SkillClaw: Agent Skills That Improve Automatically From Every User Interaction

Kabui, Charles

2026-04-22

Alibaba's AMAP-ML team released SkillClaw, an open-source framework that makes LLM agent skills improve continuously from real user interactions. Agent frameworks like OpenClaw and Hermes rely on reusable skills (code snippets, tool-use patterns) to handle complex tasks, but those skills stay frozen after deployment. When one user finds a better approach or workaround, that knowledge stays trapped with them. SkillClaw adds two components: a local

API proxy that records session artifacts, and an evolve server with an autonomous evolver that aggregates trajectories across users, identifies recurring behavioral patterns through clustering, and translates them into concrete skill updates. Updated skills sync to all connected users through shared storage (S3, OSS, or local filesystem). The system integrates natively with 10+ agent platforms, including Hermes, Codex, and Claude Code. On WildClawBench's 60 real-world agent tasks, it significantly improved Qwen3-Max performance with limited interaction data.

Most agent skill libraries are write-once. If your coding agent learns to handle a tricky deployment, that insight dies with your session. SkillClaw makes skills compound: a teammate's debugging workaround becomes your agent's capability, automatically. The framework runs silently as a local proxy, requiring zero changes to how you interact with your agent. A validation gate defers updates until safe rollout is confirmed, preventing untested patterns from propagating. For teams running multiple agents across devices, isolated experience silos become a shared, continuously improving library.

The direction is notable: agent capabilities are shifting from static tool definitions toward evolving knowledge bases shaped by real usage. SkillClaw treats multi-user interaction data as the primary signal for improvement, not curated datasets or manual configuration. If this approach scales, the gap between freshly deployed agents and battle-tested ones could shrink fast.

Sources:

- SkillClaw Paper (arXiv)
- SkillClaw GitHub Repository
- WildClawBench: Real-World Agent Evaluation
- SkillClaw on HuggingFace Papers
