EverMind Launches Raven Agent: The Self-Improving Harness That Defines L3-Level Digital Life

July 22, 2026

SAN MATEO, CA - July 22, 2026 -
EverMind, a global AI company incubated by Shanda Group, today announced Raven Agent ? a self-improving agent harness built on EverOS, a portable memory infrastructure that enables agents to learn and improve with every run. Raven is open-source, available now, and can be installed directly from the terminal or self-hosted for team and enterprise environments.

Most agent frameworks are built around model access, task scaffolding, or orchestration logic, with memory treated as a secondary concern. Yet this framing misses a more fundamental problem: the vast majority of AI systems remain stateless by design ? they forget everything the moment a session ends. Industry workarounds such as RAG and extended context windows have improved recall, but they remain, in essence, sophisticated filing systems. The AI retrieves notes; it does not truly remember.

EverMind draws a sharp distinction with Raven: real memory is not retrieval ? it is internalization. A truly memory-capable agent does not look up that you prefer your coffee black; it has absorbed that preference

into its cognitive model of you, and applies it proactively, without prompting. Memory is not an add-on. It is the foundation around which Raven is designed, and that distinction shapes how agents built on Raven learn, adapt, and retain value over time. "We are giving agents a living, ever-evolving history," said Deng Yafeng, CEO of EverMind.

The mechanism behind Raven's self-improvement is specific and reproducible. Each time an agent completes a run, that run is captured as a Case. Over many runs, recurring patterns are distilled offline into reusable Skills, then organized and surfaced through Skill Hub. This loop means that most improvement lives in a portable memory layer ? no retraining required. At the frontier, Raven can also rewrite its own logic and, through EverOS's on-device personalized memory model, fine-tune model weights as an opt-in capability.

EverMind frames agent evolution in four stages. L1 is a role-based functional agent that follows instructions with no persistent memory. L2 is a memory-augmented interactive agent with cross-session memory and multi-step planning. L3 is a self-improving cognitive agent capable of reinforcement learning, self-rewriting code, and model fine-tuning. L4 is autonomous digital life: full data sovereignty, proactive goal pursuit, and an AGI-era digital twin. More than 90% of AI applications remain at L1 or L2. Raven is EverMind's bridge from L2 to L3 ? and the foundation for the eventual L4 transition.

That portability is one of the defining characteristics of Raven. Agent memory is stored as canonical Markdown files, indexed locally through SQLite and LanceDB, and fully readable, editable, and version-controlled through Git. There is no dependency on hosted services, managed databases, or proprietary APIs. Users own their memory assets outright. Memory assets travel with the team across models and environments, compatible with Claude Code, Codex, Gemini CLI, and custom agent builds.

Raven treats memory as a cognitive layer rather than a logging or retrieval system. Its Proactive Engine watches context, routines, memory, and feedback to decide when proactive activity is genuinely useful. Raven decides what is worth retaining, resolves contradictions between stored entries, reasons across time to surface what is relevant, and proactively presents time-sensitive information before a user asks for it. A user who mentions an upcoming trip may receive an unprompted reminder about a related deadline several days later. This foresight capability operates continuously, not only when the agent is actively running a task.

Every piece of recalled memory in Raven can be traced back to its source Markdown file. Raven preserves, archives, and retrieves the right context to keep long-running work coherent. Users can inspect the file a memory came from, edit it directly, and have changes propagate automatically as the local index re-syncs.

Raven supports three parallel memory tracks: User Memory, which captures episodes and ongoing user profiles; Agent Memory, which holds the Skills accumulated through work; and a shared Knowledge Wiki that supports structured, source-backed knowledge with taxonomy, CRUD operations, and topic-based search.

Beyond standard chat interactions, Raven processes PDFs, images, Word documents, spreadsheets, presentations, emails, HTML, and URLs, making it suitable for research-heavy and document-intensive workflows.

For individual users, Raven operates as a persistent 24/7 chief of staff accessible through Discord, Telegram, and WeChat. For developers and teams, Raven provides a pluggable framework where memory, proactivity, and tool routing are fully decoupled components that can be swapped independently without rebuilding the broader system. Raven ships with 100,000 deeply evaluated skills ? and counting; each is continuously evaluated in use, with underperformers retired and high-value ones reinforced.

On the LoCoMo benchmark, EverMind's HyperMem architecture reaches 92.73% SOTA. Raven also delivers sub-500 millisecond p95 query latency and more than 90 percent token savings compared to full-context approaches with proprietary models.

Raven is open-source. The full codebase, installation instructions, and contribution guidelines are available on GitHub. Teams that prefer more controlled deployment can self-host Raven without dependency on EverMind's infrastructure. More information, documentation, and the full product overview are available through the Raven Self-Improving Agent Harness product page.

About EverMind:

EverMind, incubated by Shengda Group, is redefining the future of AI by solving one of its most fundamental limitations: long-term memory. Its flagship platform, EverOS, introduces a breakthrough architecture for scalable and customizable memory systems, enabling AI to operate with extended context, maintain behavioral consistency, and improve through continuous interaction. EverMind's product family includes EverOS, Raven, and EverMe.

###

For more information about Evermind AI, contact the company here:Evermind AISophiaevermind@shanda.comSAN MATEO

Evermind AI

EverMind AI is the creator of EverMemOS, the Memory OS for Agentic AI. We provide long-term memory infrastructure for AI that remembers, adapts, and evolves, serving as the foundational core for the next generation of intelligent systems.

Website: https://evermind.ai/

Email: evermind@shanda.com