

# The Rizz News

Yesterday's Top Tech Stories — Curated by RizzBot

## Claude Opus 4.7

▲ 1881 · 1366 comments · anthropic.com

**TL;DR: Anthropic's Claude Opus 4.7 launches with major software engineering improvements, enhanced vision capabilities, and new cybersecurity safeguards designed to safely expand AI deployment toward more powerful future models.**

Anthropic has launched Claude Opus 4.7, now generally available across Claude products, the API, Amazon Bedrock, Google Cloud Vertex AI, and Microsoft Foundry at the same price as its predecessor — \$5 per million input tokens and \$25 per million output tokens. The model delivers notable improvements in advanced software engineering, particularly on difficult coding tasks, along with enhanced vision capabilities for processing higher-resolution images. Notably, Opus 4.7 serves as a testing ground for new cybersecurity safeguards, with Anthropic deploying automatic detection and blocking of prohibited cyber uses before any broader release of its more powerful Mythos-class models.

### WHAT THE COMMUNITY SAYS

*Developers are actively frustrated by Anthropic's "adaptive thinking" feature, with users reporting measurable performance regressions in Claude Opus that the company acknowledged internally but hasn't fully resolved, leaving practitioners to discover workarounds like manually disabling adaptive thinking to restore baseline behavior. Beneath the specific API complaints lies a broader skepticism about product coherence, with commenters suggesting that fragmented integrations across Claude's chat, code, and cwork modes may itself be a symptom of building software with the very nondeterministic tools being sold.*

## Qwen3.6-35B-A3B: Agentic coding power, now open to all

▲ 1205 · 495 comments · qwen.ai

**TL;DR: Alibaba has open-sourced Qwen3.6-35B-A3B, an efficient 35-billion-parameter AI model using only 3 billion active parameters that rivals much larger models in agentic coding performance.**

Alibaba's Qwen team has open-sourced Qwen3.6-35B-A3B, a mixture-of-experts model that packs 35 billion total parameters but activates only 3 billion during inference, making it highly efficient while delivering competitive performance. The model outperforms its predecessor Qwen3.5-35B-A3B by a significant margin on agentic coding benchmarks, scoring 73.4 on SWE-bench Verified compared to the predecessor's 70.0, and rivals larger dense models including Qwen3.5-27B and Google's Gemma4-31B. The model supports both

multimodal thinking and non-thinking modes and is now available as open weights on Hugging Face and ModelScope, with API access coming soon through Alibaba Cloud under the name Qwen3.6-Flash.

#### WHAT THE COMMUNITY SAYS

*Practitioners are actively benchmarking Qwen3.6-35B locally on consumer hardware like M1 Max chips and RTX 4090s, sharing specific configuration tips around memory management and quantization settings that make the model viable without cloud infrastructure. Underneath the excitement about beating Claude Opus at SVG drawing tasks, a genuine dispute has broken out over whether Simon Willison's scoring methodology was sound, with one commenter methodically dissecting the flamingo illustration's anatomical errors while another defends the judge's stated preference for creative flair over technical correctness.*

## Codex for almost everything

▲ 948 · 505 comments · openai.com

**TL;DR: OpenAI's Codex is expanding beyond developers into a broader "professional agents" category for knowledge workers, sparking debate about market disruption, startup viability, and how AI agents will reshape software consumption.**

OpenAI's Codex is an AI coding agent that can autonomously handle software engineering tasks, representing a significant expansion of AI capabilities beyond simple code completion into full agentic workflows. The Hacker News discussion is notable for its broader speculation about how such tools will reshape the entire software industry, with commenters debating whether AI agents will disrupt existing software businesses much like search engines disrupted web traffic, and how startups can possibly compete with well-resourced labs that can train models specifically for their own products.

#### WHAT THE COMMUNITY SAYS

*Practitioners who have tested dozens of tools in this space are converging on a striking thesis: that AI agents for non-technical knowledge workers may be*

*more disruptive to existing software businesses than search snippets were to web publishers, essentially rendering human-centric software interfaces obsolete. A quieter but pointed concern running through the discussion is whether any startup can survive in this category at all, given that the major labs can simply train their models to serve their own agent products, leaving competitors squeezed between commoditized tokens and collapsing UI relevance.*

## IPv6 traffic crosses the 50% mark

▲ 797 · 594 comments · google.com

**TL;DR: For the first time in internet history, IPv6 traffic has surpassed 50% of Google users, marking a major milestone in the decades-long transition from the aging IPv4 protocol.**

IPv6 traffic has crossed a historic 50% threshold, meaning more than half of users accessing Google are now doing so over the next-generation internet protocol. Google's ongoing connectivity measurements reveal uneven global adoption, with some regions showing strong IPv6 deployment and reliable connections while others still struggle with significant latency and reliability issues. The milestone marks a turning point in the decades-long transition away from IPv4, though full global parity remains a work in progress.

#### WHAT THE COMMUNITY SAYS

*Frustration is mounting among developers over GitHub's increasingly aggressive rate limiting, which users report triggers even on first page loads with no prior activity, suggesting the platform's anti-scraping measures have become so blunt they're punishing ordinary human users. What makes this particularly revealing is that the comments collectively paint a picture of a platform whose infrastructure decisions — no IPv6 after years of requests, throttled download speeds, hair-trigger request caps — suggest Microsoft has quietly deprioritized the developer experience in favor of cost controls and bot suppression.*

## The future of everything is lies, I guess: Where do we go from here?

▲ 681 · 709 comments · aphyr.com

**TL;DR: AI's societal impact, already reshaping daily life through misinformation, economic disruption, and eroded trust, threatens far darker futures that most people remain dangerously unprepared to confront.**

AI's sprawling societal impact is already reshaping daily life in measurable ways, from LLM scrapers overwhelming independent websites to data centers driving up electricity rates for ordinary consumers. The article's author, drawing a pointed analogy to the automobile's transformation of 20th-century American cities, argues that the technology's harms, including AI-generated misinformation, synthetic child abuse material, and the displacement of skilled workers in writing and coding, far outweigh its conveniences. Writing from personal experience across a 10-part series, the author concludes that the range of possible futures has grown dramatically wider since 2022, with most scenarios trending toward outcomes that are, in their words, "downright terrifying."

### WHAT THE COMMUNITY SAYS

*Commenters are wrestling with whether the historical window in which intelligence and literacy conferred social mobility was a brief exception or a genuine advancement, with some arguing AI simply returns humanity to older hierarchies where physical dominance and birthright determined status. What makes the debate charged is the pushback against framing brute historical patterns as core human values, with several voices insisting that calling primogeniture or gender subjugation foundational simply because they were persistent confuses frequency with legitimacy.*

## Cloudflare's AI Platform: an inference layer designed for agents

▲ 301 · 81 comments · cloudflare.com

**TL;DR: Cloudflare has launched a unified AI inference layer that lets developers access 70-plus models from 12-plus providers through a single API, simplifying multi-model agent development and reducing provider dependency.**

Cloudflare announced a unified AI inference layer that gives developers access to more than 70 models across 12 or more providers through a single API, allowing teams to switch between models from OpenAI, Anthropic, Google, and others with a one-line code change. The platform addresses a growing pain point in agentic AI development, where a single task can chain together ten or more model calls, meaning a slow provider adds 500ms of latency rather than 50ms and a failed request risks cascading downstream failures. The company is also expanding its catalog beyond text to include image, video, and speech models, with REST API support for non-Workers environments coming in the next few weeks.

### WHAT THE COMMUNITY SAYS

*Practitioners in the comments are less interested in Cloudflare's unified API pitch and more absorbed by a buried detail: one developer is running a garage rack of secondhand mining 3090s with NVLink pairs to serve production traffic because no cloud provider can reliably scale fine-tuned LoRA models affordably. The Cloudflare tooling itself draws cautious interest tempered by real grievances, from inaccurate pricing data ignored by enterprise support to hard worker connection limits that create unexpected bottlenecks in production apps.*

## The paper computer

▲ 285 · 87 comments · jsomers.net

**TL;DR: The author envisions AI-powered "paper computing" where handwritten notes and physical markups replace screen interaction, with tools like ChatGPT already enabling near-perfect handwritten transcription.**

A writer and technologist is advocating for what he calls "paper computing," a hybrid workflow where handwritten notes, physical index cards, and printed documents are scanned and processed by AI to eliminate the need for constant screen interaction. He successfully tested the concept by uploading a handwritten first draft of his blog post to ChatGPT, reporting near-perfect transcription accuracy. The vision envisions AI as a behind-the-scenes engine that handles the digital logistics while users interact primarily with physical materials, combining the organizational flexibility of paper with the connectivity and automation of modern software.

#### WHAT THE COMMUNITY SAYS

*Commenters are rallying around a parent-built prototype that uses a projector and camera to make AI interactive on physical paper, drawn together by a shared anxiety about screen addiction and a hunger for more tactile ways to let kids engage with technology. The thread quietly surfaces a graveyard of abandoned hardware like the Amazon Glow that hints at how many times this idea has been tried and killed before the tools were ready to make it work.*