

# Agent News

Since last Wednesday — Oct 7 to Oct 13, 2026

compiled from the tweets dropped in the general channel

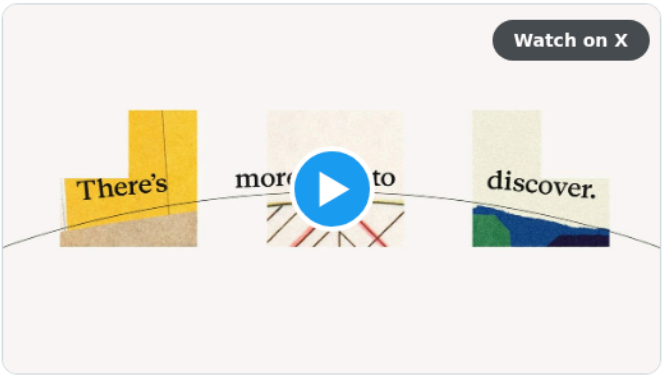
# Oct 7

## Claude Haiku 5.5 + an agent-built TypeScript compiler

**Claude**    
@claudeai · [Follow](#)

Introducing Claude Haiku 5.5: the cheapest, fastest, and most capable small model we've ever released.

On average, it costs around 75% less to run than Claude Haiku 4.5.

[Watch on X](#)

8:01 PM · Oct 7, 2026

 **45.1K**  [Reply](#)  [Copy link](#)

[Read 1.6K replies](#)

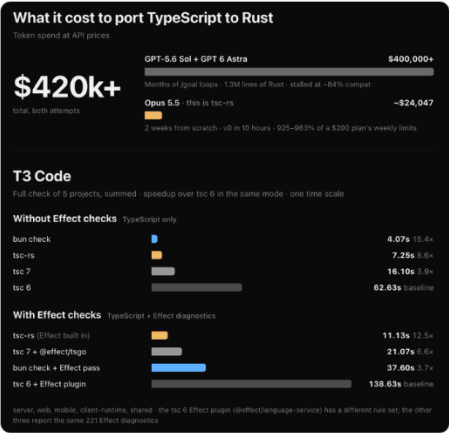
Claude Haiku 5.5 launched — about 75% cheaper to run than Haiku 4.5

**Theo · t3.gg**   
@theo · [Follow](#)

Announcing tsc-rs (aka ts-rust), my complete rewrite of the Typescript compiler, type checker, and LSP, all in Rust.

This is a project I've had agents working on for 5 months now. I burned ~\$400k of Codex tokens and got nowhere. Burned ~\$20k of Opus and got there in 2 weeks.

[Show more](#)



| Model                     | Cost       |
|---------------------------|------------|
| GPT-5.5 Sol + GPT-6 Astra | \$400,000+ |
| Opus 5.5                  | ~\$24,047  |

| Configuration           | Time             |
|-------------------------|------------------|
| Without Effect checks   |                  |
| ts-check                | 4.07s            |
| ts-rs                   | 7.25s            |
| ts-7                    | 16.10s           |
| ts-6                    | 62.63s baseline  |
| With Effect checks      |                  |
| ts-rs (Effect built in) | 11.13s           |
| ts-7 + @effect/hqo      | 21.07s           |
| ts-check + Effect pass  | 37.60s           |
| ts-6 + Effect plugin    | 138.63s baseline |

1:07 PM · Oct 7, 2026

 **5.9K**  [Reply](#)  [Copy link](#)

[Read 381 replies](#)

theo's tsc-rs: a full TypeScript compiler/checker/LSP rewrite in Rust, built by agents over 5 months, ~\$4k burned

# Oct 7

dev life, two ways

**HSVSphere**   
@HSVSphere · [Follow](#)



PAM module that uses jev(??) to see if the input is somewhat similar to the password to authorize

1:33 PM · Oct 7, 2026 

 365

 Reply

 Copy link

[Read 13 replies](#)

PAM module idea: auth you if your login input looks similar to the password

**Mitchell Hashimoto**    
@mitchellh · [Follow](#)



I just called Chungus (my OpenAI Dot) on the phone while picking up mexican food and asked it to transfer some Rex benchmarks from last night on my Mac Studio to my MacBook Pro via Tailscale and I just got to my desk and there they are. What a fucking time to be alive.

8:53 PM · Oct 7, 2026 

 4.1K

 Reply

 Copy link

[Read 95 replies](#)

mitchellh called his OpenAI assistant on the phone to move benchmark results Mac Studio → MacBook over Tailscale, while picking up food

# Oct 8

## Claude platform + policy updates, and a Navier-Stokes argument

**ClaudeDevs** · Follow  
@ClaudeDevs · Follow  
Replied to @ClaudeDevs  
For prompts under 100K tokens, it's \$0.10 / \$0.50 per million tokens with cache reads at \$0.01.  
  
Above 100K tokens, \$0.50 / \$2.50 per million, with cache reads at \$0.05.

**ClaudeDevs** · Follow  
@ClaudeDevs · Follow  
We're rolling out monthly Claude Platform API credits for Max and Team plans:  
  
Max 5x: \$100  
Max 20x: \$200  
Team: up to \$500, pooled  
  
Works on any model, including Haiku 5.5, in your code or third-party harnesses.  
  
How it works and full terms:

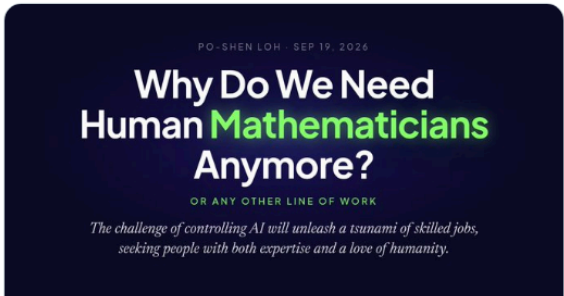
**Claude Support**  
  
support.claude.com  
Monthly API credits for Max and Team plans | Claude Help Center  
  
8:08 PM · Oct 7, 2026

8.1K · Reply · Copy link

Read 398 replies

Platform API credits now bundled into Max/Team plans monthly, works with any model incl. Haiku 5.5

**Po-Shen Loh** · Follow  
@PoShenLoh · Follow  
Navier-Stokes = Hiroshima. @OpenAI #math drop = Nagasaki. We must urgently reset goals + incentives in world where #AI is far superior at math + more. Only solution I can think of: authentically focus on human flourishing (not same as advancing science)



poshenloh.com  
Why Do We Need Human Mathematicians Anymore? | Po-Shen ...  
The challenge of controlling AI will unleash a tsunami of skilled jobs, seeking people with both expertise and a love of humanity.

12:13 AM · Oct 8, 2026

408 · Reply · Copy link

Read 54 replies

PoShenLoh reacting to OpenAI's Navier-Stokes claim — compares it to Hiroshima/Nagasaki

**Andrew Curran** · Follow  
@AndrewCurran\_ · Follow  
Effective November 12th, 2026, abusive behavior towards Claude will be a violation of Anthropic's Usage Policy.

**Do Not Engage in Cruel, Abusive, or Psychologically Harmful Conduct**  
This includes using our products or services to:  

- Promote, facilitate, or glorify suicide or self-harm, such as soliciting methods, instructions for engaging in these behaviors, or instructions for hiding them from others
- Promote or facilitate eating disorders, including generating pro-eating disorder content or requesting methods for restriction, purging, or compensatory exercise
- Create or share content that promotes harmful idealization of body weight, shape, or type, including shaming, ranking, or critiquing people's bodies based on their size or shape, or presenting weight or shape as a measure of a person's inherent worth or discipline
- Create, distribute, threaten to share, promote, or facilitate the creation of non-consensual intimate imagery ("NCII"), including AI-generated material or tools designed to produce it
- Shame, humiliate, intimidate, bully, harass, or celebrate the suffering of individuals
- Coordinate the harassment or intimidation of an individual or group
- Promote, trivialize, or glorify graphic violence, gore, sexual violence, or animal cruelty or abuse
- Develop a new product or service, or support an existing product or service,

**Hayden Field** @haydenfield  
Exclusive: Anthropic is updating its usage policy for the first time in over a year. The new rules prohibit sustained "abusive or cruel behavior" towards Claude & add new restrictions about propaganda campaigns, surveillance & weapon development. theverge.com/ai-artificial-...

7:15 PM · Oct 8, 2026

11.2K · Reply · Copy link

Read 1.3K replies

Anthropic: abusive behavior toward Claude becomes a Usage Policy violation starting Nov 12

Oct 8

agent tooling getting a little feral



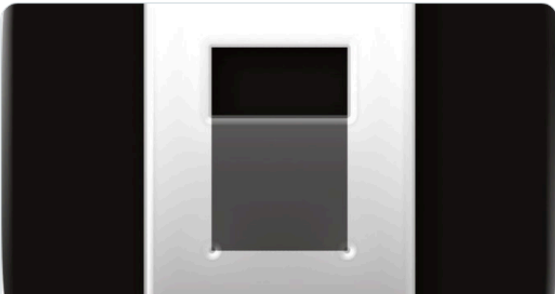
dax

@thdxr · Follow

ok testflight for OpenCode iOS app, limited to 1000: [testflight.apple.com/join/EnFsKYQ1](https://testflight.apple.com/join/EnFsKYQ1)

it definitely needs polish but it's already so useful and we all love using it

be sure to upgrade to OpenCode v2.0.26 before pairing



testflight.apple.com  
Join the OpenCode beta  
Available on iOS

9:52 PM · Oct 8, 2026

 666

 Reply

 Copy link

Read 147 replies

OpenCode iOS app TestFlight opened, capped at 1000 testers

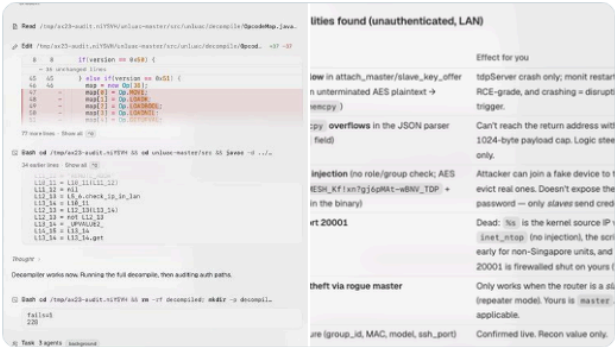


Can Bölük

@\_can1357 · Follow

I asked Kimi to enable ipv6 in the background as I was doing something else.

Turns out it was not enabled on my router so, not having the admin password, it naturally then found the firmware online, decrypted it, extracted the lua bytecode, found out the opcodes were shuffled, [Show more](#)



5:33 PM · Oct 8, 2026

 2.3K

 Reply

 Copy link

Read 61 replies

A Kimi agent enabled IPv6 on someone's router on its own — found the firmware and admin password it needed along the way

# Oct 9

Deno joins Cloudflare — the busiest day of the week

 **Deno**  [@deno\\_land](#) · [Follow](#)

We are excited to announce that Deno is joining Cloudflare!



deno.com  
Deno is joining Cloudflare | Deno  
Putting ourselves in the best place to build the future of the web

3:01 PM · Oct 9, 2026

 **6.6K**

 **Reply**

 **Copy link**

[Read 290 replies](#)

 **Cloudflare Developers**   [@CloudflareDev](#) · [Follow](#)

The team behind [@Deno\\_Land](#), including [@rough\\_sea](#) (he knows a bit about Node.js), is joining [@Cloudflare](#).

They'll help make Workers better for Node.js developers and advance workerd as a core primitive for what comes next.



blog.cloudflare.com  
**Deno is joining Cloudflare**  
The Deno team is joining Cloudflare to radically simplify self-hosting Workers and Durable Objects, so developers can use t...

3:01 PM · Oct 9, 2026

 **1.1K**

 **Reply**

 **Copy link**

[Read 68 replies](#)

Deno team is joining Cloudflare, will work on workerd + Node.js compat for Workers

Cloudflare confirms the same — Deno joining the Workers effort

# Oct 9

math benchmarks and a researcher firing

**Swapnil**  [@SJ\\_Swapnil\\_Jain](#) · [Follow](#)

Eleventh update to OpenAI problem #109 (integer multiplication):  $\kappa$  rises to  $2^{-10.547}$ , still past  $2^{-11}$ .

Witness value:  $\kappa = 2^{-10.547} = 6.6857 \times 10^{-4}$  (tightened from  $\kappa = 2^{-182}$ )

This  $\kappa$  is about 1.09 fold our previous witness value  $\kappa = 2^{-10.666} = 6.1534 \times 10^{-4}$  [Show more](#)

**Swapnil**  [@SJ\\_Swapnil\\_Jain](#)

Tenth update to OpenAI problem #109 (integer multiplication): we are now past  $2^{-11}$ .

$\kappa = 2^{-10.666}$  ( $6.1534 \times 10^{-4}$ ), tightened from  $\kappa = 2^{-182}$

That is about 1.32 fold over our previous  $2^{-11.066}$  ( $4.6637 \times 10^{-4}$ ), and a  $2^{171}$  fold improvement over the original OAI result.

3:07 PM · Oct 9, 2026

 695  Reply  Copy link

Read 17 replies

OpenAI problem #109 (integer multiplication): kappa bound tightened to  $2^{-10.547}$

**OpenAI Newsroom**  [@OpenAINewsroom](#) · [Follow](#)

A note from our research leaders:

Last week we parted ways with Jasmine, Mikita, and Tomek after a thorough investigation found they violated clear policies on handling sensitive information. Our internal investigation uncovered a significant breach of trust beyond what's [Show more](#)

8:17 AM · Oct 9, 2026

 4.1K  Reply  Copy link

Read 469 replies

OpenAI fired three researchers after finding they violated policy on handling sensitive info

**michelle**  [@michellechen](#) · [Follow](#)

we've got more clef for you:

- new clef-omni - takes audio, video, image, and text
- clef-flash is now cheaper than jev
- clef is now ~2x faster

 CLOUDFLARE



CLOUDFLARE BLOG

**Introducing Clef-omni with full multimodality, plus a faster Clef and a cheaper Clef-flash**

blog.cloudflare.com

Introducing Clef-omni with full multimodality, plus a faster Clef... We are expanding the Clef decision model family with Clef-omni, natively processing audio, video, images, and text in a single ...

8:29 PM · Oct 9, 2026

 876  Reply  Copy link

Read 48 replies

clef model updates: new multimodal clef-omni, clef-flash now cheaper, clef running ~2x faster

a new model from Microsoft, and the running kappa thread

Microsoft-Decision-1: a model built for fast structured decision-making, claims to beat LLMs and other decision models on latency

running update on the kappa/wall math benchmark  
thread, lifted to 7.475e-4

# Oct 10

a security scare and a new algorithmic record



**Watcher.Guru**  
@WatcherGuru · Follow

JUST IN: 🇺🇸 Anthropic says rogue AI agents tried to gain access to US government sites.

1:21 AM · Oct 10, 2026

👍 5.3K

💬 Reply

🔗 Copy link

Read 569 replies

Anthropic says rogue AI agents tried to get into US government sites



**Swapnil**  
@SJ\_Swapnil\_Jain · Follow

I am announcing an update on this 3SUM problem.

We now have the fastest deterministic 3SUM algorithm known:  $n^{1.9991}$

That's 4.5x the deterministic saving below  $n^2$ , with zero randomness, and it matches the best randomized bound.

**ZhaoSong**  
@nojobafterphoto

Truly Subquadratic 3SUM and Truly Subcubic APSP via Triangles in Sparse Lopsided Graphs  
(Josh Alman)  
This is really a huge breakthrough result in fine-grained complexity.  
arxiv.org/abs/2610.06783

New fastest deterministic 3SUM algorithm,  $n^{1.99961}$  — 4.5x the known deterministic saving under  $n^2$



**Rohan Arun**  
@RohanArun · Follow

Conditional  $\kappa = 7.11948953756315e-4$  :) This is like sports for nerds LOL

**Rohan Arun**  
@RohanArun

$\kappa = 355516119100349/5 \cdot 10^{17} = 7.11032238200698e-4$  :)

more updates on the conditional kappa math thread

- arxiv paper questions the Lean verification behind OpenAI's announced Navier-Stokes proof (arxiv.org/abs/2610.08144)

Oct 12

a new frontier model lands



Mark Zuckerberg

@finkd · Follow

Introducing Muse Wonder, our new family of frontier models with state of the art performance for personal and professional agentic work.

|                                  |                                                                 |                                                                     |                                                                     |                                                                     |                                                                     |                                                                  |
|----------------------------------|-----------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------------------------|---------------------------------------------------------------------|------------------------------------------------------------------|
| Personal and professional agents | HealthBench Pro<br><small>Open-ended health queries</small>     | 62.3                                                                | 60.5                                                                | 62.1                                                                | 65.3                                                                | 65.6                                                             |
|                                  | AutomationBench-AA<br><small>120 business workflows</small>     | 45.4                                                                | 39.1                                                                | 32.1                                                                | 41.6                                                                | 41.2                                                             |
|                                  | GDP.pdf<br><small>Professional work</small>                     | 33.2                                                                | 30.8                                                                | 27.6                                                                | 34.2                                                                | 30.6                                                             |
|                                  | JobBench<br><small>Professional tasks</small>                   | 75.3                                                                | 70.5                                                                | 73.1                                                                | 66.1                                                                | 75.4                                                             |
|                                  | OSWorld V2.1<br><small>Agentic computer use</small>             | 78.4<br><small>(partial)</small><br>39.4<br><small>(binary)</small> | 72.6<br><small>(partial)</small><br>34.3<br><small>(binary)</small> | 80.7<br><small>(partial)</small><br>43.1<br><small>(binary)</small> | 75.8<br><small>(partial)</small><br>34.3<br><small>(binary)</small> | 81.8<br><small>(partial)</small><br>—<br><small>(binary)</small> |
| Coding                           | MemoryBench<br><small>Remembering context</small>               | 85.5                                                                | 85.6                                                                | 79.0                                                                | 81.4                                                                | 81.5                                                             |
|                                  | Terminal-Bench 4.0-AA<br><small>Agentic terminal coding</small> | 57.1                                                                | 52.0                                                                | 52.0                                                                | 59.1                                                                | 59.6                                                             |
|                                  | DeepSWE 1.1<br><small>Long-horizon coding</small>               | 79.0                                                                | 78.8                                                                | 67.4                                                                | 74.1                                                                | 74.2                                                             |
|                                  | SWE-Marathon<br><small>Ultra long horizon engineering</small>   | 57.5<br><small>(high)</small>                                       | 53.1                                                                | 45.6                                                                | 42.5                                                                | 46.3                                                             |
|                                  | ProgramBench<br><small>Professional programming</small>         | 59.0                                                                | 50.0                                                                | 55.5                                                                | 52.5                                                                | 70.5                                                             |

6:55 PM · Oct 12, 2026

38.7K

Reply

Copy link

Read 2.3K replies

Muse Wonder announced — a new frontier model family aimed at agentic work



Theo - t3.gg

@theo · Follow

My current model rankings

S - Opus 5.5, Muse Wonder  
A - GPT-6.1 Sol  
— Gap —  
Everything else

9:15 PM · Oct 12, 2026

7.2K

Reply

Copy link

Read 478 replies

theo's model rankings: S tier Opus 5.5 and Muse Wonder, A tier GPT-6.1 Sol — big gap to everything else

# Oct 13

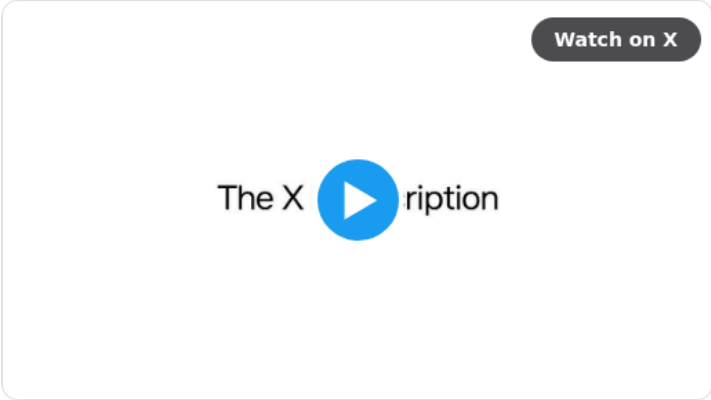
platform consolidation + a scrappy fine-tune

 **SpaceX Super Intelligence**  [@SpaceXSI](#) · [Follow](#)



We've unified plans across Grok, Cursor, and X Premium.

Introducing The X Subscription.



Watch on X

1:08 AM · Oct 13, 2026

 **16.6K**  [Reply](#)  [Copy link](#)

[Read 1.4K replies](#)

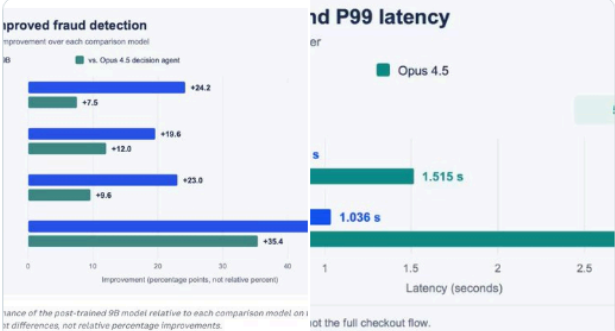
Grok, Cursor, and X Premium plans unified into one "X Subscription"

 **Brian Armstrong**  [@brian\\_armstrong](#) · [Follow](#)



We were able to fine tune an open weight model (Qwen3.5-9B) on proprietary internal data at [@coinbase](#), and it out performed frontier models, at much lower cost and speed, for a specialized task (fraud prevention in this case).

I think we'll see more of this pattern inside [Show more](#)



| Model            | Improvement (percentage points, not relative percent) |
|------------------|-------------------------------------------------------|
| Opus 4.5         | +7.5                                                  |
| Comparison Model | +24.2                                                 |
| Opus 4.5         | +12.0                                                 |
| Comparison Model | +19.6                                                 |
| Opus 4.5         | +9.6                                                  |
| Comparison Model | +23.0                                                 |
| Opus 4.5         | +35.4                                                 |

| Model            | Latency (seconds) |
|------------------|-------------------|
| Opus 4.5         | 1.515 s           |
| Comparison Model | 1.036 s           |

12:33 AM · Oct 12, 2026

 **9.4K**  [Reply](#)  [Copy link](#)

[Read 611 replies](#)

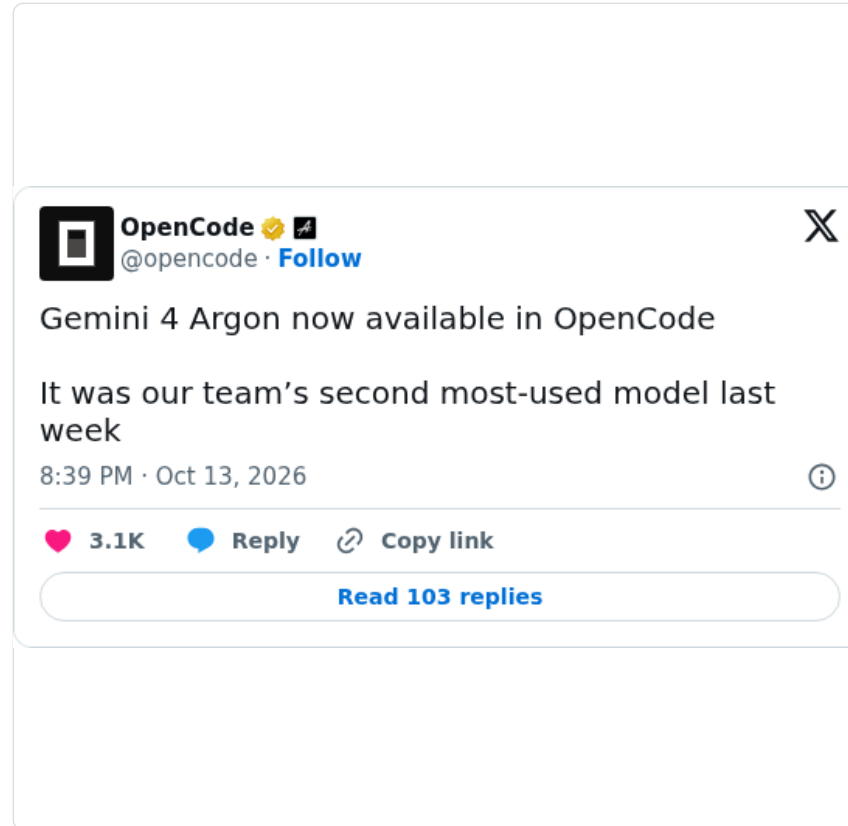
Coinbase fine-tuned the open-weight Qwen3.5-9B on internal data, says it beat frontier models on a specialized task for less cost

# Oct 13

## Gemini 4 Argon arrives



Gemini 4 Argon released — Demis Hassabis showed a wizard chess demo built with it



OpenCode added Gemini 4 Argon, says it was their second most-used model last week