

# Claude Tag Moves Coding Agents Into Slack; LangSmith Engine Closes the Repair Loop

Coding Agents Alpha Tracker

2026-06-24

## Claude Tag Moves Coding Agents Into Slack; LangSmith Engine Closes the Repair Loop

*By Coding Agents Alpha Tracker • June 24, 2026*

Anthropic's Claude Tag is the strongest workflow signal today: multiple insiders describe a Slack-native coding agent that lives in threads, handles incidents, and writes a large share of code. Also worth attention: LangSmith Engine's trace-to-PR loop, Cursor's new team skill surfaces, and a few OSS agent-infra projects getting real discussion.

### TOP SIGNAL

Anthropic's **Claude Tag** is the clearest workflow shift today: Boris Cherny says the Slack-native agent launched today, and internally it's already used to write PRs, address user feedback, investigate incidents, run data analyses, and answer company-knowledge questions; he says **65% of the product team's new code** is created by Anthropic's internal version [1, 2, 3]. Anthropic engineer @\_catwu says the internal system **merges 65% of product PRs**, and Andrej Karpathy's reaction is that this is not Slack-wrapped RAG but a **deeply integrated, multiplayer** coding-agent product that changes how teams work [4, 5].

“it's not some LLM Q&A with RAG over Slack... it's a different way of working entirely... I work from Slack now.” [5]

### TRY THIS

- **Run incident response in the same thread the humans use.**  
@\_catwu's flow: when the page lands, tag Claude in the incident thread; it pulls graphs, diffs the deploy, identifies root cause, and tags the author. The team approves in-thread, then Claude opens the fix, lands it, watches the metric recover, and resolves the page [6].

- **Use a per-thread agent as both search and executor.** Boris Cherny’s setup starts with @.Claude in a Slack channel. Each thread gets its own sandbox, memory, and permissions; Claude can clone repos, write/test/compile code, answer questions like **What’s the status on X?** or **Who owns this service?**, proactively watch channels, draft PRs, and react with / when a thread is resolved [7, 8, 2].
- **Steal LangChain’s closed-loop improvement pattern.** The durable sequence is simple: **(1)** run the agent and mine weaknesses, **(2)** propose harness improvements, **(3)** confirm the changes help without regressions, then loop [9]. LangSmith Engine turns that into an operational workflow: build the agent, test with datasets/evals, deploy + monitor traces/online evals, then patch, regress-test, redeploy; when traces show errors, eval failures, negative feedback, or new bad behaviors, Engine clusters them into issues and drafts targeted PRs plus new tests/evals for review [10].
- **For messy browser/API exploration, have the agent build the harness first.** Simon Willison used Claude Code for web to build a small playground UI for testing OPFS + Pyodide behavior across browsers—a good pattern when you need a disposable test bench before committing to product code [11].

## WHAT SHIPPED

- **Claude Tag (Slack launch)** — Launched today; tag Claude in-channel and each thread gets its own sandbox, memory, and permissions. Cherny says it can proactively monitor channels, and Anthropic built security layers around model training, classifiers/auto mode, access controls, and channel/workspace boundaries. More: [introducing-claude-tag](#) [1, 8, 7, 12, 13]
- **LangSmith Engine (available now)** — Connect a tracing project and Engine will inspect traces, cluster repeated failures into issues, draft targeted PRs, and propose new examples/evals for the test suite. LangChain says teams like Vanta, Campfire, and Cogent are already catching regressions earlier and cutting triage time [10].
- **Cursor customize updates** — Plugins can now ship prebuilt canvases (example: Atlassian canvas for live issues/projects/docs); Cursor also added a team leaderboard for popular plugins/skills/MCPs with one-click add, and extended team marketplaces to GitLab, Bitbucket, and Azure DevOps alongside local repos. Changelog: [cursor.com/changelog/customize](#) [14, 15, 16]
- **Google Interactions API GA** — One API for Gemini models and agents, with dedicated coding-agent skills, the Antigravity Agent remote Linux sandbox, multimodal tool use, and `background=True` for async long-running interactions [17].

- **Notable comparison: GLM 5.2 vs Opus 4.8** — Jason Zhou shared a same-prompt, same-reference-image frontend test where both models produced a working Three.js logistics dashboard. He also cites GLM 5.2 pricing at **\$1.40 / 1M input** and **\$4.40 / 1M output**, with Opus about **5x more expensive** [18, 19].
- **OSS agent infra worth a look** — In Matthew Berman’s roundup, **Deer Flow** (~74k GitHub stars) stood out as a long-horizon harness built around sub-agents, memory, sandboxes, and skills; **Codebase Memory MCP** (~12k) claims Linux-kernel-scale indexing in 3 minutes with sub-millisecond structural queries and 120x fewer tokens; **Skill Specter** (<10k) scans skills for 65 vulnerability patterns before install [20].

## GO DEEPER

- **1:36–2:19** — **LangSmith Engine’s trace** → issue → PR loop. Best short walkthrough today of a real agent-improvement pipeline: watch traces, cluster patterns, draft the fix, add evals, then keep monitoring after merge [10].



*The Agent for Your Agent. (1:35)*

- **6:20–7:26** — **Codebase Memory MCP on huge repos**. If your agents keep burning tokens just to understand code structure, this is the pitch to examine: Linux kernel indexed in 3 minutes, structural queries in

under 1 ms, 158 languages, 11 harnesses [20].



*You NEED to try these 12 open-source AI projects RIGHT NOW (6:19)*

- **Study Deep Agents alongside LangChain’s loop engineering write-up.** The repo matters, but the bigger takeaway is the pattern: weakness mining, harness edits, regression checks, repeat [9, 21].

*Editorial take: the serious setups are starting to look the same—put the agent inside the real workstream, give it bounded tools and memory, then close the loop with traces, tests, and human approval. [8, 6, 10]*

---

## Sources

1. X post by @bcherny
2. X post by @bcherny
3. X post by @bcherny
4. X post by @\_catwu
5. X post by @karpathy
6. X post by @ClaudeDevs
7. X post by @bcherny
8. X post by @bcherny
9. X post by @sydneyrunkle
10. The Agent for Your Agent.

11. OPFS + Pyodide test harness
12. X post by @bcherny
13. X post by @bcherny
14. X post by @cursor\_ai
15. X post by @cursor\_ai
16. X post by @cursor\_ai
17. X post by @\_philschmid
18. X post by @aibuilderclub\_
19. X post by @jasonzhou1993
20. You NEED to try these 12 open-source AI projects RIGHT NOW
21. X post by @LangChain