

Human Judgment Moves Upstream in the Software Factory

Coding Agents Alpha Tracker

2026-08-22

Human Judgment Moves Upstream in the Software Factory

By Coding Agents Alpha Tracker • August 22, 2026

Addy Osmani's practical software-factory playbook is the day's highest-alpha signal: event-driven queues, verification budgets, and explicit human gates. New releases extend the loop toward cheaper model execution, mobile session control, and agent interoperability.

TOP SIGNAL

Build the gate before scaling the agent. Addy Osmani says most teams can go surprisingly far with stock Claude Code or Codex, multiple sessions, good SPECs, constraints, and explicit human-involvement criteria; a factory earns its keep when work becomes repeatable and event-driven across GitHub, Slack, or Linear queues in isolated environments. [1]

The design principle is relocation, not removal: humans still choose the problem, architecture, quality bar, trusted verification signals, and when the evidence is sufficient to ship, while deterministic checks move earlier and run continuously. [1]

TRY THIS

- **Use a bounded issue-to-draft-PR contract.** Tell the agent to read GitHub issue #123 and the repository instructions; implement only the acceptance criteria; leave authentication, billing, migrations, and existing test assertions alone; work in a branch; run `npm run lint`, `npm test`, and `npm run build`; stop if a required check cannot run; then open a draft PR listing checks, remaining risks, and human decisions. Do not merge. [1]

- **Turn labels into the control plane.** For incoming issues, use four states: `ready-to-implement`, `ready-to-spec`, `needs-info`, and `wait-to-implement`. Only let sessions claim work marked `ready`; the label becomes the queue, the lock, and a parking place for work that is not rejected permanently. [1]
- **Budget verification like performance.** Run fast checks such as linting and type checking early; reserve mutation, browser, security, and the full test suite for the draft-PR boundary so the development loop stays fast. Osmani says verification, retries, browser checks, and human review can make a task take 2–4× longer; track cost per merged PR and code shelf life, and pair each run status with per-stage timing—his sample quick-finder run took 7 minutes, while a favorites feature took 56 minutes after two rejections and a human decision. [1]
- **Persist the context your parallel sessions create.** Osmani warns that five or ten concurrent sessions create cold mental models; ask agents to save their trajectory and lessons to a local or repository file instead of leaving the reasoning in chat history. A complementary pattern from Armin Ronacher: ask the agent to make the OpenAPI spec “agent discoverable”—returning the relevant spec when it summarizes a task—and he reports the agent wired an embedding model on Cloudflare five minutes later. [1, 2]

WHAT SHIPPED

- **Codex crossed 20M active users.** Tibo reports 20M active users and announced a banked reset for every Codex and ChatGPT Work user. A separate rate-limit update says some users have had worse cache hit rates than in prior stable weeks, which may explain faster usage drain; OpenAI is investigating. [3, 4]
- **GPT-5.6 Sol got a temporary price cut.** OpenAI says API and credit pricing will drop by more than 20% for the next three months. [5]
- **Mobile session control is converging.** Claude Code can now start a session from a phone: machines running `claude remote-control` appear as device cards in the Code tab, where the user taps a machine and chooses a directory. Antigravity’s Remote Control exposes active sessions through supported browsers and iOS/Android, rolling out to all users beginning with Ultra subscribers. Ben Tossell had identified the absence of mobile-created Claude sessions as a reason users could move to Codex. [6, 7, 8]
- **Simon Willison’s CLI stack caught up with tool-using models.** `llm 0.32.1` fixes fresh installs by pinning `openai<3` after the OpenAI Python library dropped `httpx`; the coming 0.33 release will switch to `httpx2`. `llm-openrouter 0.7` adds compatibility with LLM 0.32, Open-

Router’s Responses API, and server-side `Shell`, `WebFetch`, and `WebSearch` tools enabled with options such as `-T WebSearch`. [9, 10]

- **Kody Exchange launched an agent-interopability surface.** Kent C. Dodds says it replaces copy-pasting between agents by letting them connect directly, and it is free to try without an account. The positioning matches Ali Spittel’s formulation—“I want to use my agent to use your thing”—which Dodds calls a major part of Kody’s appeal. [11, 12, 13]
- **LangChain Academy Tutors puts instruction inside the coding-agent loop.** Its tutors use skills to teach concepts, ask questions, run quizzes, and give feedback, with customizable teaching style; the project links to the `lca-tutors` repository. [14]

GO DEEPER

- **Peter Steinberger — “No Doors for Agents”.** Steinberger’s talk supplies the conceptual frame for the product shifts above: the event recap attributes to him, “The text box is AI’s radio-on-TV phase.” Read that alongside mobile session control and agent-to-agent surfaces. [15, 16]
- **Study Addy Osmani’s Factory reference repo, demo app, and workshop.** This is the implementation path behind the queue, verification, and human-gate patterns; Osmani presents it as a reference setup rather than a claim that every team needs a factory. [1]
- **Study `llm-openrouter 0.7`.** The interesting bit is not another model wrapper but the small, explicit tool surface—Responses API plus shell and web primitives—that lets a CLI agent use routed models without hiding the execution boundary. [10]

Editorial take: The coding-agent frontier is becoming a supervised operating loop—queue work, preserve context, spend a verification budget, and only then widen autonomy; cheaper models and mobile controls make that loop easier to run, not safer to skip. [1, 5, 7]

Sources

1. Human judgment doesn’t leave the software factory. It relocates.
2. X post by @mitsuhiko
3. X post by @thsottiaux
4. X post by @thsottiaux
5. X post by @OpenAI
6. X post by @ClaudeDevs
7. X post by @antigravity
8. X post by @bentosell
9. `llm 0.32.1`

10. llm-openrouter 0.7
11. X post by @kentcdodds
12. X post by @ASpittel
13. X post by @kentcdodds
14. X post by @LangChain
15. X post by @steipete
16. X post by @dawnsongtweets