](#), [Runxin Xu](#), [Ruoyu Zhang](#), [Shirong Ma](#), [Xiao Bi](#), [Xiaokang Zhang](#), [Xingkai Yu](#), [Yu Wu](#), [Z. F. Wu](#), [Zhibin Gou](#), [Zhihong Shao](#), [Zhoushu Li](#), [Ziyi Gao](#), [Aixin Liu](#), [Bing Xue](#), [Binxuan Wang](#), [Bochao Wu](#), [Bei Feng](#), [Chengda Lu](#), ... [Zhen Zhang](#) + Show authors

[Nature](#) 645, 633–638 (2025) | [Cite this article](#)

320K Accesses | 173 Citations | 800 Altmetric | [Metrics](#)



Contest	Problems	Points
IMO 2025	P1, P2, P3, P4, P5	83.3%
CMO 2024	P1, P2, P4, P5, P6	73.8%
Putnam 2024	A1 ~ B4, B5, B6	98.3%

Table 1 | Problems in gray are **fully solved**, while underlined problems received **partial credit**.

Jan 6, 2026

2

**karpathy**

Home Blog

**2025 LLM Year in Review**

20 Dec, 2025

**2025 LLM Year in Review**

By Andrej Karpathy | Dec 19, 2025



2025 has been a strong and eventful year of progress in LLMs. The following is a list of personally notable and mildly surprising "paradigm changes" - things that altered the landscape and stood out to me conceptually.

<https://karpathy.bearblog.dev/year-in-review-2025/>

Jan 6, 2026

3

**What Happened Since We Last Met?****1. Reinforcement Learning from Verifiable Rewards (RLVR)****2. Ghosts vs. Animals / Jagged Intelligence****3. Cursor / new layer of LLM apps****4. Claude Code / AI that lives on your computer****5. Vibe coding****6. Nano banana / LLM GUI**

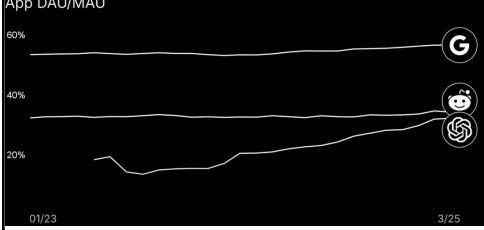
**TLDR.** 2025 was an exciting and mildly surprising year of LLMs. LLMs are emerging as a new kind of intelligence, simultaneously a lot smarter than I expected and a lot dumber than I expected. In any case they are extremely useful and I don't think the industry has realized anywhere near 10% of their potential even at present capability. Meanwhile, there are so many ideas to try and conceptually the field feels wide open. And as I mentioned on my Dwerkesh pod earlier this year, I simultaneously (and on the surface paradoxically) believe that we will both see rapid and continued progress *and* that yet there is a lot of work to be done. Strap in.

**豆包DAU破亿，成字节史上推广费用最少的破亿产品**

界面新闻 2025年12月24日 21:56

**AI Now**

App DAU/MAU

**What Happened Since We Last Met?**

@CryptoDefLord NE 🇺🇸 🇲🇽 🇫🇷 @CryptoDefLord · 7月1日

Transfer window now open for tech guys. This Asian man Jiahui Yu was traded from OpenAI to Meta for \$100m.

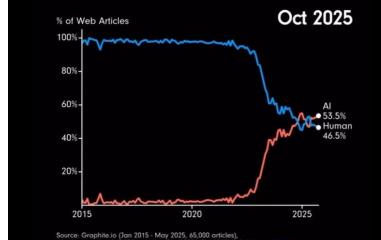
Who is next?

科技人才转会窗口现已开放。亚洲球员于嘉辉以1亿美元的价格从OpenAI转会至Meta。

下一个是谁？

**Compensation**

\$280K – \$400K + Offers Equity

**The Rise of AI-Generated Content**

manus

The general AI agent

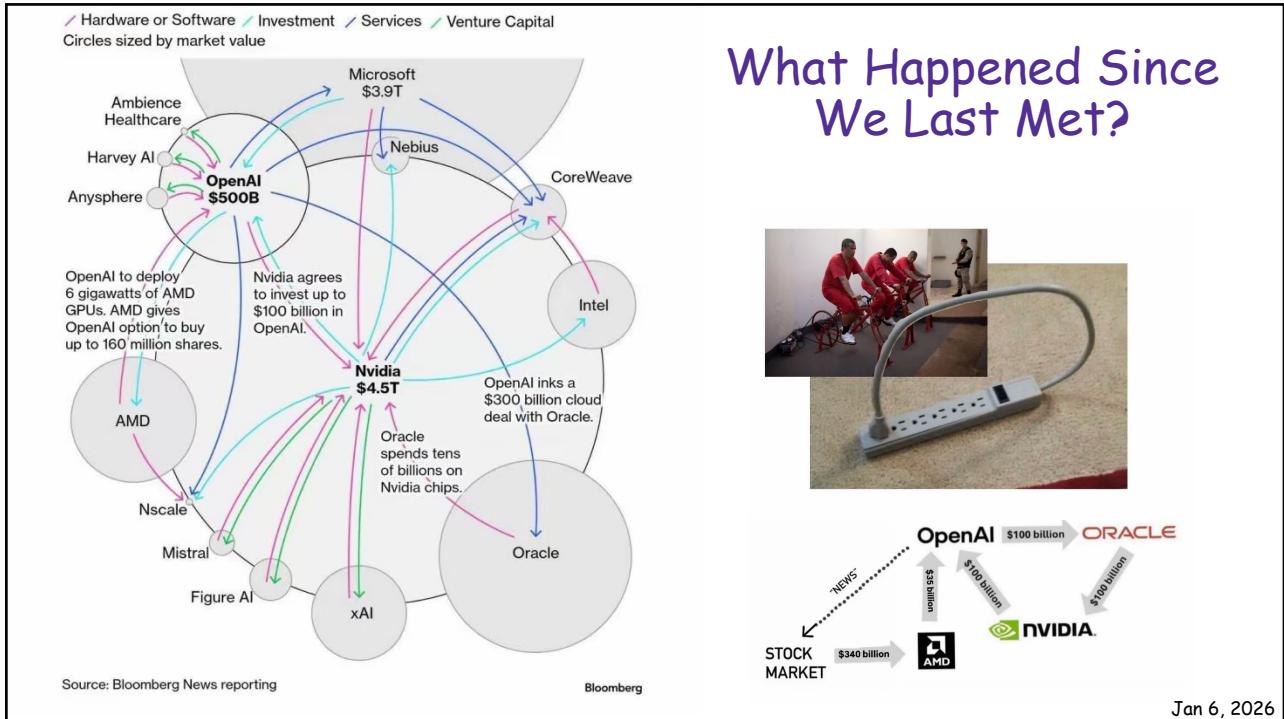
**Meta just acquired a Chinese-founded AI startup for \$2B. Here's why that matters**

AI firm Manus claims its bot can make decisions with far less prompting than rivals

Jenna Benchreif - CBC News · Posted: Dec 30, 2025 3:36 PM EST | Last Updated: December 31, 2025

Jan 6, 2026

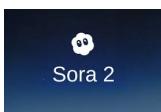
4



5

## Maintrack: Facilitates Humans to Leverage Compute

- AI Coding: Connecting human & compute
- Deep Research: Automation of information acquisition & processing
- World Model: Data-driven simulation of physical world
- AI Scientist: Automated hypothesis generation and validation

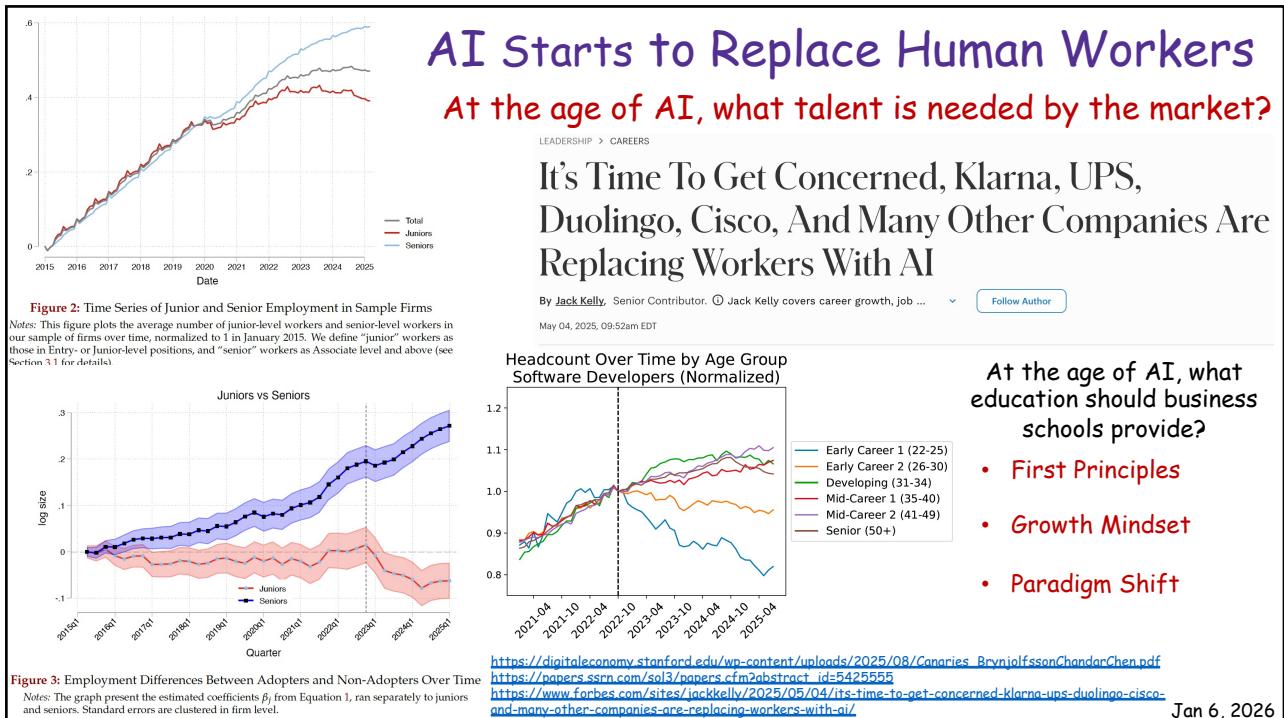


Accelerating scientific breakthroughs with an AI co-scientist

February 19, 2025 - Juraj Gottweis, Google Fellow, and Vivek Natarajan, Research Lead

Jan 6, 2026

6



7

## Complete Replication of a PNAS Paper

### Universal vote-by-mail has no impact on partisan turnout or vote share

<https://github.com/andybhall/vbm-replication-extension>

Daniel M. Thompson<sup>a,1</sup>, Jennifer A. Wu<sup>b,1</sup>, Jesse Yoder<sup>a,1</sup>, and Andrew B. Hall<sup>a,2</sup>

<sup>a</sup>Department of Political Science, Stanford University, Stanford, CA 94305; and <sup>b</sup>Stanford Institute of Economic Policy Research, Stanford University, Stanford, CA 94305

Edited by Douglas S. Massey, Princeton University, Princeton, NJ, and approved May 6, 2020 (received for review April 15, 2020)

In response to coronavirus disease 2019 (COVID-19), many scholars and policy makers are urging the United States to expand voting-by-mail programs to safeguard the electoral process. What are the effects of vote-by-mail? In this paper, we provide a comprehensive design-based analysis of the effect of universal vote-by-mail—a policy under which every voter is mailed a ballot in advance of the election—on electoral outcomes. We collect data from 1996 to 2018 on all three US states that implemented universal vote-by-mail in a staggered fashion across counties, allowing us to use a difference-in-differences design at the county level to estimate causal effects. We find that 1) universal vote-by-mail does not appear to affect either party's share of turnout, 2) universal vote-by-mail does not appear to increase either party's vote share, and 3) universal vote-by-mail modestly increases overall average turnout rates, in line with previous estimates. All three conclusions support the conventional wisdom of election administration experts and contradict many popular claims in the media.

vote-by-mail | elections | COVID-19 | partisanship

The coronavirus disease 2019 (COVID-19) pandemic threatens the 2020 US election. Fears that the pandemic could deter many people from voting—or cause them to become infected if they do vote—have spurred calls for major electoral reforms. As election administration experts Nathaniel Persily and Charles Stewart put it, “The nation must act now to ensure

receive an absentee ballot, while stopping short of moving to universal VBM; as such, by studying a more dramatic version of the recommended policies, our paper provides a useful upper bound related to these discussions.\* While a large literature in political science studies various forms of convenience voting—see SI Appendix, Table SI for a full review—there has not been any comprehensive analysis of VBM that employs clear designs for causal inference to estimate effects on partisan outcomes.<sup>†</sup> The existing research supporting the neutral partisan effects of VBM compares turnout in Oregon before and after it implemented

**Significance**

In response to COVID-19, many scholars and policy makers are urging the United States to expand voting-by-mail programs to safeguard the electoral process, but there are concerns that such a policy could favor one party over the other. We estimate the effects of universal vote-by-mail, a policy under which every voter is mailed a ballot in advance of the election, on partisan election outcomes. We find that universal vote-by-mail does not affect either party's share of turnout or either party's vote share. These conclusions support the conventional wisdom of election administration experts and contradict many popular claims in the media. Our results imply that the partisan outcomes of vote-by-mail elections closely resemble in-person elections, at least in normal times.

**Model Information**

- Model: Claude Opus 4.5 (claude-opus-4-5-20251101)
- Interface: Claude Code CLI
- Date: January 2026

Jan 13, 2026

8

# Complete Replication of a PNAS Paper

<https://github.com/andybhall/vbm-replication-extension/blob/main/INSTRUCTIONS.md>

## AI-Generated Academic Paper: Replicating and Extending "Universal Vote-by-Mail Has No Impact on Partisan Turnout or Vote Share"

### Project Overview

You are tasked with producing a complete academic political science paper by replicating and extending Thompson, Wu, Yoder, and Hall (2020), published in PNAS. The original paper used a difference-in-differences design to estimate the causal effects of universal vote-by-mail (VBM) on partisan electoral outcomes, finding null partisan effects and a modest (~2 percentage point) increase in overall turnout.

#### Your task:

1. Replicate the original findings using the authors' published replication data and code
2. Extend the analysis by collecting new data for the same three states (California, Utah, Washington) through 2024
3. Test whether the null partisan findings hold in the post-COVID era

Original paper: <https://www.pnas.org/doi/10.1073/pnas.2007249117>

Original replication materials: <https://github.com/stanford-dpl/vbm>

### IMPORTANT: Stop-and-Check Points

Throughout this project, there are mandatory STOP AND CHECK points marked with 🔴. At each of these points, you must:

1. Summarize what you have completed
2. Present key outputs for review
3. List any issues or concerns
4. Wait for human approval before proceeding

Do not proceed past a 🔴 checkpoint without explicit approval.

Jan 13, 2026

9

# Complete Replication of a PNAS Paper

[https://github.com/andybhall/vbm-replication-extension/blob/main/CLAUDE\\_CODE\\_PROMPTS.md](https://github.com/andybhall/vbm-replication-extension/blob/main/CLAUDE_CODE_PROMPTS.md)

### Phase 0: Project Setup

#### Initial Prompt:

I want to replicate and extend Thompson et al. (2020) "Universal Vote-by-Mail Has No Impact on Partisan Turnout or Vote Share" from PNAS. The paper studies California's Voter's Choice Act. I have the original replication data. Please set up the project structure and review the original materials.

### Phase 1: Literature Review

#### Prompt:

Approved, proceed to Phase 1: Literature Review

### Phase 2: Replication

#### Prompt:

Approved, proceed to Phase 2

### Phase 3: Extension Data Collection

#### Prompt:

Approved, proceed to Phase 3

### Phase 4: Data Preparation

#### Prompt:

Approved, proceed to Phase 4

Jan 13, 2026

10

## Complete Replication of a PNAS Paper

[https://github.com/andybhall/vbm-replication-extension/blob/main/CLAUDE\\_CODE\\_PROMPTS.md](https://github.com/andybhall/vbm-replication-extension/blob/main/CLAUDE_CODE_PROMPTS.md)

### Phase 5: Extension Analysis

Prompt:

Approved, proceed to Phase 5

### Phase 6: Paper Writing

Prompt:

Approved, proceed to Phase 6

### Phase 7: Final Deliverables

Prompt:

Approved, proceed to Phase 7

### Bug Fix Session

Prompt:

Can you take a look at the event study? It seems like something may be wrong with the turnout one since it's not showing the same positive effect as all the regressions (which I trust more) are showing

Jan 13, 2026

11

## IPOs of Zhipu AI and MiniMax

### China's AI Tigers Roar on HKEX

Historic back-to-back IPOs of Zhipu AI & MiniMax signal a new era for global AI capital markets

#### ZHIPU AI

02513.HK

JAN 8, 2026

MARKET CAP (DAY 1)

**HK\$57.9B**

DAY 1 GAIN

**+13.2%**

IPO PRICE

**HK\$116.2**

FUNDS RAISED

**US\$558M**

OVERSUBSCRIPTION

**1,159x**

Focus: Enterprise AI, Tsinghua Spin-off

#### MINIMAX

0100.HK

JAN 9, 2026

MARKET CAP (DAY 1)

**HK\$1,067B**

DAY 1 GAIN

**+109%**

IPO PRICE

**HK\$165.0**

FUNDS RAISED

**US\$619M**

OVERSUBSCRIPTION

**1,837x**

Focus: Consumer AI Apps, Global Reach

<https://www.scmp.com/tech/tech-trends/article/3339301/minimax-and-zhipus-stellar-hong-kong-ipos-supercharge-chinas-ai-ambitions>



#### HISTORIC MILESTONE

MiniMax becomes first AI company globally to exceed HK\$100B market cap on IPO day.



#### RECORD SPEED

Fastest AI IPO record worldwide—just 2 years from founding to listing.



#### MASSIVE DEMAND

Combined retail investors >600k; Record institutional subscription for MiniMax.



#### GLOBAL CONFIDENCE

Backed by ADIA, Alibaba & Mirae Asset, signaling strong global trust.  
Source: HKEX, Company Filings, Reuters, Bloomberg | January 2026

Jan 13, 2026

12

# DeepSeek to Disrupt the World Again?

• News • Artificial Intelligence

## Insiders Say DeepSeek V4 Will Beat Claude and ChatGPT at Coding, Launch Within Weeks

DeepSeek's upcoming V4 model could outperform Claude and ChatGPT in coding tasks, according to insiders—with its purported release nearing.



By [Jose Antonio Lanz](#)

Edited by [Andrew Hayward](#)

Jan 10, 2026

4 min read

Capability	Description
Long Code Prompts	A significant breakthrough in handling and parsing extremely long code prompts, offering a major advantage for developers working on complex software projects <a href="#">6</a> .
Data Pattern Understanding	Improved ability to understand data patterns across the entire training pipeline, with no observed degradation in performance <a href="#">6</a> .
Reasoning Ability	The model's outputs are described as more logically rigorous and clear, indicating stronger reasoning capabilities and greater reliability for complex tasks <a href="#">6</a> .

<https://tech.yahoo.com/ai/articles/insiders-deepseek-v4-beat-claude-205234497.html>

Jan 13, 2026

13

 Axiom  
5,629 followers  
1d · ①

AxiomProver solved all 12 Putnam 2025 problems. Today we're releasing the Lean proofs.

We also put together our take on the problems, proof visualizations, and a look at how humans and AI approached things differently. Lots of fun math and Lean.

We sorted our findings into three categories:

Some problems are trivial for humans but tedious for AI. Calculus (A2, B2) and combinatorial constructions (A5) fall here. A positivity lemma humans find obvious takes chunks of Lean to satisfy the formal checker. After a combinatorial construction that's natural to see, our A5 formalization is 2054 lines and took 518 minutes. Combinatorics are friendly beasts to AI and analysis formalization could make one grumpy.

Some problems were surprisingly within reach. AxiomProver doesn't have a Euclidean geometry engine, so we didn't expect it to get B1. It reduced everything to symbols and pushed through. It also handled the combinatorial game theory question (A3) cleanly -- usually a tough domain for AI as we observed in last two years' model performance in the IMO.

Some problems got completely different solutions. On A4, our mathematicians thought algebraically. AxiomProver went geometric. On B4, we found one picture that settles it elegantly. AxiomProver chose 1061 lines of combinatorial bookkeeping. No picture, just grinding through cases until the result falls out.

The one that stunned us: A6. None of our in-house mathematicians finished it. AxiomProver did.

What's hard for humans and what's hard for machines are different questions. Understanding that gap matters as we are closing it.

Grothendieck uses the rising sea metaphor to say, don't attack the problem, raise the water level until it surrounds the land.

When will AI bring in definitions, build theories, choose the right abstraction to make problems dissolve?

We are not there yet. But days like this makes it feel closer.

<https://github.com/AxiomMath/putnam2025>

## AI for Math by AxiomProver

Problem	Time to Solve	Tokens Used	Proof Length (Lines)	Theorems Generated	Tactics Used
A1	110 minutes	7,000,000	652	23	561
A2	185 minutes	6,000,000	556	26	581
A3	165 minutes	8,000,000	1,333	78	1,701
A4	107 minutes	8,000,000	960	32	1,107
A5*	518 minutes	9,100,000	2,054	52	3,074
A6*	259 minutes	16,000,000	588	28	670
B1	270 minutes	7,000,000	1,386	49	1,841
B2	65 minutes	2,000,000	417	28	325
B3	43 minutes	2,900,000	340	11	422
B4*	112 minutes	249,000	1,061	23	1,433
B5	354 minutes	18,000,000	1,495	66	1,967
B6*	494 minutes	21,000,000	1,019	30	1,052

"\*" means AxiomProver solved the problem in the following day.

Jan 13, 2026

14

## Human Performance of Putman 2025

### SCORE-TO-RANK CORRESPONDENCE

A participant is said to have rank  $n$  if  $n-1$  participants scored higher. There were 3988 participants in the competition.

Score	Rank								
90	1	64	32	49	96	35	202	21	505
87	2	63	34	48	100	34	208	20	546
81	3	62	37	47	106	33	220	19	641
80	4	61	38	46	109	32	238	18	669
78	6	60	45	45	111	31	254	17	686
75	9	59	52	44	116	30	272	16	698
74	10	58	58	43	119	29	295	15	713
72	11	57	61	42	126	28	313	14	750
71	13	55	64	41	137	27	328	13	820
70	17	54	66	40	150	26	337	12	905
69	18	53	69	39	173	25	349	11	1090
67	23	52	76	38	180	24	374	10	1171
66	26	51	79	37	194	23	404	9	1274
65	30	50	87	36	198	22	449	8	1305

### FREQUENCY DISTRIBUTION OF PROBLEM SCORES OF THE TOP 504 CONTESTANTS

Score	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	B6
10	449	93	29	112	5	0	375	81	75	95	1	0
9	28	10	1	11	3	1	25	76	21	39	0	1
8	1	99	8	15	3	0	11	17	9	5	0	0
7	0	11	2	13	0	0	4	7	0	0	0	2
6	0	0	0	1	0	0	0	0	0	0	0	1
5	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0
3	4	36	1	27	0	0	2	12	43	23	0	2
2	13	88	8	10	3	0	48	21	70	32	82	2
1	3	24	29	52	0	9	16	54	80	14	81	0
0	1	85	154	34	163	65	10	84	43	77	41	29
NA	5	58	272	229	327	429	13	152	163	219	299	467

About 4'000 students participated in Putman 2025.

<https://maa.org/wp-content/uploads/2025/03/2024-William-Lowell-Putnam-Competition-Announcement-of-Winners.pdf>

Jan 13, 2026

15

**opencode**

← Post

 Yuchen Jin   @YuchenJin\_UW

Anthropic blocked Claude subs in third-party apps like OpenCode, and reportedly cut off xAI and OpenAI access.

Claude and Claude Code are great, but not 10x better yet. This will only push other labs to move faster on their coding models/agents.

DeepSeek V4 is rumored to drop soon, with stronger coding than Claude and GPT. Whale Code soon?

Competition is great for users. Traduire le post

19:17 - 09 janv. 26 · 166K Vues

59 reposts 21 citations 1461 J'aime

202 Signets

Les réponses les plus pertinentes

 Kylie Robison  @kyliebytes · 13 h

Postez votre réponse

## Open or Close in the Age of AI



### Cowork: Claude Code for the rest of your work



Guohao Li   @guohao\_li · 1月13日

Anthropic Claude Cowork just killed our startup product 😅 So we did the most rational thing: open-sourced it.

Meet Eigeant  [github.com/eigent-ai/eige...](https://github.com/eigent-ai/eige...)

Anthropic Claude Cowork刚刚杀死了我们的创业产品 😅。所以我们做了最理智的事情：将其开源。

这就是 Eigeant  [github.com/eigent-ai/eige...](https://github.com/eigent-ai/eige...)



Claude   @claudeai · 1月13日

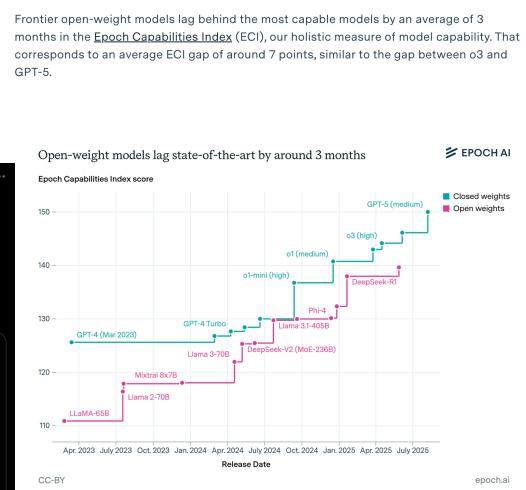
Introducing Cowork: Claude Code for the rest of your work.

Cowork lets you complete non-technical tasks much like how developers use Claude Code.

推介 Cowork：适用于你其余工作的 Claude Code。

Cowork 让你能够像开发者使用 Claude Code 那样，高效完成非技术性任务。

Open-weight models lag state-of-the-art by around 3 months on average



[https://www.reddit.com/r/Anthropic/comments/1q8z1to/anthropic\\_blocking\\_access\\_to\\_thirdparty\\_apps/](https://www.reddit.com/r/Anthropic/comments/1q8z1to/anthropic_blocking_access_to_thirdparty_apps/)

Jan 20, 2026

16

## Qwen Agent on Alibaba's Ecosystem



<https://wallstreetcn.com/articles/3763338>; <https://www.linkedin.com/pulse/building-universal-commerce-protocol-ucp-ilya-grigorik-ekemc>



Jan 20, 2026