Harness Leaderboard: Nine Coding Agents Tested on the Same AI Models

Kabui, Charles

2026-11-09

Read at ToKnow.ai



The Harness Leaderboard is a public scoreboard for developers choosing a coding agent. It keeps the AI model fixed and swaps the program around it, the part that lets a model read files, run commands and fix code (a harness). Hugging Face engineer Joel Niklaus runs nine such programs on three open-weights models across 198 tasks, three attempts each. On the board's November 8 update, mini-swe-agent leads with an average pass rate of 51.4% and

OpenClaw trails at 27.3%. Kimi Code, Hermes Agent, Pi and Claude Code are tied with the leader, because the plausible range for each gap (a confidence interval) includes zero.

Pi, one of the tied five, costs $0.29 per successful attempt, partly because it summarizes conversations early, while OpenCode costs $3.35. A task flips when it passes in one run and fails in another. Rerunning the same program flips 23.4% of tasks, swapping the program flips 29.1%, and running it on another model flips 33.6%, so rerun noise nearly matches the effect of changing the program.

One person runs every trial, and a reviewer notes that no third party has repeated them. No Claude or GPT model took part, so Anthropic's Claude Code and OpenAI's Codex stay untested with their own makers' models. OpenClaw's last place is provisional: an 8,192-token reply cap cut its answers short on two of three models, and a re-run is pending. DeepSWE trials ran online before October 23 and offline since, so that test mixes both.

Read More: DeepSeek's agent program, built so its model, tools and loop are swappable plugins and Google's research on rewriting the instructions and tools around an unchanged AI model

Sources:

- Harness Leaderboard, live board (Hugging Face Space)
- Harness Leaderboard changelog (Hugging Face)
- Pasquale Pillitteri's review of the Harness Leaderboard
- Reliability ranking update (HuggingNews)

Disclaimer: *For information only. Accuracy or completeness not guaranteed. Illegal use prohibited. Not professional advice or solicitation.* ***Read more:*** */terms-of-service*