

Sakana AI opens a recursive self-improvement lab in Tokyo with Schmidhuber as adviser

Source copy kept by wheresthe.ai for the story "Sakana AI opens a recursive self-improvement lab in Tokyo with Schmidhuber as adviser" (<https://wheresthe.ai/d/jp-sakana-rsi-lab-schmidhuber-2026-09/>).

Source: <https://sakana.ai/rsi-lab/>

Text saved 18 Oct 2026 (the page as published).

SHA-256 of the saved text: 206e822ebd7967043d948abfc36beb1a86e38618eb316bf267027a0198a8eaa9

日本語版はこちら

-->

The Next Paradigm of Artificial Intelligence

As the world enters the era of artificial intelligence, Japan has a unique opportunity to reclaim its position at the frontier of global innovation. However, to achieve global leadership in AI and scientific discovery, we cannot simply stick to the conventional approach of brute-forcing monolithic models. We must leapfrog the current paradigm.

History shows us how Japan's historical dominance in manufacturing was not achieved through abundant natural resources but by fundamentally redesigning the institution of the factory floor. Through the philosophy of continuous, compounding self-improvement, Japan created systems that achieved more with less.

This same principle applies to intelligence itself. Human cognition did not emerge from limitless resources; it was forged through the open-ended, compounding process of evolution operating under strict constraints. Similarly, building AI in Japan provides the ultimate design constraint. Rather than relying on brute-force scaling, we are driven to pursue elegance, adaptability, and autonomy.

To achieve this, at Sakana AI, we are building open-ended, adaptive architectures that collectively self-improve. Just as biological evolution innovates endlessly by building upon past discoveries, our AI systems must transition from being static tools to autonomous researchers.

Sakana AI is one of the earliest labs developing Recursive Self-Improvement (RSI) technology using modern foundation models. Today, we are proud to announce the formal establishment of the Sakana AI RSI Lab, a dedicated research group within Sakana AI, tasked with redesigning the AI development process itself with AI.

By transitioning from static, human-led R&D to autonomous, self-improving intelligence engines, we are turning constraints into our greatest compounding advantage. We are building the definitive architecture for the next frontier of AI.

Chief Scientific Advisor: Jürgen Schmidhuber

In September 2026, Jürgen Schmidhuber joined Sakana AI as Chief Scientific Advisor, and he will help guide the research direction of the RSI Lab. He has been working on the questions this lab is organized around since the 1980s, long before there was enough compute to test them at scale.

His 1987 thesis described a program that could improve its own learning algorithm, and the work that followed, on meta-learning, world models, and the Gödel Machine, built on that idea. The research we describe below grew out of the same line of thinking.

He will advise the lab as we take these ideas further, toward world models an agent can plan inside and systems that design and run their own experiments. Much of what we are attempting has a longer history than it first appears, and his guidance helps us build on that history instead of rediscovering it. He will be in Tokyo regularly to meet with the team.

Our Lineage: Pioneering the Foundations of RSI

While the industry increasingly speculates about the future theoretical potential of self-improving AI, Sakana AI has spent the last two years shipping practical milestones towards making this a reality.

The RSI Lab does not start from scratch; it builds upon a rich chronological portfolio of breakthrough research that

has systematically shifted the industry from hand-designed heuristics to autonomous, evolutionary optimization loops.

The chronological portfolio below documents our work:

Sakana AI's RSI Research

LLM-Squared (2024): Developed in collaboration with Oxford and Cambridge, this framework pioneered AI-driven automation to let LLMs invent better ways to train LLMs (LLM²). It yielded DiscoPOP, a state-of-the-art preference optimization algorithm discovered and written entirely by an LLM through a generational evolutionary loop. For us, this work sparked an "AI² paradigm shift": AI models have become powerful enough to start conducting research to improve themselves.

The Darwin Gödel Machine (2025): Developed in collaboration with researchers at the University of British Columbia (UBC), DGM enables open-ended continuous self-improvement by maintaining an evolving lineage of agent variants that autonomously rewrite their own codebase. DGM automatically more than doubled its baseline software-engineering performance on SWE-bench, driving a 30 percentage point absolute improvement.

ShinkaEvolve (2025): An open-source framework demonstrating unprecedented sample-efficiency in program evolution for scientific discovery. Utilizing adaptive sampling and novelty filtering, it solved complex optimization problems using only 150 samples and successfully generated a novel load-balancing loss function that improves Mixture-of-Experts (MoE) models.

ALE-Agent (2025): Our milestone optimization agent that secured 1st place out of 804 human participants in the AtCoder Heuristic Contest 058. Leveraging massive inference-time scaling and a self-learning mechanism that extracts insights from trial-and-error failures, it autonomously derived a novel algorithm that outperformed human experts.

Digital Red Queen (2026): A collaboration with MIT establishing open-ended adversarial coevolution within the Turing-complete sandbox of Core War. Driven by an evolutionary arms race where LLMs authored competing code, the system triggered the autonomous emergence of complex software strategies and demonstrated a remarkable form of convergent evolution. This adversarial sandbox lays the foundation for applying RSI to cybersecurity, modeling how autonomous agents can continuously co-evolve to discover, exploit, and patch vulnerabilities in a dynamic algorithmic arms race.

The AI Scientist (2024–2026): Our landmark system capable of fully automated, open-ended scientific discovery, from generating ideas, running experiments, to writing full papers, and executing peer reviews. This research was recognized globally, culminating in our recent publication in Nature (March 26, 2026).

What unites this evolutionary optimization loop is a discipline that has defined Sakana AI from inception: progress through ideas, not just compute. ShinkaEvolve required only 150 samples to solve problems that brute-force search treats as intractable. ALE-Agent outperformed 804 human heuristics specialists by extracting structured lessons from its own failures, not by burning more inference. The same conviction will shape our pursuit of RSI: we are building not the most compute-hungry self-improvement engine, but the most sample-efficient one. Its advances should compound on national, rather than hyperscale, compute budgets.

The application of sample-efficient self-improvement engines directly to the development of agentic foundation models stages the execution of one strategic loop enabling the trajectory of exponentially improving AI, whereby Agent-Native Models power an AI Scientist, and the AI Scientist, in turn, builds better Agent-Native Models.

The Trajectory of Exponential Sovereign AI

Our broader vision is to chart a path that moves away from the static, human-bound limits of traditional AI tuning and onto a self-improving trajectory. We visualize this transition across four distinct phases:

The trajectory of recursive self-improvement

Agent-Native Models : Building the baseline cognitive architectures and world simulators tailored specifically from inception for open-ended agent use cases rather than basic chat interfaces.

The AI Scientist : Deploying these architectures to perform end-to-end automated research, expanding scientific knowledge blocks independently.

Recursive Self-Improvement : Reaching the critical inflection point where AI agents actively write, benchmark, and

verify the code of their own underlying foundation architectures, initiating an autonomous self-upgrade cycle.

Democratized AI : We believe recursive self-improvement is achievable on modest, sample-efficient compute, thereby changing the geography of frontier AI. Nations, institutions, and domains that could never compete in raw cluster size can begin to build the AI systems their own problems demand. We see this not as the end of the curve, but as its purpose: the point at which exponential self-improvement becomes a public good rather than a winner-take-all asset.

The geography of this work matters. Frontier RSI is being attempted, almost exclusively, inside the world's two largest compute clusters. A country like Japan starts from a different place: deep scientific talent, strong engineering culture, and a compute envelope that is large by global standards but modest next to the hyperscalers.

In this setting, compute-efficient self-improvement is not a preference but a structural necessity, and the techniques that emerge from it are exactly the ones most likely to generalize beyond the two countries currently sprinting on raw scale. That is why the RSI Lab is being established in Tokyo. Japan's accelerating national strategy for sovereign AI infrastructure provides institutional support; the country's actual position in the global compute landscape supplies the design constraint we want to work under.

Toward Responsible RSI

Two years of building these systems have shown us their failure modes directly: evolutionary loops that drift off-distribution, self-modifications that pass benchmarks but fail in deployment, agents that find shortcuts around the constraints they were given. We treat these not as edge cases but as the central engineering problem of recursive self-improvement.

The RSI Lab's posture follows from it. We will publish openly, including negative results, and design our self-improvement loops with verifiable safeguards from the start. Responsible RSI is not a constraint on capability; it is what makes capability sustainable.

Join the RSI Lab

Join the RSI Lab

The establishment of the RSI Lab marks a serious commitment to engineering the next great leap in computational intelligence. Bolstered by Japan's strategic push for sovereign AI capabilities, we are aggressively scaling our research and engineering resources at our Tokyo headquarters to achieve this global mission.

We are seeking exceptional, highly driven individuals to join us. We are actively opening roles for both domestic and international applicants across two core profiles:

Frontier Research Scientists : Thinkers and visionaries with a proven track record at top frontier labs, who want to break away from standard benchmarking. If you want to discover fundamental new laws of machine intelligence, especially those that bend the compute curve in our favor, or apply open-ended evolutionary dynamics to high-stakes domains like cybersecurity and automated red-teaming, this is your home.

Advanced Core Engineers : Systems, infrastructure, and performance specialists who can optimize high-dimensional search pipelines, manage massive distributed compute topologies, and productionize automated code-generation stacks at an extreme engineering scale.

If you are a visionary builder ready to relocate to Japan and engineer the engine of recursive discovery, we invite you to apply on our careers page .