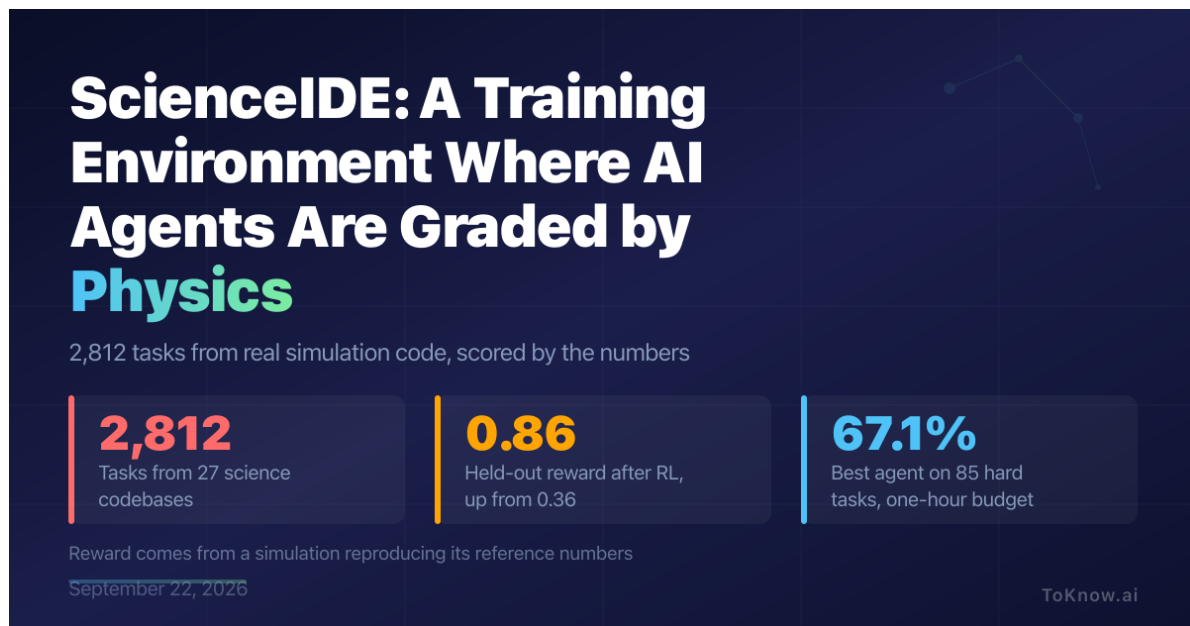


ScienceIDE: A Training Environment Where AI Agents Are Graded by Physics

Kabui, Charles

2026-09-22

 Read at ToKnow.ai



ScienceIDE is a training environment for AI agents built out of real scientific simulation code. Researchers at PhAI Labs and the AItonomy Foundation turned 27 codebases, including MIT-gcm for ocean circulation and Athena++ for astrophysical plasma, into 64 environments holding 2,812 tasks, nearly all repair and implementation, plus 1,076 executable checks of numerical results. Each task starts from a pinned copy of the simulator. The team injects a defect, or

deletes one routine, and a verifier recompiles the code and compares its physics output against reference numbers. Reward comes from the simulation being numerically right again, so an untouched repository scores zero. Fifteen environments are published so far.

Most coding agents are graded against a reference patch or test file, which invites shortcuts. Physics is harder to fake: the simulation either reproduces the numbers or it does not. Training on those trajectories produced the PhAI-IDE family, built on Qwen bases, with the 4B and 9B weights under Apache-2.0. On held-out tasks the 9B model lifted its pass rate on the LAPS plasma code from 31.25% to 50%, and reinforcement learning raised held-out reward there from 0.357 to 0.857. CodeXGLUE defect detection improved 6.99% and BBH word-sorting 36%.

ScienceIDE-Hard shows how far this has to go: 85 one-hour tasks with no hints, where the best agent, Claude Fable 5.1, solved 67.1% and eleven of the fifteen agents tested scored below 40%. The team is explicit that these gains come from selected benchmarks and hinted tasks inside codebases the models already saw, not from unseen scientific code.

Read More: [SWE-rebench V2 mined real repositories into coding-agent tasks the same way](#)

Sources:

- [ScienceIDE: Turning World’s Scientific Codebase into Agent Learnable Environments \(arXiv\)](#)
- [ScienceIDE on GitHub \(published environments, tasks and RL recipe\)](#)
- [ScienceIDE project page \(AItonomy Foundation\)](#)
- [PhAI-IDE model series on Hugging Face](#)
- [PhAI Labs technical report page](#)

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