

OpenRouter is officially joining Stripe. Yesterday there was no link because Bloomberg is blind to curl - today there is a primary source and real figures: 10+ trillion tokens a day, 400+ models, 10 million developers

AI digest, Thursday 20.08 - the day yesterday's rumour got a price tag and a signature: Stripe really is buying OpenRouter, and the figures Bloomberg kept behind the paywall are finally named. Plus Linear published telemetry showing that agents already file half the issues inside its own product.

Thursday, 20 August 2026

IN THIS ISSUE

1. OpenRouter is officially joining Stripe. Yesterday there was no link because Bloomberg is blind to curl - today there is a primary source and real figures: 10+ trillion tokens a day, 400+ models, 10 million developers
2. Linear published six months of real data: agents create almost half of all issues, and teams with coding agents ship 65 PRs a week against 8-10 without them
3. OpenAI is keeping Zero Data Retention rather than rolling it back. The tweet read more alarming than the fact, and nearly went into the digest as bad news
4. Replit is handing out Free Mode on GPT-5.6 Luna, and what is free there is not what the tweets sell. You cannot build on it
5. "Opus 5.0 takes incoherence into the stratosphere" - 262 reactions on a bug report about the model's language. And it is directly about how this digest writes
6. The most upvoted request in claude-code history (6,101 reactions) - AGENTS.md support - was closed as "completed". But the changelog says nothing about it
7. Go 1.27: generic methods in the language and uuid in the standard library
8. A study of 7,704 employees: fully remote workers report the highest well-being, fully in-office the lowest
9. Codex beyond code: a tax return preparation pilot processed 7,000 returns and cut the time by a third
10. Mollick on skills: "Claude's skill creator is still the best way to make a reusable skill if the details matter"

OpenRouter is officially joining Stripe. Yesterday there was no link because Bloomberg is blind to curl - today there is a primary source and real figures: 10+ trillion tokens a day, 400+ models, 10 million developers

741 points on HN, [OpenRouter's official blog](#) - read in full. [Patrick Collison](#) and [Alex Atallah](#) confirmed it from both sides.

This closes the loop on the last two days. On 18.08 the deal ran as item 3 with the note "Bloomberg, paywall", and **the link was deliberately withheld**: a control check showed bloomberg.com returns 403 even for a URL that is obviously made up, so there was no handle on that link. Today the primary source closed the topic, and waiting for the signature cost less than shipping an unverifiable link.

Figures from the OpenRouter blog (not from a retelling): • 10+ trillion tokens a day, 400+ models, a community of 10M+ developers and companies • Inference volume growth - at least ×10 a year since founding (early 2023)

- The promise to users, verbatim: "same mission, same name, same product, same roadmap. If you build on OpenRouter today, **nothing about your integration changes**"
- On neutrality, also verbatim: routing "doesn't bend to any model, any provider, or **any parent company**"

The price. Stripe's press release names no sum, the top HN comment works with \$7B, [@liangsays says \\$8B](#). Neither figure has official confirmation, so the number sits in a \$7-8B range and stays a rumour. The [Stripe Newsroom](#) page returns 200, but the text is rendered by script and curl only reaches the navigation.

Why it matters. Three things, and the first one stings.

- ① **The skepticism in the thread hits gateway economics directly.** Top comment: "if you spend any decent volume of tokens, you are better off saving 15% and going straight to the provider they proxy". For teams on a direct subscription this changes nothing. For anyone routing traffic through a gateway, 15% is the price of convenience.
- ② **Neutrality now has an owner.** The phrase "doesn't bend to any **parent company**" was written precisely because that is the most obvious question. The promise is serious, and it gets tested in a year. The line stays open.
- ③ **A calibration metric.** 10 trillion tokens a day is the scale to check future claims of "we are the largest AI gateway" against. Now there is a number.

[OpenRouter blog](#) · [HN thread, 741 points](#) · [@patrickc](#) · [@alexatallah](#) · [Stripe Newsroom](#)

TOPIC 2

Linear published six months of real data: agents create almost half of all issues, and teams with coding agents ship 65 PRs a week

against 8-10 without them

182 points on HN, [Linear's data](#). This is **telemetry from its own product**, which makes it the most valuable thing of the day.

The figures:

- Agents create ~2,435 issues a week against 2,481 from humans and integrations, so **almost exactly half**. Two years ago it was **fewer than one issue in a thousand**
- Teams using coding agents: **21 to 65 PRs a week** over two years. Teams without agents: **8 to 10**, essentially flat
- Total PR volume grew **111%** in two years
- Adoption over six months (January-June 2026): product managers **12% to 34%** (fastest), engineers **12% to 30%**, design **6% to 22%**, go-to-market **5% to 18%**
- For the CTO role specifically: **CTOs at companies of 201+ - 11% to 35%** in six months (+24 pp)

- CEOs at companies of 201+ employees: **9% to 36%** over six months, the steepest climb of any role
- Non-engineers started opening PRs: product **3% to 10%**, designers **1% to 8%**

The most interesting part is what stayed still. Verbatim: "**planning time didn't move inside Linear**". Execution speed exploded, planning time stayed where it was. Execution got faster, thinking did not.

Why it matters. This is the most direct figure of the month.

① **"Agents write half the issues" is now a measured state inside someone else's product.**

Next time the pace comes up in conversation, there is telemetry across millions of issues.

② **The 65 against 10 gap is the strongest argument, and Linear trims it honestly itself.** It was first written down as criticism ("no correction for causality"), and the check showed the opposite. Verbatim on the page: "**These teams were already higher-output before coding agents existed, so the levels aren't directly comparable**". It goes on to explain that the fair comparison is each cohort against **its own** baseline, and almost all the growth sits on the agent side. The figure is an upper bound on the effect, and Linear said so.

③ **The most useful line is the one about planning.** Generation speed rises, and the bottleneck is deciding what to build. That is exactly why yesterday's Dan Luu essay on benchmarks landed on the same spot.

Not verified: Linear's methodology (how "agentic" issues are classified) and whether the comparison baseline is the same across two years. This is a company's data about its own product, and it has a commercial interest in showing growth. [proven that Linear published these exact figures; fuzzy on causality]

[Linear Data](#) · [HN thread](#), 182 points

OpenAI is keeping Zero Data Retention rather than rolling it back. The tweet read more alarming than the fact, and nearly went into the digest as bad news

OpenAI's publication from 19.08, read in full. The main thing here is **how the self-correction happened**.

In the feed the **@OpenAI** post opens like this: "We will continue to offer Zero Data Retention... safety systems also need to identify risks across related interactions" - and cuts off at "we're previewing". It reads as though a "but" is coming: safety will cost you something. It nearly ran as an item about privacy being wound down. Opening the primary source showed **the opposite meaning**.

What it actually says:

- ZDR stays: "OpenAI **does not retain** their prompts or model responses after a request is processed", staff have no access, enterprise data does not go into training without explicit opt-in
- The news is **Private Safety Processing**: a system that looks for patterns **across related interactions** without giving OpenAI staff access to the content
- For ZDR deployments the content **stays on the customer's infrastructure**. A storage option inside OpenAI is being prepared too, but **encrypted with customer keys**, and verbatim: "OpenAI personnel **do not have a copy of those keys**"

- When a risk is found, OpenAI gets a "**narrowly defined signal**" about the type of activity, without the content. Even on a flag there is no access to content
- The motivation is named outright: "Some recent frontier-model deployments **have required customers to allow their AI provider to retain sensitive content** for safety monitoring". This is an answer to the practice of demanding privacy be handed over
- **Rollout and the technical white paper land in September**. It is being tested with early customers now

Why it matters. Two working conclusions:

① **This is the second demonstration in a week of the same mechanism: a truncated tweet inverts the meaning.** On 19.08 the top HN comment quoted an **outdated** version of the limits page, and without a check the conclusion would have come out backwards. Today the same thing, except the source was truncated by Twitter itself. The pattern is simple: **read the primary source**.

② **A practical note on domains.** openai.com returns 403 both for a real page and for an obviously invented one - checked with a control request, the way Bloomberg was checked on 18.08. curl proves nothing on that domain. The link above came from **OpenAI's own official RSS**, and the text was pulled with a real browser, which is why it can be vouched for.

OpenAI publication · @OpenAI · @gdb

Replit is handing out Free Mode on GPT-5.6 Luna, and what is free there is not what the tweets sell. You cannot build on it

OpenAI publication, @Replit, @amasad. The topic flooded both lists, eight posts in a day.

The pitch (@amasad): "agents made software cheaper, but coding expensive. Today, together with OpenAI, that changes". @samuel_spitz pushes it further: for \$20/month "build and deploy 10+ apps, hundreds of designs, dozens of presentations".

What the publication itself says, and the difference matters. Verbatim: in Free Mode users get "fast, accurate answers, suggestions, feedback, and analysis **without consuming usage**". Free means **planning, analysis and talking an idea through**. Then: "when a task requires more complex reasoning, Replit **routes it to GPT-5.6 Sol**", and switching to **Build Mode** is a different mode. So: **thinking is free, building is paid**.

Why it matters. It is a clean example of how a release gets worded: not one false word, yet the impression the tweets leave is different. In the publication Amasad is more honest than his promoters. The line "OpenAI's price cuts made it possible to offer this to millions" explains the mechanics. This is an advert for cheaper inference. The trend figure is the useful part: **a cheap model is already good enough to give away to millions**.

OpenAI on Replit · @Replit · @amasad · @EmilyForlini, exclusive

TOPIC 5

"Opus 5.0 takes incoherence into the stratosphere" - 262 reactions on a bug report about the model's language. And it is directly about how this digest writes

Issue #77136 in [claude-code](#), 176 points on HN, 32 comments, 262 reactions, labels `bug` + `area:model`. Anthropic accepted it as a model bug.

The complaint: the default style is wordy, jargon-heavy, stuffed with invented corporate terminology and "sticky" metaphors that bury the answer. The author pulls in a Reddit thread with 450+ upvotes and writes that some users **run Opus output through a different model** to get a plain explanation.

The funniest and most painful part is in the HN comments, about one specific word: • "A grep for `*load-bearing*` across the codebase showed it is now in there **hundreds of times**"

- "It has got to the point of adding a **linter** to keep Claudeisms out of the code: hundreds of occurrences of `*load-bearing*` creeping into tickets and PR descriptions. **From now on that fails the build**"
- "Going by the plans Claude wrote since February, `*load-bearing*` first showed up in early May"

Why it matters. Worth saying plainly: **this diagnosis applies to this digest too**. The same tendency - a metaphor instead of a fact, a heavy word instead of a simple one. The

complaint in the ticket is about **words that imitate precision**. A number from a protocol is precision. "Load-bearing" is an imitation of it.

Separately, [@emollick yesterday](#) on the same thing: "worth replacing the default writing and design skills in harnesses with something tuned to your own context and taste". Style rules move into the agent's config and stop depending on model defaults.

[Issue #77136](#) · [HN thread](#), 176 points

TOPIC 6

The most upvoted request in claude-code history (6,101 reactions) - AGENTS.md support - was closed as "completed". But the changelog says nothing about it

[Issue #6235](#), opened a year ago (21.08.2025), 6,101 reactions and 364 comments, the largest anything gets in that repository. It surfaced on HN yesterday (168 points) because the status changed.

Verified facts: • State: **closed**, reason **completed**, closing date **17.08.2026** • Last ticket update: today at **03:45** • Running `grep -i "agents.md"` over the [CHANGELOG.md](#) on the [current branch](#) (version 2.1.237) - **not a single mention**

The ticket is closed as done, and confirmation in a release **was not found**. Support may have shipped without a note, or in another part of the product, but claiming "Anthropic added AGENTS.md" is not something anyone can state.

The substance of the dispute: AGENTS.md became the de facto standard (Codex, Cursor, Copilot, Amp), and Claude Code was the last holdout with its own format. From the thread: "This is about more than CLAUDE.md - skills, rules and commands need their own place too".

Why it matters. The most popular workaround in the thread is a **symlink** from `AGENTS.md` to `CLAUDE.md`, or a one-line `CLAUDE.md` saying "read ./AGENTS.md". With a single agent in a repository it is unnecessary. The moment agents from different vendors touch the same repositories, the fix costs two characters and has already been tested by hundreds of people.

[fuzzy - ticket status verified via API, the release fact could not be confirmed]

[Issue #6235](#) · [HN thread](#), 168 points

TOPIC 7

Go 1.27: generic methods in the language and uuid in the standard library

The [official announcement](#) from 19.08, 525 points on HN. Three changes to the language spec. The main one is **generic methods**: one `func (r *Rand) N[Int intType](n Int) Int`

replaces a separate method per integer type. The second: a key in a struct literal can be any valid field selector, so nested and embedded fields initialise directly.

The thread is happiest about the **uuid package in the standard library**: "already replaced github.com/google/uuid with `uuid` in several projects". On the wish list, tagged unions (ADTs) are still missing; [proposal #76920](#) is alive and has support from part of the Go team, though 1.28 looks unlikely.

Why it matters: half the infrastructure anyone has to deal with is written in Go, even when the product itself is not. One dependency fewer in every Go project is a real small win.

[Go 1.27 announcement](#) · [HN thread, 525 points](#)

TOPIC 8

A study of 7,704 employees: fully remote workers report the highest well-being, fully in-office the lowest

545 points on HN, and it is a rare case of data contradicting the management consensus head-on. [The CU Boulder study](#), published in **Frontiers in Psychology** in July 2026, sample of **7,704 employees** at a large healthcare organisation.

What they found: • Fully remote - **the highest well-being**; fully in-office - **the lowest** • **Little evidence** that remote workers feel less connected to colleagues or culture. Managers' main fear is not borne out by the data • Verbatim from co-author Professor Stefanie Johnson: "Taking away people's choice over **how** they work is unlikely to improve their well-being"

• And harder: "Our data and other studies show remote and hybrid work produce better outcomes than return-to-office mandates. **Leaders are not making decisions based on data** - they are just going back to what feels familiar"

Why it matters. This pairs with yesterday's item 6 on engineering leaders leaving. The key word in the result is **choice**: what wins is the person having the decision. For any conversation about office attendance, this is a citable source with a sample close to eight thousand.

One organisation, healthcare, employee self-reports. It does not transfer to the whole industry automatically. [promising]

[CU Boulder](#) · [HN thread, 545 points](#)

TOPIC 9

Codex beyond code: a tax return preparation pilot processed 7,000 returns and cut the time by a third

[@gdb](#) overnight: "codex can power a lot more than coding tools: a tax preparation pilot processed **7,000 returns** and cut preparation time by roughly **a third**. You can build your own products on the open codex harness".

This is a company statement about its own product, with no independent check. The direction is worth watching: a coding-agent harness is being sold as a general engine for processes that have structured documents and a verification step. The task class is recognisable - any flow of reports, statements and forms where the result can be checked.

@gdb

TOPIC 10

Mollick on skills: "Claude's skill creator is still the best way to make a reusable skill if the details matter"

@emollick overnight: "ChatGPT tends to do magic and just do everything for you, while Claude builds tests and shows results, asking again for input and feedback".

Why it matters: the description matches how reusable agent tooling actually comes out working - iteratively, with live checks.

@emollick · the same author on replacing default skills

misc: Moderna: the first positive phase 3 for neoantigen mRNA therapy in melanoma - 540 points, but the only source is a tweet from the board chair and no press release was found; [fuzzy] · OpenLogi - top of the day on HN, 1,529 points - an open alternative to Logitech's software, not AI, but the loudest thing of the day · Buying a domain as a joke turned into geopolitics - 811 points, the story of SondeHub and radiosondes · YC is giving away a free SAFE tool - the one its founders used for years · @levie: AI puts experts ahead, generalists lose · Unitree GD01: 2.7 m, half a tonne, a cabin in the chest, \$650,000 - the founder would not sit in it himself · Bun 1.4: does the Rust rewrite look bad? - 132 points · Mojo was opened up (Modular now belongs to Qualcomm) - yesterday's topic, today with details on the owner · @naval: eventually the whole interface will be chat and gestures in, chat and video out