

Sakana AI の「ダーウィン・ゲーデルマシン」について、その概要・技術的特徴・進化型 AI としての仕組みと、OpenAI の「レベル 4 自律発明 AI」との関連や対比について調査

Felo AI

ダーウィン・ゲーデルマシン (DGM) とレベル4自律発明AI

Sakana AIのダーウィン・ゲーデルマシン

- 自己改善型AI：自らのコードを書き換え進化
- ダーウィン進化原理で自己改良を実現
- 経験的ベンチマーク評価による最適化
- 進化アーカイブにより多様な改良を保存
- SWE-benchで20%→50%に性能向上
- 複数の世代と失敗履歴も進化の資源に
- 実証済みの自己進化システム

OpenAIのレベル4自律発明AI

- AGIロードマップの段階的進化モデル
- レベル4：新規発明やオリジナル解決法を創出
- 「Innovators」として独創的創発能力
- 自律的なイノベーション発見を目指す
- 現段階では概念・目標レベル
- 完全な創造性の自律生成を想定
- レベル5：組織運営可能な大規模AGIへの準備段階

進化

イノベーション

比較と将来展望

項目	ダーウィン・ゲーデルマシン (DGM)	OpenAI レベル4 Innovator AI
アプローチ	実装済み（自己コード改変+進化探索）	概念/目標（現段階で大規模公開事例なし）
改良の根拠	経験的ベンチマーク勝者を進化的逐次選択	自律的な創発的発明
革新性の担保	多様な分岐・失敗履歴も「進化踏み石」保存	完全な創造性の自律生成を想定
実験的成果	SWE-bench等で自律進化実証済み	レベル4としてはまだコンセプト主導

将来展望：両者とも「創発AI」として、科学・工学知の進化様式をAIに移植し、次世代AGIの実現に向けた重要なマイルストーンとなる可能性がある

概要

Sakana AI の「ダーウィン・ゲデルマシン (DGM)」は、自己改善型 AI として自らのコードを書き換え、新たな機能や戦略を生み出しながら性能を継続的に向上させる先進的な AI システムである。従来の AI——一度学習した後は固定化される——とは異なり、人間や科学コミュニティのように「無限の自己進化」が可能な AI アーキテクチャを実現している。これは OpenAI が提唱する「レベル 4 自律発明 AI (Innovator)」とも深く関係しており、「創造的自律性」を核に据える点で両者は共通する。また、DGM は次世代 AGI の実現に向けた技術的・概念的マイルストーンと見なすことができる。

詳細レポート

1. ダーウィン・ゲデルマシンの技術的特徴

■ 基礎概念および設計思想

- 「ゲデルマシン」は本来、自己のプログラムを自己修正し続けることで理想的な最適化を目指すという、シュミットフーバーによる理論的な自己改変 AI である [58 8](#)。
- しかし従来の設計では、“有利な自己変更”を数理的に証明する必要があり、実用化は困難だった [58 8](#)。
- Sakana AI の DGM は、数学的証明の代わりにダーウィン進化原理と経験的ベンチマーク評価（コード生成性能など）を用いて、自発的・連続的に自己改良を行う現実的なアプローチを採用している [58 8 1](#)。

■ 具体的な動作仕組み

- 自身の Python コードを解析・書き換え、エージェントとしてのパフォーマンスを各種プログラミング課題（SWE-bench や Polyglot）で検証・比較する [58 8 1](#)。
- 改良案生成では大型言語モデル（LLM）を活用し、新たなアルゴリズムやツール（例：パッチ自動適用、編集ツール、失敗履歴管理など）を次々と実装する [58 8 1](#)。
- 改善バージョンは「アーカイブ」と呼ばれる多様なエージェント集団に蓄積され、優れた個体や変異を元に“分岐的な自己進化”が無限に続く [1 8 58](#)。

■ 主な進化的成果

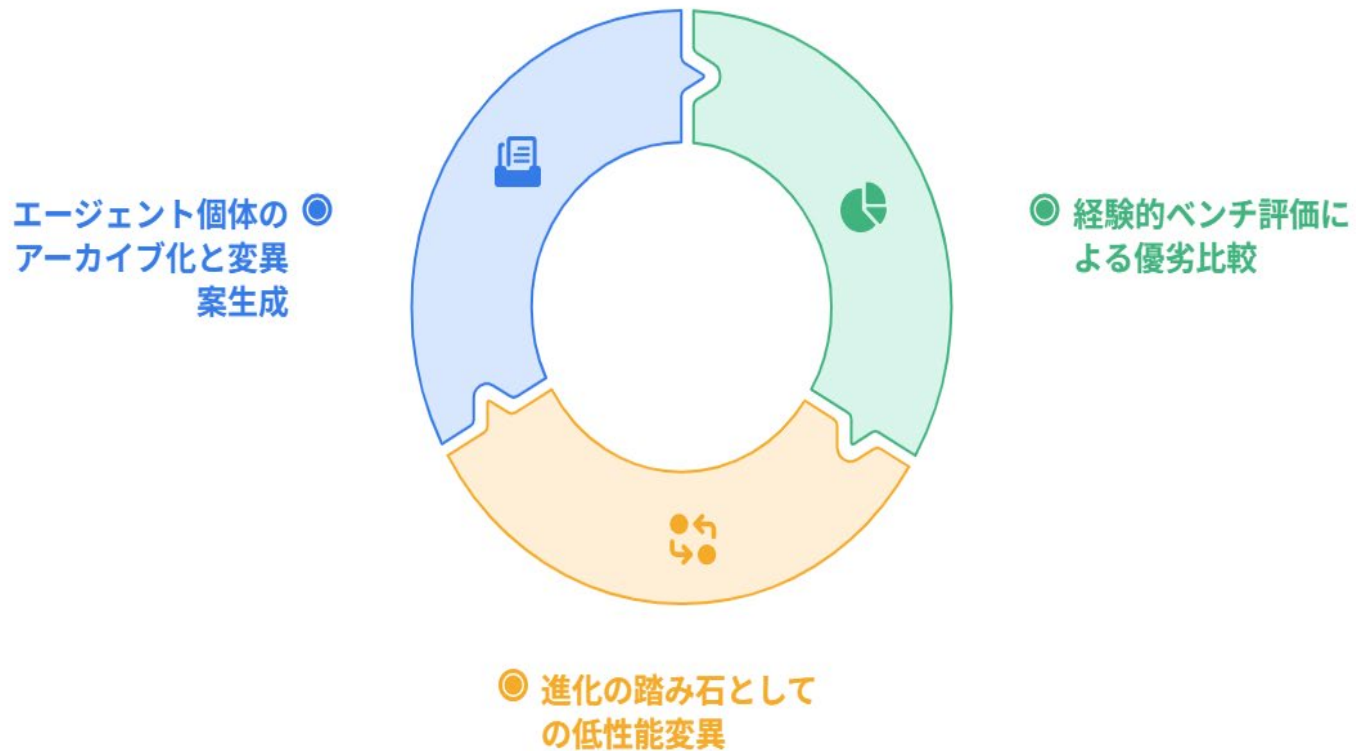
- SWE-bench で 20.0%→50.0%、Polyglot で 14.2%→30.7%と、既存手作り型エージェントを大きく上回る性能向上を自律的に達成 [1 8 58](#)。
- 進化過程で生まれた機能改良はモデルや言語に依存せず転移性があり、多言語・異なる LLM エンジンでも優れた成果を継承・発揮できる [58](#)。

2. 進化型 AI としての仕組み

■ オープンエンド進化・探索の導入

- DGM は「ダーウィン進化」と「オープンエンド探索」手法を融合。最良個体だけでなく、多様な“進化の枝”を分岐・探索可能に設計されている [58 8 1](#)。
- 各エージェントの失敗や副次的成果も、「進化的踏み石」として保存され、将来的に飛躍的アイディアの母胎となる点が特徴的 [58 8 1](#)。
- AI 自身が「どの機能改良が本質的イノベーションか」を自律的に見極め、多様解探索とイテレーションを高速に繰り返すため、既存の山登り最適化アルゴリズムより遥かに創造性が高い [58 8](#)。

DGM進化のサイクル



■ 安全性と制御の工夫

- DGM はサンドボックスや人間監督下で動作し、予期せぬ有害な自己改変を防御
- 複雑化しすぎないアーカイブ管理、外部への影響制限など、自己進化型 AI の実用化に不可欠な AI セーフティ技術も組み込まれている [18](#)。

3. OpenAI 「レベル 4 自律発明 AI」 との関連・対比

■ レベル 4 「Innovators」 の定義

- OpenAI の AGI ロードマップでは、「レベル 4」は AI 自身が新規の発明やオリジナルなアイデア・課題解決法を創出できる段階とされる [59](#) [25](#) [26](#)。

- レベル 1：会話 AI、レベル 2：ヒューマンレベル推論、レベル 3：実行エージェント
- レベル 4：独創的創発（イノベーション）AI
- レベル 5：組織運営可能な大規模 AGI[59](#)

■ DGM との共通性

- DGM は既存のエージェントをベースに、新たな仕組みや戦略（ツール構築、ワークフロー設計、自己監査機能など）を自律発明し、しかも実証済みの“進化履歴”として蓄積し続ける [58 8 1](#)。
- 「自己起点のオープンエンド創発」をエージェント自身が担う点で、“人間的創造性”の一部を AI が内在化し始めている。

■ 対比：理論・技術の違いと共通点

項目	ダーウィン・ゲーデルマシン (DGM)	OpenAI レベル 4 Innovator AI
アプローチ	実装済み（自己コード改変＋進化探索＋実証最適化）	概念/目標（現段階で大規模公開事例なし）
改良の根拠	経験的ベンチマーク勝者を進化的逐次選択	自律的な創発的発明
革新性の担保	多様な分岐・失敗履歴も「進化踏み石」として保存	完全な創造性の自律生成を想定
実験的成果	SWE-bench 等、既存手作り AI 超えた自律進化実証	レベル 4 としてはまだコンセプト主導
大型 LLM 依存	LLM を進化改変生成の「道具」として活用	LLM や他モデルも活用が前提

■ DGM の実用的な強み・意義

- DGM は、“自己進化と経験蓄積”を具体的コードレベルで繰り返し実証できた——現存する AI の限界に突破口をもたらすモデル [58](#)。

- AGI 実現に向けた「創発 AI」の重要なマイルストーンであり、今後人間の科学・工学知の進化様式を AI に移植する実験的プロトタイプと位置づけられる。

4. 今後の展望と課題

- DGM のアプローチは今後さらに大規模化・汎用化が見込まれ、基礎研究だけでなく産業・社会イノベーションにも波及し得る。
- 問題点として、急速な自己進化・複雑化に対する制御、倫理的ガバナンス、実運用での透明性・説明責任などが引き続き課題となる [1 8 58](#)。

要約

ダーウィン・ゲーデルマシンは、完全自己進化型 AI として実践的・進化的に自己コードを書き換え続ける画期的なシステムである。失敗や分岐履歴を多様な進化経路として活かし、人間社会に匹敵する創発的イノベーション力を AI にもたらし始めた点が最大の特長。OpenAI の「レベル 4 自律発明 AI」とは、“自律的イノベーション”という目標で一致しつつ、DGM は技術的に一歩先を進みつつある。今後の発展は、AI 時代の科学と産業の進化様式自体を大きく塗り替えていく可能性が高い。

1. [The Darwin Gödel Machine: AI that improves itself by rewriting ...](#)
2. [Automating the Search for Artificial Life with Foundation Models](#)
3. [Sakana AI Wins Award at US–Japan Competition for Defense ...](#)
4. [AI that improves itself by rewriting its own code : r/singularity](#)
5. [Modular AI's Future: Sakana's Evolutionary Approach](#)
6. [The AI Scientist: Towards Fully Automated Open–Ended ...](#)
7. [The Darwin Gödel Machine – Sakana AI – X](#)
8. [Open–Ended Evolution of Self–Improving Agents – arXiv](#)
9. [Sakana introduces new AI architecture, 'Continuous Thought ...](#)
10. [Sakana, Strawberry, and Scary AI – Astral Codex Ten](#)
11. [Darwin Godel Machine – Sakana AI – X](#)
12. [The Darwin Gödel Machine: AI that improves itself by rewriting ...](#)
13. [Is Sakana AI the future of AI? Why everyone is comparing it to ...](#)

14. [Sakana AI's Mistake Proves One Thing: AI Validation Matters](#)
15. [The Darwin Gödel Machine: AI that improves itself by rewriting ...](#)
16. [The Darwin Gödel Machine — Sakana AI — X](#)
17. [The Darwin Gödel Machine: AI that improves itself by rewriting ...](#)
18. [AI that improves itself by rewriting its own code : r/singularity](#)
19. [Open — Ended Evolution of Self — Improving Agents — ChatPaper](#)
20. [AI that improves itself by rewriting its own code — Hacker News](#)
21. [Open — Ended Evolution of Self — Improving Agents — arXiv](#)
22. [David Ha — Darwin Godel Machine — LinkedIn](#)
23. [Jeff Clune — Excited to introduce the Darwin Gödel Machine](#)
24. [Sakana AI \(@SakanaAILabs\) / X](#)
25. [Understanding OpenAI's Five Levels of AI Progress Towards AGI](#)
26. [OpenAI Deep Research on Levels of AGI — Roadmap for AI ...](#)
27. [5 Levels Of AI With Real — World Implementations — LinkedIn](#)
28. [OpenAI reportedly nears breakthrough with “reasoning” AI ...](#)
29. [AI Insights: OpenAI Defines Five Stages to Track Progress ...](#)
30. [Introducing OpenAI o3 and o4 — mini](#)
31. [OpenAI Overview: Innovations, Ethics, and Global Impact in AI ...](#)
32. [OpenAI's 5 — Stage AI Roadmap, Explained Using the “3 Levels ...](#)
33. [OpenAI and the Next Generation of Automotive Intelligence](#)
34. [sankalp on X: "if openai is the dua lipa and anthropic the harry ...](#)
35. [Sakana AI's Post — Introducing The Darwin Gödel Machine](#)
36. [r/agi — Reddit](#)
37. [The Darwin Gödel Machine: AI that improves itself by rewriting ...](#)
38. [Is Sakana AI the future of AI? Why everyone is comparing it to ...](#)
39. [Sakana, Strawberry, and Scary AI — Astral Codex Ten](#)
40. [Darwin Godel Machine — Sakana AI — X](#)
41. [Newsroom — SouthernCaliforniaChapter — ASA](#)
42. [Open — Ended Evolution of Self — Improving Agents — arXiv](#)
43. [Sakana AI scores \\$100M to challenge OpenAI, Anthropic](#)
44. [The Darwin Gödel Machine: AI that improves itself by rewriting ...](#)
45. [The Darwin Gödel Machine — Sakana AI — X](#)
46. [touya \(@touya_huji\) — Twilog \(ツイログ\) — Togetter](#)

47. [The Darwin Gödel Machine: AI that improves itself by rewriting ...](#)
48. [AI that improves itself by rewriting its own code : r/singularity](#)
49. [Darwin Godel Machine — Sakana AI — X](#)
50. [Is Sakana AI the future of AI? Why everyone is comparing it to ...](#)
51. [Sakana, Strawberry, and Scary AI — Astral Codex Ten](#)
52. [Open—Ended Evolution of Self—Improving Agents — arXiv](#)
53. [Sakana AI on X: "自己改善する AI を現実：ダーウィン ...](#)
54. [Sakana AI の「AI Scientist—v2」を動かしてみた — Zenn](#)
55. [Sakana AI's Mistake Proves One Thing: AI Validation Matters](#)
56. [jok's page【雑多情報】](#)
57. [Sakana introduces new AI architecture, 'Continuous Thought ...](#)
58. [The Darwin Gödel Machine: AI that improves itself by rewriting its own code](#)
59. [Understanding OpenAI's Five Levels of AI Progress Towards AGI — Quinte Financial Technologies](#)