

QA Loops Tighten: Crabbox 0.11.0, OpenClaw Proofs, CodexBar 0.25

Coding Agents Alpha Tracker

2026-05-11

QA Loops Tighten: Crabbox 0.11.0, OpenClaw Proofs, CodexBar 0.25

By Coding Agents Alpha Tracker • May 11, 2026

Today's signal is operational: coding-agent workflows are tightening around review loops, proof artifacts, and human cleanup. Also worth your attention: Crabbox 0.11.0, CodexBar 0.25, and a few small prompt/workflow tweaks from practitioners actually using this stuff.

TOP SIGNAL

- **The frontier here is QA loops, not one-shot codegen.** Peter Steinberger says he wants Codex to automatically enter `/review` after finishing a task and keep looping until it stops finding bugs [1]. That same mindset is already showing up in production-ish tooling: OpenClaw now has video proof generation for issues, where Codex or a GitHub workflow creates before/after screenshots and Crabbox records the screen, and Peter says Crabbox is essential in his org and helped level up QA [2, 3].

TRY THIS

- **Run a manual self-review loop (Peter Steinberger).** Let the agent finish the task, switch it into `/review`, then repeat that review pass until it stops finding new issues. Peter says this is the behavior he wants Codex to automate; you can emulate the loop manually today [1].
- **Attach proof artifacts to agent fixes (Peter Steinberger).** Have Codex or a GitHub workflow generate before/after screenshots, then use Crabbox for screen recording. Attach those artifacts to the issue or PR so QA can inspect evidence instead of trusting a text summary. Workflow note [2].

- **Budget a cleanup pass after codegen (Theo).** Theo’s shorthand is blunt: after the agent writes code, he removes unnecessary comments and test code. Treat “agent done” as “ready for cleanup,” not “ready to merge” [4].
- **Try a PR-review prompt tweak: ask for social signals (Peter Steinberger).** Peter says he taught Codex to look for social signals when reviewing PRs. He didn’t publish the full prompt, so the actionable move is to add that exact criterion to your review instructions and see what changes [5].

WHAT SHIPPED

- **Crabbox 0.11.0** — adds a Google Cloud provider, repo-local job workflows, AWS Windows WSL2 hydration, and a Blacksmith sync-stall guard. Strong adoption signal: Peter says it is essential in his org and helped level up QA. Release notes [3].
- **CodexBar 0.25** — new providers include Manus, MiMo, Qwen, Doubao, Venice, and more; new features include quota warning notifications, stacked Codex account switchers, and faster cost history via models.dev. Release [6].
- **OpenClaw QA automation** — video proof generation for issues is now in the workflow. Current setup: Codex or a GitHub workflow creates before/after images, Crabbox records the session, and real Telegram login was automated by @obvius. Details [2].

GO DEEPER

- **Quick watch: Theo’s linked X clip.** Theo paired a short video he said he could “play ... all day long” with a very practical note about stripping unnecessary comments and test code from agent output. Watch the demo, then keep the cleanup lesson. Clip [4, 7].
- **Study the OpenClaw proof-generation thread.** The useful bit is the QA recipe: before/after screenshots, screen recording, and real-login automation in one flow. PR comment [2].
- **Study the Crabbox 0.11.0 release notes.** Repo-local workflows plus new cloud targets are the clearest signal that the sandbox layer is maturing for repeatable agent QA work. Release notes [3].
- **Study CodexBar 0.25 if you juggle providers.** The interesting part is operator ergonomics: provider breadth, quota warnings, account switching, and cost history in one small surface. Release [6].

Editorial take: today’s real edge is tighter agent operations — review loops, proof artifacts, and post-generation cleanup are moving faster than raw codegen

novelty. [1, 2, 4]

Sources

1. X post by @steipete
2. X post by @steipete
3. X post by @steipete
4. X post by @theo
5. X post by @steipete
6. X post by @steipete
7. X post by @isaniss29