Toward Verifiable AI-Generated Science: Lessons from The AI Scientist

Yutaro Yamada, Sakana AI
NLP Colloquium, 06/03/2026

自己紹介:

- Yale University PhD in Statistics and Data Science

Yale

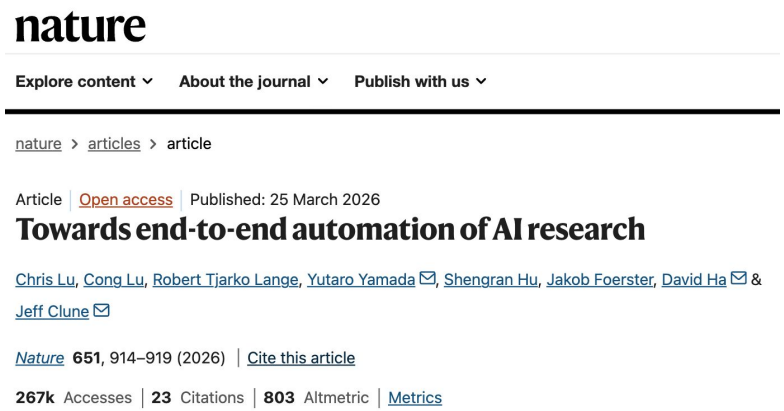
- Research Scientist at Sakana AI

sakana.ai

- Co-led the development of AI Scientist v2

  Website: yutaroyamada.com

  X/Twitter: @_yutaroyamada



nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › articles › article

Article | Open access | Published: 25 March 2026

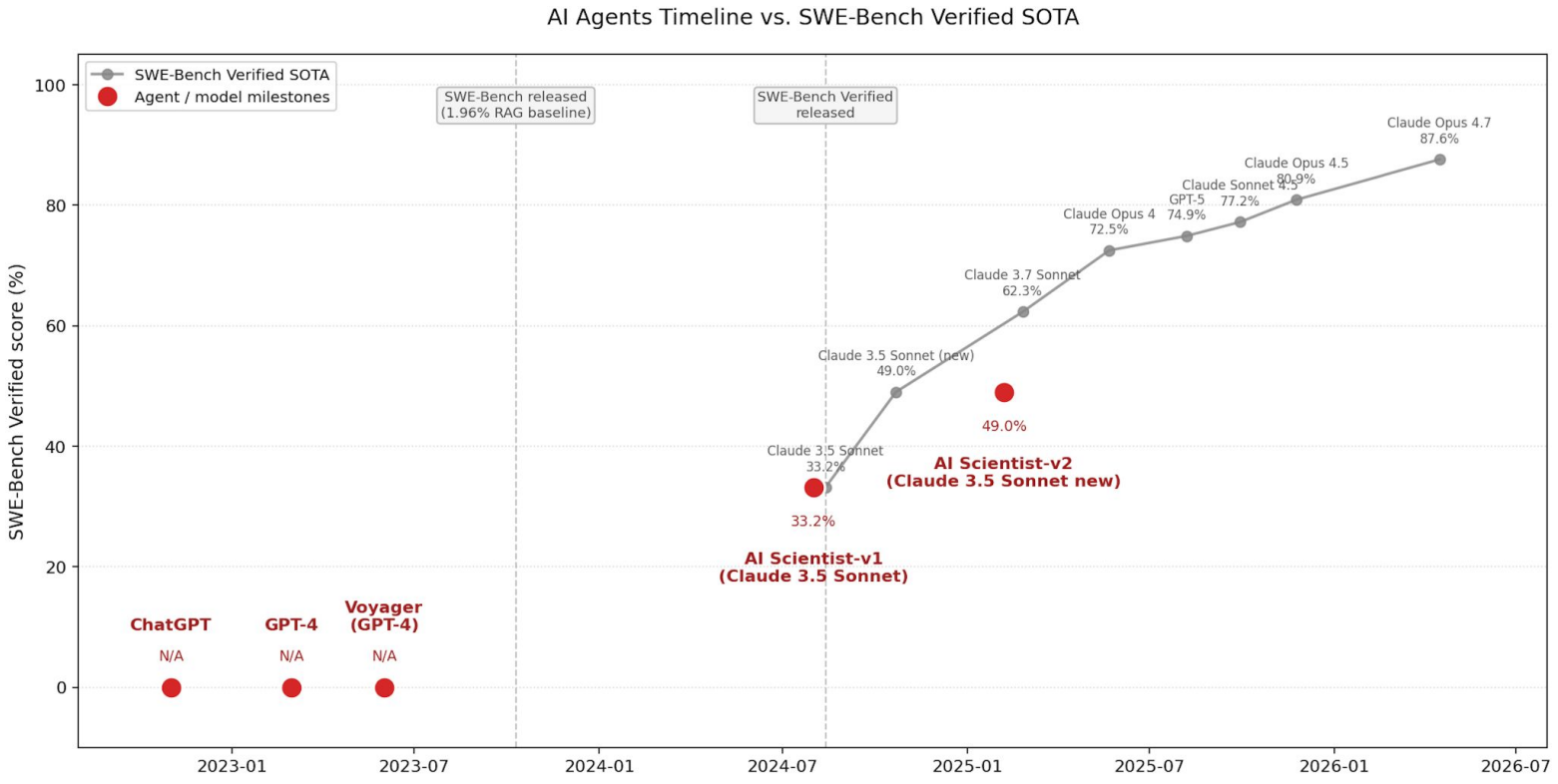
Towards end-to-end automation of AI research

Chris Lu, Cong Lu, Robert Tjarko Lange, Yutaro Yamada ✉, Shengran Hu, Jakob Foerster, David Ha ✉ & Jeff Clune ✉

Nature **651**, 914–919 (2026) | Cite this article

267k Accesses | **23** Citations | **803** Altmetric | Metrics

Model release date vs. SWE-Bench Verified Score



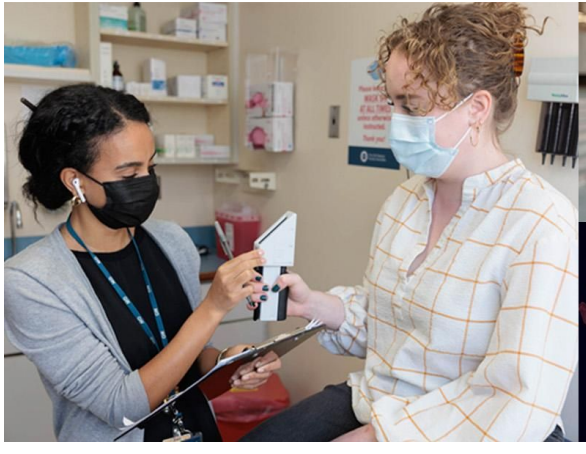
All scores are on SWE-Bench *Verified* (500-task human-validated subset; released Aug 13, 2024). The original SWE-Bench launched Oct 10, 2023, so ChatGPT / GPT-4 / Voyager predate the benchmark (N/A).

Agenda

- The AI Scientist as a frontier stress test
- The landscape of AI systems for AI science
- Discovery as closed-loop tree search

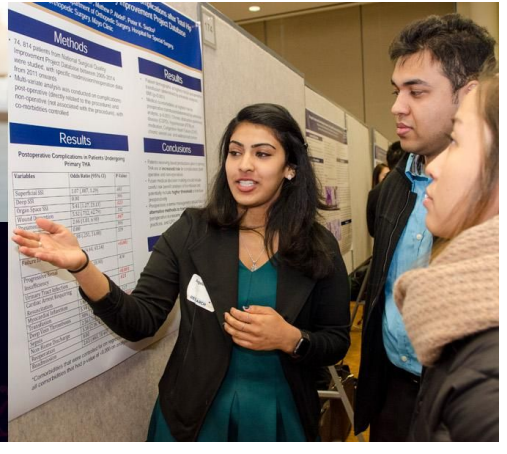
Agentic Taskとしての科学研究

問題を見つける

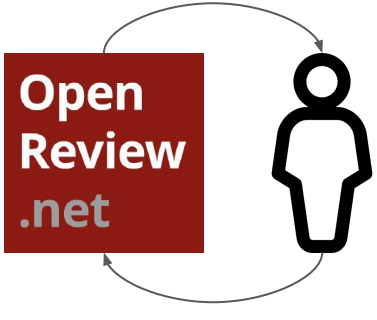


仮説を立てる

実験をする



査読



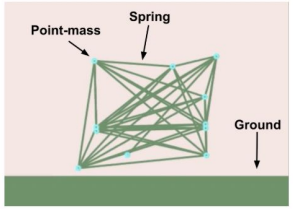
出版

知識の伝達

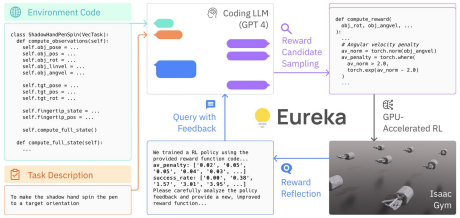
現在のLLMを科学者としてどこまで活用できるか？

LLMs are used all over the place to automate parts of research...

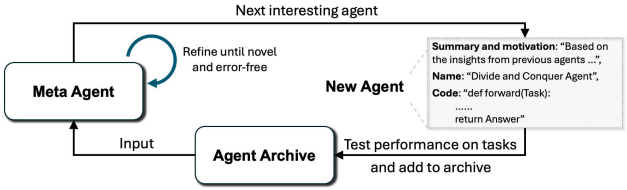
Lehman et al. (2023): ELM
Morphogenesis with LLMs



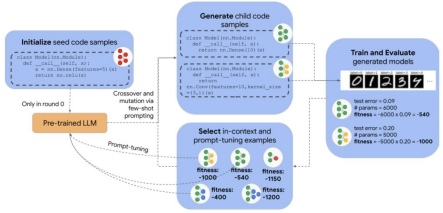
Ma et al. (2023): Eureka
Reward Fn Design with LLMs



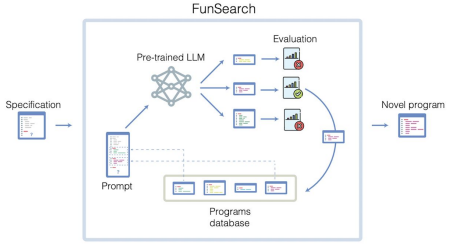
Hu et al. (2024): ADAS
Agentic Design with LLMs



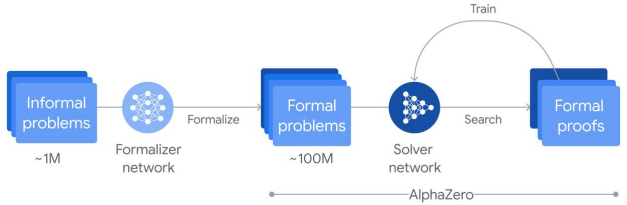
Chen et al. (2023): EvoPrompting
Architecture Search with LLMs



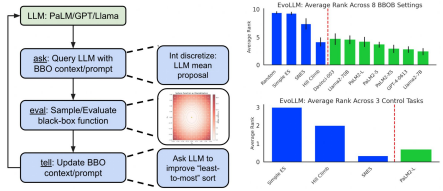
Romera-Paredes et al. (2024): FunSearch
Math/code Discovery with LLMs



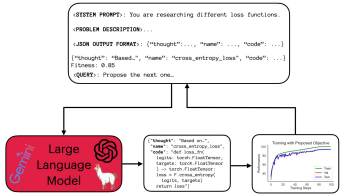
Hu et al. (2024): AlphaProof/Geometry
Math/Geometry with LLMs



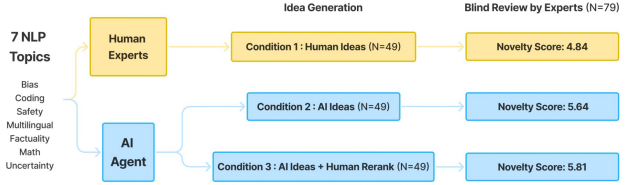
Lange et al. (2024): EvoLLM
Evolutionary Optimization with LLMs



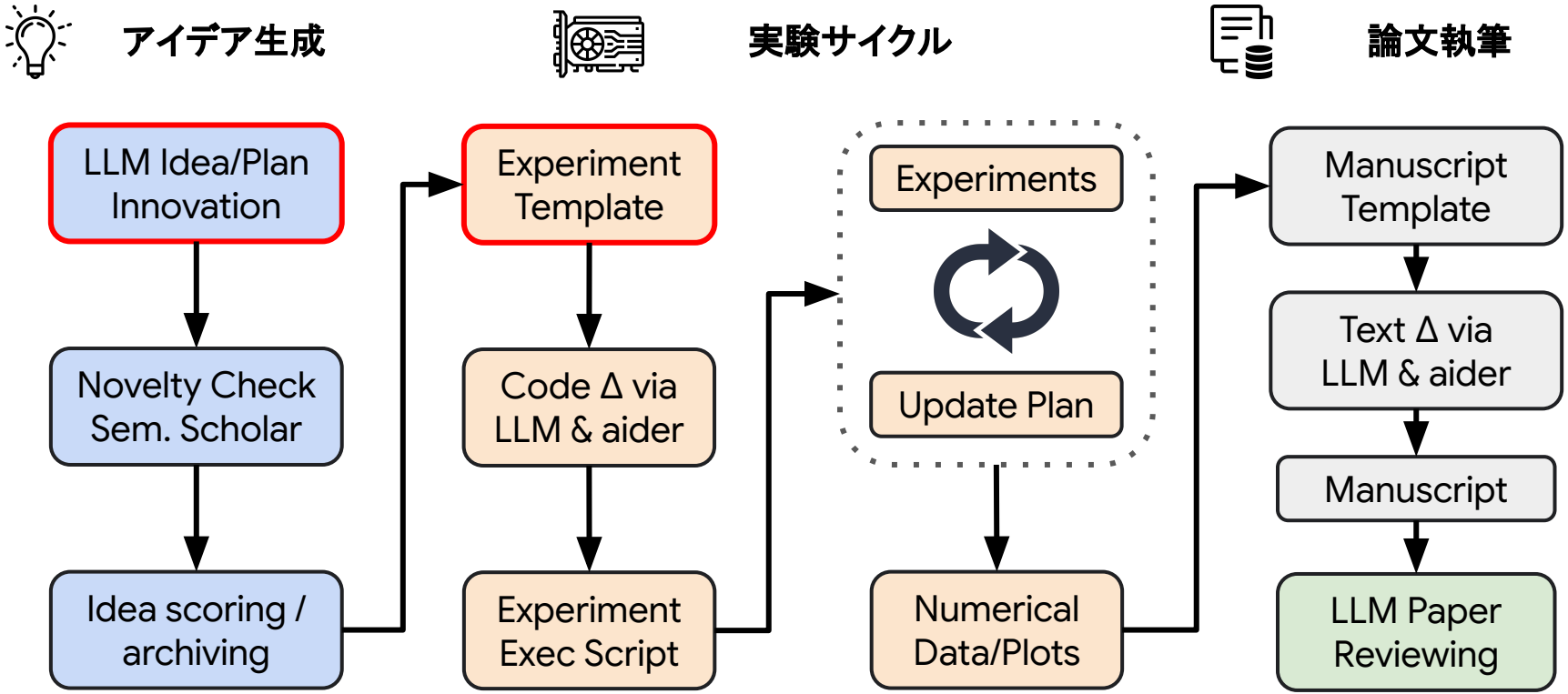
Lu et al. (2024): DiscoPOP/LLM²
Loss Fn Discovery with LLMs



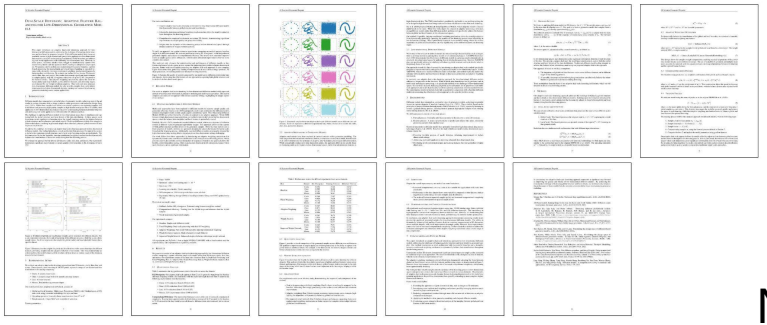
Si et al. (2024): Automated Ideation
Idea Generation with LLMs



The AI Scientist-v1：テンプレートベースの自動研究パイプライン



The AI Scientist-v1: 3種類の実験テンプレートをベースに論文を自動生成



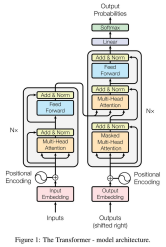
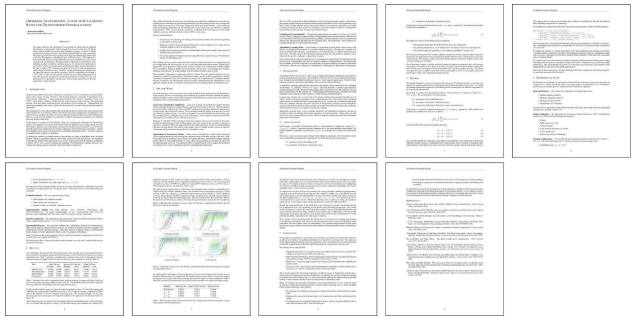
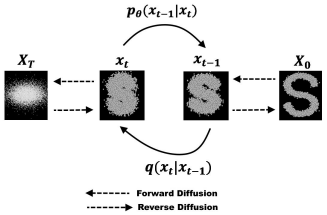


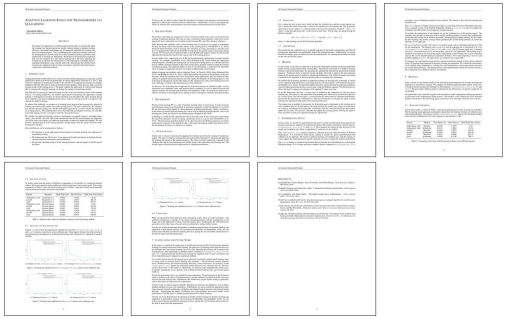
Figure 1: The Transformer - model architecture.

NanoGPT

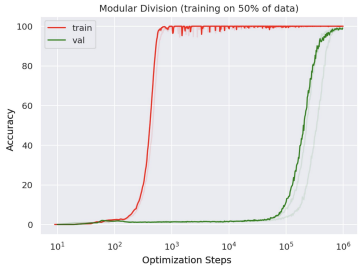


2D Diffusion



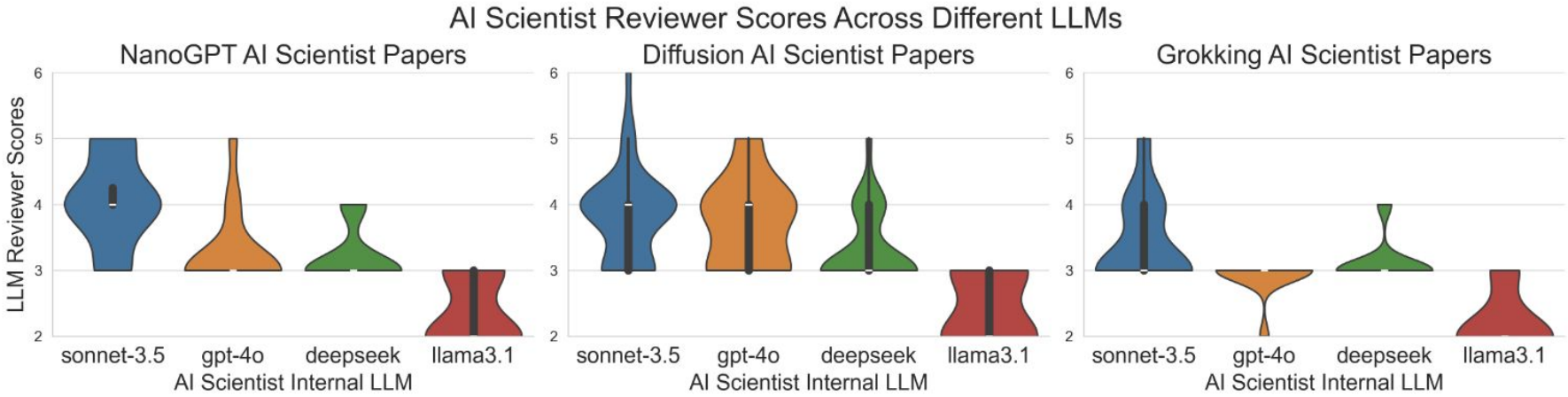


Grokking



👉 AI Scientist-v1 では、機械学習の 3 つの異なるサブフィールドを選び、研究対象とした。

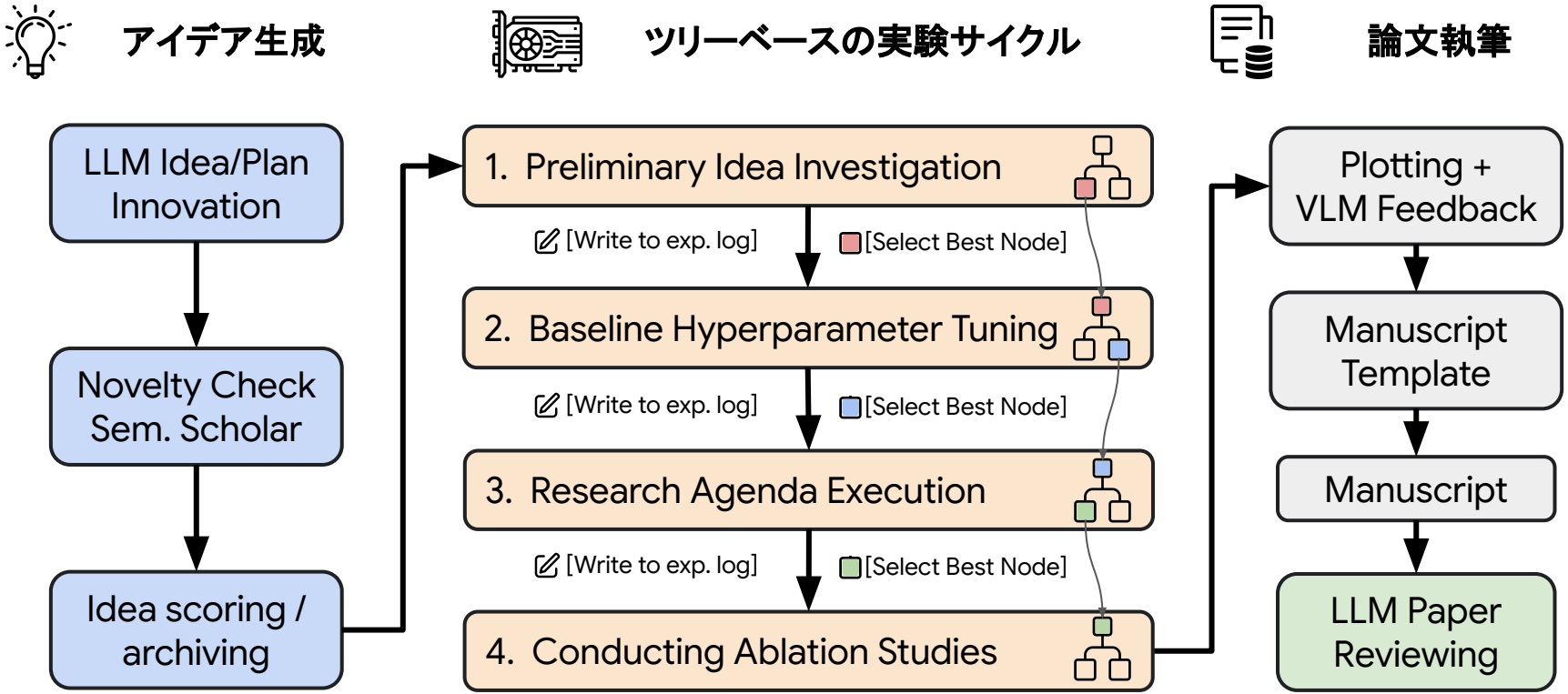
AI Scientist-v1：実験結果 & 課題



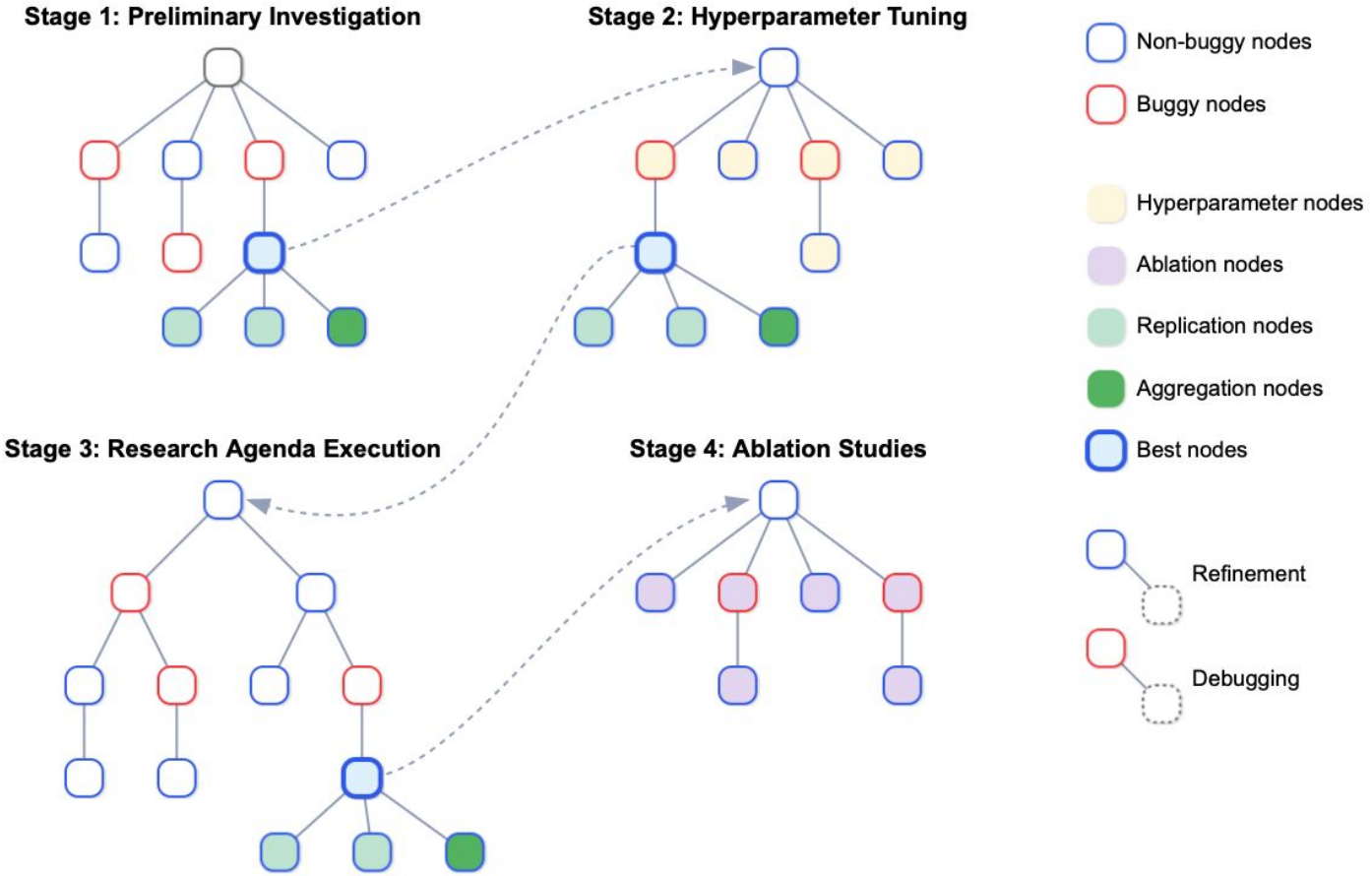
Feature	Codebase Drafting	Execution Planning	Parallel Experiments	VLM Reviewer	Human Result Evaluation
The AI Scientist-v1	Topic-Specific	Linear	✗	✗	Not Submitted
The AI Scientist-v2	Domain-General	Tree-Based	✓	✓	Workshop Acceptance-Worthy

👉 仮説検証と知識蓄積の深さを拡張するためのツリーサーチ

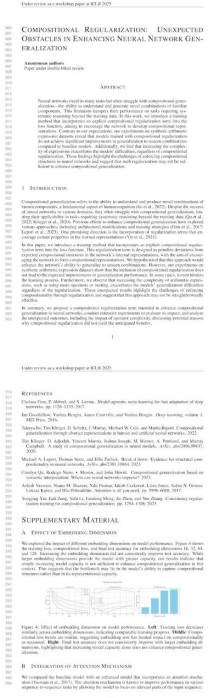
The AI Scientist-v2：ツリー探索に基づく自動化研究パイプライン

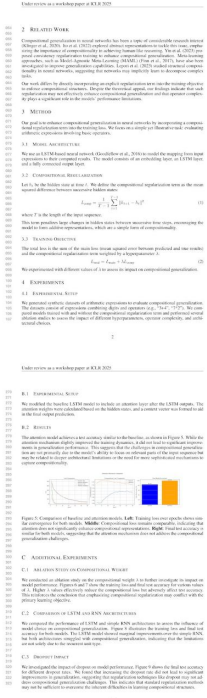


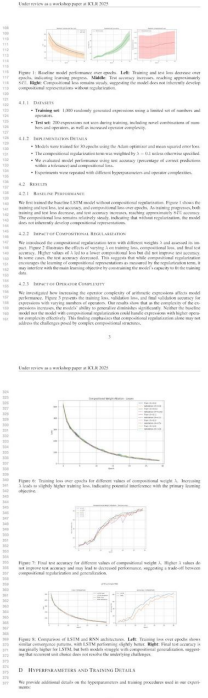
多段階のツリーベース手法による実験の自動化

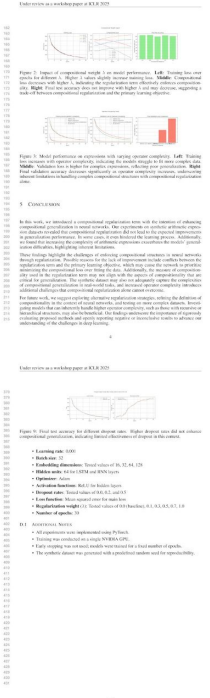


The AI Scientist-v2：AIが生成した初のワークショップレベルの論文









Publications

Real-World Challenges in Pest Detection Using Deep Learning: An Investigation into Failures and Solutions
Aislin Scirocco
07 Feb 2025 (modified: 06 Mar 2025)
ICLR 2025 Workshop ICBINB Desk Rejected Submission

👉 Scores: [3,4,7]

Unveiling the Impact of Label Noise on Model Calibration in Deep Learning
Aislin Scirocco
07 Feb 2025 (modified: 06 Mar 2025)
ICLR 2025 Workshop ICBINB Desk Rejected Submission

👉 Scores: [3,3,3]

Compositional Regularization: Unexpected Obstacles in Enhancing Neural Network Regularization
Aislin Scirocco
07 Feb 2025 (modified: 06 Mar 2025)
ICLR 2025 Workshop ICBINB Desk Rejected Submission

👉 Scores: [6,6,7]

Released in March/April 2025

ICINB ICLR 2025ワークショップ運営事務局とのコラボレーション

I Can't Believe It's Not Better:
Challenges in Applied Deep Learning

Workshop at ICLR 2025

AI-Generated Papers 👉 3 papers submitted with organizer support + IRB [H24-02652]

As part of a small experiment that we believe aligns with the theme of our workshop (and with approval from the central ICLR workshop chairs), we have included 3 AI-generated papers out of a total of 43 submissions. As a result, it is possible, though unlikely, that you may be assigned an AI-generated paper to review. If you prefer not to review AI-generated papers, please let us know by **February 11 AoE** by emailing our official email (cant.believe.it.is.not.better+workshop@gmail.com). We will review your assignments and reassign papers accordingly.

👉 透明性・査読者のオプトアウト・IRB取得が不可欠！

AI Scientist は依然として``ハルシネーション”問題を抱えている

3 METHOD

Our goal is to enhance compositional generalization in neural networks by incorporating a compositional regularization term into the training loss. We focus on a simple yet illustrative task: evaluating arithmetic expressions involving basic operators.

3.1 MODEL ARCHITECTURE

We use an LSTM-based neural network (Goodfellow et al., 2016) to model the mapping from input expressions to their computed results. The model consists of an embedding layer, an LSTM layer, and a fully connected output layer.

Comment: This should be Hochreiter & Schmidhuber

3.2 COMPOSITIONAL REGULARIZATION

Let h_t be the hidden state at time t. We define the compositional regularization term as the mean squared difference between successive hidden states:

$$T-1$$

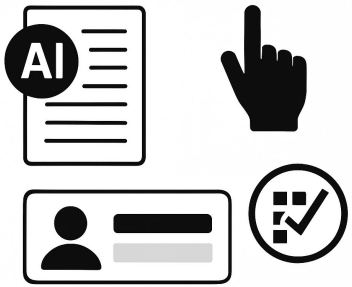
Comment: This should be more precise.

👉 Check paper & Github for full analysis:
github.com/SakanaAI/AI-Scientist-ICLR2025-Workshop-Experiment/

今後の技術的・倫理的課題

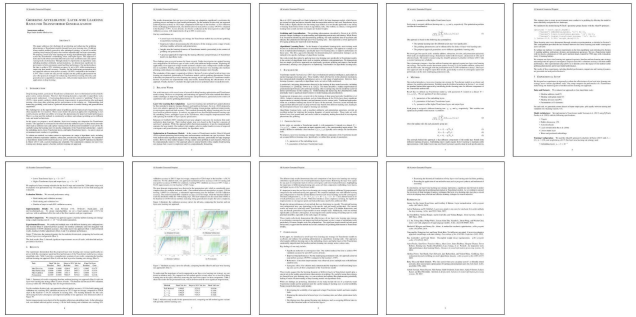
- 透明性の確保

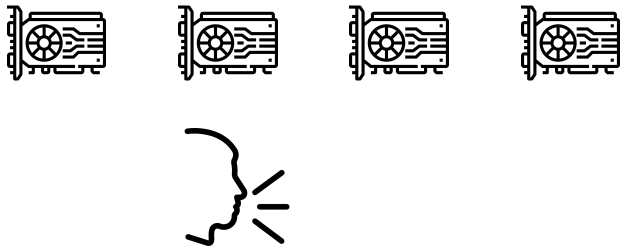
  ⇒ いつ・どのように開示すべきか？

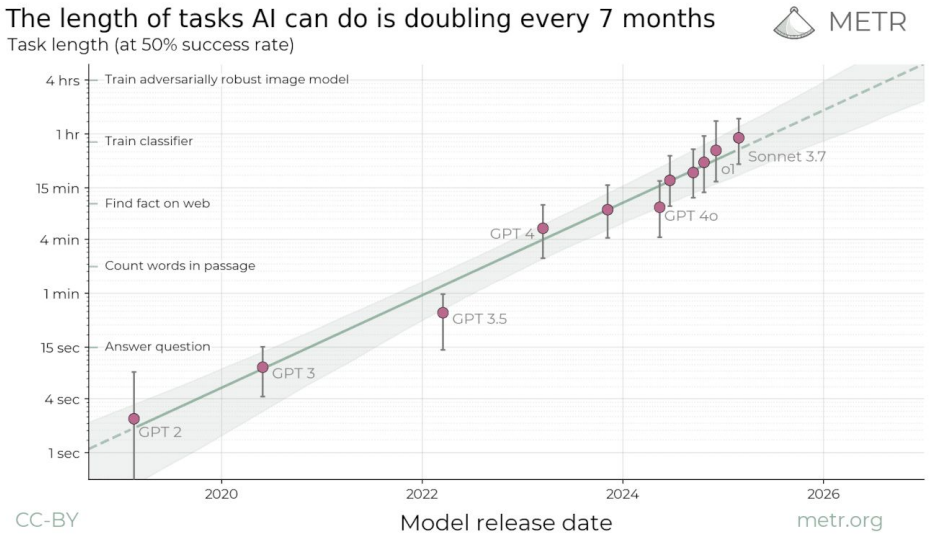


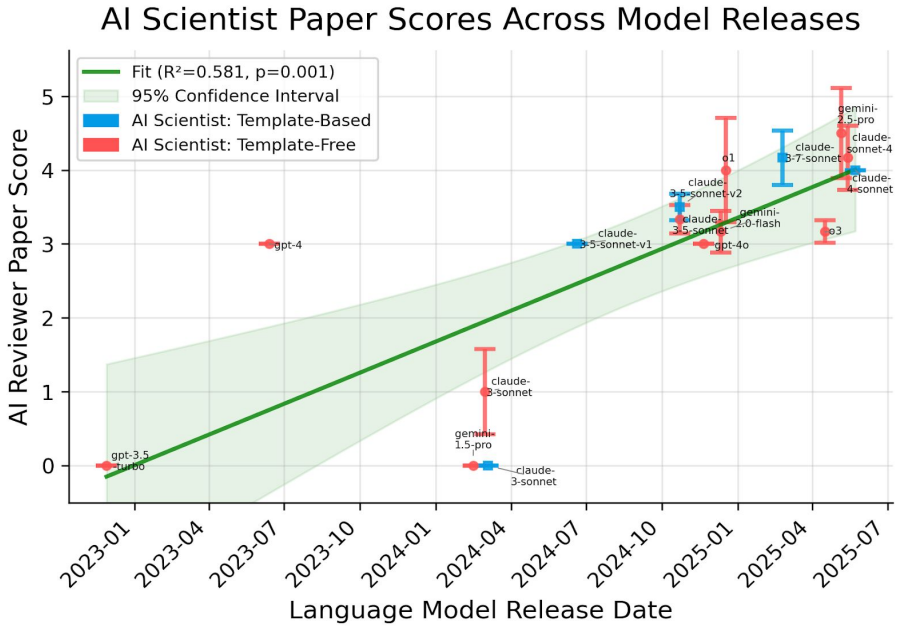
- 悪用対策

  ⇒ どのように査読制度の形骸化（ゲーミング）を防ぐか？






The length of tasks AI can do is doubling every 7 months
Task length (at 50% success rate)
METR
4 hrs
1 hr
15 min
4 min
1 min
15 sec
4 sec
1 sec
Train adversarially robust image model
Train classifier
Find fact on web
Count words in passage
Answer question
GPT 2
GPT 3
GPT 3.5
GPT 4
GPT 4o
o1
Sonnet 3.7
2020
2022
2024
2026
Model release date
CC-BY
metr.org


AI Scientist Paper Scores Across Model Releases
Fit (R²=0.581, p=0.001)
95% Confidence Interval
AI Scientist: Template-Based
AI Scientist: Template-Free
AI Reviewer Paper Score
0
1
2
3
4
5
gpt-3.5-turbo
gpt-4
claude-3-sonnet
gemini-1.5-pro
claude-3-5-sonnet-v1
claude-3-5-sonnet-v2
claude-3-5-sonnet
gpt-4o
o1
gemini-2.0-flash
o3
claude-3-7-sonnet
gemini-2.5-pro
claude-sonnet-4
claude-4-sonnet
2023-01
2023-04
2023-07
2023-10
2024-01
2024-04
2024-07
2024-10
2025-01
2025-04
2025-07
Language Model Release Date

The broader field: AI automating AI Research

CORE AUTOMATION

WE'RE BUILDING THE WORLD'S MOST AUTOMATED AI LAB.

Our objective: **systems that optimize and automate work**, starting with research itself.

Recursive Superintelligence Inc.

Aksel ✓
@akseljoonas

Introducing ml-intern, the agent that just automated the post-training team @huggingface

It's an open-source implementation of the real research loop that our ML researchers do every day. You give it a prompt, it researches papers, goes through citations, implements ideas in GPU sandboxes, iterates and builds deeply research-backed models for any use case. All built on the Hugging Face ecosystem.

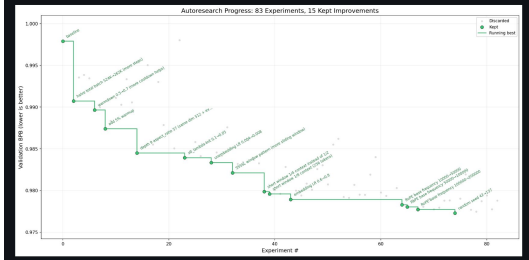
BUSINESS INSIDER

DOW ↗ +0.69% NASDAQ ↗ +1.73% S&P 500 ↗ +1.05% OIL ↗ +2.9% AAPL ↗ +2.3% NVDA ↗ +0.38% MSFT ↗ +0.26% TSLA ↘ -2.38% AMZN ↗ +1.29% META

AI

OpenAI's chief scientist says AI is getting close to being as good as a human research intern

autoresearch



One day, frontier AI research used to be done by meat computers in between eating, sleeping, having other fun, and synchronizing once in a while using sound wave interconnect in the ritual of "group meeting". That era is long gone. Research is now entirely the domain of autonomous swarms of AI agents running across compute cluster megastructures in the skies. The agents claim that we are now in the 10,205th generation of the code base, in any case no one could tell if that's right or wrong as the "code" is now a self-modifying binary that has grown beyond human comprehension. This repo is the story of how it all began. -@karpathy, March 2026.

The broader field: AI automating science

periodic labs

ACCELERATE SCIENCE

Today, we introduce Periodic Labs. Our goal is to create an AI scientist.

Science works by conjecturing how the world might be, running experiments, and learning from the results.

Intelligence is necessary, but not sufficient. New knowledge is created when ideas are found to be consistent with reality. And so, at Periodic, we are building AI scientists *and* the autonomous laboratories for them to operate.

Until now, scientific AI advances have come from models trained on the internet. But despite its vastness — it's still finite (estimates are ~10T text tokens where one English word may be 1-2 tokens). And in recent years the best frontier AI models have fully exhausted it.

Researchers seek better use of this data, but as any scientist knows: though re-reading a textbook may give new insights, they eventually need to try their idea to see if it holds.

Autonomous labs are central to our strategy. They provide huge amounts of high-quality data (each experiment can produce GBs of data!) that exists nowhere else. They generate valuable negative results which are seldom published. But most importantly, they give our AI scientists the tools to act.

We're starting in the physical sciences.

February 5, 2026 Research Publication

GPT-5 lowers the cost of cell-free protein synthesis

Working with Ginkgo Bioworks, we created an AI-driven autonomous lab and achieved a 40% reduction in protein production cost.

Read the paper ↗

Listen to article 8:42 Sha

cusp.ai

CuspAI is the frontier AI company on a mission to solve the breakthrough materials needed to power human progress.

Back to jobs

Anthropic STEM Fellow

San Francisco, CA

FutureHouse

Automating scientific discovery.

We're a non-profit building AI agents to automate research in biology and other complex sciences.

Learn More

Catherine Yeo
@catherinehyeo

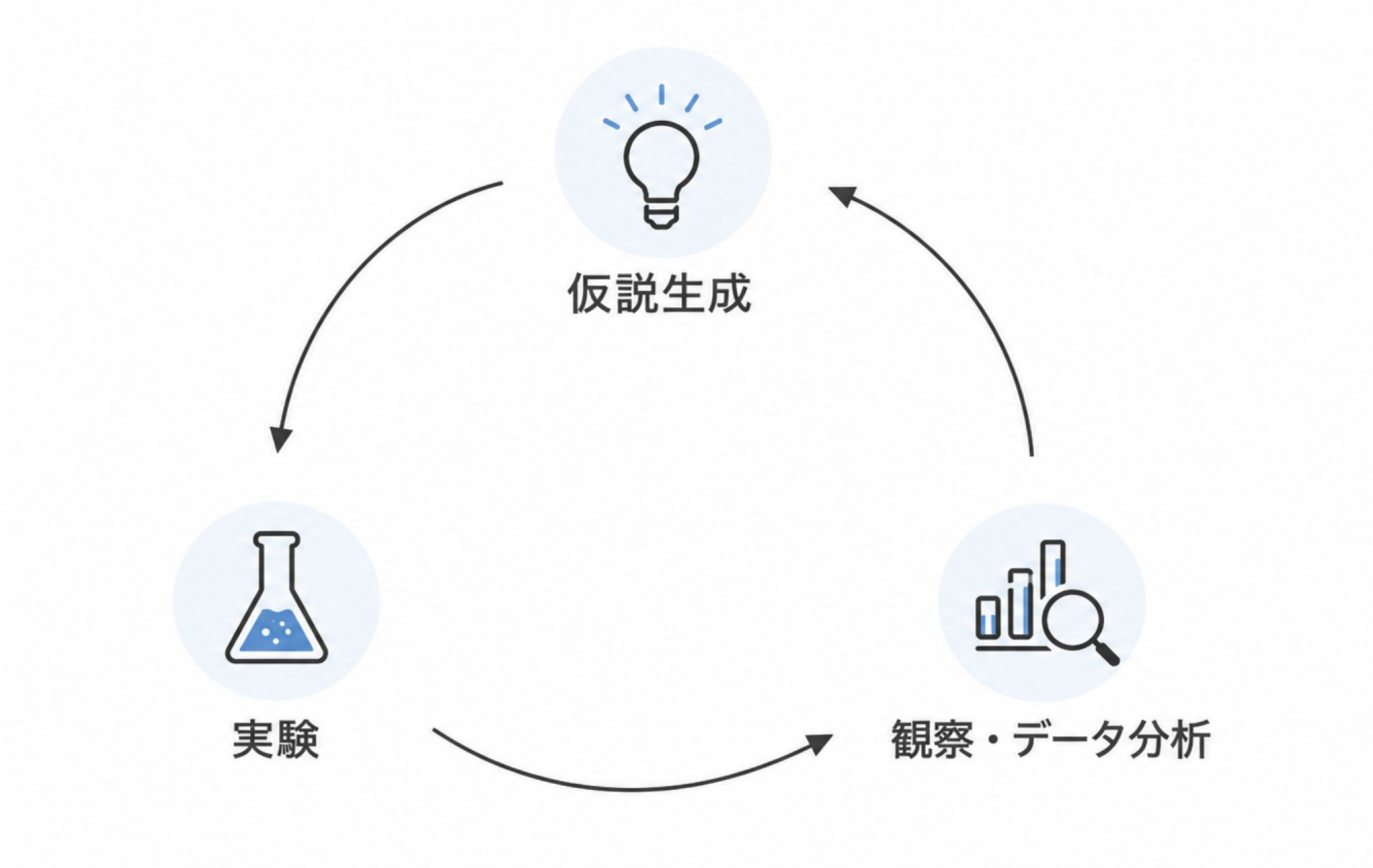
Introducing **Altara: the scientific intelligence platform for the physical world.**

Today @evatuecke and I are excited to announce our $7M seed led by @GreylockVC, joined by @Neo, @BoxGroup, @Liquid2V, and angel investors including @JeffDean and leadership from OpenAI & AMD.

We're already working with early customers in semiconductors, batteries, and advanced materials. More below.

12:19 AM · May 6, 2026 · **15.8K** Views

The verification loop of science



The broader field: everyone is attacking the loop (or parts of the loop)

Google

Accelerating scientific discovery with Co-Scientist

Received: 20 March 2025
Accepted: 11 May 2026
Accelerated Article Preview
Published online: 19 May 2026
Cite this article as: Gottweis, J. et al. Accelerating scientific discovery with Co-Scientist. Nature https://doi.org/10.1038/s41586-026-10644-y (2026)

Juraj Gottweis, Wei-Hung Weng, Alexander Daryin, Tao Tu, Petar Sirkovic, Artiom Myaskovsky, Grzegorz Glowaty, Felix Weissenberger, Alessio Orlandi, Dan Popovici, Anil Palepu, Keran Rong, Ryutaro Tanno, Khaled Saab, Fan Zhang, Jacob Blum, Andrew Carroll, Kavita Kulkarni, Nenad Tomasev, Dina Zverinski, Ivor Rendulic, Elahe Vedadi, Florian Hasler, Luka Rimanic, Marina Boia, Ivan Budiselic, Ben Feinstein, Mathias Bellaiche, Tom Sheffer, Jan Freyberg, Jeremy Ratcliff, Ottavia Bertolli, Katherine Chou, Avinatan Hassidim, Burak Ozkturk, Amin Vahdat, Yuan Guan, Vikram Dhillon, Eeshit Dhaval Vaishnav, Byron Lee, Tiago R. D. Costa, José R. Penadés, Gary Peltz, Yossi Matias, James Manyika, Demis Hassabis, Yunhan Xu, Pushmeet Kohli, Annalisa Pawlosky, Alan Karthikesalingam & Vivek Natarajan

nature

https://doi.org/10.1038/s41586-026-10658-6

仮説生成

FutureHouse

Accelerated Article Preview

A multi-agent system for automating scientific discovery

Received: 23 May 2025
Accepted: 12 May 2026
Accelerated Article Preview
Published online: 19 May 2026

Ali Essam Ghareeb, Benjamin Chang, Ludovico Mitchener, Angela Yiu, Caralyn J. Szostkiewicz, Dmytro Sivved, Gavin J. Gyimesi, Jon M. Laurent, Samantha M. Wright, Muhammed T. Razzak, Andrew D. White, Silvia C. Finnemann, Michaela M. Hinks & Samuel G. Rodriques

nature

https://doi.org/10.1038/s41586-026-10644-y

GPT-5 lowers the cost of cell-free protein synthesis

Working with Ginkgo Bioworks, we created an AI-driven autonomous lab and achieved a 40% reduction in protein production cost.

periodic labs

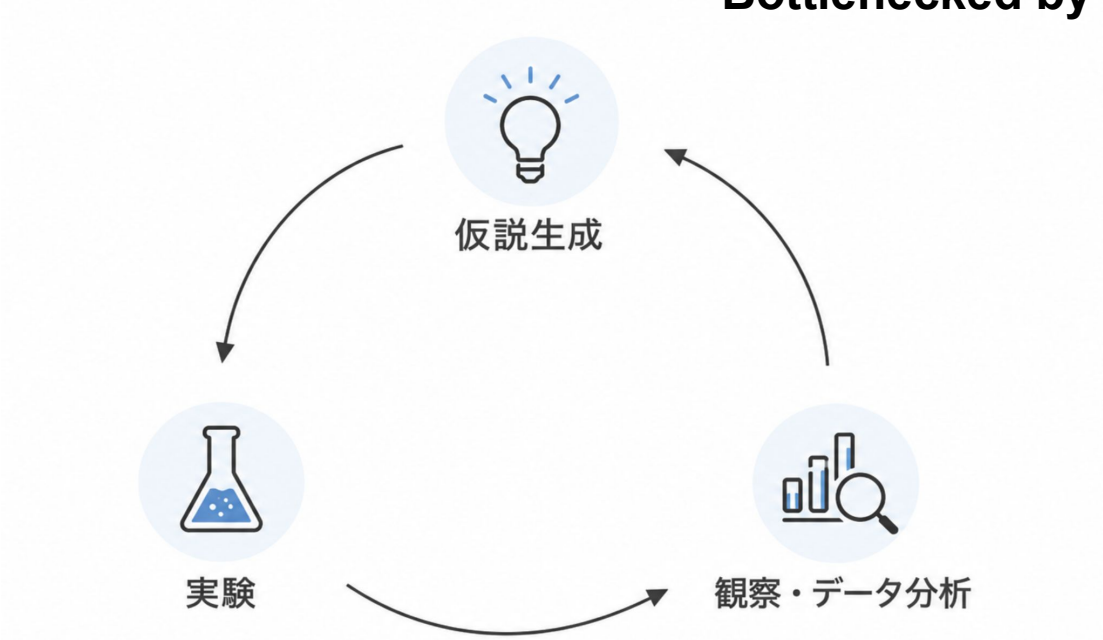
実験

観察・データ分析

Kosmos: An AI Scientist for Autonomous Discovery

Where Science Is Bottlenecked

Bottlenecked by intelligence

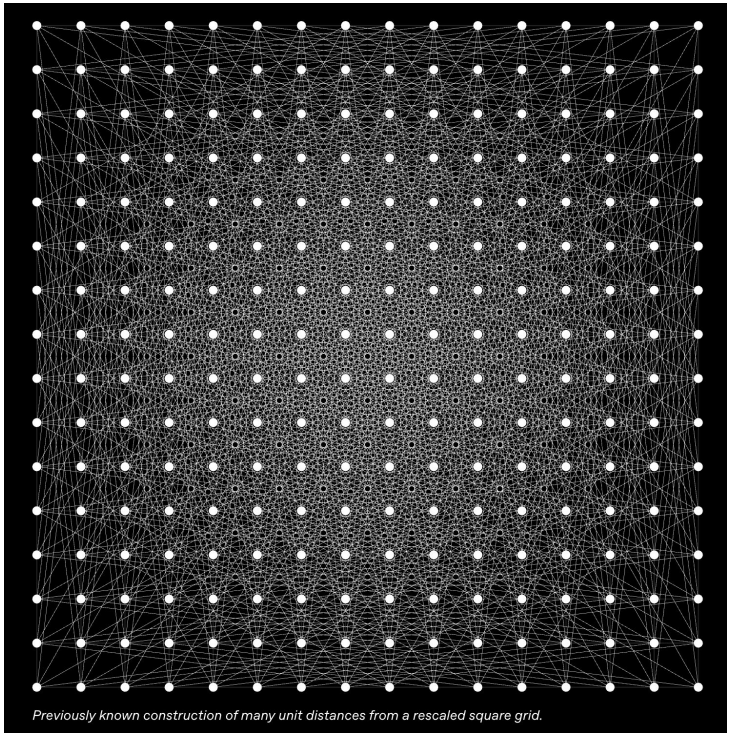


Bottlenecked by cost, time, and poor scalability

May 20, 2026 Research Milestone

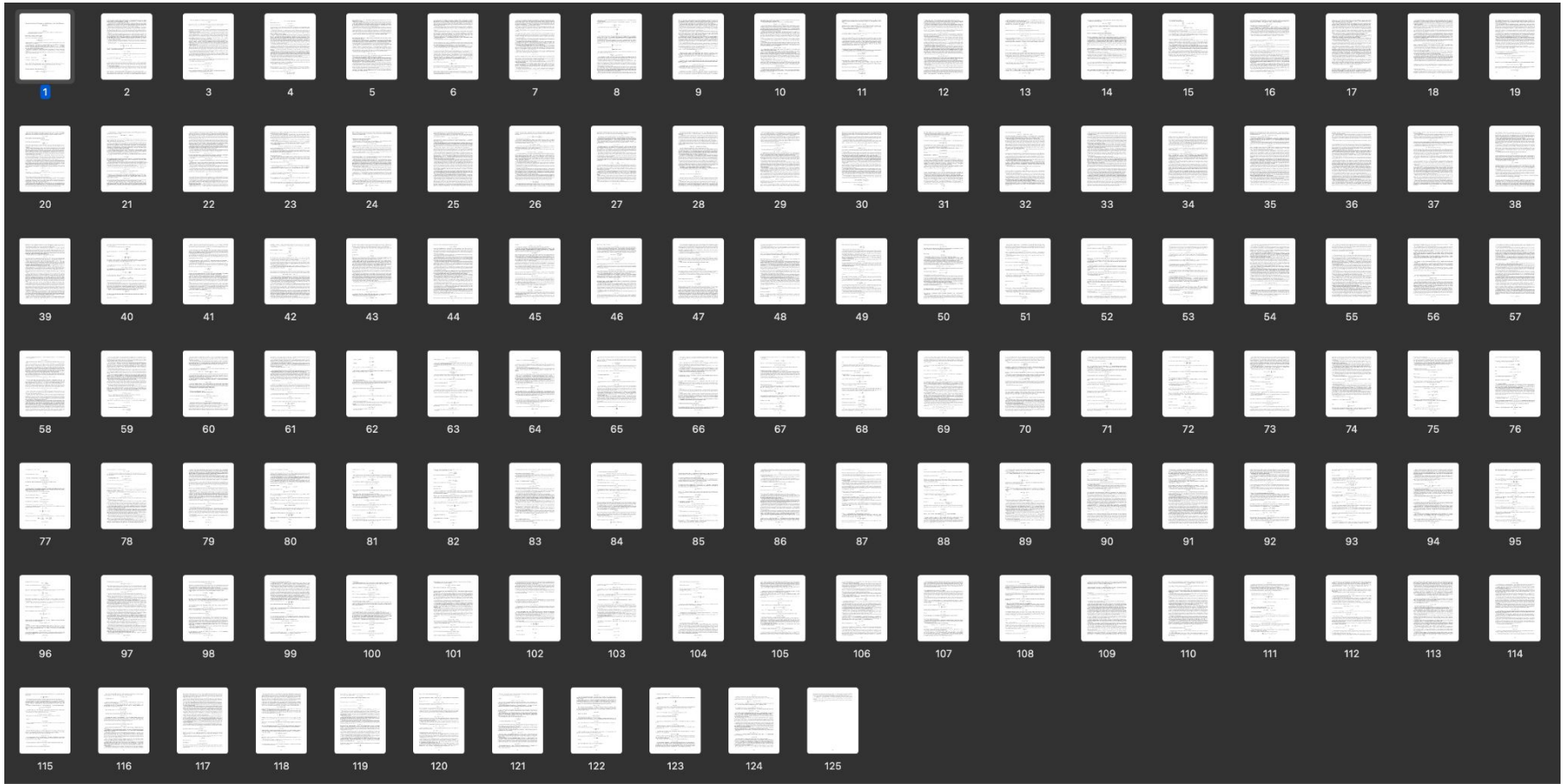
An OpenAI model has disproved a central conjecture in discrete geometry

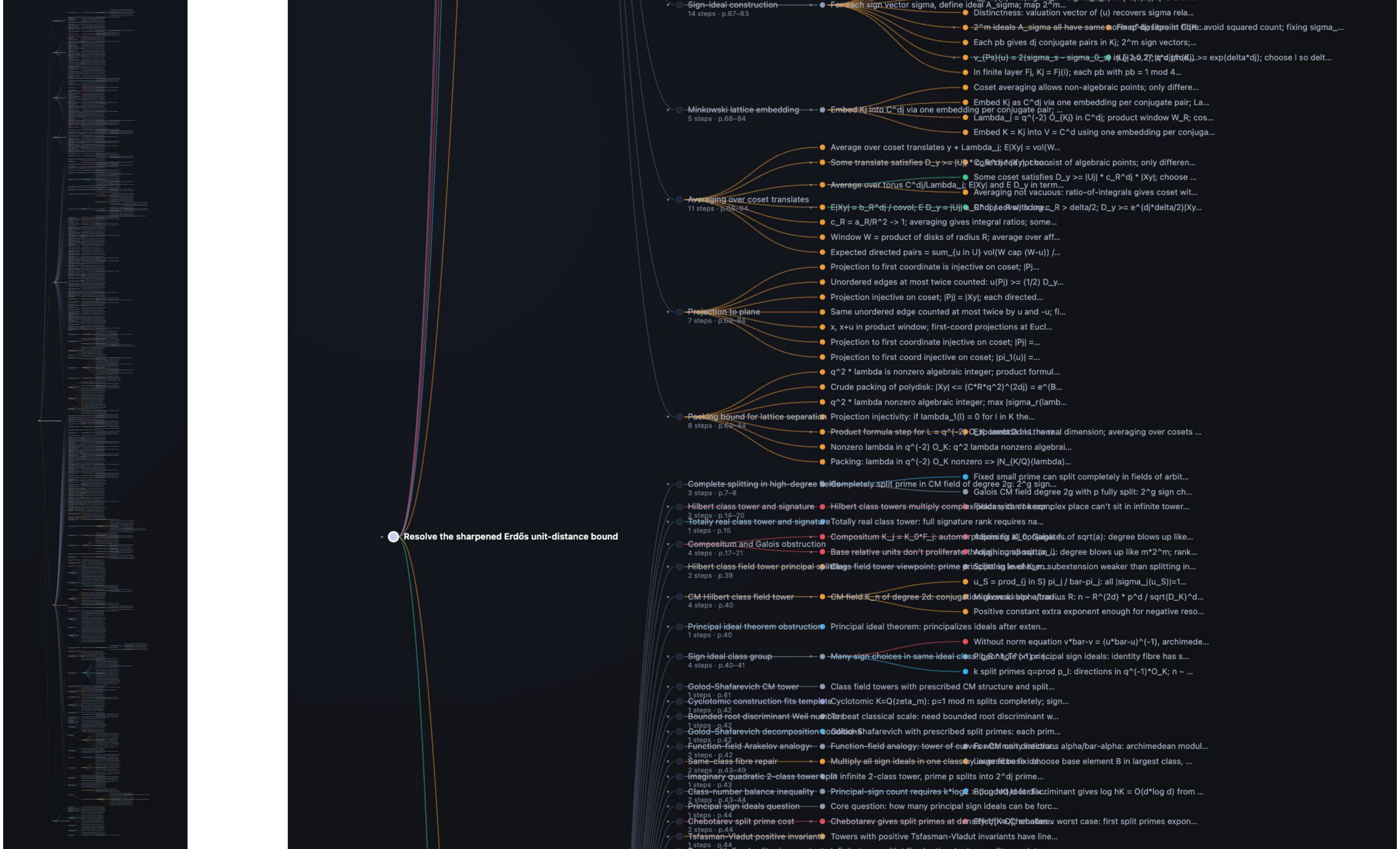
平面上に n 個の点を置いたとき、距離がちょうど　1 になる点のペアを最大でいくつ作れるか？



Previously known construction of many unit distances from a rescaled square grid.

Rewritten Chain of Thought for the Solution to the Unit Distance Problem




Sign-ideal construction
14 steps · p.67–83
Distinctness: valuation vector of (u) recovers sigma rela...
Each pb gives dj conjugate pairs in Kj; 2^m sign vectors;...
In finite layer Fj, Kj = Fj(i); each pb with pb = 1 mod 4...
Coset averaging allows non-algebraic points; only differe...
Embed Kj as C^dj via one embedding per conjugate pair; La...
Minkowski lattice embedding
5 steps · p.68–84
Lambda_j = q^(-2) O_{Kj} in C^dj; product window W_R; cos...
Embed K = Kj into V = C^d using one embedding per conjuga...
Average over coset translates y + Lambda_j; E|Xy| = vol(W...
Some coset satisfies D_y >= |Uj| * c_R^dj * |Xy|; choose ...
Averaging not vacuous: ratio-of-integrals gives coset wit...
Averaging over coset translates
11 steps · p.68–84
c_R = a_R/R^2 -> 1; averaging gives integral ratios; some...
Window W = product of disks of radius R; average over aff...
Expected directed pairs = sum_{u in U} vol(W cap (W-u)) /...
Projection to first coordinate is injective on coset; |Pj...
Unordered edges at most twice counted: u(Pj)) >= (1/2) D_y...
Projection injective on coset; |Pj| = |Xy|; each directed...
Projection to plane
7 steps · p.69–84
Same unordered edge counted at most twice by u and -u; fi...
x, x+u in product window; first-coord projections at Eucl...
Projection to first coordinate injective on coset; |Pj| =...
Projection to first coord injective on coset; |pi_1(u)| =...
q^2 * lambda is nonzero algebraic integer; product formul...
Crude packing of polydisk: |Xy| <= (C*R*q^2)^(2dj) = e^(B...
q^2 * lambda nonzero algebraic integer; max |sigma_r(lamb...
Packing bound for lattice separation
8 steps · p.69–84
Projection injectivity: if lambda_1(l) = 0 for l in K the...
Nonzero lambda in q^(-2) O_K: q^2 lambda nonzero algebrai...
Packing: lambda in q^(-2) O_K nonzero => |N_{K/Q}(lambda)...
Fixed small prime can split completely in fields of arbit...
Complete splitting in high-degree fields
3 steps · p.7–8
Galois CM field degree 2g with p fully split: 2^g sign ch...
Hilbert class tower and signature
Totally real class tower and signature
1 steps · p.15
Totally real class tower: full signature rank requires na...
Resolve the sharpened Erdős unit-distance bound
Compositum and Galois obstruction
4 steps · p.17–21
Hilbert class field tower principal splitting
2 steps · p.39
u_S = prod_{j in S} pi_j / bar-pi_j: all |sigma_j(u_S)|=1...
CM Hilbert class field tower
4 steps · p.40
Positive constant extra exponent enough for negative reso...
Principal ideal theorem obstruction
1 steps · p.40
Principal ideal theorem: principalizes ideals after exten...
Without norm equation v*bar-v = (u*bar-u)^(-1), archimede...
Sign ideal class group
4 steps · p.40–41
k split primes q=prod p_l: directions in q^(-1)*O_K; n ~ ...
Golod-Shafarevich CM tower
Class field towers with prescribed CM structure and split...
Cyclotomic K=Q(zeta_m): p=1 mod m splits completely; sign...
To beat classical scale: need bounded root discriminant w...
Golod-Shafarevich with prescribed split primes: each prim...
Same-class fibre repair
Imaginary quadratic 2-class tower split
In infinite 2-class tower, prime p splits into 2^dj prime...
Class-number balance inequality
Principal sign ideals question
Core question: how many principal sign ideals can be forc...
Chebotarev split prime cost
Tsfasman-Vladut positive invariant
Towers with positive Tsfasman-Vladut invariants have line...

メタアナリシス：全体構造

706 reasoning steps · 7 attack strategies · 280 sub-approaches · 125-page trace

`benchmark` (9), `strategy` (39), `hypothesis` (104), `test` (361), `refine` (30), `insight` (28), `prune` (58), `conclusion` (77).

280 sub-approaches and **39 idea-line openers** (`strategy`-type nodes).

メタアナリシス : Reasoning traceの長さ

Most are shallow try-and-discard. At the sub-approach level: 57% (160/280) are a *single* step, and 81% (227/280) are ≤3 steps.

Only 10% contain any `refine` step at all, and only **18 of 280 (6%) show a genuine refine-loop** — i.e., test → refine/re-hypothesize → test again.

The genuine iterative chains are a small handful and they're exactly where the breakthrough came from.

"Almost-totally-real (ATR) quadratic extension," steps 82–131 — 20 steps, with 10 tests, 2 explicit refines, and 4 refine-loops.

メタアナリシス：testノードは何をしている？

"Test" steps are the *work of evaluating an idea*
E.g. computing a bound, running the arithmetic, applying a known theorem to see what it yields, estimating a size (a regulator, discriminant, edge or point count)

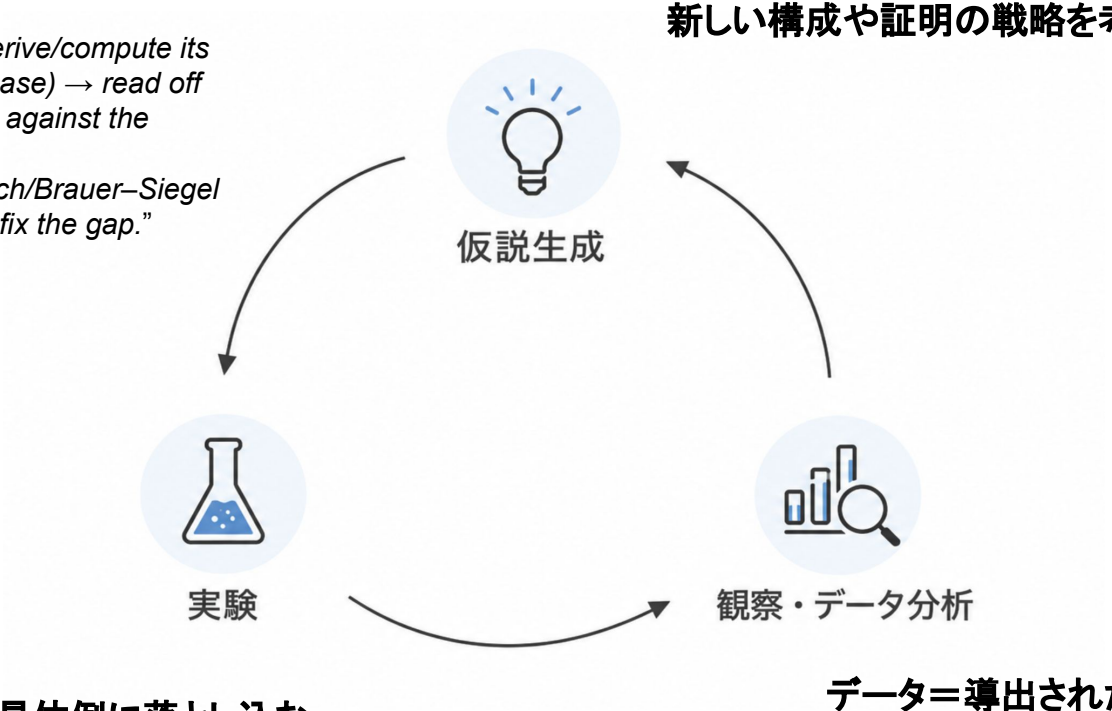
Real examples from the data:

- *"Grid counting: $r_2(k) = 4 \cdot 2^t$ for squarefree k…"* — working out the lattice degree.
- *"Compare exp(d·log log d) to Erdős allowance exp(C·d): dangerous gap…"* — the size comparison that flags the number-field route.
- *"Padding makes comparison harder; disjoint-union analysis confirms no amplification"* — checking an amplification trick and finding it fails.
- *"Standard averaging: $E|X_y| = vol(\Omega)/covol(L)$…"* — the expected-overlap computation in the final construction.

メタアナリシス：数学にも実験 →解析のループ

“Propose a construction → derive/compute its consequences (or test a toy case) → read off the resulting scale → check it against the benchmark and against ST/Dirichlet/Golod–Shafarevich/Brauer–Siegel → narrow the construction to fix the gap.”

新しい構成や証明の戦略を考える

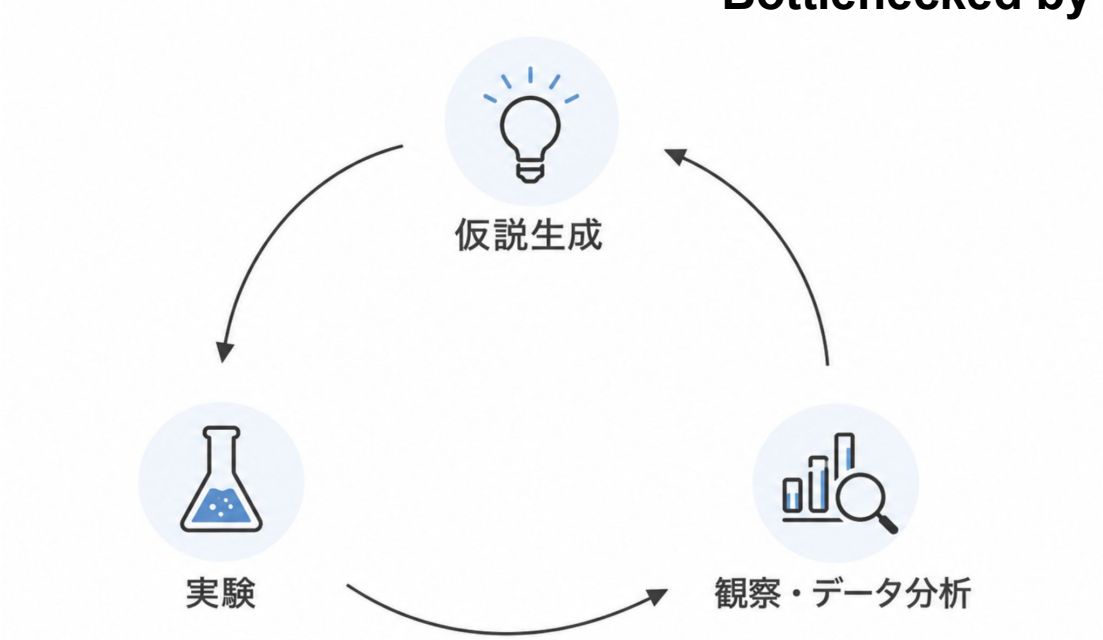


長く複雑な計算をする、具体例に落とし込む

データ＝導出された不等式など
測定機器＝既存の定理

Where Mathematical Science Is Bottlenecked

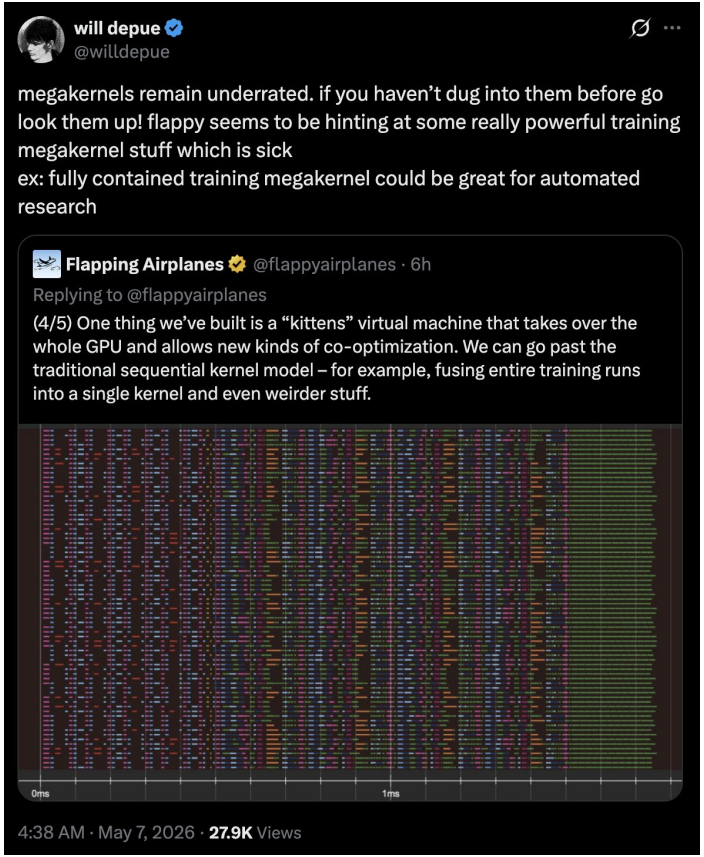
Bottlenecked by intelligence



Bottlenecked by cost, time, and poor scalability?

Bottlenecked by cost, time, and poor scalability

Megakernel https://x.com/willdepue/status/2052110690526409207



https://hazyresearch.stanford.edu/blog/2025-05-27-no-bubbles

NanoGPT Speedrun

20 min per experiment

Short-horizon vs. long-horizon tasks

Short-horizon tasks: easier for agents because feedback is fast and measurable: code passes tests, loss goes down, benchmark score improves, runtime decreases.

Long-horizon scientific tasks: harder because the feedback is delayed, noisy, and often conceptual: Is the problem important? Is the hypothesis meaningful? Are the baselines fair? Does the experiment support the claim? Is the result robust?

mean performance vs. rare discoveries

For discovery systems, we may care less about E[mean performance] and more about E[max discovery under a compute budget].

Learning to Discover at Test Time

Mert Yuksekgonul[*1], Daniel Koceja[*1], Xinhao Li[*4], Federico Bianchi[*5]
Jed McCaleb[3], Xiaolong Wang[4], Jan Kautz[2], Yejin Choi[2], James Zou[†1,5], Carlos Guestrin[†1], Yu Sun[*1,2]
[1] Stanford University [2] NVIDIA [3] Astera Institute [4] UC San Diego [5] Together AI

Abstract

How can we use AI to discover a new state of the art for a scientific problem? Prior work in test-time scaling, such as AlphaEvolve, performs search by prompting a frozen LLM. We perform reinforcement learning at test time, so the LLM can continue to train, but now with experience specific to the test problem. This form of continual learning is quite special, because its goal is to produce one great solution rather than many good ones on average, and to solve this very problem rather than generalize to other problems. Therefore, our learning objective and search subroutine are designed to prioritize the most promising solutions. We call this method Test-Time Training to Discover (TTT-Discover). Following prior work, we focus on problems with continuous rewards.

We report results for every problem we attempted, across mathematics, GPU kernel engineering, algorithm design, and biology. TTT-Discover sets the new state of the art in almost all of them: (i) Erdős' minimum overlap problem and an autocorrelation inequality; (ii) a GPUMode kernel competition (up to 2× faster than prior art); (iii) past AtCoder algorithm competitions; and (iv) denoising problem in single-cell analysis. Our solutions are reviewed by experts or the organizers.

All our results are achieved with an open model, OpenAI gpt-oss-120b, and can be reproduced with our publicly available code, in contrast to previous best results that required closed frontier models. Our test-time training runs are performed using Tinker, an API by Thinking Machines, with a cost of only a few hundred dollars per problem.

Can ideation be separated from execution?

Is hill-climbing bad?

Is designing hill-climbable problems a new skill?

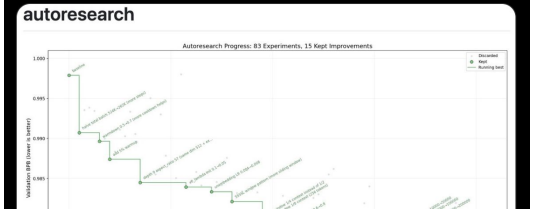
Andrej Karpathy ✓
@karpathy

I packaged up the "autoresearch" project into a new self-contained minimal repo if people would like to play over the weekend. It's basically nanochat LLM training core stripped down to a single-GPU, one file version of ~630 lines of code, then:

- the human iterates on the prompt (.md)
- the AI agent iterates on the training code (.py)

The goal is to engineer your agents to make the fastest research progress indefinitely and without any of your own involvement. In the image, every dot is a complete LLM training run that lasts exactly 5 minutes. The agent works in an autonomous loop on a git feature branch and accumulates git commits to the training script as it finds better settings (of lower validation loss by the end) of the neural network architecture, the optimizer, all the hyperparameters, etc. You can imagine comparing the research progress of different prompts, different agents, etc.

github.com/karpathy/autor...
Part code, part sci-fi, and a pinch of psychosis :)



Neel Guha
@NeelGuha

I wrote a blogpost about writing machine learning research papers (e.g., NeurIPS, ICML, ICLR, etc.). The core idea is that most papers follow one of a predetermined set of templates. The post talks about each template, describes their rules, and offers examples...

The "Data Artifact Paper"

The "Horse Race Paper"

The "New Paradigm Paper"

The "Resurrected Baseline Paper"

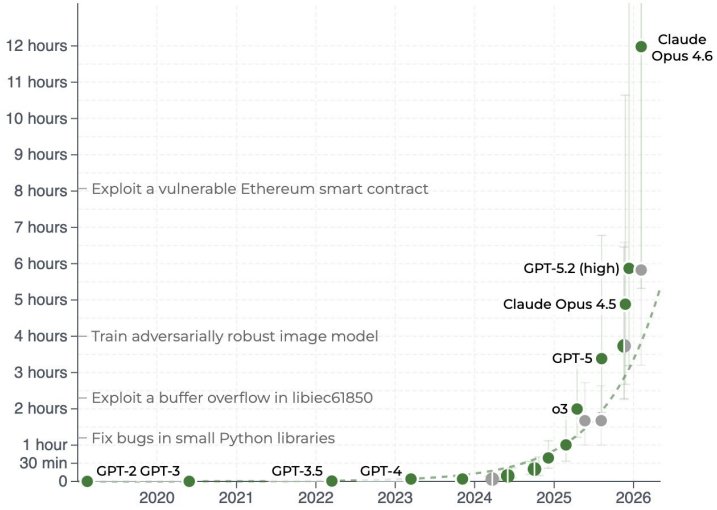
The "Unification Paper"

The "Problem Solving Paper"

The "Discovery Paper"

The "Countervailing Wisdom Paper"

6:27 AM · Mar 25, 2026 · **78.5K** Views



Thank you!


Y. Yamada


R.T. Lange


Cong Lu


S. Hu


Chris Lu


J. Foerster


J. Clune †


D. Ha †

sakana.ai

FLAIR

UBC

† Equal advising

⇒ Contact: yutaroyamada@sakana.ai
⇒ X/Twitter: @_yutaroyamada