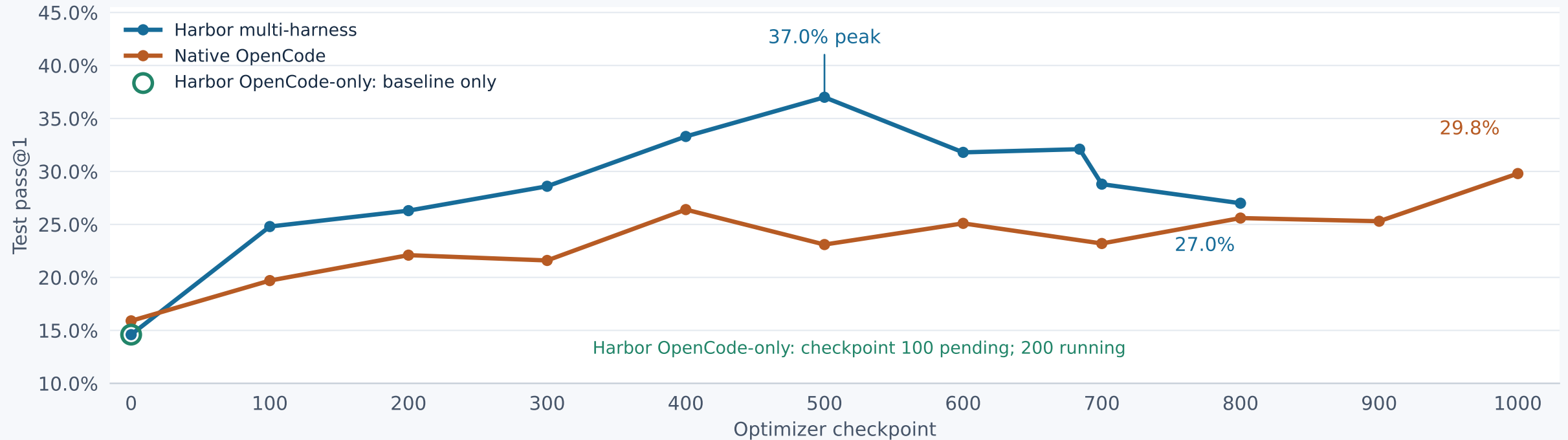


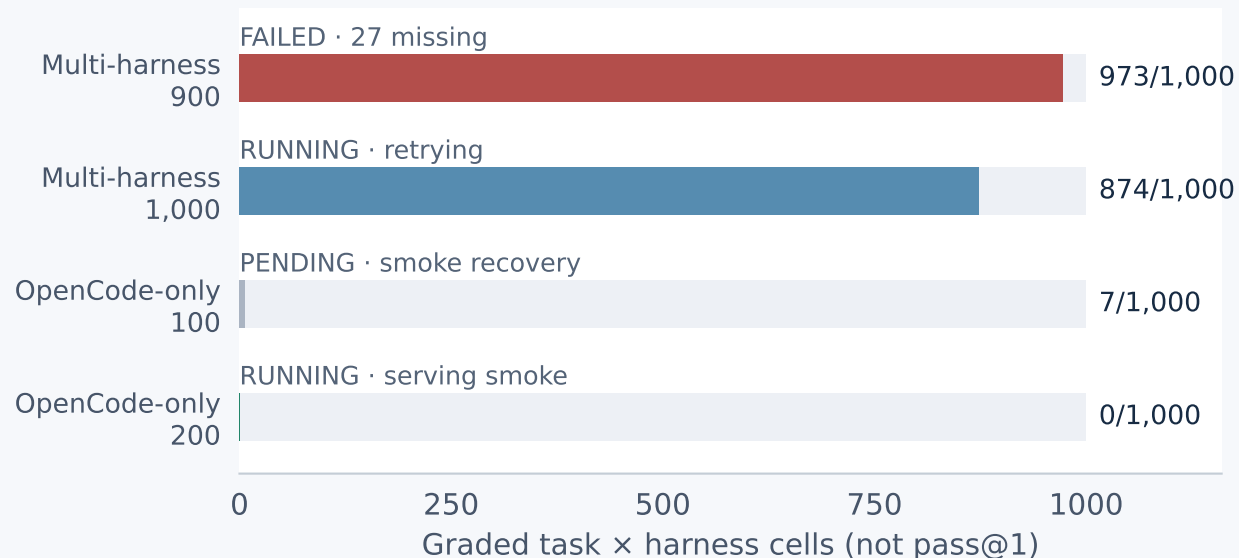
Data Agent · evaluation and training update

16 September 2026, 18:17 UTC | Qwen3.5-2B | Fixed 250-task test set

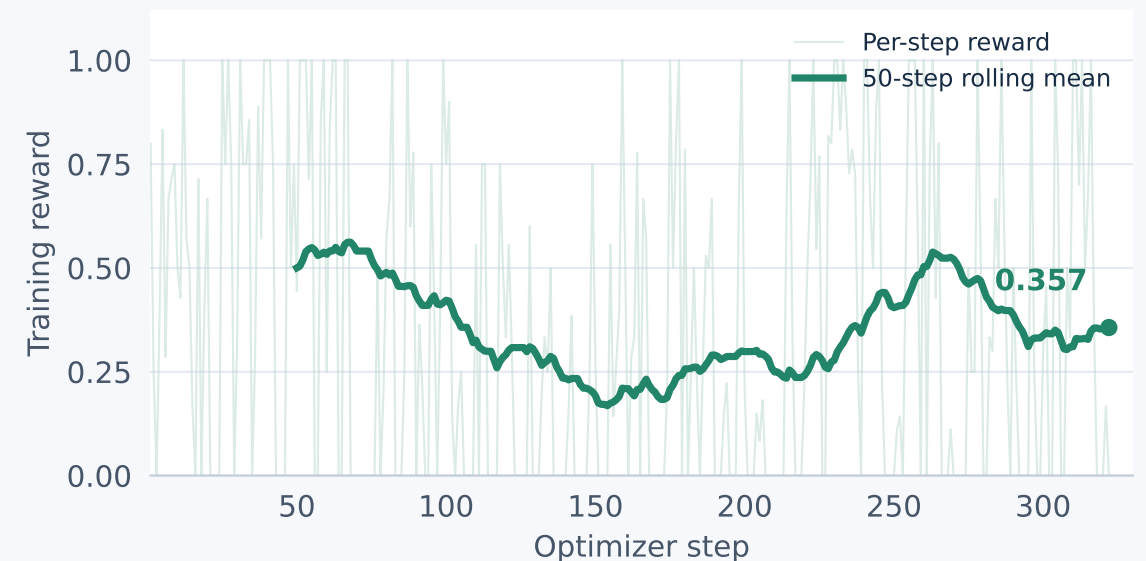
Accepted checkpoint evaluations



Unfinished Harbor evaluations



Harbor OpenCode-only · step 322



Pass@1 curves include only complete, audited 1,000-cell evaluations. Incomplete results are excluded.

Baselines/infrastructure differ across runs. Training reward changes with the task mix and is not held-out evaluation.