DHH and Poteto ship agent code without reading every line, as OpenAI, Meta and Factory buy agent-tool makers

Coding Agents Alpha Tracker

2026-11-05

DHH and Poteto ship agent code without reading every line, as OpenAI, Meta and Factory buy agent-tool makers

By Coding Agents Alpha Tracker • November 5, 2026

Two practitioners explain how they decide which agent-written code gets human review. Also: a wave of acquisitions of coding-agent tool makers, two new US open-weight coding models, OpenAI's agent pipeline that found and patched Linux kernel bugs, and practical tool updates.

Shipping agent code without reading all of it

Two practitioners explained this period how they decide what agent-written code a human actually reviews.

DHH (Omarchy). In a live demo of Omarchy 4.5, DHH said he hasn't written "a single line of serious production code" for Omarchy by hand since the Quattro release. Development has been agent-driven ever since [1]. His loop: tell the agent what he wants, then keep tweaking and giving feedback "until everything looked just right." Building a traditional desktop mode on top of Hyprland took about a day this way [1]. His review rule: - Changes that are critical or touch userland get reviewed. - Work that can be checked by a program, such as keeping up the test suite, runs through automated agent updates and "is not even fully human reviewed" [1]. - He says he adds little over "a pair of frontier agents" at judging whether code is correct. What he adds is "taste, direction, and scope" [1].

His tip for building new apps: point the agent at something that already exists and say "I want it in that structure" [1]. A PDF-signing app came from one prompt ("something like Apple PDF signing where it can save the signatures").

It had no layout instructions and needed only one follow-up fix [1]. A translation PR that Muse Wonder produced in a day added 350,000 lines. He notes they are not 350,000 lines of code, and the result is Omarchy in 50 languages [1].

Poteto. In slides for an upcoming AI Engineer talk, Poteto says his team ships "30x'd" amounts of code "without reading it," and performance is as good as or better than at launch. He credits "agent friendly architecture bets" with keeping slop under control. About 230 people landed at least one PR to Grok bot in October [2].

Coding-agent tool makers acquired

Ben Tossell listed three deals on one day: Humanlayer → Factory, Conductor → Meta, Pierre → OpenAI [3]. - **Pierre → OpenAI:** Greg Brockman says Pierre's open-source developer tools "already power parts of codex today" [4]. - **Conductor → Meta:** the team will build "the next interface" with Meta Superintelligence Labs (MSL) [5]. - **Cat Wu → OpenAI:** after leaving Anthropic, where she worked on Claude Code, Cat Wu is joining OpenAI to lead product management for Core Products [6].

New open-weight coding models

- **SWE-2 Flash (Cognition):** a fast coding model that Cognition says beats all American open-weight models on leading evals, including models 3x its size. Cognition isn't releasing the weights openly. Instead it is building a program to share them securely with partners and researchers [7]. swyx says it is 3x faster and 5x cheaper with the same score on FrontierCode 2 [8].
- **Ripple-65B Preview (RicursiveAI):** a 65B dense model pretrained from scratch. RicursiveAI claims a new state of the art among US open-weight models on agentic coding, beating much larger models such as Inkling-975B and Nemotron 3 Ultra-550B [9]. It says the full R&D cost was $32M [10].

OpenAI's agent pipeline for finding and patching vulnerabilities

OpenAI built an autonomous pipeline that discovers vulnerabilities, exploits them and writes patches. In one run with an upcoming cyber model, it validated and generated patches for 948 Linux kernel local privilege escalation (LPE) vulnerabilities, at a total API cost of $1.2M [11]. Brockman puts that at about $1,000 per vulnerability [12]. OpenAI's Eric Wallace says security researchers used to find about 20 Linux LPEs in a typical year. He says the model found about 1,000 in one week and discovered at least 10 new classes of exploitation technique. OpenAI is putting "a significant amount" of frontier compute into

patching open source for free [13]. Jarred Sumner predicts that "dependencies are over" because rewriting one is "one prompt away" [14].

Tool updates

- **Codex `/goal`:** local threads used to use more and more memory as they got long. Brent Traut says the leaks in long-running turns are fixed, so "you should be able to /goal as long as you like now" [15, 16].
- **Browser use in LangChain Managed Deep Agents:** add a `browser.py` file that calls `define_browser` [17]. Set `scope="thread"` to give each thread a fresh browser, which suits end-to-end QA. Set `scope="user"` to share cookies and logins across threads [17]. The agent gets four Stagehand tools: snapshot, run (code-mode actions), screenshot, and handoff, which passes the browser to you to sign in [17].
- **Sourcegraph Agentic Batch Changes** can now connect to MCP servers such as Linear, Confluence and Buildkite [18].
- **Antigravity** now renders `.ipynb` notebooks natively [19]. A DeepMind engineer has also added T3 Code as a recognized client in Antigravity's ACP server (PR) [20].
- **T3 Code:** Theo claims users "more than 3x their token throughput" after switching [21].

Smaller techniques

- **Skip duplicate CI after a squash merge (Kent C. Dodds):** when PR CI passes, record the Git *tree* hash (not the commit SHA) along with the workflow inputs. If the squash commit on main has the same tree hash, skip tests, lint and typecheck before deploying. If not, run the full CI suite [22].
- **Clone a SaaS product instead of buying it (swyx, Latent Space):** swyx offered $10,000 for an open-source clone of a $40,000 SaaS product and covered entrants' coding-agent subscriptions. He got 69 submissions, about five he would pay for, all within a week [23]. One entry replaced clicking through 1,000 talk submissions one at a time with a single sortable table [23]. The AI judge only checked whether entries were valid. Picking the best stayed with humans, because "I cannot make an AI judge for things that I don't know what I want" [23].
- **Cost controls for agent tool calls (treg tutorial, shared by Jason Zhou):** give each agent its own limited token, kept in its environment and never in a prompt. Try a tool on 20 rows and check the output by hand before automating, and log every call [24]. Cap each paid tool before it goes into a loop [24]. Use free sources first and a gate before each paid step, and run long chains on something that stays awake: closing a laptop paused a 21-hour run for 16.7 hours [24].

Sources

1. Omarchy Quattro RS 4.5 Live Demo - DHH
2. X post by @poteto
3. X post by @bentossell
4. X post by @gdb
5. X post by @charlieholtz
6. X post by @_catwu
7. X post by @cognition
8. X post by @swyx
9. X post by @RicursiveAI
10. X post by @swyx
11. X post by @OpenAI
12. X post by @gdb
13. X post by @Eric_Wallace_
14. X post by @jarredsumner
15. X post by @btraut
16. X post by @ajambrosino
17. Browser Use with Managed Deep Agents
18. X post by @Sourcegraph
19. X post by @antigravity
20. X post by @PatrickToulme
21. X post by @theo
22. X post by @kentcdodds
23. Behind Latent Space: A Rare Look Inside the Pod | AIEi Shanghai
24. X article by @agustingrowth