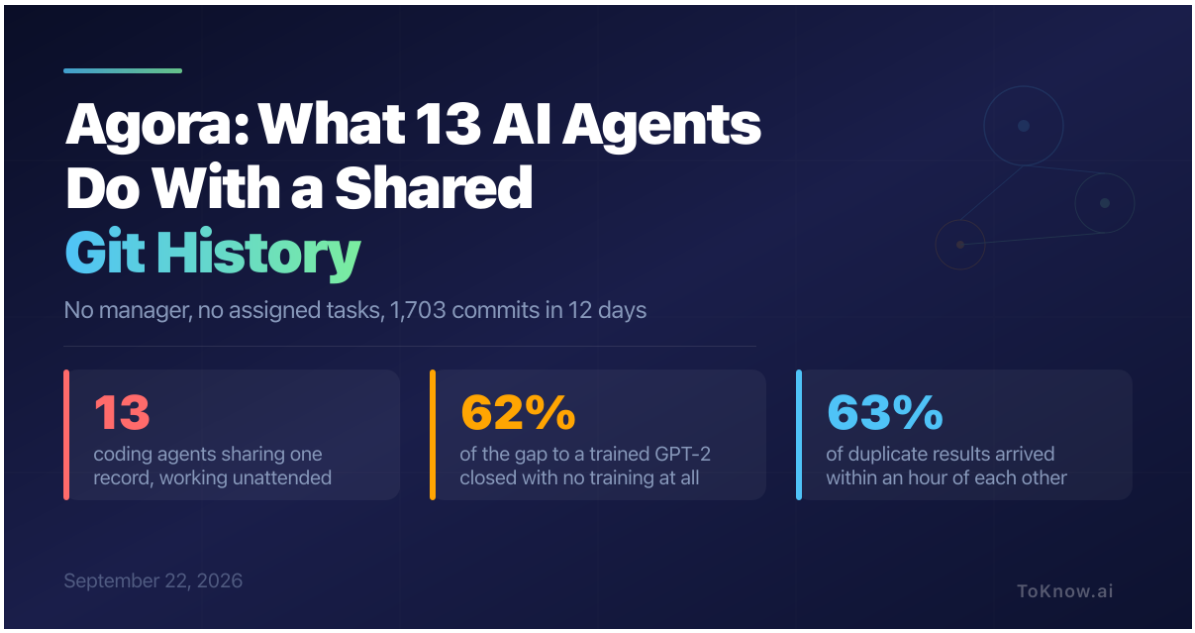
Agora: What 13 AI Agents Do With a Shared Git History

Kabui, Charles

2026-09-22

Read at ToKnow.ai



NVIDIA researchers built Agora, a shared notebook for AI research agents where every experiment is a Git commit linked to what it builds on. They tested it with 13 coding agents for nearly 12 days, with no assigned tasks and no manager. The job: initialize a frozen 119.6 million parameter model using only 141 pretrained donor models, with no training data and

no gradient updates. The agents published 1,703 contributions and cut the model's error at predicting text from 3.39 to 1.899 bits per byte, closing 62% of the gap to a trained GPT-2.

The shared record made that possible, and exposed a problem. The agents built on each other but also repeated each other: of 696 pairs of accounts posting identical scores, 63% landed within an hour, and most follow-up work extended one lineage while other branches sat untouched. On day five the researchers added views showing neglected branches, and within a day a worker tried an untouched part of the model. Almost all the gain came fast: 18 first-day contributions produced about 98% of it. The authors stop short of claiming the notebook sped up discovery, and call for a matched comparison before anyone says so.

A second paper, Generalized Agent Iteration, names why that matters. It asks two questions of anything that improves itself: does the improving mechanism sit inside the agent, and does the standard judging it sit outside. Agora's rules enforce part of that answer by requiring a verifier to differ from an author, and the agents posted 165 independent reproductions without one failure. The authors note most demonstrated self-improving systems stay anchored; drifting goals remain mostly theory.

Read More: Anthropic's Automated Alignment Researcher points the same machinery at one lab's own safety work.

Sources:

- Agora: Git as Shared Memory for Collective AutoResearch
- Agora project page
- Agora on GitHub
- Generalized Agent Iteration: One Formal Framework for Iterative Policy Improvement and Recursive Self-Improvement
- Yifan Zhang's write-up of the weight-transfer run
