

Codex Goes Mobile and Practical Agent Loops Get Sharper

Coding Agents Alpha Tracker

2026-05-15

Codex Goes Mobile and Practical Agent Loops Get Sharper

By Coding Agents Alpha Tracker • May 15, 2026

OpenAI's Codex mobile preview was the clearest workflow shift today: coding agents are becoming remote operators you can steer from anywhere. Inside: Riley Brown's deploy-on-every-change setup, Thibault Sottiaux's recurring Codex workflows, and the key releases and harness signals worth tracking.

TOP SIGNAL

- **Codex just moved from “terminal tool” to “remote operator.”** OpenAI's preview puts Codex inside the ChatGPT mobile app so you can start work, review outputs, steer execution, and approve next steps from iOS/Android while the agent keeps running on your laptop, Mac mini, or devbox [1, 2]. Greg Brockman calls it a “huge step forward for universal usage of agents,” and Riley Brown's day-one demo makes that concrete with voice prompting, long-running task notifications, and deploy-on-every-change app iteration from the phone [3, 4].

TRY THIS

- **Set up Codex mobile as a real remote console.** Riley Brown's exact sequence: update Codex desktop and ChatGPT iOS, restart Codex, keep both on the same Wi-Fi, connect from the mobile prompt, and authorize the same account so chats sync [4]. Then switch to **Chats first** for one-off/non-coding tasks, keep **Projects** for coding, use **voice mode** for hands-free prompting, and leave notifications on because some agents run **10-30 minutes** on longer jobs [4]. Plugins are available via @; skills do not show up in the mobile picker yet, but Brown says natural-language requests still hit desktop-side skills, and if you're comfortable with it he

recommends **full access** permissions instead of approving every action [4].

- **Create a deploy-on-every-change skill for phone-based app iteration.** Brown’s setup: enable the Vercel plugin in Codex desktop, then prompt: **make a skill called YOLO Mode ... every single change is deployed to prod ... the public link is sent** [4]. After that, add “please YOLO it” to prompts; he shows it creating a landing page, returning a public link, and reusing the same link after a dark-mode revision request [4]. He also demos **create a full notes app on YOLO mode. Mobile optimized. Use Supabase for DB... like Trello and gets a deployed app with auth and persisted data** [4].
- **Schedule a daily “chief of staff” agent.** OpenAI Codex lead Thibault Sottiaux says he hands off **100+** tasks/day and runs a recurring automation: have Codex go through **Gmail, Notion, and calendar**, summarize the day, flag what is at risk, and schedule it for **9am daily** so the report lands in the inbox [5]. This is a good pattern if your real bottleneck is launch coordination, on-call visibility, or project drift rather than writing the next function.
- **Prompt like a manager; review like a database engineer.** Sottiaux says the biggest lift comes from defining what “good” and “solved” look like, including exact output structure [5]. Mike Krieger says he now hashes out the spec with Claude *before* it writes code so the model converges on a clear North Star [6]. But Simon Eskildsen still manually reviews every line touching TurboPuffer’s database, and Peter Steinberger’s looped **codex /review** skill explicitly stops short of architecture decisions — good reminder to use agent loops for local cleanup, not irreversible system design [7, 8].

WHAT SHIPPED

- **Codex in ChatGPT mobile app (preview).** Start work, review outputs, steer execution, and approve next steps from mobile while Codex runs on laptop/Mac mini/devbox; Greg Brockman called it a “huge step forward for universal usage of agents,” and Romain Huet says it’s live on iOS and Android [1, 3, 2]. OpenAI post [2].
- **OpenClaw v2026.5.12.** OpenAI setup now defaults to **Codex** login; runtime fallbacks and stalled-stream recovery were added; Telegram polling survives stalls; installs/startup got leaner/faster [9]. Steipete says the team has been pushing performance, reliability, security, stability, new **crabbox** automation flows, and automated video QA [10]. Release notes [9].
- **CodexBar 0.26.0.** New integrations for **Kiro, Antigravity, OpenRouter, Kimi**; calmer menus + keyboard nav; better **Codex/Claude** limits and cost scoping; macOS asset and CLI/Homebrew fixes [11]. Re-

lease [11].

- **mcporter 0.11.0.** Steipete says he now uses it mainly as a more stable browser-automation CLI and for agents testing MCPs without restarts; he expects it to matter less as **code mode** spreads across harnesses [12]. Release [12].
- **Harness quality signal from Theo.** His current ranking is rough on Claude Code: he calls it the worst harness, says **OpenCode** has better UX, multi-model support, and cheaper/faster tool-call pruning, says **Cursor** performs better with Opus, and notes that most of his own T3 Code usage is with **Codex 5.5 low/fast** anyway [13].
- **xAI Grok Build (early beta).** New CLI for coding, app building, and workflow automation for SuperGrok Heavy subscribers; xAI says the beta is meant to improve from user feedback, and Theo’s immediate take was “fast and flicker-free” [14, 15]. Try it [14].
- **LangChain Deep Agents 0.6.** New **harness profiles** for open models, **code interpreter** inside the loop, streaming typed projections, **DeltaChannel** checkpoints, and **ContextHubBackend** for skills/policies/memories [16, 17]. Blog [16].

GO DEEPER

- **08:35-09:17 — Riley Brown: YOLO Mode setup.** Best short clip if you want phone-native vibe coding today: Brown shows the exact Vercel-plugin + custom-skill prompt that makes every change auto-deploy to a public URL [4].



Codex Mobile Released and It's Insane (8:34)

- **17:57-19:05** — **Thibault Sottiaux: Codex as daily chief of staff.** Good clip for anyone thinking beyond codegen. He walks through handing Codex recurring coordination work across Gmail, Notion, and calendar, then scheduling the summary for 9am daily [5].
- **27:36-28:37** — **Thibault Sottiaux: define “done” precisely.** Short, reusable prompting lesson: make the output shape explicit and help Codex evaluate its own success instead of giving a fuzzy objective [5].
- **46:40-47:53** — **Mike Krieger: spec before code.** Worth watching if your agents keep producing something technically correct but strategically off. Krieger’s fix is to collaborate on the spec first, then let Claude implement against a crisp North Star [6].
- **Repo to study** — **steipete’s codex-review skill.** Tiny repo, important pattern: iterative review loops are cheap now, but the author explicitly warns they do not replace architecture judgment [8].
- **Release notes to study** — **OpenClaw v2026.5.12.** Read this if you care about real-world agent reliability; the changelog is mostly about recovery paths, defaults, and startup friction, which is where production agent systems actually leak time [9].

Editorial take: the durable edge right now is not “more autonomous” by itself — it’s better remote control, tighter review loops, and much clearer task definitions.

[1, 6, 5, 8]

Sources

1. X post by @OpenAI
2. X post by @romainhuet
3. X post by @gdb
4. Codex Mobile Released and It's Insane
5. Codex for Everyday Work: AI Agents Beyond Coding
6. What Anthropic is building after Claude Code
7. Simon Eskildsen on scaling Shopify, building turbopuffer, and the future of databases
8. X post by @steipete
9. X post by @openclaw
10. X post by @steipete
11. X post by @steipete
12. X post by @steipete
13. I'm done.
14. X post by @xai
15. X post by @theo
16. X post by @LangChain_OSS
17. X post by @LangChain