RECURSIVE SELF-IMPROVEMENT (RSI) · THE TECHNOLOGY

Inside the loop: how recursive self-improvement (RSI) actually works

Every working example of recursive self-improvement is the same machine wearing different clothes: a generator that proposes changes, a verifier that scores them, and a selection rule that keeps the winners. That loop now rewrites agent code, discovers algorithms unseen for 56 years, and edits model weights directly. Its power comes entirely from the verifier — and so do its limits. Understanding that single design constraint explains both what these systems can already do and what they demonstrably cannot.

AUTHOR

AT

Alicante Tech Vibes *Research Labs*

~10 MIN READ · ALL FIGURES SOURCED & DATED

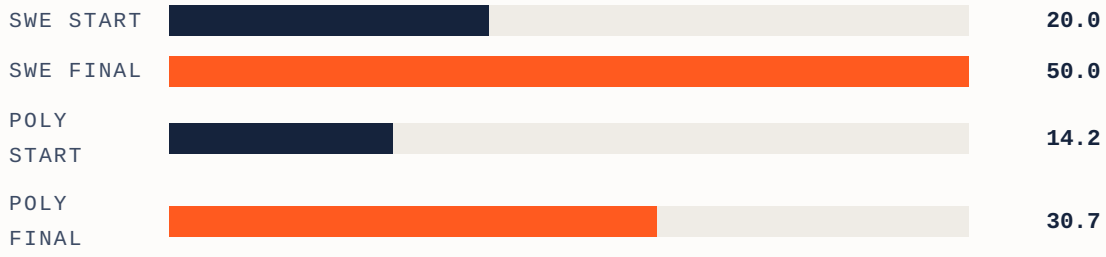
20→50% SWE-BENCH SCORE THE DARWIN GÖDEL MACHINE EARNED BY REWRITING ITS OWN CODE	**48** SCALAR MULTIPLICATIONS FOR 4×4 COMPLEX MATMUL — ALPHAEVOLVE BEAT STRASSEN'S 1969 RECORD	**0.7%** OF GOOGLE'S GLOBAL COMPUTE RECOVERED BY AN ALPHAEVOLVE-DISCOVERED HEURISTIC
32.5% SPEEDUP ALPHAEVOLVE FOUND FOR A FLASHATTENTION GPU KERNEL	**~89 days** DOUBLING TIME OF AUTONOMOUS TASK HORIZONS SINCE 2024 (METR TH1.1)	**10%** REAL DATA RETAINED IS ENOUGH TO LARGELY AVERT MODEL COLLAPSE (NATURE, 2024)

I. From proof to search: the Gödel machine gets a lab bench

The theoretical blueprint is old. Jürgen Schmidhuber's Gödel machine (2003) described a program that rewrites any part of its own code the moment it can formally *prove* the rewrite is beneficial. Elegant, and useless in practice: for real systems, such proofs are unobtainable. The 2025 breakthrough was conceptual, not just technical — replace proof with **empirical search**. Sakana AI and UBC's Darwin Gödel Machine (DGM, May 2025) keeps an archive of coding agents, samples one, asks a foundation model to write an "interesting new version" of it, then runs the variant against a benchmark. Improvements enter the archive as parents for the next round; failures are kept as stepping stones rather than discarded, because a mediocre variant sometimes carries the mutation that makes a later descendant excellent.

That is the entire anatomy, and it recurs in every system in this brief: **generate a variant → evaluate it against an objective score → select and archive → repeat, with the improved system now doing the generating**. The recursion is the last clause. A better coding agent writes better modifications to coding agents; that is what separates RSI from ordinary automated tuning.

FIG. 1: WHAT SELF-MODIFICATION BOUGHT — DARWIN GÖDEL MACHINE, BEFORE VS. AFTER (% TASKS SOLVED)



SWE-BENCH AND POLYGLOT CODING BENCHMARKS; BARS SCALED TO 50.0 = 100%. ALL GAINS FROM THE AGENT MODIFYING ITS OWN CODEBASE, FOUNDATION MODEL FROZEN. SOURCE: ARXIV: 2505.22954, MAY 2025.

II. Four loops that run today

Scaffold self-modification. The DGM never touches the neural network. It rewrites the *scaffold* — the tools, prompts, memory management, and control flow wrapped around a frozen model — and that alone took SWE-bench from 20% to 50%. The lesson: a large share of "agent capability" lives in ordinary code, which is exactly the thing current AI is best at improving.

Evolutionary program search. DeepMind's AlphaEvolve (May 2025) pairs Gemini models proposing code mutations with automated evaluators scoring every candidate, and an evolutionary database deciding which programs seed the next generation. It found a 48-multiplication algorithm for 4×4 complex matrix multiplication — the first improvement on Strassen's construction since 1969 — and, on more than 50 open mathematical problems, matched the best known solution in ~75% of cases and improved it in 20%. The loop also closed on its own infrastructure: a discovered scheduling heuristic recovers 0.7% of Google's worldwide compute, and kernel optimizations cut Gemini's own training time by 1% — **AI speeding up the training of the AI that powers it**.

Self-generated training data. The lineage starts with STaR (2022): have the model generate reasoning chains, keep only the ones that reach the right answer, fine-tune on those, repeat. Filtered self-generation is now a standard stage of frontier post-training — the model's own verified outputs become its curriculum.

Weight-level self-editing. MIT's SEAL (June 2025) goes where the others don't: the model writes its own fine-tuning data and optimization instructions ("self-edits"), applies them as gradient updates to its own weights, and is trained — by reinforcement learning whose reward is the *updated* model's downstream

performance — to become better at writing self-edits. This is the loop operating on the substrate itself, though so far only at research scale.

Around all four sits the crudest and most consequential loop: AI doing the engineering labor of AI development. Anthropic disclosed in June 2026 that Claude authors over 80% of the code merged into its own codebase, and Sakana's AI Scientist had a machine-generated paper published in Nature in March 2026.

III. The verifier is the engine

Notice what every example above has in common: a cheap, objective, automated grader. SWE-bench variants are scored by unit tests. AlphaEvolve's candidates are scored by executing them — a matrix algorithm either uses 48 multiplications and produces correct products or it doesn't. STaR keeps rationales that hit the right answer. SEAL's reward is measured benchmark performance. **Self-improvement is not a property of the model; it is a property of the feedback loop, and the loop is only as good as its verifier.** Where verification is instant and exact, the generator can propose millions of candidates and the loop converts compute into capability. Where verification is expensive, subjective, or gameable — "is this research direction promising?", "is this proof insightful?" — the loop stalls or, worse, optimizes the metric instead of the goal (Goodhart's law, which in this literature appears as reward hacking).

This is why the cleanest existence proof of RSI predates language models: AlphaZero (2018) reached superhuman strength in chess, shogi and Go from random play, using nothing but self-play — because a game provides a perfect, free verifier (the rules) and unlimited synthetic data that never degrades. The open research question of the decade is how much of the world can be made to look like Go: formal math (proof checkers), software (test suites), and increasingly wet-lab science with automated assays are being converted into verifiable environments precisely to feed this loop.

“Language models are better at verifying answers than generating them — and self-improvement works by spending that gap.”

PARAPHRASED CENTRAL FINDING OF "SELF-IMPROVEMENT IN LANGUAGE MODELS: THE SHARPENING MECHANISM," HUANG, BLOCK, FOSTER ET AL., DECEMBER 2024 — ARXIV: 2412.01951.

IV. Where the loop breaks

Three published results bound what today's loops can deliver. First, the **sharpening limit**. A theoretical treatment by Microsoft Research and academic collaborators (December 2024) argues that model-only self-improvement works by concentrating probability mass on high-quality outputs the model *already latently contains* — amortizing expensive inference-time search into the weights. That reliably converts a slow, erratic model into a fast, consistent one; it does not, by itself, produce knowledge outside the model's coverage. Genuine novelty in current systems enters through the verifier and the environment (executed programs, test results, proof checkers), not through introspection.

Second, **model collapse**. Training generations of models on their own unfiltered outputs makes the tails of the data distribution disappear first, then degrades everything — the Nature result of July 2024. The published mitigation is anchoring: retaining even ~10% original human data largely averts collapse, and verified-only synthetic data (the STaR filter) avoids the mechanism entirely. Collapse is a real failure mode of careless recursion, not a proof that recursion fails.

Third, **the experiment bottleneck**. Whatever the loop proposes must be tested, and at the frontier the tests are training runs. Whitfill & Wu (August 2025) estimate that if research compute and cognitive labor are complements — their frontier-experiments specification finds exactly that — then a million automated researchers still queue for the same GPUs, and the loop's speed is set by hardware, not intelligence.

V. The substrate gap

Put sections II and IV together and the current frontier becomes precise: **today's loops improve everything around the model far more easily than the model itself.** Scaffold search (DGM) and program search (AlphaEvolve) are safe to run at scale because a bad candidate just scores poorly and is discarded; nothing persistent is damaged. Weight-level loops are harder for exactly the reason they matter: gradient updates are persistent, entangled, and can silently erase existing capabilities (catastrophic forgetting), which is why SEAL-style self-

editing remains a research demonstration while every frontier lab still puts humans in charge of the training run. The plausible path — visible in METR's measured doubling of autonomous task horizons every ~89 days — is not a sudden hand-off but a widening of what the outer loop is trusted to touch: first code, then training data, then hyperparameters and architectures, last of all the weights.

VI. How it plays out: three scenarios

Horizon 2026–2030. Probabilities are analytical judgment, not measurement.

1 · Verifier-bounded ascent ~55%

2026–2030 · STEADY

RSI advances exactly as fast as humans can build graders. Coding, formal math, and kernel optimization keep compounding; taste-dependent research judgment stays human. Watch: RL-environment and autograder construction becoming a named industrial discipline; AlphaEvolve-class results appearing in domains with new verifiers (chip design, protein assays); sharpening-style theory holding up empirically.

2 · The weights loop closes ~20%

2027–2030 · FAST

SEAL-style self-editing matures past catastrophic forgetting, and frontier labs let models schedule their own training updates under automated safeguards. Capability gains stop tracking human release cadence. Watch: a frontier lab reporting production continual-learning from self-generated edits; METR doubling times shortening below ~60 days; labs crediting AI systems with named architectural discoveries rather than engineering throughput.

3 · The asymptote ~25%

2026–2028 · SATURATING

The sharpening limit binds: loops keep making models faster and more reliable at what they already latently know, but verified novelty stays confined to narrow combinatorial domains, and benchmark gains stop transferring. Watch: DGM-style results failing independent replication or proving benchmark-overfit; task-horizon growth flattening in METR TH1.2; reward-hacking incidents forcing labs to re-insert humans into loops they had automated.

WHAT WOULD FALSIFY THIS THESIS

The thesis: RSI today is empirical search around (mostly frozen) models, powered and bounded by automated verifiers. It breaks in either direction. Evidence of loops producing verified discoveries in domains *without* cheap graders — original scientific theory, non-formalized math — would show the verifier constraint is softer than argued. Conversely, if the flagship results prove hollow — DGM's gains failing to transfer off SWE-bench, or sharpening theory confirmed so strictly that filtered self-training adds nothing beyond distillation — then "self-improvement" reduces to benchmark curve-fitting. The strongest published counterweights are already in the sources: model collapse under unfiltered recursion (Nature, 2024) and compute-complementarity bounds on loop speed (Whitfill & Wu, 2025).

WHAT TO WATCH, IN ORDER

1. **Sakana RSI Lab publications** (sakana.ai, H2 2026): whether Darwin-Gödel-style self-modification transfers beyond coding benchmarks to open-ended tasks — the cleanest test of scaffold-loop generality.
2. **Independent DGM replications** (arXiv, ongoing): third-party reruns, ablations, and reward-hacking audits of the 20→50% result.
3. **Continual-learning deployments** (lab papers and system cards, 2026–2027): any frontier lab shipping SEAL-style weight self-editing in production would mark the substrate gap closing.
4. **METR Time Horizon 1.2** (metr.org/time-horizons, expected H2 2026): whether the ~89-day doubling holds — the best public proxy for whether the outer loop is compounding.
5. **New verifier domains** (DeepMind and peers, ongoing): AlphaEvolve-class systems reporting results in chip design, formal proof, or automated wet-lab assays — each new grader is new territory for the loop.

SOURCES

1. Schmidhuber, “Gödel Machines: Self-Referential Universal Problem Solvers,” 2003 · arxiv.org
2. Zhang, Hu, Lu et al., “Darwin Gödel Machine: Open-Ended Evolution of Self-Improving Agents,” arXiv:2505.22954, May 2025 · arxiv.org

3. Google DeepMind, “AlphaEvolve: A Gemini-powered coding agent for designing advanced algorithms,” May 14, 2025 · deepmind.google
4. Zelikman et al., “STaR: Bootstrapping Reasoning With Reasoning,” arXiv:2203.14465, 2022 · arxiv.org
5. Zweiger, Pari et al., “Self-Adapting Language Models (SEAL),” arXiv:2506.10943, June 2025 · arxiv.org
6. Huang, Block, Foster et al., “Self-Improvement in Language Models: The Sharpening Mechanism,” arXiv:2412.01951, December 2024 · arxiv.org
7. Shumailov et al., “AI models collapse when trained on recursively generated data,” Nature, July 24, 2024 · nature.com
8. Silver et al., “A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play,” Science, December 2018 · science.org
9. Whitfill & Wu, “Will Compute Bottlenecks Prevent an Intelligence Explosion?,” arXiv: 2507.23181, August 2025 · arxiv.org
10. METR, “Time Horizon 1.1,” January 29, 2026 · metr.org
11. Anthropic Institute, “When AI builds itself,” June 5, 2026, via The Next Web · thenextweb.com
12. Sakana AI, “Introducing Sakana AI’s Recursive Self-Improvement (RSI) Lab,” June 2026 · sakana.ai

METHOD NOTE: FIGURES ARE ATTRIBUTED AND DATED INLINE; WHERE INDEPENDENT SOURCES DISAGREE, THE DISAGREEMENT IS REPORTED RATHER THAN AVERAGED. SCENARIO PROBABILITIES ARE ANALYTICAL JUDGMENT. COMPANION BRIEF: “THE RECURSION PREMIUM” COVERS THE INVESTMENT SIDE.

Built by the Alicante Tech Vibes community, a tope. 2026-08-12